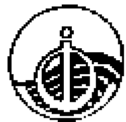


REGIONAL SEAS

UNITED NATIONS ENVIRONMENT PROGRAMME JUNE 1995

Statistical analysis and interpretation of marine community data

Reference Methods For Marine Pollution Studies No. 64

Prepared in co-operation with

FAO

IMO

UNEP 1995

NOTE: This document has been prepared by the Plymouth Marine Laboratory, U.K., in collaboration with the Intergovernmental Oceanographic Commission (IOC), the Food and Agricultural Organization (FAO), the International Atomic Energy Agency, Marine Environment Laboratory (IAEA-MEL), and the United Nations Environment Programme (UNEP) under the project FP/ME/5101-93-03(3033).

For bibliographic purposes this document may be cited as:

UNEP/IOC/FAO/IMO/IAEA: Statistical analysis and interpretation of marine community data. Reference Methods for Marine Pollution Studies No. 64, UNEP 1995.

INTRODUCTION

Aim

This manual describes a strategy for the statistical analysis and interpretation of biological data on community structure, consisting of abundance or biomass readings for a set of species and a number of samples. The latter usually consist of one or more replicates taken:

a) at a number of sites at one time (spatial analysis),

b) at the same site at a number of times (temporal analysis),

c) for a community subject to different manipulative "treatments" (laboratory or field experiments),

or some combination of these. The species-by-samples arrays are typically large, and patterns in community structure are often not readily apparent. Statistical analysis therefore centres around reducing the complexity of these matrices, usually by some graphical representation of the biological relationships between the samples. This is followed by statistical testing to identify and characterise changes in community structure in time or space and relate these to changing environmental or experimental conditions.

Emphasis

Of principle concern are the biological effects of contaminants, though, since the same analysis techniques are appropriate to wider studies of community structure, a number of examples are included which are not pollution-related. In general these illustrate some important aspect of the methodology which is applicable to pollution studies. The scope of the examples is specifically marine (though the techniques have wider application) and, though the examples range over different community types (benthic infauna, corals, plankton, fish etc.), there is a bias towards soft-sediment benthos, reflecting both the authors' own research interests and the widespread use of such community data in pollution monitoring.

Scope

There is a vast array of sophisticated statistical techniques for handling species-by-samples matrices, ranging from their reduction to simple diversity indices, through curvilinear or distributional representations of richness, dominance, evenness etc., to a plethora of multivariate approaches involving clustering or ordination methods. This manual does *not* attempt to give an overview of all the options, or even the majority of them. Instead it presents a strategy which has evolved over several years, within the Community Ecology group at Plymouth Marine Laboratory, and which has a proven track record in published analysis and interpretation of a wide range of marine community data. The manual attempts to explain *how* and *why* the selected techniques work, to a level of understanding sufficient to appreciate when they are (and are not) applicable, and to interpret their outcome. It is aimed at ecologists with no more than an introductory background in statistics, who need to apply these statistical techniques to answer specific questions about changes in community structure.

This volume is also not a software manual, describing how to use a particular computer program or package to carry out the analyses discussed here, though the advocated approach is mirrored in the software package PRIMER (Plymouth Routines In Multivariate Ecological Research), developed at the Plymouth Marine Laboratory and available commercially. Footnotes in the text make brief reference to the PRIMER modules which have been used to obtain the analyses presented. The PRIMER package has been used throughout (though the figures have in many cases been subjected to further annotation etc. using a presentation graphics program, Harvard Graphics). Note, however, that PRIMER is not the only option for computation. The major statistical packages such as SAS, BMDP, SPSS, GENSTAT, etc. have always included multivariate options, as do some PC packages such as SYSTAT, STATGRAPHICS etc. In addition, more specialised software, such as CLUSTAN, the Cornell Ecology programs, KYST, CANOCO, PATN etc., is fairly widespread. None of these packages will offer precisely the combination of options discussed here, and some take a rather different (but equally valid) approach to the problems posed. An over-riding thrust of the current exposition is, however, to retain as great a simplicity of explanation and transparency of interpretation as is possible in what, conventionally, has been regarded as a difficult area for practitioners lacking a strong statistical background.

PREFACE

The Regional Seas Programme was initiated by UNEP in 1974. Since then the Governing Council of UNEP has repeatedly endorsed a regional approach to the control of marine pollution and the management of marine and coastal resources and has requested the development of regional action plans. The Regional Seas Programme at present includes 12 regions and has some 140 coastal States participating in it (1), (2).

One of the basic components of the action plans sponsored by UNEP in the framework of the Regional Seas Programme is the assessment of the state of the marine environment and of its resources and of the sources and trends of the pollution, and the impact of pollution on human health, marine ecosystems and amenities. In order to assist those participating in this activity and to ensure that the data obtained through this assessment can be compared on a world-wide basis and thus contribute to the Global Environment Monitoring System (GEMS) of UNEP, a set of Reference Methods and Guidelines for marine pollution studies is being developed as part of a programme of comprehensive technical support which includes the provision of expert advice, reference methods and material, training and data quality assurance (3). The methods are recommended to be adopted by Governments participating in the Regional Seas Programme.

The methods and guidelines are prepared in co-operation with the relevant specialized bodies of the United Nations systems as well as other organizations and are tested by a number of experts competent in the field relevant to the methods described.

In the description of the methods and guidelines the style used by the International Organization for Standardization (ISO) is followed as closely as possible.

The methods and guidelines, as published in UNEP's series of Reference Methods for Marine Pollution Studies, are not considered as final. They are planned to be periodically revised taking into account the development of our understanding of the problems, of analytical instrumentation and the actual need of the users. In order to facilitate these revisions the users are invited to convey their comments and suggestions to:

Marine Environmental Studies Laboratory
IAEA Marine Environment Laboratory
B.P. No. 800
MC-98012 MONACO Cedex

which is responsible for the technical co-ordination of the development, testing and intercalibration of Reference Methods.

---

(1) UNEP: Achievements and planned development of the UNEP's Regional Seas Programme and comparable programmes sponsored by other bodies. UNEP Regional Seas Reports and Studies No. 1, UNEP, 1982.

(2) P. HULM: A strategy for the Seas. The Regional Seas Programme: Past and Future, UNEP 1983.

(3) UNEP/IAEA/IOC: Reference Methods and Materials: A Programme of comprehensive support for regional and global marine pollution assessments. UNEP, 1990.

CONTENTS

PREFACE

ACKNOWLEDGEMENTS

INTRODUCTION

CHAPTER 1:	A framework for studying changes in community structure
CHAPTER 2:	Measures of similarity of species abundance/biomass between samples
CHAPTER 3:	Hierarchical clustering
CHAPTER 4:	Ordination of samples by Principal Components Analysis (PCA)
CHAPTER 5:	Ordination of samples by Multi-Dimensional Scaling (MDS)
CHAPTER 6:	Testing for differences between groups of samples
CHAPTER 7:	Species analyses
CHAPTER 8:	Diversity measures, dominance curves and other graphical analyses
CHAPTER 9:	Transformations
CHAPTER 10:	Species removal and aggregation
CHAPTER 11:	Linking community analyses to environmental variables
CHAPTER 12:	Causality: community experiments in the field and laboratory
CHAPTER 13:	Data requirements for biological effects studies: which components and attributes of the biota to examine?
CHAPTER 14:	Relative sensitivities and merits of univariate, graphical/distributional and multivariate techniques
CHAPTER 15	Multivariate measures of community stress
APPENDIX 1	Index of example data
APPENDIX 2	Principal literature sources and further reading
APPENDIX 3	References cited

CHAPTER 1: A FRAMEWORK FOR STUDYING CHANGES IN COMMUNITY STRUCTURE

The purpose of this opening chapter is twofold:

a) to introduce a few of the data sets which are used most extensively, as illustrations of techniques, throughout the manual;

b) to outline a framework for the various possible stages in a community analysis[1].

Examples are given of some core elements of the recommended approaches, foreshadowing the analyses explained in detail later and explicitly referring forward to the relevant chapters. Though at this stage the details are likely to remain mystifying, the intention is that this opening chapter should give the reader some feel for where the various techniques are leading and how they slot together. As such, it may serve both as an introduction and a summary.

Stages

It is convenient to categorise possible analyses broadly into four main stages.

1) *Representing communities* by graphical description of the relationships between the biota in the various samples. This is thought of as "pure" description, rather than explanation or testing, and the emphasis is on reducing the complexity of the multivariate information in typical species/samples matrices, to obtain some form of low–dimensional picture of how the biological samples interrelate.

2) *Discriminating sites/conditions* on the basis of their biotic composition. The paradigm here is that of the hypothesis test, examining whether there are "proven" community differences between groups of samples identified *a priori*, for example demonstrating differences between control and putatively impacted sites, establishing before/after impact differences at a single site, etc.

3) *Determining levels of "stress"* or disturbance, by attempting to construct biological measures from the community data which are indicative of disturbed conditions. These may be absolute measures ("this observed structural feature is indicative of pollution") or relative criteria ("under impact, this coefficient is expected to decrease in comparison with control levels"). Note the contrast with the previous stage, however, which is restricted to demonstrating differences between groups of samples, not ascribing directionality to the change (e.g. deleterious consequence).

4) *Linking to environmental variables* and examining issues of *causality* of any changes. Having allowed the biological information to "tell its own story", any associated physical or chemical variables, matched to the same set of samples, can be examined for their own structure and its relation to the biotic pattern (its "explanatory power"). The extent to which identified environmental differences are actually *causal* to observed community changes can only really be determined by manipulative experiments, either in the field or through laboratory/mesocosm studies.

Techniques

The spread of methods for extracting workable representations and summaries of the biological data can be grouped into three categories.

1) *Univariate methods* collapse the full set of species counts for a sample into a single coefficient, for example a *diversity index*. This might be some measure of the numbers of different species for a fixed number of individuals (species richness) or the extent to which the community counts are dominated by a small number of species (dominance/evenness index), or some combination of these. Clearly, the *a priori* selection of a single taxon as an *indicator species*, amenable to specific inferences about its response to a particular environmental gradient, also gives rise to a univariate analysis.

2) *Distributional techniques*, also termed graphical or curvilinear plots (when they are not strictly distributional), are a class of methods which summarise the set of species counts for a single sample by a curve or histogram. One example is *k–dominance curves* (Lambshead *et al.*, 1983), which rank the species in decreasing order of abundance, convert the values to percentage abundance relative to the total number of

---

1. *The term **community** is used throughout the manual, somewhat loosely, to refer to any assemblage data (samples leading to counts or biomass for a range of species); the usage does not necessarily imply indigenous structuring of the biota, for example by competitive interactions.*

be condensed into one (or a small number of) key summary statistics. Simple (or multiple) regression of Shannon diversity as the dependent variable, against the environmental descriptors as independent variables, is then technically feasible, though in practice rarely very informative given the over–condensed nature of the information utilised.

For impact studies, much has been written about the effect of pollution or disturbance on diversity measures: whilst the response is not necessarily unidirectional (under the hypothesis of Huston, 1979, diversity is expected to rise at intermediate disturbance levels before its strong decline with gross disturbance), there is a sense in which *determining stress levels* is possible, through relation to historical diversity patterns for particular environmental gradients. Similarly, empirical evidence may exist that particular indicator taxa (e.g. Capitellids) change in abundance along specific pollution gradients (e.g. of organic enrichment). Note though that, unlike the diversity measures constructed from abundances across species, averaged in some way[3], indicator species levels or the number of species in a sample (S) may not initially satisfy the assumptions necessary for classical statistical analysis. For the number of species, S, the normality and constant variance conditions can usually be produced by transformation of the variable (e.g. log S). However, for most individual species, abundance across the set of samples is likely to be a very poorly–behaved variable, statistically speaking. Typically, a species will be absent from many of the samples and, when it is present, the counts are often highly variable, with an abundance probability distribution which is heavily right–skewed[4]. Thus, for all but the most common individual species, transformation is no real help and parametric statistical analyses *cannot* be applied to the counts, in any form. In any case, it is not valid to "snoop" in a large data matrix, of typically 100–250 taxa, for one or more "interesting" species to analyse by univariate techniques (any indicator or keystone species selection must be done *a priori*). Such arguments lead to the tenets underlying this manual:

a) community data is inherently multivariate (highly so) and usually needs to be analysed *en masse* in order to elicit the important biological structure and its relation to the environment;

b) standard parametric modelling is totally invalid.

Thus throughout, rather little emphasis is given to representing communities by univariate measures, though some possibilities for construction can be found at the start of Chapter 8, some brief remarks on hypothesis testing (ANOVA) at the start of Chapter 6, a discussion of transformations (to approximate normality and constant variance) at the start of Chapter 9, and an example given of a univariate regression between biota and environment in Chapter 11. Finally, Chapter 14 gives a series of detailed comparisons of univariate with distributional and multivariate techniques, in order to gauge their relative sensitivities and merits in a range of practical studies.

EXAMPLE: Frierfjord macrofauna

The first example is from the IOC/GEEP practical workshop on biological effects of pollutants (Bayne *et al.*, 1988), held at the University of Oslo, August 1986. This attempted to contrast a range of biochemical, cellular, physiological and community analyses, applied to field samples from potentially contaminated and control sites, in a fjordic complex (Frierfjord/Langesundfjord) linked to Oslofjord (*F*, Fig. 1.1). For the benthic macrofaunal component of this study (Gray *et al.*, 1988), four replicate 0.1 m^2 Day grab samples were taken at each of six sites (A–E and G, Fig. 1.1) and, for each sample, organisms retained on a 1.0 mm sieve were identified and counted. Wet weights were determined for each species in each sample, by pooling individuals within species.

Part of the resulting data matrix can be seen in Table 1.2: in total there were 110 different taxa categorised from the 24 samples. Such matrices (abundance, *A*, and biomass, *B*) are the starting point for *all* the analyses of this manual, and this example is typical in respect of the relatively high ratio of species to samples (always >> 1) and the prevalence of zeros. Here, as elsewhere, even an undesirable reduction to the 30 "most important" species (see Chapter 2) leaves more than 50% of the matrix consisting of zeros. Standard multivariate normal analyses (e.g. Mardia *et al.*, 1979) of these counts are clearly ruled out; they require both

3. *And thus subject to the central limit theorem, which will tend to induce statistical normality.*

4. *It is the authors' experience, certainly in the study of benthic communities, that the individuals of a species are not distributed at random in space (a Poisson process) but are often highly clustered, either through local variation in forcing environmental variables or mechanisms of recruitment, mortality and community interactions. This leads to counts which, in statistical terms, are described as over–dispersed, combined with a high prevalence of zeros, causing major problems in attempting parametric modelling by categorical/log–linear methods.*

Table 1.3. Distributional techniques. Summary of analyses for the four stages.

Stages	Distributional examples	
	ABC (k-dominance) curves (Ch 8)	Species abundance distributions (Ch 8)
1) Representing communities	*Curves for each site/condition (or preferably replicate)*	
2) Discriminating sites/conditions	*ANOVA on univariate summaries (e.g. W, Ch 8), or:* *ANOSIM test (Ch 6) on "distances" between every pair of curves*	*Test for commonality of distributions (e.g. chi-squared), if valid*
3) Determining stress levels	*Biomass curve drops below numbers curve under disturbance*	*Species abundance distribution has "longer tail" with disturbance*
4) Linking to environment	*Difficult, except for univariate summaries of the curves (by regression) (Causality: see Ch 12)*	

DISTRIBUTIONAL TECHNIQUES

A less condensed form of summary of each sample is offered by the distributional/graphical methods, outlined for the four stages in Table 1.3.

Representation is by curves or histograms (Chapter 8), either plotted for each replicate sample separately or for pooled data within sites or conditions. The former permits a visual judgement of the sampling variation in the curves and, as with diversity indices, replication is required to ***discriminate sites***, i.e. test the null hypothesis that two or more sites (/conditions etc.) have the same curvilinear structure. The easiest approach to testing is then to summarise each replicate curve by a single statistic and apply ANOVA as before: for the ABC method, mentioned earlier, the *W* statistic (Chapter 8) is a convenient measure of the extent to which the biomass curve "dominates" the abundance curve, or vice-versa. This is effective in practice though, in theory, it simply amounts to computing another diversity index and is therefore just a univariate approach. A more general test, which honours the curvilinear structure, could be constructed by the ANOSIM procedure (described later under multivariate techniques), computed between every pair of replicate ABC curves.[6]

The distributional/graphical techniques have been proposed specifically as a way of ***determining stress levels***. For the ABC method, the strongly polluted (/disturbed) state is indicated if the abundance *k*-dominance curve falls above the biomass curve throughout its length (e.g. see the later plots in Fig. 1.4): the phenomenon is linked to the loss of large-bodied "climax" species and the rise of small-bodied opportunists. Note that the ABC procedure claims to give an *absolute* measure, in the sense that disturbance status is attributable on the basis of samples from a single site; in practice however it is always wise to design collection from (matched) impacted and control sites to confirm that the control condition exhibits the "undisturbed" ABC pattern (biomass curve above the abundance curve, throughout). Similarly, the species abundance distribution has features characteristic of disturbed status (e.g. see the middle plots in Fig. 1.6), namely a move to a less "J-shaped" distribution by a reduction in the first one or two abundance classes (loss of rarer species), combined with the gain of some higher abundance classes (very numerous opportunist species).

The distributional/graphical methods may thus have particular merits in allowing recognition of "stressed" states (Chapter 14), though they have the disadvantage of being more difficult to work with statistically.

6. This is somewhat esoteric and is not pursued in this manual; for details see Clarke (1990). Similarly outside the current scope are tests of equality for two or more observed histograms arising from species abundance distributions. Again, the most straightforward approach to testing is probably to summarise each distribution by two or three measures (of location, spread, skewness etc.) and carry out ANOVA on the summary statistics for each replicate. Another possibility is a chi-squared test (or some form of Cramer-von Mises approach), for testing equality of two or more frequency distributions, but this is unlikely to be valid given the species interdependencies in a single sample.

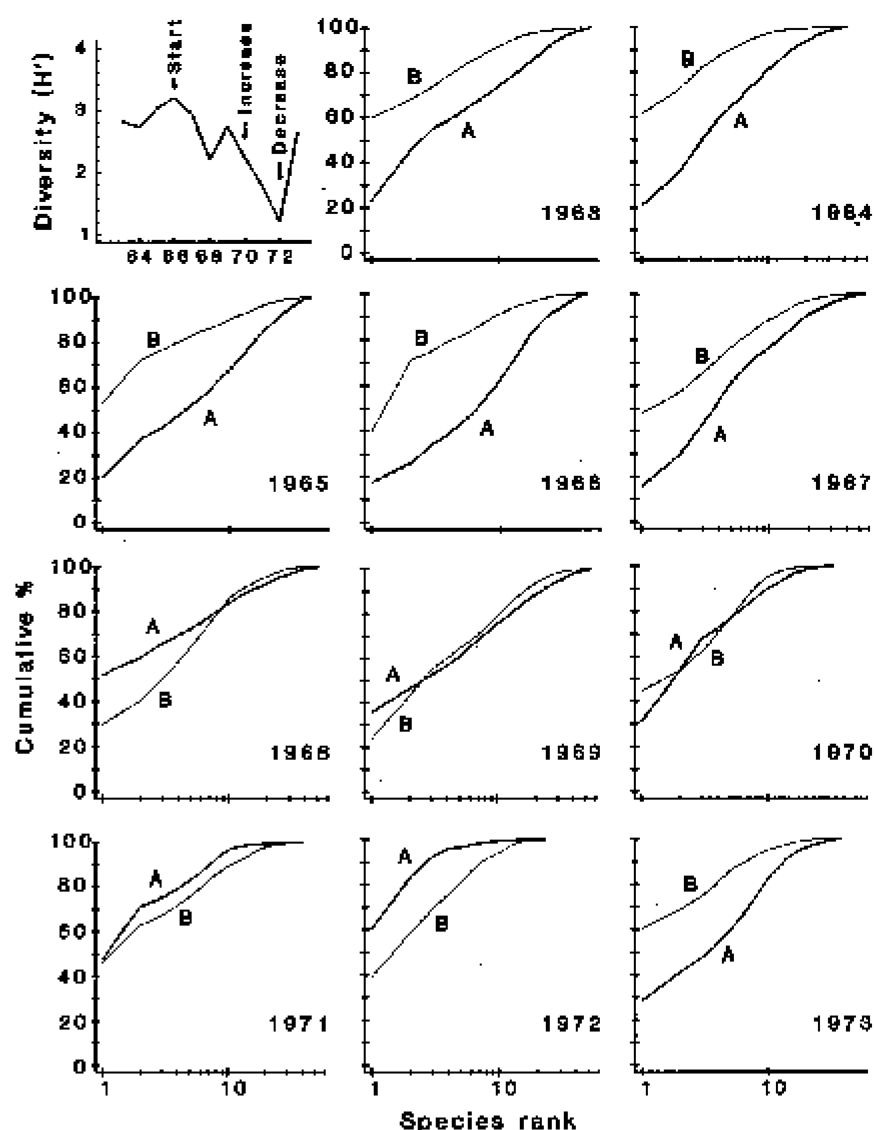


Fig. 1.4. Loch Linnhe macrofauna (L). Top left: Shannon diversity over the 11 annual samples, also indicating timing of start of effluent discharge and a later increase and decrease in level; remaining plots show ABC curves for the separate years 1963–1973 (B = biomass, thin line; A = abundance, thick line).

twelve sites, i.e. at site 1, twelve species were represented by a single individual, two species by 2–3 individuals, three species by 4–7 individuals, etc. (Gray and Pearson, 1982). For the middle sites close to the dump centre, the hypothesised loss of less-abundant species, and gain of a few species in the higher geometric classes, can clearly be seen.

MULTIVARIATE TECHNIQUES

Table 1.5 summarises the analyses possible under the four stages, when adopting one of three multivariate methods: hierarchical clustering (CLUSTER), multi-dimensional scaling (MDS) and principal component analysis (PCA).

The first two methods start explicitly from a triangular matrix of similarity coefficients computed between every pair of samples (e.g. Table 1.6). The coefficient is usually some simple algebraic measure (Chapter 2) of how close the abundance levels are for each species, averaged over all species, and defined such that 100% represents total similarity and 0% complete dissimilarity. There is a range of properties that such a coefficient should possess but still some flexibility in its choice: it is important to realise that the definition of what constitutes similarity of two communities may vary, depending on the biological question under consideration. As with the earlier methods, a multivariate analysis too must attempt to reduce the complexity of the (high-dimensional) community data by taking a

Table 1.5. **Multivariate techniques.** *Summary of analyses for the four stages.*

	Multivariate examples		
Stages	**Hierarchical clustering** *(Ch 2, 3)*	**MDS ordination** *(Ch 5)*	**PCA ordination** *(Ch 4)*
1) Representing communities	*Dendrogram of samples*	*Configuration plot of samples (often 2–dimensional)*	
2) Discriminating sites/conditions	*ANOSIM on sample similarity matrix (Ch 6)* *Similarity percentage breakdown (Ch 7) gives species responsible*		*ANOSIM on Euclidean distances (or multinormal tests, rarely valid)*
3) Determining stress levels	*Recent suggestions: "meta–analyses", variability measures, breakdown of seriation (Ch 15)*		
4) Linking to environment	*Visual: superimposing environmental variables on biotic ordinations (Ch 11)* *Analytical: finding subset of environmental variables whose ordination "best matches" the biotic ordination (Ch 11)* *(Causality: see Ch 12)*		

Table 1.6. **Frierfjord macrofauna** *(F). Bray–Curtis similarities, after √√–transformation of counts, for every pair of replicate samples from sites A, B, C only (four replicates per site).*

	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
A1	–											
A2	61	–										
A3	69	60	–									
A4	65	61	66									
B1	37	28	37	35	–							
B2	42	34	31	32	55	–						
B3	45	39	39	44	66	66	–					
B4	37	29	29	37	59	63	60	–				
C1	35	31	27	25	28	56	40	34	–			
C2	40	34	26	29	48	69	62	56	56	–		
C3	40	31	37	39	59	61	67	53	40	66	–	
C4	36	28	34	37	65	55	69	55	38	64	74	

coefficient" which reflects the extent to which the two sets of ranks do not agree, the ordination gives a simple and compelling visual representation of "closeness" of the species composition for any two samples.

The PCA technique (Chapter 4) takes a different starting position, and makes rather different assumptions about the definition of (dis)similarity of two samples, but again ends up with an ordination plot, usually in two or three dimensions, which approximates the continuum of relationships between samples (e.g. Fig. 1.8). In fact, PCA is a rather unsatisfactory procedure for most species–by–samples matrices, for at least two reasons:

a) it defines dissimilarity of samples in an inflexible way (Euclidean distance in the full–dimensional species space, Chapter 4), not well–suited to the rather special nature of species abundance data, with its predominance of zero values;

b) it requires exclusion of the species which are less common, so that the number of species retained is comparable with the number of samples.

However, a description of its operation is included in this manual because it is an historically important technique, the first ordination method to be devised and one which is still commonly encountered[10], and because it comes into its own in the analysis of environmental samples. Abiotic variables (sediment grain size, salinity, contaminant levels etc.) are usually relatively few in number, are continuously scaled, and their distributions can be transformed so that standard correlation coefficients (and Euclidean distances) are appropriate ways of describing their inter–relationships. PCA is then a fully satisfactory means of producing a low–dimensional summary, and even has some advantages over MDS in providing an interpretation of the main axes of the plot.

10. In fact, rather a bewildering array of ordination techniques are in common use (e.g. Principal Co–ordinates Analysis, Correspondence Analysis, Detrended Correspondence Analysis, etc). Chapter 5 has some brief remarks on their relation to PCA and MDS but this manual concentrates only on the two ordination methods available in PRIMER.

tested by a standard multivariate equivalent of ANOVA (MANOVA, e.g. Mardia *et al.*, 1979).

Part of the process of discriminating sites, times, treatments etc., where successful, is the ability to identify the species that are principally responsible for these distinctions: it is all too easy to lose sight of the basic data matrix in a welter of sophisticated multivariate analyses! Similarly, one might as a result of a cluster analysis determine certain sites/times that group together, and again wish to identify which species are mainly responsible for the observed clustering. Note the distinction here between *a priori* groups, identified before examination of the data, and *a posteriori* groups, identified as a result of the data analysis (the ANOSIM tests are *only* applicable to *a priori* hypotheses). These ideas are pursued in Chapter 7, both through rearrangement of the data matrix and through a possible partition of the average Bray–Curtis dissimilarity between groups of samples, into components from different species (similarity percentage breakdown, SIMPER, Clarke, 1993).

In the ***determination of stress levels***, whilst the multivariate techniques are sensitive (Chapter 14) and well-suited to establishing community differences associated with different sites/times/treatments etc., their species–specific basis would appear to make them unsuitable for drawing general inferences about the "pollution status" of an isolated group of samples. Even in comparative studies, on the face of it there is not a clear sense of "directionality" of change (e.g. deleteriousness), when it is established that communities at putatively impacted sites differ from those at control sites. Nonetheless, there are a number of ways in which such directionality *has* been ascribed in recent studies, whilst retaining an essentially multivariate form of analysis (Chapter 15):

a) a "meta–analysis" – a combined ordination of data from NE Atlantic shelf waters, at a coarse level of taxonomic discrimination[12] – suggests a common directional change in the balance of taxa under a variety of types of pollution/disturbance (Warwick and Clarke, 1993a);

b) a number of studies demonstrate increased "multivariate dispersion" among replicates under impacted conditions, in comparison to controls (Warwick and Clarke, 1993b);

c) another feature of disturbance, demonstrated in a particular coral community study, but with the potential for wider applicability, is a loss of smooth "seriation" patterns along transects (e.g. of increasing depth), again in comparison to controls in time or space (Clarke *et al.*, 1993).

Two methods of ***linking multivariate biotic patterns to environmental variables*** are explored in Chapter 11; these are illustrated here by the Garroch Head dump-ground study described earlier (Fig. 1.5). The MDS of the macrofaunal communities from the 12 sites is shown in Fig. 1.9a; this is based on Bray–Curtis similarities computed from (transformed) species biomass values.[13] A steady change in the community is apparent as the dump centre (site 6) is approached along the West arm of the transect (sites 1 to 6), with a mirrored structure along the East arm (sites 6 to 12), so that the samples from the two ends of the transect have similar species composition. That this biotic pattern correlates with the organic loading of the sediments can best be seen by superimposing the values for a single environmental variable, such as Carbon concentration, on the MDS configuration. Fig. 1.9b represents C values by circles of differing diameter, placed at the corresponding site locations on the MDS, and the pattern across sites of the 11 available environmental variables (sediment concentrations of C, N, Cu, Cd, Zn, Ni, etc.) can be viewed in this way (Chapter 11).[14]

A different approach is required in order to answer questions about *combinations* of environmental variables, for example to what extent the biotic pattern can be "explained" by knowledge of the full set, or a

12. *The effect of carrying out the various graphical and multivariate analyses at taxonomic levels higher than species is the subject of Chapter 10.*

13. *Chapter 13, and the meta–analysis section in Chapter 15, discuss the relative merits and drawbacks of using species abundance or species biomass when both are available; in fact, Chapter 13 is a wider discussion of the relative advantages of sampling particular components of the biota, for a study on the effects of pollutants.*

14. *The flexibility is clearly needed to plot an MDS configuration several times, superimposing different environmental variables. Such situations are the main motivation for the modular construction of the PRIMER package, with its stand–alone routines that exchange information via files. Thus, a similarity matrix is output by CLUSTER and input to MDS (and BIOENV, ANOSIM etc.), and configuration co–ordinates are output by MDS (and PCA) and input to the plotting routine CONPLOT. This can then be run repeatedly with differing conversion files of site designations, or different columns of an environmental file, without the need to re–run the similarity or MDS computations.*

Table 1.7. Nutrient enrichment experiment, Solbergstrand mesocosm, Norway (N). Meiofaunal abundances (shown for copepods only) from four replicate boxes for each of three treatments (Control, Low and High levels of added nutrients).

	Control				Low dose				High dose			
Species	**C1**	**C2**	**C3**	**C4**	**L1**	**L2**	**L3**	**L4**	**H1**	**H2**	**H3**	**H4**
Halectinosoma gothiceps	0	0	1	1	16	23	8	16	0	1	0	0
Danielssenia fusiformis	1	1	1	1	1	3	8	5	1	0	0	3
Tisbe sp. 1 (gracilis group)	0	0	0	0	0	0	0	0	2	27	119	31
Tisbe sp. 2	0	0	0	0	45	22	39	25	6	0	3	32
Tisbe sp. 3	0	0	0	0	86	83	88	0	5	29	0	20
Tisbe sp. 4	0	0	0	0	151	249	264	87	8	0	0	34
Tisbe sp. 5	0	0	0	0	129	0	0	115	4	0	1	40
Typhlamphiascus typhlops	4	2	2	4	5	8	4	3	0	0	0	0
Bulbamphiascus imus	1	0	0	2	0	0	0	0	0	0	0	0
Stenhelia reflexa	3	1	0	1	2	0	0	0	0	0	0	0
Amphiascus tenuiremis	1	0	0	0	0	0	2	6	0	0	0	0
Ameira parvula	0	0	0	0	4	2	3	2	2	0	1	2
Proameira simplex	0	0	0	0	0	2	0	5	0	0	0	0
Leptopsyllus paratypicus	0	0	1	0	0	0	0	0	0	0	0	0
Enhydrosoma longifurcatum	2	2	1	2	3	1	0	0	0	0	0	0
Laophontidae indet.	0	0	0	0	0	0	1	0	0	0	0	0
Ancorabolis mirabilis	3	0	4	4	2	18	3	3	27	3	1	0
Unidentified Copepodites	0	0	1	0	1	1	1	3	0	1	0	0
....												

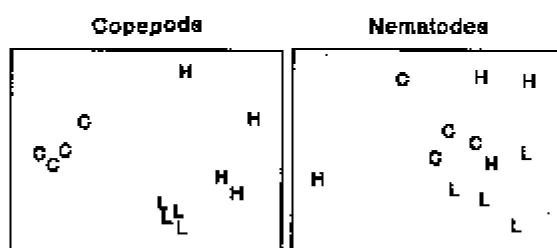


Fig. 1.10. Nutrient enrichment experiment (N). Separate MDS ordinations of √√-transformed abundances for copepod and nematode species, in four replicate boxes from each of three treatments (C, L, H).

SUMMARY

A framework has been outlined of three categories of technique (***univariate, graphical/distributional*** and ***multivariate***) and four analysis stages (***representing communities, discriminating sites/conditions, determining levels of stress*** and ***linking to environmental variables***). The least familiar tools, and the most powerful, are in the multivariate category, and those that underlie the PRIMER programs in particular are now examined from first principles.

CHAPTER 2: MEASURES OF SIMILARITY OF SPECIES ABUNDANCE / BIOMASS BETWEEN SAMPLES

SIMILARITY FOR QUANTITATIVE DATA MATRICES

Data matrix

The available biological data is assumed to consist of an array with p rows (species) and n columns (samples), whose entries are counts of each species for each sample, or the total biomass of all individuals of each species in each sample. For the moment nothing further is assumed about the structure of the samples. They might consist of one or more replicates (repeated samples) from a number of different sites, times or experimental "treatments" but this information is *not* used in the initial analysis. The strategy outlined in Chapter 1 is to *observe* any pattern of similarities and differences across the samples (i.e. let the biology "tell its own story") and, only later, compare this with known or hypothesised inter-relations between the samples based on environmental or experimental factors.

Similarity coefficient

The starting point for many of the analyses that follow is the concept of ***similarity*** (S) between any pair of samples, in terms of the biological communities they contain. Inevitably, because the information for each sample is multivariate (many species), there are many ways of defining similarity, each giving different weight to different aspects of the community. For example, some definitions might concentrate on the similarity in abundance of the few commonest species whereas others pay more attention to concurrence of rare species.

