

129971

CHANGE IN MARINE COMMUNITIES:
An Approach to Statistical Analysis and Interpretation
2nd Edition

K. R. Clarke & R. M. Warwick

 **PRIMER-E Ltd**

Plymouth Marine Laboratory, UK

VLIZ (vzw)
VLAAMS INSTITUUT VOOR DE ZEE
FLANDERS MARINE INSTITUTE
Oostende - Belgium

Published 2001
by
PRIMER-E Ltd

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

Business Office: 6 Hedingham Gardens
Roborough
Plymouth PL6 7DX
United Kingdom



Directors: K Robert Clarke MSc PhD Raymond N Gorley MA

First edition 1994
Reprinted 1997
Second edition 2001

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

© Copyright 2001 PRIMER-E Ltd, all rights reserved

Cover photos: Steve Smith (University of New England, Armidale, NSW, Australia), Paul Somerfield (Plymouth Marine Laboratory, UK), and Ashley Rowden (NIWA, Wellington, New Zealand)

CONTENTS

INTRODUCTION

- CHAPTER 1** A framework for studying changes in community structure
- CHAPTER 2** Measures of similarity of species abundance/biomass between samples
- CHAPTER 3** Hierarchical clustering
- CHAPTER 4** Ordination of samples by Principal Components Analysis (PCA)
- CHAPTER 5** Ordination of samples by Multi-Dimensional Scaling (MDS)
- CHAPTER 6** Testing for differences between groups of samples
- CHAPTER 7** Species analyses
- CHAPTER 8** Diversity measures, dominance curves and other graphical analyses
- CHAPTER 9** Transformations
- CHAPTER 10** Species removal and aggregation
- CHAPTER 11** Linking community analyses to environmental variables
- CHAPTER 12** Causality: community experiments in the field and laboratory
- CHAPTER 13** Data requirements for biological effects studies: which components and attributes of the marine biota to examine?
- CHAPTER 14** Relative sensitivities and merits of univariate, graphical/distributional and multivariate techniques
- CHAPTER 15** Multivariate measures of community stress
- CHAPTER 16** Further comparison of multivariate patterns
- CHAPTER 17** Biodiversity measures based on relatedness of species
- APPENDIX 1** Index of example data
- APPENDIX 2** Principal literature sources and further reading
- APPENDIX 3** References cited

INTRODUCTION

Purpose

This manual accompanies the computer software package PRIMER (Plymouth Routines In Multivariate Ecological Research), obtainable from PRIMER-E Ltd, Plymouth (see www.primer-e.com). Its scope is the analysis of data arising in community ecology and environmental science which is multivariate in character (many species, multiple environmental variables), and it is intended for use by ecologists with no more than a minimal background in statistics. As such, this *methods* manual complements the PRIMER *user* manual, by giving the background to the statistical techniques employed by the analysis programs (Table 0.1), at a level of detail which should allow the ecologist to understand the output from the programs, be able to describe the results in a non-technical way to others and have confidence that the right methods are being used for the right problem.

This may seem a tall order, in an area of statistics (primarily multivariate analysis) which has a reputation as esoteric and mathematically complex! However, whilst it is true that the computational details of some of the core techniques described here (for example, non-metric multidimensional scaling) are decidedly non-trivial, we maintain that all of the methods that have

Table 0.1. Chapters in this manual in which the main methodology underlying specific PRIMER routines are covered.[¶]

Routines	Chapters
<i>Similarity</i>	2
CLUSTER	3
PCA	4, 11
MDS	5
ANOSIM, ANOSIM2	6
SIMPER	7
DIVERSE	8, 17
CASWELL, Geometric & Dominance Plots	8
Transform	9
Aggregate	10
BIO-ENV, Draftsman Plot	11
MVDISP, RELATE	15
BVSTEP, 2STAGE	16
TAXDTEST	17

[¶] PRIMER also contains a number of other routines dealing with file and data manipulation (e.g. averaging, joining, defining factors, selecting subsets), graph options, output of results etc. The running of these routines, and those in Table 0.1, is fully described in the accompanying PRIMER User Manual/Tutorial.

been adopted or developed within PRIMER are so *conceptually* straightforward as to be amenable to simple explanation and transparent interpretation. In fact, the adoption of non-parametric and permutation approaches for display and testing of multivariate data requires, paradoxically, a lower level of statistical sophistication on the part of the user than does a satisfactory exposition of standard (parametric) hypothesis testing in the *univariate* case.

The principal aim of this manual is therefore to describe a coherent strategy for the interpretation of data on community structure, namely values of abundance, biomass, % cover, presence/absence etc. for a set of species and one or more replicate samples 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 uncontrolled or manipulative "treatments";
- or some combination of these.

These species-by-samples arrays are typically large, and patterns in community structure are often not readily apparent from simple inspection of the data. 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.

Material covered

It should be made clear at the outset that the title "Change in Marine Communities" does not in any way reflect a restriction in the scope of the techniques in the PRIMER package to the marine environment. The first edition of this manual *was* intended primarily for a marine audience and, given that its examples are still drawn entirely from marine contexts, it would be disingenuous to change the title now to a more general one. However, it will be self-evident to the reader that there is very little in the following pages that is exclusively marine. Indeed, the PRIMER package is now not only used world-wide for all types of marine community surveys and experiments, of benthic fauna, algae, corals, plankton, fish, diet studies etc., but is increasingly found in freshwater, terrestrial and palaeontology contexts, and sometimes solely in multivariate studies of physico-chemical characteristics.

As a result of the authors' own research interests and the widespread use of community data in pollution monitoring, a major thrust of the manual is the biological effects of contaminants but, again, most of the methods are much more generally applicable. This is reflected in a range of more fundamental ecological studies among the real data sets exemplified here.

The literature contains a large 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/Biodiversity group at Plymouth Marine Laboratory (PML), and which has a proven track record in interpretation of a wide range of marine community data; see, for example, papers listed under Clarke or Warwick in Appendix 3 (which have attained four-figure total citations in SCI journals). The analyses and displays in these papers, and in this manual, almost all draw upon the wide range of routines available in the PRIMER package (though in many cases annotations etc in plots have been further edited by simple importing into graphics programs such as Microsoft Powerpoint).

Note also that, whilst other software packages will not encompass this specific combination of routines, several of the individual techniques can be found elsewhere. For example, the core clustering and ordination methods described here are available in several mainstream statistical packages (SAS, S-Plus, Systat, Statgraphics etc.), and more specialised statistical programs (CANOCO, PATN, PC-ORD, the Cornell Ecology programs, etc.) tackle essentially similar problems, though usually employing different techniques and a different strategy.

Practical use

The arrangement of topics, and level of exposition, have benefited from experience gained at several training workshops funded jointly by FAO, UNEP and UNESCO/IOC, and a series of commercially-run PRIMER courses at Plymouth and venues outside the UK. The advocacy of these techniques thus springs not only from regular use and development within PML's Community Ecology/Biodiversity group but also from valuable feedback from a series of workshops in which practical data analyses were central.

Throughout the manual, extensive use is made of data sets from the published literature to illustrate the techniques. Appendix 1 gives the original literature source for each of these 25 or so data sets and an index to all the pages on which they are analysed. Each data set is allocated a single letter designation and, to avoid confusion, referred to in the text of the manual by that letter, placed in curly brackets (e.g. $\{A\}$ = Amoco-Cadiz oil spill, macrofauna; $\{B\}$ = Bristol Channel, zooplankton; $\{C\}$ = Celtic Sea, zooplankton etc).

Literature citation

This 2nd edition of the manual follows the 1st edition closely in respect of the first 15 chapters, though minor revisions have been made throughout. Chapters 16 and 17 are entirely new. Appendix 2 lists some background papers appropriate to each chapter, including the source of specific analyses, and a full listing of references cited is in Appendix 3.

Whilst the manual is genuinely collaboratively authored, for the purposes of directing queries on specific topics it is broadly true that the first author (KRC) bears the responsibility for the chapters on statistical methods (1–7, 9, 11) and the second author (RMW) is mainly responsible for the chapters on interpretation (10, 12–14), the responsibility for Chapters 8 and 15 being shared more or less equally. Chapters 16 and 17 were written by KRC, drawing on the results of joint papers in various authorship combinations by KRC, RMW and Paul Somerfield (also of the Plymouth Marine Laboratory). Since this manual is not accessible within the published literature, referral to the methods it describes would properly be by citing the primary papers on which it is based; these are indicated in the text and Appendix 2. Alternatively, comprehensive discussion of the philosophy (and many of the details) of the multivariate and univariate approaches advocated can be found in Clarke (1993, 1999) and Warwick (1993), respectively, with the newer methods in this edition best summarised in Clarke and Warwick (1998a), Somerfield and Clarke (1995) and Warwick and Clarke (2001).

Acknowledgements

We are grateful to a large number of individuals and institutions for their help and support – please see the detailed list at the end of the manual.

K R Clarke
R M Warwick
2001

CHAPTER 1: A FRAMEWORK FOR STUDYING CHANGES IN COMMUNITY STRUCTURE

The purpose of this opening chapter is twofold:

- a) to introduce some of the data sets which are used extensively, as illustrations of techniques, throughout the manual;
- b) to outline a framework for the various possible stages in a community analysis[¶].

Examples are given of some core elements of the recommended approaches, foreshadowing the analyses explained in detail later and 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 is intended to 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

[¶] The term **community** is used throughout the manual, somewhat loosely, to refer to any **assemblage** data (samples leading to counts, biomass, % cover, etc. for a range of species); the usage does not necessarily imply internal structuring of the species composition, for example by competitive interactions.

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 **species 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. Also included are **biodiversity indices** which measure the degree to which species or organisms in a sample are taxonomically or phylogenetically related to each other. 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** (Lambshhead *et al.*, 1983), which rank the species in decreasing order of abundance, convert the values to percentage abundance relative

to the total number of individuals in the sample, and plot the cumulated percentages against the species rank. This, and the analogous plot based on species biomass, are superimposed to define *ABC (abundance-biomass comparison) curves* (Warwick, 1986), which have proved a useful construct in investigating disturbance effects. Another example is the *species abundance distribution* (sometimes termed the *distribution of individuals amongst species*), in which the species are categorised into geometrically-scaled abundance classes and a histogram plotted of the number of species falling in each abundance range (e.g. Gray and Pearson, 1982). It is then argued, again from empirical evidence, that there are certain characteristic changes in this distribution associated with community disturbance.

Such distributional techniques relax the constraint in the previous category that the summary from each sample should be a *single* variable; here the emphasis is more on diversity *curves* than single diversity indices, but note that both these categories share the property that comparisons between samples are not based on particular species identities: two samples can have exactly the same diversity or distributional structure without possessing a single species in common.

3) *Multivariate methods* are characterised by the fact that they base their comparisons of two (or more) samples on the extent to which these samples share particular species, at comparable levels of abundance. Either explicitly or implicitly, all multivariate techniques are founded on such *similarity coefficients*, calculated between every pair of samples. These then facilitate a *classification* or *clustering* (these terms are interchangeable) of samples into

groups which are mutually similar, or an *ordination plot* in which, for example, the samples are “mapped” (usually in two or three dimensions) in such a way that the distances between pairs of samples reflect their relative dissimilarity of species composition.

Techniques described in detail in this manual are a method of *hierarchical agglomerative clustering* (e.g. Everitt, 1980), in which samples are successively fused into larger groups, as the criterion for the similarity level defining group membership is relaxed, and two ordination techniques: *principal components analysis* (PCA, e.g. Chatfield and Collins, 1980) and *non-metric multi-dimensional scaling* (NMDS, usually shortened to MDS, Kruskal and Wish, 1978).

For each broad category of analysis, the techniques appropriate to each stage are now discussed, and pointers given to the relevant chapters.

UNIVARIATE TECHNIQUES

For diversity indices and other single-variable extractions from the data matrix, standard statistical methods are usually applicable and the reader is referred to one of the many excellent general statistics texts (e.g. Sokal and Rohlf, 1981). The requisite techniques for each stage are summarised in Table 1.1. For example, when samples have the structure of a number of replicates taken at each of a number of sites (or times, or conditions), computing the means and 95% confidence intervals gives an appropriate *representation* of the Shannon diversity (say) at each site, with *discrimination between sites* being demonstrated by one-way analysis of variance (ANOVA), which is a test of the null hypothesis that there are no differences in mean diversity between

Table 1.1. *Univariate techniques. Summary of analyses for the four stages.*

Stages	Univariate examples		
	Diversity indices (Ch 8)	Indicator taxa	Biodiversity indices (Ch 17)
1) Representing communities	<i>Means and 95% confidence intervals for each site/condition (Ch 8, 9, 17)</i>		
2) Discriminating sites/conditions	<i>One-way analysis of variance, ANOVA (Ch 6)</i>		
3) Determining stress levels	<i>By reference to historical data for sites (Ch 14, 15) and regional “species pool” (Ch 17) Ultimately a decrease in diversity Initial increase in opportunists Loss of taxonomic distinctness</i>		
4) Linking to environment	<i>Regression techniques (Ch 11); for causality issues see Ch 12</i>		

sites. *Linking to the environment* is then also relatively straightforward, particularly if the environmental variables can 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 rarely very informative in practice, 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[¶], indicator species levels or the number of species in a sample (S) may not initially satisfy the assumptions necessary for classical statistical analysis. For 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[†]. 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

[¶] And thus subject to the central limit theorem, which will tend to induce statistical normality.

[†] 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.

be done *a priori*). Such arguments lead to the tenets underlying this manual:

- a) community data are usually highly multivariate (large numbers of species, each subject to high statistical *noise*) and need to be analysed *en masse* in order to elicit the important biological *signal* and its relation to the environment;
- b) standard parametric modelling is totally invalid.

Thus, throughout, little emphasis is given to representing communities by univariate measures, though some definitions of indices 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, an example given of a univariate regression between biota and environment in Chapter 11, and a more extensive discussion of sampling properties of diversity indices, and biodiversity measures based on taxonomic relatedness, makes up Chapter 17. 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 0.1m² 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/or biomass, B) are the starting point for the biotic analyses of this manual, and this example is typical in respect of the relatively high ratio of species to samples (always $\gg 1$) and the prevalence of zeros. Here, as elsewhere, even an undesirable reduction to the 30 “most important” species (see Chapter 2) leaves more

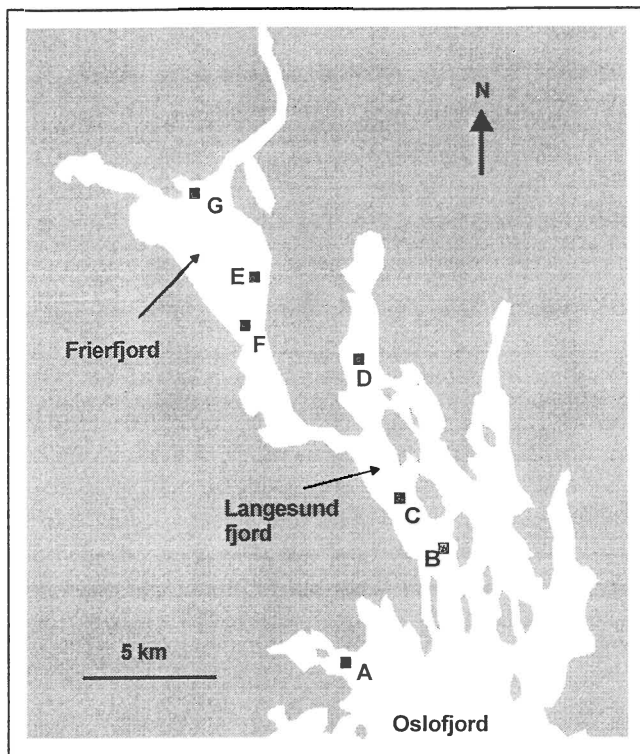


Fig. 1.1. Frierfjord, Norway {F}. Benthic community sampling sites (A-G) for the IOC/GEEP Oslo Workshop; site F omitted for macrobenthos.

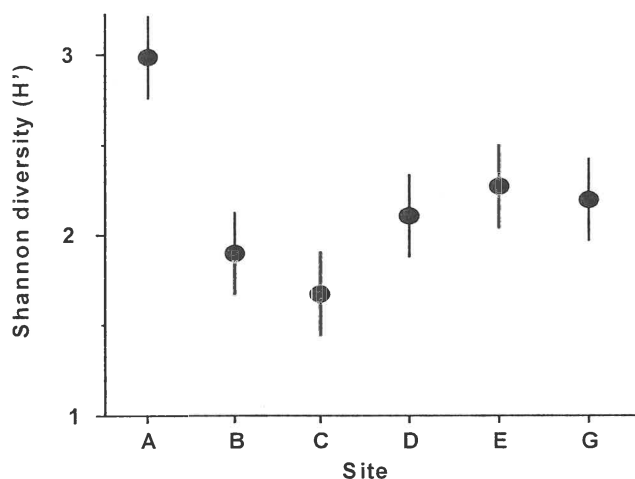


Fig. 1.2. Frierfjord macrofauna {F}. Means and 95% confidence intervals for Shannon diversity (H'), from four replicates at each of six sites (A-E, G).

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 that the number of species (variables) be small in relation to the number of samples, and that the abundance or biomass values are transformable to approximate normality: neither is possible.

As discussed above, one easy route to simplification of this “high-dimensional” complexity is to reduce each column of the matrix (each sample) to a single,

Table 1.2. Frierfjord macrofauna {F}. Abundance and biomass matrices (part only) for the 110 species in 24 samples (four replicates at each of six sites A-E, G); abundance in numbers per $0.1m^2$, biomass in mg per $0.1m^2$.

Species	Samples							
	A1	A2	A3	A4	B1	B2	B3	B4
Abundance								
<i>Cerianthus lloydi</i>	0	0	0	0	0	0	0	0
<i>Halicryptus sp.</i>	0	0	0	1	0	0	0	0
<i>Onchnesoma</i>	0	0	0	0	0	0	0	0
<i>Phascolion strombi</i>	0	0	0	1	0	0	1	0
<i>Golfingia sp.</i>	0	0	0	0	0	0	0	0
<i>Holothuroidea</i>	0	0	0	0	0	0	0	0
<i>Nemertina, indet.</i>	12	6	8	6	40	6	19	7
<i>Polycaeta, indet.</i>	5	0	0	0	0	0	1	0
<i>Amaena trilobata</i>	1	1	1	0	0	0	0	0
<i>Amphicteis gunneri</i>	0	0	0	0	4	0	0	0
<i>Ampharetidae</i>	0	0	0	0	1	0	0	0
<i>Anaitides groenl.</i>	0	0	0	1	1	0	0	0
<i>Anaitides sp.</i>	0	0	0	0	0	0	0	0
....								
Biomass								
<i>Cerianthus lloydi</i>	0	0	0	0	0	0	0	0
<i>Halicryptus sp.</i>	0	0	0	26	0	0	0	0
<i>Onchnesoma</i>	0	0	0	0	0	0	0	0
<i>Phascolion strombi</i>	0	0	0	6	0	0	2	0
<i>Golfingia sp.</i>	0	0	0	0	0	0	0	0
<i>Holothuroidea</i>	0	0	0	0	0	0	0	0
<i>Nemertina, indet.</i>	1	41	391	1	5	1	2	1
<i>Polycaeta, indet.</i>	9	0	0	0	0	0	0	0
<i>Amaena trilobata</i>	144	14	234	0	0	0	0	0
<i>Amphicteis gunneri</i>	0	0	0	0	45	0	0	0
<i>Ampharetidae</i>	0	0	0	0	0	0	0	0
<i>Anaitides groenl.</i>	0	0	0	7	11	0	0	0
<i>Anaitides sp.</i>	0	0	0	0	0	0	0	0
....								

univariate description. Fig. 1.2 shows the results of computing the Shannon diversity (H' , see Chapter 8) of each sample[¶], and plotting for each site the mean diversity and its 95% confidence interval, based on a pooled estimate of variance across all sites from the ANOVA table, Chapter 6. (An analysis of the type outlined in Chapter 9 shows that prior transformation of H' is not required; it already has approximately constant variance across the sites, a necessary prerequisite for standard ANOVA). The most obvious feature of Fig. 1.2 is the relatively higher diversity at the “control” location, A.

[¶] Using the PRIMER DIVERSE routine.

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

Stages	Distributional examples	
	ABC (<i>k</i> -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. <i>W</i> , 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[†].

[†] 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.

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 distribution/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, for example in *linking to environmental variables*, where the only viable course again seems to be reduction of the curve(s) for each sample to a summary statistic (such as *W*), which can be regressed on particular abiotic variables.

Table 1.4. Loch Linnhe macrofauna {L}. Abundance/biomass matrix (part only); one (pooled) set of values per year (1963-1973).

Species	1963		1964		1965		1966 ...	
	A	B	A	B	A	B	A	B
<i>Scutopus ventrolineatus</i>	0	0	0	0	11	0.05	0	0
<i>Nucula tenuis</i>	2	0.01	13	0.07	16	0.10	6	0.04
<i>Mytilus edulis</i>	0	0	0	0	5	0.09	0	0
<i>Modiolus sp. indet.</i>	0	0	0	0	0	0	0	0
<i>Thyasira flexuosa</i>	93	3.57	210	7.98	28	1.06	137	5.17
<i>Myrtea spinifera</i>	214	27.39	136	17.41	2	0.26	282	36.10
<i>Lucinoma borealis</i>	12	0.39	26	1.72	0	0	22	0.73
<i>Montacuta ferruginosa</i>	1	0	0	0	4	0.02	0	0
<i>Mysella bidentata</i>	0	0	0	0	0	0	0	0
<i>Abra sp. indet.</i>	0	0	0	0	12	0.26	0	0
<i>Corbula gibba</i>	2	0.13	8	0.54	9	0.27	2	0.13
<i>Thracia sp. indet.</i>	0	0	0	0	0	0	0	0
...								

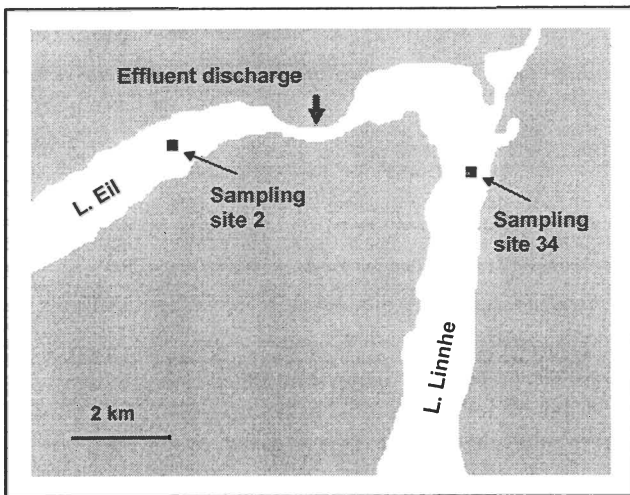


Fig. 1.3. Loch Linnhe and Loch Eil, Scotland {L}. Map of site 34 (Linnhe) and site 2 (Eil), sampled annually over 1963-1973.

EXAMPLE: Loch Linnhe macrofauna

Pearson (1975) describes a time series of macrobenthic community samples, taken over the period 1963-1973 inclusive, at two sites in a sea loch system on the west coast of Scotland ({L}, Fig. 1.3.) Pooling to a single sample for each of the 11 years resulted in abundance and biomass matrices of 115 rows (species) and 11 columns (samples), a small part of which is shown in Table 1.4.[¶] Starting in 1966, pulp-mill effluent was discharged to the sea lochs (Fig. 1.3),

[¶] It is displayed in this form purely for illustration; this is not a valid file format for PRIMER, which requires the abundance and biomass information to be in separate (same-shape) arrays.

with the rate increasing in 1970 and a significant reduction taking place in 1972 (Pearson, 1975). The top left-hand plot of Fig 1.4 shows the Shannon diversity of the macrobenthic samples over this period, and the remaining plots the ABC curves for each year.[†] There appears to be a consistent change of structure from one in which the biomass curve dominates the abundance curve in the early years, to the curves crossing, reversing altogether and then finally reverting to their original form.

EXAMPLE: Garroch Head macrofauna

Pearson and Blackstock (1984) describe the sampling of a transect of 12 sites across the sewage-sludge disposal ground at Garroch Head in the Firth of Clyde, SW Scotland ({G}, Fig. 1.5). The samples considered here were taken during 1983 and consisted of abundance and biomass values of 84 macrobenthic species, together with associated contaminant data on the extent of organic enrichment and the concentrations of heavy metals in the sediments. Fig. 1.6 shows the resulting species abundance distributions for the 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.

[†] Computed from the PRIMER Dominance plot routine.

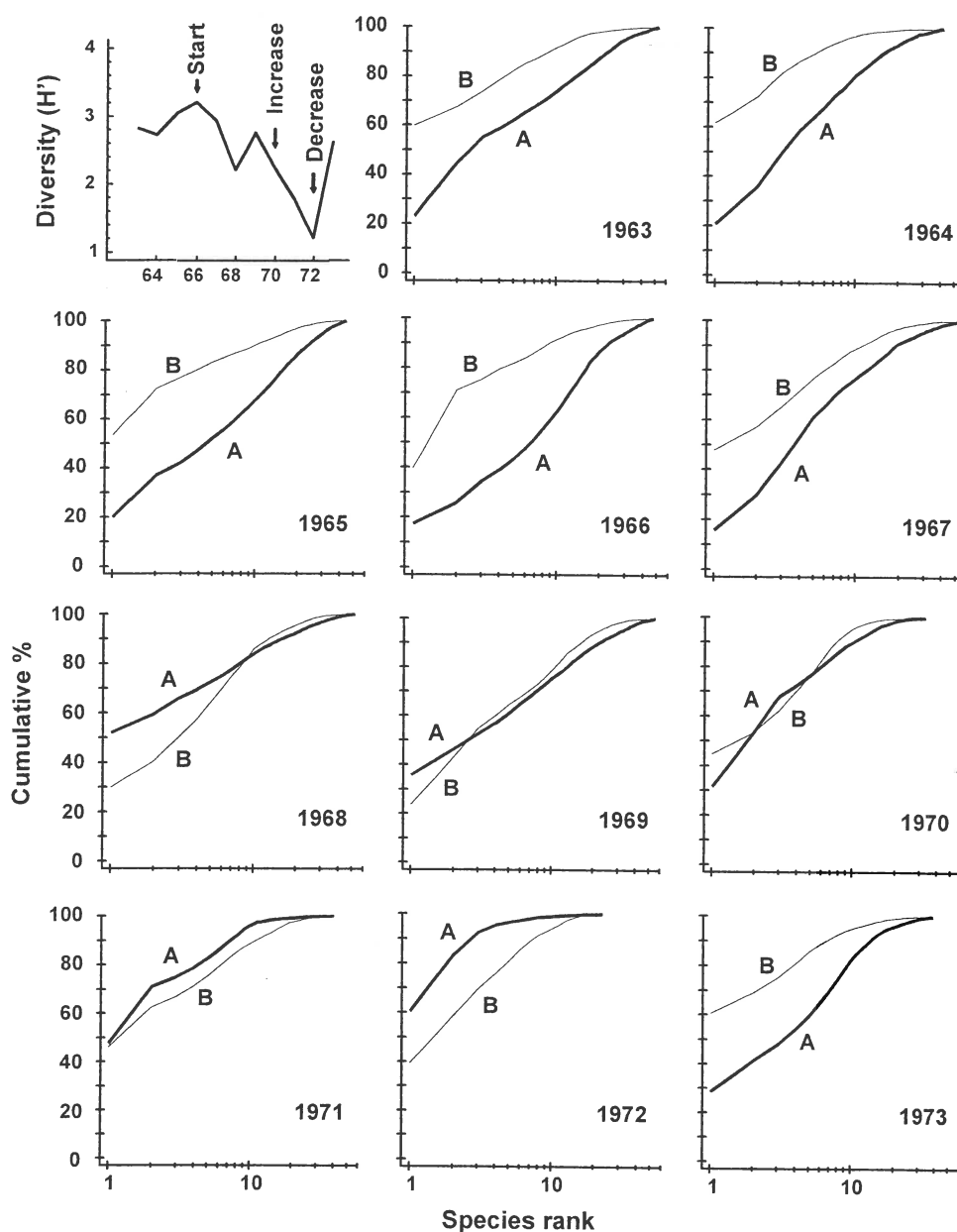


Fig. 1.4. Loch Linnhe macrofauna {L}. Top left: Shannon diversity over the 11 annual samples, also indicating timing of start of effluent discharge and a later increase and decrease in level; remaining plots show ABC curves for the separate years 1963-1973 (B = biomass, thin line; A = abundance, thick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

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 components 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 particular (low-dimensional) "view" of the structure it exhibits. A view in which most of the emphasis is on the pattern of occurrence of rare species may be very different than one in which the emphasis is wholly on the handful of species that numerically dominate most of the samples. One convenient way of providing this spectrum of choice, to match the biological imperatives whilst retaining desirable theoretical properties, is to restrict attention to a single similarity coefficient (specifically that of Bray and Curtis, 1957) but allow a choice of prior transformation of the data. A useful transformation continuum (Chapter 9) ranges through: no transform, square root, fourth root, logarithmic and

finally, reduction of the sample information to the recording only of presence or absence for each species.[¶] At the former end of the spectrum all attention will be focused on the dominant counts, at the latter end on the rarer species.

For the clustering technique, *representation* of the communities for each sample is by a dendrogram (e.g. Fig. 1.7a), linking the samples in hierarchical groups on the basis of some definition of similarity between each cluster (Chapter 3). This is a particularly relevant

representation in cases where the samples are expected to divide into well-defined groups, perhaps structured by some clear-cut environmental distinctions. Where, on the other hand, the community pattern is responding to abiotic gradients which are more continuous, then representation by an ordination is usually more appropriate. The method of non-metric MDS (Chapter 5) attempts to place the samples on a “map”, usually in two dimensions (e.g. see Fig. 1.7b), in such a way that the *rank order* of the distances between samples on the map exactly agrees with the rank order of the matching (dis)similarities, taken from the triangular similarity matrix. If successful, and success is measured by a *stress coefficient* which reflects lack of agreement in the two sets of ranks, 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:

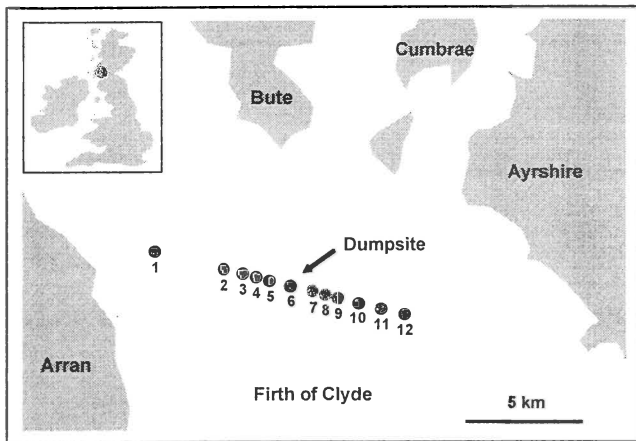


Fig. 1.5. Garroch Head, Scotland {G}. Location of sewage sludge dump ground and position of sampling sites (1–12); the dump centre is at site 6.

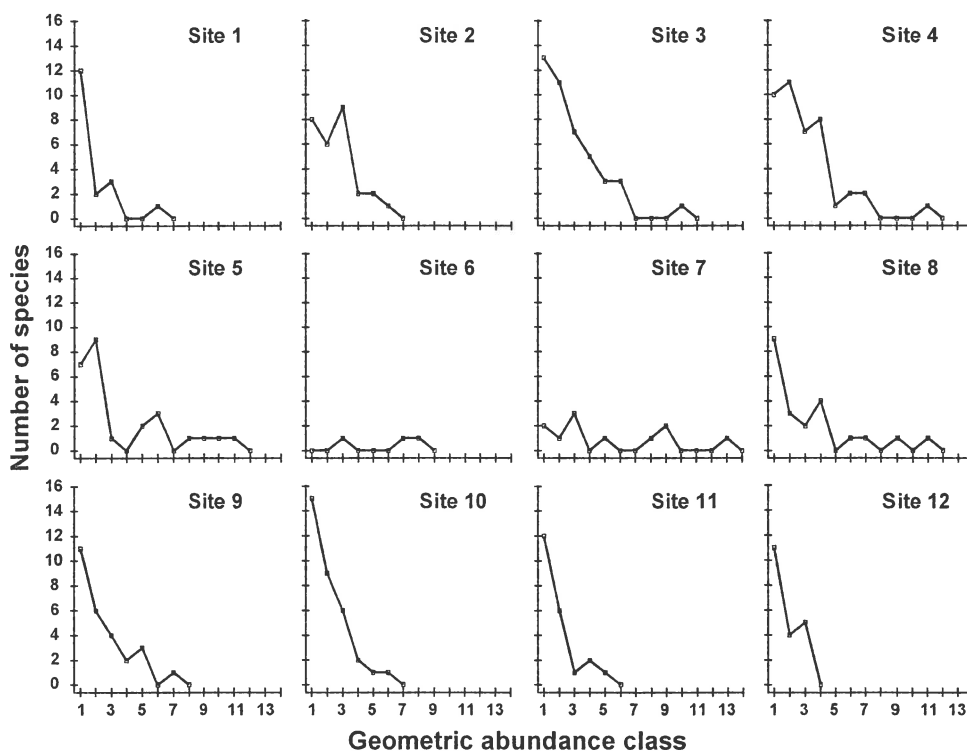


Fig. 1.6. Garroch Head macrofauna {G}. Plots of number of species against number of individuals per species in $\times 2$ geometric classes, for the 12 sampling sites of Fig. 1.5.

[¶] The PRIMER routines automatically offer this set of transformation choices, applied to the whole data matrix, but also cater for more selective transformations of particular sets of variables, as is often appropriate to environmental rather than species data.

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

Stages	Multivariate examples		
	Hierarchical clustering (Ch 2, 3)	MDS ordination (Ch 5)	PCA ordination (Ch 4)
1) Representing communities	<i>Dendrogram of samples</i>	<i>Configuration plot of samples (often 2-dimensional)</i>	
2) Discriminating sites/conditions	<i>ANOSIM on sample similarity matrix (Ch 6)</i> <i>Similarity percentage breakdown (Ch 7) gives species responsible</i>	<i>ANOSIM on Euclidean distances (or multinormal tests, rarely valid)</i>	
3) Determining stress levels	<i>“Meta-analyses”, variability measures, breakdown of seriation (Ch 15)</i>		
4) Linking to environment	<i>Visual: superimposing environmental variables on biotic ordinations (Ch 11)</i> <i>Analytical: finding subset of environmental variables whose ordination “best matches” the biotic ordination (Ch 11, 16)</i> <i>(Causality: see Ch 12)</i>		

Table 1.6. *Frierfjord macrofauna {F}. Bray-Curtis similarities, after $\sqrt{\sqrt{\cdot}}$ 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	–

- 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;
- it requires exclusion of the species which are less common, so that the number of species retained is more 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[†], and

[†] 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.

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.

Discriminating sites/conditions from a multivariate analysis requires non-“classical” hypothesis testing ideas, since it is totally invalid to make the standard assumptions of normality (which in this case would need to be multivariate normality of the 100+ dimensions of the different species!). Instead, Chapter 6 describes a simple permutation or randomisation test (of the type first studied by Mantel, 1967), that makes very few assumptions about the data and is therefore widely applicable. In Fig. 1.7b for example, it is clear without further testing that site A has a different community composition across its replicates than the groups (E, G) or (B, C, D); much less clear is whether there is any statistical evidence of a distinction between the B, C and D sites. A non-parametric test of the null hypothesis of “no site differences between B, C and D” can be constructed by defining a statistic which contrasts between-site and within-site “distances”, then recomputing it for all possible permutations of the 12 labels (4 Bs, 4 Cs and 4 Ds) between the 12 locations on the MDS. If these arbitrary site relabellings generate values of the statistic which are similar

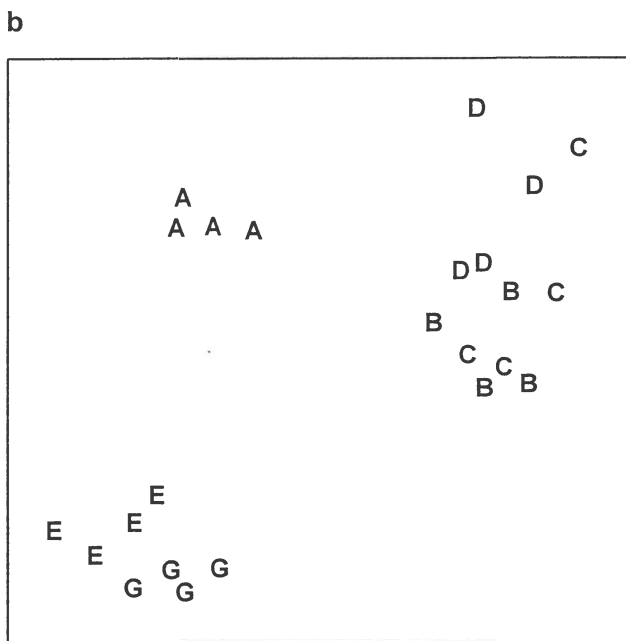
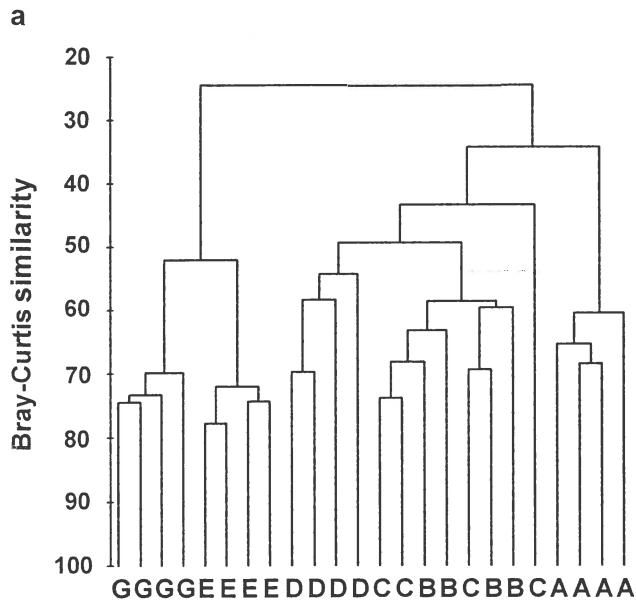


Fig. 1.7. Frierfjord macrofauna {F}. a) Dendrogram for hierarchical clustering (group-average linking); b) non-metric multidimensional scaling (MDS) ordination in two dimensions; both computed for the four replicates from each of the six sites (A–E, G), using the similarity matrix partially shown in Table 1.4.

to the value for the real labelling, then there can clearly be little evidence that the sites are biologically distinguishable. This idea is formalised and extended to more complex sample designs in Chapter 6; for reasons which are described there it is preferable to compute a “between versus within site” summary statistic directly from the (rank) similarity matrix rather than the distances on the MDS plot. This, and the analogy with ANOVA, suggests the term ANOSIM for the test (Analysis of Similarities, Clarke and Green, 1988;

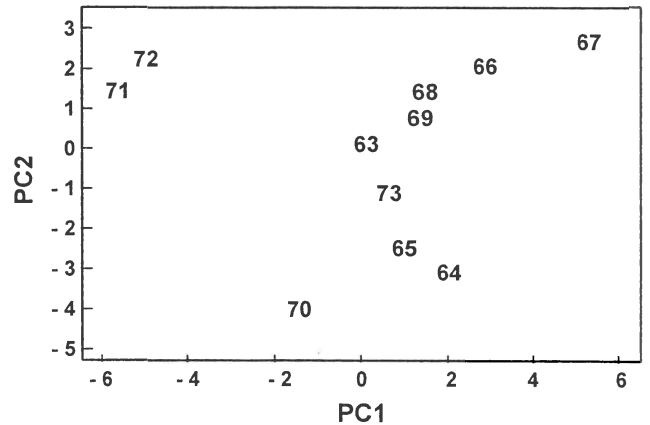


Fig. 1.8. Loch Linnhe macrofauna {L}. Two-dimensional principal components analysis (PCA) ordination of the \sqrt{d} -transformed abundances from the 11 years 1963–1973, omitting the less-common species.

Clarke, 1993).[¶] It is possible to employ the same test in connection with PCA, using an underlying dissimilarity matrix of Euclidean distances, though when the ordination is of a relatively small number of environmental variables, which can be transformed into approximate multivariate normality, then abiotic differences between sites can be 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, as a result of a cluster analysis, one might 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, primarily through a 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

[¶] PRIMER performs tests for one- and two-way crossed and nested designs in the routine ANOSIM, with a special type of two-way crossed design tested in ANOSIM2.

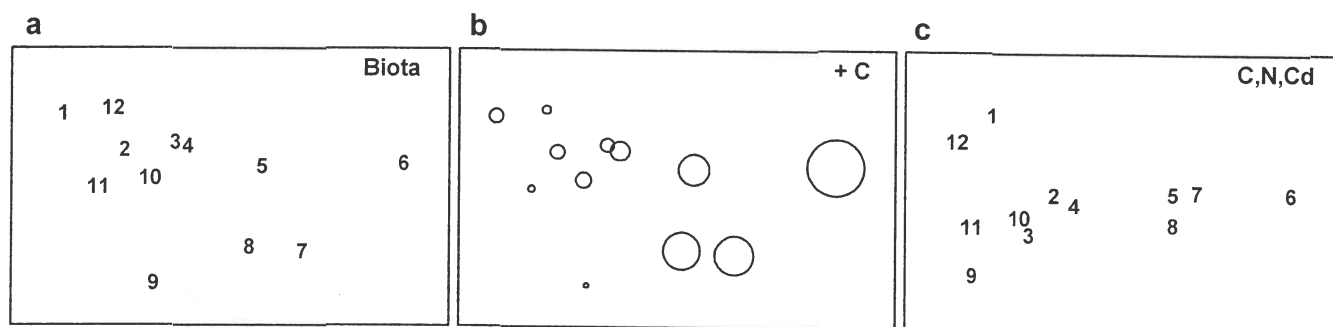


Fig. 1.9. Garroch Head macrofauna {G}. a) MDS ordination of Bray-Curtis similarities from $\sqrt{}$ -transformed species biomass data for the sites shown in Fig. 1.5; b) the same MDS but with superimposed circles of increasing size, representing increasing carbon concentrations in matched sediment samples; c) ordination of (log-transformed) carbon, nitrogen and cadmium concentrations in the sediments at the 12 sites.

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 published 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[†] – 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 spatial coral community study (but with wider applicability to other spatial and temporal patterns), is a loss of smooth *seriation* along transects of increasing depth, again in comparison to controls in time and 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.[‡] A steady change in the community is apparent as the dump centre (site 6) is approached along the western arm of the transect (sites 1 to 6), with a mirrored structure along the eastern 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).[§]

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 subset, of the abiotic variables. Though there is clearly one strong underlying gradient in Fig. 1.9a (horizontal axis), corresponding to an increasing level of organic enrichment, there are nonetheless secondary community differences (e.g. on the vertical axis) which may be amenable to explanation by metal concentration differences, for example. The heuristic approach adopted

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

[‡] 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 marine biota, for a study on the effects of pollutants.

[§] The PRIMER MDS routine allows simple and flexible superimposition of individual variables in this way, whether environmental (as here) or species values.

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).

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

here is to display the *multivariate* pattern of the environmental data, ask to what extent it matches the between-site relationships observed in the biota, and then maximise some *matching coefficient* between the two, by examining possible subsets of the abiotic variables (the BIO-ENV or BVSTEP procedures, Chapters 11 and 16 respectively)[¶].

Fig. 1.9c is based on this optimal subset for the Garroch Head sediment variables, namely (C, N, Cd). It is an MDS plot, using Euclidean distance for its dissimilarities,[†] and is seen to replicate the pattern in Fig. 1.9a rather closely. In fact, the optimal match is determined

[¶] The BIOENV routine in PRIMER optimises the match over all combinations of abiotic variables. Where this is not computationally feasible, the BVSTEP routine performs a stepwise search, adding (or subtracting) single abiotic variables at each step, much as in stepwise multiple regression. This also allows generalisation to pattern-matching scenarios other than abiotic-to-biotic. For example, BVSTEP allows selection of a subset of species whose multivariate structure matches, to a high degree, the pattern for the full set of species (Chapter 16); this provides a more general alternative to the SIMPER procedure (of Chapter 7), for identifying influential species.

[†] It is, though, virtually indistinguishable in this case from a PCA, because of the small number of variables and the implicit use of the same dissimilarity matrix for both techniques.

by correlating the underlying dissimilarity matrices rather than the ordinations themselves, in parallel with the reasoning behind the ANOSIM tests, discussed earlier.

The suggestion is therefore that the biotic pattern of the Garroch Head sites is associated not just with an organic enrichment gradient but also with a particular heavy metal. It is important, however, to realise the limitations of such an “explanation”. Firstly, there are usually other combinations of abiotic variables which will correlate nearly as well with the biotic pattern, particularly as here when the environmental variables are strongly inter-correlated amongst themselves. Secondly, there can be no direct implication of *causality* of the link between these abiotic variables and the community structure, based solely on field survey data: the real driving factors could be unmeasured but happen to correlate highly with the variables identified as producing the optimal match. This is a general feature of inference from purely observational studies and can only be avoided formally by “randomising out” effects of unmeasured variables; this requires random allocation of treatments to observational units for field or laboratory-based community *experiments* (Chapter 12).

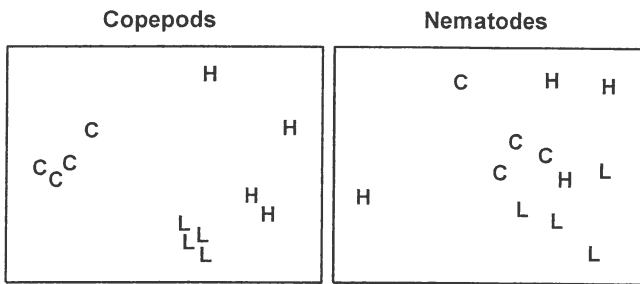


Fig. 1.10. Nutrient enrichment experiment {N}. Separate MDS ordinations of \sqrt{x} -transformed abundances for copepod and nematode species, in four replicate boxes from each of three treatments (C, L, H).

EXAMPLE: Nutrient enrichment experiment, Solbergstrand

An example is given in Table 1.7 of meiofaunal community data from a nutrient enrichment experiment in the Solbergstrand mesocosm, Norway {N}, in which 12 undisturbed box cores of sediment were transferred into the mesocosm basins and separately “dosed” with two levels of increased nutrients (low, L, and high, H),

with some boxes remaining undosed (control, C). Fig. 1.10 shows the MDS plots of the four replicate boxes from each treatment, separately for the copepod and nematode components of the meiofaunal communities (see also Chapter 12). For the copepods, there is a clear imputation of a (causal) response to the treatment, though this is less apparent for the nematodes, and requires a test of the null hypothesis of “no treatment effect”, using the ANOSIM test of Chapter 6.

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 most powerful tools are in the multivariate category, and those that underlie the PRIMER routines 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 of p rows (species) and n columns (samples), whose entries are counts of each species for each sample, or the total biomass of all individuals or their percentage cover, or some other "quantity" of each species in each sample. This includes the special case where only presence (1) or absence (0) of each species is known. 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 rarer species.

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

- a) The absolute numbers (/biomass/cover), 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 (/biomass/cover) are sometimes used, i.e. the data is *standardised* to give the percentage of total abundance (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 and known, abundances are adjusted to a unit sample volume), it may still sometimes be biologically relevant to define two samples as being perfectly similar when they have the same % *composition* of species, fluctuations in total abundance (/biomass /cover) being of no interest. This is not the normal situation, however, changes in total abundance usually having meaningful interpretation in quantitative sampling.

- 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, most biologists would feel that S should equal zero when two samples have no species in common and S must equal 100% if two samples have identical entries (after modification, 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 analysis and in some associated statistical tests. A similarity matrix can be used to:

- discriminate sites (or times) from each other, by noting that similarities between replicates within a site are consistently higher than similarities between replicates at different sites (ANOSIM test, Chapter 6);
- cluster sites into groups that have similar communities, so that similarities within each group of sites are usually higher than those between groups (Clustering, Chapter 3);
- allow a gradation of sites to be represented graphically, in the case where site A has some similarity with site B, B with C, C with D but A and C are less similar, A and D even less so etc. (Ordination, Chapter 4).

Species similarity matrix

In a complementary way, the original data matrix can be thought of as describing the pattern of occurrences of each species across the given set of samples, and a matching triangular array of similarities can be constructed between every pair of species. Two species are “similar” (S' near one) if they have significant representation at the same set of sites, and totally “dissimilar” ($S' = 0$) if they never co-occur. Species similarities are discussed later in this chapter, and the resulting clustering and ordination diagrams in Chapter 7, but for the bulk of this manual “similarity” refers to *between-sample* similarity.

Bray-Curtis coefficient

Of the numerous similarity measures that have been suggested over the years[¶], one has become particularly common in ecology, usually referred to as the **Bray-Curtis** coefficient, since Bray and Curtis (1957) were primarily responsible for introducing this coefficient into ecological work. The similarity between the j th and k th samples, S_{jk} , has two definitions (they are

entirely equivalent, as can be seen from some simple algebra or by calculating a few examples):

$$S_{jk} = 100 \left\{ 1 - \frac{\sum_{i=1}^p |y_{ij} - y_{ik}|}{\sum_{i=1}^p (y_{ij} + y_{ik})} \right\} \quad (2.1)$$

$$= 100 \frac{\sum_{i=1}^p 2 \min(y_{ij}, y_{ik})}{\sum_{i=1}^p (y_{ij} + y_{ik})}$$

Here y_{ij} represents the entry in the i th row and j th column of the data matrix, i.e. the abundance (/biomass /cover) for the i th species in the j th sample ($i = 1, 2, \dots, p$; $j = 1, 2, \dots, n$). Similarly, y_{ik} is the count for the i th species in the k th sample. $|\dots|$ represents the absolute value of the difference (the sign is ignored) and $\min(\dots)$ the minimum of the two counts; the separate sums in the numerator and denominator are both over all rows (species) in the matrix.

EXAMPLE: Loch Linnhe macrofauna

A trivial example, used in this and the following chapter to illustrate simple manual computation of similarities and hierarchical clusters, is provided by extracting six species and four years from the Loch Linnhe macrofauna data $\{L\}$ of Pearson (1975), seen already in Fig. 1.3 and Table 1.4. (Of course, arbitrary extraction of “interesting” species and years is not a legitimate procedure in a real application; it is done here simply as a means of showing the computational steps.)

Table 2.1. *Loch Linnhe macrofauna* $\{L\}$ subset. (a) Abundance (untransformed) for some selected species and years. (b) The resulting Bray-Curtis similarities between every pair of samples.

(a) Year:	64	68	71	73	(b)				
(Sample:	1	2	3	4)	Sample	1	2	3	4
Species					1	–			
<i>Echinoca.</i>	9	0	0	0	2	8	–		
<i>Myrioche.</i>	19	0	0	3	3	0	42	–	
<i>Labidopl.</i>	9	37	0	10	4	39	21	4	–
<i>Amaeana</i>	0	12	144	9					
<i>Capitella</i>	0	128	344	2					
<i>Mytilus</i>	0	0	0	0					

Table 2.1a shows the data matrix of counts and Table 2.1b the resulting lower triangular matrix of Bray-Curtis similarity coefficients. For example, using the first form of equation (2.1), the similarity between samples 1 and 4 (years 1964 and 1973) is:

[¶] Legendre and Legendre (1998), in their excellent and invaluable reference text on all things multivariate in ecology, give definitions of about 25 similarity coefficients!

$$S_{14} = 100 \left\{ 1 - \frac{9 + 16 + 1 + 9 + 2 + 0}{9 + 22 + 19 + 9 + 2 + 0} \right\} = 39.3$$

The second form of equation (2.1) can be seen to give the same result:

$$S_{14} = 100 \left\{ \frac{2[0 + 3 + 9 + 0 + 0 + 0]}{9 + 22 + 19 + 9 + 2 + 0} \right\} = 39.3$$

Computation is therefore simple and it is easy to verify that the coefficient possesses the following desirable properties.

- $S = 0$ if the two samples have no species in common, since $\min(y_{ij}, y_{ik}) = 0$ for all i (e.g. samples 1 and 3 of Table 2.1a). Of course, $S = 100$ if two samples are identical, since $|y_{ij} - y_{ik}| = 0$ for all i .
- 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 cm² of sediment surface to numbers per m²; all y values are simply multiplied by the same constant and this cancels in the numerator and denominator terms of equation (2.1).
- “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”. Note that a lack of dependence on 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 will 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 usually 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 (the counts, biomass, % cover or whatever) *before* computing the Bray-Curtis similarities. Two useful transformations are the *root* transform, \sqrt{y} , and the *double root* (or *4th root*) transform, $\sqrt[4]{y}$. There is more on the effects of transformation later in the manual; for now it is only necessary to note that the root transform, \sqrt{y} , has the effect of down-weighting the importance of the highly abundant species, so that similarities depend not only on their values but also those of less common (“mid-range”) species. The 4th root transform, $\sqrt[4]{y}$, takes this process further, with a more severe down-weighting of the abundant species, allowing not only the mid-range but also the rarer species to exert some influence on the calculation of similarity. An alternative severe transformation, with very similar effect to the 4th root, is the *log* transform, $\log(1+y)$.

The result of the 4th root 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.[¶] 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 (e.g. $S_{24} > S_{14}$ and $S_{34} > S_{12}$ now), showing that transformations *can* have a significant effect on the final ordination or clustering.

Table 2.2. Loch Linnhe macrofauna {L} subset. (a) $\sqrt[4]{y}$ -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	(b)				
(Sample:	1	2	3	4)	Sample	1	2	3	4
Species					1	–			
<i>Echinoca.</i>	1.7	0	0	0	2	26	–		
<i>Myrioche.</i>	2.1	0	0	1.3	3	0	68	–	
<i>Labidopl.</i>	1.7	2.5	0	1.8	4	52	68	42	–
<i>Amaeana</i>	0	1.9	3.5	1.7					
<i>Capitella</i>	0	3.4	4.3	1.2					
<i>Mytilus</i>	0	0	0	0					

[¶] Bray-Curtis is the main coefficient calculated by the PRIMER Similarity routine, which also allows a range of transformations of the data.

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 *balances* the weighting given to each species when computed on original counts (biomass/cover). 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:

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

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 scaling term is inside not outside the summation over species. For example, from Table 2.1a, the Canberra similarity between samples 1 and 4 is:

$$S_{14} = 100 \left\{ 1 - \frac{1}{5} \left(\frac{9}{9} + \frac{16}{22} + \frac{1}{19} + \frac{9}{9} + \frac{2}{2} \right) \right\} = 24.4$$

Note that joint absences have no effect here because they are deliberately excluded (since 0/0 is undefined) and p is reset to be the number of species that are present in at least one of the two samples under consideration.