The data matrix itself may first be modified; there are three main possibilities.

a) The absolute numbers (or biomass), i.e. the fully quantitative data observed for each species, are most commonly used. In this case, two samples are considered perfectly similar only if they contain the same species in *exactly* the same abundance.

b) The relative numbers (or biomass) are sometimes used, i.e. the data is ***standardised*** to give the percentage of total abundance/biomass (over all species) that is accounted for by each species. Thus each matrix entry is divided by its column total (and multiplied by 100) to form the new array. Such standardisation will be essential if, for example, differing *and unknown* volumes of sediment or water are sampled, so that absolute numbers of individuals are not comparable between samples. Even if sample volumes are the same (or, if different, abundances are adjusted to a unit sample volume), it may still sometimes be biologically more relevant to define two samples as being perfectly similar when they have the same % ***composition*** of species, fluctuations in total abundance (or biomass) being of no interest.

c) A reduction to simple presence or absence of each species may be all that is justifiable. For example, sampling artefacts may make quantitative counts totally unreliable, or concepts of abundance may be difficult to define for some important faunal components.

A similarity coefficient S is conventionally defined to take values in the range (0,100%), or less commonly (0,1), with the ends of the range representing the extreme possibilities:

S = 100% (or 1) if two samples are totally similar;

S = 0 if two samples are totally dissimilar.

What constitutes total similarity, and particularly total dissimilarity, of two samples depends on the specific similarity coefficient adopted but there are clearly some properties that it would be desirable for a coefficient to possess. For example, S should equal zero when two samples have no species in common and S must equal 100% if two samples have identical entries (after data reduction, in cases b and c above).

Similarity matrix

Similarities are calculated between every pair of samples and it is conventional to set these $n(n-1)/2$ values out in a lower triangular matrix. This is a square array, with row and column labels being the sample numbers 1 to n, but it is not necessary to fill in either the diagonals (similarity of sample j with itself is always 100%!) or the upper right triangle (the similarity of sample j to sample k is the same as the similarity of sample k to sample j, of course).

Similarity matrices are the basis (explicitly or implicitly) of many multivariate methods, both in the representation given by a clustering or ordination

Table 2.1a). Of course, $S = 100$ if two samples are identical, since $|y_{ij} - y_{ik}| = 0$ for all i.

b) A scale change in the measurements does not change S. For example, biomass could be expressed in g rather than mg or abundance changed from numbers per cm^2 of sediment surface to numbers per m^2; all y values are simply multiplied by the same constant and this cancels in the numerator and denominator terms of equation (2.1).

c) "Joint absences" also have no effect on S. In Table 2.1a the last species is absent in all samples; omitting this species clearly makes no difference to the two summations in equation (2.1). That similarity should depend on species which are present in one or other (or both) samples, and *not* on species which are absent from both, is usually a desirable property. As Field *et al.* (1982) put it: "Taking account of joint absences has the effect of saying that estuarine and abyssal samples are similar because both lack outer–shelf species". Nonetheless, independence of joint absences is a property not shared by all similarity coefficients.

Transformation of raw data

In one or two ways, the similarities of Table 2.1b are not a good reflection of the overall match between the samples, taking all species into account. To start with, the similarities all appear too low; samples 2 and 3 would seem to deserve a similarity rating higher than 50%. As will be seen later, this is not an important consideration since the most useful multivariate methods depend on the relative order (*ranking*) of the similarities in the triangular matrix, rather than their absolute values. More importantly, the similarities of Table 2.1b are unduly dominated by the counts for the two most abundant species (4 and 5), as can be seen from studying the form of equation (2.1): terms involving species 4 and 5 dominate the sums in both numerator and denominator. Yet the larger abundances in the original data matrix will often be extremely variable in replicate samples (in statistical terms, variance is often found to increase with the square of the mean) and it is quite undesirable to base an assessment of similarity of two communities only on the counts for a handful of very abundant species.

The answer is to transform the original y values (counts or biomass) *before* computing the Bray–Curtis similarities. Two useful transformations are the ***log transform***, $\log(1 + y)$, and the ***double root*** (or ***4th root***) transform $\sqrt{\sqrt{y}}$. There is more on the effects of transformation later in the manual; for now it is only necessary to note that the $\log(1 + y)$ and $\sqrt{\sqrt{y}}$ transforms have an approximately similar and fairly severe effect in down–weighting the importance of the very abundant species so that the less dominant, and even the rare species, play some role in determining similarity of two samples. The result of the $\sqrt{\sqrt{}}$ transform for the previous example is shown in Table 2.2a and the Bray–Curtis similarities computed from these transformed abundances, using equation (2.1), are given in Table 2.2b.[1]

Table 2.2. Loch Linnhe macrofauna (L) subset. (a) $\sqrt{\sqrt{}}$–transformed abundance for the four years and six species of Table 2.1. (b) Resulting Bray–Curtis similarity matrix.

(a) Year:	64	68	71	73
(Sample:	1	2	3	4)
Species				
Echinoca.	1.7	0	0	0
Myrioche.	2.1	0	0	1.3
Labidopl.	1.7	2.5	0	1.8
Amaeana	0	1.9	3.5	1.7
Capitella	0	3.4	4.3	1.2
Mytilus	0	0	0	0

(b) Sample	1	2	3	4
1	–			
2	26	–		
3	0	68	–	
4	52	68	42	–

There is a general increase in similarity levels but, of more importance, the rank order of similarities is no longer the same as in Table 2.1b (eg $S_{24} > S_{14}$ and $S_{34} > S_{12}$ now), showing that transformations *can* have a significant effect on the final ordination or clustering display. In fact, for very variable data, choice of transformation can sometimes be more critical than choice of similarity coefficient or ordination technique, and the subject therefore merits a chapter to itself (Chapter 9).

Canberra coefficient

An alternative to transformation is to select a similarity coefficient that *automatically* adjusts the weighting given to each species when computed on original counts (or biomass). One such possibility given by Lance and Williams (1967) and referred to as the ***Canberra coefficient***, defines similarity between sample j and sample k as:

$$S_{jk} = 100\left(1 - p^{-1}\sum_{i=1}^{p} \frac{|y_{ij} - y_{ik}|}{(y_{ij} + y_{ik})}\right) \qquad (2.2)$$

Clearly, this has a strong likeness to the Bray–Curtis coefficient, but the absolute differences in counts for each species are *separately* scaled, i.e. the denominator

1. Bray–Curtis is the main coefficient calculated by the PRIMER CLUSTER program, which also allows a range of transformations of the data.

The *"simple matching" similarity* between samples j and k is defined as:

$$S_{jk} = 100.(a+d)/(a+b+c+d) \qquad (2.5)$$

so called because it represents the probability (×100) of a single species picked at random (from the full species list) being present in both samples or absent in both samples. Note that S is a function of d here, and thus depends on joint absences.

If the "simple matching" coefficient is adjusted, by first removing all species which are jointly absent from samples j and k, one obtains the *Jaccard coefficient:*

$$S_{jk} = 100.a/(a+b+c) \qquad (2.6)$$

i.e. S is the probability (×100) that a single species picked at random (from the reduced species list) will be present in both samples.

A popular coefficient found under several names, commonly *Sorenson* or *Dice*, is:

$$S_{jk} = 100.2a/(2a+b+c) \qquad (2.7)$$

Note that this is identical to the Bray–Curtis coefficient when the latter is calculated on (0, 1) presence/absence data, as can be seen most clearly from the second form of equation (2.1).[2] For example, reducing Table 2.1a to (0,1) data, and comparing samples 1 and 4 as previously, equation (2.1) gives:

$$S_{14} = 100\left(\frac{2[0+1+1+0+0+0]}{1+2+2+1+1+0}\right) = 57.1$$

This is clearly the same construction as substituting $a = 2$, $b = 1$, $c = 2$ into equation (2.7).

Among the many other coefficients that have been proposed, one that can be found occasionally in marine ecological studies is that of *McConnaughey* (1964):

$$S_{jk} = 100[a(2a+b+c)]/[2(a+b)(a+c)] \qquad (2.8)$$

RECOMMENDATIONS

1) In most ecological studies, it seems to make sense to use a coefficient which does not depend on the number of species which are jointly absent from both samples.

2) Similarities calculated on original abundance (or biomass) values can often be over-dominated by a small number of highly abundant (or large-bodied) species, so that they fail to reflect similarity of overall community composition.

3) Some coefficients (such as the Canberra) which separately scale the contribution of each species to adjust for this, have a tendency to over-compensate, i.e. rare species, which may be arbitrarily distributed across the samples, are given equal weight to very common ones. The same criticism applies to reduction of the original matrix to simple presence/absence of each species. In addition, the latter loses potentially valuable information about the approximate prevalence of a species (absent, rare, present in modest numbers, common, very abundant etc).

4) A balanced compromise is often to apply a similarity coefficient such as Bray–Curtis to counts or biomass values which have been moderately ($\sqrt{y}$) or fairly severely transformed ($\log(1 + y)$ or $\sqrt{\sqrt{y}}$). All species then contribute something to the definition of similarity whilst the retention of some information on the prevalence of a species ensures that the commoner species are generally given greater weight than the rare ones.

5) Initial standardisation is occasionally desirable, dividing each count by the total abundance of all species in that sample; this is essential when non-comparable, *unknown* sample volumes have been taken. Without this column standardisation, the Bray–Curtis coefficient will reflect differences between two samples due both to differing community composition and/or differing *total* abundance. The standardisation removes any effect of the latter; whether this is desirable is a biological rather than statistical question. (Experience with benthic communities suggests that the standardisation should usually be avoided, valuable biological information being contained in the abundance or biomass totals). Note, however, that column standardisation does not remove the need subsequently to transform the data matrix, if the similarities are to take account of more than just the few commonest species.[3]

---

2. *Thus the Sorensen coefficient can be obtained in the PRIMER CLUSTER program by "transforming" the data to presence/absence and selecting Bray–Curtis similarity.*

3. *In CLUSTER, standardisation is not the default option for sample similarities but, if selected, it is therefore carried out before any transformation.*

arbitrary to some degree. Field *et al.* (1982) suggest removal of all species that never constitute more than p% of the total abundance (/biomass) of any sample, where p is (arbitrarily) chosen to leave in around 50 or 60 species (typically $p = 3$% or so). This is preferable to simply retaining the 50 or 60 species with the highest total abundance across all samples, since the latter strategy may result in omitting several species which are key constituents of a site which is characterised by a low total number of individuals.[4] It is important to note, however, that this inevitably arbitrary process of omitting species is *not necessary* for the more usual between-sample similarity calculations. There the computation of the Bray-Curtis coefficient down-weights the contributions of the less common species in an entirely natural and continuous fashion (the rarer the species the less it contributes, on average), and all species should be retained in these calculations.

DISSIMILARITY COEFFICIENTS

The converse concept to similarity is that of ***dissimilarity***, the degree to which two samples are unlike each other. Though similarity and dissimilarity are just opposite sides of the same coin, the latter is a more natural starting point in constructing ordinations, in which *dissimilarities* (δ) between pairs of samples are turned into *distances* (d) between sample locations on a "map". Thus large dissimilarity implies that samples should be located at a large distance from each other, and dissimilarities near 0 imply nearby location; δ must therefore always be positive, of course.

Similarities can easily be turned into dissimilarities, by:

$$\delta = 100 - S \qquad (2.11)$$

For example, for the Bray-Curtis coefficient this gives:

$$\delta_{jk} = 100. \frac{\sum_{i=1}^{p} |y_{ij} - y_{ik}|}{\sum_{i=1}^{p} (y_{ij} + y_{ik})} \qquad (2.12)$$

which has limits $\delta = 0$ (no dissimilarity) and $\delta = 100$ (total dissimilarity).

4. *The PRIMER CLUSTER program will compute Bray-Curtis species similarities, with or without row standardisation and transformation (though the default is as recommended here), and allowing prior reduction by the p% criterion, either by specifying p or the number N of retained species.*

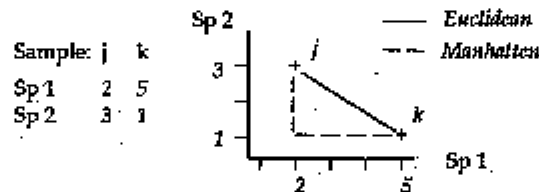
However, rather than conversion from similarities, other important dissimilarity measures arise in the first place as distances. Their role as implicit dissimilarity matrices underlying particular ordination techniques will be seen more clearly later (e.g. in Principal Components Analysis, Chapter 4).

Euclidean distance

The natural distance between any two points in space is referred to as ***Euclidean*** distance (from classical or Euclidean geometry). In the context of a species abundance matrix, the Euclidean distance between samples j and k is defined algebraically as:

$$d_{jk} = \sqrt{\left[\sum_{i=1}^{p} (y_{ij} - y_{ik})^2\right]} \qquad (2.13)$$

This can best be understood, geometrically, by taking the special case where there are only two species so that samples can be represented by points in 2-dimensional space, namely their position on the two axes of Species 1 and Species 2 counts. This is illustrated below for a specific two samples by two species abundance matrix. The co-ordinate points (2, 3) and (5, 1) on the (Sp. 1, Sp. 2) axes are the two samples j and k. The direct distance d_{jk} between them of $\sqrt{[(2-5)^2 + (3-1)^2]}$ (Pythagoras) clearly corresponds to equation (2.13).



It is easy to envisage the extension of this to a matrix with three species; the two points are now simply located on 3-dimensional species axes and their straight line distance apart is a natural geometric concept. Algebraically, it is the root of the sums of squared distances apart along the three axes, equation (2.13). Extension to four and higher numbers of species (dimensions) is harder to envisage geometrically (in our 3-dimensional world) but the concept remains unchanged and the algebra is no more difficult to understand in higher dimensions than three: additional squared distances apart on each new species axis are added to the summation under the square root in (2.13). In fact, this concept of representing a species-by-samples matrix as points in high-dimensional ***species space*** is a very fundamental and important one and will be met again in Chapter 4,

CHAPTER 3: HIERARCHICAL CLUSTERING

CLUSTER ANALYSIS

The previous chapter has shown how to replace the original data matrix with pairwise similarities, chosen to reflect the particular aspect of similarity in community structure (similarity in counts of abundant species, similarity in general disposition of rare species etc) which the biologist requires to emphasise for the study in question. Typically, the number of pairwise similarities is large, $n(n-1)/2$ for n samples, and it can often be no easier to detect a pattern in the resulting lower triangular similarity matrix than it is in the original data. Table 3.1 illustrates this for just a portion (roughly a quarter) of the similarity matrix for the Frierfjord macrofauna data *(F)*. Close examination shows that the four replicates within site A generally have higher within–site similarities than do pairs of replicates within sites B and C, or replicates *between* sites, but the pattern is far from clear. What is needed is a graphical display linking samples that have mutually high levels of similarity.

Table 3.1. Frierfjord macrofauna counts (F). Bray–Curtis similarities, after $\sqrt{\sqrt{}}$ transformation of counts, for every pair of replicate samples from sites A, B, C only (four replicate samples per site).

	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
A1	–											
A2	67	–										
A3	69	60	–									
A4	65	61	66									
B1	37	28	37	35	–							
B2	42	34	31	32	55	–						
B3	45	39	39	44	66	66	–					
B4	37	29	29	37	59	63	60	–				
C1	35	31	27	25	28	56	40	34	–			
C2	40	34	26	29	48	69	62	56	56	–		
C3	40	31	37	39	59	61	67	53	40	66	–	
C4	36	28	34	37	65	55	69	55	38	64	74	–

Cluster analysis (or ***classification***) aims to find "natural groupings" of samples such that samples within a group are more similar to each other, generally, than samples in different groups. Cluster analysis is used in the present context in the following ways.

a) Different sites (or different times at the same site) can be seen to have differing community compositions by noting that replicate samples within a site form a cluster that is distinct from replicates within other sites. This can be an important hurdle to overcome in any analysis; if replicates for a site are clustered more or less randomly with replicates from every other site then further interpretation is likely to be dangerous. (A more formal statistical test for distinguishing sites is the subject of Chapter 6).

b) When it is established that sites can be distinguished from one another (or, when replicates are not taken, it is assumed that a single sample is representative of that site or time), sites or times can be partitioned into groups with similar community structure.

c) Cluster analysis of the *species* similarity matrix can be used to define species assemblages, ie groups of species that tend to co–occur in a parallel manner across sites.

Range of methods

Literally hundreds of clustering methods exist, some of them operating on similarity/dissimilarity matrices whilst others are based on the original data. Everitt (1980) and Cormack (1971) give excellent and readable reviews. Clifford and Stephenson (1975) is another well-established text on classification methods, from an ecological viewpoint.

Five classes of clustering methods can be distinguished, following the categories of Cormack (1971).

1) ***Hierarchical methods.*** Samples are grouped and the groups themselves form clusters at lower levels of similarity.

2) ***Optimising techniques.*** A single set of mutually exclusive groups (usually a pre–specified number) is formed by optimising some clustering criterion, for example minimising a within–cluster distance measure in the species space.

3) ***Mode–seeking methods.*** These are based on considerations of *density* of samples in the neighbourhood of other samples, again in the species space.

4) ***Clumping techniques.*** The term "clumping" is reserved for methods in which samples can be placed in more than one cluster.

5) ***Miscellaneous techniques.***

Table 3.2. Loch Linnhe macrofauna (L) subset. Abundance array after $\sqrt{\sqrt{}}$–transform, the resulting Bray–Curtis similarity matrix and the successively fused similarity matrices from a hierarchical clustering, using group average linking.

Year:	64	68	71	73
Sample:	1	2	3	4
Species -				
Echinoca.	1.7	0	0	0
Myrioche.	2.1	0	0	1.3
Labidopl.	1.7	2.5	0	1.8
Amaeana	0	1.9	3.5	1.7
Capitella	0	3.4	4.3	1.2
Mytilus	0	0	0	0

→

Sample	1	2	3	4
1	–			
2	25.6	–		
3	0.0	67.9	–	
4	52.2	68.1	42.0	–

→

Sample	1	2&4	3
1	–		
2&4	38.9	–	
3	0.0	55.0	–

→

Sample	1	2&3&4
1	–	
2&3&4	25.9	–

a) ***Single linkage.*** $S(1, 2\&4)$ is the *maximum* of $S(1, 2)$ and $S(1, 4)$, *i.e.* 52.2%.

b) ***Complete linkage.*** $S(1, 2\&4)$ is the *minimum* of $S(1, 2)$ and $S(1, 4)$, *i.e.* 25.6%.

c) ***Group–average link.*** $S(1, 2\&4)$ is the *average* of $S(1, 2)$ and $S(1, 4)$, *i.e.* 38.9%.

Table 3.2 adopts group–average linking, hence

$$S(2\&4, 3) = [S(2, 3) + S(4, 3)]/2 = 55.0$$

The new matrix is again examined for the highest similarity, defining the next fusing; here this is between "2&4" and "3", at similarity level 55.0%. The matrix is again reformed for the two new clusters "1" and "2&3&4" and there is only a single similarity, $S(1, 2\&3\&4)$, to define. For *group–average linking*, this is the mean of $S(1, 2\&4)$ and $S(1, 3)$ but it must be a *weighted* mean, allowing for the fact that there are twice as many samples in cluster "2&4" as in cluster "3". Here:

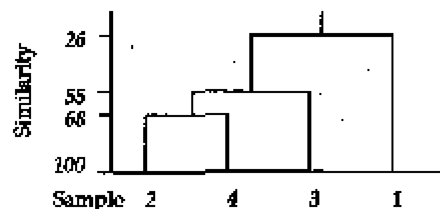
$$S(1, 2\&3\&4) = [2 \times S(1, 2\&4) + 1 \times S(1, 3)]/3$$
$$= (2 \times 38.9 + 1 \times 0)/3 = 25.9$$

Though it is computationally efficient to form each successive similarity matrix by taking weighted averages of the similarities in the previous matrix, an alternative which is entirely equivalent (and perhaps conceptually simpler) is to define the similarity between two groups as the *simple* (unweighted) average of all between–group similarities in the initial triangular matrix. Thus:

$$S(1, 2\&3\&4) = [S(1, 2) + S(1, 3) + S(1, 4)]/3$$
$$= (25.6 + 0.0 + 52.2)/3 = 25.9,$$

the same answer as above.

The final merge of all samples into a single group therefore takes place at similarity level 25.9%, and the clustering process for the group–average linking shown in Table 3.2 can be displayed in the following dendrogram.



Dendrogram features

This example raises a number of more general points about the use and appearance of dendrograms.

1) Samples need to re–ordered along the x axis, for clear presentation of the dendrogram; it is always possible to arrange samples in such an order that none of the dendrogram branches cross each other.
2) The resulting order of samples on the x axis is not unique. A simple analogy is with a child's "mobile"; the vertical lines are strings and the horizontal lines rigid bars. When the whole structure is suspended by the top string, the bars can rotate freely, generating many possible re–arrangements of samples on the x axis. For example, in the above figure, samples 2 and 4 could switch places (sequence 4, 2, 3, 1) or sample 1 move to the opposite side of the diagram (sequence 1, 2, 4, 3), but a sequence such as 1, 2, 3, 4 is not possible. It follows that to use the x axis sequence as an *ordering* of samples is misleading.
3) Cluster analysis attempts to group samples into discrete clusters not display their inter–relationship on a continuous scale; the latter is the province of ***ordination*** and this would be preferable for the simple example above. Clustering imposes a rather arbitrary grouping on what appears to be a continuum of change from an unpolluted year (1964), through steadily increasing impact (loss of some species, increase in abundance of "opportu-

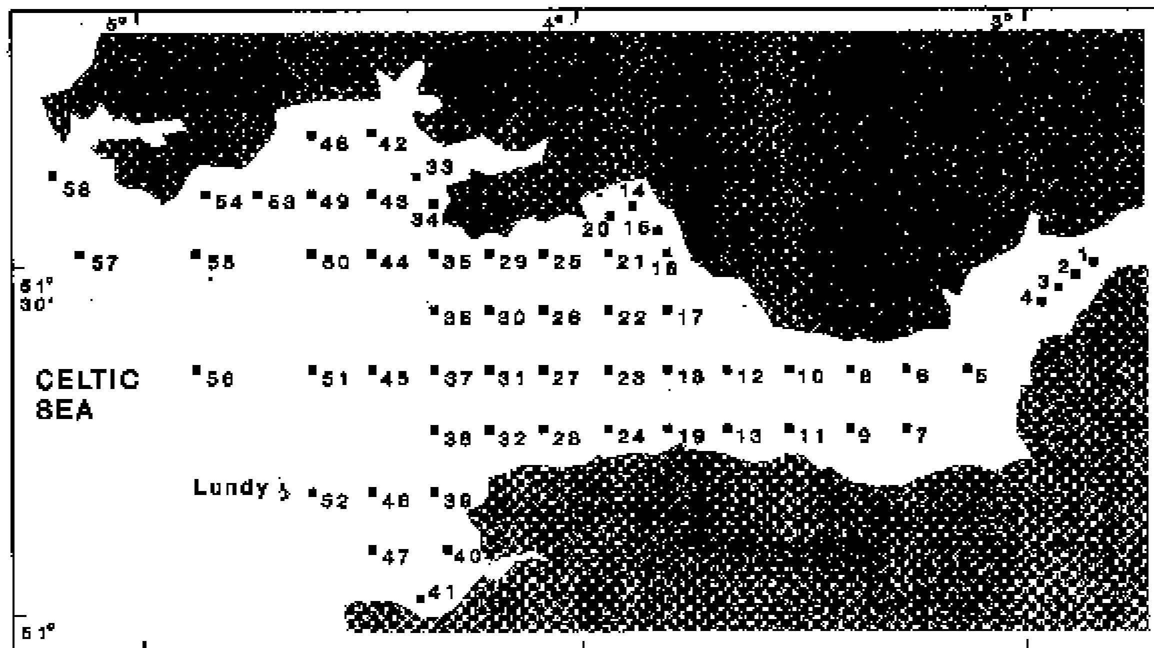


Fig. 3.2 Bristol Channel zooplankton (B). Sampling sites.

EXAMPLE: Bristol Channel zooplankton

Collins and Williams (1982) perform hierarchical cluster analyses of zooplankton samples, collected by double oblique hauls at 57 sites in the Bristol Channel UK, for three different seasons in 1974 *{B}*. This is not a pollution study but a baseline survey carried out by the Plymouth laboratory, as part of a major programme to understand and model the ecosystem of the estuary. Fig. 3.2 is a map of the sample locations, sites 1–58 (site 30 not sampled).

Fig. 3.3 shows the results of a hierarchical clustering using group-average linking on data sampled during April 1974. The raw data were expressed as numbers per cubic metre for each of 24 holozooplankton species, and Bray-Curtis similarities calculated on $\sqrt{\sqrt{}}$-transformed abundances. From the resulting dendrogram, Collins and Williams select the four groups determined at a 55% similarity level and characterise these as *true estuarine* (sites 1–8, 10, 12), *estuarine and marine* (9, 11, 13–27, 29), *euryhaline marine* (28, 31, 33–35, 42–44, 47–50, 53–55) and *stenohaline marine* (32, 36–41, 45, 46, 51, 52, 56–58). A corresponding clustering of *species* and a re-ordering of the rows and columns of the original data matrix allows the identification of a number of species groups characterising these main site clusters, as is seen later (Chapter 7).

The dendrogram provides a sequence of fairly convincing groups; once each of the four main groups has formed it remains separate from other groups over a relatively large drop in similarity. Even so, a cluster analysis gives an incomplete and disjointed picture of the sample pattern. Remembering the analogy of the "mobile", it is not clear from the dendrogram alone whether there is any natural sequence of community change across the four main clusters (implicit in the designations true estuarine, estuarine and marine, euryhaline marine, stenohaline marine). For example, the stenohaline marine group could just as correctly have been rotated to lie between the estuarine and marine and euryhaline marine groups. In fact, there is a strong (and more-or-less continuous) gradient of community change across the region, associated with the changing salinity levels. This is best seen in an ordination of the 57 samples on which are superimposed the salinity levels at each site; this example is therefore returned to in Chapter 11.

RECOMMENDATIONS

1) Hierarchical clustering with group-average linking, based on sample similarity or dissimilarity matrices such as Bray-Curtis, has proved a useful technique in a number of ecological studies of the last two decades. It is appropriate for delineating groups of sites with distinct community structure

CHAPTER 4: ORDINATION OF SAMPLES BY PRINCIPAL COMPONENT ANALYSIS (PCA)

ORDINATIONS

An *ordination* is a *map* of the samples, usually in two or three dimensions, in which the placement of samples, rather than representing their simple geographical location, reflects the similarity of their biological communities. To be more precise, *distances* between samples on the ordination attempt to match the corresponding *dissimilarities* in community structure: nearby points have very similar communities, samples which are far apart have few species in common or the same species at very different levels of abundance (or biomass). The word "attempt" is important here since there is no uniquely defined way in which this can be achieved. (Indeed, when a large number of species fluctuate in abundance in response to a wide variety of environmental variables, each species being affected in a different way, the community structure is essentially *high–dimensional* and it may be impossible to obtain a useful two or three–dimensional representation).

So, as with cluster analysis, several methods have been proposed, each using different forms of the original data and varying in their technique for approximating high–dimensional information in low–dimensional plots. They include:

a) ***Principal Components Analysis,*** PCA (see, for example, Chatfield and Collins, 1980);

b) ***Principal Co–ordinates Analysis,*** PCoA (Gower, 1966);

c) ***Correspondence Analysis*** and ***Detrended Correspondence Analysis,*** DECORANA (Hill and Gauch, 1980);

d) ***Multi–Dimensional Scaling,*** MDS; in particular ***non–metric MDS*** (see, for example, Kruskal and Wish, 1978).

A comprehensive survey of ordination methods is outside the scope of this volume. As with clustering methods, detailed explanation is given only of the techniques required for the analysis strategy adopted throughout the manual. This is not to deny the validity of other methods but simply to affirm the importance of applying, with *understanding*, one or two techniques of proven utility. The two ordination methods selected are therefore the simplest (arguably) of the various options, at least in concept.

a) PCA is the longest–established method, though the relative inflexibility of its definition limits its practical usefulness more to multivariate analysis of environmental data rather than species abundances or biomass; nonetheless it is still widely encountered and is of fundamental importance.

b) Non–metric MDS is a more recent development, whose complex algorithm could only have been contemplated in an era of advanced computational power; however, its rationale can be very simply described and understood, and many people would argue that the need to make few (if any) assumptions about the data make it the most widely applicable and effective method available.

PRINCIPAL COMPONENTS ANALYSIS

The starting point for a PCA is the original data matrix rather than a derived similarity matrix (though there is an *implicit* dissimilarity matrix underlying PCA, that of Euclidean distance). The data array is thought of as defining the positions of samples in relation to axes representing the full set of species, one axis for each species. This is the very important concept introduced in Chapter 2 (following equation (2.13)). Typically, there are many species so the samples are points in a very high–dimensional space.

A simple 2–dimensional example

It helps to visualise the process by again considering an (artificial) example in which there are only two species (and nine samples).

	Sample	1	2	3	4	5	6	7	8	9
Abundance	Sp. 1:	*6*	*0*	*5*	*7*	*11*	*10*	*15*	*18*	*14*
	Sp. 2:	*2*	*0*	*8*	*6*	*6*	*10*	*8*	*14*	*14*

The nine samples are therefore points in two dimensions, and labelling these points with the sample number gives the following plot.

(perpendicular) distances of the points from the line.[1] The second approach comes from noting in the above example that the biggest differences between samples take place along the PC1 axis, with relatively small changes in the PC2 direction. The PC1 axis is therefore defined as that direction in which the *variance* of sample points projected perpendicularly onto the axis *is maximised*. In fact, these two separate definitions of the PC1 axis turn out to be *totally equivalent* and one can use whichever concept is easier to visualise.

Extension to 3–dimensional data

Suppose that the simple example above is extended to the following matrix of counts for *three* species.

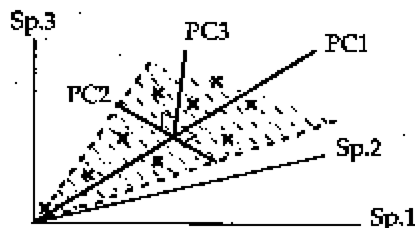
	Sample	1	2	3	4	5	6	7	8	9
Abundance	Sp. 1:	6	0	5	7	11	10	15	18	14
	Sp. 2:	2	0	8	6	6	10	8	14	14
	Sp. 3:	3	1	6	6	9	11	10	16	15

Samples are now points in three dimensions (Sp.1, Sp.2 and Sp.3 axes) and there are therefore three principal component axes, again simply a rotation of the three species axes. The definition of the (PC1, PC2, PC3) axes generalises the 2–dimensional case in a natural way:

PC1 is the axis which *maximises* the *variance* of points projected perpendicularly onto it;

PC2 is constrained to be perpendicular to PC1, but is then again chosen as the direction in which the variance of points projected perpendicularly onto it is maximised;

PC3 is the axis perpendicular to both PC1 and PC2 (there is no choice remaining here).



1. *This type of idea may be familiar from ordinary linear regression, except that the latter is formulated asymmetrically: the regression of y on x minimises the sum of squared* vertical *distances of points from the line.*

An equivalent way of visualising this is again in terms of "best fit": PC1 is the "best fitting" line to the sample points and, together, the PC1 and PC2 axes define a plane (stippled in the above diagram) which is the "best fitting" plane.

Algebraic definition

The above geometric formulation can be expressed algebraically. The three new variables (PCs) are just ***linear combinations*** of the old variables (species), such that PC1, PC2 and PC3 are ***uncorrelated***. In the above example:

$$\begin{aligned} PC1 &= \;\;\,0.62\times Sp.1 + 0.52\times Sp.2 + 0.58\times Sp.3 \\ PC2 &= -0.73\times Sp.1 + 0.65\times Sp.2 + 0.20\times Sp.3 \quad (4.1) \\ PC3 &= \;\;\,0.28\times Sp.1 + 0.55\times Sp.2 - 0.79\times Sp.3 \end{aligned}$$

The principal components are therefore interpretable (in theory) in terms of the counts for each original species axis. Thus PC1 is a sum of roughly equal (and positive) contributions from each of the species; it is essentially ordering the samples from low to high total abundance. At a more subtle level, for samples with the same total abundance, PC2 then mainly distinguishes relatively high counts of Sp.2 (and low Sp.1) from low Sp.2 (and high Sp.1); Sp.3 values do not feature strongly in PC2 because the corresponding coefficient is small. Similarly the PC3 axis mainly contrasts Sp.3 and Sp.2 counts.

Variation explained by each PC

The definition of principal components given above is in terms of successively maximising the variance of sample points projected along each axis, with the variance therefore decreasing from PC1 to PC2 to PC3. It is thus natural to quote the values of these variances (in relation to their total) as a measure of the amount of "information" contained in each axis. Furthermore, it turns out that the total of the variances along all PC axes is equal to the total variance of points projected successively onto each of the original species axes. That is, letting *var (PCi)* denote variance of samples on the *i*th PC axis and *var (Sp.i)* denote variance of points on the *i*th species axis (i = 1,2,3):

$$\Sigma_i\, var(PCi) = \Sigma_i\, var(Sp.i) \qquad (4.2)$$

Thus, the relative variation of points along the *i*th PC axis (as a percentage of the total), namely

$$P_i = 100.\frac{var(PCi)}{\Sigma_i\, var(PCi)} = 100.\frac{var(PCi)}{\Sigma_i\, var(Sp.i)} \qquad (4.3)$$

has a useful interpretation as the % of the original total *variance explained* by the *i*th PC. For the simple

nal abundance data *(L)* is shown in Fig. 4.1. The original matrix contained a total of 115 species for the 11 samples, one for each year of the period 1963–1973. Pulp–mill effluent was first discharged to the loch in 1966 with an increased discharge in 1969/70 and a subsequent decrease in 1972/73.

Exclude less–common species

The retention of rarer species in a PCA ordination will have a strongly distorting effect, even supposing that the matrix operations to construct the ordination are possible. For the Loch Linnhe data there are 11 samples in 115–dimensional species space! An initial and drastic reduction in the number of species is necessary for the PCA algorithm to work. In fact, many of the species are represented only by a single individual in a single year and their omission will not be a serious loss to interpretation, but the necessity of making an (essentially arbitrary) decision about which species to exclude is one of the problems with applying PCA to biological community data. By contrast, the clustering methods of the last chapter were applied to a similarity matrix which could be constructed from *all* species, the rarer ones either being emphasised, as in reduction to presence/absence, or down–weighted automatically (though not ignored totally) by the choice of similarity coefficient and transformation. An ordination method based on this similarity matrix (for example, the MDS method of Chapter 5) clearly scores over PCA, in this respect.

In fact, Fig 4.1 is based on a data matrix of only 29 species, those making up more than 3% of the total abundance in at least one of the samples. (The rationale for this type of selection procedure was

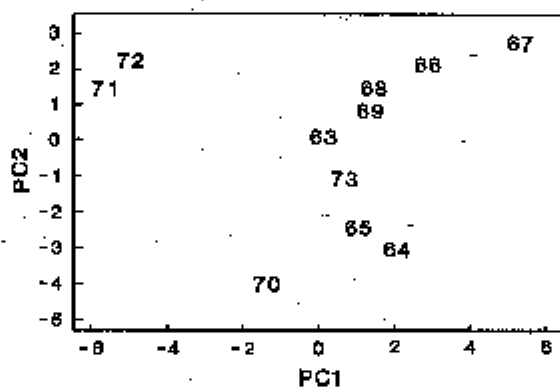


Fig. 4.1. Loch Linnhe macrofauna (L). 2–dimensional PCA ordination of sample abundances (√√– transformed) from the 11 years 1963–1973. PC1 (x–axis) and PC2 (y–axis) together account for 57% of the total sample variability.

discussed in the section on species similarities in Chapter 2). Calculation of the principal components is now possible though, even so, the software package needs to handle its computations carefully. A total of 11 sample points will always fit perfectly into 10 dimensions (think of the lower–dimensional analogy again: 3 points in 3–dimensional space will always lie on a 2–dimensional plane). Thus, only 10 (at most) PC axes can be constructed, or to put it another way, all the sample variance can be explained by the first 10 PCs. In fact, the first two PCs in Fig. 4.1 explain 57% of the total variability so the 2–dimensional ordination does not give a fully satisfactory picture of the changing community pattern over the years. If this example were being pursued further, it would be advisable to look also at the third PC (at least), perhaps with some form of 3–dimensional perspective plot or by the three separate 2–dimensional plots of (PC1, PC2), (PC1, PC3) and (PC2, PC3). Nonetheless, one main feature is clear from Fig. 4.1: the relatively large change in community composition between 1970 and 1971, and the reversion in 1973 to a community which is more like the earlier years.

Transformation of abundance/biomass

In much the same way as was seen for the calculation of similarity coefficients in Chapter 2, it may be necessary to make an initial transformation of the abundance or biomass values to avoid over–domination of the resulting analysis by the very common species. For the Loch Linnhe data, *Capitella* numbers in a yearly sample range from 0 to over 4,000 individuals, whereas the bulk of the other species have counts in single or double figures. For untransformed data (and using a covariance–based analysis, as discussed below), the *Capitella* axis will clearly contain a substantial part of the overall variation of samples in the species space, so that the direction of the PC1 axis will tend to be dictated by that species alone. A more balanced picture will emerge after transformation: Fig. 4.1 is based on √√–transformed abundances.

Scale and location changes

The data matrix can also be ***normalised*** (after any transformation has taken place). For each species abundance, subtract the mean count and divide by the standard deviation over all samples for that species. This makes the variance of samples along all species axes the same (= 1) so all species are of potentially equal importance in determining the principal components. This normalised analysis is referred to as ***correlation–based PCA*** rather than the ***covariance–based PCA*** obtained when the data is not normalised

different variables in the environmental analysis, e.g. contaminant concentrations will often be right-skewed (and require something like log transformation), salinity may be left-skewed (reverse log transformation) and sediment granulometry measures like "% mud" or "silt/clay" may need no transformation at all. These issues are returned to in Chapter 9.

PCA STRENGTHS

1) ***PCA is conceptually simple.*** Whilst the algebraic basis of the PCA algorithm requires a facility with matrix algebra for its understanding, the geometric concepts of a "best fitting" plane in the species space, and perpendicular projection of samples onto that plane, are relatively easily grasped. Some of the more recently proposed ordination methods, which either extend or supplant PCA (eg Principal Co-ordinates Analysis, Detrended Correspondence Analysis) can be very much harder to understand for practitioners without a mathematical background.

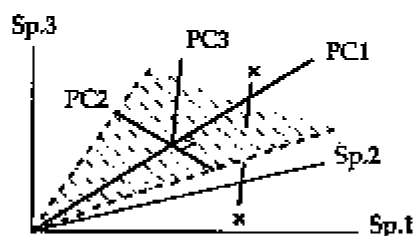
2) ***It is computationally straightforward.*** Again, this statement needs to be seen in relative terms. *Provided* the number of species is reduced, usually drastically, the required matrix operations pose no real problems to modern computing power and packages are widely available which carry out the necessary eigenvalue (***latent root***) extraction. That multivariate methods have only come to the fore as a practical data analysis tool in the last two decades should not be a surprise to anyone. Even the computationally simplest of techniques, PCA, could never be carried out manually in any realistic example. Nonetheless, PCA tends to take only seconds, rather than minutes or hours, of processing time on a personal computer. The constraints are mainly on the number of *species* handled, and large numbers of samples can usually be accommodated. This is in contrast to cluster and MDS analyses which tend to be more constrained by the number of *samples* they can handle; once the data is reduced to a similarity matrix between samples (the input form to both clustering and MDS) the number of species in the original matrix is irrelevant. PCA could therefore have a role, when there are large numbers of samples, in providing an initial picture which would suggest separation of the data into two (or more) distinct sets of samples, each of which is analysed by more accurate (but more computationally–intensive) ordinations such as MDS.

3) ***Ordination axes are interpretable.*** The PC axes are simple linear combinations of the values for each species, as in equation (4.1), so in theory have some potential for interpretation. In practice though, when there are more than a handful of species (as is usual), this rarely leads to any useful information. Environmental data arrays often contain a smaller number of variables however, and interpretation of the PCA axes may be informative in that case (see, for example, Chapter 11).

PCA WEAKNESSES

1) ***There is little flexibility in defining dissimilarity.*** An ordination is essentially a technique for converting *dissimilarities* of community composition between samples into (Euclidean) *distances* between these samples in a 2– or higher–dimensional ordination plot. Implicitly, PCA defines dissimilarity between two samples as their Euclidean distance apart in the full p–dimensional species space; however, as was seen in Chapter 2, this is rather a poor way of defining sample dissimilarity: something like a Bray–Curtis coefficient would be preferred but standard PCA cannot accommodate this. The only flexibility it has is in transforming (and/or normalising) the species axes so that dissimilarity is defined as Euclidean distance on these new scales.

2) ***Its distance–preserving properties are poor.*** Having defined dissimilarity as distance in the p–dimensional species space, PCA converts these distances by *projection* of the samples onto the 2–dimensional ordination plane. This may distort some distances rather badly. Taking the usual visual analogy of a 2–dimensional ordination from three species, it can be seen that samples which are relatively far apart on the PC3 axis can end up being co–incident when projected (perhaps from "opposite sides") onto the (PC1, PC2) plane.



CHAPTER 5: ORDINATION OF SAMPLES BY MULTI-DIMENSIONAL SCALING (MDS)

OTHER ORDINATION METHODS

Principal Co-ordinates Analysis

The two main weaknesses of PCA, identified at the end of Chapter 4, are its inflexibility of dissimilarity measure and its poor distance-preservation. The first problem is addressed in an important paper by Gower (1966), describing an extension to PCA termed ***Principal Co-ordinates Analysis*** (PCoA), also sometimes referred to as ***classical scaling***. This allows a much wider definition of dissimilarity than simple Euclidean distance in the species space (the basis of PCA). Other dissimilarity measures are converted to distances, in high-dimensional space, but the final step is again a projection onto a low-dimensional ordination space (eg a 2-dimensional plane), as in ordinary PCA. Thus, PCA is a special case of PCoA, when the original dissimilarity is just Euclidean distance. It follows that PCoA is still subject to the second criticism of PCA: its lack of emphasis on distance-preservation when the information is difficult to represent in a low number of dimensions.

Detrended Correspondence Analysis

Correspondence analyses are a class of ordination methods featuring strongly in French data-analysis literature (for a review in English see Greenacre, 1984). Key papers in ecology are Hill (1973a) and Hill and Gauch (1980), who introduced ***detrended correspondence analysis*** (DECORANA). The methods start from the data matrix, rather than a set of dissimilarity coefficients, so are rather inflexible in their definition of sample dissimilarity; in effect, multinomial assumptions generate an implicit dissimilarity measure of "chi-squared" distance. Basic correspondence analysis (CA) has its genesis in a particular model of unimodal species response to underlying (but unmeasured) environmental gradients; an account is outside the scope of this manual but a comprehensive exposition (by C. F. J. ter Braak) of CA and related techniques can be found in Jongman *et al.* (1987).[1]

The popular DECORANA version of CA has a primary motivation of straightening out an "arch effect" in a CA ordination, which is expected on theoretical grounds if species abundances have unimodal (Gaussian) responses along a single strong environmental gradient. Where such models are not appropriate, it is unclear what artefacts the algorithms may introduce into the final picture. In the Hill and Gauch (1980) procedure, the detrending is essentially carried out by first splitting the ordination space into segments, stretching or shrinking the scale in each segment and then realigning the segments to remove wide-scale curvature. For some people, this is uncomfortably close to attacking the data with scissors and glue and, though the method is not as subjective as this would imply, some arbitrary decisions about where and how the segmentation and rescaling are defined are rather hidden from the user in the program code. Thus Pielou (1984) and others have criticized DECORANA for its "overzealous" manipulation of the data. It is also a pity that the multivariate techniques which historically have been applied most frequently in the ecological literature are often either inadequately suited to the data or are based on conceptually complex algorithms (e.g. DECORANA and TWINSPAN, Hill 1979a, b), erecting a communication barrier between data analyst and ecologist.

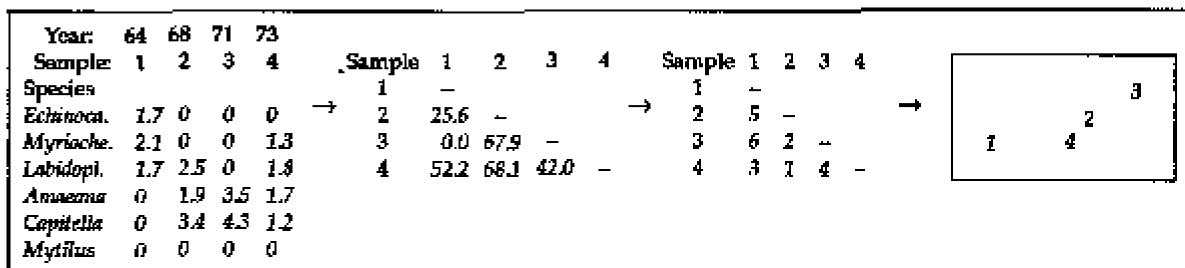
The ordination technique which is adopted in this manual's strategy, ***non-metric MDS***, is itself a complex numerical algorithm but it can (and will) be argued that it is *conceptually* simple. It makes few (if any) model assumptions about the form of the data or the inter-relationship of the samples, and the link between the final picture and the user's original data is relatively transparent and easy to explain. It addresses both the major criticisms of PCA made earlier: it has *great flexibility* both in the definition and conversion of dissimilarity to distance and its rationale *is* the preservation of these relationships in the low-dimensional ordination space.

NON-METRIC MULTIDIMENSIONAL SCALING

The method of ***non-metric MDS*** was introduced by Shepard (1962) and Kruskal (1964), for application to problems in psychology; a useful introductory text is

1. *A convenient way of carrying out correspondence analyses, and related canonical methods, is to use ter Braak's excellent CANOCO package.*

Table 5.1. ***Loch Linnhe macrofauna (L) subset.*** *Abundance array after √√–transform, the Bray–Curtis similarities (as in Table 3.2), the rank similarity matrix and the resulting 2–dimensional MDS ordination.*



Year:	64	68	71	73
Sample:	1	2	3	4
Species				
Echinoca.	1.7	0	0	0
Myrioche.	2.1	0	0	1.3
Labidopl.	1.7	2.5	0	1.8
Amaema	0	1.9	3.5	1.7
Capitella	0	3.4	4.3	1.2
Mytilus	0	0	0	0

→

Sample	1	2	3	4
1	–			
2	25.6	–		
3	0.0	67.9	–	
4	52.2	68.1	42.0	–

→

Sample	1	2	3	4
1	–			
2	5	–		
3	6	2	–	
4	3	1	4	–

→

Three replicate sediment cores were taken for meiofaunal analysis on each occasion, and nematodes identified and counted. This analysis considers only the mean nematode abundances across replicates and season (no seasonal differences were evident in a more detailed analysis), so the data matrix consists of 182 species and 19 samples.

This is not an example of a pollution study: the Exe estuary is a relatively unimpacted environment. The aim here is to display the biological relationships among the 19 stations and then to link these to a set of environmental variables (granulometry, interstitial salinity etc.) measured at these sites, to reveal potential determinants of nematode community structure. Fig. 5.1 shows the 2–dimensional MDS ordination of the 19 samples, based on √√–transformed abundances and a Bray–Curtis similarity matrix. Distinct clusters of sites emerge (in agreement with those from a matching cluster analysis), bearing no clear–cut relation to geographical position or tidal level of the samples.

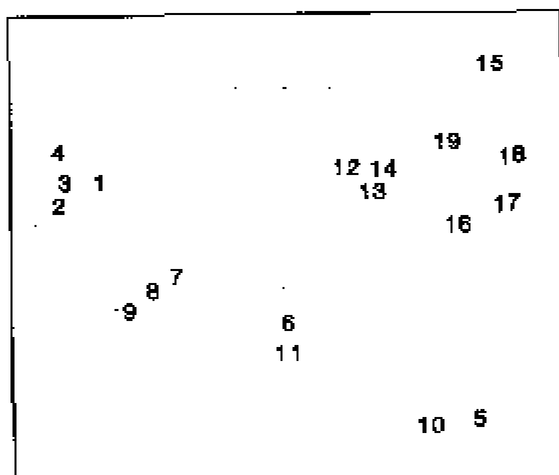


Fig. 5.1. Exe estuary nematodes (X). *MDS ordination of the 19 sites based on √√–transformed abundances and Bray–Curtis similarities (stress = 0.05).*

Instead they appear to relate to variables such as sediment type and organic content, and these links are discussed further in Chapter 11. For now the question is: what are stages in the construction of Fig. 5.1?

MDS ALGORITHM

The non–metric MDS algorithm, as employed in Kruskal's original MDSCAL program for example, is an *iterative procedure*, constructing the MDS plot by successively refining the positions of the points until they satisfy, as closely as possible, the dissimilarity relations between samples.[3] It has the following steps.

1) ***Specify the number of dimensions*** (m) required for the final ordination plot. If, as will sometimes be desirable, one wishes to compare configurations in two and three dimensions then they have to be constructed separately. For the moment think of m as 2.

2) ***Construct a starting configuration of the n samples.*** This could be the result of an ordination by another method, for example PCA or PCoA, or it could literally be just a random set of n points in m (= 2) dimensions.

3) ***Regress the interpoint distances from this plot on the corresponding dissimilarities.*** Let $\{d_{jk}\}$ denote the distance between the jth and kth sample points on the current ordination plot, and $\{\delta_{jk}\}$ the corresponding dissimilarity in the original dissimilarity matrix (of, say, Bray–Curtis coefficients). A scatter plot is then drawn of distance against dissimilarity for all $n(n-1)/2$ such pairs of values. This is termed a ***Shepard diagram*** and Fig 5.2 shows the type of graph that results. (In fact, this is at a late

3. *This is also the algorithm used in the PRIMER program MDS. The required input is a similarity matrix (e.g. as produced by CLUSTER), and the output includes a plot file which can be input to CONPLOT to display the 2–d MDS configuration.*

three dimensions, with just a 2–dimensional parameter space (the x,y plane) and the vertical axis (z) denoting the stress at each (x,y) point. In reality the stress surface is a function of more parameters than this of course, but we have seen before how useful it can be to visualise high–dimensional algebraic operations in terms of 3–dimensional geometry. An appropriate analogy is to imagine a rambler walking across a range of hills in a thick fog(!), attempting to find the *lowest* point within an encircling range of high peaks. A good strategy is always to walk in the direction in which the ground slopes away most steeply (the method of steepest descent, in fact) but there is no guarantee that this strategy will necessarily find the lowest point overall, ie the ***global minimum*** of the stress function. The rambler may reach a low point from which the ground rises in all directions (and the thus the steepest descent algorithm converges) but there may be an even lower point on the other side of an adjacent hill. He is then trapped in a ***local minimum*** of the stress function. Whether he finds the global or a local minimum depends very much on where he starts the walk, i.e. the starting configuration of points in the ordination plot.

Such local minima do occur in many MDS analyses, usually corresponding to configurations of sample points which are only slightly different from one another. Often this may be because there are one or two points which bear little relation to any of the other samples and there are several choices as to where they may be placed, or perhaps they have a more complex relationship with other samples and may be difficult to fit into (say) a 2–dimensional picture. There is no guaranteed method of ensuring that a global minimum of the stress function has been reached; the practical solution is therefore to repeat the MDS analysis several times starting with different random positions of samples in the initial configuration (step 2 above). If the same (lowest stress) solution re–appears from a number of different starts then there is a strong assurance, though never a total guarantee, that this is indeed the best solution. Note that the easiest way to determine whether the same solution has been reached as in a previous attempt is simply to check for equality of the stress values; remember that the configurations themselves could be arbitrarily rotated or reflected with respect to each other.[4] In genuine applications, converged stress values are rarely precisely the same if configurations differ materially.

Degenerate solutions can also occur, in which groups of samples collapse to the same point (even though they are not 100% similar), or to the vertices of a triangle, or are strung out round a circle. In these cases the stress may go to zero. (This is akin to our rambler starting his walk outside the encircling hills, so that he sets off in totally the wrong direction and ends up at the sea!). Artefactual solutions of this sort are relatively rare and easily detected: repetition from different random starts will find many solutions which are more sensible. (In fact, a more likely cause of an ordination in which points tend to be placed around the circumference of a circle is that the input matrix is of similarities when the program is expecting dissimilarities, or vice–versa; in such cases the stress will also be very high.) A much more common form of degenerate solution is repeatable and is a genuine result of a disjunction in the data. For example, if the data divide into two groups, which have 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.

Another feature of MDS mentioned earlier is that, unlike PCA, there is not any direct relationship between ordinations in different numbers of dimensions. In PCA, the 2–dimensional picture is just a projection of the 3–dimensional one, and all PC axes can be generated in a single analysis. With MDS, the minimisation of stress is clearly a quite different optimisation problem for each ordination of different dimensionality; indeed, this explains the greater success of MDS in distance–preservation. Samples that are in the same position with respect to (PC1, PC2) axes, though are far apart on the PC3 axis, will be projected on top of each other in a 2–dimensional PCA but they will remain separate, to some degree, in a 2–dimensional as well as a 3–dimensional MDS.

If the ultimate aim is a 2–dimensional ordination, it may still be useful to carry out a 3–dimensional MDS initially. Its first two dimensions will often provide a reasonable ***starting point*** to the iterative computations

---

4. The arbitrariness of orientation can be a practical nuisance when comparing different ordinations, and it can be helpful to rotate an MDS so that its direction of maximal variation always lies along the x axis. This can be simply achieved by applying PCA to the 2–d MDS co–ordinates (this is not the same thing as applying PCA to the original data matrix of course!); the PRIMER MDS routine does this automatically but the CONPLOT program also permits user–specified orientation/reflections.

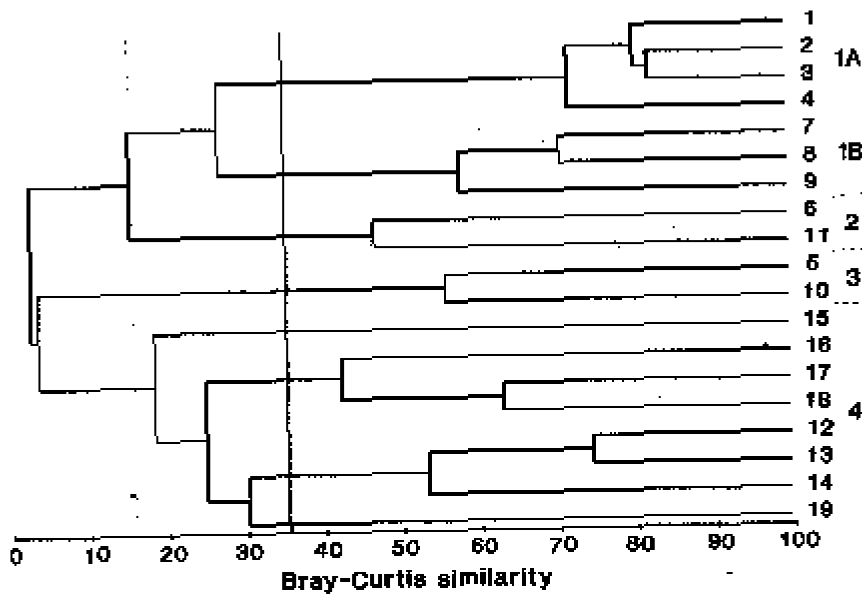


Fig. 5.3. Exe estuary nematodes (X). Dendrogram of the 19 stations, using group-average clustering from Bray–Curtis similarities on √√-transformed abundances. The four groups of stations separated at a 15% similarity threshold (dotted line) are indicated (the two tightly clustered sub-groups within group 1 were designated 1A and 1B by Field et al. 1982).

accurate placement, or simply corresponds to a major error in the data matrix.

3) Is there distortion when similar samples are connected in the ordination plot? One simple check on the success of the ordination in dissimilarity–preservation is to identify the top 10% or 20% (say) of values in the similarity matrix and draw a line between the corresponding points on the MDS configuration. An inaccurate representation is indicated if several connections are made between points which are further apart on the plot than other unconnected pairs of points.

4) Is the "minimum spanning tree" consistent with the ordination picture? A similar idea to the above is to construct the *minimum spanning tree* (MST, Gower and Ross, 1969). All samples are "connected" by a single line which is allowed to branch but does not form a closed loop, such that one minimises the sum along this line of *dissimilarities* (taken from the original dissimilarity matrix not the distance matrix from the ordination, note). This line is then plotted on the 2–dimensional ordination and inadequacy is again indicated by connections which look unnatural in the context of placement of samples in the MDS configuration.

5) Do superimposed groups from a cluster analysis distort the ordination plot? The combination of clustering and ordination analyses can be a very effective way of checking the adequacy and mutual consistency of both representations. Fig. 5.3 shows the dendrogram from a cluster analysis of the Exe estuary nematode data *(X)* of Fig. 5.1. Two or more (arbitrary) similarity values are chosen at a spread of hierarchical levels, each determining a particular grouping of samples. In Fig. 5.3, four groups are formed at around a 15% similarity level and eight groups would be determined for any similarity threshold between 30 and 45%. These two sets of groupings are superimposed on the MDS ordination, Fig. 5.4, and it is clear that the agreement between the two techniques is excellent: the clusters are sharply defined and would be determined in much the same way if one were to select clusters by eye from the 2–dimensional ordination alone. The stress for Fig. 5.4 is also low, at 0.05, giving confidence that the 2–dimensional plot is

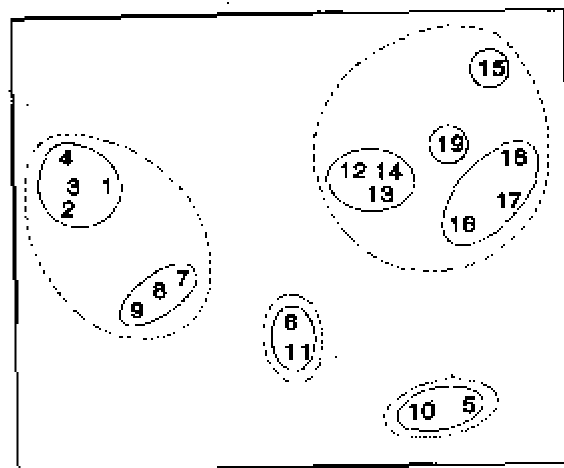


Fig. 5.4. Exe estuary nematodes (X). 2–dimensional MDS configuration, as in Fig. 5.1, with superimposed clusters from Fig. 5.3, at similarity levels of 15% (dashed line) and 30%–45% (continuous line).

2-dimensional PCA of Fig. 4.2 but with superimposed groups from a cluster analysis of the Euclidean distance matrix[10] between the 16 samples (Fig. 5.5b). With the same division into five clusters (thin lines) and ten clusters (thick lines), a much more distorted picture results, with samples that are virtually coincident in the PCA plot being placed in separate groups and samples appearing distant from each other forming a common group.

The outcome one would expect on theoretical grounds is therefore apparent in practice here: MDS can provide a more realistic picture in situations where PCA gives a distorted representation of the true "distances" between samples. In fact, the biological conclusions from this particular study are entirely negative: the test described in Chapter 6 shows that there are no statistically significant differences in community structure between any of the four dosing levels in this experiment.

EXAMPLE: Celtic Sea zooplankton

In situations where the samples are strongly grouped, as in Fig. 5.3 and 5.4, both clustering and ordination analyses will demonstrate this, usually in equally adequate fashion. The strength of ordination is in displaying a *gradation* of community composition across a set of samples. An example is provided by Fig. 5.6, of zooplankton data from the Celtic Sea *{C}*. Samples were collected from 14 depths, separately for day and night time studies at a single site. The changing community composition with depth can be traced on the resulting MDS (from Bray–Curtis similarities). There is a greater degree of variability in community structure of the near-surface samples, with a marked change in composition at about 20–25m; deeper than this the changes are steady but less pronounced and they step in parallel for day and night time samples. Another obvious feature is the strong difference in composition between day and night near-surface samples, contrasted with their relatively higher similarity at greater depth. Cluster analysis of the same data would clearly not permit the accuracy and subtlety of interpretation that is possible from ordination of a gradually changing community pattern.

10. *As previously noted, Euclidean distance is the dissimilarity measure implicit in a PCA ordination.*

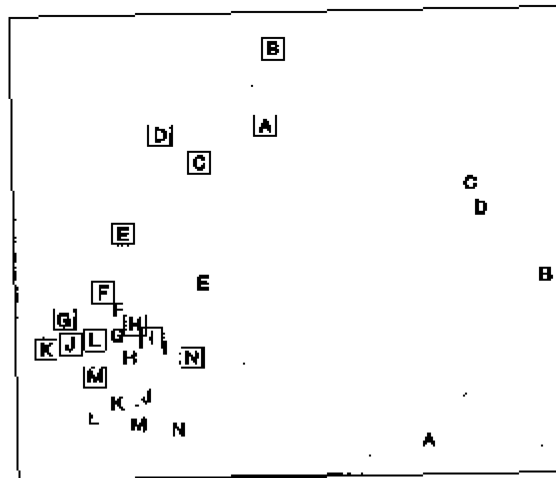


Fig. 5.6. Celtic Sea zooplankton {C}. MDS plot for night (boxed) and day time samples from 14 depths (5 to 70m, denoted A, B, ..., N), taken at a single site during September 1978.

MDS STRENGTHS

1) ***MDS is simple in concept.*** Though the numerical algorithm is undeniably complex, it is always clear what MDS is trying to achieve: the construction of a sample map whose inter-point distances have the *same rank order* as the corresponding dissimilarities between samples.

2) ***It is based on the relevant sample information.*** MDS works on the sample dissimilarity matrix not on the original data array, so there is complete freedom of choice to define similarity of community composition in whatever terms are biologically most meaningful.

3) ***Species deletions are unnecessary.*** Another advantage of starting from the sample dissimilarity matrix is that the number of species on which it was based is largely irrelevant to the amount of calculation required. Of course, if the original matrix contained a large number of species whose patterns of abundance across the samples varied widely, and prior transformation (or choice of similarity coefficient) dictated that all species were given equal weight, then the structure in the sample dissimilarities might be more difficult to represent in a low number of dimensions. More usually, the similarity measure will automatically downweight the contribution of species that are rarer (and thus more prone to random and uninterpretable fluctuations). There is then no necessity to delete species, either to obtain realistic low-dimensional

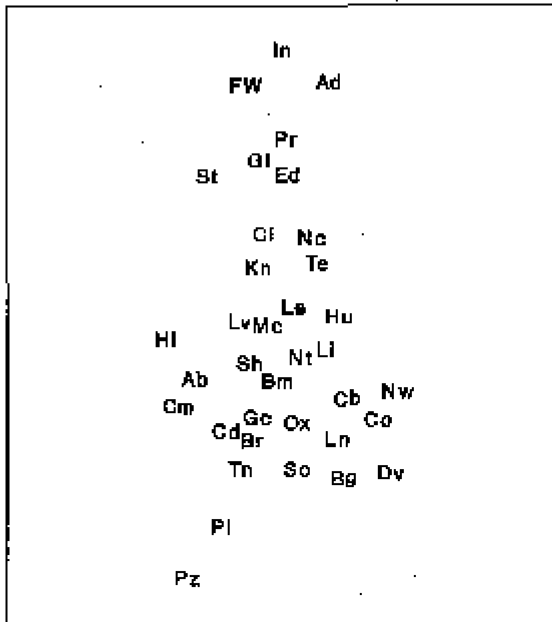


Fig. 5.7. Non–metric MDS configuration of the road distances (partly given in Table 5.2) between selected UK towns and cities (stress = 0.04).

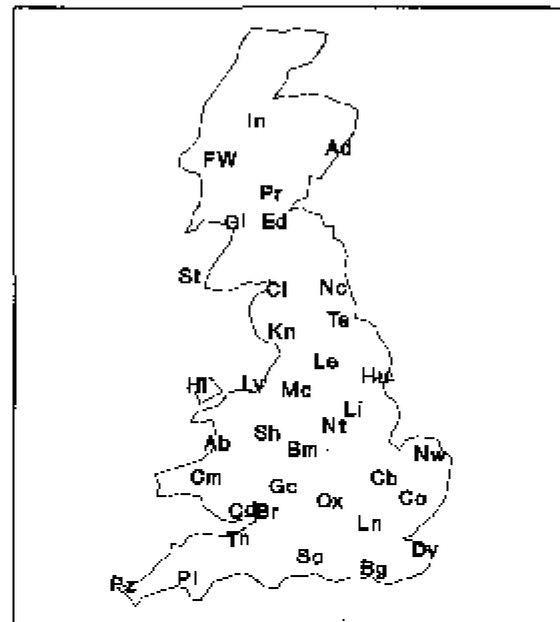


Fig. 5.8. Non–metric MDS configuration of the same towns and cities as in Table 5.2, but starting from the matrix of direct ("as the crow flies") distances between every pair (stress = 0).

5) ***Similarities can be given unequal weight.*** If some samples are inherently less reliable than others because they are based on smaller amounts of material sampled (perhaps combining the results of fewer replicates), then similarities involving these samples can be given less influence in the construction of the MDS configuration: a weighting term could be added to the definition of stress in equation (5.1). It is also true, though not of practical significance here, that the algorithm can operate perfectly successfully when the similarity matrix is subject to a certain amount of missing data.[12]

MDS WEAKNESS

1) ***MDS is computationally demanding.*** To generate a single configuration with moderate to large numbers of samples takes some time on a modern personal computer, though speed has become much less of a problem than it once was. However, MDS on much more than $n = 100$ samples is not only rather computationally intensive (processor time increases roughly proportional to n^2) but also increasing sample size generally brings increasing complexity of the sample relationships, and a 2–dimensional representation is unlikely to be adequate in any case. (Of course this last point is just as true, if not more true, for other ordination methods). This scenario was touched on in Chapter 4, where it was suggested that large data sets can often be sub–divided *a priori*, or on the basis of well–defined subsets from a cluster analysis, and the groups analysed separately by MDS. Representatives (or averages) from each group can then be input to an MDS to display the large–scale structure.

2) ***Convergence to the global minimum of stress is not guaranteed.*** As we have seen, the iterative nature of the MDS algorithm makes it necessary to repeat each analysis a number of times, from different starting configurations, to be fairly confident that a solution that re–appears several times (with the lowest observed stress) is indeed the global minimum of the stress function. Generally, with higher stress, the greater is the likelihood of

---

12. This could only be of importance if data were to arise directly as similarities constructed from pairwise comparisons of biological material, and some of those comparisons are not made or are lost. It is not of relevance if similarities are generated from a species–by–samples data matrix since, usually, either all or none of the similarities involving a particular sample can be calculated; if the latter, then there is clearly no way the sample could feature in the ordination!

CHAPTER 6: TESTING FOR DIFFERENCES BETWEEN GROUPS OF SAMPLES

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

UNIVARIATE TESTS

When the species abundance (or biomass) information in a sample is reduced to a single index, such as Shannon diversity (see Chapter 8), the existence of replicate samples from each of the groups (sites/times etc.) allows formal statistical treatment by analysis of variance (ANOVA). This requires the assumption that the univariate index is normally distributed and has constant variance across the groups, conditions which are normally not difficult to justify (perhaps after transformation, see Chapter 9). A so-called ***global test*** of the ***null hypothesis*** (H_0), that there are no differences between groups, involves computing a particular ratio of variability in the group means to variability among replicates within each group. The resulting ***F statistic*** takes values near 1 if the null hypothesis is true, larger values indicating that H_0 is false; standard tables of the F distribution yield a significance level (p) for the observed F statistic. Roughly speaking, p is interpreted as the probability that the group means we have observed (or a set of means which appear to differ from each other to an even greater extent) could have occurred if the null hypothesis H_0 is actually true.