The separate scaling constrains each species to make equal contribution (potentially) to the similarity between two samples. However abundant a species is, its contribution to S can never be more than $100/p$, and a rare species with a single individual in each of the two samples contributes the same as a common species with 1000 individuals in each. Whilst there may be circumstances in which this is desirable, more often it leads to *overdomination* of the pattern by a large number of rare species, of no real significance. (Often the sampling strategy is incapable of adequately quantifying the rarer species, so that they are distributed arbitrarily, to some degree, across the samples.)

Correlation coefficient

A common statistical means of assessing the relationship between two columns of data (samples j and k here) is the standard *product moment*, or *Pearson*, correlation coefficient:

$$r_{jk} = \frac{\sum_i (y_{ij} - \bar{y}_{.j})(y_{ik} - \bar{y}_{.k})}{\sqrt{\sum_i (y_{ij} - \bar{y}_{.j})^2 \sum_i (y_{ik} - \bar{y}_{.k})^2}} \quad (2.3)$$

where $\bar{y}_{.j}$ is defined as the mean value over all species for the j th sample. In this form it is not a similarity coefficient, since it takes values in the range $(-1, 1)$, not $(0, 100)$, with positive correlation (r near +1) if high counts in one sample match high counts in the other, and negative correlation ($r < 0$) if high counts match absences. There are a number of ways of converting r to a similarity coefficient, the most obvious for community data being $S = 50(1+r)$.

Whilst correlation is sometimes used as a similarity coefficient explicitly in this form, and more often implicitly as the similarity measure underlying certain ordination techniques (e.g. Principal Components Analysis, Chapter 4), it is not particularly suitable for much biological community data, with its plethora of zero values. For example, it violates the criterion that S should not depend on joint absences; here two columns are more highly positively correlated (and give S nearer 100) if species are added which have zero counts for both samples. If correlation *is* to be used a measure of similarity, it makes good sense to transform the data initially, exactly as for the Bray-Curtis computation, so that large counts or biomass do not totally dominate the coefficient.

General suitability of Bray-Curtis

Why does the Bray-Curtis coefficient have such a dominant role in ecological studies? The answer is simple; it is one of the very few measures that satisfy all of the following, practically desirable criteria:

- a) it takes the value 100 when two samples are identical (as do most coefficients);
- b) it takes the value 0 when two samples have no species in common (this is a much tougher condition and most coefficients fail it);
- c) a change of measurement unit does not affect its value (most coefficients pass this one);
- d) its value is unchanged by inclusion or exclusion of a species which is jointly absent from the two samples (another difficult condition to satisfy, and many coefficients fail);
- e) inclusion (or exclusion) of a third sample, C, in the data array makes no difference to the similarity between samples A and B (surprisingly, many coefficients fail this, because they depend on some form

of standardisation carried out for each *species*, either by the species total or maximum value across all samples);

- f) it has the flexibility to register differences in *total* abundance for two samples as a less-than-perfect similarity when the *relative* abundances for all species are identical (some coefficients standardise automatically by sample totals, so cannot reflect this component of similarity/difference).

In addition, Faith *et al* (1987) use a simulation study to look at the robustness of various similarity coefficients in reconstructing a (non-linear) ecological response gradient. They find that Bray-Curtis and a very closely-related modification, the *Kulczynski* coefficient (Kulczynski 1928)

$$S_{jk} = 100 \frac{\sum_{i=1}^p \min(y_{ij}, y_{ik})}{2 / [(\sum_{i=1}^p y_{ij})^{-1} + (\sum_{i=1}^p y_{ik})^{-1}]} \quad (2.4)$$

perform most satisfactorily.

Coefficients other than Bray-Curtis, which satisfy all of the above conditions, either have counterbalancing drawbacks, or are so closely related to Bray-Curtis as to be virtually indistinguishable in most practical applications. An example of the former is the Canberra coefficient, with its forced equal weighting of rare and common species. The latter is exemplified by Kulczynski, which clearly reverts exactly to Bray-Curtis for standardised samples (when the column totals are all 100). Comparing equations (2.1) and (2.4), it is seen only to differ from Bray-Curtis, for non-standardised data, in respect of the form of average used in the denominator term, employing a harmonic rather than arithmetic mean of the column totals. This can only have a substantial influence on the outcome in cases where total abundance (/biomass/cover) is very variable and close to zero for one or more samples, which will also usually be restricted to analyses on untransformed data.

PRESENCE/ABSENCE DATA

As discussed at the beginning of this chapter, quantitative uncertainty may make it desirable to reduce the data simply to presence or absence of each species in each sample, or this may be the only feasible or cost-effective option for data collection in the first place. Alternatively, reduction to presence/absence may be thought of as the ultimate in severe transformation of counts; the data matrix (e.g. in Table 2.1a) is replaced

by 1 (presence) or 0 (absence) and Bray-Curtis similarity (say) computed. This will have the effect of giving potentially equal weight to all species, whether rare or abundant (and will thus have somewhat similar effect to the Canberra coefficient).

Many similarity coefficients have been proposed based on (0, 1) data arrays; see for example, Sneath and Sokal (1973) or Legendre and Legendre (1998). When computing similarity between samples *j* and *k*, the two columns of data can be reduced to the following four summary statistics without any loss of relevant information:

- a* = the number of species which are present in both samples;
- b* = the number of species present in sample *j* but absent from sample *k*;
- c* = the number of species present in sample *k* but absent from sample *j*;
- d* = the number of species absent from both samples.

For example, when comparing samples 1 and 4 from Table 2.1a, these frequencies are:

		Sample 4:	
		1	0
Sample 1:	1	<i>a</i> = 2	<i>b</i> = 1
	0	<i>c</i> = 2	<i>d</i> = 1

In fact, because of the symmetry, coefficients must be a symmetric function of *b* and *c*, otherwise S_{14} will not be equal to S_{41} . Similarly, similarity measures that are not affected by joint absences will not contain *d*. The following are some of the more commonly advocated coefficients.

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

$$S_{jk} = 100[(a + d)/(a + b + c + d)] \quad (2.5)$$

so called because it represents the probability ($\times 100$) 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:

$$S_{jk} = 100[a/(a + b + c)] \quad (2.6)$$

i.e. S is the probability ($\times 100$) 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

$$S_{jk} = 100[2a/(2a + b + c)] \quad (2.7)$$

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).[¶] 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).

Several other coefficients have been proposed; Legendre and Legendre (1998) list at least 15, but only one further measure is given here. In the light of the earlier discussion on coefficients satisfying desirable, biologically-motivated criteria, note that there is a presence/absence form of the *Kulczynski* coefficient (2.4), a close relative of Bray-Curtis, namely:

$$S_{jk} = 50 \left(\frac{a}{a+b} + \frac{a}{a+c} \right) \quad (2.8)$$

RECOMMENDATIONS

- 1) In most ecological studies, some intuitive axioms for desirable practical behaviour of a similarity coefficient lead inexorably to the use of the Bray-Curtis measure (or a closely-related coefficient such as that of Kulczynski).
- 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.

[¶] Thus the Sorenson coefficient can be obtained in the PRIMER Similarity routine by “transforming” the data to presence/absence and selecting Bray-Curtis similarity.

- 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 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 the Bray-Curtis similarity to counts (/biomass/cover values) which have been moderately, \sqrt{y} , or fairly severely transformed, $\log(1+y)$ or $\sqrt[3]{y}$. 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 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 standardisation should usually be avoided, valuable biological information being contained in the abundance, biomass or cover 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.[†]

SPECIES SIMILARITIES

Starting with the original data matrix of abundances (or biomass, % cover etc), the similarity between any pair of *species* can be defined in an analogous way to that for samples, but this time involving comparison of the i th and l th row (species) across all $j = 1, \dots, n$ columns (samples).

[†] In the PRIMER Similarity routine, standardisation is not the default option for sample similarities but, if selected, it is therefore carried out before any transformation.

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.[¶] 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 downweights 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:

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

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

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

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

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).

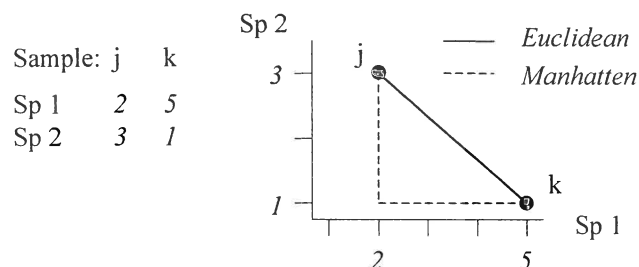
[¶] The PRIMER Similarity routine will compute Bray-Curtis *species* similarities, with or without row standardisation and transformation (though the default is as recommended here). Prior to this, the Select Variables option allows reduction of the number of species, by retaining those that contribute $p\%$ or more to at least one of the samples, or by specifying the number n of “most important” species to retain. The latter uses the same $p\%$ criterion but gradually increases p until only n species are left.

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:

$$d_{jk} = \sqrt{\sum_{i=1}^p (y_{ij} - y_{ik})^2} \quad (2.13)$$

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 simple 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 $\sqrt{[(2-5)^2 + (3-1)^2]}$ (from 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, where it is crucial to an understanding of Principal Components Analysis.

Manhattan distance

Euclidean distance is not the only way of defining distance apart of two samples in species space; an alternative is to sum the distances along each species axis:

$$d_{jk} = \sum_{i=1}^p |y_{ij} - y_{ik}| \quad (2.14)$$

This is often referred to as *Manhattan* (or *city-block*) distance because in two dimensions it corresponds to the distance you would have to travel to get between any two locations in a city whose streets are laid out in a rectangular grid. It is illustrated for the simple example above by the dashed lines. Manhattan distance is of interest here because of its obvious close affinity to Bray-Curtis dissimilarity, equation (2.12). In fact, when a data matrix has initially been column standardised (but not transformed), Bray-Curtis dissimilarity is just (half) the Manhattan distance, since the summation in the bottom line of (2.12) then always takes the value 200.

[It is worth noting a point of terminology in passing, though not one of any great practical consequence. Euclidean and Manhattan measures, (2.13) and (2.14), are called distances or *metrics* because they obey the triangle inequality, i.e. for any three samples j, k, r :

$$d_{jk} + d_{kr} \geq d_{jr} \quad (2.15)$$

Bray-Curtis dissimilarity does not, in general, satisfy the triangle inequality, so should not be called a metric. However, many other useful dissimilarity coefficients are also not metrics. For example, the square of Euclidean distance (i.e. equation (2.13) without the $\sqrt{\quad}$ sign) is another natural definition of "distance" which is not

a metric, yet dissimilarities from this would have the same rank order as those from Euclidean distance and therefore give rise, for example, to identical MDS ordinations (see Chapter 5). It follows that whether a dissimilarity coefficient is, or is not, a metric is likely to be of limited practical significance for the strategy this manual advocates.]

RECOMMENDATION

There are thus a variety of means of generating a similarity or dissimilarity (/distance) matrix to input to the next stage of a multivariate analysis, which might be either a clustering or ordination of samples, Fig. 2.1. For comparative purposes it may sometimes be of interest to use Euclidean distance in the species space as input to a cluster analysis[¶] (an example is given later in Fig. 5.5) but, in general, the recommendation remains unchanged: Bray-Curtis similarity/dissimilarity, computed after suitable transformation, will often be a satisfactory coefficient for *biological* data on community structure. Background physical or chemical data is a different matter since it is usually of a rather different type, and Chapter 11 shows the usefulness of the concept of Euclidean distance in the (normalised) *environmental variable space*. Initially though, concentration is on analysing the biological data in isolation, and the next stage will often be to perform a cluster analysis (Fig 2.1).

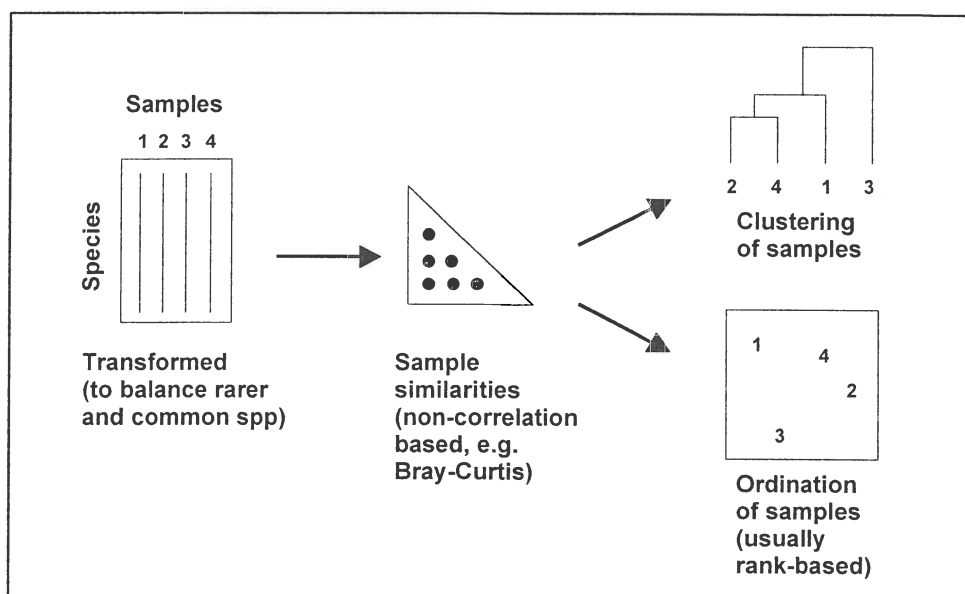


Fig. 2.1. Stages in a multivariate analysis based on similarity coefficients.

[¶] The PRIMER Similarity routine can generate Euclidean distances (normalised or not, see page 4-6), on either biotic or environmental input matrices.



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{x} -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	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	-

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, i.e. 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.**

Cormack (1971) also warned against the indiscriminate use of cluster analysis: “availability of ... classification techniques has led to the waste of more valuable scientific time than any other ‘statistical’ innovation”. The

ever larger number of techniques and their increasing accessibility on modern computer systems makes this warning no less pertinent today. The policy adopted here is to concentrate on a single technique that has been found to be of widespread utility in ecological studies, whilst emphasising the potential arbitrariness in all classification methods and stressing the need to perform a cluster analysis in conjunction with a range of other techniques (e.g. ordination, statistical testing) to obtain balanced and reliable conclusions.

HIERARCHICAL AGGLOMERATIVE CLUSTERING

The most commonly used clustering techniques are the *hierarchical agglomerative* methods. These usually take a similarity matrix as their starting point and successively *fuse* the samples into groups and the groups into larger clusters, starting with the highest mutual similarities then gradually lowering the similarity level at which groups are formed.[¶] The process ends with a single cluster containing all samples. Hierarchical *divisive* methods perform the opposite sequence, starting with a single cluster and *splitting* it to form successively smaller groups.

The result of a hierarchical clustering is represented by a tree diagram or *dendrogram*, with the *x* axis representing the full set of samples and the *y* axis defining a similarity level at which two samples or groups are considered to have fused. Note that there is no firm convention for which way up the dendrogram should be portrayed (increasing or decreasing *y* axis values) or even whether the tree can be placed on its side; all three possibilities can be found in this manual!

Fig. 3.1 shows a dendrogram for the similarity matrix from the Frierfjord macrofaunal abundances, a subset of which is shown in Table 3.1. It can be seen that all four replicates from sites A, D, E and G fuse with each other to form distinct site groups *before* they amalgamate with samples from any other site; that, conversely, site B and C replicates are not distinguished, and that A, E and G do not link to B, C and D until quite low levels of between-group similarities are reached.

[¶] The PRIMER CLUSTER routine displays the dendrogram from hierarchical agglomerative clustering, allowing a choice from the three linkage possibilities described below and various options for axis labelling, orientation etc.

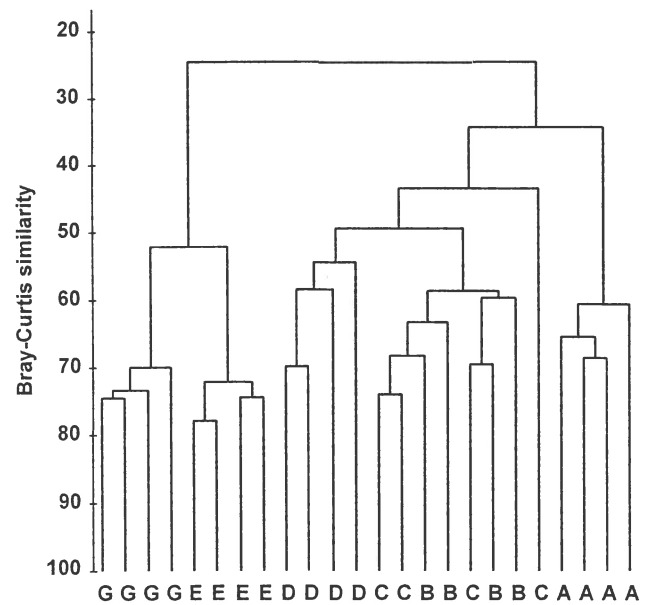


Fig. 3.1. Frierfjord macrofauna counts {F}. Dendrogram for hierarchical clustering (using group-average linking) of four replicate samples from each of sites A-E, G, based on the Bray-Curtis similarity matrix shown (in part) in Table 3.1.

The mechanism by which Fig. 3.1 is extracted from the similarity matrix, including the various options for defining what is meant by the similarity of two groups of samples, is best described for a simpler example.

Construction of dendrogram

Table 3.2 shows the steps in the successive fusing of samples, for the subset of Loch Linnhe macrofaunal abundances used as an example in the previous chapter. The data matrix has been $\sqrt{}$ -transformed, and the first triangular array is the Bray-Curtis similarity of Table 2.2.

Samples 2 and 4 are seen to have the highest similarity (underlined) so they are combined, at similarity level 68.1%. (Above this level there are considered to be four clusters, simply the four separate samples.) A new similarity matrix is then computed, now containing three clusters: "1", "2&4" and "3". The similarity between cluster "1" and cluster "3" is unchanged at 0.0 of course but what is an appropriate definition of similarity $S(1, 2\&4)$ between clusters "1" and "2&4", for example? This will be some function of the similarities $S(1,2)$, between samples 1 and 2, and $S(1,4)$, between 1 and 4; there are three main possibilities here.

Table 3.2. Loch Linnhe macrofauna {L} subset. Abundance array after $\sqrt{4}$ -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		Sample	1	2&4	3		Sample	1	2&3&4
Sample:	1	2	3	4	→	1	–				→	1	–			→	1	–	
Species						2	25.6	–				2&4	38.9	–			2&3&4	25.9	–
<i>Echinoca.</i>	1.7	0	0	0		3	0.0	67.9	–			3	0.0	55.0	–				
<i>Myrioche.</i>	2.1	0	0	1.3		4	52.2	68.1	42.0	–									
<i>Labidopl.</i>	1.7	2.5	0	1.8															
<i>Amaeana</i>	0	1.9	3.5	1.7															
<i>Capitella</i>	0	3.4	4.3	1.2															
<i>Mytilus</i>	0	0	0	0															

- 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:

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

$$= (2 \times 38.9 + 1 \times 0)/3 = 25.9$$

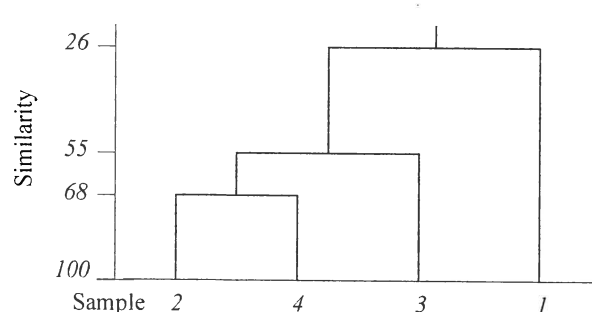
Though it is computationally efficient to form each successive similarity matrix by taking weighted averages of the similarities in the previous matrix, an alternative which is entirely equivalent (and perhaps conceptually simpler) is to define the similarity between two groups as the *simple* (unweighted) 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 be 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 would be with a child’s “mobile”; the vertical lines are strings and the horizontal lines rigid bars. When the 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 (new sequence 4, 2, 3, 1) or sample 1 move to the opposite side of the diagram (new 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-relations 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 “opportunists” like *Capitella*), to the start of a reversion to an

improved condition in 1973. Of course it is unwise and unnecessary to attempt serious interpretation of such a small subset of data but, even so, the equivalent MDS ordination for this subset (met in Chapter 5) contrasts well with the relatively unhelpful information in the above dendrogram. (A PCA ordination of the full data set can be seen in Fig. 4.1.)

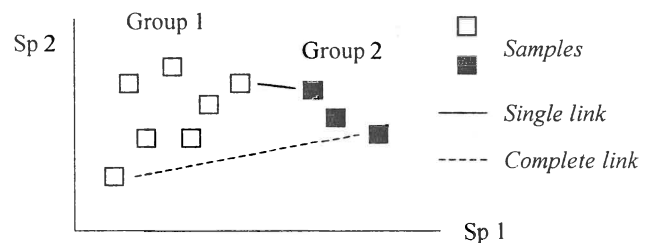
- 4) The hierarchical nature of this clustering procedure dictates that, once a sample is grouped with others, it will never be separated from them in a later stage of the process. Thus, early borderline decisions which may be somewhat arbitrary are perpetuated through the analysis and may sometimes have a significant effect on the shape of the final dendrogram. For example, similarities $S(2, 3)$ and $S(2, 4)$ above are very nearly equal. Had $S(2, 3)$ been just greater than $S(2, 4)$, rather than the other way round, the final picture would have been a little different. In fact, the reader can verify that had $S(1, 4)$ been around 56% (say), the same marginal shift in the values of $S(2, 4)$ and $S(2, 3)$ would have had radical consequences, the final dendrogram now grouping 2 with 3 and 1 with 4 before these two groups come together in a single cluster. From being the first to be joined, samples 2 and 4 now only link up at the final step. Such situations are certain to arise if, as here, one is trying to force what is essentially a steadily changing pattern into discrete clusters.

Dissimilarities

Exactly the converse operations are needed when clustering from a dissimilarity rather than a similarity matrix. The two samples or groups with the *lowest* dissimilarity at each stage are fused. The single linkage definition of dissimilarity of two groups is the *minimum* dissimilarity over all pairs of samples between groups; complete linkage selects the *maximum* dissimilarity and group-average linking involves just an unweighted mean dissimilarity.

Linkage options

The differing consequences of the three linkage options are most easily seen for the special case used in Chapter 2, where there are only two species (rows) in the original data matrix. Samples are then points in the *species space*, with the (x,y) axes representing abundances of (Sp.1, Sp.2) respectively. Consider also the case where dissimilarity between two samples is defined simply as their (Euclidean) distance apart in this plot.



In the above diagram, the single link dissimilarity between Groups 1 and 2 is then simply the minimum distance apart of the two groups, giving rise to an alternative name for the single linkage, namely *nearest neighbour* clustering. Complete linkage dissimilarity is clearly the maximum distance apart of any two samples in the different groups, namely *furthest neighbour* clustering. Group-average dissimilarity is the mean distance apart of the two groups, averaging over all between-group pairs.

Single and complete linkage have some attractive theoretical properties. For example, they are effectively *non-metric*. Suppose that the Bray-Curtis (say) similarities in the original triangular matrix are replaced by their *ranks*, i.e. the highest similarity is given the value 1, the next highest 2, down to the lowest similarity with rank $n(n-1)/2$ for n samples. Then a single (or complete) link clustering of the ranked matrix will have the exactly the same *structure* as that based on the original similarities (though the y axis similarity scale in the dendrogram will be transformed in some non-linear way). This is a desirable feature since the precise similarity values will not often have any direct significance; what matters is their relationship to each other and any non-linear (monotonic) rescaling of the similarities would ideally not affect the analysis. This is also the stance taken for the preferred ordination technique in this manual's strategy, the method of non-metric multi-dimensional scaling (MDS, see Chapter 5).

However, in practice, single link clustering has a tendency to produce chains of linked samples, with each successive stage just adding another single sample onto a large group. Complete linkage will tend to have the opposite effect, with an emphasis on small clusters at the early stages. (These characteristics can be reproduced by experimenting with the special case above, generating nearest and furthest neighbours in a 2-dimensional species space). Group-averaging, on the other hand, is often found empirically to strike a balance in which a moderate number of medium-sized clusters are produced, and only grouped together at a later stage.

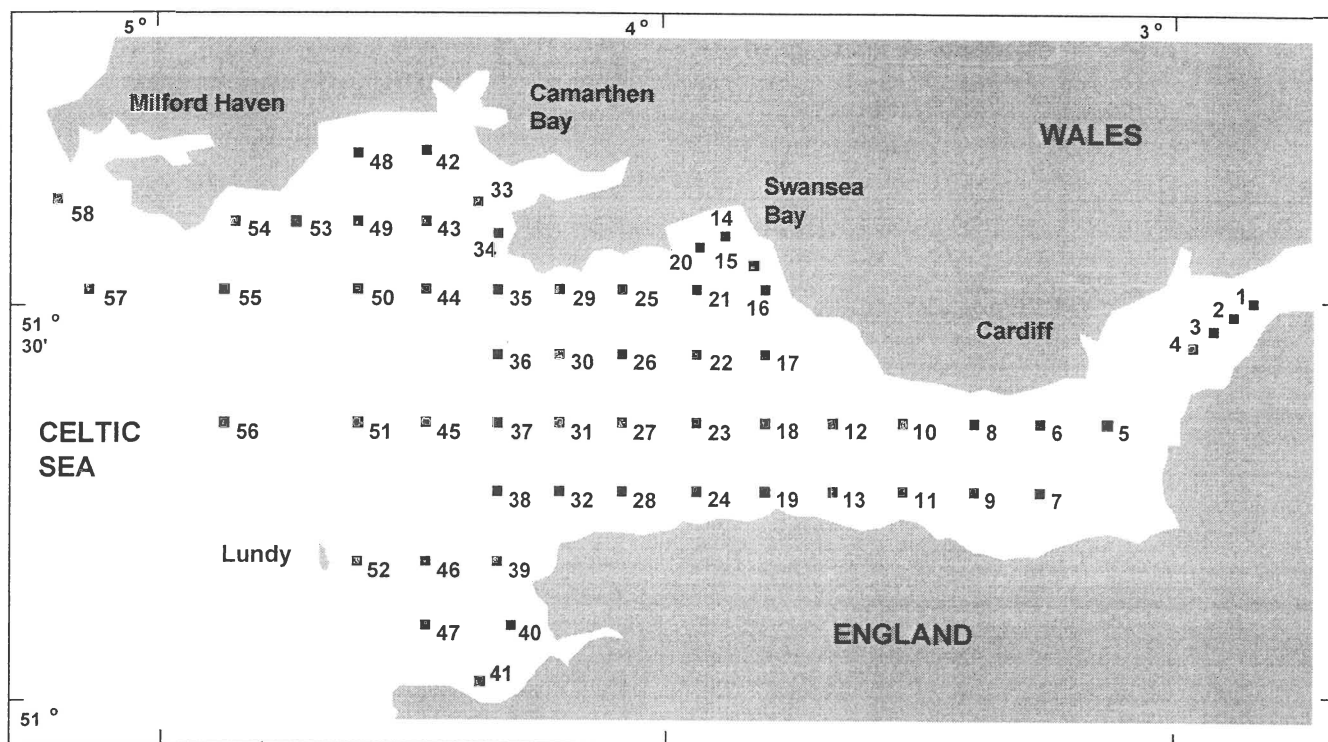


Fig. 3.2 Bristol Channel zooplankton {B}. Sampling sites.