Fig. 6.1 and Table 6.1 provide an illustration, for the 6 sites and 4 replicates per site of the Frierfjord macrofauna samples. The mean Shannon diversity for the 6 sites is seen in Fig. 6.1, and Table 6.1 shows that the F ratio is sufficiently high that the probability of observing means as disparate as this by chance is $p < 0.001$ (or $p < 0.1\%$), if the true mean diversity at all sites is the same. This is deemed to be a sufficiently unlikely chance event that the null hypothesis can safely be rejected. Convention dictates that values of $p < 5\%$ (say) are sufficiently small, *in a single test*, to discount the possibility that H_0 is true, but there is nothing sacrosanct about this figure: clearly, values of $p = 4\%$ and 6% should elicit the same inference. It is equally clear that repeated significance tests, each of which has

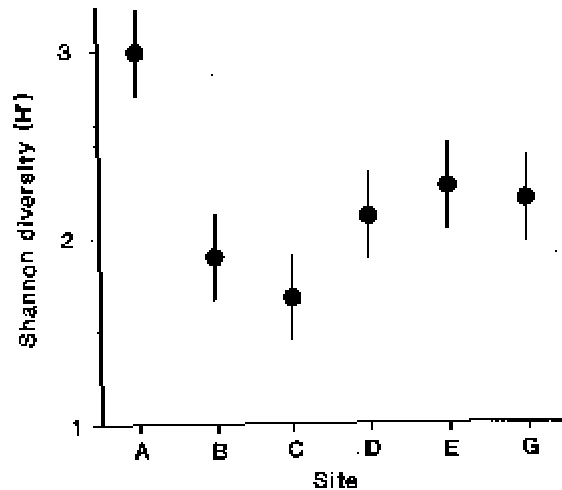


Fig. 6.1. Frierfjord macrofauna (F). Means and 95% confidence intervals of Shannon diversity (H') at the 6 field sites (A–E, G) shown in Fig. 1.1.

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

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

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

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

ANOSIM test (***analysis of similarities***)[1], by analogy with the acronym ANOVA (analysis of variance). The history of such permutation tests dates back to the epidemiological work of Mantel (1967), and this is combined with a general randomization approach to the generation of significance levels (***Monte Carlo tests***, Hope 1968). In the context below, it was described by Clarke and Green (1988).

'ANOSIM' FOR THE 1-WAY LAYOUT

Fig. 6.3 displays the MDS based only on the 12 samples (4 replicates per site) from the B, C and D sites of the Frierfjord macrofauna data. The null hypothesis (H_0) is that there are no differences in community composition at these 3 sites. In order to examine H_0, there are 3 main steps:

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

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

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

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

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

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

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

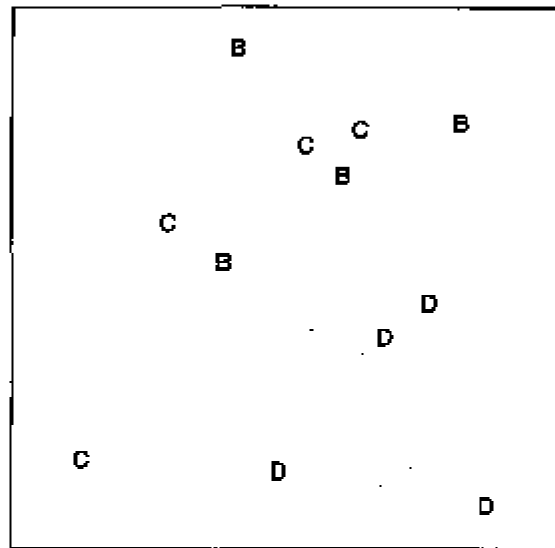


Fig. 6.3. Frierfjord macrofauna (F). MDS ordination as for Fig. 6.2 but computed only from the similarities involving sites B, C and D (stress = 0.11).

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

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

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

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

R will usually fall between 0 and 1, indicating some degree of discrimination between the sites. R substantially less than zero is unlikely since it would correspond to similarities across different sites being *higher* than those within sites; such an occurrence is more likely to indicate an incorrect labelling of samples. The R statistic itself is a useful comparative measure of the degree of separation of sites, though the main interest usually centres on whether it is

---

1. *The PRIMER program ANOSIM covers tests for replicates from 1–way <u>and</u> 2–way (nested or crossed) layouts; the program ANOSIM2 tackles the special case of a 2–way layout with <u>no</u> replication, which needs a modified style of test described at the end of this chapter.*

computed; 12% of these values are equal to or larger than 0.23 so H_0 cannot be rejected. By contrast, $R = 0.54$ for the comparison of B against D, which is the most extreme value possible under the 35 permutations. B and D are therefore inferred to be significantly different at the $p < 3\%$ level. For C against D, $R = 0.57$ similarly leads to rejection of the null hypothesis ($p < 3\%$).

There is a danger in such repeated significance tests which should be noted (although little can be done to ameliorate it here). To reject the null hypothesis at a significance level of 3% implies that a 3% risk is being run of drawing an incorrect conclusion (a *Type I error*, in statistical terminology). If many such tests are performed this risk will cumulate. For example, all pairwise comparisons between 10 sites, each with 4 replicates (allowing 3% level tests at best), would involve 45 tests, and the overall risk of drawing at least one false conclusion is high. For the analogous pairwise comparisons following the global *F* test in a univariate ANOVA, there exist ***multiple comparison*** tests which attempt to adjust for this repitition of risk. No such constructs are possible here, and the pragmatic course is to exercise appropriate caution in interpretation and/or enhance the potential significance of the individual tests by a modest increase in the number of replicates. Equation (6.2) shows that 5 replicates from each site would allow a 1% level test for a pairwise comparison (126 permutations), and 6 replicates gives close to a 0.2% level test (462 permutations); compounding these smaller values is clearly preferable to cumulating 3% risks (or the 10% Type I error, at best, from pairwise comparisons of only 3 replicates!).

This also raises the issue of *Type II error* of such a permutation test, related to its *power* to detect a difference between sites if one genuinely exists. Such concepts are not easily examined for non–parametric procedures of this type, which make no distributional assumptions and for which it is difficult to specify a precise non–null hypothesis; all that can be obviously said is that power will improve with increasing replication.

Generality of application

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

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

EXAMPLE: Indonesian reef-corals

Warwick *et al.* (1990b) examined data from 10 replicate transects across a single coral–reef site in S. Tikus Island, Thousand Islands, Indonesia, for each of the six years 1981, 1983, 1984, 1985, 1987 and 1988. The community data are in the form of % cover of a transect by each of the 58 coral species identified, and the analysis used Bray–Curtis similarities on untransformed data to obtain the MDS of Fig. 6.5. There appears to be a strong change in community pattern between 1981 and 1983 (putatively linked to the 1982/3 El Niño) and this is confirmed by a 1–way ANOSIM test for these two years alone: $R = 0.43$ ($p < 0.1\%$). Note that, though not really designed for this situation, the test is perfectly valid in the face of much greater "variability" in 1983 than 1981; in fact it is mainly a change in variability rather than location in the MDS plot that distinguishes the 1981 and 1983 groups (a point returned to in Chapter 15). This is in contrast with the standard univariate ANOVA (or multivariate MANOVA) test, which will have no power to detect a variability change; indeed it is invalid without an assumption of approximately equal variances (or variance–covariance matrices) across the groups.

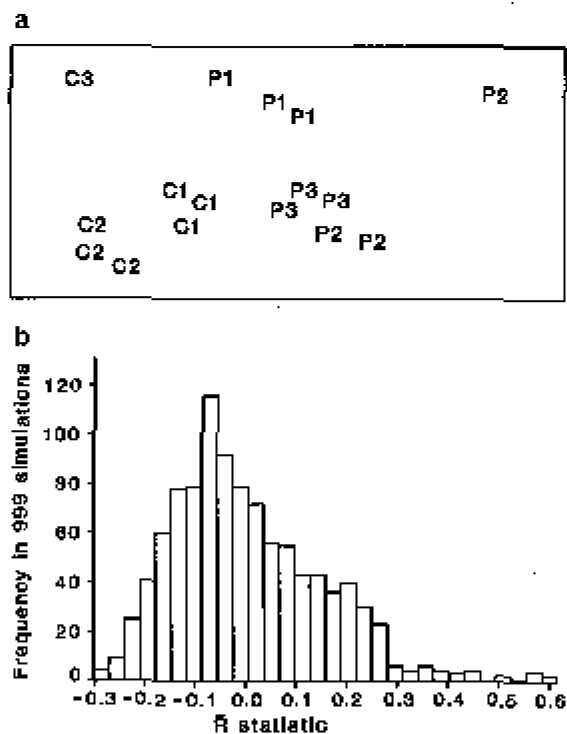


Fig. 6.6. Clyde nematodes (Y). a) MDS of species abundances from three 'polluted' (P1–P3) and three 'control' sites (C1–C3), with three replicate samples at most sites (stress = 0.09). b) Simulated distribution of the test statistic $\bar{R}$, under the hypothesis H1 of 'no site differences' within each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

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

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

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

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

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

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

An example of a two–way crossed design is given in Warwick *et al.* (1990a) and is introduced more fully here in Chapter 12. This is a so–called ***natural experiment***, studying disturbance effects on meiobenthic communities by the continual reworking of sediment by soldier crabs. Two replicate samples were taken from each of four disturbed patches of sediment, and from adjacent undisturbed areas, on a sand flat at Eaglehawk Neck, Tasmania; Fig. 6.7a is a schematic representation of the 16 sample locations. There are two factors: the presence or absence of disturbance by the crabs and the "block effect" of the four different disturbance patches. It might be anticipated that the community will change naturally across the sand flat, from block to block, and it is important to be able to separate this effect from any changes associated with the disturbance itself. There are parallels here with impact studies in which pollutants affect sections of several bays so that matched control and polluted conditions can be compared against a background of changing community pattern across a wide spatial scale. There are presumed to be replicate samples from

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

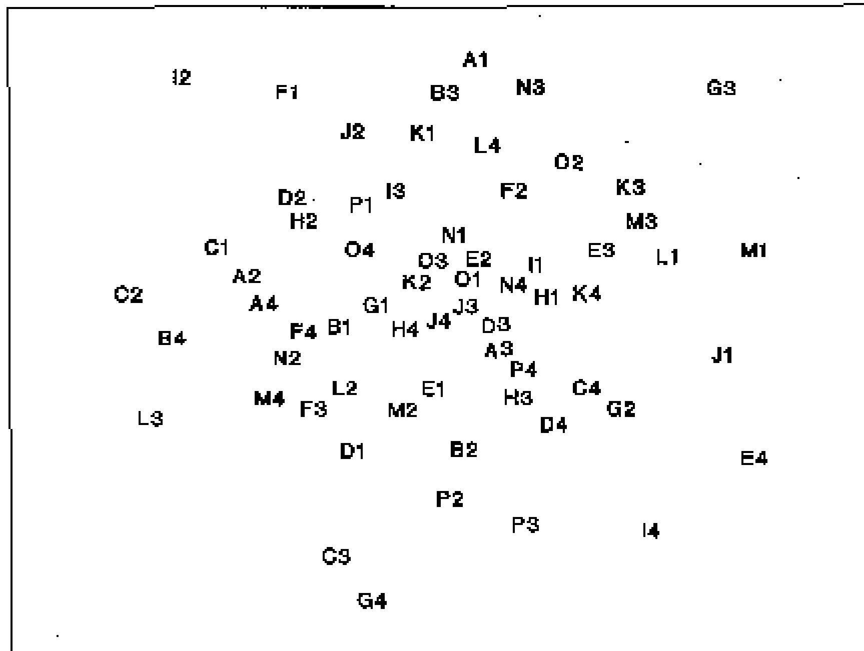


Fig. 6.8. Westerschelde nematodes experiment {W}. MDS of species abundances from 16 different nutrient–enrichment treatments, A to P, applied to sediment cores in each of four mesocosm basins, 1 to 4 (stress = 0.28).

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

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

is clearly necessary before any interpretation is attempted.

In the absence of replication, a test is still possible in the *univariate* case, under the assumption that interaction effects are small in relation to the main treatment or block differences (Scheffé, 1959). In a similar spirit, a *global* test of H_o is possible here, relying on the observation that *if* certain treatments are responsible for community changes, in a more–or–less consistent way across blocks, *separate* MDS analyses for each block should show a repeated treatment pattern. This is illustrated schematically in the top half of Fig. 6.9: the fact that treatment A is consistently close to B (and C to D) can only arise if H_o is false. The analogy with the univariate test is clear: large interaction effects imply that the treatment pattern differs from block to block and there is little chance of identifying a treatment effect; on the other hand, for a treatment x block design such as the current mesocosm experiment there is no reason to expect treatments to behave very differently in the different basins.

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

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

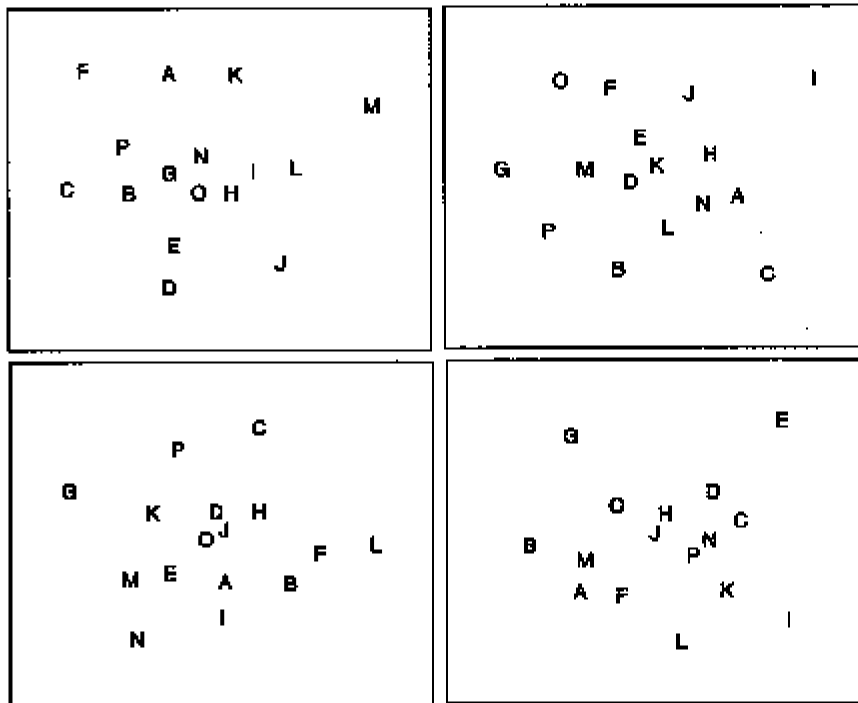


Fig. 6.10. Westerschelde nematodes experiment {W}. MDS for the 16 treatments (A to P), performed separately for each of the four basins; no shared treatment pattern is apparent (stress ranges from 0.16 to 0.20).

nematode communities at 19 sites in the Exe estuary, seen in Chapter 5. In fact, this is based on an average of data over 6 successive bi–monthly sampling occasions. For the individual times, the samples remain strongly clustered into the 4 or 5 main groups apparent from Fig. 6.11. Less clear, however, is whether any structure exists within the largest group (sites 12 to 19) or whether the scatter in Fig. 6.11 is simply the consequence of sampling variation.

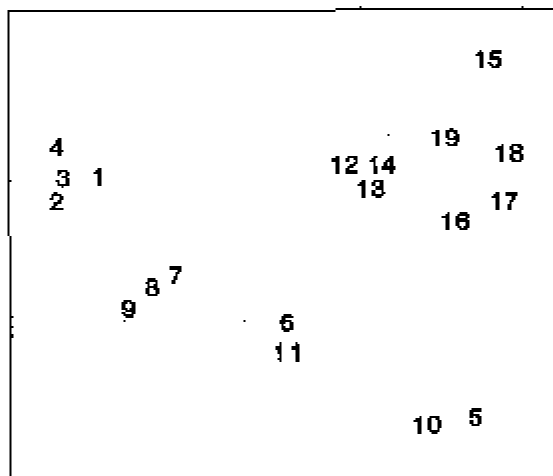


Fig. 6.11. Exe estuary nematodes {X}. MDS, for 19 inter–tidal sites, of species abundances averaged over 6 bi–monthly sampling occasions; see also Fig. 5.1 (stress = 0.05).

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

Figure 6.12 shows evidence of a consistent site pattern, for example in the proximity of sites 12 to 14 and the tendency of site 15 to be placed on its own; the fact that site 15 is missing on one occasion does not undermine this perceived structure. Pairwise computat n gives $\rho_{av} = 0.36$ and its significance can be determined by a Monte Carlo test, as before. The (non–missing) site labels are permuted amongst the available samples, separately for each time, and these designations fixed whilst all the paired ρ values are computed (using pairwise removal) and averaged. Here, the largest

CHAPTER 7: SPECIES ANALYSIS

SPECIES CLUSTERING AND MDS

Chapter 2 (page 2–6) describes how the original data matrix can be used to define similarities between every pair of *species*; two species are thought of as "similar" if their numbers (or biomass) tend to fluctuate in parallel across sites. The resulting ***species similarity matrix*** can be input to a cluster analysis or ordination in exactly the same way as for *sample* similarities.[1]

Fig. 7.1 displays the results of a cluster analysis on Exe estuary nematode data *(X)*, extensively illustrated in Chapter 5. The dendrogram is based on Bray–Curtis similarities computed on standardised abundances, as given in equations (2.9) and (2.10). Following the recommendations on page 2–6, the number of species was first reduced, retaining only those that accounted for more than 4% of the total abundance at any one site. Cluster analysis with a greater number of species is possible but the "hit–and–miss" occurrence of the rarer species across the sites tends to confuse the picture. In fact, at a similarity of around 10%, the dendrogram divides fairly neatly into 5 clusters of species, and these groups can be identified with the 5 clusters that emerge from the sample dendrogram, Fig. 5.3. (This identification comes simply from categorising the species by the site groups in which they have the greatest abundance; the correspondence between site and species groupings on this basis is seen to be very close.)

Fig 7.2 shows the 2–dimensional MDS plot of the same species similarities. The groups determined from the cluster analysis are superimposed and indicate a good measure of agreement. However, both clustering and MDS have worked well here because the sites are strongly grouped, with many species characteristic of only one site group. Typically, species cluster analyses are less clearly delineated than this and the corresponding MDS ordinations have high stress. A more informative approach is often to concentrate on the *sample* similarities and highlight the species principally responsible for determining the *sample* groupings in the cluster or ordination analyses.

1. *Computation of species similarities is an option available in the PRIMER program CLUSTER, and is referred to as inverse analysis by Field et al. (1982).*

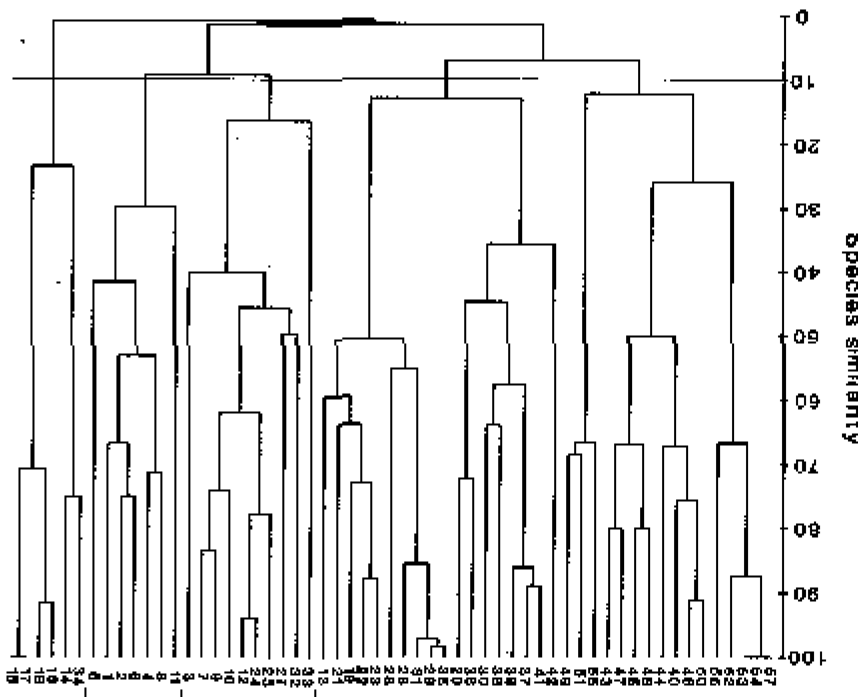
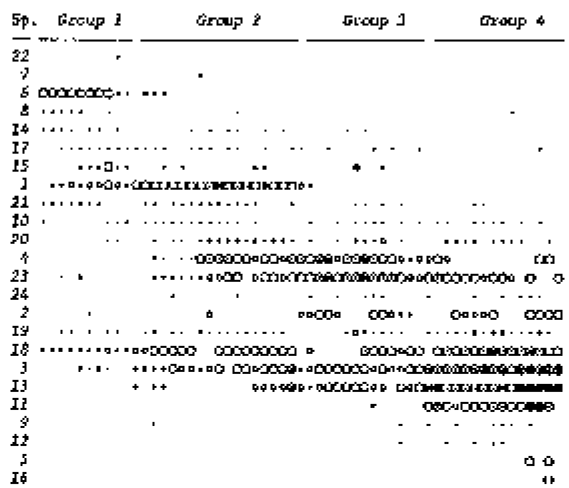


Fig. 7.1. ***Exe estuary nematodes (X).*** *Dendrogram using group–average linking on Bray–Curtis species similarities from standarized abundance data; the 57 most important species were retained from an original list of 182. The 5 groups defined at arbitrary similarity level of 10% are indicated.*



Group 1: 1,2,4,5,3,6,7,8,10,12
Group 2: 9,24,11,19,27,17,11,20,15,16,14,21,18,25,29,22,26,23
Group 3: 42,34,48,49,50,53,44,43,33,35,54,55,47,31
Group 4: 51,41,45,37,32,36,38,57,56,58,28,39,40,46,52

Fig. 7.3. ***Bristol Channel zooplankton {B}.*** *Shade matrix for the 24 species and 57 sites. The original abundances have been categorized and represented by symbols of increasing density, and the rows and columns of the array re–ordered on the basis of cluster and MDS analyses of the sites and species.*

Similarity breakdown

An alternative, more analytical way of achieving the same characterisation is to compute the ***average dissimilarity*** $\bar{\delta}$ between *all* pairs of inter–group samples (i.e. every sample in group 1 paired with every sample in group 2, say) and then break this average down into the separate ***contributions from each species*** to $\bar{\delta}$.[3]

For Bray–Curtis dissimilarity δ_{jk} between two samples j and k, the contribution from the ith species, $\delta_{jk}(i)$, could simply be defined as the ith term in the summation of equation (2.11), namely:

$$\delta_{jk}(i) = 100 . | y_{ij} - y_{ik} | / \Sigma_{i=1}^{p} \ (y_{ij} + y_{ik}) \qquad (7.1)$$

$\delta_{jk}(i)$ is then averaged over all pairs (j, k), with j in the first and k in the second group, to give the ***average contribution*** $\bar{\delta}_i$ from the ith species to the overall dissimilarity $\bar{\delta}$ between groups 1 and 2.[4] 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 $\bar{\delta}_i$ across all such pairs is the ***standard deviation*** $SD(\delta_i)$ of the $\delta_{jk}(i)$ values.[5] 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 it also does so consistently in inter–comparisons of all samples in the two groups; it is thus a good ***discriminating species.***

Table 7.1. ***Bristol Channel zooplankton {B}.*** *Breakdown of average dissimilarity between groups 1 and 2 into contributions from each species; species are ordered in decreasing contribution (part only given).*

Sp.	Name	$\bar{\delta}_i$	$SD(\delta_i)$	$\bar{\delta}_i/SD(\delta_i)$	$\Sigma\bar{\delta}_i\%$
6	*Eurytemora affinis*	7.7	2.8	2.7*	13.0
4	*Centropages hamatus*	7.3	4.4	1.7*	25.2
8	*Calanus helgolandicus*	6.8	4.0	1.7*	36.7
1	*Acartia bifilosa*	5.7	4.0	1.4*	46.3
23	*Temora longicornis*	5.6	3.3	1.7*	55.6
18	*Pseudocalanus elongatus*	4.7	1.5	3.1*	63.5
13	*Paracalanus parvus*	3.3	4.2	0.8	69.1
15	*Pleurobrachia pileus ju*	3.1	2.8	1.1	74.3
20	*Sagitta elegans ju*	2.9	1.9	1.6*	79.1
19	*Sagitta elegans*	2.1	1.6	1.3	82.5
8	*Gastrosaccus spinifer*	2.0	1.8	1.1	85.9
14	*Pleurobrachia pileus*	1.9	1.6	1.2	89.0
10	*Mesopodopsis slabberi*	1.7	1.4	1.3	91.9
21	*Schistomysis spiritus*	1.6	1.4	1.1	94.5
17	*Polychaete larvae*	1.5	1.3	1.2	97.1
2	*Acartia clausi*	0.7	1.8	0.4	98.3
.	....	...	...	...	....

For the Bristol Channel zooplankton data *{B}* of Fig. 7.3, Table 7.1 shows the results of breaking down the dissimilarities between sample groups 1 and 2 into species contributions. Species are ordered by their average contribution $\bar{\delta}_i$ to the total average dissimilarity $\bar{\delta} = \Sigma\bar{\delta}_i = 59.5$. Species which are likely to

3. This is implemented in the PRIMER program SIMPER ("similarity percentages"), both in respect of contributions to average similarity within a group and average dissimilarity between groups.

4. Though this is a natural definition, it should be noted that there is no unambiguous partition of δ_{jk} into contributions from each species, since the standardising term in the denominator of equation (7.1) is a function of all species values.

5. The usual definition of standard deviation from elementary statistics is a convenient measure of variability here, but there is no sense in which the $\delta_{jk}(i)$ values are "independent observations", and one cannot use standard statistical inference to define, say, "95% confidence intervals" for the mean contribution from the ith species.

CHAPTER 8: DIVERSITY MEASURES, DOMINANCE CURVES AND OTHER GRAPHICAL ANALYSES

UNIVARIATE MEASURES

A variety of different indices (single numbers) can be used as measures of some attribute of community structure in a sample. These include the total number of individuals (N), total number of species (S), the total biomass (B), and also ratios such as B/N (the average size of an organism in the sample) and N/S (the average number of individuals per species). These indices tend to be less informative than some measure of the way in which the total number of individuals is divided up among the different species, i.e. ***diversity indices***.

Indices of diversity and evenness

A single index of species (or higher taxon) diversity is commonly employed in community studies, and is amenable to simple statistical analysis. A bewildering variety of diversity indices has been used, and it is not appropriate here to discuss their relative merits and disadvantages. Good accounts can be found in Heip *et al.* (1988)[1] and Magurran (1991).

Two different aspects of community structure contribute to the concept of community diversity:

a) ***Species richness.*** This is a measure related to the total number of species present. Obviously we would consider a sample containing more species than another to be the more diverse.

b) ***Equitability.*** This expresses how evenly the individuals are distributed among the different species, and is often termed ***evenness***. For example, if two samples each comprising 100 individuals and four species had species abundances of 25, 25, 25, 25 and 97, 1, 1, 1, we would intuitively consider the former to be more diverse although the species richness is the same. The former has high evenness, but low ***dominance*** (essentially the reverse of evenness), while the latter has low evenness and high dominance (the sample being highly dominated by one species).

Different diversity indices may emphasize the species richness or equitability components of diversity to varying degrees. Several of these indices are included as special cases in a unified series of diversity numbers of different orders proposed by Hill (1973b).[2] However, these numbers do not as yet seem to have been widely adopted. The most commonly used diversity measure is the ***Shannon–Wiener diversity index:***

$$H' = -\Sigma_i p_i (\log p_i) \qquad (8.1)$$

where p_i is the proportion of the total count (or biomass etc) arising from the ith species.

This incorporates both the species richness and equitability components. Note that logarithms to the base 2 are often used in the calculation, giving the diversity units as 'bits per individual'. Log_e is also frequently used, so when comparing published indices it is important to check that the same logarithm base has been used in each case.

Species richness

Species richness is often given simply as the total number of species (S), which is obviously very dependent on sample size (the bigger the sample, the more species there are likely to be). More commonly ***Margalef's index*** (d) is used, which also incorporates the total number of individuals (N) and is a measure of the number of species present for a given number of individuals:

$$d = (S-1) / \log N \qquad (8.2)$$

Equitability

This is most commonly expressed as ***Pielou's evenness index:***

$$J' = H'(observed) / H'_{max} \qquad (8.3)$$

where H'_{max} is the maximum possible diversity which would be achieved if all species were equally abundant (= $\log S$).

---

1. *Although this book relates specifically to the meiobenthos, the treatment of statistical methods is applicable to all community studies.*

2. *The PRIMER program DIVERSE permits selection from a dozen or so indices, including Hill numbers and the other richness and equitability measures given here.*

Determining stress levels

Increasing levels of environmental stress have generally been considered to *decrease* diversity (e.g. H'), *decrease* species richness (e.g. d) and *decrease* evenness (e.g. J'), i.e. *increase* dominance. This interpretation may, however, be an over-simplification of the situation. More recent theories on the influence of disturbance or stress on diversity suggest that in situations where disturbance is minimal, species diversity is reduced because of competitive exclusion between species; with a slightly increased level or frequency of disturbance competition is relaxed, resulting in an increased diversity, and then at still higher or more frequent levels of disturbance species start to become eliminated by stress, so that diversity falls again. Thus it is at intermediate levels of disturbance that diversity is highest (Connell, 1978; Huston, 1979). Therefore, depending on the starting point of the community in relation to existing stress levels, increasing levels of stress (e.g. induced by pollution) may either result in an increase *or* decrease in diversity. It is difficult, if not impossible, to say at what point on this continuum the commmunity under investigation exists, or what value of diversity one might expect at that site if the community were not subjected to any anthropogenic stress. Thus, changes in diversity can only be assessed by *comparisons* between stations along a spatial contamination gradient (e.g. Fig. 8.1) or with historical data (Fig. 8.2).

Caswell's neutral model

The equitability component of diversity can, however, be compared with some theoretical expectation of diversity, given the number of individuals and species present. Observed diversity has been compared with predictions from ***Caswell's neutral model*** (Caswell, 1976). This model constructs an ecologically 'neutral' community with the same number of species and individuals as the observed community, assuming certain community assembly rules (random births/deaths and random immigrations/emigrations) and no interactions between species. The deviation statistic V is then determined which compares the observed diversity (H') with that predicted from the neutral model ($E(H')$):

$$V = \frac{[H' - E(H')]}{S.D.(H')} \qquad (8.4)$$

A value of zero for the V statistic indicates neutrality, positive values indicate greater diversity than predicted and negative values lower diversity. Values >+2 or <–2 indicate significant departures from neutrality. The computer program of Goldman & Lambshead (1989) is useful.[4]

Table 8.1 gives the V statistics for the macrobenthos and nematode component of the meiobenthos from Hamilton Harbour, Bermuda (c.f. Fig. 8.1). Note that the diversity of the macrobenthos at stations H4 and H3 is significantly below neutral model predictions, but the nematodes are close to neutrality at all stations. This indicates that the macrobenthic communities are under some kind of stress at these two stations. However, it must be borne in mind that deviation in H' from the neutral model prediction depends only on differences in equitability, since the species richness is fixed, and that the equitability component of diversity may behave differently from the species richness component in response to stress (see, for example, Fig. 8.2). Also, it is quite possible that the 'intermediate disturbance hypothesis' will have a bearing on the behaviour of V in response to disturbance, and increased disturbance may either cause it to decrease or increase. Using this method, Caswell found that the flora of tropical rain forests had a diversity below neutral model predictions!

Table 8.1. Hamilton Harbour, Bermuda (H). *V statistics for summed replicates of macrobenthos and meiobenthic nematode samples at six stations.*

Station	Macrobenthos	Nematodes
H2	+0.5	–0.1
H3	–5.4	+0.4
H4	–4.5	–0.5
H5	–1.9	0.0
H6	–1.3	–0.4
H7	–0.2	–0.4

GRAPHICAL/DISTRIBUTION PLOTS

The purpose of graphical/distributional representations is to extract information on patterns of relative species abundances without reducing that information to a single summary statistic, such as a diversity index. This class of techniques can be thought of as intermediate between *univariate* summaries and full *multivariate* analyses. Unlike multivariate methods, these distributions may extract universal features of community structure which are not a function of the

4. *This is implemented in the PRIMER program CASWELL.*

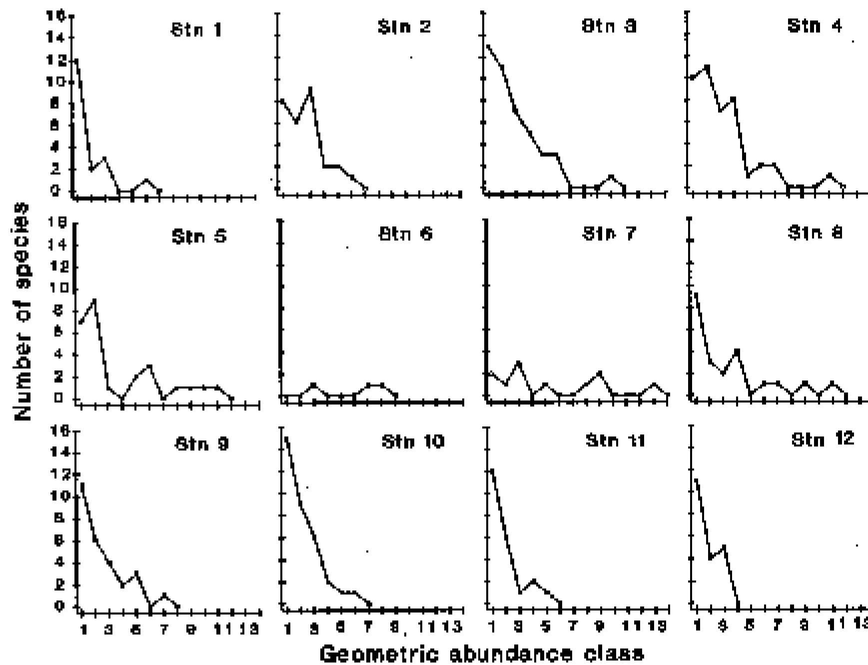


Fig. 8.4. ***Garroch Head macrofauna (G).*** *Plots of ×2 geometric species abundance classes for the 12 sampling stations shown in Fig. 8.3.*

polluted (outer) stations is much flatter, with low dominance. Fig. 8.5b shows *k*–dominance curves for the same data. Here the curve for the inner stations is elevated, indicating lower diversity than at the 250 m – 1 km stations.

Abundance/biomass comparison (ABC) plots

The advantage of distribution plots such as *k*–dominance curves is that the distribution of species abundances among individuals and the distribution of species biomasses among individuals can be compared on the same terms. Since the two have different units of measurement, this is not possible with diversity indices.

This is the basis of the ***Abundance/Biomass Comparison (ABC)*** method of determining levels of disturbance (pollution-induced or otherwise) on benthic macrofauna communities. Under stable conditions of infrequent disturbance the competitive dominants in macrobenthic communities are *K*–selected or conservative species, with the attributes of large body size and long life-span: these are rarely dominant numerically but are dominant in terms of biomass. Also present in these communities are smaller *r*–selected or opportunistic species with a short life-span, which are usually numerically dominant but do not represent a large proportion of the community biomass. When pollution perturbs a community, conservative species are less favoured and opportunistic species often become the biomass dominants as well as the numerical dominants. Thus,

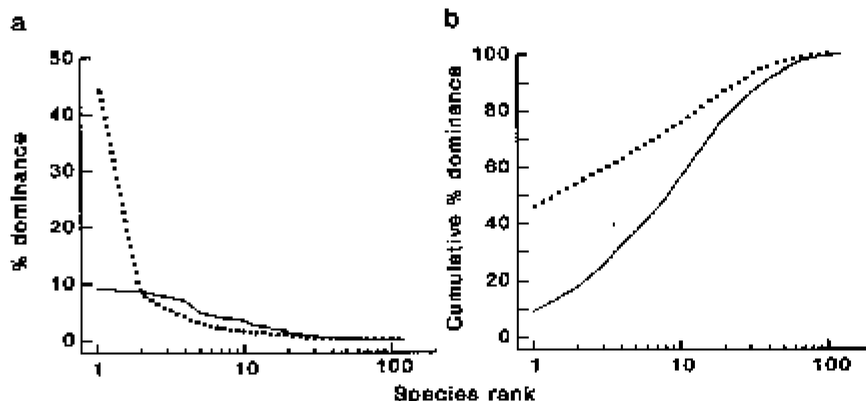


Fig. 8.5. ***Ekofisk macrobenthos (E).*** *a) Average ranked species abundance curves (x–axis logged) for 6 stations within 250 m of the centre of drilling activity (dotted line) and 10 stations between 250 m and 1 km from the centre (solid line); b) k–dominance curves for the same groups of stations.*

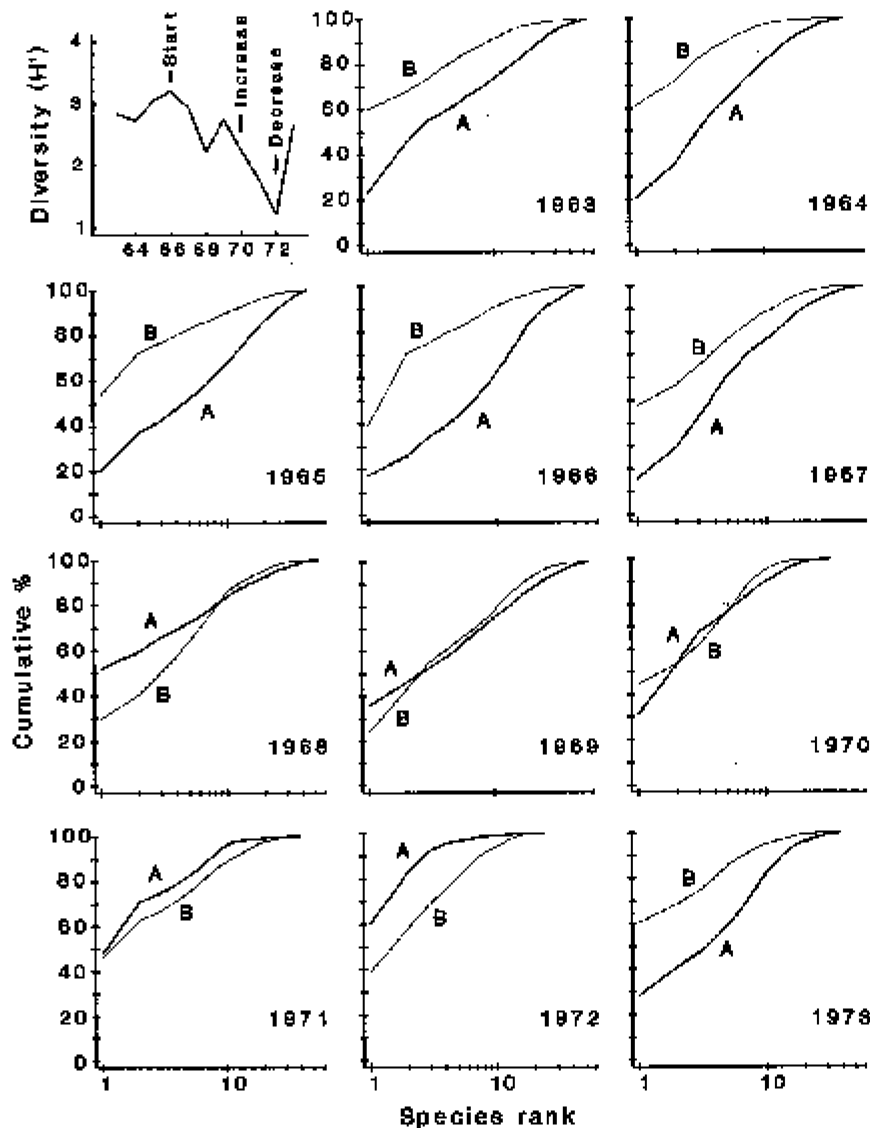


Fig. 8.7. Loch Linnhe macrofauna (L). Shannon diversity (H') and ABC plots over the 11 years, 1963 to 1973. Abundance = thick line, biomass = thin line.

method of data analysis would have indicated gross pollution. However, the biomass and abundance curves start to become transposed at some distance from the dump-centre, when species diversity is still high.

Transformations of k–dominance curves

Very often k–dominance curves approach a cumulative frequency of 100% for a large part of their length, and in highly dominated communities this may be after the first two or three top–ranked species. Thus, it may be difficult to distinguish between the forms of these curves. The solution to this problem is to transform the y–axis so that the cumulative values are closer to linearity. Clarke (1990) suggests the ***modified logistic*** transformation:

$$y_i' = \log[(1 + y_i)/(101 - y_i)] \qquad (8.5)$$

An example of the effect of this transformation on ABC curves is given in Fig. 8.9 for the macrofauna at two stations in Frierfjord, Norway *{F}*, A being an unimpacted reference site and C a potentially impacted site. At site C there is an indication that the biomass and abundance curves cross at about the tenth species, but since both curves are close to 100% at this point, the crossover is unclear. The logistic transformation enables this crossover to be better visualised, and illustrates more clearly the differences in the ABC configurations between these two sites.

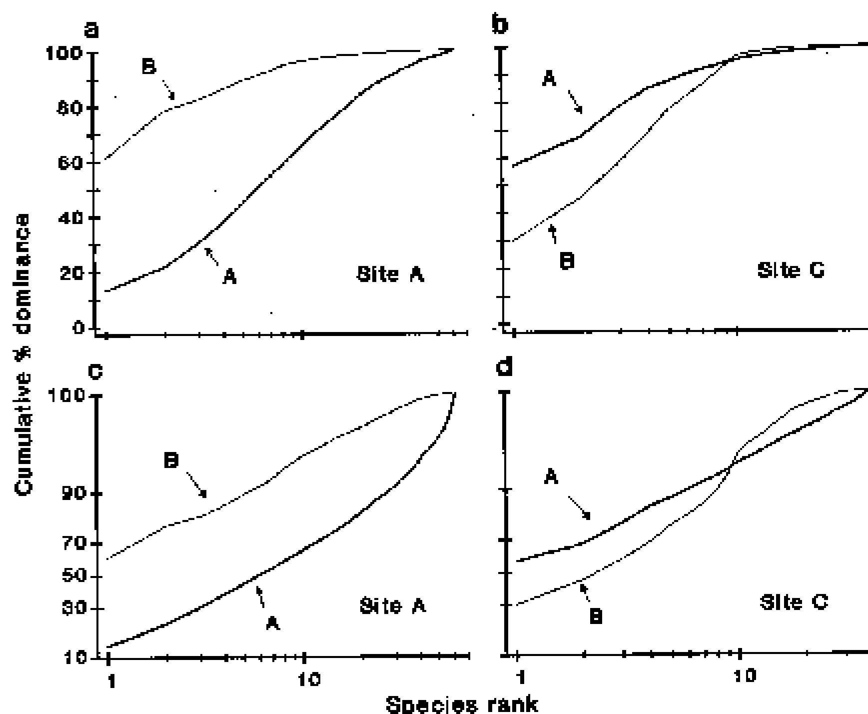


Fig. 8.9. Frierfjord macrofauna (F). a), b) Standard ABC plots for sites A (reference) and C (potentially impacted). c), d) ABC plots for sites A and C with the y-axis subjected to modified logistic transformation. Abundance = thick line, biomass = thin line.

the biomass curve, showing a slight and steady decline before the inevitable final rise.

Under polluted conditions there is still a change in position of partial dominance curves for abundance and biomass, with the abundance curve now above the biomass curve in places, and the abundance curve becoming much more variable. This implies that pollution effects are not just seen in changes to a few dominant species but are a phenomenon which pervades the complete suite of species in the community. For example, the time series of macrobenthos data from Loch Linnhe (Fig. 8.11) shows that in the most polluted years 1971 and 1972 the abundance curve is above the biomass curve for most of its length (and the abundance curve is very atypically erratic), the curves cross over in the moderately polluted years 1968 and 1970 and have an unpolluted configuration prior to the pollution impact in 1966. In 1967, there is perhaps the suggestion of incipient change in the initial *rise* in the abundance curve. Although these curves are not so smooth (and therefore not so visually appealing!) as the original ABC curves, they may provide a useful alternative aid to interpretation and are certainly more robust to random fluctuations in the abundance of a small-sized, numerically dominant species.

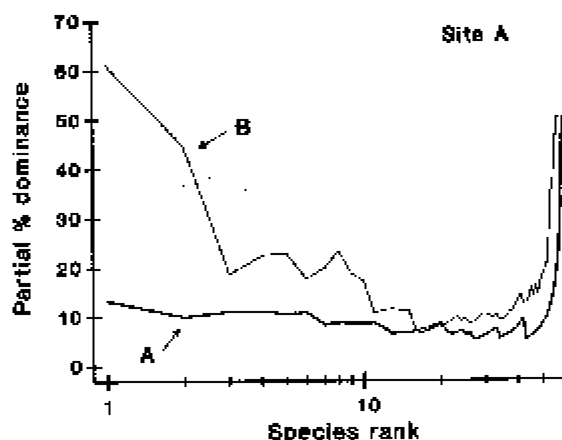


Fig. 8.10. Frierfjord macrofauna (F). Partial dominance curves (abundance/biomass comparison) for reference site A (c.f. Figs 8.9a and c for corresponding standard and transformed ABC plots).

Phyletic role in ABC method

Warwick and Clarke (1994) have recently shown that the ABC response results from (i) a shift in the proportions of different phyla present in communities, some phyla having larger-bodied species than others, and (ii) a shift in the relative distributions of abundance and biomass among species within the Polychaeta but not within any of the other major phyla

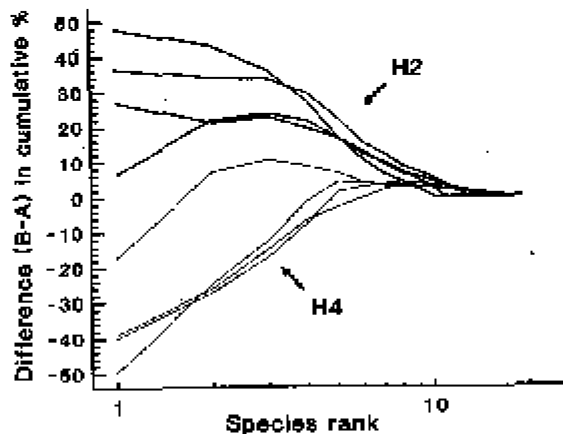


Fig. 8.12. **Hamilton Harbour macrobenthos {H}.** *Difference (B–A) between cumulative dominance curves for biomass and abundance for four replicate samples at stations H2 (thick line) and H4 (thin line).*

W statistics

When the number of sites, times or replicates is large, presenting ABC plots for *every* sample can be cumbersome, and it would be convenient to reduce each plot to a single summary statistic. Clearly, some information must be lost in such a condensation: one plots cumulative dominance curves rather than quoting a diversity index precisely because of a reluctance to reduce the diversity information to a single statistic. Nonetheless, Warwick's (1986) contention that the biomass and abundance curves increasingly overlap with moderate disturbance, and transpose altogether for the grossly disturbed condition, is a unidirectional hypothesis and very amenable to quantification by a single summary statistic.

Fig. 8.12 displays the ***difference curves*** *B–A* for each of four replicate macrofauna samples from two stations (H2 and H4) in Hamilton Harbour, Bermuda; these are simply the result of subtracting the abundance (A_i) from the biomass (B_i) value for each species rank (i) in an ABC curve.[7]

For all four replicates from H2, the biomass curve is above the abundance curve throughout its length, so the sum of the B_i–A_i values across the ranks i will be strongly positive. In contrast, this sum will be strongly negative for the replicates at H4, for which abundance and biomass curves are largely transposed. Intermediate cases in which A and B curves are intertwined will tend to give $\Sigma(B_i - A_i)$ values near zero. The summation requires some form of standardisation to a common scale, so that comparisons can be made between samples with differing numbers of species, and Clarke (1990) proposes the W (for Warwick) statistic:

$$W = \sum_{i=1}^{S} (B_i - A_i)/[50(S-1)] \qquad (8.7).$$

It can be shown algebraically that W takes values in the range (–1, 1), with $W \rightarrow +1$ for even abundance across species but biomass dominated by a single species, and $W \rightarrow -1$ in the converse case (though neither limit is likely to be attained in practice).

An example is given by the changing macrofauna communities along the transect across the sludge-dumping ground at Garroch Head {G}. Fig. 8.13 plots the W values for each of the 12 stations against the station number. These summarise the 12 component ABC plots of Fig. 8.8 and clearly delineate a similar pattern of gradual change from unpolluted to disturbed conditions, as the centre of the dumpsite is approached.

Hypothesis testing for dominance curves

There are no replicates in the Garroch Head data to allow testing for statistical significance of observed changes in ABC patterns but, for studies involving replication, the W statistic provides an obvious route to hypothesis testing. For the Bermuda samples of Fig. 8.12, W takes values 0.431, 0.253, 0.250 and 0.349 for the

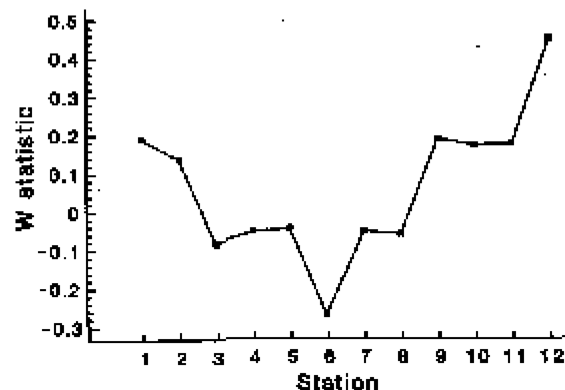


Fig. 8.13. Garroch Head macrofauna {G}. W values corresponding to the 12 ABC curves of Fig. 8.8, plotted against station number; station 6 is the centre of the dump ground (Fig. 8.3).

---

7. Note that, as always with an ABC curve, B_i and A_i do not necessarily refer to values for the <u>same</u> *species; the ranking is performed separately for abundance and biomass.*

CHAPTER 9: TRANSFORMATIONS

There are two distinct roles for transformations in community analyses:

a) to validate statistical assumptions for parametric techniques – in the approach of this manual such methods are restricted to *univariate* tests;

b) to weight the contributions of common and rare species in the (non-parametric) *multivariate* representations.

The second reason is the only one of relevance to the preceding chapters, with the exception of Chapter 8 where it was seen that standard parametric analysis of variance (ANOVA) could be applied to diversity indices computed from replicate samples at different sites or times. Being composite indices, derived from all species counts in a sample, some of these will already be approximately continuous variates with symmetric distributions, and others can be readily transformed to the normality and constant variance requirements of standard ANOVA. Also, there may be interest in the abundance patterns of individual species, specified *a priori* (e.g. keystone species), which are sufficiently common across most sites for there to be some possibility of valid parametric analysis after transformation.

UNIVARIATE CASE

For purely illustrative purposes, Table 9.1 extracts the counts of a single *Thyasira* species from the Frierfjord macrofauna data *(F)*, consisting of four replicates at each of six sites.

Table 9.1. Frierfjord macrofauna (F). Abundance of a single species (Thyasira sp.) in four replicate grabs at each of the six sites (A–E, G).

Site:	A	B	C	D	E	G
Replicate						
1	1	7	0	1	62	66
2	4	0	0	6	102	68
3	3	3	0	5	93	52
4	11	2	3	13	69	36
Mean	4.8	3.0	0.8	6.8	81.8	55.5
Stand. dev.	4.3	2.9	1.5	5.1	18.7	14.8

Two features are apparent:

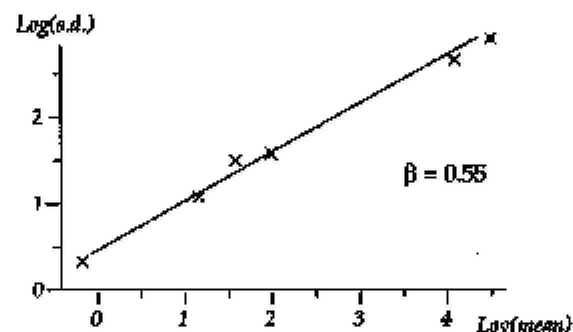
1) the replicates are not symmetrically distributed (they tend to be right–skewed);

2) the replication variance tends to increase with increasing mean, as is clear from the mean and standard deviation (*s.d.*) values given in Table 9.1.

The lack of symmetry (and thus approximate normality) of the replication distribution is probably of less importance than the large difference in variability; ANOVA relies on an assumption of constant variance across the groups. Fortunately, both defects can be overcome by a simple transformation of the raw data; a power transformation (such as a square root), or a logarithmic transformation, have the effect both of reducing right–skewness and stabilising the variance.

Power transformations

The *power transformations* $y^* = y^\lambda$ form a simple and useful family, in which decreasing values of λ produce increasingly severe transformations. The log transform, $y^* = \log_e(y)$, can also be encompassed in this series (technically, $(y^\lambda - 1)/\lambda \rightarrow \log_e y$ as $\lambda \rightarrow 0$). Box and Cox (1964) give a formal maximum likelihood procedure for optimal selection of λ but, in practice, a precise value is not important, and indeed rather artificial if one were to use slightly different values of λ for each new analysis. The aim should be to select a transformation of the right order for all data of a particular type, choosing only from, say: none, square root, 4th root or logarithmic. It is *not* necessary for a valid ANOVA that the variance be precisely stabilised or the non–normality totally removed, just that gross departures from the parametric assumptions (e.g. the order of magnitude change in s.d. in Table 9.1) are avoided. One useful technique is to plot *log s.d.* against *log mean* and estimate the approximate slope of this relationship (β). This is shown here for the data of Table 9.1.



It can be shown that, approximately, if λ is set roughly equal to 1 – β, the transformed data will have constant variance. That is, a slope of zero implies no transformation, 0.5 implies the square root, 0.75 the 4th root and 1 the log transform. Here, the square root is indicated and Table 9.2 gives the mean and standard

transform. However, in this form, the transformation is impractical because the (many) zero values produce $\log(0) \rightarrow -\infty$. Thus, common practice is to use $\log(1+y)$ rather than $\log(y)$, since $\log(1+y)$ is always positive for positive y and $\log(1+y) = 0$ for $y = 0$. The modified transformation no longer falls strictly within the power sequence; on large abundances it does produce a more severe transformation than the 4th root but for small abundances it is less severe than the 4th root. In fact, there are rarely any practical differences between cluster and ordination results performed following $y^{0.25}$ or $\log(1+y)$ transformations; they are effectively equivalent in focusing attention on patterns within the whole community, mixing contributions from both common and rare species.[1]

Table 9.4. Loch Linnhe macrofauna (L) subset. The changing similarity between samples 2 and 4 (of Table 9.3) as each of the six species is omitted in turn, for both untransformed and 4th root–transformed abundances.

Untransformed							
Species omitted:	None	1	2	3	4	5	6
Bray–Curtis (*S*):	*21*	*21*	*21*	*14*	*13*	*64*	*21*

√√–transformed							
Species omitted:	None	1	2	3	4	5	6
Bray–Curtis (*S*):	*68*	*68*	*75*	*61*	*59*	*76*	*68*

The logical end–point of this transformation sequence is therefore not the log transform but a reduction of the quantitative data to ***presence/absence***, the Bray–Curtis coefficient (say) being computed on the resulting matrix of 1's (presence) and 0's (absence). This computation is illustrated in Table 9.5 for the subset of the Loch Linnhe macrofauna data used earlier. Comparing with Table 9.3, note that the rank order of similarities again differs, though it is closer to that for the 4th root transformation than for the untransformed data. In fact, reduction to presence/absence can be thought of as the ultimate transformation in down-weighting the effects of common species. Species which are sufficiently ubiquitous to appear in *all* samples (producing a 1 in all columns) clearly cannot discriminate between the samples in any way, and thus do not contribute to the final multivariate description. The emphasis is therefore shifted firmly towards patterns in the intermediate and rarer species, the generally larger numbers of these tending to over-ride the contributions from the few numerical or biomass dominants.

Table 9.5. Loch Linnhe macrofauna (L) subset. Presence (1) or absence (0) of the six species in the four samples of Table 9.3, and the resulting Bray–Curtis similarities.

Presence/absence

Sample:	1	2	3	4
Species				
Echinoca.	*1*	*0*	*0*	*0*
Myrioche.	*1*	*0*	*0*	*1*
Labidopl.	*1*	*1*	*0*	*1*
Amaeana	*0*	*1*	*1*	*1*
Capitella	*0*	*1*	*1*	*1*
Mytilus	*0*	*0*	*0*	*0*

Sample	1	2	3	4
1	–			
2	*33*	–		
3	*0*	*80*	–	
4	*57*	*86*	*67*	–

One inevitable consequence of "widening the franchise" in this way, allowing many more species to have a say in determining the overall community pattern, is that it will become increasingly harder to obtain 2–d ordinations with low stress: the "view" we have chosen to take of the community is inherently high–dimensional. This can be seen in Fig. 9.1, for the dosing experiment *(D)* in the Solbergstrand mesocosm (GEEP Oslo workshop), previously met in Figs 4.2 and 5.5. Four levels of contaminant dosing (designated Control, Low, Medium, High) were each represented by four replicate samples of the resulting nematode communities, giving the MDS ordinations of Fig. 9.1. Note that as the severity of transformation increases, through none, root, 4th root and presence/absence (Fig. 9.1a to 9.1d respectively), the stress values rise from 0.08 to 0.19. It is important to realise that this is *not* an argument for deciding against transformation of the data. Fig. 9.1a is not a *better* representation of the between–sample relationships than the other plots: it is a *different* one. The choice of transformation is determined by which aspects of the community we wish to study. If interest is in the response of the whole community then we have to accept that it may be more difficult to capture this in a low–dimensional picture (a 3–d or higher–dimensional MDS may be desirable). On the other hand, if the data are totally dominated by one or two species, and it is these that are of key biological interest, then of course it will be possible to visualise in a 1– or 2–d picture how their numbers (or

---

1. Though practical differences are likely to be negligible, on purely theoretical grounds it could be argued that the 4th root is the more satisfactory of the two transformations because Bray–Curtis similarity is then invariant to a scale change in y. Similarity values *__would__* *be altered under a $\log(1+y)$ transformation if abundances were converted from absolute values to numbers per m^2 of the sampled substrate, or if biomass readings were converted from mg to g. This does not happen with a strict power transformation; it is clear from equation (2.1) that any multiplying constant applied to y will cancel on the top and bottom lines of the summations.*

CHAPTER 10: SPECIES REMOVAL AND AGGREGATION

SPECIES REMOVAL

For some univariate and graphical/distributional methods of data analysis it is important to include all species present at each site, since the omission of some of them will affect the outcome of the analysis. (This is obviously true for diversity measures such as species richness, for example). In certain circumstances, however, it is not possible or not advisable to include all species in multivariate analyses. There are two main circumstances where ***eliminating species*** is necessary:

a) *Sample* PCA (*not* MDS) ordinations. The species number must be reduced to (say) <50 species, or else there will be problems with computing eigenvalues (see Chapter 4).[1]

b) *Species* ordinations. Although MDS and cluster analyses are possible for all species, the rarer species, whose occurrence at a particular station may largely be due to chance, must be excluded for an interpretable outcome (see Chapter 7).[2]

The way in which species are eliminated requires careful consideration. A commonly employed method is to remove those species which are rare in respect of their total abundance at all stations in the survey, for example those species comprising less than 1 or 2% of the total number of individuals. This however can be dangerous in situations where total abundance beween stations is very variable, as is often the case. Situations frequently arise where certain stations have a very low overall abundance of organisms, but there may be many species which are absolutely characteristic of those stations. Using the above method of species reduction, all these species could be eliminated! To overcome this problem it is recommended that species accounting for >*p% of the total score* (abundance or biomass) in *any one sample* are retained (*p* is chosen to reduce species to the required number; typically *p* = 3 or 4).

---

1. As discussed in Chapter 4, PCA is not normally recommended for species data. If required, however, PRIMER can perform it by first running REDUCE, to retain only the 50 most important species (in the sense defined below), and then using SWAP to transpose the matrix for entry into the PCA routine, which expects variables as columns.

2. This could also be carried out by an initial run of REDUCE but a better option generally is to specify reduction of species (/samples) directly in CLUSTER. This will retain knowledge of the original row (/column) numbers in the derived species (/samples) similarity matrix

SPECIES REDUNDANCY

We have already seen (Chapters 4 & 5) that sample relationships can often be well summarised in a 2-dimensional ordination, which is reduced from a very much higher-dimensional species space. This implies that many species must be ***interchangeable*** in the way they characterise the samples, and that an analysis of a small subset of the total number of species may give a similar result to that for the full species analysis. This can be confirmed by performing MDS on a randomly chosen subset of species. Gray *et al.* (1988), for example, compared the configurations produced from an MDS of 110 species of macrobenthos at six stations in Frierfjord, Norway with a similar analysis using just 19 randomly selected species (Fig. 10.1). Note that the ordinations are remarkably similar in the way in which they discriminate between sites (although there is a slight difference in that the replicate samples at stations G and E are transposed in location).

Thus, there appears to be considerable ***redundancy*** in the species which characterise the community composition. Although the above example, extracting a random subset of species, is of no real practical interest, attempts have been made to exploit this redundancy in the context of taxonomic aggregation.

SPECIES AGGREGATION

The painstaking work involved in sorting and identifying samples to the species level has resulted in community analysis for environmental impact studies being traditionally regarded as labour-intensive, time-consuming and therefore relatively expensive. One practical means of overcoming this problem is to exploit the redundancy in community data by analysing the samples to higher taxonomic levels such as family or phyla, rather than to species. If results from identifications to higher taxonomic levels are comparable to a full species analysis, this means that:

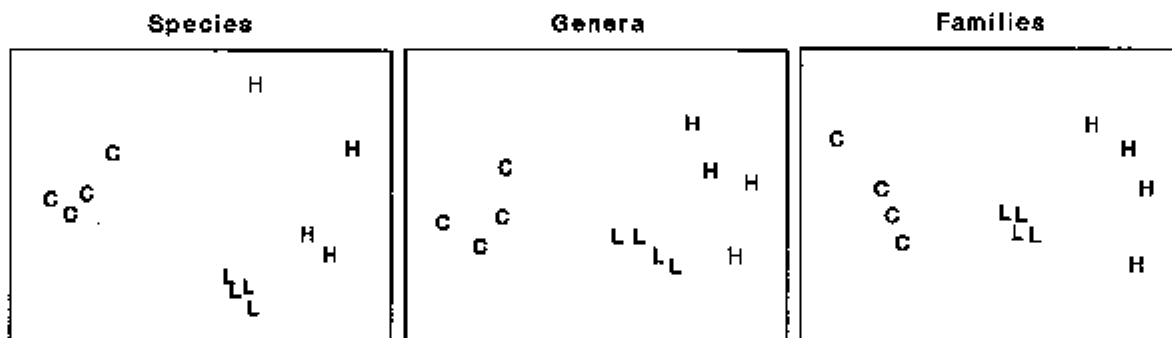


Fig. 10.2. Nutrient–enrichment experiment, Solbergstrand (N). MDS plot of copepod abundances (√√–transformed, Bray–Curtis similarities) for four replicates from each of three treatments; species data aggregated into genera and families (stress = 0.09, 0.09, 0.06).

2) ***Distributional methods.*** Aggregation for ABC curves is possible, and family level analyses are often identical to species level analyses (see Fig. 10.7).

3) ***Univariate methods.*** The concept of pollution ***indicator groups*** rather than indicator species is well–established. For example, at organically enriched sites, polychaetes of the family Capitellidae become abundant (not just *Capitella capitata*), as do meiobenthic nematodes of the family Oncholaimidae. The ***nematode/copepod ratio*** (Raffaelli and Mason, 1981) is an example of a pollution index based on higher taxonomic levels. Such indices are likely to be of more general applicability than those based on species level information. Diversity indices themselves can be defined at hierarchical taxonomic levels for internal comparative purposes, although this is not commonly done in practice.

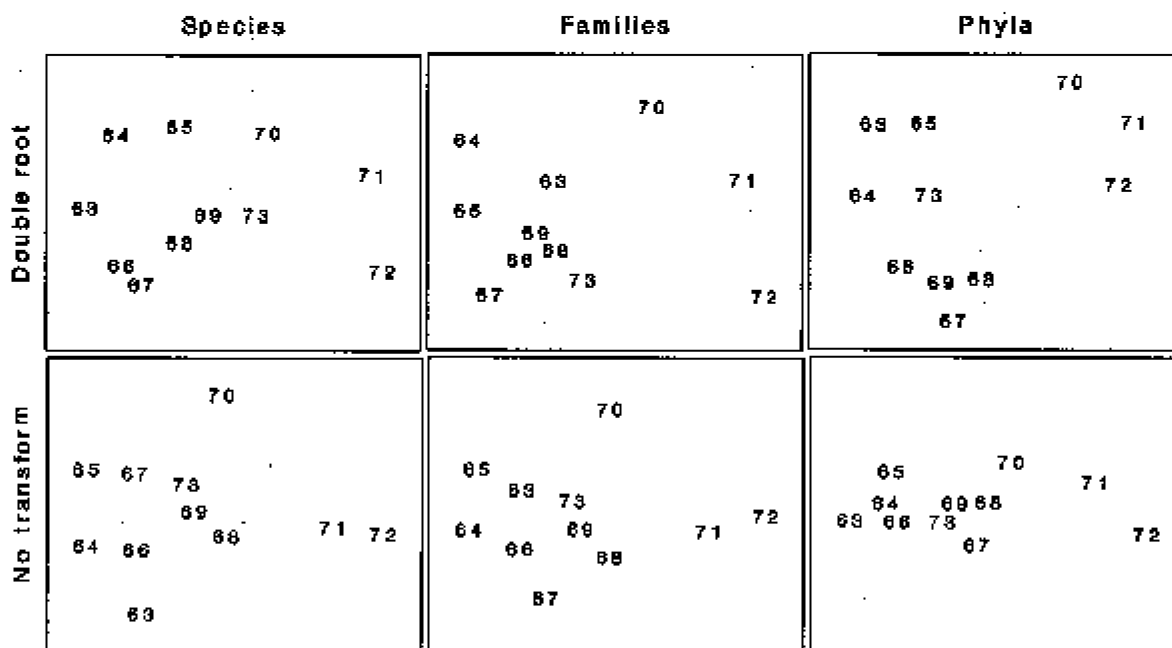


Fig. 10.3. Loch Linnhe macrofauna (L). MDS (using Bray–Curtis similarities) of samples from 11 years. Abundances are √√–transformed (top) and untransformed (bottom), with 115 species (left), aggregated into 45 families (middle) and 9 phyla (right). (Reading across rows, stress = 0.09, 0.09, 0.10, 0.09, 0.09, 0.02).

the left in 1973 associated with reduced pollution levels and community stresss. This pattern is equally clear at all levels of taxonomic aggregation. Again, the separation of the most polluted years is most distinct at the phylum level, at least for the double square root transformed data (and the configuration is more linear with respect to the pollution gradient at the phylum level for the untransformed data).