EXAMPLE: Bristol Channel zooplankton

Collins and Williams (1982) perform hierarchical cluster analyses of zooplankton samples, collected by double oblique net hauls at 57 sites in the Bristol Channel UK, for three different seasons in 1974 {B}. This was 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 $\sqrt{\sqrt{\text{transformed abundances}}}$. 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 similarities or dissimilarities such as Bray-Curtis, has proved a useful technique in a number of ecological studies of the last three decades. It is appropriate for delineating groups of sites with

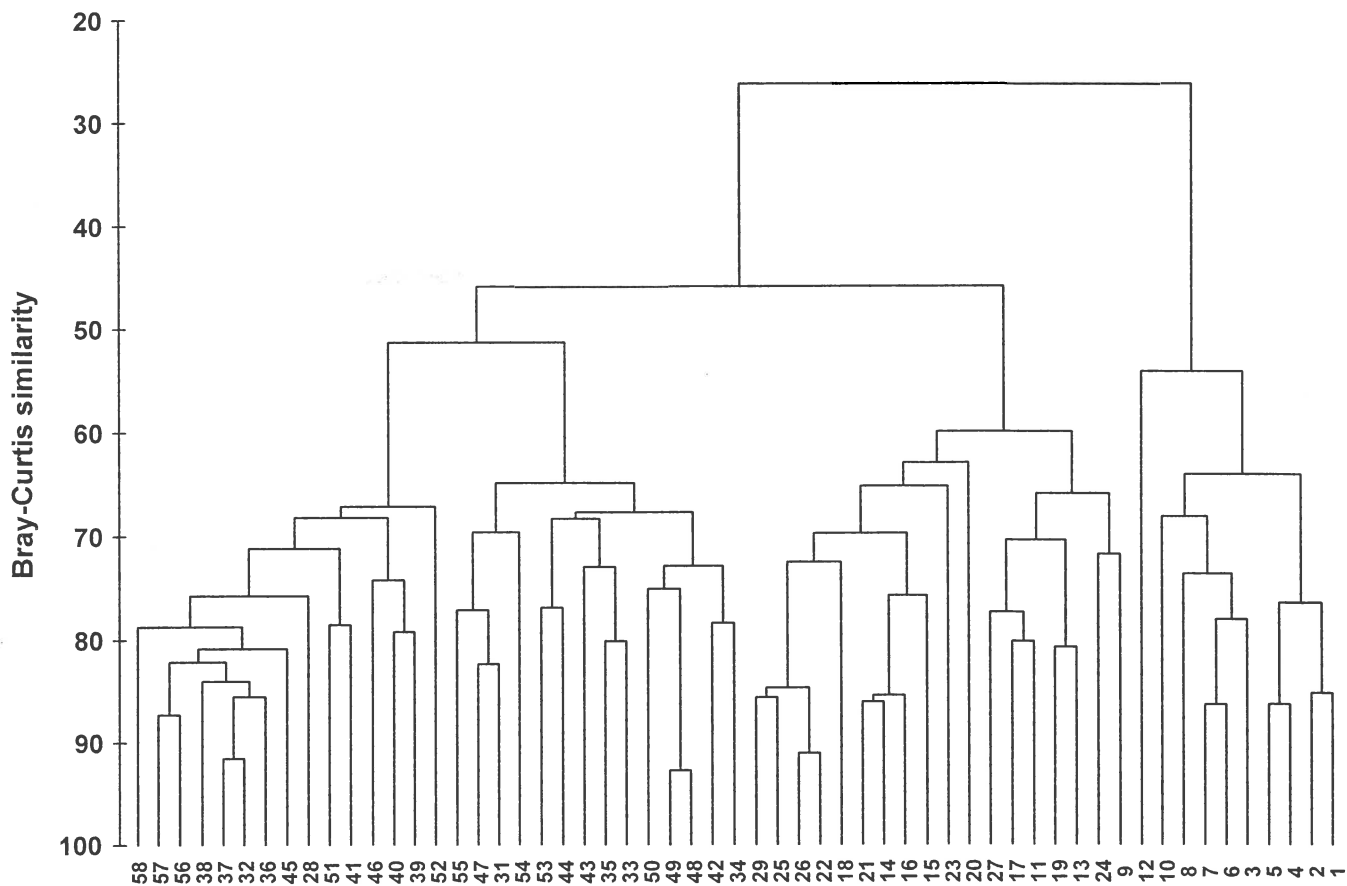


Fig. 3.3. Bristol Channel zooplankton {B}. Dendrogram for hierarchical clustering of the 57 sites, using group-average linking of Bray-Curtis similarities calculated on $\sqrt{\sqrt{\cdot}}$ -transformed abundance data.

distinct community structure (this is not to imply that groups have no species in common, of course, but that different characteristic patterns of abundance are found consistently in different groups).

2) Clustering is less useful (and could sometimes be misleading) where there is a steady gradation in community structure across sites, perhaps in response to strong environmental forcing (e.g. large range of salinity, sediment grain size, depth of water column, etc.).

Ordination is preferable in these situations.

3) Even for samples which are strongly grouped, cluster analysis is often best used in conjunction with ordination. Superimposition of the clusters (at various levels of similarity) on an ordination plot will allow any relationship between the groups to be more informatively displayed, and it will be seen later (Chapter 5) that agreement between the two representations strengthens belief in the adequacy of both.

CHAPTER 4: ORDINATION OF SAMPLES BY PRINCIPAL COMPONENTS 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 (arguably) the simplest 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

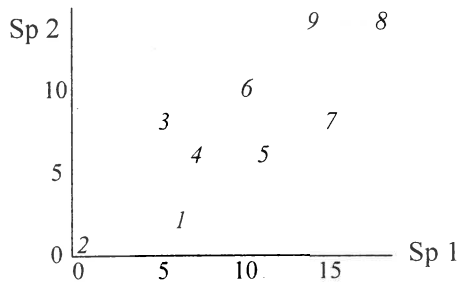
The starting point for PCA is the original data matrix rather than a derived similarity matrix (though there is an *implicit* dissimilarity matrix underlying PCA, that of Euclidean distance). The data array is thought of as defining the positions of samples in relation to axes representing the full set 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

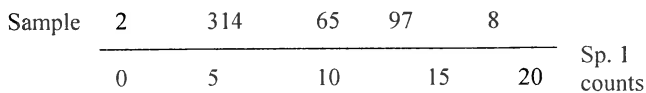
It helps to visualise the process by again considering an (artificial) example in which there are only two species (and nine samples).

	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

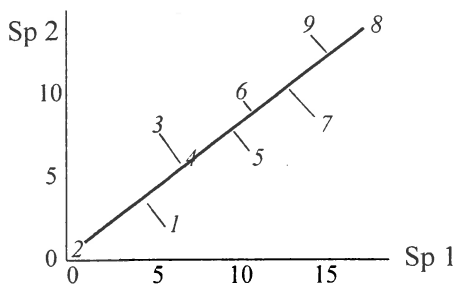
The nine samples are therefore points in two dimensions, and labelling these points with the sample number gives the following plot.



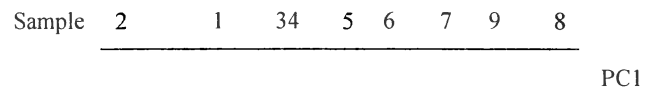
This is an ordination already, of 2-dimensional data on a 2-dimensional map, and it summarises pictorially all the relationships between the samples, without needing to discard any information at all. However, suppose for the sake of example that a 1-dimensional ordination is required, in which the original data is reduced to a genuine *ordering* of samples along a line. How do we best place the samples in order? One possibility (though a rather poor one!) is simply to ignore altogether the counts for one of the species, say Species 2. The Species 1 axis then automatically gives the 1-dimensional ordination (Sp.1 counts are again labelled by sample number):



(Think of this as projecting the points in the 2-dimensional space down onto the Sp.1 axis). Not surprisingly, this is a rather inaccurate 1-dimensional summary of the sample relationships in the full 2-dimensional data, e.g. samples 7 and 9 are rather too close together, certain samples seem to be in the "wrong order" (9 should be closer to 8 than 7 is, 1 should be closer to 2 than 3 is) etc. More intuitively obvious would be to choose the 1-dimensional picture as the (perpendicular) projection of points onto the line of "best fit" in the 2-dimensional plot.

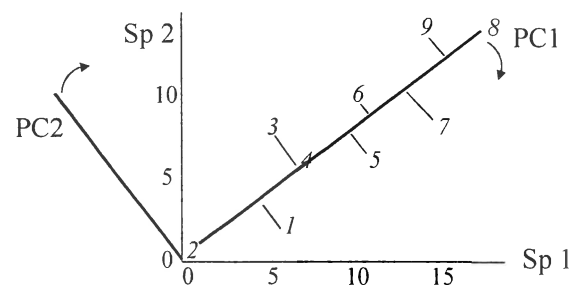


The 1-dimensional ordination, called the *first principal component axis (PC1)*, is then:

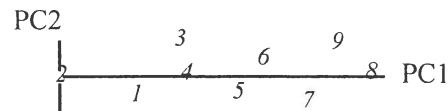


and this picture is a much more realistic approximation to the 2-dimensional sample relationships (e.g. 1 is now closer to 2 than 3 is, 7, 9 and 8 are more equally spaced and in the "right" sequence etc).

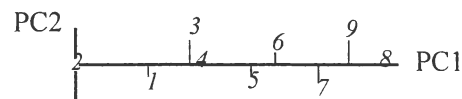
The *second principal component axis (PC2)* is defined as the axis *perpendicular* to PC1, and a full *principal component analysis* then consists simply of a *rotation* of the original 2-dimensional plot:



to give the following *principal component plot*.



Obviously the (PC1, PC2) plot contains exactly the same information as the original (Sp.1, Sp.2) graph. The whole point of the procedure though is that, as in the current example, we may be able to *dispense* with the second principal component (PC2): the points in the (PC1, PC2) space are projected onto the PC1 axis and relatively little information about the sample relationships is lost in this reduction of dimensionality.



Definition of PC1 axis

Up to now we have been rather vague about what is meant by the "best fitting" line through the sample points in 2-dimensional species space. There are two natural definitions. The first chooses the PC1 axis as the line which minimises the sum of squared perpend-

icular distances of the points from the line.[¶] 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.

Extensions to 3-dimensional data

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

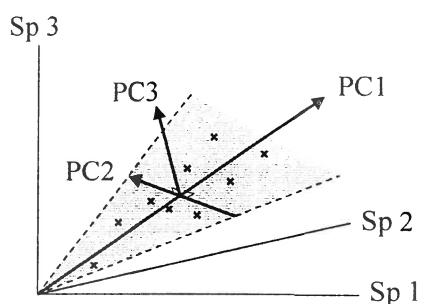
	Sample	1	2	3	4	5	6	7	8	9
Abundance	Sp.1	6	0	5	7	11	10	15	18	14
	Sp.2	2	0	8	6	6	10	8	14	14
	Sp.3	3	1	6	6	9	11	10	16	15

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).



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:

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

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.

Variance 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 $var(PC_i)$ denote variance of samples on the i th PC axis and $var(Sp.i)$ denote variance of points on the i th species axis ($i = 1, 2, 3$):

$$\sum_i var(PC_i) = \sum_i var(Sp.i) \quad (4.2)$$

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

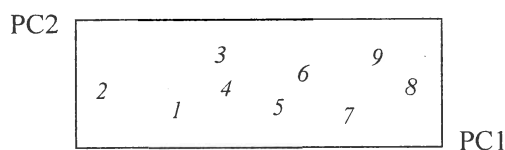
$$P_i = 100 \frac{var(PC_i)}{\sum_i var(PC_i)} = 100 \frac{var(PC_i)}{\sum_i var(Sp.i)} \quad (4.3)$$

[¶] 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.

has a useful interpretation as the % of the original total *variance explained* by the *i*th PC. For the simple 3-dimensional example above, PC1 explains 93%, PC2 explains 6% and PC3 only 1% of the variance in the original samples.

Ordination plane

This brings us back finally to the reason for rotating the original three species axes to three new principal component axes. The first two PCs represent a plane of “best fit”, encompassing the maximum amount of variation in the sample points. The % *variance explained* by PC3 may be small and we can dispense with this third axis, projecting all points perpendicularly onto the (PC1, PC2) plane to give the 2-dimensional *ordination plane* that we have been seeking. For the above example this is:



and it is almost a perfect 2-dimensional summary of the 3-dimensional data, since PC1 and PC2 account for 99% of the total variation. In effect, the points lie on a plane (in fact, nearly on a line!) in the original species space, so it is no surprise to find that this PCA ordination differs negligibly from that for the initial 2-species example: the counts added for the third species were highly correlated with those for the first two species.

Higher-dimensional data

Of course there are many more species than three in a normal species by samples array, typically at least 30, but the approach to defining principal components and an ordination plane is the same. Samples are now points in (say) a 30-dimensional species space and the “best fitting” 2-dimensional plane is found and samples “projected” onto it to give the 2-dimensional PCA ordination. The PC axes are the “perpendicular directions” in this 30-dimensional space along which the variances of the points are (successively) maximised. The degree to which a 2-dimensional PCA succeeds in representing the full 30-dimensional information is seen in the percentage of total variance explained by the first two (of 30) principal components. Often PC1 and PC2 may not explain more than 40-50% of the total variation, and a 2-dimensional PCA ordination gives an inadequate and potentially misleading picture

of the relationship between the samples. A 3-dimensional sample ordination, using the first three PC axes, may give a fuller picture or it may be necessary to invoke PC4, PC5 etc. before a reasonable percentage of the total variation is encompassed. Guidelines for an acceptable level of “% variance explained” are difficult to set, since they depend on the objectives of the study, the number of species and samples etc., but an empirical rule-of-thumb might be that a picture which accounts for as much as 70-75% of the original variation is likely to describe the overall structure rather well.

The geometric concepts of fitting planes and projecting points in (say) 30-dimensional space are not ones that most people are comfortable with (!) so it is important to realise that, *algebraically*, the central ideas are no more complex than in three dimensions. Equations like (4.1) simply extend to *p* (=30) principal components, each a linear function of the *p* species counts. The “perpendicularity” (*orthogonality*) of the principal component axes is reflected in the zero values for all sums of cross-products of coefficients, e.g. for equation (4.1):

$$(0.62) \times (-0.73) + (0.52) \times (0.65) + (0.58) \times (0.20) = 0$$

$$(0.62) \times (0.28) + (0.52) \times (0.55) + (0.58) \times (-0.79) = 0$$

etc.,

the coefficients being scaled so that their sum of squares adds to one, e.g.

$$(0.62)^2 + (0.52)^2 + (0.58)^2 = 1$$

$$(-0.73)^2 + (0.65)^2 + (0.20)^2 = 1$$

etc.

There is clearly no difficulty in extending such relations to 4, 5 or any number of coefficients.

The algebraic construction of coefficients satisfying these conditions but also defining which “perpendicular directions” maximise variation of the samples in the species space, is outside the scope of this manual. It involves calculating eigenvalues and eigenvectors of a *p* by *p* matrix, see Chatfield and Collins (1980), for example. (Note that a knowledge of matrix algebra is essential to understanding this construction). The advice to the reader is to hang on to the geometric picture: all the essential ideas of PCA are present in visualising the construction of a 2-dimensional ordination plot from a 3-dimensional species space.

EXAMPLE: Loch Linnhe macrofauna

The various options available, and the limitations imposed when constructing an ordination using PCA, are best appreciated in the context of a real data set. A 2-dimensional PCA of the full Loch Linnhe macrofaunal 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 the 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

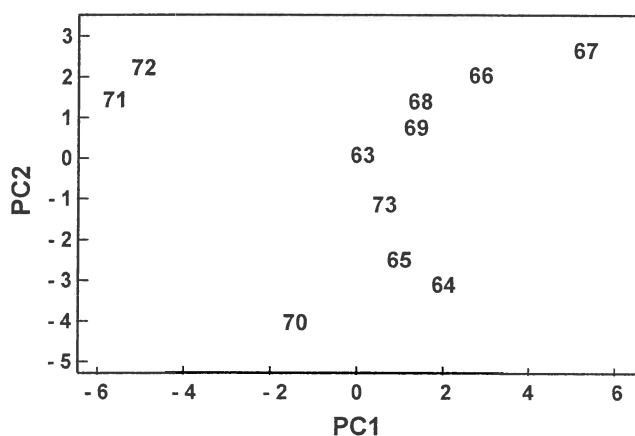


Fig. 4.1. Loch Linnhe macrofauna $\{L\}$. 2-dimensional PCA ordination of sample abundances ($\sqrt{\sqrt{\cdot}}$ -transformed) from the 11 years 1963-1973. PC1 (x-axis) and PC2 (y-axis) together account for 57% of the total sample variability.

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 discussed in the section on species similarities in Chapter 2). Calculations 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 taken 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 $\sqrt{\sqrt{\cdot}}$ -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 (the terminology comes from whether the algebraic extraction of eigenvalues and eigenvectors takes place on the correlation or covariance matrix between species).

When does one use normalisation and when transformation (or both)? In fact, the arguments are somewhat analogous to those seen in Chapter 2 for the computation of similarities. There, techniques which tended to weight all species equally (for example, the calculation of Canberra coefficients) were rejected in favour of methods that maintained a greater (though not overwhelmingly greater) importance for common species than rarer ones. This was achieved through initial $\sqrt{\cdot}$, $\sqrt[3]{\cdot}$ or log transformation; the equivalent option here would be to take the same transformation and apply covariance-based PCA. It is true that, here, the set of species has first been drastically reduced so that all the rarer species are eliminated; nonetheless there seems no compelling reason why the remaining species should be given equal weight, as they would in a correlation-based PCA.

Note that even if normalisation is used, it still makes sense to perform an initial transformation. This has the effect of reducing the inevitable *right-skewness* of the spread of sample counts along a species axis (i.e. abundances tend to bunch at smaller values with a long “tail” of occasional large counts). Computation of variances and the resulting normalisation are really designed to cope with data which are, as the name implies, approximately *normally* distributed; transformation makes this more likely.

PCA OF ENVIRONMENTAL DATA

The conclusion above is that covariance-based PCA would probably be preferred to correlation-based PCA for species abundance matrices, though the summary at the end of this chapter makes it clear that neither is a very satisfactory ordination method for such data. There is one important situation, however, where PCA is a more useful tool and where normalisation is usually *essential*. This is in the multivariate analysis of environmental rather than species data. In conventional statistical notation, one has a matrix of p columns

(variables) by n rows (samples)[¶], with the variables perhaps being a mixture of physical parameters (grain size, salinity, depth of the water column) and other environmental or chemical measurements (nutrient levels, heavy metal, hydrocarbon or PCB concentrations etc). Patterns in the environmental data across samples can be displayed in an analogous way to species data, by a multivariate ordination, and techniques for linking the biological and environmental summaries are discussed in Chapter 11.

PCA is more appropriate to environmental variables because of the form of the data (there are no large blocks of zero counts needing to be treated in a special way: it is no longer necessary to select a dissimilarity coefficient that ignores “joint absences” and some form of Euclidean distance measure makes more sense for environmental data). However, a crucial difference between species and environmental data is that the latter will usually have a complete mix of measurement scales (salinity in ‰, grain size in ϕ units, depth in m, metal concentrations in $\mu\text{g/g}$, PCBs in ng/g etc). In a multi-dimensional visualisation of the environmental data matrix, the samples are points referred to *environmental axes* rather than species axes, but what does it mean now to talk about (Euclidean) distance between two sample points in the *environmental variable space*? If the units on each axis differ, and have no natural connection with each other, then point A can be made to appear closer to point B than point C, or closer to point C than point B, simply by a change of scale on one of the axes (e.g. measuring PCBs in $\mu\text{g/g}$ not ng/g , or depth in fathoms rather than m). Obviously it would be entirely wrong for the PCA ordination to vary with such arbitrary scale changes. There is one natural solution to this: perform a *correlation-based PCA*, i.e. normalise all axes (after transformations, if any) so that they have comparable, dimensionless scales.

The problem does not generally arise for species data, of course, because though a scale change on the axes may be contemplated (e.g. changing counts from numbers of individuals per core to numbers per m^2 of sediment surface), the *same* scale change is made on each axis and the PCA ordination will be unaffected. The overall guideline would therefore be to use

[¶] Note that this is the opposite convention to that used in ecology for abundance matrices, where the rows (species) are the variables (the reason for the different ecological convention is clear: binomial species names are much more neatly displayed as row than column labels!). Input to the PRIMER PCA routine can use either convention. It is unnecessary to transpose the matrix before entry; one simply needs to be careful to specify whether rows or columns are to be treated as the variables.

correlation-based PCA for environmental variables and covariance-based PCA for species data (if no alternative, preferable ordination method is available for the latter). In *both* cases, prior transformation is likely to be beneficial; different transformations may be desirable for different variables in the environmental analysis, e.g. contaminant concentrations will often be right-skewed (and require something like 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

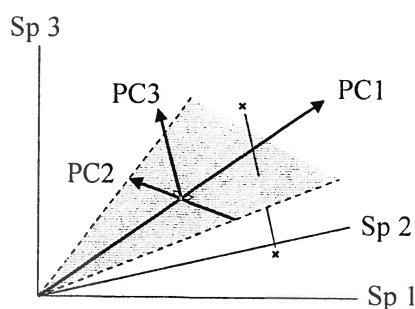
- 1) **PCA is conceptually simple.** Whilst the algebraic basis of the PCA algorithm requires a facility with matrix algebra for its understanding, the geometric concepts of a “best fitting” plane in the species space, and perpendicular projection of samples onto that plane, are relatively easily grasped. Some of the more recently proposed ordination methods, which either extend or supplant PCA (e.g. Principal Co-ordinates Analysis, Detrended Correspondence Analysis) can be harder to understand for practitioners without a mathematical background.
- 2) **It is computationally straightforward.** Again, this statement needs to be seen in relative terms. After the number of species has been reduced, usually rather drastically, the required matrix operations pose no 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 since the 1970’s is not surprising. Even the computationally simplest of the techniques, PCA, which dates back to the earlier period of classical statistics, could not be carried out manually in most realistic examples. Nonetheless, PCA tends to take only fractions of seconds of processing time on a modern PC. 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 very 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.



**EXAMPLE: Dosing experiment,
Solbergstrand mesocosm**

An example of this final point for a real data set can be seen in Fig. 4.2. This is of nematode data for the dosing experiment {D} in the Solbergstrand mesocosms, at the GEEP Oslo Workshop. Intact box core samples were collected from Oslofjord and held for three months under four dosing regimes: control, low, medium and high levels of a hydrocarbon and Cu contaminant mixture, continuously dosed to the basin waters. Four replicate box cores were subjected to each treatment and at the end of the period cores for all 16 boxes were examined for nematode communities (amongst other faunistic components). Fig. 4.2 shows the resulting PCA, based on log-transformed counts for 26 nematode genera (those making up more than 3% of the total individuals in at least one core). The interest here, of course, is in whether all replicates from one of the four treatments separate out from other treatments, indicating a change in community composition which could be attributed to a directly *causal* effect of the contaminant dosing. A cursory glance suggests that the high dose replicates (H) may indeed do this. However, closer study shows that the % of variance explained by the first two PC axes is very low: 22% for PC1 and 15% for PC2. The picture is likely to be *very unreliable* therefore, and an examination of the third and higher PCs confirms the distortion: some of the H replicates are much further apart in the full species space than this projection into two dimensions would imply. For example, the right-hand H sample is actually closer to the nearest M sample than it is to other H samples. The distances in the full species space are therefore *poorly-preserved* in the 2-dimensional ordination.

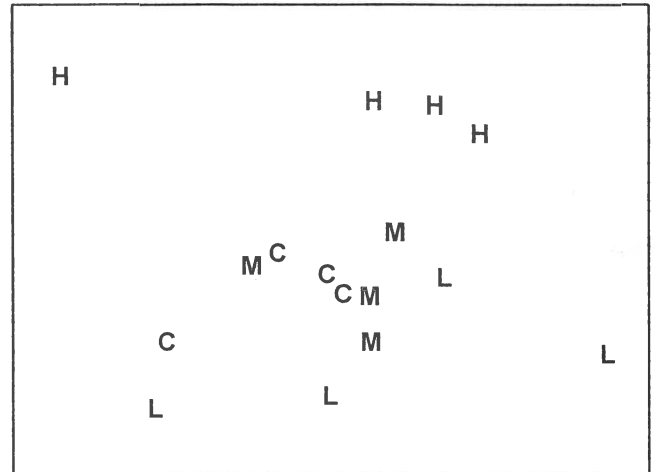


Fig. 4.2. Dosing experiment, Solbergstrand {D}. 2-dimensional PCA ordination of log-transformed nematode abundances from 16 box cores (4 replicates from each of control, low, medium and high doses of a hydrocarbon and Cu contaminant mixture). PC1 and PC2 account for 37% of the total variance.

This example is returned to again in Chapter 5, Fig. 5.5, where it is seen that an MDS of the same data makes a much better stab at “dissimilarity preservation”, though the data is such that no method will find it easy to represent in two dimensions. The moral here is clear:

- a) be very wary of interpreting any PCA plot which explains so little of the total variability in the original data;
- b) statements about apparent differences in a multivariate community analysis of one site (or time or treatment) from another should be backed-up by appropriate statistical tests; this is the subject of Chapter 6.

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 (e.g. 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).[¶]

The popular DECORANA version of CA has a primary motivation of straightening out an “arch effect” in a

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

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 re-aligning 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 DEOCRANA 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 Kruskal and Wish (1978), though again the applications described are not ecological. Throughout this manual

the term MDS refers to Kruskal's non-metric procedure (metric MDS is possible, and is akin to PCoA, but will not be discussed in any detail).

The starting point is the similarity or dissimilarity matrix among samples (Chapter 2). This can be whatever similarity matrix is biologically relevant to the questions being asked of the data. Through choice of coefficient and possible transformation or standardisation, one can choose whether to ignore joint absences, to emphasise similarity in common or rare species, to compare only % composition or allow sample totals to play a part, etc. In fact, the flexibility of MDS goes beyond this. It recognises the essential arbitrariness of *absolute* similarity values; Chapter 2 showed that the range of values taken could alter dramatically with transformation (often, the more severe the transformation, the higher and more compressed the similarity values become) and there is no absolute interpretation of a statement like "the similarity of sample 1 to sample 2 is 1.5 times that of sample 1 to sample 3". The natural interpretation is in terms of the *relative* values of similarity to each other, e.g. simply that "sample 1 is more similar to sample 2 than it is to sample 3". This is an intuitively appealing and very generally applicable base from which to build a graphical representation of the sample patterns and, in effect, the *ranks* of the similarities are the only information used by a successful non-metric MDS ordination.

The purpose of MDS can thus be simply stated: it constructs a "map" or *configuration* of the samples, in a specified number of dimensions, which attempts to satisfy all the conditions imposed by the rank (dis)similarity matrix, e.g. if sample 1 has higher similarity to sample 2 than it does to sample 3 then sample 1 will be placed closer on the map to sample 2 than it is to sample 3.