Amoco–Cadiz oil–spill

Macrofauna species were sampled at station 'Pierre Noire' in the Bay of Morlaix on 21 occasions between April 1977 and Febrary 1982, spanning the period of the wreck of the 'Amoco–Cadiz' in March 1978. The species abundance MDS has been repeated with the data aggregated into five 'phyla': Annelida, Mollusca, Arthropoda, Echinodermata and 'others' (Fig. 10.4). The analysis of phyla closely reflects the timing of pollution events, the configuration being slightly more linear than in the species analysis. All pre–spill samples (A–E) are in the top left of the configuration, the immediate post–spill sample (F) shifts abruptly to the bottom right after which there is a gradual recovery in the pre–spill direction. Note that in the species analysis, although results are similar, the immediate post–spill response is rather more gradual. The community reponse at the phylum level is remarkably sensitive, considering that the sampling site was some 40 km away from the oil–spill.

Indonesian reef corals

The El Niño of 1982–3 resulted in extensive bleaching of reef corals throughout the Pacific. Fig. 10.5 shows the coral community response at South Pari Island over six years in the period 1981–1988, based on ten replicate line transects along which coral species cover was determined. Note the immediate post–El Niño location shift on the species MDS and a circuitous

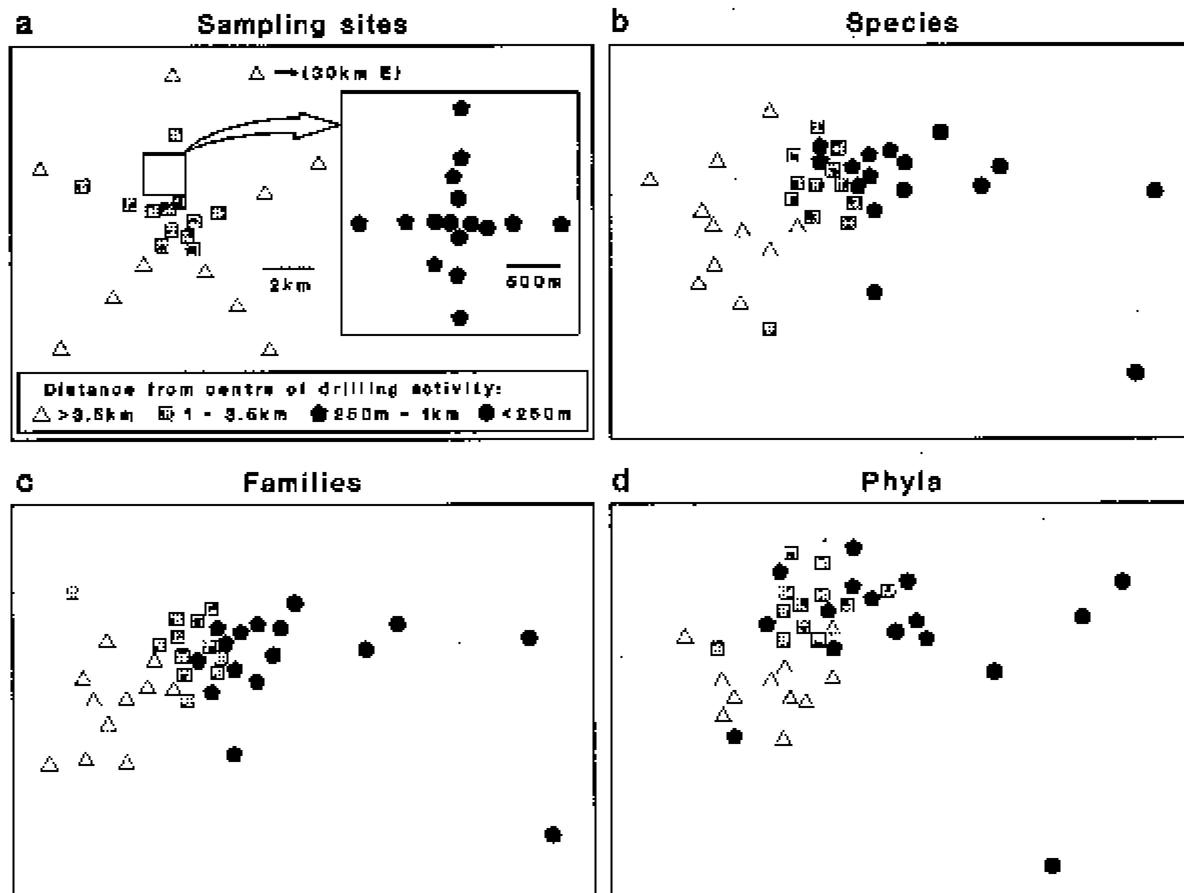


Fig. 10.6. Ekofisk oil–platform macrobenthos (E). a) Map of station positions, indicating symbol/shading conventions for distance zones from the centre of drilling activity; b)–d) MDS for root–transformed species, family and phyla abundances respectively (stress = 0.12, 0.11, 0.13).

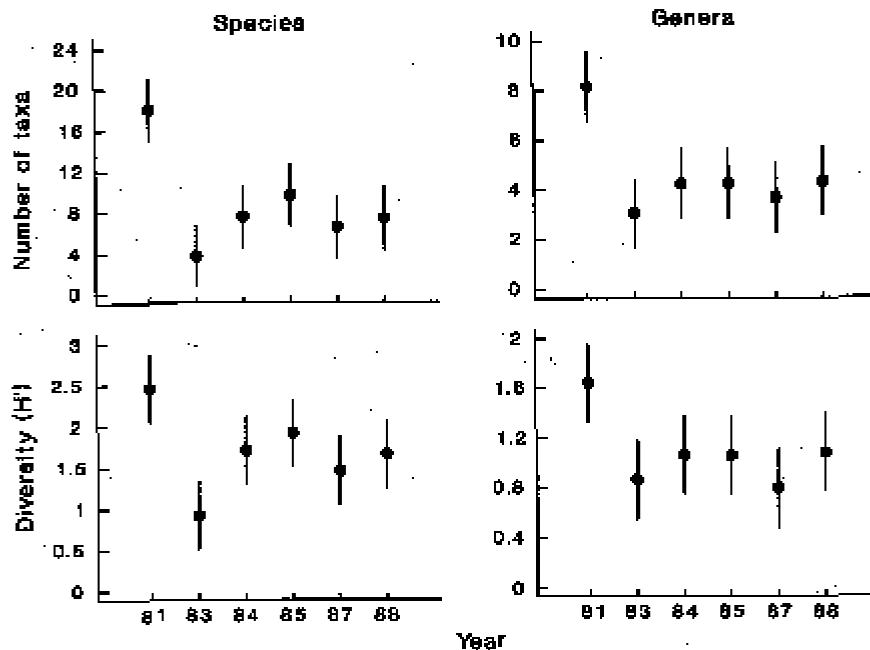


Fig. 10.8. Indonesian reef corals {I}. Means and 95% confidence intervals for number of taxa and Shannon diversity at South Tikus Island, showing the impact and partial recovery from the 1982–3 El Nino. Species data (left) have been aggregated into genera (right).

again shown to be surprisingly sensitive in detecting pollution-induced community change.

GRAPHICAL EXAMPLES

Loch Linnhe macrofauna

ABC plots for the Loch Linnhe macrobenthos species data are given in Chapter 8, Fig. 8.7, where the performance of these curves with respect to the time-course of pollution events is discussed. In Fig. 10.7 the species data are aggregated to family level, and it is seen that the curves are virtually identical to the species level analysis, so that there would have been no loss of information had the samples only been sorted originally into families.

Similar results were produced by replotting the ABC curves for the Garroch Head sewage sludge dumping ground macrobenthos *{G}* (Fig. 8.8) at the family level (Warwick, 1988b).

UNIVARIATE EXAMPLE

Indonesian reef corals

Fig. 10.8 shows results from another survey of 10 replicate line transects for coral cover over the period 1981–1988, in this case at South Tikus Island, Indonesia *{I}*. Note the similarity of the species and genus analyses for the number of taxa and Shannon diversity, with an immediate post-El Niño drop and subsequent suggestion of partial recovery.

RECOMMENDATION

Clearly the operational taxonomic level for environmental impact studies is another factor to be considered when planning such a survey, along with decisions about the number of stations to be sampled, number of replicates, types of statistical analysis to be employed etc. The choice will depend on several factors, particularly the time, manpower and expertise available and the extent to which that component of the biota being studied is known to be robust to taxonomic aggregation, for the type of statistical analysis being employed and the type of perturbation expected. Thus, it is difficult to give any firm recommendations and each case must be treated on its individual merits. For routine monitoring of organic enrichment situations using macrobenthos, one can be relatively certain that family level analysis will be perfectly adequate, but for other components of the fauna, and for other types of perturbation, sufficient evidence has not yet accumulated to be sure of this.

CHAPTER 11: LINKING COMMUNITY ANALYSES TO ENVIRONMENTAL VARIABLES

APPROACH

In many studies, the biotic data is matched by a suite of environmental variables measured at the same set of sites. These could be ***natural variables*** describing the physical properties of the substrate (or water) from which the samples were taken, e.g. median particle diameter, depth of the water column, salinity etc., or they could be ***contaminant variables*** such as sediment concentrations of heavy metals. The requirement here is to examine the extent to which the physico–chemical data is related to ("***explains***") the observed biological pattern.

The approach adopted is firstly to analyse the biotic data and then ask how well the information on environmental variables, taken either singly (Field *et al.*, 1982) or in combination (Clarke and Ainsworth, 1993), matches this community structure[1]. The motivation here, as in earlier chapters, is to retain simplicity and transparency of analysis, by letting the species and environmental data "tell their own stories" (under minimal model assumptions) before judging the extent to which one provides an "explanation" of the other.

ANALYSIS OF ENVIRONMENTAL DATA

An analogous range of multivariate methods is available for display and testing of environmental samples as has been described for faunistic data: species are simply replaced by physical/chemical variables. However, the matrix entries are now of a rather different type and lead to different analysis choices. No longer do zeros predominate; the readings are usually more nearly continuous and, though their distributions are often right–skewed (with variability increasing with the mean), it is often possible to transform them to approximate normality (and stabilise the variance) by a simple root or logarithmic transformation, see Chapter 9. Under these conditions, Euclidean distance is an appropriate measure of dissimilarity and PCA (Chapter 4) is an effective ordination technique, though note that this will need to be performed on the correlation rather than the covariance matrix, i.e. the variables will usually have different units of measurement and need normalising to a common scale (see the discussion on p4–6).

In the typical case of samples from a spatial contaminant gradient, it is also usually true that the number of variables is either much smaller than for a biotic matrix or, if a large number of chemical determinations has been made (e.g. GC/MS analysis of a range of specific aromatic hydrocarbons, PCB congeners etc.), they are often highly inter–correlated, tending to preserve a fixed relation to each other in a simple dilution model. A PCA can thus be expected to do an adequate job of representing in (say) two dimensions a pattern which is inherently low–dimensional to start with.

In a case where the samples are replicates from different groups, defined *a priori*, the ANOSIM tests of Chapter 6 are equally available for testing environmental hypotheses, e.g. establishing differences between sites, times, conditions etc., where such tests are meaningful.[2] The appropriate (rank) dissimilarity matrix would use Euclidean distances.

EXAMPLE: Garroch Head macrofauna

For the 12 sampling stations (Fig. 8.3) across the sewage–sludge dump ground at Garroch Head (G), the biotic information was supplemented by sediment chemical data on metal concentrations (Cu, Mn, Co, ...) and organic loading (% carbon and nitrogen); also recorded was the water depth at each station. The data matrix is shown in Table 11.1; it follows the normal convention in classical multivariate analysis of the

1. *Methods such as canonical correlation (e.g. Mardia et al., 1979), and the important technique of canonical correspondence (ter Braak, 1986), take the rather different stance of embedding the environmental data within the biotic analysis, motivated by specific gradient models defining the species–environment relationships.*

2. *The ANOSIM tests in the PRIMER package are not now the only possibility; the data will have been transformed to approximate normality so, if the number of variables is not large, classical multivariate (MANOVA) tests such as Wilks' Λ (e.g. Mardia et al., 1979) are valid, and will generally have greater power.*

strong pattern of incremental change on moving from the ends of the transect to the centre of the dump site, which (unsurprisingly) has the greatest levels of organic enrichment and metal concentrations (a significant exception being Mn).

LINKING BIOTA TO UNIVARIATE ENVIRONMENTAL MEASURES

Univariate community measures

If the biotic data are best summarised by one, or a few, simple univariate measures (such as diversity indices), one possibility is to attempt to correlate these with a similarly small number of environmental variables, taken one at a time. The summary provided by a principal component from a PCA of environmental variables can be exploited in this way. In the case of the Garroch Head dump ground, Fig. 11.2 shows the relation between Shannon diversity of the macrofauna samples at the 12 sites and the overall contaminant load, as reflected in the first PC of the environmental data (Fig. 11.1). Here the relationship appears to be a simple linear decrease in diversity with increasing load, and the fitted linear regression line clearly has a significantly non–zero slope ($\beta = -0.29$, $p<0.1\%$).

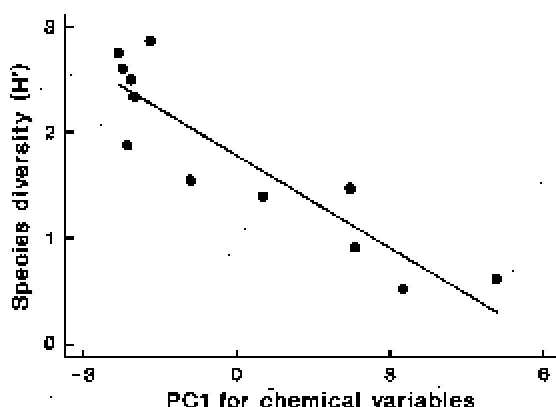


Fig. 11.2. Garroch Head macrofauna (G). Linear regression of Shannon diversity (H'), at the 12 sampling stations, against the first PC axis score from the environmental PCA of Fig. 11.1, which broadly represents an axis of increasing contaminant load (equation 11.1).

Multivariate community measures

In most cases however, the biotic data is best described by a multivariate summary, such as an MDS ordination. Its relation to a univariate environmental measure can then be visualized by representing the values of this variable as symbols of differing size and superimposing these symbols on the biotic ordination of the corresponding samples. This, or the simpler superimposition of coded values for the variable, can be an effective means of noting any consistent differences in the environmental variable between biotic clusters or observing a smooth relationship with ordination gradients (Field *et al.*, 1982).[4]

EXAMPLE: Bristol Channel zooplankton

The cluster analysis of zooplankton samples from 57 sites in the Bristol Channel (*B*) was seen in Chapter 3, and the dendrogram suggested a division of the samples into 4 or 5 main clusters (Fig. 3.3). The matching MDS (Fig. 11.3), whilst in good agreement with the cluster analysis, reveals a more informative picture of a strong gradient of change from the Inner Channel to the Celtic Sea sites. This is seen most graphically by superimposing a code representing the salinity levels for each sample (Fig. 11.4). Biological considerations suggest that a simple linear coding is not appropriate: one would expect species turnover to be much greater through a salinity differential of 1 ppt in fully saline water than the turnover from a similar 1 ppt change at (say) 25 ppt. This motivates application of a ***reverse logarithmic*** transformation, log (36–*s*), or more precisely:

$$s^* = a - b \log(36 - s) \qquad (11.2)$$

where $a = 8.33$, $b = 3$ are simple constants chosen for this data to constrain the transformed variable s^* to the range 1 (low) to 9 (high salinity). Fig. 11.4 then clearly

4. Superimposing environmental data onto an ordination is an option provided in the PRIMER program CONPLOT, which displays MDS configurations. The technique can also be useful in a wider context: Field et al. (1982) superimpose morphological characteristics of each species onto a species MDS of the type seen in Chapter 7, and Warwick and Clarke (1993a; see also Fig. 15.3) give an example of superimposition of biotic variables drawn from the same data matrix as used to create the MDS. The latter can provide insight into the role of individual taxa in shaping the biotic picture, especially when the number of taxa is small, as is the case for the phylum-level "meta-analysis" of Chapter 15.

EXAMPLE: Garroch Head macrofauna

The macrofauna samples from the 12 stations on the Garroch Head transect *{G}* lead to the MDS plot of Fig. 11.5a. For a change, this is based not on abundance but biomass values (root–transformed).[6] Earlier in the chapter, it was seen that the contaminant gradient induced a marked response in species diversity (Fig. 11.2), and there is an even more graphic representation of steady community change in the multivariate plots as the dump centre is approached (stations 1 through to 6), with gradual reversion to the original community structure on moving away from the centre (stations 6 through to 12). The correlation of the biotic pattern with particular contaminant variables is clearly illustrated by the superimposition technique introduced above; Fig. 11.5b displays the values of % carbon in the sediment (Table 11.1) as circles of varying diameter, which confirms the main axis of the biotic MDS as one of increasing organic enrichment. Several of the metal concentrations from Table 11.1 show a similar pattern, one exception being Mn, which displays a strong gradient in the other direction (Fig. 11.5c). In fact, some of the metal and organic variables are so highly correlated with each other (e.g. compare the plot for Pb in Fig. 11.5d with 11.5b) that there is little point in retaining all of them in the environmental data matrix. Clearly, when two abiotic variables are so strongly related (*collinear*), separate putative effects on the biotic structure could never be disentangled (their effects are said to be *confounded*).

6. Chapter 14 argues that, where it is available, biomass can sometimes be more biologically relevant than abundance, though in practice MDS plots from both will be broadly similar, especially under heavy transformation as the data tends towards presence/absence (Chapter 9).

EXAMPLE: Exe estuary nematodes

The Garroch Head data is an example of a smooth gradation in faunal structure reflected in a matching gradation in several contaminant variables. In contrast, the Exe estuary nematode communities {X}, discussed extensively in Chapter 5, separate into five well–defined clusters of samples (Fig. 11.6a). For each of the 19 intertidal sites, six environmental variables were also recorded: the median particle diameter of the sediment (MPD), its percentage organic content (% Org), the depth of the water table (WT) and of the blackened hydrogen sulphide layer (H_2S), the interstitial salinity (Sal) and the height of the sample on the shore, in relation to the inter–tidal range (Ht). When each of these is superimposed in turn on the biotic ordination, some instructive patterns emerge. MPD, represented appropriately by circles of differing size (Fig. 11.6b), appears to increase monotonically along the main MDS axis but cannot be responsible for the division, for example, between sites 1–4 and 7–9. On the other hand, the relation of salinity to the MDS configuration is non–monotonic (Fig. 11.6c), with larger values for the "middle" groups, but now providing a contrast between the 1–4 and 7–9 clusters. Other variables, such as the height up the shore,

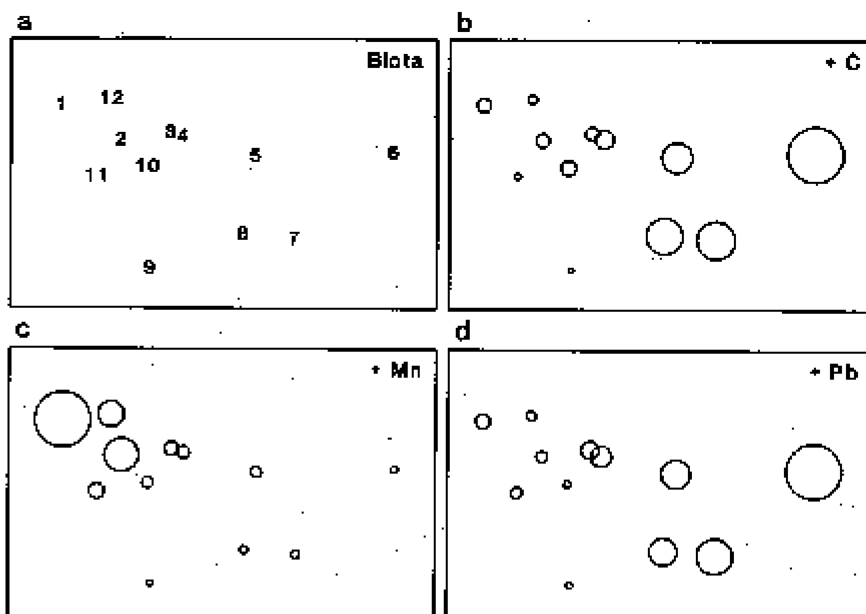


Fig. 11.5. Garroch Head macrofauna {G}. a) MDS of Bray–Curtis similarities from √–transformed species biomass data at the 12 stations (Fig. 8.3); b)–d) the same MDS but with superimposed circles of increasing size with increasing sediment concentration of C, Mn and Pb, from Table 11.1. (Stress = 0.05).

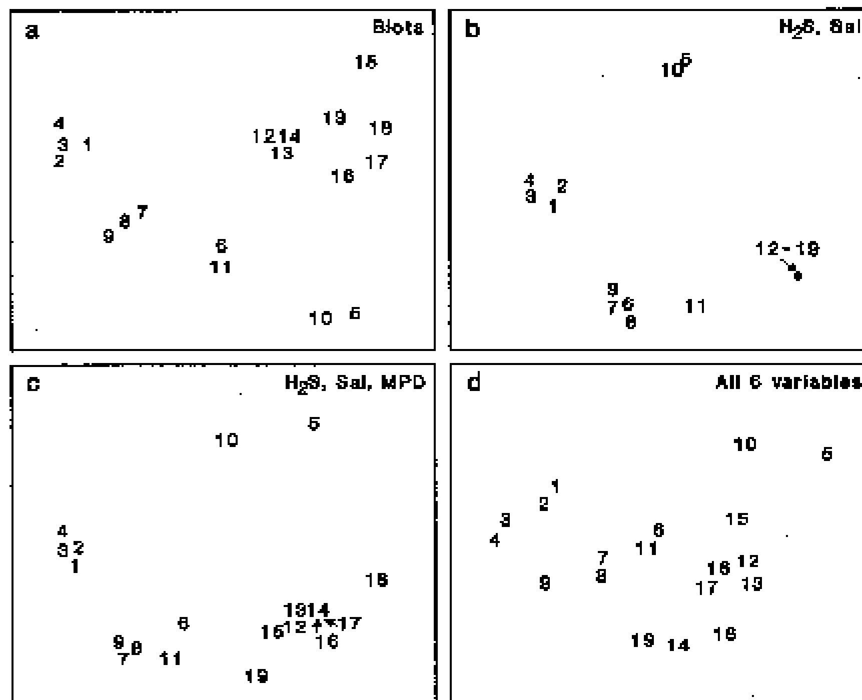


Fig. 11.7. Exe estuary nematodes {X}. MDS ordinations of the 19 sites, based on: a) species abundances, as in Fig. 5.1; b) two sediment variables, depth of the H_2S layer and interstitial salinity; c) the environmental combination "best matching" the biotic pattern: H_2S, salinity and median particle diameter; d) all six abiotic variables. (Stress = 0.05, 0, 0.04, 0.06).

Fig. 11.7b is effectively just a scatter plot, since it involves only two variables).

The point to notice here is the remarkable degree of concordance between biotic and abiotic plots, particularly Figs. 11.7a and c; both group the samples in very similar fashion. Leaving out MPD (Fig. 11.7b), 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. 11.7d), the (1–4) group becomes more widely spaced than is in keeping with the biotic plot, sample 9 is separated from 7 and 8, sample 14 split from 12 and 13 etc., and the fit again deteriorates. In fact, Fig. 11.7c represents the *best fitting* environmental combination, in the sense defined below, and therefore best "explains" the community pattern.

Measuring agreement in pattern

Quantifying the match between any two plots could be accomplished by a Procrustes analysis (Gower, 1971), in which one plot is rotated, scaled or reflected to fit the other, in such a way as to minimize a sum of squared distances between the superimposed configurations. This is not wholly consistent, however, with the approach in earlier chapters; for exactly the same reasons as advanced in deriving the ANOSIM statistic in Chapter 6, the "best match" should not be dependent on the dimensionality one happens to choose to view the two patterns. The more fundamental constructs are, as usual, the similarity matrices underlying both biotic and abiotic ordinations.[7] 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, a very natural measure to adopt bearing in mind that a successful MDS is a function only of the similarity ranks.

The procedure is summarised schematically in Fig. 11.8, and Clarke and Ainsworth (1993) describe the approach in detail. Two possible matching coefficients are defined between the (unravelled) elements of the respective rank similarity matrices $\{r_i; i = 1, ..., N\}$ and $\{s_i; i = 1, ..., N\}$, where $N = n(n-1)/2$ and n is the number of samples. These are the simple *Spearman coefficient* (e.g. Kendall, 1970):

---

7. For example, in spite of the very low stress in Fig. 11.7, a 2-d Procrustes fit of 11.7a and c will be rather poor, since the (5, 10) and (12–19) groups are interchanged between the plots. Yet, the interpretation of the two analyses is fundamentally the same (five clusters, with the (5, 10) group out on a limb etc.), and this will be fully expressed, without arbitrary dimensionality constraints, in the underlying similarity matrices.

and 6 to 9 have lower values than for samples 5, 10 and 12 to 19, with sample 11 intermediate).

The best 2–variable combination also involves depth of the H_2S layer but adds the interstitial salinity. The correlation ($\rho_w = 0.76$) is markedly better than for any other 2–variable subset, and this is the combination shown in Fig. 11.7b. The best 3–variable combination retains these two but adds the median particle diameter, and gives the overall optimum value for ρ_w of 0.80 (Fig. 11.7c); ρ_w drops slightly to 0.79 for the best 4– and higher–way combinations. The results in Table 11.2 do therefore seem to accord with the visual impressions in Fig. 11.7.[10] In this case, the first column of Table 11.2 has a hierarchical structure: the best combination at one level is always a subset of the best combination on the line below. This is not guaranteed (although it seems to happen surprisingly often) since *all* combinations have been evaluated and simply ranked.

Table 11.2 Exe estuary nematodes {X}. Combinations of the 6 environmental variables, taken k at a time, yielding the best matches of biotic and abiotic similarity matrices for each k, as measured by weighted Spearman rank correlation ρ_w; bold type indicates overall optimum. See earlier text for variable abbreviations.

k	Best variable combinations (ρ_w)				
1	H_2S (.62)	%Org (.54)	Sal (.53)	...	
2	H_2S, Sal (.76)	H_2S, MPD (.67)	H_2S, %Org (.65)	Sal, %Org (.61)	...
3	**H_2S, Sal, MPD (.80)**	H_2S, Sal, %Org (.75)	H_2S, Sal, WT (.73)	...	
4	H_2S, Sal, MPD, %Org (.79)	H_2S, Sal, MPD, Ht (.78)	...		
5	H_2S, Sal, MPD, %Org, Ht (.79)	...			
6	H_2S, Sal, MPD, %Org, Ht, WT (.77)				

10. *This will not always be the case if the 2–d faunal ordination has non–negligible stress. It is the matching of the similarity matrices which is definitive, although it would usually be a good idea to plot the abiotic ordination for the best combination at each value of k, in order to gauge the effect of a small change in ρ_w on the interpretation. Experience so far suggests that combinations giving the same value of ρ_w to two decimal places do not give rise to ordinations which are distinguishable in any practically important way, thus it is recommended that ρ_w is quoted only to this accuracy, as in Table 11.2.*

An exhaustive search over v variables involves

$$\Sigma_{k=1}^{v} \frac{v!}{k!(v-k)!} = 2^{v} - 1 \qquad (11.5)$$

combinations, i.e. 63 for the Exe estuary study, though this number quickly becomes prohibitive when v is larger than 11 or 12. Above that level, one could consider stepwise (and related) procedures which search in a more hierarchical fashion, adding and deleting variables one at a time. These are not guaranteed to find the global minimum of ρ, and run the significant risk of focussing attention on a single "best" combination when, in reality, there may be very many combinations giving an essentially similar match to the biota. In practice, it may be desirable to limit the scale of the search initially, for a number of reasons, e.g. always to include a variable known from previous experience or external information to be potentially causal. Alternatively, as discussed earlier, scatter plots of the environmental variables may demonstrate that some are highly inter–correlated and nothing in the way of improved "explanation" could be achieved by entering them all into the analysis.

An example is given by the Garroch Head macrofauna study {G}, for which the 11 abiotic variables of Table 11.1 are first transformed, to validate the use of Euclidean distances and standard product–moment correlations (page 11–2), and then examined for evidence of collinearity (page 11–5). A possible rule–of–thumb would be to reduce all subsets of (transformed) variables which have mutual correlations averaging more than about 0.95 (say) to a single representative. Here, this leaves 8 abiotic variables in the full BIO–ENV search, which results in an optimal match of the biotic pattern with C, N and Cd (ρ_w = 0.78). The corresponding ordination plots are seen in Fig. 11.9. The biotic MDS of Fig. 11.9a, though structured mainly by a single strong gradient towards the dump centre (e.g. the organic enrichment gradient seen in Fig. 11.9b), is not wholly 1–dimensional. Additional information, on a heavy metal, appears to improve the "explanation".

Design

Two final points can be made about the sampling design. The general subject of experimental and field survey design is an immense one, requiring a manual of its own[14]. It is also a problematic area for many of the (non–parametric) multivariate techniques because the lack of formal model structures makes it difficult to define *power* of statistical procedures, such as the randomisation tests described above and in Chapters 6 and 15. In the context of linking biotic and abiotic patterns, it is intuitively clear that this has the greatest prospect of success if there are a moderately large number of sample conditions, and the closest possible matching of environmental with biological data. In the case of a number of replicates from each of a number of sites, this could imply that the biotic samples, which would be well–separated in order to represent genuine variation at a site, would each have a closely–matched environmental replicate.

Another lesson of the earlier Garroch Head example is the difficulty of drawing conclusions about causality from *any* observational study. In that case, a subset of abiotic variables were so highly correlated with each other that it was desirable to omit all but one of them from the computations. There may sometimes be good external reasons for retaining a particular number of the set but, in general, one of them is chosen arbitrarily as a *proxy* for the rest. If that variable does appear to be linked to the biotic pattern then any member of the subset could be implicated, of course. More importantly, there cannot be a definitive *causal* implication here, since each retained variable is also a proxy for any potentially causal variable which correlates highly with it, but remains *unmeasured*. Clearly, in an environmental impact study, a design in which the main pollution gradient (e.g. chemical) is highly correlated with variations in some natural environmental measures (e.g. salinity, sediment structure), cannot be very informative, whether the latter variables are measured or not. A desirable strategy, particularly for the non–parametric multivariate analyses considered here, is to limit the influence of important natural variables by attempting to select sites which have the same environmental conditions but a range of contaminant impacts (including control sites[15] of course). Even then, in a purely observational study one can never entirely escape the stricture that any apparent change in community, with changing pollution impact, could be the result of an unmeasured natural variable with which the contaminant levels happen to correlate. Such issues of causality motivate the following chapter on experimental approaches.

---

14. Green (1979) provides some useful guidelines, mainly in the context of univariate analyses.

15. Note the plurality; Underwood (1992) argues persuasively that impact is best established against a baseline of site–to–site variability in control conditions.

CHAPTER 12: CAUSALITY: COMMUNITY EXPERIMENTS IN THE FIELD AND LABORATORY

In Chapter 11 we have seen how both univariate and multivariate community attributes can be correlated with natural and anthropogenic environmental variables. With careful sampling design, these methods can provide strong evidence as to which environmental variables appear to affect community structure most, but they cannot actually *prove* cause and effect. In experimental situations we can investigate the effects of a single factor (the *treatment*) on community structure, while other factors are held constant or controlled, thus establishing cause and effect. There are three main categories of experiments that can be used:

1) *'Natural experiments'*. Nature provides the treatment: i.e. we compare places or times which differ in the intensity of the environmental factor in question.
2) *Field experiments.* The experimenter provides the treatment: i.e. environmental factors (biological, chemical or physical) are manipulated in the field.
3) *Laboratory experiments.* Environmental factors are manipulated by the experimenter in laboratory mesocosms or microcosms.

The degree of 'naturalness' (hence realism) *decreases* from 1–3, but the degree of control which can be exerted over confounding environmental variables *increases* from 1–3.

In this chapter, each class of experiments is illustrated by a single example. Unfortunately all these concern the meiobenthos, since this component of the biota is very amenable to community level experiments (see Chapter 13), whereas experiments with other components of the biota have mainly been concerned with populations of individual species, rather than communities.

In all cases care should be taken to avoid *pseudoreplication,* i.e. the *treatments* should be replicated, rather than a series of 'replicate' samples taken from a single treatment (pseudoreplicates, e.g. Hurlbert, 1984). This is because other confounding variables, often unknown, may also differ between the treatments. It is also important to run experiments long enough for community changes to occur: this favours components of the fauna with short generation times (see Chapter 13).

'NATURAL EXPERIMENTS'

It is arguable whether so called *natural experiments* are actually experiments at all, and not simply well-designed field surveys, since they make comparisons of places or times which differ in the intensity of the particular environmental factor under consideration. The obvious logical flaw with this approach is that its validity rests on the assumption that places or times differ only in the intensity of the selected environmental factor (treatment); there is no possibility of *randomly allocating treatments to experimental units,* the central tool of experimentation and one that ensures that the potential effects of unmeasured, uncontrolled variables are averaged out across the experimental groups. Design is often a problem, but statistical techniques such as two-way ANOVA, e.g. Sokal and Rohlf (1981), or two-way ANOSIM (Chapter 6), may enable us to examine the treatment effect allowing for differences between sites, for example. This is illustrated in the first example below.