EXAMPLE: Loch Linnhe macrofauna

This is illustrated in Table 5.1 for the subset of the Loch Linnhe macrofauna data used to demonstrate hierarchical clustering (Table 3.2). Similarities between $\sqrt{\cdot}$ -transformed counts of the four year samples are given by Bray-Curtis similarity coefficients, and the corresponding rank similarities are also shown. (The highest similarity has the *lowest* rank, 1, and the lowest similarity the highest rank, $n(n-1)/2$.) The MDS configuration is constructed to *preserve* the similarity ranking as Euclidean distances in the 2-dimensional plot: samples 2 and 4 are closest,

2 and 3 next closest, then 1 and 4, 3 and 4, 1 and 2, and finally, 1 and 3 are furthest apart. The resulting figure is a more informative summary than the corresponding cluster analysis in Chapter 3, showing, as it does, a gradation of change from clean (1) to progressively more impacted years (2 and 3) then a reversal of the trend, though not complete recovery to the initial position (4).

Though the mechanism for constructing such MDS plots has not yet been described, two general features of MDS can already be noted.

- 1) MDS plots can be arbitrarily *scaled, located, rotated or inverted*. Clearly, rank order information about which samples are most or least similar can say nothing about which direction in the MDS plot is "up" or "down", or the absolute "distance apart" of two samples: what can be interpreted is *relative* distances apart, of course.
- 2) It is not difficult in the above example to see that four points could be placed in two dimensions in such a way as to satisfy the similarity ranking exactly.[¶] For more realistic data sets, though, this will not usually be possible and there will be some distortion or *stress* between the similarity rankings and the corresponding distance rankings in the ordination plot (even in a high-dimensional configuration). This motivates the principle of the MDS algorithm: to choose a configuration of points which minimises this degree of stress, appropriately measured.

EXAMPLE: Exe estuary nematodes

The construction of an MDS plot is illustrated with data collected by Warwick (1971) and subsequently analysed in this way by Field *et al* (1982). A total of 19 stations from different sites and tide-levels in the Exe estuary, UK, were sampled bi-monthly at low spring tides between October 1966 and September 1967.

[¶] In fact, there are rather too many ways of satisfying it and the algorithm described in this chapter will find slightly different solutions each time it is run, all of them equally correct. However, this is not a problem in genuine applications with (say) six or more points. The number of similarities increases roughly with the square of the number of samples and a position is reached very quickly in which not all rank orders can be preserved and this particular indeterminacy disappears.

Table 5.1. Loch Linnhe macrofauna {L} subset. Abundance array after $\sqrt{\cdot}$ -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		Sample	1	2	3	4		Sample	1	2	3	4		
Species					→	1	–				→	1	–										3	
<i>Echinoca.</i>	1.7	0	0	0		2	25.6	–				2	5	–										2
<i>Myrioche.</i>	2.1	0	0	1.3		3	0.0	67.9	–			3	6	2	–									1
<i>Labidopl.</i>	1.7	2.5	0	1.8		4	52.2	68.1	42.0	–		4	3	1	4	–								4
<i>Amaeana</i>	0	1.9	3.5	1.7																				
<i>Capitella</i>	0	3.4	4.3	1.2																				
<i>Mytilus</i>	0	0	0	0																				

Three replicate sediment cores were taken for meio-faunal 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 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 $\sqrt{\cdot}$ -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 relationship to geographical position or tidal level of the samples.

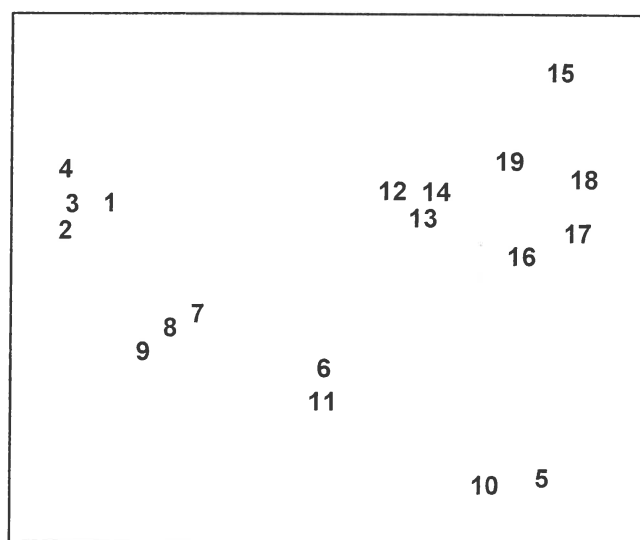


Fig. 5.1. Exe estuary nematodes {X}. MDS ordination of the 19 sites based on $\sqrt{\cdot}$ -transformed abundances and Bray-Curtis similarities (stress = 0.05).

Instead they appear to relate to variables such as sediment type and organic content, and these links are discussed 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.[†] 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 j th and k th 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.

[†] This is also the algorithm used in the PRIMER MDS routine. The required input is a similarity matrix, either as calculated in PRIMER or read in directly from Excel, for example.

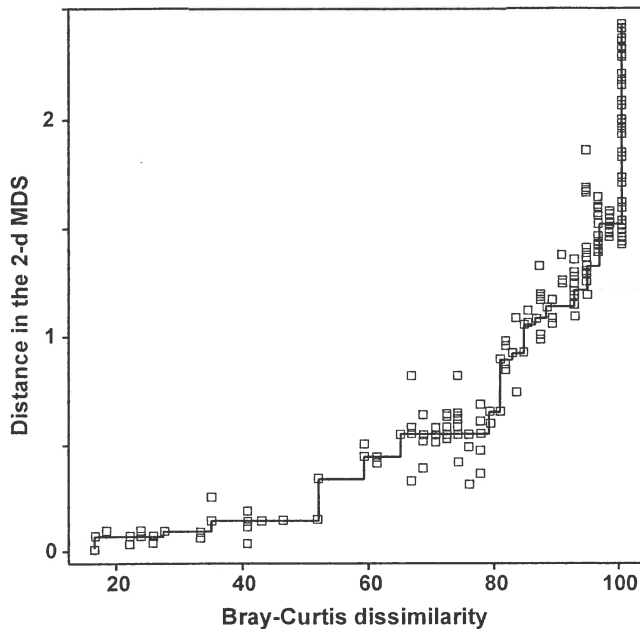


Fig. 5.2. *Exe estuary nematodes {X}*. Shepard diagram of the distances (d) in the MDS plot (Fig. 5.1) against the dissimilarities (δ) in the Bray-Curtis matrix. The line is the fitted non-parametric regression; stress (=0.05) is a measure of scatter about the line.

(In fact, Fig. 5.2 is at a late stage of the iteration, corresponding to the final 2-dimensional configuration of Fig. 5.1; at earlier stages the graph will appear similar though with a greater degree of scatter). The decision that characterises different ordination procedures must now be made: how does one define the underlying relation between distance in the plot and the original dissimilarity? There are two main approaches.

a) Fit a standard **linear regression** of d on δ , so that final distance is constrained to be *proportional* to original dissimilarity. More flexible would be to fit some form of **curvilinear regression** (maybe a quadratic, cubic or other polynomial). These are **parametric** models giving rise to the term **metric MDS** for this approach.

b) Perform a **non-parametric regression** of d on δ giving rise to a **non-metric MDS**. Fig. 5.2 illustrates the non-parametric (monotonic) regression line. This is a “best-fitting” line which moulds itself to the shape of the scatter plot, but is always constrained to increase (and therefore consists of a series of steps). The relative success of non-metric MDS, in preserving the sample relationships in the distances of the ordination plot, comes from the flexibility in shape of this non-parametric regression line. A perfect MDS was defined previously as one in which the rank order of dissimilarities was totally preserved in the rank order of distances; individual

points on the Shepard plot must then all be (monotonic) increasing; the larger a dissimilarity, the larger (or equal) the corresponding distance, and the non-parametric regression line is a perfect fit. It follows that the extent to which the scatter points deviate from the line measures the failure to match the rank order dissimilarities, motivating the following definition of **stress**.

4) **Measure goodness-of-fit of the regression** by calculating the **stress** value

$$Stress = \sqrt{\sum_j \sum_k (d_{jk} - \hat{d}_{jk})^2 / \sum_j \sum_k d_{jk}^2} \quad (5.1)$$

where \hat{d}_{jk} is the distance predicted from the fitted regression line corresponding to dissimilarity δ_{jk} . If $d_{jk} = \hat{d}_{jk}$ for all the $n(n-1)/2$ distances in this summation, the stress is zero. Large scatter clearly leads to large stress and this can be thought of as measuring the difficulty involved in compressing the sample relationships into two (or a small number) of dimensions. Note that the denominator is simply a scaling term: distances in the final plot have only relative not absolute meaning and the squared distance term in the denominator makes sure that stress is a dimensionless quality.

5) **Perturb the current configuration in a direction of decreasing stress**. This is perhaps the most difficult part of the algorithm to visualise and will not be detailed; it is based on established techniques of numerical optimisation, in particular the method of **steepest descent**. The essential idea is that the regression relation is used to evaluate stress for (small) changes in the position of points on the ordination plot, and points are then moved to new positions in directions which look like they will decrease the stress most rapidly.

6) **Repeat steps 3 to 5 until convergence is achieved**. The iteration now cycles around the two stages of a new regression of distance on dissimilarity for the new ordination positions, then further perturbation of the positions in directions of decreasing stress. In most cases, the cycle will stop when further adjustment of the points leads to no improvement in stress.

Features of the algorithm

Like all iterative procedures, especially ones this complex, things can go wrong! By a series of minor adjustments to the parameters at its disposal (the co-

ordinate positions in the configuration), the method gradually finds its way down to a minimum of the stress function. This is most easily envisaged in 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, i.e. the *global minimum* of the stress function. The rambler may reach a low point from which the ground rises in all directions (and 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.[¶] 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 has been told to expect 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, or for which *all* dissimilarities within the groups are smaller than *any* dissimilarity between groups, then there is clearly no yardstick within our non-parametric approach for determining how far apart the groups should be placed in the MDS plot. It is then not surprising to find that the samples in each group collapse to a point (a commonly met special case is when one of the two groups consists of a single outlying point). The solution is to split the data and carry out an ordination separately on the two groups (or, in the latter case, re-run the MDS omitting the outlier).

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

[¶] *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 also then permits easy user-control of final orientation and reflection.*

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 for the 2-dimensional configuration.[¶] In fact, this strategy will tend to reduce the risk of finding local minima or degenerate solutions. The samples are likely to fit more easily into three dimensions, itself reducing the risk of finding a local minimum; the 2-dimensional iteration will then be constrained to start much nearer a global minimum than it would for a purely random initial configuration. Another reason for obtaining higher-dimensional solutions is to compare their stress with that from two dimensions: this is one of several ways in which the accuracy of a 2-dimensional MDS can be assessed.

ADEQUACY OF MDS REPRESENTATION

1) *Is the stress value small?* By definition, stress increases with reducing dimensionality of the ordination (or in rare cases where a low-dimensional ordination is a perfect representation, stress remains constant). It has therefore been suggested that stress values in 2, 3, 4 etc. dimensions should be compared: if there is a particularly large drop in stress passing from two to three dimensions (say) and only a modest, steady decrease thereafter, this would imply that a 3-dimensional ordination is likely to be a more satisfactory representation than a 2-dimensional one. However, experience with ecological data suggests that clear-cut “shoulders” such as this, in the plot of minimum stress against dimensionality, are rarely seen. It is also undeniable that a 2-dimensional picture will usually be a more useful and accessible summary, so the question is often turned around: not “What is the true dimensionality of the data?” but “Is a 2-dimensional plot a *usable* summary of the sample relationships, or is it likely to be sufficiently misleading to force its abandonment in favour of a 3- or higher-dimensional

[¶] This procedure is adopted by the PRIMER MDS routine, which also allows the user to specify the number of random re-starts (ideally at least 10). The PRIMER results log contains the coordinates of the best (lowest stress) 2-dimensional and 3-dimensional solutions and the stress values for all 2-d and 3-d repeats. It can plot either the optimal 2-d or 3-d configuration.

plot?” One answer to this is through empirical evidence and simulation studies of stress values. Stress increases not only with reducing dimensionality but also with increasing quantity of data, but a rough rule-of-thumb for 2-dimensional ordinations, using the stress formula (5.1), is as follows.[†]

Stress <0.05 gives an excellent representation with no prospect of misinterpretation (a perfect representation would probably be one with stress <0.01 since numerical iteration procedures often terminate when stress reduces below this value[‡]).

Stress <0.1 corresponds to a good ordination with no real prospect of a misleading interpretation; 3- or higher-dimensional solutions will not add any additional information about the *overall* structure (though the fine structure of any compact groups may bear closer examination).

Stress <0.2 still gives a potentially useful 2-dimensional picture, though for values at the upper end of this range too much reliance should not be placed on the detail of the plot; a cross-check of any conclusions should be made against those from an alternative technique (e.g. the superimposition of cluster groups suggested in point 5 below).

Stress >0.3 indicates that the points are close to being arbitrarily placed in the 2-dimensional ordination space. In fact, the totally random positions used as a starting configuration for the iteration usually give a stress around 0.35–0.45. Values of stress in the range 0.2–0.3 should therefore be treated with a great deal of scepticism and certainly discarded in the upper half of this range, especially for a small to moderate number of points (<50 say). Other techniques will be certain to highlight inconsistencies and higher-dimensional ordinations should be examined.

2) *Does the Shepard diagram appear satisfactory?*

The stress value totals the scatter around the regression line in a Shepard diagram, for example the low stress of 0.05 for Fig. 5.1 is reflected in the low scatter in Fig. 5.2. Outlying points in the plot could

[†] There are alternative definitions of stress, for example the “stress formula 2” option provided in the MDSCAL and KYST programs. This differs only in the denominator scaling term in (5.1) but is believed to increase the risk of finding local minima and to be more appropriate for other forms of multivariate scaling, e.g. multidimensional unfolding, which are outside the scope of this manual.

[‡] This is true of the MDS routine in PRIMER, for example.

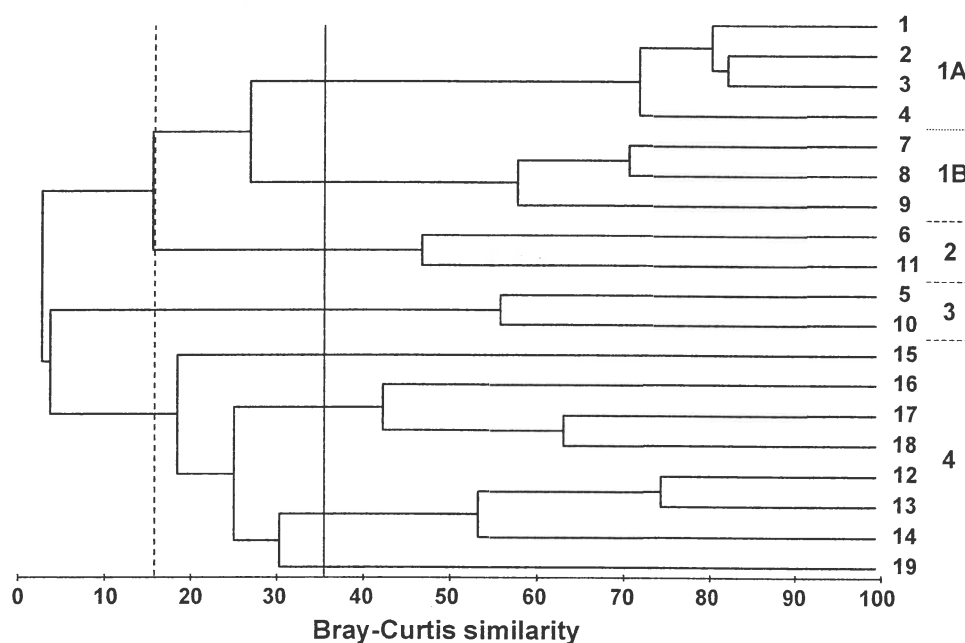


Fig. 5.3. Exe estuary nematodes {X}. Dendrogram of the 19 stations, using group-average clustering from Bray-Curtis similarities on $\sqrt{1}$ -transformed abundances. The four groups of stations separated at a 15% similarity threshold (dotted line) are indicated (the two tightly clustered sub-groups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

be identified with the samples involved; often there are a range of *outliers* all involving dissimilarities with a particular sample and this can indicate a point which really needs a higher-dimensional representation for 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%.

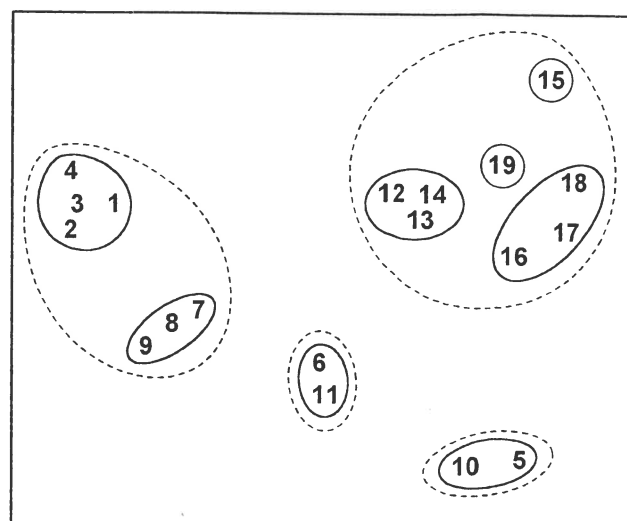


Fig. 5.4. Exe estuary nematodes {X}. 2-dimensional MDS configuration, as in Fig. 5.1, with superimposed clusters from Fig. 5.3, at similarity levels of 15% (dashed line) and 30%-45% (continuous line).

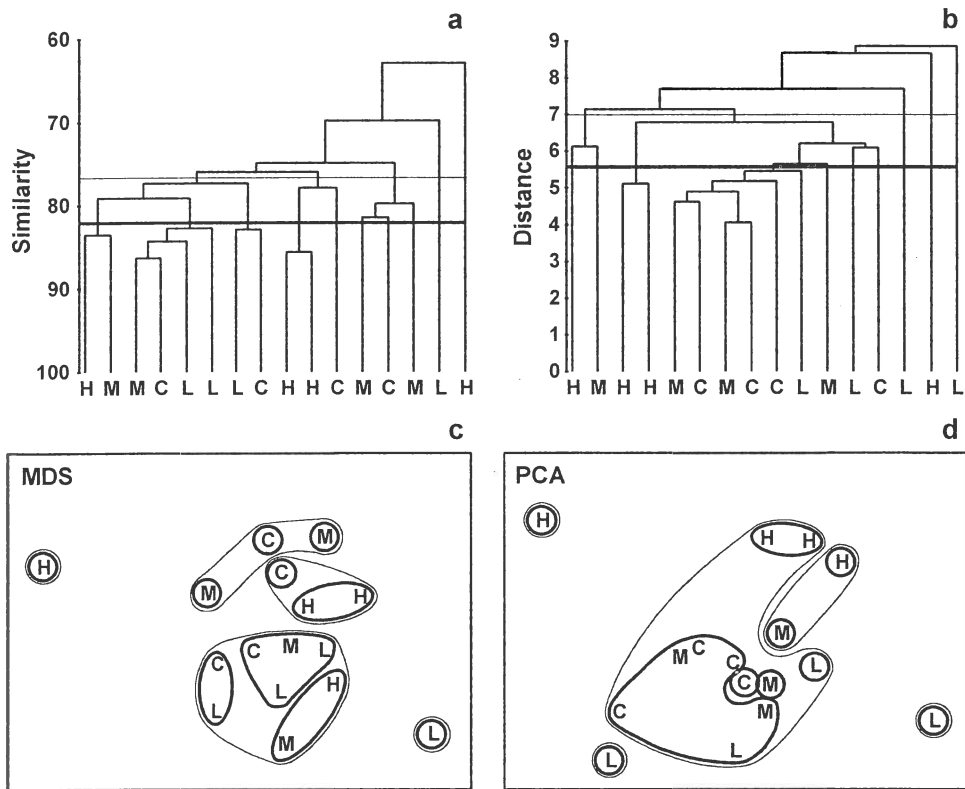


Fig. 5.5. Dosing experiment, Solbergstrand mesocosm {D}. Nematode abundances for four replicates from each of four treatments (control, low, medium and high dose of hydrocarbons and Cu) after species reduction and log transformation as in Fig. 4.2. a), c) Group-averaged clustering from Bray-Curtis similarities; clusters formed at two arbitrary levels are superimposed on the 2-dimensional MDS obtained from the same similarities (stress = 0.16). b), d) Group-average clustering from Euclidean distances; clusters from two levels are superimposed on the 2-dimensional PCA of Fig. 4.2. Note the greater degree of distortion in the latter.

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 an accurate representation of the sample relationships. One is not always as fortunate as this, and a more revealing example of the benefits of viewing clustering and ordination in combination is provided by the data of Fig. 4.2.[¶]

EXAMPLE: Dosing experiment, Solbergstrand

The nematode abundance data from the dosing experiment {D} at the GEEP Oslo Workshop was previously analysed by PCA, see Fig. 4.2 and accompanying text. The analysis was likely to be unsatisfactory, since the % of variance explained by the first two

principal components was very low, at 37%. Fig. 5.5c shows the MDS ordination from the same data, and in order to make a fair comparison with the PCA the data matrix was treated in exactly the same way prior to analysis.[†] The stress for the 2-dimensional MDS configuration is moderately high (at 0.16), indicating some difficulty in displaying the relationships between these 16 samples in two dimensions. However, the PCA was positively misleading in its apparent separation of the four high dose (H) replicates in the 2-dimensional space; by contrast the MDS does provide a usable summary which is not likely to lead to serious misinterpretation. This can be seen by superimposing the corresponding cluster analysis results, Fig. 5.5a, onto the MDS. Two similarity thresholds have been chosen in Fig. 5.5a such that they (arbitrarily) divide the samples into 5 and 10 groups, the corresponding hierarchy of clusters being indicated in Fig. 5.5c by thin and thick lines respectively. Whilst it is clear that there are no natural groupings of the samples in the MDS plot, and the groupings provided by the cluster analysis must therefore be regarded with some caution, the two analyses are not markedly inconsistent.

[¶] One option within PRIMER is to run CLUSTER on the ranks of the similarities rather than the similarities themselves. Whilst not of any real merit in itself (and not the default option), Clarke (1993) argues that this could have marginal benefit when performing a group-average cluster analysis solely to see how well the clusters agree with the MDS plot: the argument is that the information utilised by both techniques is then made even more comparable.

[†] The same 26 species were retained and a log transformation applied before computation of Bray-Curtis similarities, though, of course, a species reduction would not normally be necessary with MDS or clustering of samples.

In contrast, the parallel operation for the PCA ordination clearly illustrates the poorer distance-preserving properties of this method. Fig. 5.5d repeats the 2-dimensional PCA of Fig. 4.2 but with superimposed groups from a cluster analysis of the Euclidean distance matrix[‡] 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 that would be expected 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 20-25 m; 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 community composition between day and night near-surface samples, contrasted with their

[‡] As previously noted, Euclidean distance is the dissimilarity measure implicit in a PCA ordination.

[§] The precise relationships between the day and night samples for the larger depths (F-N) would now best be examined by an MDS of that data alone, the greater precision resulting from the MDS then not needing to cater, in the same 2-d picture, for the relationships to (and between) the A-E samples. This re-analysis of subsets should be a commonly-used strategy in the constant battle to display high-dimensional information in low dimensions.

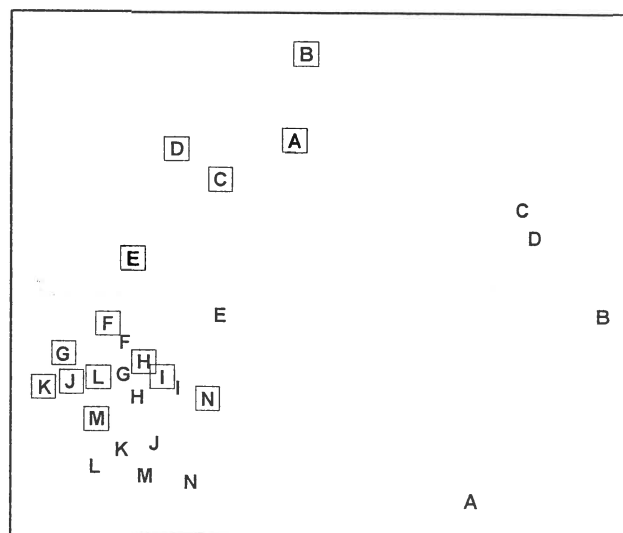


Fig. 5.6. Celtic Sea zooplankton {C}. MDS plot for night (boxed) and day time samples from 14 depths (5 to 70m, denoted A, B, ..., N), taken at a single site during September 1978.

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 such a gradually changing community pattern.

MDS STRENGTHS

- 1) **MDS is simple in concept.** The numerical algorithm is undeniably complex, but 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 is likely to be more difficult to represent in a low number of dimensions. More usually, the similarity

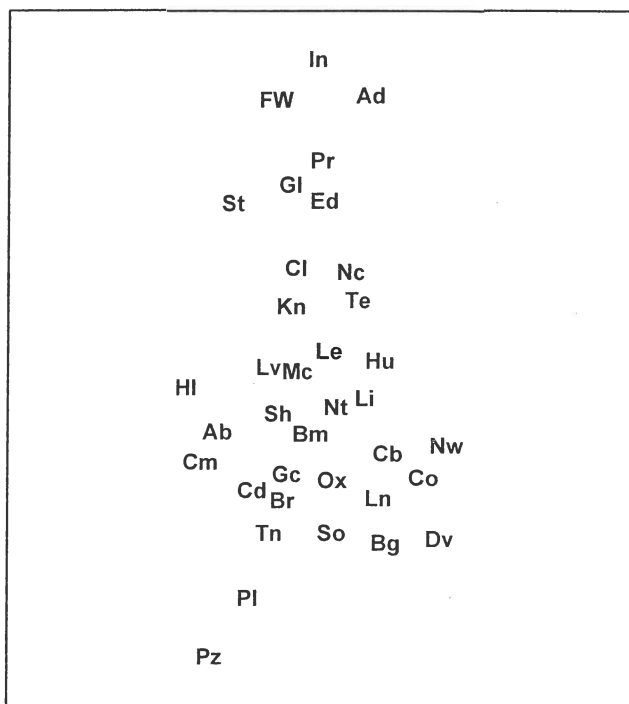


Fig. 5.7. Non-metric MDS configuration of the road distances (partly given in Table 5.2) between selected UK towns and cities (stress = 0.04).

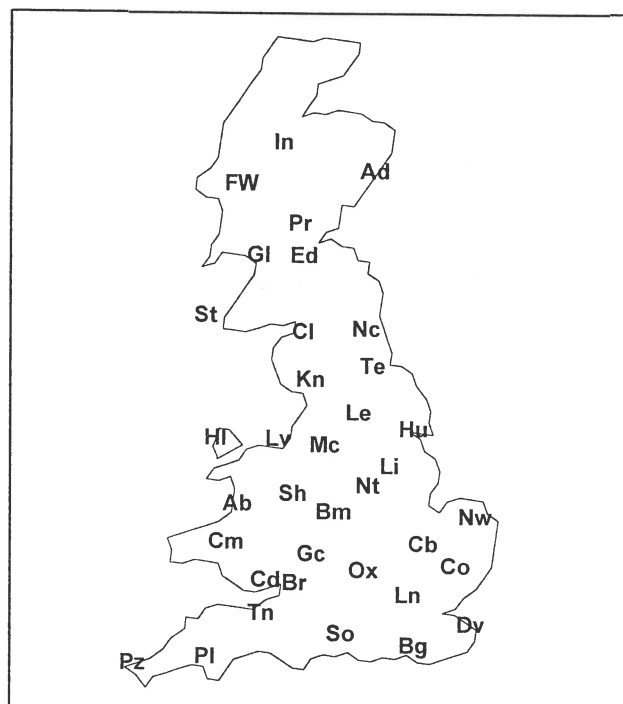


Fig. 5.8. Non-metric MDS configuration of the same towns and cities as in Table 5.2, but starting from the matrix of direct ("as the crow flies") distances between every pair (stress = 0).

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 much practical significance here, that the algorithm can operate perfectly successfully when the similarity matrix is subject to a certain amount of missing data.[†]

MDS WEAKNESSES

1) **MDS is computationally demanding.** To generate a configuration with a moderate number of samples takes some time on a modern PC (for up to $n = 100$ samples, a few seconds for each of 10 random starts,

[†] Neither of these options are currently implemented in PRIMER. They 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!

say), though speed has become much less of an issue than it once was. MDS on more than about $n = 1000$ 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 and in the discussion of Fig. 5.6, where it was suggested that large data sets can often, with benefit, 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 another MDS to display the large-scale structure across groups.

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

[‡] There are no longer absolute constraints on matrix sizes within PRIMER though there will be effective constraints imposed by available memory and processor speed. However, the pertinent question should now be not "how many samples will it handle?" but "how many samples does it make sense to try to represent, approximately, in a single 2-d picture?"

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, the higher the stress, the greater the likelihood of non-optimal solutions, so a larger number of repeats is required; this adds to the computational burden.

- 3) *The algorithm places most weight on the large distances.* A common feature of most ordination methods (including MDS and PCA) is that more attention is given to correct representation of the *overall* structure of the samples than their *local* structure. For MDS, it is clear from the form of equation (5.1) that the largest contributions to stress will come from incorrect placement of samples which are very distant from each other. Where distances are small, the sum of squared difference terms will also be relatively small and the minimisation process will not be as sensitive to incorrect positioning. This is another reason therefore for repeating the ordination *within* each large cluster: it will lead to a more accurate display of the fine structure, if this is important to interpretation. An example is given later in Figs. 6.2a and 6.3, and is fairly typical of the generally minor differences that result: the subset of points are given more freedom to expand in a particular direction but their relative positions are usually only marginally changed.

RECOMMENDATIONS

- 1) MDS can be recommended as one of the best (*the* best, arguably) ordination techniques available (e.g. Everitt, 1978). The few comprehensive studies that have compared ordination methods for community data give non-metric MDS a high rating (e.g. Kenkel and Orloci, 1986). In comparison with earlier techniques, such as PCA, MDS has a number of practical advantages stemming from its flexibility and (lack of) assumptions.
- 2) When the inter-sample relationships are relatively simple, e.g. there are a few strong clusters or one strong gradient of change across all samples, most ordination methods will perform adequately and give comparable pictures. The main advantage of MDS is its greater ability to represent more complex relations accurately in low-dimensional space.
- 3) If the stress is low (say <0.1), an MDS ordination is probably a more useful representation than a cluster analysis: when the samples are strongly

grouped the MDS will reveal this anyway, and when there is a more gradual continuum of change, or some interest in the placement of major groups with respect to each other, MDS will display this in a way that a cluster analysis is quite incapable of doing. For higher values of stress, the techniques should be thought of as *complementary* to each other; neither may present the full picture so the recommendation is to *perform both* and *view them in combination*. This may make it clear which points on the MDS are problematic to position (examining some of the local minimum solutions can help here[¶]) and an ordination in a higher dimension may prove more consistent with the cluster groupings. Conversely, the MDS plots may make it clear that some groups in the cluster analysis are arbitrary subdivisions of a natural continuum.

[¶] For example, run the PRIMER MDS routine several times, with a single random starting position on each occasion, and examine the plots that give a higher stress than originally detected.

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 $p < 5\%$ 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 also clear

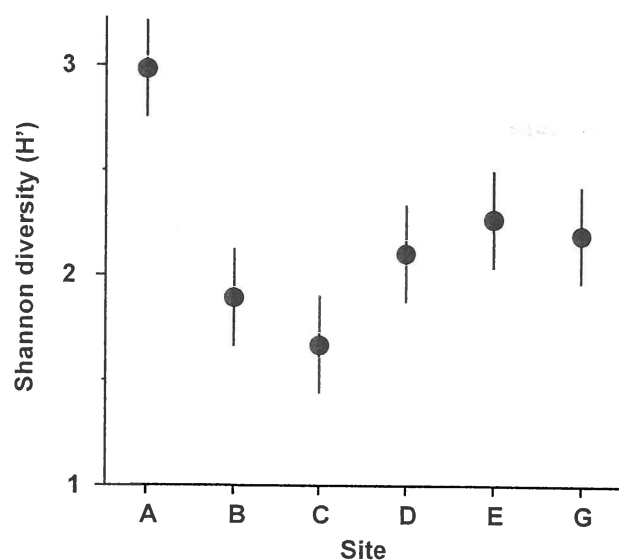


Fig. 6.1. Frierfjord macrofauna {F}. Means and 95% confidence intervals of Shannon diversity (H') at the 6 field sites (A-E, G) shown in Fig. 1.1.

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

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

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

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

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

MULTIVARIATE TESTS

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

This simple structure of groups, and replicates within groups, is referred to as a *1-way layout*, and it was seen above that 1-way ANOVA would provide the appropriate testing framework *if* the data were univariate (e.g. diversity or total abundance across all species). There *is* an analogous multivariate analysis of variance (MANOVA, e.g. Mardia *et al*, 1979), in which the F test is replaced by a test known as Wilks' Λ , but its assumptions will never be satisfied for typical multi-species abundance (or biomass) data. This is the problem referred to in the earlier chapters on choosing similarities and ordination methods; there are typically many more species (variables) than

samples and the probability distribution of counts could never be reduced to approximate (multivariate) normality, by any transformation, because of the dominance of zero values. For example, for the Frierfjord data, as many as 50% of the entries in the species/samples matrix are zero, even after reducing the matrix to only the 30 most abundant species!

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

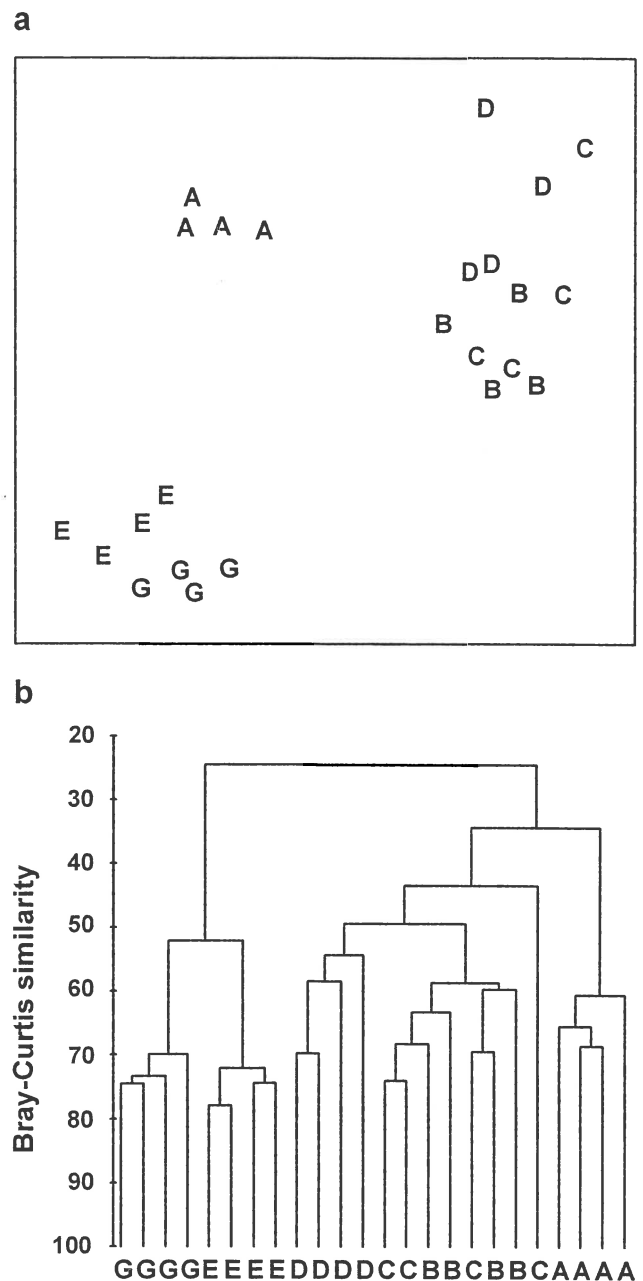


Fig. 6.2 Frierfjord macrofauna {F}. a) MDS plot, b) dendrogram, for 4 replicates from each of the 6 sites (A-E and G), from Bray Curtis similarities computed for $\sqrt[4]{}$ -transformed species abundances (MDS stress = 0.05).

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

'ANOSIM' FOR THE 1-WAY LAYOUT

Fig.6.3 displays the MDS based only on the 12 samples (4 replicates per site) from the B, C and D sites of the Frierfjord macrofauna data. The null hypothesis (H_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 it would have a number of drawbacks.

- a) Such a statistic could only apply to a situation in which the method of display was an MDS rather than, say, a cluster analysis.
- b) The result would depend on whether the MDS was constructed in two, three or higher dimensions. There is often no "correct" dimensionality and one may end up viewing the picture in several different dimensions – it would be unsatisfactory to generate different test statistics in this way.
- c) The configuration of B, C and D replicates in Fig. 6.3 also differs slightly from that in Fig. 6.2a, which includes the full set of sites A-E, G. It is again undesirable that a test statistic for comparing *only* B, C and D should depend on which other sites are included in the picture.

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

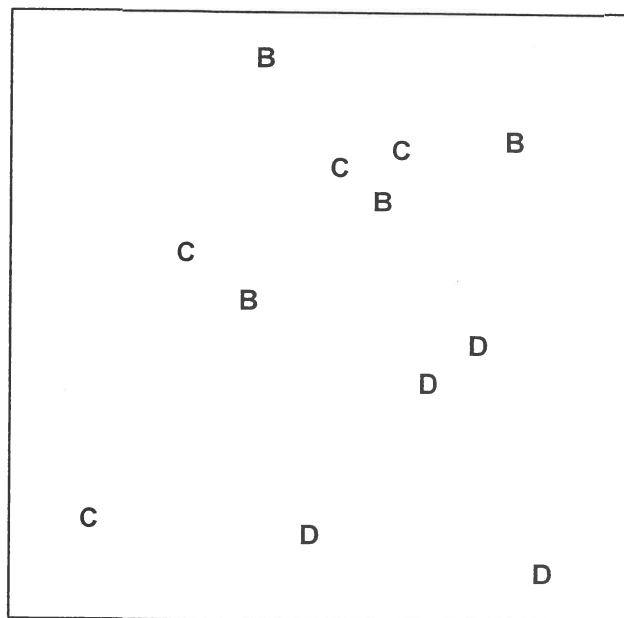


Fig. 6.3. Frierfjord macrofauna {F}. MDS ordination as for Fig. 6.2 but computed only from the similarities involving sites B, C and D (stress = 0.11).

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

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

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

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

- a) R can never technically lie outside the range (-1,1);
- b) $R = 1$ only if *all* replicates within sites are more similar to each other than *any* replicates from different sites;
- c) R is approximately zero if the null hypothesis is true, so that similarities between and within sites will be the same on average.

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

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

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

2) **Recompute the statistic under permutations** of the sample labels. Under the null hypothesis H_0 : “no difference between sites”, there will be little effect on average to the value of R if the labels identifying which replicates belong to which sites are arbitrarily rearranged; the 12 samples of Fig. 6.3 are just replicates from a single site if H_0 is true. This is the rationale for a **permutation test** of H_0 ; all possible allocations of four B, four C and four D labels to the 12 samples are examined and the R statistic recalculated for each. In general there are

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

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

3) **Calculate the significance level** by referring the observed value of R to its permutation distribution. If H_0 is true, the likely spread of values of R is given by the random rearrangements, so that if the *true* value of R looks unlikely to have come from this distribution there is evidence to reject the null hypothesis. Formally, if only t of the T simulated values of R are as large (or larger than) the *observed* R then H_0 can be rejected at a **significance level** of $(t+1)/(T+1)$, or in percentage terms, $100(t+1)/(T+1)\%$.

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

EXAMPLE: Frierfjord macrofauna

The rank similarities underlying Fig. 6.3 are shown in Table 6.2 (note that these are the similarities involving only sites B, C and D, extracted from the matrix for *all* sites and re-ranked). Averaging across the 3 diagonal sub-matrices (within groups B, C and D) gives $\bar{r}_W = 22.7$, and across the remaining (off-diagonal) entries gives $\bar{r}_B = 37.5$. Also $n = 12$ and $M = 66$, so that $R = 0.45$. In contrast, the spread of R values possible from random re-labelling of the 12 samples can be seen in the histogram of Fig. 6.4: the largest of $T = 999$ simulations is less than 0.45 ($t=0$). An observed value of $R = 0.45$ is seen to be a most unlikely event, with a probability of less than 1 in a 1000 if H_0 is true, and we can therefore reject H_0 at a significance level of $p < 0.1\%$ (at least, because $R = 0.45$ may still have been the most extreme outcome observed had we chosen an even larger number of simulations).

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

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

Pairwise tests

The above is a **global** test, indicating that there are site differences *somewhere* that may be worth examining further. Specific pairs of sites can then be compared: for example, the similarities involving only sites B and C are extracted, re-ranked and the test procedure repeated, giving an R value of 0.23. This time there are only 35 distinct relabellings so, under the null hypothesis H_0 that sites B and C do not differ, the full permutation distribution of possible values of R can be computed; 12% of these values are equal to or larger than 0.23 so H_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.

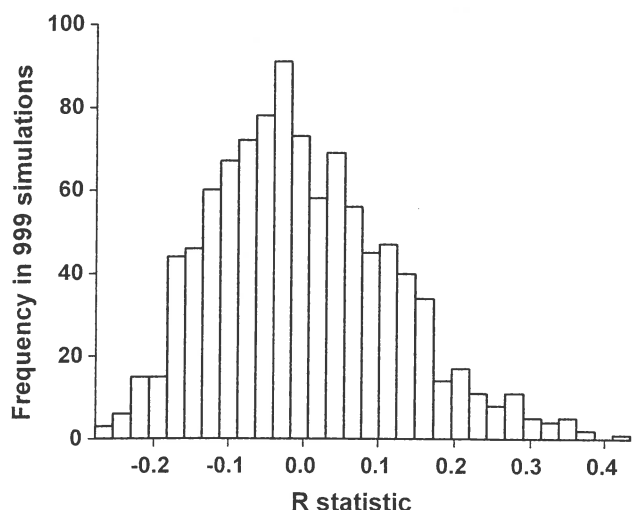


Fig. 6.4. Frierfjord macrofauna {F}. Simulated distribution of the test statistic R (equation 6.1) under the null hypothesis of 'no site differences'; this contrasts with an observed value for R of 0.45.

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

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

However, the difficulty with such a Bonferroni correction is clear from the above example: with only 4 replicates in each group, and thus only 35 possible permutations, a significance level of $0.05/35$ ($\approx 1.7\%$) can never be achieved! It may be possible to plan for a modest improvement in the number of replicates: 5 replicates from each site would allow a 1% level test

for a pairwise comparison, equation (6.2) showing that there are then 126 permutations, and two groups of 6 replicates would give close to a 0.2% level test. However, this may not be realistic in some practical contexts, or it may be inefficient to concentrate effort on too many replicates at one site, rather than (say) increase the spatial coverage of sites. Also, for a fixed number of replicates, a too demanding low Type I error (significance level) will be at the expense of a greater risk of *Type II error*, the probability of *not* detecting a difference when one genuinely exists.

Strategy for interpretation

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

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

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

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

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

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

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

Generality of application

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

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

EXAMPLE: Indonesian reef-corals

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

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

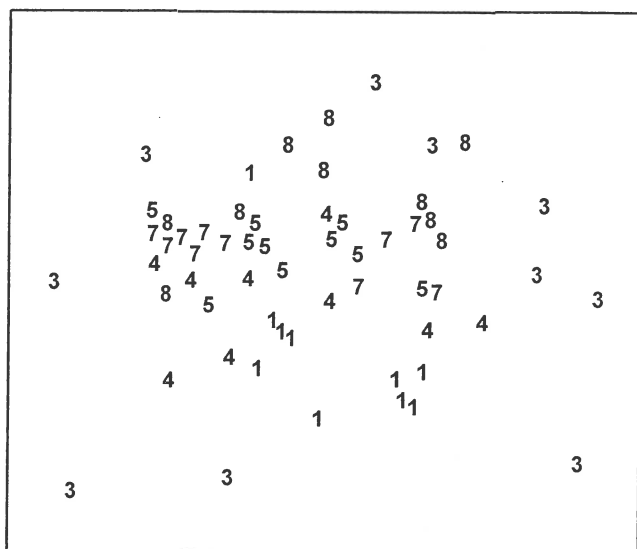


Fig. 6.5. Indonesian reef corals, *S. Tikus Island* {I}. MDS of % species cover from 10 replicate transects in each of 6 years: 1 = 1981, 3 = 1983 etc (stress = 0.19).

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

ANOSIM FOR TWO-WAY LAYOUTS

Three types of field and laboratory designs are considered here:

- the **2-way nested** case can arise where two levels of spatial replication are involved, e.g. sites are grouped *a priori* to be representative of two “treatment” categories (control and polluted) but there are also replicate samples taken within sites;
- the **2-way crossed** case can arise from studying a fixed set of sites at several times (with replicates at each site/time combination), or from an experimental study in which the same set of “treatments”

(e.g. control and impact) are applied at a number of locations (“blocks”), for example in the different mesocosm basins of a laboratory experiment;

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

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

EXAMPLE: Clyde nematodes (2-way nested case)

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

- H1: there are no differences among sites *within* each “treatment” (control or polluted conditions);
- H2: there are no differences *between* control and polluted conditions.

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

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

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

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

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

H2, which will usually be the more interesting of the two hypotheses, can now be examined. The test of H1 demonstrated that there are, in effect, only three

genuine “replicates” (the sites 1-3) at each of the two conditions (C and P).

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:

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

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

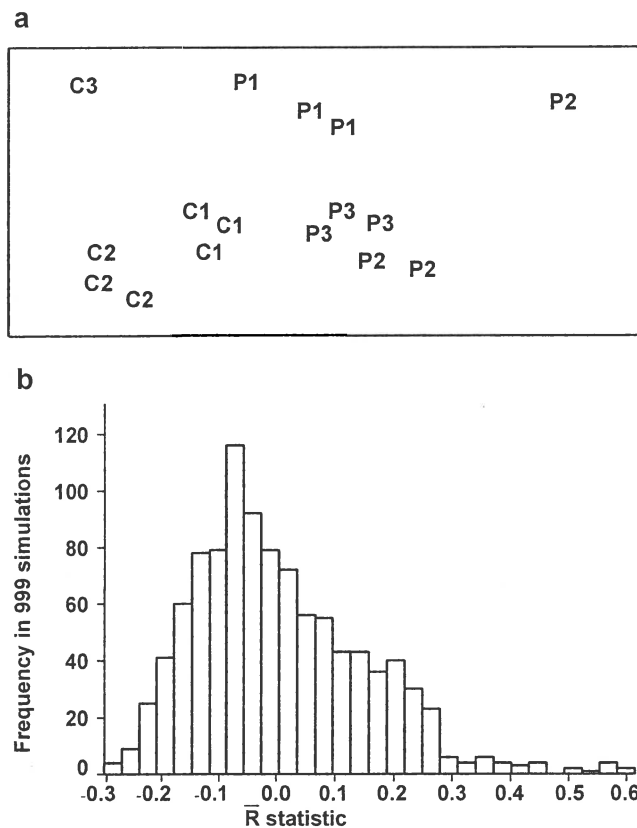


Fig. 6.6. *Clyde nematodes* {Y}. a) MDS of species abundances from three “polluted” (P1-P3) and three “control” sites (C1–C3), with three replicate samples at most sites (stress = 0.09). b) Simulated distribution of the test statistic \bar{R} under the hypothesis H1 of “no site differences” within each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

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

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

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

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

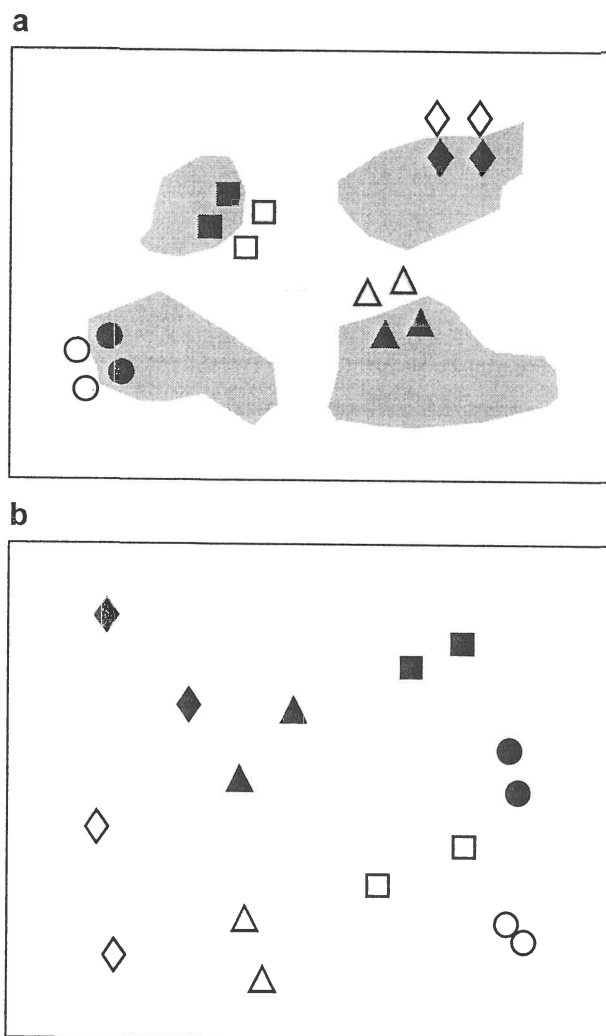


Fig. 6.7. Tasmania, Eaglehawk Neck {T}. a) Schematic of the ‘2-way crossed’ sampling design for 16 meiofaunal cores with two disturbed and two undisturbed replicates from each of four patches of burrowing activity by soldier crabs (shaded). b) MDS of species abundances for the 16 samples, showing separation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

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

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

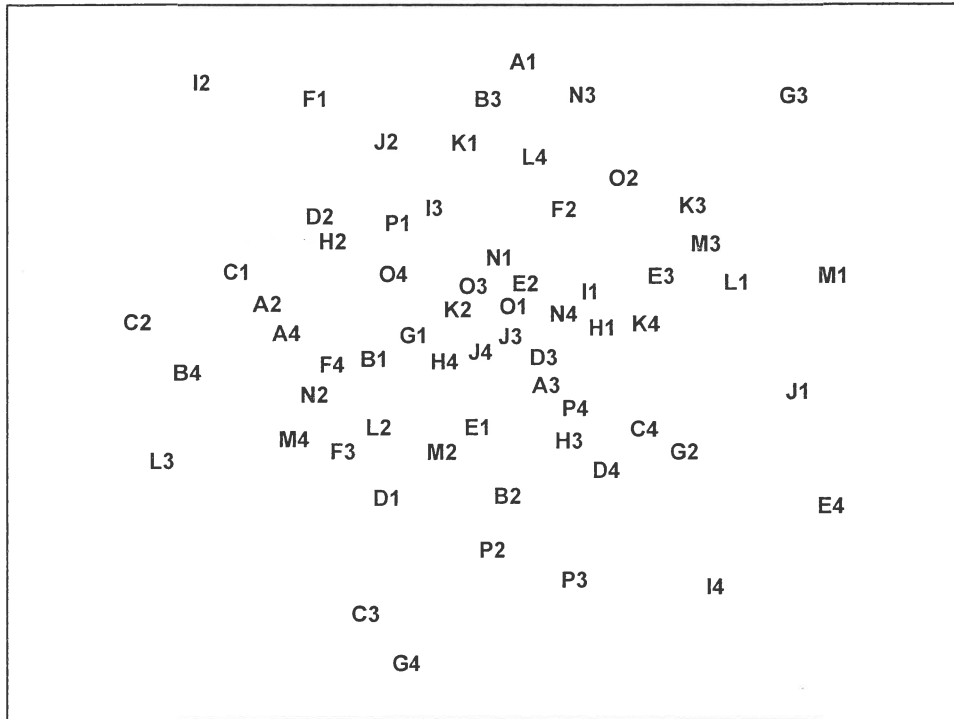


Fig. 6.8. Westerschelde nematodes experiment {W}. MDS of species abundances from 16 different nutrient-enrichment treatments, A to P, applied to sediment cores in each of four mesocosm basins, 1 to 4 (stress = 0.28).

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

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

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

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

The experiment involved 15 different nutrient enrichment conditions and one control, the treatments being applied to the surface of the undisturbed sediment cores. After 16 weeks controlled exposure in the mesocosm environment, the meiofaunal communities in the 64 cores were identified, and Bray-Curtis similarities on root-transformed abundances gave the MDS of Fig. 6.8. The full set of 16 treatments was repeated in each of the 4 basins (blocks), so the structure is a 2-way treatments \times 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_0 : there are no treatment differences (but allowing the possibility of basin effects)

is clearly necessary before any sort of interpretation is attempted.

In the absence of replication, a test is still possible in the *univariate* case, under the assumption that interaction effects are small in relation to the main treatment or block differences (Scheffé, 1959). In a similar spirit, a *global* test of H_0 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_0 is false. The analogy

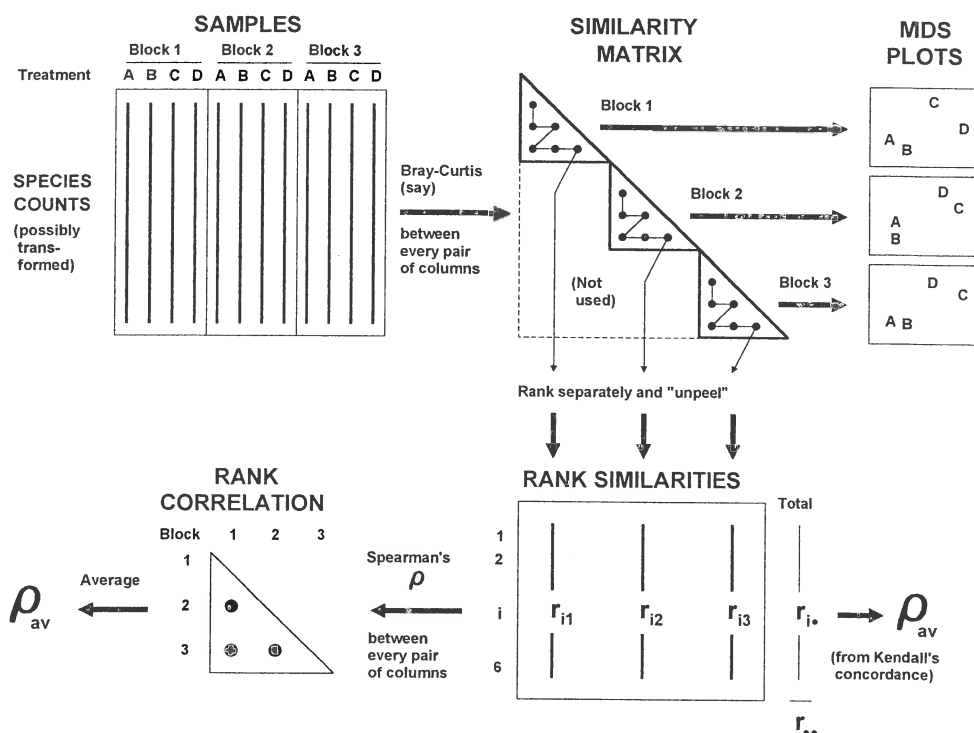


Fig. 6.9. Schematic diagram illustrating the stages in defining concordance of treatment patterns across the blocks, and the two computational routes for ρ_{av} .

with the univariate test is clear: large interaction effects imply that the treatment pattern differs from block to block and there is little chance of identifying a treatment effect; on the other hand, for a treatment \times 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_0 . Note that, as required by the statement of H_0 , the test makes no assumption about the absence of block effects; between-block similarities are irrelevant to a statistic based only on agreement in within-block patterns.