In some cases natural experiments may be the only possible approach for hypothesis testing in community ecology, because the attribute of community structure under consideration may result from *evolutionary mechanisms* rather than *ecological mechanisms,* and we obviously cannot conduct manipulative field or laboratory experiments over evolutionary time. One example of a community attribute which may be determined by evolutionary mechanisms relates to size spectra in marine benthic communities. Several hypotheses, some complementary and some contradictory, have been invoked to explain biomass size spectra and species size distributions in the metazoan benthos, both of which have bimodal patterns in shallow temperate shelf seas. Ecological explanations involve physical constraints of the sedimentary environment, animals needing to be small enough to move between the particles (i.e. interstitial) or big enough to burrow, with an intermediate size range capable of neither (Schwinghamer, 1981). Evolutionary explanations invoke the optimisation of two size-related sets of reproductive and feeding traits: for example small animals (meiobenthos) have direct benthic development and can be dispersed as adults, large animals (macrobenthos) have planktonic larval development and

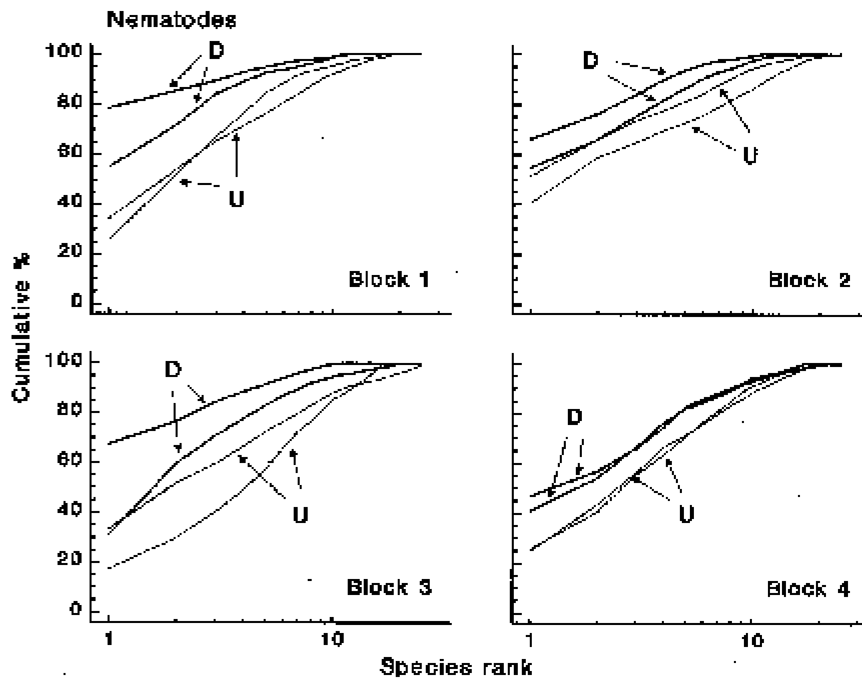


Fig. 12.2. Tasmania, Eaglehawk Neck (T). Replicate k–dominance curves for nematode abundance in each sampling block. D = disturbed, U = undisturbed.

For the nematodes, species richness, species diversity and evenness were significantly reduced in disturbed as opposed to undisturbed areas, although total abundance was unaffected. For the copepods, however, there were no significant differences in any of these univariate measures.

Graphical/distributional plots. *k*–dominance curves (Fig. 12.2) also revealed significant differences in the relative species abundance distributions for nematodes (using both the ANOVA and ANOSIM-based tests referred to briefly at the end of Chapter 8, and detailed in Clarke, 1990). For the copepods, however, (plots given in Chapter 13, Fig. 13.4), *k*–dominance curves are intermingled and crossing, and there is no significant treatment effect.

Multivariate ordinations. MDS revealed significant differences in species composition for both nematodes and copepods: the effects of crab disturbance were similar within each block and similar for nematodes and copepods. Note the similarities in Fig. 12.3 between the nematode and copepod configurations: both disturbed samples within each block are above both undisturbed (except for one block for the copepods), and the blocks are arranged in sequence

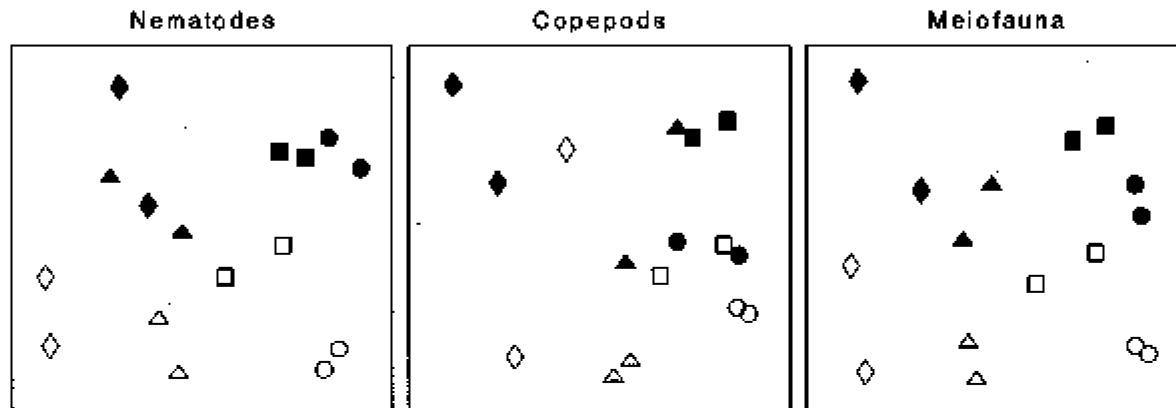


Fig. 12.3. Tasmania, Eaglehawk Neck (T). MDS configurations for nematode, copepod and 'meiofauna' (nematode + copepod) abundance (root–transformed). Different shapes represent the four blocks of samples. Open symbols = undisturbed, filled = disturbed (stress = 0.12, 0.09, 0.11 respectively).

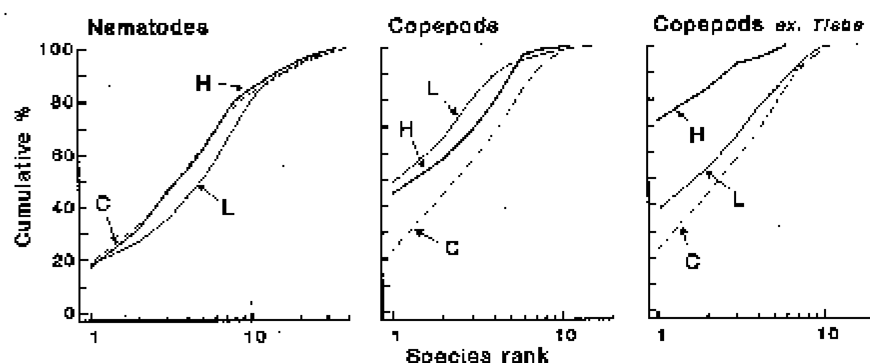


Fig. 12.5. Nutrient enrichment experiment (N). k–dominance curves for nematodes, total copepods and copepods omitting the 'weed' species of Tisbe, for summed replicates of each treatment. C = control, L = low and H = high dose.

LABORATORY EXPERIMENTS

More or less natural communities of some components of the biota can be maintained in laboratory (and also outdoor) experimental containers and subjected to a variety of manipulations. Many types of experimental systems have been used for marine studies, ranging from microcosms (containers less than 1 m^3) to mesocosms (1–1000 m^3). Macrocosms (larger than 10^3 m^3), usually involving the artificial enclosure of natural areas in the field, have also been used, but so far mainly for research on fish.

Effects of organic enrichment on meiofaunal community structure *(N)*

Gee *et al.* (1985) collected undisturbed box cores of sublittoral sediment and transferred them to the experimental mesocosms established at Solbergstrand, Oslofjord, Norway. They effected organic enrichment by the addition of powdered *Ascophyllum nodosum* in quantities equivalent to 50 g C m^{-2} (four replicate boxes) and 200 g C m^{-2} (four replicate boxes), with four undosed boxes as controls, in a randomised design within one of the large mesocosm basins. After 56 days, five small core samples of sediment were taken from each box and combined to give one sample. The structure of the meiofaunal communities in these samples was then compared.

Univariate indices. Table 12.3 shows that, for the *nematodes,* there were no significant differences in species richness or Shannon diversity between treatments, but evenness was significantly higher in enriched boxes than controls. For the *copepods,* there were significant differences in species richness and evenness between treatments, but not in diversity.

Graphical/distributional plots. Fig. 12.5 shows the average *k*–dominance curves over all four boxes in each treatment. For the *nematodes* these are closely coincident, suggesting no obvious treatment effect. For the copepods, however, there are apparent differences between the curves. A feature of the copepod assemblages in the enriched boxes was the presence, in highly variable numbers, of several species of the large epibenthic harpacticoid *Tisbe,* which are 'weed' species often found in old aquaria and associated with organic enrichment. If this genus is omitted from the analysis, a clear sequence of increasing elevation of the *k*–dominance curves is evident from control to high dose boxes.

Table 12.3. Nutrient–enrichment experiment (N). Univariate measures for all replicates at the end of the experiment, with the F–ratio and significance levels from one–way ANOVA.

	Species richness (*d*)	Shannon diversity (*H'*)	Species evenness (*J'*)
Nematodes			
Control	3.02	2.25	0.750
	3.74	2.39	0.774
	3.36	2.47	0.824
	4.59	2.76	0.747
Low dose	4.39	2.86	0.877
	2.65	2.47	0.840
	4.67	2.89	0.875
	2.33	2.27	0.860
High dose	2.86	2.17	0.782
	2.82	2.39	0.843
	4.30	2.40	0.829
	4.09	2.47	0.853
F ratio	0.04	1.39	5.13
Significance (p)	*ns*	*ns*	<5%
Copepods			
Control	2.53	1.93	0.927
	1.92	1.56	0.969
	2.50	1.77	0.908
	2.47	1.94	0.931
Low dose	1.80	1.60	0.643
	1.66	1.28	0.532
	1.66	1.16	0.484
	1.79	1.54	0.640
High dose	1.75	1.59	0.767
	0.97	1.00	0.620
	1.03	0.30	0.165
	1.18	1.70	0.872
F ratio	17.72	2.65	4.56
Significance (p)	<0.1%	*ns*	<5%

CHAPTER 13: DATA REQUIREMENTS FOR BIOLOGICAL EFFECTS STUDIES: WHICH COMPONENTS AND ATTRIBUTES OF THE BIOTA TO EXAMINE?

COMPONENTS

The biological effects of pollutants can be studied on assemblages of a wide variety of organisms:

Pelagos
- plankton (both phytoplankton and zooplankton)
- fish (pelagic and demersal)

Benthos (soft–bottom)
- macrobenthos
- meiobenthos
- (microbenthos, not much used for community studies)

Benthos (hard–bottom)
- epifauna (encrusting forms, eg. corals)
- motile fauna (both macrofauna and meiofauna in e.g. algae, holdfasts and epifauna)

These various components of the biota each have certain practical and conceptual advantages and disadvantages for use in biological effects studies. These are discussed in this chapter, and an example is given for each of the components (although not all of these examples are directly concerned with pollution effects).

PLANKTON

The *advantages* of plankton are that:

a) Long tows over relatively large distances result in community samples which reflect integrated ecological conditions over large areas. They are therefore useful in monitoring more global changes.

b) Identification of macro–planktonic organisms is moderately easy, because of the ready availability of appropriate literature.

The *disadvantage* of plankton is that, because the water masses in which they are suspended are continually mobile, they are not useful for monitoring the local effects of a particular pollutant source.

Example: Continuous Plankton Recorder

Plankton samples have been collected from 'ships of opportunity' plying their usual commercial routes across the NE Atlantic since the late 1940s (Colebrook, 1986). The plankton recorders collect samples through a small aperture, and these are trapped on a continuously winding roll of silk so that each section of silk contains an integrated sample from a relatively large area. This has enabled long term trends in plankton abundance to be assessed: there has been a gradual decline in both zooplankton and phytoplankton since the early 1950s, with an upturn in the 1980s (Fig. 13.1).

FISH

The *advantages* of fish are that:

a) Because of their mobility they are again more useful for studying general rather than local effects, but some demersal fish communities may show site fidelity, such as the coral–reef fish in the example below.

b) The taxonomy of fish is relatively easy, at least in Europe and N. America.

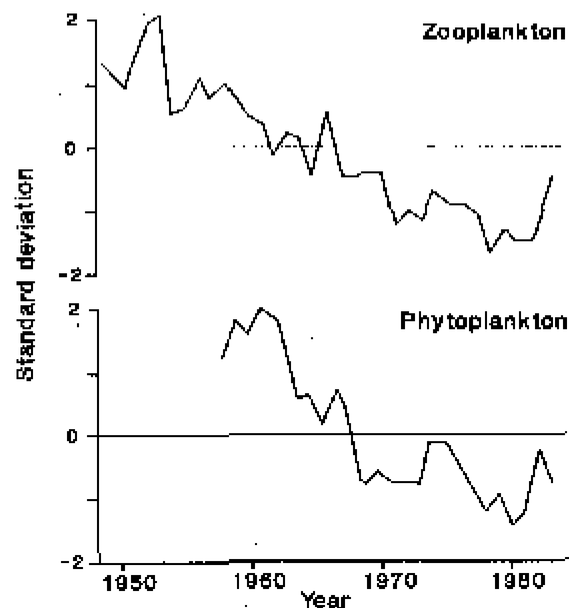


Fig. 13.1. Continuous Plankton Recorder Survey of the NE Atlantic (P). First principal components for zooplankton and phytoplankton, over the years of the survey (from Colebrook, 1986). Graphs scaled to zero mean and unit variance.

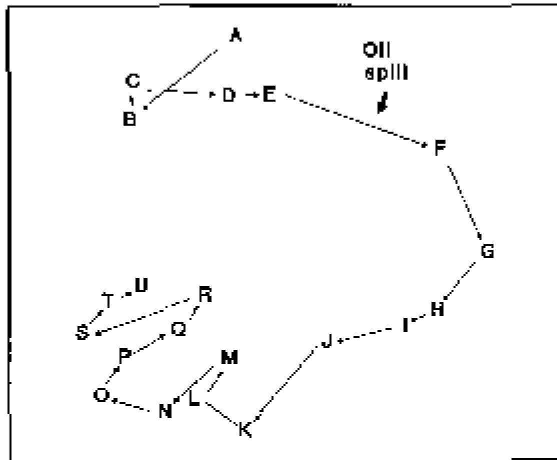


Fig. 13.3. Amoco–Cadiz oil spill, Bay of Morlaix (A). MDS for macrobenthos at station "Pierre Noire", at approximately 3–monthly sampling intervals (stress = 0.09).

a) Because of their small size and high density in marine sediments, quantitative sampling of the meiobenthos is easy from small ships, open boats etc.

b) The small volume of the samples means that they can easily be transported to the laboratory, and need not be processed on board ship.

c) Their generation times are usually measured in months rather than years, so that their potential response time to pollution events is much faster than that of the macrobenthos.

d) Because of this fast response time, and direct benthic rather than planktonic development, the meiobenthos are good candidates for causality experiments in experimental microcosms and mesocosms.

The *disadvantages* of meiobenthos are that:

a) Their taxonomy is considered difficult. Identification of almost all the meiobenthic taxa to species level presents difficulties even in Europe and North America, and in many parts of the world the fauna is almost completely unknown. However, three factors mitigate to a considerable degree against this problem:

i. The robustness of community analyses to the use of taxonomic levels higher than species (see Chapter 10).

ii. The cosmopolitan nature of most meiobenthic *genera*.

iii. The increasing availability of easily used keys to meiobenthic genera. For example, the pictorial keys to marine nematodes of Platt and Warwick (1988) have been used successfully worldwide.

b) Community responses of the meiobenthos to pollution are not well documented, so that there is not an extensive body of information in the literature against which particular case–histories can be evaluated.

Example: Soldier crab disturbance of nematode assemblages, Tasmania

This natural field experiment was described in Chapter 12. It will be remembered that the nematode diversity profiles were affected by the crab disturbance (Fig. 12.2), whereas no significant effect was noted for copepods (Fig. 13.4). Many nematode species are more sedentary in habit than copepods, often adhering to sand–grains by secretions from their caudal glands, and some species prefer conditions of low oxygen concentration or are obligate anaerobes. The so called 'thiobiotic' meiofaunal community contains many nematode species, but apparently no copepods. Non–bioturbated sediments will have a vertical gradient in physical and chemical conditions ranging from wave–disturbed sediments with an oxiphilic meiofauna community near the surface to a stable sediment with a thiobiotic community deeper down. Dramatic disturbance by crabs, of the kind found at this site, will inevitably destroy this gradient, so that the whole sediment column will be well aerated and unstable. This reduction in habitat complexity is probably the most parsimonious explanation for the reduction in nematode species diversity.

The differential response of these two components of the meiobenthos has been elaborated here in order to demonstrate how a knowledge of the biology of these components can aid in the interpretation of community responses to perturbation. The macrobenthos and meiobenthos may also respond differently to different kinds of perturbation (e.g. physical disturbance, "pollution") so that a comparative study of both may be indicative of the cause.

Example: Macrobenthos and meiobenthos in Hamilton Harbour, Bermuda

Fig. 13.5 shows the average k–dominance curves for the macrobenthos and the nematode component of the meiobenthos at six stations in Hamilton Harbour. For the macrobenthos, the curves at three of the stations (H3, H4 & H6) are much more elevated than the other three, suggesting some kind of perturbation at these sites. For the nematodes, however, all curves are closely coincident. There must therefore be some form

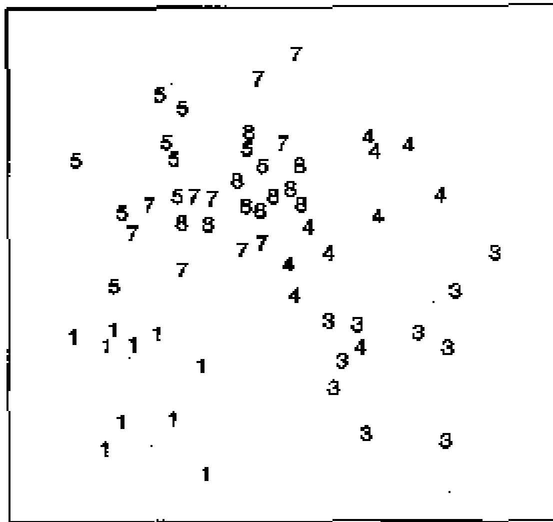


Fig. 13.6. Indonesian reef–corals (I). MDS for coral species percentage cover data for South Pari Island (10 replicate transects in each year). 1=1981, 3=1983 etc. (stress = 0.25).

is seen in community composition between 1981 and 1983, with a more steady pattern of change thereafter, though without full reversion to the initial state.

HARD-BOTTOM MOTILE FAUNA

Hard–bottom motile fauna

The motile fauna living on rocky substrates and associated with algae, holdfasts, hydroids etc. has rarely been used in pollution impact studies because of its many *disadvantages*:

a) Remote sampling is difficult.

b) Quantititative extraction from the substrate, and comparative quantification of abundances between different substrate types, are difficult.

c) Responses to perturbation are largely unknown.

d) A suitable habitat (e.g. algae) is not always available. A solution to this problem, and also problem (b), might be to deploy standardised artificial substrates, e.g. plastic mesh pan-scrubbers, along suspected pollution gradients in the field, allowing these to become colonised.

Example: Metazoan fauna of intertidal seaweed samples from the Isles of Scilly

The entire metazoan fauna (macrofauna + meiofauna) was examined from five species of intertidal macro–algae (*Chondrus, Laurencia, Lomentaria, Cladophora, Polysiphonia*) each collected at eight sites near low water from rocky shores on the Isles of Scilly, U.K. (Gee and Warwick, 1994). The MDS plots for meiobenthos and macrobenthos were very similar, with the algal species showing very similar relationships to each other in terms of their meiofaunal and macrofaunal community structure (Fig. 13.7). The structure of the weed therefore clearly influenced community structure in both these components of the benthic fauna.

ATTRIBUTES

Species *abundance* data are by far the most commonly used in environmental impact studies at the community level. However, the abundance of a species is perhaps the least ecologically relevant measure of its relative importance in a community, and we have already seen in Chapter 10 that higher taxonomic

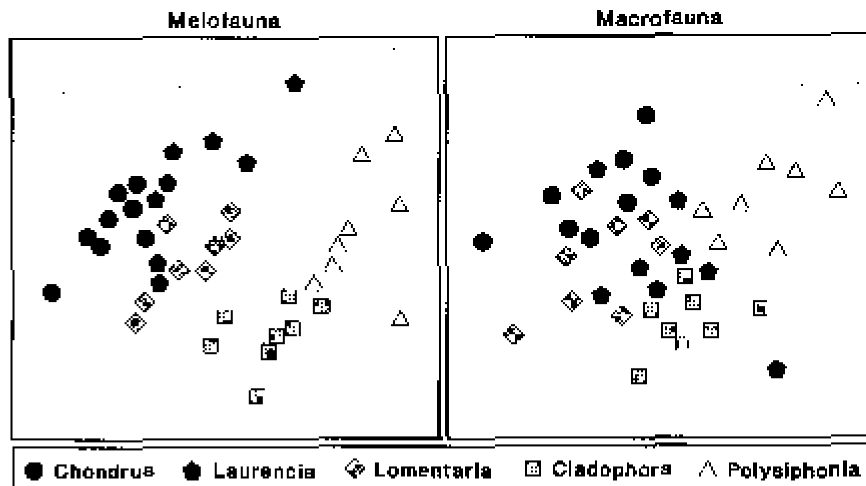


Fig. 13.7. Isles of Scilly seaweed fauna (S). MDS of standardised √√–transformed meiofauna and macrofauna species abundance data. The five seaweed species are indicated by different symbol and shading conventions (stress = 0.19, 0.18).

CHAPTER 14: RELATIVE SENSITIVITIES AND MERITS OF UNIVARIATE, GRAPHICAL/DISTRIBUTIONAL AND MULTIVARIATE TECHNIQUES

Two communities with a completely different taxonomic composition may have identical univariate or graphical/distributional structure, and conversely those comprising the same species may have very different univariate or graphical structure. This chapter compares univariate, graphical and multivariate methods of data analysis by applying them to a broad range of studies on various components of the marine biota from a variety of localities, in order to address the question of whether species dependent and species independent attributes of community structure behave the same or differently in response to environmental changes, and which are the most sensitive. Within each class of methods we have seen in previous chapters that there is a very wide variety of different techniques employed, and to make this comparative exercise more tractable we have chosen to examine only one method for each class:

Shannon–Wiener diversity index H' (see Chapter 8),

k–dominance curves including ABC plots (Chapter 8),

non–metric MDS ordination on a Bray–Curtis similarity matrix of appropriately transformed species abundance or biomass data (Chapter 5).

EXAMPLE 1: Macrobenthos from Frierfjord/Langesundfjord, Norway

As part of the GEEP/IOC Oslo Workshop, macrobenthos samples were collected at a series of six stations in Frierfjord/Langesundfjord *(F)*, station A being the outermost and station G the innermost (station F was not sampled for macrobenthos). For a map of the sampling locations see Fig. 1.1.

Univariate indices

Site A had a higher species diversity and site C the lowest but the others were not significantly different (Fig. 14.1).

Graphical/distributional plots

ABC plots indicated that stations C, D and E were most stressed, B was moderately stressed, and A and G were unstressed (Fig. 14.2).

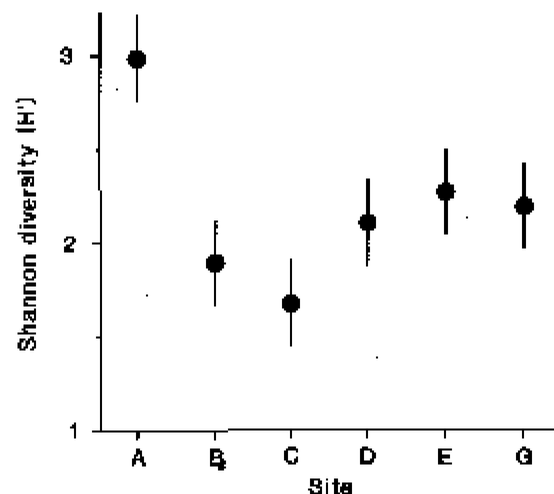


Fig. 14.1. Frierfjord macrobenthos (F). Shannon diversity (mean and 95% confidence intervals) for each station.

Multivariate analysis

An MDS of all 24 samples (4 replicates at each station), supported by the ANOSIM test, showed that only stations B and C were not significantly different from each other (Fig. 14.3). Gray *et al.* (1988) show that the clusters correlate with water depth rather than with measured levels of anthropogenic variables such as hydrocarbons or metals.

Conclusions

The MDS was much better at discriminating between stations than the diversity measure, but perhaps more importantly, sites with similar univariate or graphical/distributional community structure did not cluster together on the MDS. For example, diversity at E was not significantly different from D but they are furthest apart on the MDS; conversely, E and G had different ABC plots but clustered together. However, B, C and D all have low diversity and the ABC plots indicate disturbance at these stations. The most likely explanation is that these deep–water stations are affected by seasonal anoxia, rather than anthropogenic pollution.

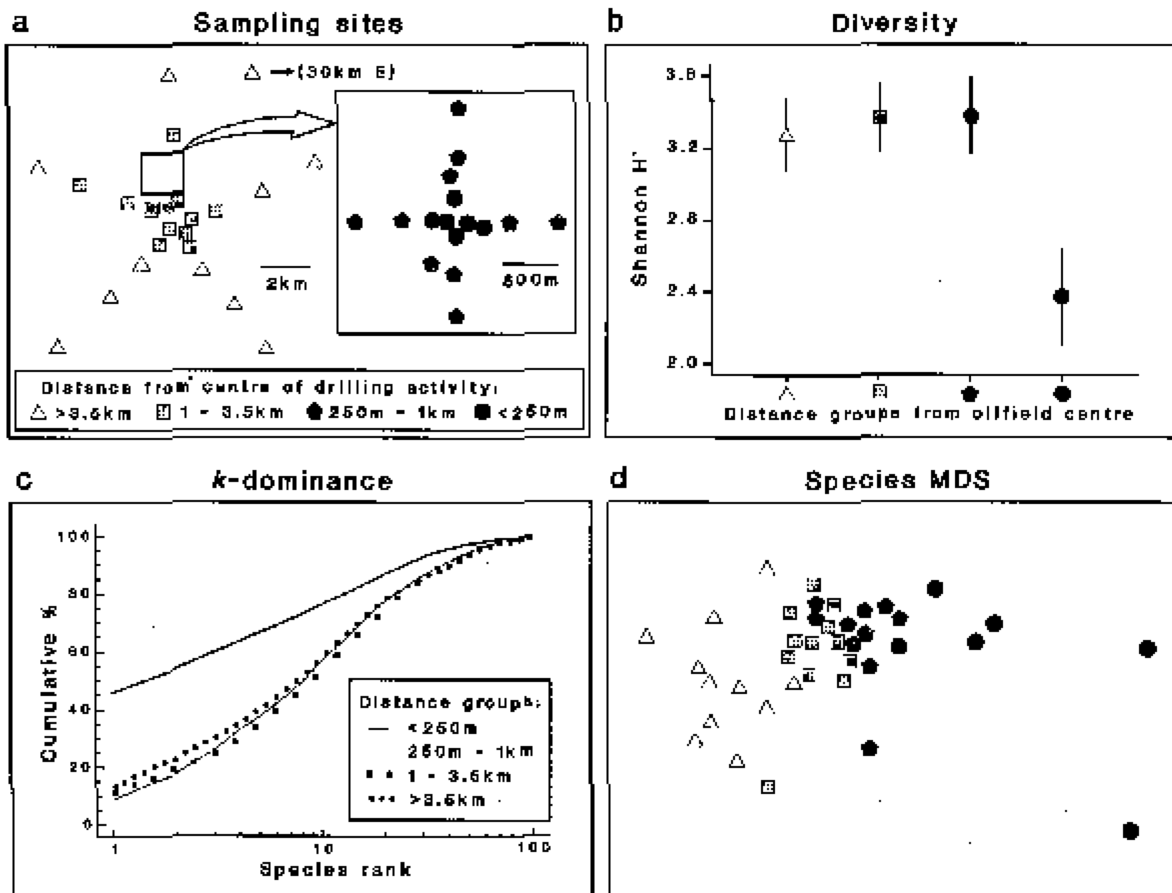


Fig. 14.4. Ekofisk macrobenthos (E). a) Map of sampling sites, represented by different symbol and shading conventions according to their distance from the 2/4K rig at the current centre of drilling activity; b) Shannon diversity (mean and 95% confidence intervals) in these distance zones; c) mean k–dominance curves; d) MDS from root–transformed species abundances (stress = 0.12).

EXAMPLE 3: Reef corals at South Pari Island, Indonesia

Warwick et al. (1990b) analysed coral community responses to the El Niño of 1982–3 at two reefs sites in the Thousand Islands, Indonesia *(I)*, based on 10 replicate line transects for each of the years 1981, 83, 84, 85, 87 and 88.

Univariate indices

At Pari Island there was an immediate reduction in diversity in 1983, apparent full recovery by 1985, with a subsequent but not significant reduction (Fig. 14.5).

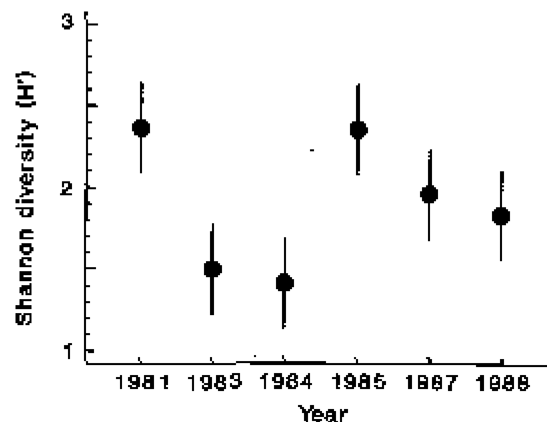


Fig. 14.5. Indonesian reef corals, Pari Island (I). Shannon diversity (means and 95% confidence intervals) of the species coral cover from 10 transects in each year.

Graphical/distributional plots

The mean k–dominance curves were similar in 1981 and 1985, with the curves for 1983, 84, 87 and 88 more elevated (Fig. 14.6). Tests on the replicate curves (see the end of Chapter 8) confirmed the significance of

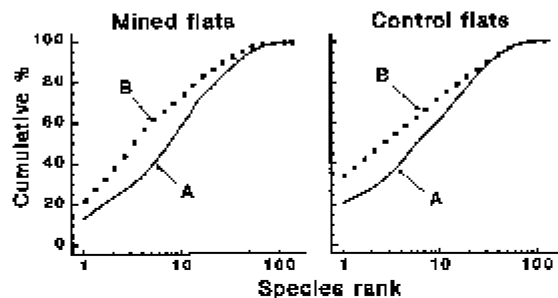


Fig. 14.9. Maldive Islands, coral–reef fish (M). Average k–dominance curves for abundance and biomass at mined and control reef–flat sites.

Conclusions

There were clear differences in community composition due to mining activity revealed by multivariate methods, even on the reef–slopes adjacent to the mined flats, but these were not detected at all by univariate or graphical/distributional techniques, even on the flats where the separation in the MDS is so obvious.

EXAMPLE 5: Macro- and meiobenthos from Isles of Scilly seaweeds

The entire metazoan fauna (macrofauna + meiofauna) has been analysed from five species of intertidal macro–algae (*Chondrus, Laurencia, Lomentaria, Cladophora, Polysiphonia*) each collected at either sites near low water from rocky shores on the Isles of Scilly *(S)* (Fig. 14.11).

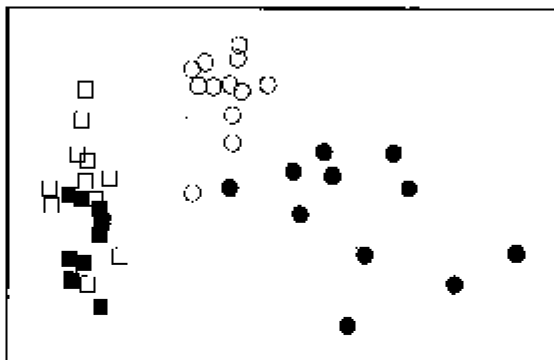


Fig. 14.10. Maldive Islands, coral–reef fish (M). MDS of 4th root–transformed species abundance data. Symbols as in Fig. 14.8, i.e. circles = reef–flat, squares = slope, solid = mined, open = control (stress = 0.09).

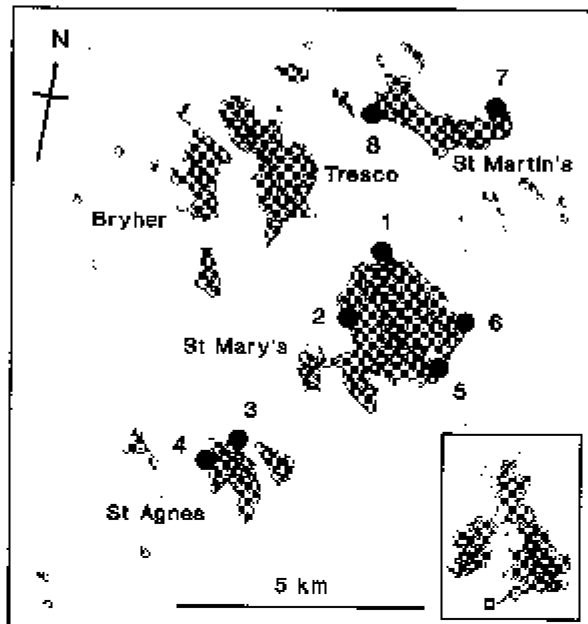


Fig. 14.11. Isles of Scilly (S). Map of the 8 sites from each of which 5 seaweed species were collected.

Univariate indices

The meiofauna and macrofauna showed clearly different diversity patterns with respect to weed type; for the meiofauna there was a trend of increasing diversity from the coarsest (*Chondrus*) to the finest (*Polysiphonia*) weed, but for the macrofauna there was no clear trend and *Polysiphonia* had the lowest diversity (Fig. 14.12).

Graphical/distributional plots

These differences in meiofauna and macrofauna diversity profiles were also reflected in the *k*-dominance curves (Fig. 14.13) which had different sequencing for these two faunal components, for example the *Polysiphonia* curve was the lowest for meiofauna and highest for macrofauna.

Multivariate analysis

The MDS plots for meiobenthos and macrobenthos were very similar, with the algal species showing very similar relationships to each other in terms of their meiofaunal and macrofaunal community structure (see Fig. 13.7, in which the shading and symbol conventions for the different weed species are the same as those in Fig. 14.12). Two–way ANOSIM (weed species/sites) showed all weed species to be significantly different from each other in the composition of both macrofauna and meiofauna.

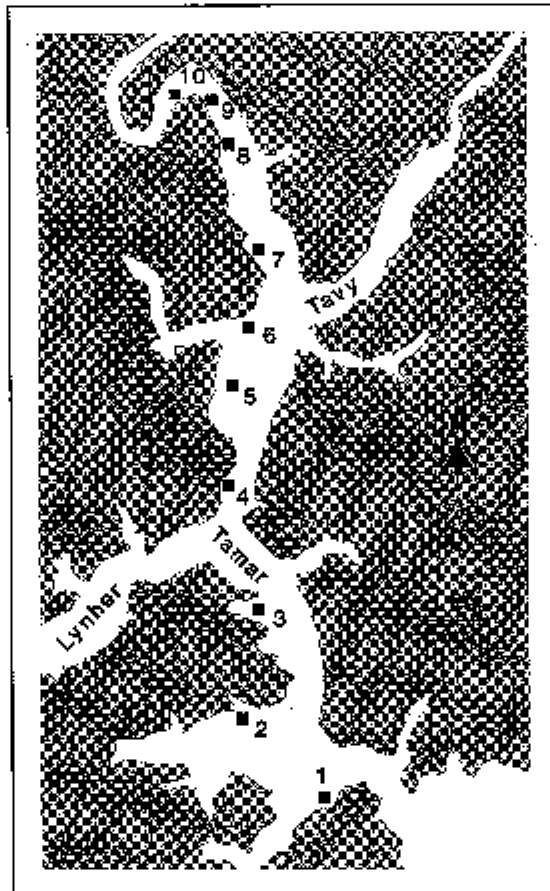


Fig. 14.14. Tamar estuary meiobenthos (R). Map showing locations of 10 intertidal mud–flat sites.

EXAMPLE 7: Meiofauna from Eaglehawk Neck sandflat, Tasmania

This example of the effect of disturbance by burrowing and feeding of soldier crabs *(T)* was dealt with in some detail in Chapter 12. For nematodes, univariate graphical and multivariate methods all distinguished disturbed from undisturbed sites. For copepods only the multivariate methods did. Univariate and graphical methods indicated different responses for nematodes and copepods, whereas the multivariate methods indicated a similar response for these two taxa.

GENERAL CONCLUSIONS

Three general conclusions emerge from these examples:

1) The similarity in community structure between sites or times based on their univariate or graphical/distributional attributes is different from their clustering in the multivariate analysis.
2) The species–dependent multivariate method is much more sensitive than the species–independent methods in discriminating between sites or times.
3) In examples where more than one component of the fauna has been studied, univariate and graphical methods may give different results for different components, whereas multivariate methods tend to give the same results.

The sensitive multivariate methods have hitherto only been used for detecting *differences* in community composition between sites. Although these differences can be correlated with measured levels of stressors such as pollutants, the multivariate methods so far described do not in themselves indicate

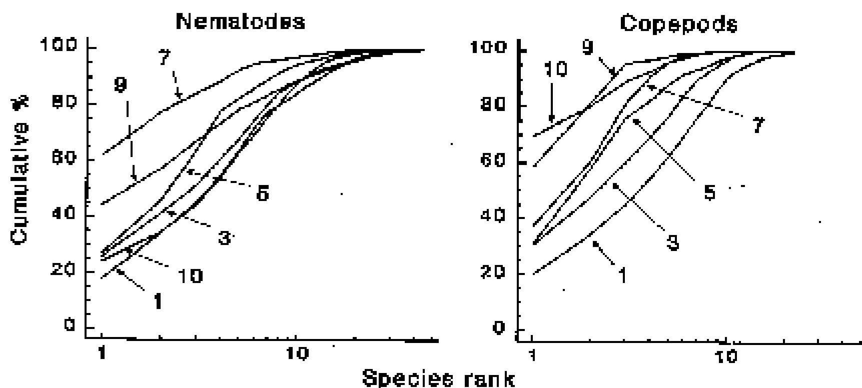


Fig. 14.15. Tamar estuary meiobenthos (R). k–dominance curves for amalgamated data from 6 replicate cores for nematode and copepod species abundances. For clarity of presentation, some sites have been omitted.

CHAPTER 15: MULTIVARIATE MEASURES OF COMMUNITY STRESS

We have seen in Chapter 14 that multivariate methods of data analysis are very sensitive for detecting differences in community structure between samples in space, or changes over time. Until recently, however, these methods have simply been used to detect *differences* between communities, and not in themselves as measures of *community stress* in the same sense that species–independent methods (e.g. diversity, ABC curves) have been used. Even using the relatively less sensitive species–independent methods there may be problems of interpretation in this context. Diversity does not behave consistently or predictably in response to environmental stress. Both current theory (Connell, 1978; Huston, 1979) and empirical observation (e.g. Dauvin 1984) suggest that increasing levels of disturbance may either decrease or increase diversity, and it may even remain the same. A monotonic response would be easier to interpret. False indications of disturbance using the ABC method may also arise when, as occasionally happens, the species responsible for elevated abundance curves are pollution sensitive rather than pollution tolerant species (e.g. small amphipods, *Hydrobia etc.*). Knowledge of the actual identities of the species involved will therefore aid the interpretation of ABC curves, and the resulting conclusions will be derived from an informal hybrid of species–independent and species–dependent information (Warwick and Clarke, in press). In this chapter we describe three possible approaches to the measurement of community stress using the fully species–dependent *multivariate* methods.

META-ANALYSIS OF MARINE MACROBENTHOS

This method was initially devised as a means of comparing the severity of community stress between various cases of both anthropogenic and natural disturbance. On initial consideration, measures of community degradation which are independent of the taxonomic identity of the species involved would be most appropriate for such comparative studies. Species composition varies so much from place to place depending on local environmental conditions that any general species–dependent response to stress would be masked by this variability. However, diversity measures are also sensitive to changes in natural environmental variables and an unperturbed community in one locality could easily have the same diversity as a perturbed community in another. Also, to obtain comparative data on species diversity requires a highly skilled and painstaking analysis of species and an unusually high degree of standardisation with respect to the degree of taxonomic rigour applied to the sample analysis: e.g. it is not valid to compare diversity at one site where one taxon is designated as "nemertines" with another at which this taxon has been divided into species.

The problem of natural variability in species composition from place to place can be overcome by working at taxonomic levels higher than species. The taxonomic composition of natural communities tends to become increasingly similar at these higher levels. Although two communities may have no species in common, they will almost certainly comprise the same phyla. For soft–bottom marine benthos, we have already seen in Chapter 10 that disturbance effects are detectable with multivariate methods at the highest taxonomic levels, even in some instances where these effects are rather subtle and are not evidenced in univariate measures even at the species level, e.g. the Amoco–Cadiz *(A)* and Ekofisk *(E)* studies.

Meta–analysis is a term widely used in biomedical statistics and refers to the combined analysis of a range of individual case–studies which in themselves are of limited value but in combination provide a more global insight into the problem under investigation. Warwick and Clarke (1993a) have combined macrobenthic data aggregated to phyla from a range of case-studies *(J)* relating to varying types of disturbance, and also from sites which are regarded as unaffected by such perturbations. A choice was made of the most ecologically meaningful units in which to work, bearing in mind the fact that abundance is a rather poor measure of such relevance, biomass is better and production is perhaps the most relevant of all (Chapter 13). Of course, no studies have measured production (*P*) of all species within a community, but many studies provide both abundance (*A*) and biomass (*B*) data. Production was therefore approximated using the allometric equation:

$$P = (B/A)^{0.73} \times A \qquad (15.1)$$

B/A is of course the mean body–size, and 0.73 is the average exponent of the regression of annual production on body–size for macrobenthic invertebrates. Since the data from each study are standardised (i.e. production of each phylum is expressed as a proportion of the total) the intercept of

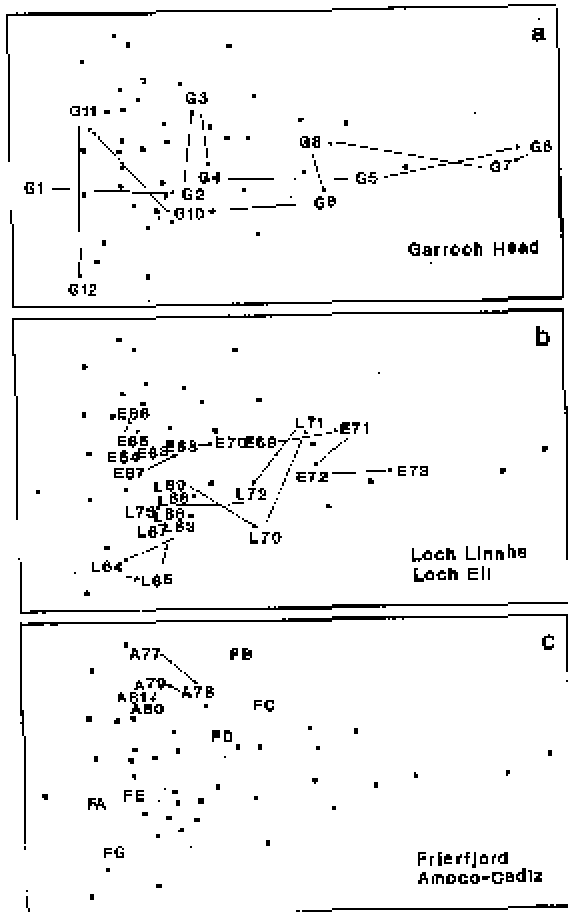


Fig. 15.2. Joint NE Atlantic shelf studies ("meta–analysis") [J]. As Fig. 15.1 but with individual studies highlighted: a) Garroch Head (Clyde) dump–ground; b) Loch Linnhe and Loch Eil; c) Frierfjord and Amoco–Cadiz spill (Morlaix).

Stations at the two extremities of the transect (1 and 12) are at the extreme left of the wedge, and stations close to the dump centre (6) are at the extreme right.

2) *Loch Linnhe and Loch Eil [L].* In the early years (1963–68) both stations are situated at the unpolluted left–hand end of the configuration (Fig. 15.2b). After this the L. Eil station moves towards the right, and at the end of the sampling period (1973) it is close to the right–hand end; only the sites at the centre of the Clyde dump–site are more polluted. The L. Linnhe station is rather less affected and the previously mentioned recovery in 1973 is evidenced by the return to the left–hand end of the wedge.

3) *Frierfjord (Oslofjord) [F].* The left to right sequence of stations in the meta–analysis is A–G–E–D–B–C (Fig. 15.2c), exactly matching the ranking in order of increasing stress. Note that the three stations affected by seasonal anoxia (B, C and D) are well to the right of the other three, but are not as severely disturbed as the organically enriched sites in 1) and 2) above.

4) *Amoco–Cadiz spill, Morlaix [A].* Note the shift to the right between 1977 (pre–spill) and 1978 (post–spill), and the subsequent return to the left in 1979–81 (Fig. 15.2c). However, the shift is relatively small, suggesting that this is only a mild effect.

5) *Skagerrak.* The biologically disturbed 300m station is well to the right of the undisturbed 100m station, although the former is still quite close to the left–hand end of the wedge.

6–8) *Unpolluted sites.* The Northumberland, Carmarthen Bay and Keil Bay stations are all situated at the left–hand end of the wedge.

An initial premise of this method was that, at the phylum level, the taxonomic composition of communities is relatively less affected by natural environmental variables than by pollution or disturbance (Chapter 10). To test this Warwick and Clarke (1993a) superimposed symbols scaled in size according to the values of the two most important environmental variables considered to influence community structure, sediment grain size and water depth, onto the meta–analysis MDS configuration (Chapter 11). Both variables were quite randomly distributed, which supports the original assumption.

With respect to individual phyla, annelids comprise a high proportion of the total "production" at the polluted end of the wedge, with a decrease at the least polluted sites. Molluscs are also present at all sites, except the two most polluted, and have increasingly higher dominance towards the non–polluted end of the wedge. Echinoderms are even more concentrated at the non–polluted end, with some tendency for higher dominance at the bottom of the configuration (Fig. 15.3a). Crustacea are again concentrated to the left, but this time completely confined to the top part of the configuration (Fig. 15.3b). Clearly, the differences in relative proportions of crustaceans and echinoderms are largely responsible for the vertical spread of samples at this end of the wedge, but these differences cannot be explained in terms of the effects of any recorded natural environmental variables. Nematoda are clearly more important at the polluted end of the wedge, an obvious consequence of the fact that species associated with organic enrichment tend to be very large in comparison with their normal meiofaunal counterparts (e.g. Oncholaimids), and are therefore

location of samples on the MDS or PCA plots and emphasise the *movement* (to the right) of putatively impacted samples relative to appropriate controls. For a new study, the spread of sample positions in the meta–analysis allows one to scale the importance of observed changes, in the context of differences between control and impacted samples for the training set.

Table 15.1. Joint NE Atlantic shelf studies ("meta–analysis") (J). Eigenvectors for first three principal components from covariance–based PCA of standardised and 4th root–transformed phylum "production" (all samples).

Phylum	PC1	PC2	PC3
Cnidaria	–0.039	0.094	0.039
Platyhelminthes	–0.016	0.026	–0.105
Nemertea	0.169	0.026	0.061
Nematoda	0.349	–0.127	–0.166
Priapulida	–0.019	0.010	0.003
Sipuncula	–0.156	0.217	0.105
Annelida	0.266	0.109	–0.042
Chelicerata	–0.004	0.013	–0.001
Crustacea	0.265	0.864	–0.289
Mollusca	–0.445	–0.007	0.768
Phoronida	–0.009	0.005	0.008
Echinodermata	–0.693	–0.404	–0.514
Hemichordata	–0.062	–0.067	–0.078
Chordata	–0.012	0.037	–0.003

It is perhaps premature, however, to make a positive recommendation that new data sets should be evaluated in either of the above ways. The training data is unlikely to be fully representative of all types of perturbation that could be encountered. For example, all the grossly polluted samples presently involve organic enrichment of some kind, which is conducive to the occurrence of the large nematodes which play some part in the positioning of these samples at the extreme right of the meta–analysis MDS or PCA. This may not happen with communities subjected to toxic chemical contamination only. Also, the training data are only from the NE European shelf, although data from a tropical locality (Trinidad, West Indies) have also been shown to conform with the same trend (Agard *et al.*, 1993).

INCREASED VARIABILITY

Warwick and Clarke (1993b) noted that, in a variety of environmental impact studies, the variability among samples collected from impacted areas was much greater than that from control sites. The suggestion was that this variability in itself may be an identifiable symptom of perturbed situations. The four examples examined were:

1) ***Meiobenthos from a nutrient–enrichment study (N);*** a mesocosm experiment to study the effects of three levels of particulate organic enrichment (control, low dose and high dose) on meiobenthic community structure (nematodes plus copepods), using four replicate box–cores of sediment for each treatment level.
2) ***Macrobenthos from the Ekofisk oil field, N Sea (E);*** a grab sampling survey at 39 stations around the oil field centre. To compare the variability among samples at different levels of pollution impact, the stations were divided into four groups (A–D) with approximately equal variability with respect to pollution loadings. These groups were selected from a scatter plot of the concentrations of two key pollution–related environmental variables, petroleum hydrocarbons and barium. Since the dose/response curve of organisms to pollutant concentrations is usually logarithmic, the values of these two variables were log–transformed.
3) ***Corals from S Tikus Island, Indonesia (I);*** changes in the structure of reef–coral communities between 1981 and 1983, along ten replicate line transects, resulting from the effects of the 1982–83 El Niño.
4) ***Reef–fish in the Maldive Islands (M);*** the structure of fish communities on reef flats at 23 coral sites, 11 of which had been subjected to mining, with the remaining 12 unmined sites acting as controls.

Data were analysed by non–metric MDS using the Bray–Curtis similarity measure and either square root (mesocosm, Ekofisk, Tikus) or fourth root (Maldives) transformed species abundance data (Fig. 15.4). While the control and low dose treatments in the meiofaunal mesocosm experiment show tight clustering of replicates, the high dose replicates are much more diffusely distributed (Fig. 15.4a). For the Ekofisk macrobenthos, the Group D (most impacted) stations are much more widely spaced than those in Groups A–C (Fig. 15.4b). For the Tikus Island corals, the 1983 replicates are widely scatterd around a tight cluster of 1981 replicates (Fig. 15.4c), and for the Maldives fish the control sites are tightly clustered entirely to the left of a more diffuse cluster of replicates of mined sites (Fig. 15.4d). Thus, the increased variability in multivariate structure with increased disturbance is clearly evident in all examples.

It is possible to construct an index from the relative variability between impacted and control samples. One obvious comparative measure of dispersion would be based on the difference in average distance

For the Ekofisk macrobenthos, strongly positive values are found in comparisons between the group D (most impacted) stations and the other three groups. It should be noted however that stations in groups C, B and A are increasingly more widely spaced geographically. Whilst groups B and C have similar variability, the degree of dispersion increases between the two outermost groups B and A, probably due to natural spatial variability. However, the most impacted stations in group D, which fall within a circle of 500 m diameter around the oil–field centre, still show a greater degree of dispersion than the stations in the outer group A which are situated *outside* a circle of 7 kilometers diameter around the oil–field. Comparison of the impacted versus control conditions for both the Tikus Island corals and the Maldives reef–fish gives strongly positive IMD values. For the Maldives study, the mined sites were more closely spaced geographically than the control sites, so this is another example for which the increased dispersion resulting from the anthropogenic impact is "working against" a potential increase in variability due to wider spacing of sites. Nonetheless, for both the Ekofisk and Maldives studies the increased dispersion associated with the impact more than cancels out that induced by the differing spatial scales. For both the mesocosm meiobenthos and the Tikus Island coral studies there are no such differences in spatial layout between the treatments to dilute the observed dispersion effects.

Application of the comparative index of multivariate dispersion suffers from the lack of any obvious statistical framework within which to test hypotheses of comparable variability between groups. As proposed, it is also restricted to the comparison of only two groups, though it can be extended to several groups in straightforward fashion. Let $\bar{r}_i$ denote the mean of the $N_i = n_i(n_i - 1)/2$ rank similarities among the n_i samples within the ith group ($i = 1, ..., g$), having (as before) re–ranked the triangular matrix ignoring all between–group similarities, and let N denote the number of similarities involved in this ranking process ($N = \Sigma_i N_i$). Then the ***dispersion sequence***

$$\bar{r}_1/k, \bar{r}_2/k, ..., \bar{r}_g/k \qquad (15.4)$$

defines the relative variability within each of the g groups, the larger values corresponding to greater within–group dispersion. The denominator scaling factor k is $(N + 1)/2$, i.e. simply the mean of all N ranks involved, so that a ***relative dispersion*** of unity corresponds to "average dispersion". (If the number of samples is the same in all groups then the values in equation (15.4) will average unity, though this will not quite be the case if the $\{n_i\}$ are unbalanced.)

As an example, the relative dispersion values given by equation (15.4) have been computed[1] for the four studies considered above (Table 15.3). This can be seen as complementary information to the IMD values; Table 15.2 provides the pairwise comparisons following on from the global picture in Table 15.3. The conclusions from Table 15.3 are, of course, consistent with the earlier discussion, e.g. the increase in variability at the outermost sites in the Ekofisk study, because of their greater geographical spread, being nonetheless smaller than the increased dispersion at the central, impacted stations.

***Table 15.3.** Variability study (N, E, I, M). Relative dispersion of the groups (equation 15.4) in each of the four studies.*

Meiobenthos	Control	*0.58*
	Low dose	*0.79*
	High dose	*1.63*
Macrobenthos	Group A	*1.34*
	Group B	*0.79*
	Group C	*0.81*
	Group D	*1.69*
Corals	1981	*0.58*
	1983	*1.42*
Reef–fish	Control reefs	*0.64*
	Mined reefs	*1.44*

BREAKDOWN OF SERIATION

Clear–cut zonation patterns in the form of a serial change in community structure with increasing water depth are a striking feature of intertidal and shallow–water benthic communities on both hard and soft substrata. The causes of these zonation patterns are varied, and may differ according to circumstances, but include environmental gradients such as light or wave energy, competition and predation. None of these mechanisms, however, will necessarily give rise to discontinuous bands of different assemblages of species, which is implied by the term ***zonation***, and the more general term ***seriation*** is perhaps more appropriate for this pattern of community change, zonation (with discontinuities) being a special case. Many of the factors which determine the pattern of

1. Both the IMD and the relative dispersion values are computed by the PRIMER program MVDISP.

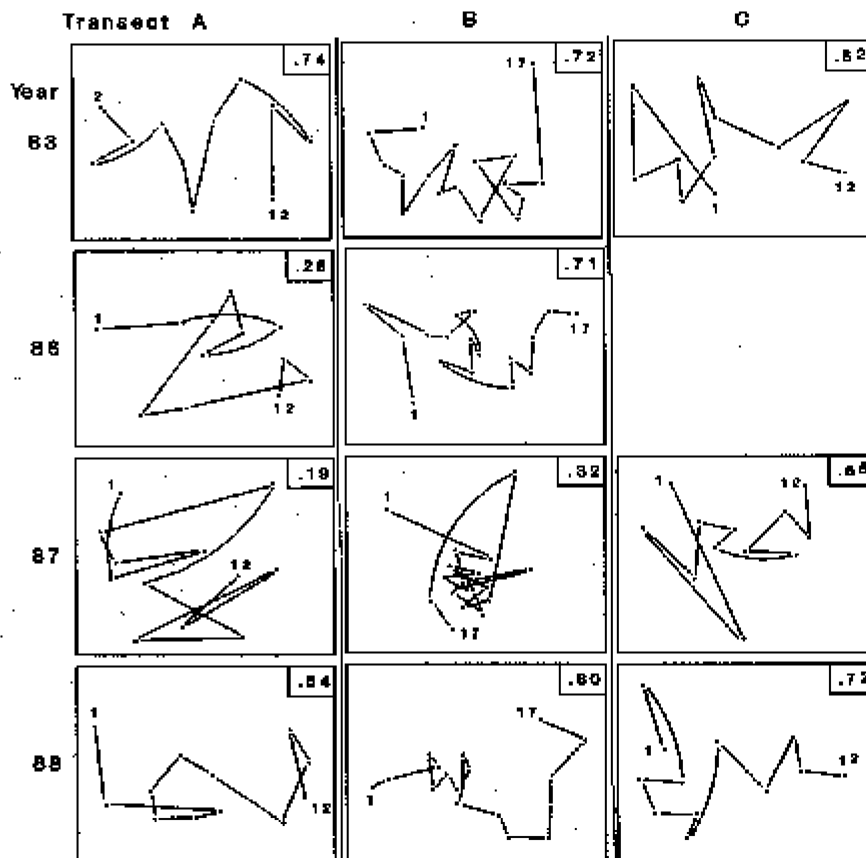


Fig. 15.6. Ko Phuket corals {K}. MDS ordination of the changing coral communities (species cover data) along three transects (A to C) at four times (1983 to 1988). The lines indicate the degree of seriation by linking successive points along a transect, from onshore (1) to offshore samples (12 or 17); IMS values are at top right. Sample 1 from transect A in 1983 is omitted (see text) and no samples were taken for transect C in 1986 (reading across rows, stress = 0.10, 0.11, 0.09; 0.10, 0.11; 0.08, 0.14, 0.11; 0.07, 0.09, 0.10).

non–monotonic – with the composition being similar at opposite ends of the transect but very different in the middle – then the IMS will be close to zero. These near–zero values can be negative as well as positive but no particular significance attaches to this.

A statistical significance test would clearly be useful, to answer the question: when is the IMS sufficiently different from zero to reject the null hypothesis of a complete absence of seriation? Such a test can be derived by a ***Monte Carlo permutation procedure***. If the null hypothesis is true then the labelling of samples along the transect $(1,2,...,n)$ is entirely arbitrary, and the spread of IMS values which are consistent with the null hypothesis can be determined by recomputing it for permutations of the sample labels in one of the two similarity matrices (holding the other fixed). For T randomly selected permutations of the sample labels, if only t of the T simulated IMS values are greater than or equal to the observed IMS, the null hypothesis can be rejected at a significance level of $100(t+1)/(T+1)\%$. In structure, the test is analogous to that considered at the end of Chapter 6 (implemented in the PRIMER program ANOSIM2), and again referred to briefly in Chapter 11 in the context of the BIO–ENV procedure.

One distinctive feature of the current test is that tied ranks will be much more prevalent, particularly in the similarities computed from the linear sequence, and it is advisable to make proper allowance for this in calculating the Spearman coefficients. Kendall (1970, equation 3.7) gives an appropriate adjustment to ρ_s, and this form is used in the analysis below.[2]

In 1983, before the dredging operations, MDS configurations (Fig. 15.6) indicate that the points along each transect conform rather closely to a linear sequence, and there are no obvious discontinuities in the sequence of community change (i.e. no discrete

2. The calculations for the tests were carried out using the PRIMER program RELATE. The similarity–based formulation, and the associated permutation test, are also readily extendable to more complex models than a linear sequence of change along a spatial transect. In a homologous way, community change could be related to a temporal trend or cyclicity, or to the sampling positions in a 2–dimensional spatial layout. There are null hypothesis tests for all these possibilities in RELATE, in addition to a general test for lack of relationship between __any__ two supplied similarity matrices with the same label sets (independently derived).

APPENDIX 1: INDEX OF EXAMPLE DATA

The following is a list of all (real) data sets used as examples in the text, where they are referenced by their indexing letter (A-Z). In addition to the pages on which these examples can be found, the entries give the source reference (see also Appendix 3) for the original publication of these data. Note that these are not always the appropriate references for the analyses presented in the text; the latter can generally be found in Appendix 2.

A – Amoco–Cadiz oil spill, Bay of Morlaix, France. *Macrofauna.* (Dauvin, 1984)
p 10–4, 10–5, 13–2, 13–3, 15–2, 15–3

B – Bristol Channel, England. *Zooplankton.* (Collins and Williams, 1982)
p 3–5, 3–6, 7–2, 7–3, 7–4, 11–3, 11–4

C – Celtic Sea. *Zooplankton.* (Collins, pers. comm.).
p 5–9

D – Dosing experiment, Solbergstrand mesocosm, Norway (GEEP Workshop). *Nematodes.* (Warwick *et al.*, 1988).
p 4–8, 5–8, 9–4

E – Ekofisk oil platform, N. Sea. *Macrofauna.* (Gray *et al.*, 1990).
p 8–4, 8–5, 10–5, 10–6, 14–2, 14–3, 15–5, 15–6, 15–7

F – Frierfjord, Norway (GEEP Workshop). *Macrofauna.* (Gray *et al.*, 1988).
p 1–3, 1–4, 1–9, 1–10, 3–1, 3–2, 6–1, 6–2, 6–3, 6–4, 8–9, 9–1, 10–1, 10–2, 13–6, 14–1, 14–2, 15–2, 15–3

G – Garroch Head, Scotland. *Macrofauna.* (Pearson and Blackstock, 1984).
p 1–6, 1–7, 1–8, 1–11, 1–12, 8–4, 8–5, 8–6, 8–8, 8–11, 11–1, 11–2, 11–3, 11–5, 11–9, 11–10, 15–2, 15–3

H – Hamilton Harbour, Bermuda (GEEP Workshop). *Macrofauna, nematodes.* (Warwick *et al.*, 1990c).
p 8–2, 8–3, 8–11, 13–3, 13–4

I – Indonesian reef corals, S. Pari and S. Tikus Islands. *Coral % cover.* (Warwick *et al.*, 1990b).
p 6–5, 6–6, 8–2, 10–4, 10–5, 10–7, 13–4, 13–5, 14–3, 14–4, 15–5, 15–6, 15–7

J – Joint NE Atlantic shelf studies ("meta–analysis"). *Macrofauna "production".* (Warwick and Clarke, 1993a)
p 15–2, 15–3, 15–4, 15–5

K – Ko Phuket coral reefs, Thailand. *Coral species cover.* (Clarke *et al.*, 1993).
p 15–8, 15–9, 15–10

L – Loch Linnhe and Loch Eil, Scotland. *Macrofauna.* (Pearson, 1975).
p 1–6, 1–7, 1–10, 4–4, 4–5, 8–6, 8–7, 8–10, 10–3, 10–4, 10–6, 10–7, 15–2, 15–3

M – Maldive Islands. *Coral reef fish.* (Dawson–Shepherd *et al.*, 1992).
p 13–2, 14–4, 14–5, 15–5, 15–6, 15–7

N – Nutrient–enrichment experiment, Solbergstrand mesocosm, Norway. *Nematodes, copepods.* (Gee *et al.*, 1985).
p 1–12, 1–13, 10–3, 10–4, 12–5, 12–6, 15–5, 15–6, 15–7

P – Plankton survey (Continuous Plankton Recorder), N.E. Atlantic. *Zooplankton, phytoplankton.* (Colebrook, 1986).
p 13–1

R – Tamar estuary mud–flat, S.W. England. *Nematodes, copepods.* (Austen and Warwick, 1989).
p 14–6, 14–7, 14–8

S – Scilly Isles, U.K. *Seaweed metazoa.* (Gee and Warwick, 1994).
p 13–5, 14–5, 14–6

T – Tasmania, Eaglehawk Neck. *Nematodes, copepods.* (Warwick *et al.*, 1990a).
p 6–7, 6–8, 12–2, 12–3, 12–4, 13–3, 13–4, 14–7

W – Westerschelde estuary cores, Netherlands; mesocosm experiment on food supply. *Nematodes.* (Austen and Warwick, in press)
p 6–8, 6–9, 6–11

X – Exe estuary, England. *Nematodes.* (Warwick, 1971).
p 5–3, 5–4, 5–7, 6–11, 6–12, 7–1, 7–2, 11–5, 11–6, 11–7, 11–9

Y – Clyde, Scotland. *Nematodes.* (Lambshead, 1986)
p 6–6, 6–7

Z – Azoic sediment recolonization experiment. *Copepods.* (Olafsson and Moore, 1992).
p 12–4

APPENDIX 2: PRINCIPAL LITERATURE SOURCES AND FURTHER READING

This manual chiefly reflects an approach to multivariate and other graphical community analyses that has been adopted and developed at the Plymouth Marine Laboratory (PML) over the last decade, and has been the subject of assessment and training at several IOC and FAO/UNEP workshops (e.g. papers in Bayne *et al.* 1988, Addison and Clarke 1990). Methodological papers involving work at PML include: Field *et al.* (1982), Warwick (1986), Clarke and Green (1988), Clarke (1990), Warwick and Clarke (1993a & b), Clarke and Ainsworth (1993), and Clarke and Warwick (1994). Clarke (1993) and Warwick (1993) review these methods, and a number of PML papers exemplify their use through the PRIMER package: see for example the papers listed under Warwick in Appendix 3.

Of course, the exposition here draws on a wider body of statistical and descriptive techniques, and there follows a brief listing of the main papers and books that can be consulted for further details of the methods and analyses of each Chapter.

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

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

Chapter 4: Ordination by PCA. This is one of the founding techniques of multivariate statistics; standard modern texts include Chatfield and Collins (1980) and Everitt (1978). The concluding example (Fig. 4.2) is from Warwick *et al.* (1988).

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

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

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

Chapter 8: Univariate/graphical analyses. Pielou (1975), Heip *et al.* (1988) and Magurran (1991) are useful texts, summarising a vast literature on a variety of diversity indices and ranked species abundance plots. The diversity examples here (Figs. 8.1 & 8.2) are discussed by Warwick *et al.* (1990c, 1990b respectively) and the Caswell *V* computations (Table 8.1) are from Warwick et al. (1990c). The Garroch Head species abundance distributions (Fig. 8.4) are first found in Pearson *et al.* (1983); Fig. 8.3 is redrawn from Pearson and Blackstock (1984). Warwick (1986) introduced Abundance–Biomass Comparison curves, and the Loch Linnhe and Garroch Head illustrations (Figs. 8.7

APPENDIX 3: REFERENCES CITED

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

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

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

Austen, M.C., Warwick, R.M. (in press). Effects of experimental manipulation of food supply on estuarine meiobenthos. *Hydrobiologia*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Warwick, R.M., Carr, M.R., Clarke, K.R., Gee, J.M., Green, R.H. (1988). A mesocosm experiment on the effects of hydrocarbon and copper pollution on a sublittoral soft-sediment meiobenthic community. *Mar. Ecol. Prog. Ser.* 46: 181–191

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

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

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

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

Warwick, R.M., Clarke, K.R., 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 1982–3 El Nino in the Thousand Islands, Indonesia. *Coral Reefs* 8: 171–179

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

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

Issued by

Oceans and Coastal Areas Programme Activity Centre
United Nations Environment Programme

Additional copies of this publication
can be obtained from:

Oceans and Coastal Areas Programme Activity Centre
United Nations Environment Programme
P.O. Box 30552
Nairobi
KENYA

or from:

Marine Environmental Studies Laboratory
International Atomic Energy Agency
Marine Environment Laboratory
B.P. No. 800 - MC 98012
MONACO CEDEX