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

for agreement of two blocks, j and k , is the Spearman correlation coefficient

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

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

$$W = \frac{12}{b^2 N(N^2 - 1)} \sum_{i=1}^N (r_{i.} - \frac{r_{..}}{N})^2 \quad (6.4)$$

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

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

As a correlation coefficient, ρ_{av} takes values in the range $(-1, 1)$, with $\rho_{av} = 1$ implying perfect agreement and $\rho_{av} \approx 0$ if the null hypothesis H_0 is true.

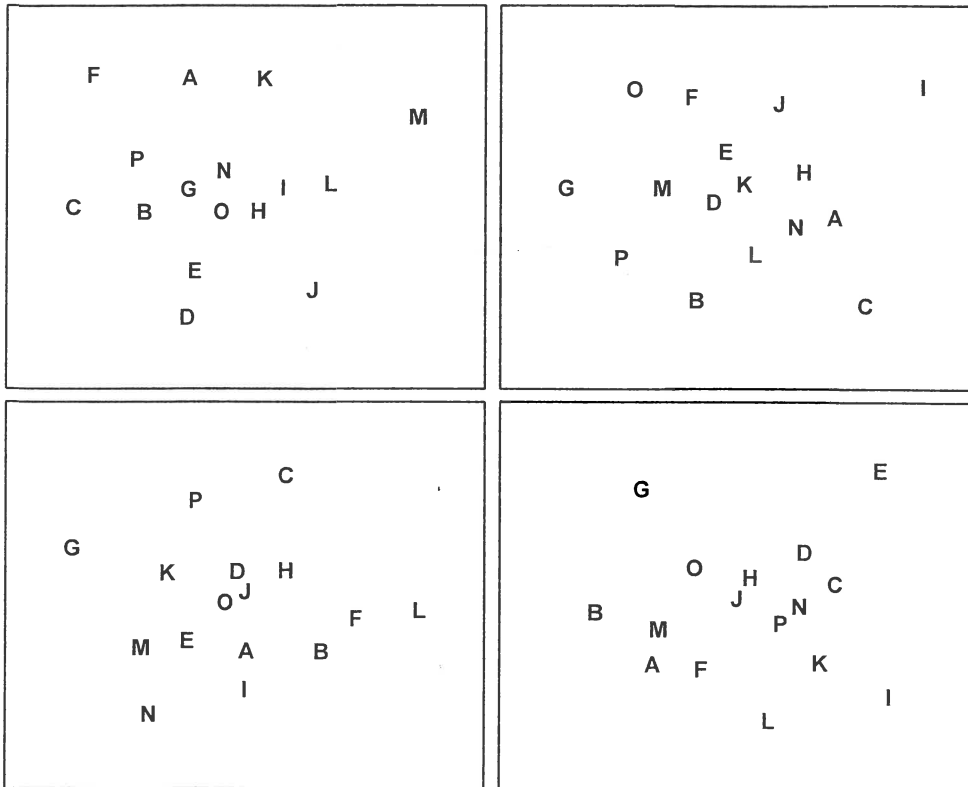


Fig. 6.10. Westerschelde nematodes experiment $\{W\}$. MDS for the 16 treatments (A to P), performed separately for each of the four basins; no shared treatment pattern is apparent (stress ranges from 0.16 to 0.20).

Note that standard significance tests and confidence intervals for ρ or W (e.g. as given in basic statistical tables) are totally invalid, since they rely on the ranks $\{r_{ij}; i=1, \dots, N\}$ being from independent variables; this is obviously not true of similarity coefficients from all possible pairs of a set of (independent) samples. This does not make ρ_{av} any the less appropriate as a measure of agreement whose departure from zero (rejection of H_0) is testable by permutation.

For the nutrient enrichment experiment, Fig. 6.10 shows the separate MDS plots for the 4 mesocosm basins. Although the stress values are rather high (and the plots therefore slightly unreliable as a summary of the among treatment relationships), there appears to be no commonality of pattern, and this is borne out by a near zero value for ρ_{av} of -0.03 . This is central to the range of simulated values for ρ_{av} under H_0 (obtained by permuting treatment labels separately for each block and recomputing ρ_{av}), so the test provides no evidence of any treatment differences. Note that the symmetry of the 2-way layout also allows a test of the (less interesting) hypothesis that there are no block effects, by looking for any consistency in the among-basin relationships across separate analyses for each of the 16 treatments. The test is again non-significant, with $\rho_{av} = -0.02$. The overall negative conclusion to the tests should bar any further attempts at interpretation of these data.

EXAMPLE: Exe nematodes (no replication and missing data)

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

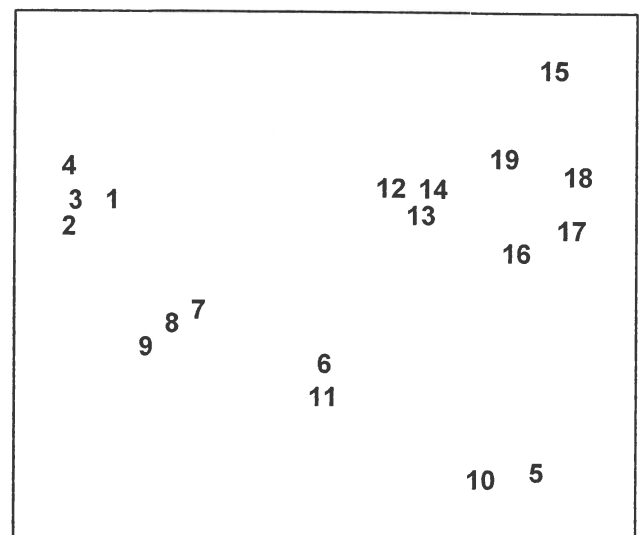


Fig. 6.11. Exe estuary nematodes $\{X\}$. MDS, for 19 inter-tidal sites, of species abundances averaged over 6 bi-monthly sampling occasions; see also Fig.5.1 (stress = 0.05).

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

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

Fig. 6.12 shows evidence of a consistent site pattern, for example in the proximity of sites 12 to 14 and the tendency of site 15 to be placed on its own; the fact that site 15 is missing on one occasion does not undermine this perceived structure. Pairwise computation gives $\rho_{av} = 0.36$ 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 such ρ_{av} value in 999 simulations was 0.30, so the null hypothesis is rejected at the $p < 0.1\%$ level.

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

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

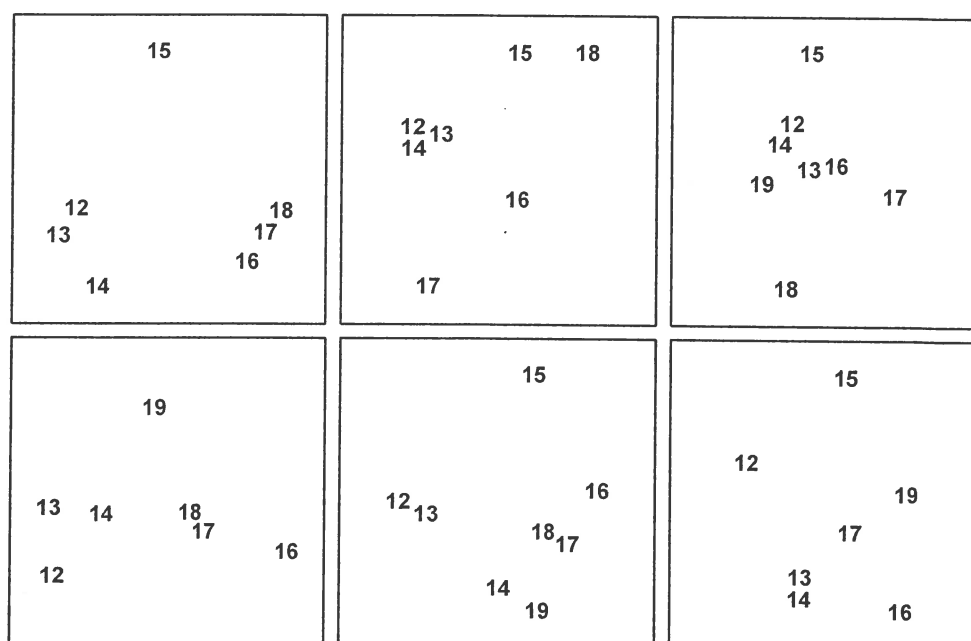


Fig. 6.12. *Exe* estuary nematodes $\{X\}$. MDS for sites 12 to 19 only, performed separately for the 6 sampling times (read across rows for time order); in spite of the occasional missing sample some commonality of site pattern is apparent (stress ranges from 0.01 to 0.08).

RECOMMENDATIONS

- 1) For typical species abundance matrices, it is much better to use an ANOSIM-type permutation procedure rather than a classical MANOVA test; the latter will almost always be totally invalid.
- 2) Choice of the level and type of replication should be carefully considered. Though it is difficult to define power for any of the ANOSIM tests, it is clearly important to take sufficient replicates to generate a large enough set of permutations for meaningful significance levels. Equally important is that replicates should *genuinely* represent the condition being sampled: *pseudo-replication* (see Hurlbert, 1984) is commonplace, e.g. analyses of sub-cores of single cores, or sets of spatially contiguous samples which are unrepresentative of the extent of a site. For pseudo-replicates in a 2-way layout, the only valid course is to average them and carry out the above global test for the case of “no replication”.
- 3) A point that cannot be over-stressed is that ANOSIM tests only apply to groups of samples specified *prior* to seeing (or collecting) the data. A dangerous misconception is that one can use a cluster analysis of the species abundance data to define sample groupings, whose statistical validity can be established by performing an ANOSIM test for differences between these groups. *This is entirely erroneous*, the argument being completely circular. The only safe course here is to use this first set of data to define potential groups of interest, i.e. to erect the hypothesis, and then to collect a further set of data to test that hypothesis.

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.[¶]

Fig. 7.1 displays the results of a cluster analysis on the Exe estuary nematode data {X} first seen 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.

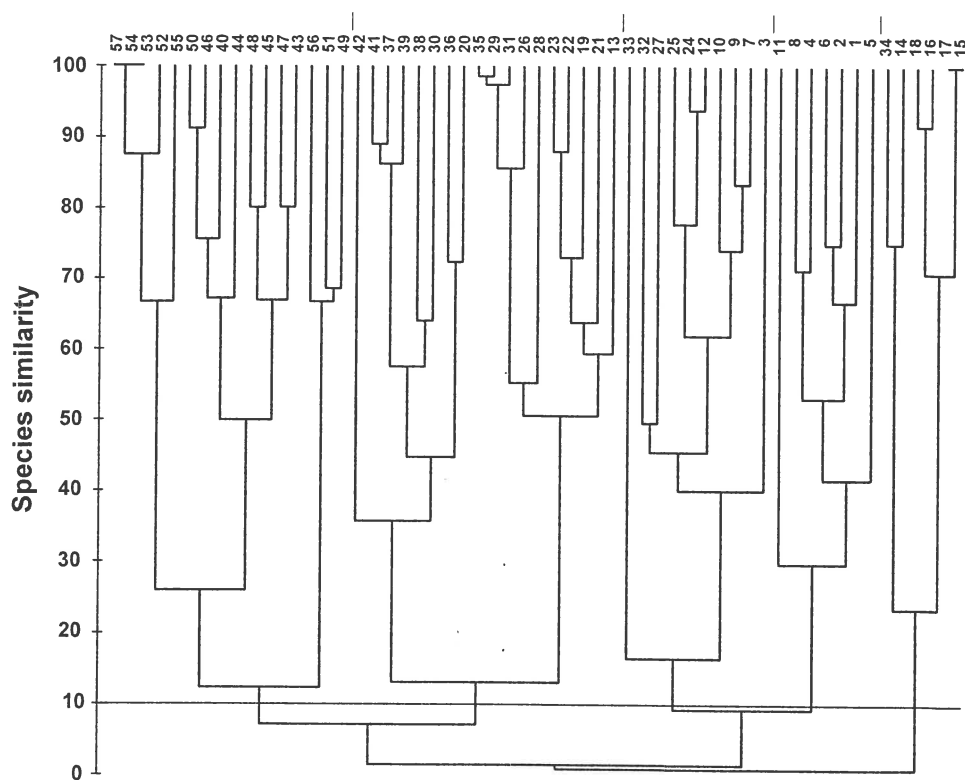


Fig. 7.1. Exe estuary nematodes {X}. Dendrogram using group-average linking on Bray-Curtis *species* similarities from standardised abundance data; the 57 most important species were retained from an original list of 182. The 5 groups defined at arbitrary similarity level of 10% are indicated.

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

The effect of the re-ordering here is to concentrate the higher abundances in the diagonal region of the matrix, and it is then relatively easy to identify species which have characteristically different abundance levels between (say) sample groups 1 and 2 (e.g. species 6, 1, 4, 23, 18, 3). However, for a matrix with larger numbers of species and a less satisfactory species ordination, a more automatic, analytical procedure for identifying influential species is preferable, as follows.

Similarity breakdown

The fundamental information on the multivariate structure of an abundance matrix is summarised in the Bray-Curtis similarities between *samples*, and it is by disaggregating these that one most precisely identifies the species responsible for particular aspects of the multivariate picture.[†] So, first compute the **average dissimilarity** $\bar{\delta}$ between *all* pairs of inter-group samples (e.g. every sample in group 1 paired with every sample in group 2) and then break this average down into separate **contributions from each species** to $\bar{\delta}$.

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

$$\delta_{jk}(i) = 100 \cdot |y_{ij} - y_{ik}| / \sum_{i=1}^p (y_{ij} + y_{ik}) \quad (7.1)$$

$\delta_{jk}(i)$ 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 i th species to the overall dissimilarity $\bar{\delta}$ between groups 1 and 2.[‡] 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.[§] If $\bar{\delta}_i$

[†] This is implemented in the PRIMER SIMPER routine ("similarity percentages"), both in respect of contribution to average similarity within a group and average dissimilarity between groups.

[‡] Though this is a natural definition, it should be noted that, in the general unstandardised case, 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.

[§] 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 i th species.

is large and $SD(\delta_i)$ small (and thus the ratio $\bar{\delta}_i/SD(\delta_i)$ is large), then the i th 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**.

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 $\bar{\delta} = \sum \bar{\delta}_i = 59.5$. Species which are likely to be good discriminators of groups 1 and 2 are indicated by an asterisk in the $\bar{\delta}_i/SD(\delta_i)$ column. The final column rescales the first column to percentages, i.e. it computes the % of the total dissimilarity $\bar{\delta}$ that is contributed by the i th species and then cumulates these percentages down the rows of the table. It can be seen that many of the species play some part in determining the dissimilarity between groups 1 and 2, and this is typical of such analyses. Here, nearly 90% of the contribution to $\bar{\delta}$ is accounted for by the first twelve species listed, with over 50% accounted for by the first five. The results are in good agreement, of course, with the pattern observed in the condensed matrix format of Fig. 7.3.

In much the same way, though perhaps of less practical significance, one can examine the contribution each species makes to the average **similarity within** a group, \bar{S} . The average contribution of the i th species, \bar{S}_i , could be defined by taking the average, over all pairs

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

of samples *within* a group, of the *i*th term in the similarity definition of equation (2.1) (the second form). The more abundant a species is within a group, the more it will contribute to the *intra*-group similarities. It *typifies* that group if it is found at a consistent abundance throughout, so the standard deviation of its contribution $SD(S_i)$ is low, and the ratio $\bar{S}_i/SD(S_i)$ high. Note that this says nothing about whether that species is a good *discriminator* of one group from another; it may be very *typical* of a number of groups.

Such a breakdown is shown for group 1 of the Bristol Channel zooplankton data in Table 7.2. The average similarity within the group is $\bar{S} = 66.3$, with more than two-thirds of this contributed by only three species (6, 18 and 1), the first two of which are found at very consistent levels within the group.

Table 7.2. Bristol Channel zooplankton {B}. Breakdown of average similarity within group 1 into contributions from each species (part only given).

Sp	Name	\bar{S}_i	$SD(S_i)$	$\bar{S}_i/SD(S_i)$	$\Sigma \bar{S}_i$ %
6	<i>Eurytemora affinis</i>	19.3	6.3	3.1*	29.1
18	<i>Pseudocalanus elongatus</i>	14.7	2.7	5.4*	51.3
1	<i>Acartia bifilosa</i>	12.2	6.4	1.9*	69.6
17	<i>Polychaete larvae</i>	3.9	3.1	1.2	75.5
14	<i>Pleuorbrachia pileus</i>	3.4	3.8	0.9	80.7
21	<i>Schistomysis spiritus</i>	3.3	3.6	0.9	85.7
15	<i>Pleuorbrachia pileus</i> <i>juv</i>	3.3	4.7	0.7	90.7
.....

Limitations of the method

The SIMPER procedure has two main constraints which, to some extent, limit its usefulness.[‡]

- a) It applies only to Bray-Curtis dissimilarities, whereas one might legitimately want to examine the influence of particular variables in a more general case, e.g. when the variables are not species abundances but environmental measures, and the dissimilarity coefficient is not Bray-Curtis but Euclidean distance.
- b) It compares two groups of samples at a time, identifying the influential species only for each specific comparison. Some multivariate patterns, however, are not so readily categorised but represent a continuum of community change in response to one or more underlying gradients.

[‡] A third constraint is the restriction to a one-way layout, allowing only comparison of levels of one factor. However, this is easily relaxed, e.g. to a two-way crossed design in which species contributions to the dissimilarities between two levels of factor 1 ('treatments') are averaged across the separate levels of factor 2 ('blocks'), see Platell et al (1998).

What is needed here is a more holistic technique, identifying the set of influential species which, *between them*, capture the full multivariate pattern (whether clustered or forming a gradation), and which operates with any appropriately-defined similarity coefficient. A possible method is suggested in a later chapter (16) on comparing multivariate patterns.

RECOMMENDATIONS

A multivariate display of the samples, either by an ordination or a cluster analysis, is not the end-point of a community analysis; it should be seen as a framework within which the patterns of individual species abundances can be interpreted.

1) This may be by simple re-examination of the data matrix, ordered and re-presented (perhaps averaged within groups) in the light of the information from the multivariate analysis.

2) In the case of a convincing clustering of samples, individual species contributions to the separation of the groups can be examined with the SIMPER procedure. Note that this is not a statistical testing framework, just an exploratory analysis. It indicates which species are principally responsible *either* for an observed clustering pattern *or* for differences between sets of samples that have been defined *a priori* and are confirmed to differ in community structure by the tests of Chapter 6.

3) Species identified in this manner (or by the more general pattern-matching procedures discussed in Chapter 16) are sometimes viewed most effectively in conjunction with the ordination. One at a time, they can be superimposed on an MDS (or PCA) plot, as circles whose varying diameters reflect the abundance changes for that species across samples (see, for example, Fig. 15.3).[†]

[†] Both the PRIMER MDS and PCA routines allow for straightforward superimposition of either individual species abundances (from the data file used to create the ordination) or from an independent data set (e.g. of environmental variables measured for the same set of samples).

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

UNIVARIATE MEASURES

A variety of different statistics (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). Abundance or biomass totals (or averages) are not dimensionless quantities so tend to be less informative than **diversity indices**, such as: richness of the sample, in terms of the number of species for a given number of individuals; dominance in the way in which the total number of individuals in the sample is divided up among the different species (technically referred to as the **species abundance distribution**).

Diversity indices

The main aim is to reduce the multivariate (multi-species) complexity of assemblage data into a single index (or small number of indices) evaluated for each sample, which can then be handled statistically by univariate analyses. It will often be possible to apply standard normal-theory tests (t -tests and ANOVA) to such derived indices, possibly after transformation (see page 6-1).

A bewildering variety of diversity indices has been used, in a vast literature on the subject, and some of the most frequently used candidates are listed below.[¶] More detail can be found in two (of several) overviews aimed specifically at the biological reader, Heip *et al* (1988) and Magurran (1991). It should be noted, however, that diversity indices of this type tend to exploit some combination of just two features of the sample information[†]:

a) **Species richness**. This measure is either simply the total number of species present or some adjusted

[¶] The *PRIMER DIVERSE* routine permits selection of a subset from this list (and more), returning the values to a worksheet for further analysis or export to a mainstream statistical package.

[†] This can often be demonstrated quite effectively by calculating a plethora of indices for a set of samples, then submitting the resulting worksheet to the *PCA* routine (treating the diversity indices as if they were environmental variables, see Chapter 4). The first two principal axes are often then found to explain most of the variability in the full suite of indices, i.e. the information content is inherently only 2-dimensional.

form which attempts to allow for differing numbers of individuals. Obviously, for samples which are strictly comparable, 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 emphasize the species richness or equitability components of diversity to varying degrees. The most commonly used diversity measure is the **Shannon** (or Shannon–Wiener) **diversity index**:

$$H' = - \sum_i p_i \log(p_i) \quad (8.1)$$

where p_i is the proportion of the total count (or biomass etc) arising from the i th species. Note that logarithms to the base 2 are often sometimes used in the calculation, reflecting the index's genesis in information theory. There is, however, no natural biological interpretation here, so the more usual natural logarithm (to the base e) is probably preferable, and commonly used. Clearly, when comparing published indices it is important to check that the same logarithm base has been used in each case. If not, it is simple to convert between results since $\log_2 x = (\log_e x) / (\log_e 2)$, i.e. all indices just need to be multiplied or divided by a constant factor. Whether it is sensible to compare H' across different studies is another matter, since Chapter 17 shows that, like many of the indices given here (Simpson being a notable exception, Fig. 17.1), it can be sensitive to the degree of sampling effort. Hence H' should only be compared across equivalent sampling designs.

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). Alternatively,

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:

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

Equitability

This is often expressed as *Pielou's evenness index*:

$$J' = H' / H'_{max} = H' / \log S \quad (8.3)$$

where H'_{max} is the maximum possible value of Shannon diversity, i.e. that which would be achieved if all species were equally abundant (namely, $\log S$).

Simpson

Another commonly used measure is the *Simpson index*, which has a number of forms:

$$\begin{aligned} \lambda &= \sum p_i^2 \\ 1-\lambda &= 1 - (\sum p_i^2) \\ \lambda' &= \{\sum_i N_i(N_i-1)\} / \{N(N-1)\} \\ 1-\lambda' &= 1 - \{\sum_i N_i(N_i-1)\} / \{N(N-1)\} \end{aligned} \quad (8.4)$$

where N_i is the number of individuals of species i . The index λ has a natural interpretation as the probability that any two individuals from the sample, chosen at random, are from the same species (λ is always ≤ 1). It is a *dominance* index, in the sense that its largest values correspond to assemblages whose total abundance is dominated by one, or a very few, of the species present. Its complement, $1 - \lambda$, is thus an equitability or *evenness* index, taking its largest value (of $1 - S^{-1}$) when all species have the same abundance. The slightly revised forms λ' and $1 - \lambda'$ are appropriate when total sample size (N) is small (they correspond to choosing the two individuals at random *without* replacement rather than *with* replacement). As with Shannon, Simpson diversity can be employed when the $\{p_i\}$ come from proportions of biomass, standardised abundance or other data which are not strictly integral counts but, in that case, the λ' and $1 - \lambda'$ forms are not appropriate.

Other count-based measures

Further well-established indices include that of Brillouin (see Pielou 1975):

$$H = N^{-1} \log_e \{N! / (N_1! N_2! \dots N_S!)\} \quad (8.5)$$

and a further model-based description, Fisher's α (Fisher *et al.*, 1943), which is the shape parameter, fitted by maximum likelihood, under the assumption that the species abundance distribution follows a log series. This has certainly been shown to be the case for some ecological data sets but can by no means be universally assumed, and (as with Brillouin) its use is clearly restricted to genuine (integral) counts.

The final option in this category is the *rarefaction* method of Sanders (1968) and Hurlbert (1971), which under the strict assumption that individuals arrive in the sample independently of each other, can be used to project back from the counts of total species (S) and individuals (N), how many species (ES_n) would have been 'expected' had we observed a smaller number (n) of individuals:

$$ES_n = \sum_{i=1}^S \left[1 - \frac{(N - N_i)!(N - n)!}{(N - N_i - n)!N!} \right] \quad (8.6)$$

The idea is thereby to generate an absolute measure of species richness, say ES_{100} (the number of different species 'expected' in a sample of 100 individuals), which can be compared across samples of very differing sizes. It must be admitted, however, that the independence assumption is practically unrealistic. It corresponds to individuals from each species being spatially randomly distributed, giving rise to independent Poisson counts in replicate samples. This is rarely observed in practice, with most species exhibiting some form of spatial clustering, which can often be extreme. Rarefaction will then be strongly biased, consistently overestimating the expected number of species for smaller sample sizes.

Hill numbers

Finally, Hill (1973b) proposed a unification of several diversity measures in a single statistic, which includes as special cases:

$$\begin{aligned} N_0 &= S \\ N_1 &= \exp(H) \\ N_2 &= 1 / \sum p_i^2 \\ N_\infty &= 1 / \max\{p_i\} \end{aligned} \quad (8.7)$$

N_1 is thus a transform of Shannon diversity, N_2 the reciprocal of Simpson's λ and N_∞ is clearly another possible evenness index, taking larger values if no species dominates the total abundance. Other variations on these *Hill numbers* are given by Heip (1988).

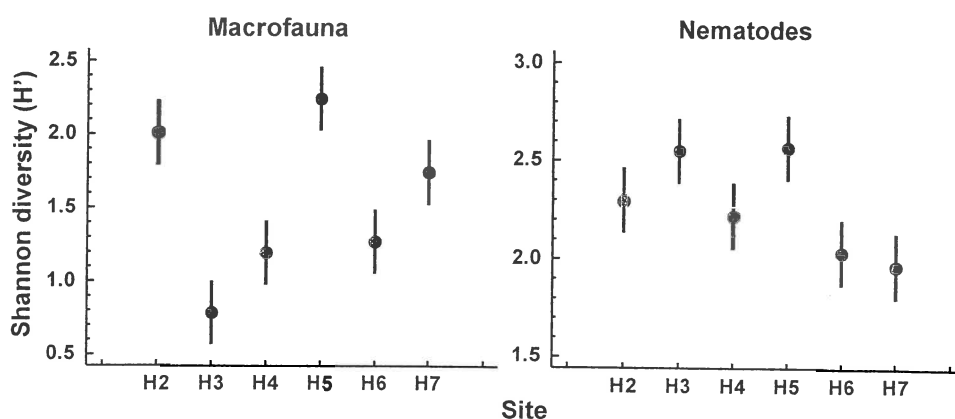


Fig. 8.1. Hamilton Harbour, Bermuda {H}. Diversity (H') and 95% confidence intervals for macrobenthos (left) and meiobenthic nematodes (right) at six stations.

Units of measurement

The numbers of individuals belonging to each species are the most common units used in the calculation of the above indices. For internal comparative purposes other units can sometimes be used, e.g. biomass or total cover of each species along a transect or in quadrats (e.g. for hard-bottom epifauna), but obviously diversity measures using different units are not comparable. Often, on hard bottoms where colonial encrusting organisms are difficult to enumerate, total or percentage cover will be much more realistic to determine than species abundances.

Representing communities

Changes in univariate indices between sites or over time are usually presented graphically simply as plots of means and confidence intervals for each site or time. For example, Fig. 8.1 graphs the differences in diversity of the macrobenthos and meiobenthic nematodes at six stations in Hamilton Harbour, Bermuda, showing that there are clear differences in diversity between sites for the former but much less obvious differences for the latter. Fig. 8.2 graphs the temporal changes in three univariate indices for reef corals at South Tikus Island, Indonesia, spanning the period of the 1982-3 El Niño (an abnormally long period of high water temperatures which caused extensive coral bleaching in many areas throughout the Pacific). Note the dramatic decline between 1981 and 1983 and subsequent partial recovery in both the number of species (S) and the Shannon diversity (H'), but no obvious changes in evenness (J).

Discriminating sites or times

The significance of differences in univariate indices between sampling sites or times can simply be tested by one-way analysis of variance (ANOVA), followed

by t-tests or (preferably) multiple comparison tests for individual pairs of sites[†]; see the discussion at the start of Chapter 6.

Determining stress levels

Increasing levels of environmental stress have historically 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. Subsequent theories on the influence of disturbance or stress on diversity have suggested 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 community 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).

[†] These tools are routinely available in all basic statistics packages and are not replicated in the current version of PRIMER; the DIVERSE routine can simply output the chosen suite of indices for each sample as a worksheet, which can be saved in Excel (or text) format for input to many such packages.

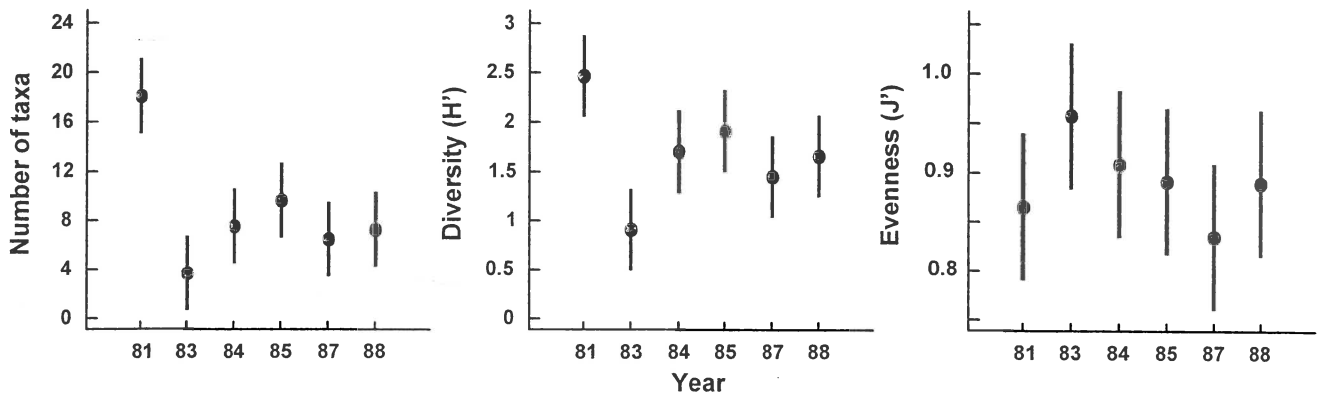


Fig. 8.2. Indonesian reef corals, South Tikus Island {I}. Total number of species (S), Diversity (H') and Evenness (J') based on coral species cover data along transects, spanning the 1982-3 El Niño.

Caswell's neutral model

In some circumstances, the equitability component of diversity can, however, be compared with a theoretical expectation for 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 ($E(H')$):

$$V = \frac{[H' - E(H')]}{SD(H')} \quad (8.8)$$

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 and Lambshead (1989) is useful.[¶]

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 might indicate 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/DISTRIBUTIONAL 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

[¶] This is implemented in the PRIMER CASWELL routine, but the significance aspects should be treated with some caution since they are inevitably crucially dependent on the neutral model assumptions. These are usually over-simplistic for real assemblages (even when genuinely neutral, in the sense that their species do not interact) because they again assume simple spatial randomness.

community structure which are not a function of the specific taxa present, and may therefore be related to levels of biological 'stress'.[†]

- 1) **Rarefaction curves** (Sanders, 1968) were among the earliest to be used in marine studies. They are plots of the number of individuals on the *x*-axis against the number of species on the *y*-axis. The more diverse the community is, the steeper and more elevated is the rarefaction curve. The sample sizes (*N*) may differ widely between stations, but the relevant sections of the curves can still be compared.
- 2) Gray and Pearson (1982) recommend plotting the number of species in $\times 2$ **geometric abundance classes** as a means of detecting the effects of pollution stress. These are plots of the number of species represented by only 1 individual in the sample (class 1), 2–3 individuals (class 2), 4–7 (class 3), 8–15 (class 4) etc. In unpolluted situations there are many rare species and the curve is smooth with its mode well to the left. In polluted situations there are fewer rare species and more abundant species so that the higher geometric abundance classes are more strongly represented, and the curve may also become more irregular or 'jagged' (although this latter feature is more difficult to quantify). Gray and Pearson further suggest that it is the species in the intermediate abundance classes 3 to 5 that are the most sensitive to pollution-induced changes and might best illustrate the differences between polluted and unpolluted sites (i.e. this is a way of selecting 'indicator species' objectively).

- 3) **Ranked species abundance (dominance) curves** are based on the ranking of species (or higher taxa) in decreasing order of their importance in terms of abundance or biomass. The ranked abundances, expressed as a percentage of the total abundance of all species, are plotted against the relevant species rank. Log transformations of one or both axes have frequently been used to emphasise or downweight different sections of the curves. Logging the *x* (rank) axis enables the distribution of the commoner species to be better visualised.

[†] Two plotting programs of this type are available in the PRIMER package: Geometric class plots, which produce a frequency distribution of geometric abundance (/biomass) classes, and Dominance plots, which generate ranked species abundance (/biomass) curves, choosing from ordinary, cumulative or partial forms, and single or dual (abundance-biomass comparison) curves, as illustrated in the remainder of this chapter.

- 4) ***k*-dominance curves** are cumulative ranked abundances plotted against species rank, or log species rank (Lambhead *et al.*, 1983). This has a smoothing effect on the curves. Ordering of curves on a plot will obviously be the reverse of rarefaction curves, with the most elevated curve having the *lowest* diversity. To compare *dominance* separately from the *number of species*, the *x*-axis (species rank) may be rescaled from 0–100 (relative species rank), to produce **Lorenz curves**.

EXAMPLES: Garroch Head and Ekofisk macrofauna

Plots of geometric abundance classes along a transect across the Garroch Head {G} sewage-sludge dump site (Fig. 8.3) are given in Fig. 8.4. Note that the curves are very steep at both ends of the transect (the relatively unpolluted stations) with many species represented by only one individual, and they extend across very few abundance classes (6 at station 1 and 3 at station 12). As the dump centre at station 6 is approached the curves become much flatter, extending over many more abundance classes (13 at station 7), and there are fewer rare species.

In Fig. 8.5a, average ranked species abundance curves (with the *x*-axis logged) are given for the macrobenthos at a group of 6 sampling stations within 250m of the current centre of oil-drilling activity at the Ekofisk field in the North Sea {E}, compared with a group of 10 stations between 250m and 1km from the centre (see inset map in Fig. 10.6a for locations of these stations). Note that the curve for the more polluted (inner) stations is J-shaped, showing high dominance of abundant species, whereas the curve for the less polluted (outer) stations is much flatter, with low dominance

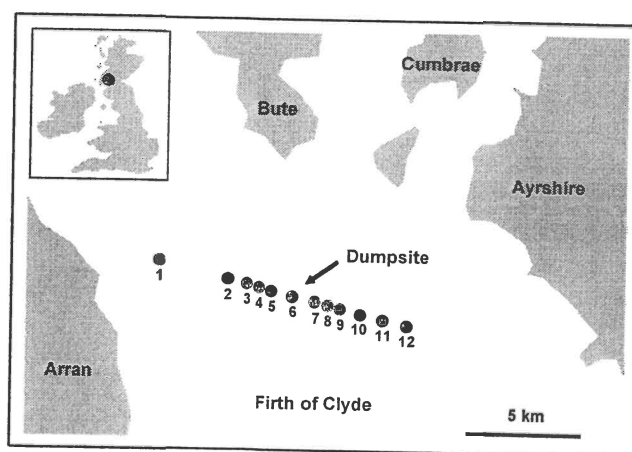


Fig. 8.3. Garroch Head macrofauna {G}. Map showing location of dump-ground and position of sampling stations (1–12); the dump centre is at station 6.

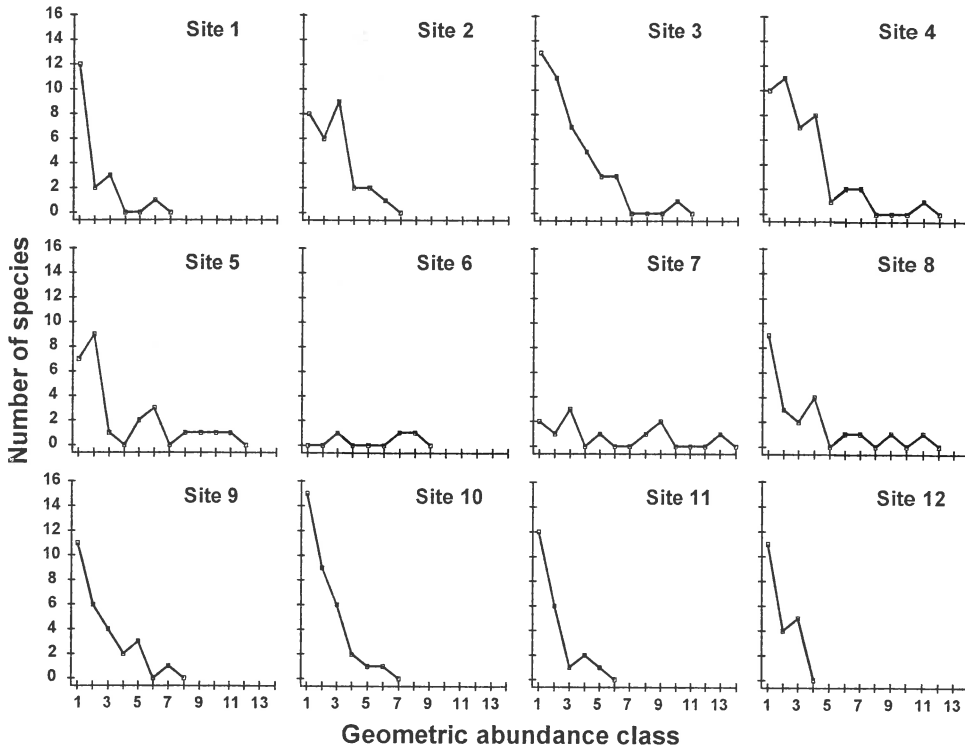


Fig. 8.4. Garroch Head macrofauna {G}. Plots of $\times 2$ geometric species abundance classes for the 12 sampling stations shown in Fig. 8.3.

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 250m–1km stations.

Abundance/biomass comparison plots

Whether k -dominance curves are plotted from the species abundance distribution or from species biomass values, the y -axis is always scaled in the same range (0 to 100). This facilitates 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 can be numerically significant but do not represent a large proportion of the community biomass. When pollution perturbs a community, conservative species are less favoured in comparison with opportunists. Thus, under pollution stress, the distribution of numbers of individuals among species behaves differently from the distribution of biomass among species.

The ABC method, as originally described by Warwick (1986), involves the plotting of separate k -dominance

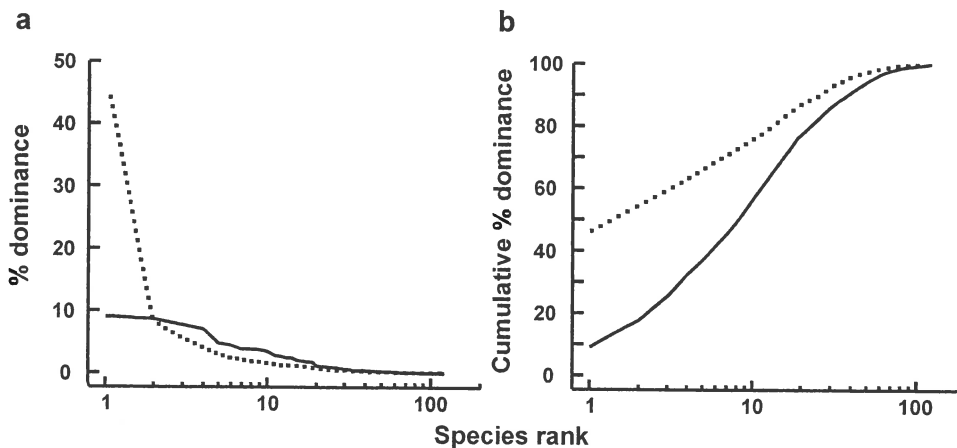


Fig. 8.5. Ekofisk macrobenthos {E}. a) Average ranked species abundance curves (x-axis logged) for 6 stations within 250m of the centre of drilling activity (dotted line) and 10 stations between 250m and 1km from the centre (solid line); b) k -dominance curves for the same groups of stations.

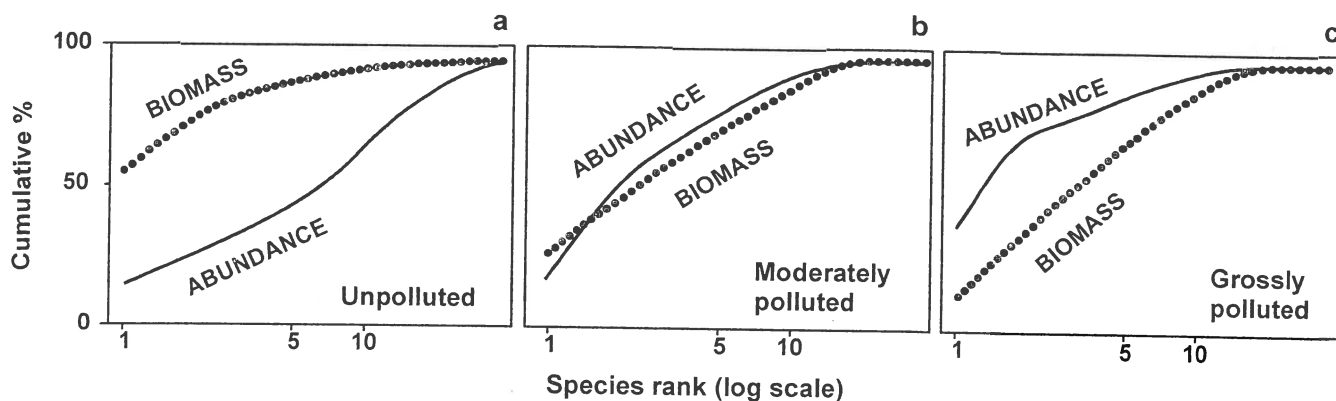


Fig. 8.6. Hypothetical *k*-dominance curves for species biomass and abundance, showing 'unpolluted', 'moderately polluted' and 'grossly polluted' conditions.

curves (Lambshhead *et al*, 1983) for species abundances and species biomasses on the same graph and making a comparison of the forms of these curves. The species are ranked in order of importance in terms of abundance or biomass on the *x*-axis (logarithmic scale) with percentage dominance on the *y*-axis (cumulative scale). In undisturbed communities the biomass is dominated by one or a few large species, leading to an elevated biomass curve. Each of these species, however, is represented by rather few individuals so they do not dominate the abundance curve, which shows a typical diverse, equitable distribution. Thus, the *k*-dominance curve for biomass lies above the curve for abundance for its entire length (Fig. 8.6a). Under moderate pollution (or disturbance), the large competitive dominants are eliminated and the inequality in size between the numerical and biomass dominants is reduced so that the biomass and abundance curves are closely coincident and may cross each other one or more times (Fig. 8.6b). As pollution becomes more severe, benthic communities become increasingly dominated by one or a few opportunistic species which whilst they dominate the numbers do not dominate the biomass, because they are very small-bodied. Hence, the abundance curve lies above the biomass curve throughout its length (Fig. 8.6c).

The contention is that these three conditions (termed *unpolluted*, *moderately polluted* and *grossly polluted*) should be recognisable in a community without reference to control samples in time or space, the two curves acting as an "internal control" against each other. Reference to spatial or temporal control samples is, however, still desirable. Adequate replication of sampling is a prerequisite of the method, since the large biomass dominants are often represented by few individuals, which will be liable to a higher sampling error than the numerical dominants.

EXAMPLES: Loch Linnhe and Garroch Head macrofauna

ABC curves for the macrobenthos at site 34 in Loch Linnhe, Scotland $\{L\}$ between 1963 and 1973 are given in Fig. 8.7. The time course of organic pollution from a pulp-mill, and changes in species diversity (*H'*), are shown top left. Moderate pollution started in 1966, and by 1968 species diversity was reduced. Prior to 1968 the ABC curves had the unpolluted configuration. From 1968 to 1970 the ABC plots indicated moderate pollution. In 1970 there was an increase in pollutant loadings and a further reduction in species diversity, reaching a minimum in 1972, and the ABC plots for 1971 and 1972 show the grossly polluted configuration. In 1972 pollution was decreased and by 1973 diversity had increased and the ABC plots again indicated the unpolluted condition. Thus, the ABC plots provide a good 'snapshot' of the pollution status of the benthic community in any one year, without reference to the historical comparative data which would be necessary if a single species diversity measure based on the abundance distribution was used as the only criterion.

ABC plots for the macrobenthos along a transect of stations across the accumulating sewage-sludge dump-ground at Garroch Head, Scotland $\{G\}$ (Fig 8.3) are given in Fig. 8.8. Note how the ABC curves behave along the transect, with the peripheral stations 1 and 12 having unpolluted configurations, those near the dump-centre at station 6 with grossly polluted configurations and intermediate stations showing moderate pollution. Of course, at the dump-centre itself there are only three species present, so that any method of data analysis would have indicated gross pollution. However, the biomass and abundance curves start to

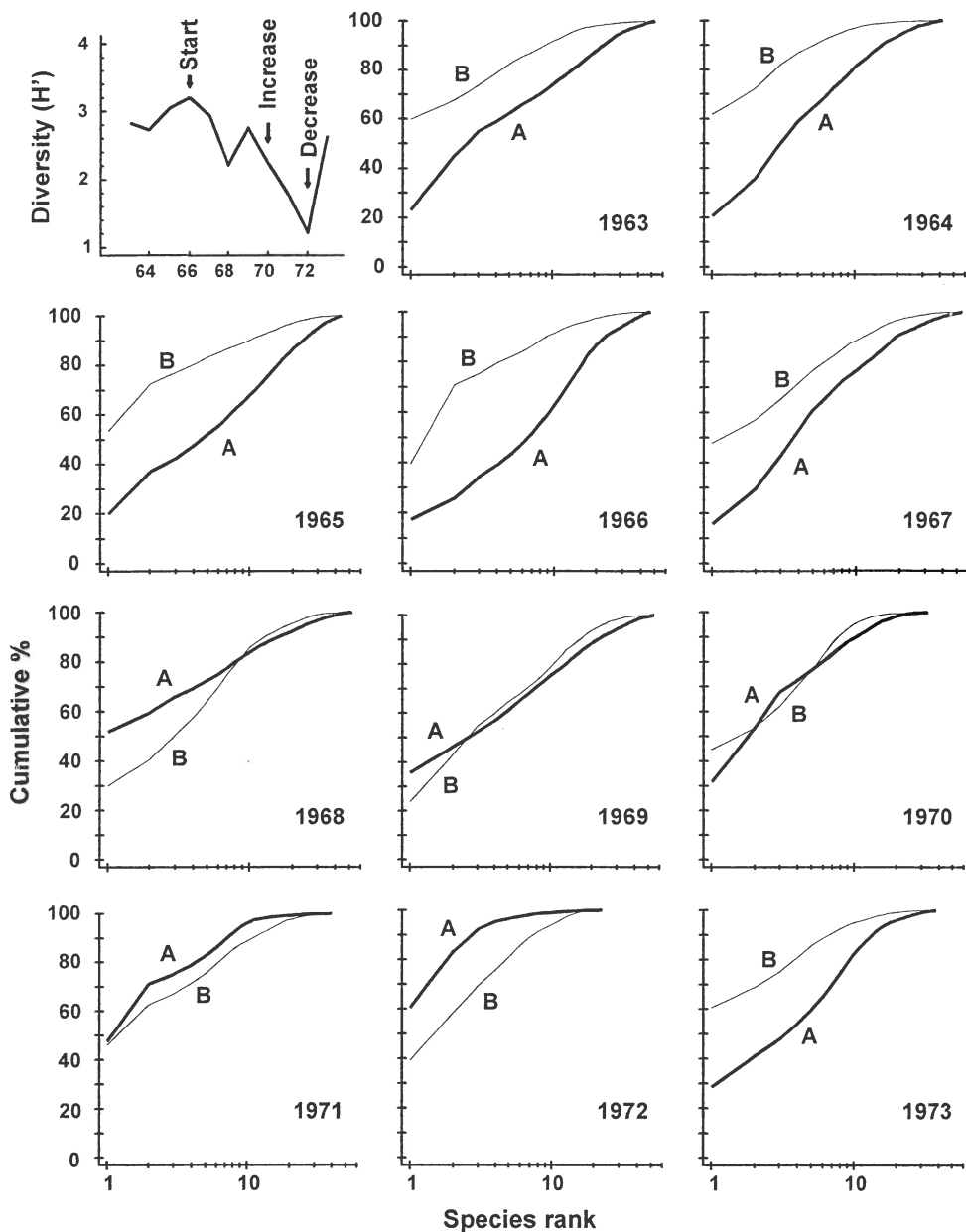


Fig. 8.7. Loch Linnhe macrofauna {L}. Shannon diversity (H') and ABC plots over the 11 years, 1963 to 1973. Abundance = thick line, biomass = thin line.

become transposed at some distance from the dump-centre, when species richness 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:

$$y_i' = \log[(1 + y_i)/(101 - y_i)] \quad (8.9)$$

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.

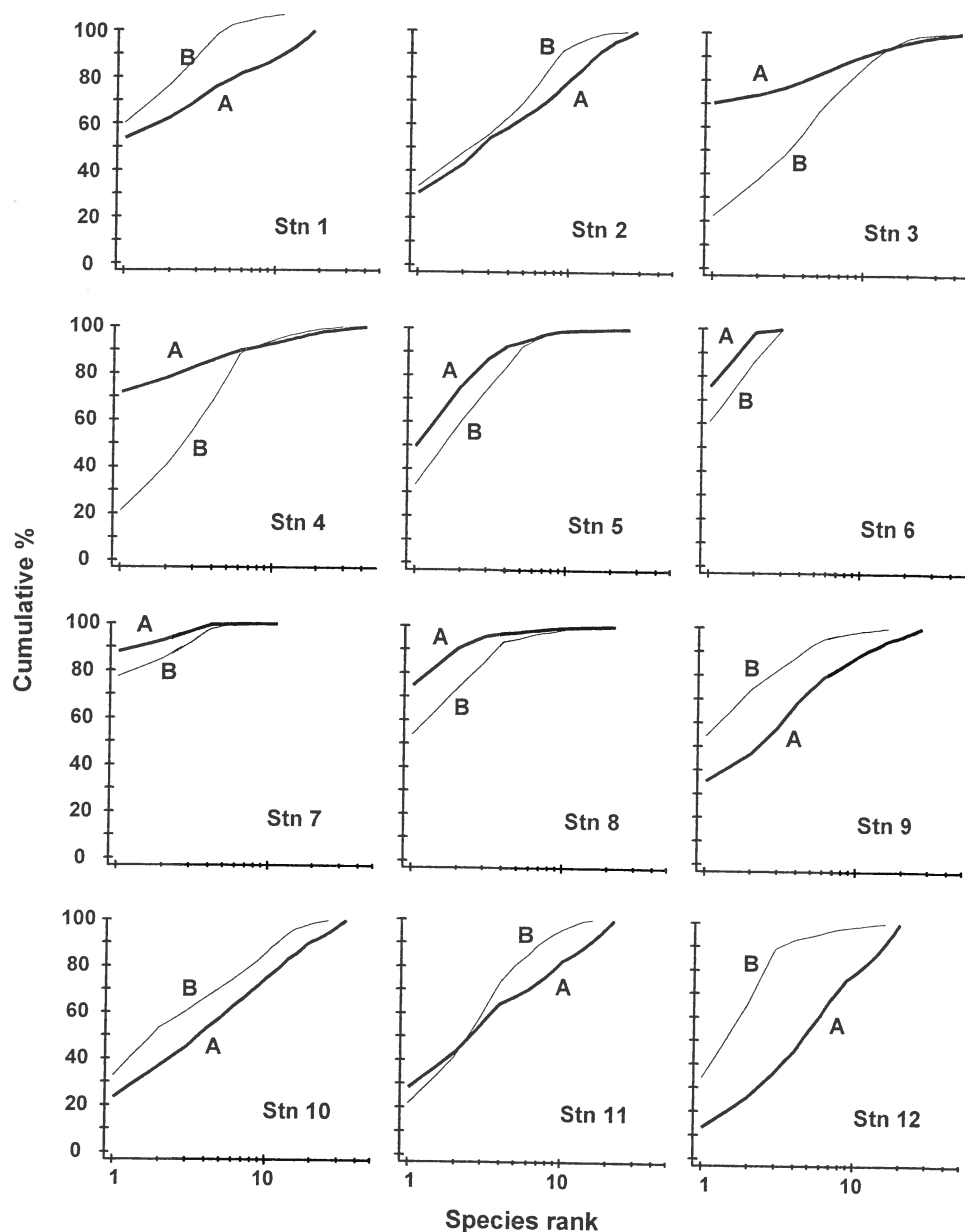


Fig. 8.8. Garroch Head macrofauna {G}. ABC curves for macrobenthos in 1983. Abundance = thick line, biomass = thin line.

Partial dominance curves

A second problem with the cumulative nature of k -dominance (and ABC) curves is that the visual information presented is over-dependent on the single most dominant species. The unpredictable presence of large numbers of a species with small biomass, perhaps an influx of the juveniles of one species, may give a false impression of disturbance. With genuine disturbance, one might expect patterns of ABC curves to be unaffected by successive removal of the one or two most dominant species in terms of abundance or biomass, and so Clarke (1990) recommended the use of *partial dominance curves*, which compute the dominance of the second ranked species *over the remainder* (ignoring the first ranked species), the

same with the third most dominant etc. Thus if a_i is the absolute (or percentage) abundance of the i th species, when ranked in decreasing abundance order, the partial dominance curve is a plot of p_i against $\log i$ ($i = 1, 2, \dots, S-1$), where

$$p_1 = 100a_1 / \sum_{j=1}^S a_j, \quad p_2 = 100a_2 / \sum_{j=2}^S a_j, \\ \dots, \quad p_{S-1} = 100a_{S-1} / (a_{S-1} + a_S) \quad (8.10)$$

Earlier values can therefore never affect later points on the curve. The partial dominance curves (ABC) for undisturbed macrobenthic communities typically look like Fig. 8.10, with the biomass curve (thin line) above the abundance curve (thick line) throughout its length. The abundance curve is much smoother than

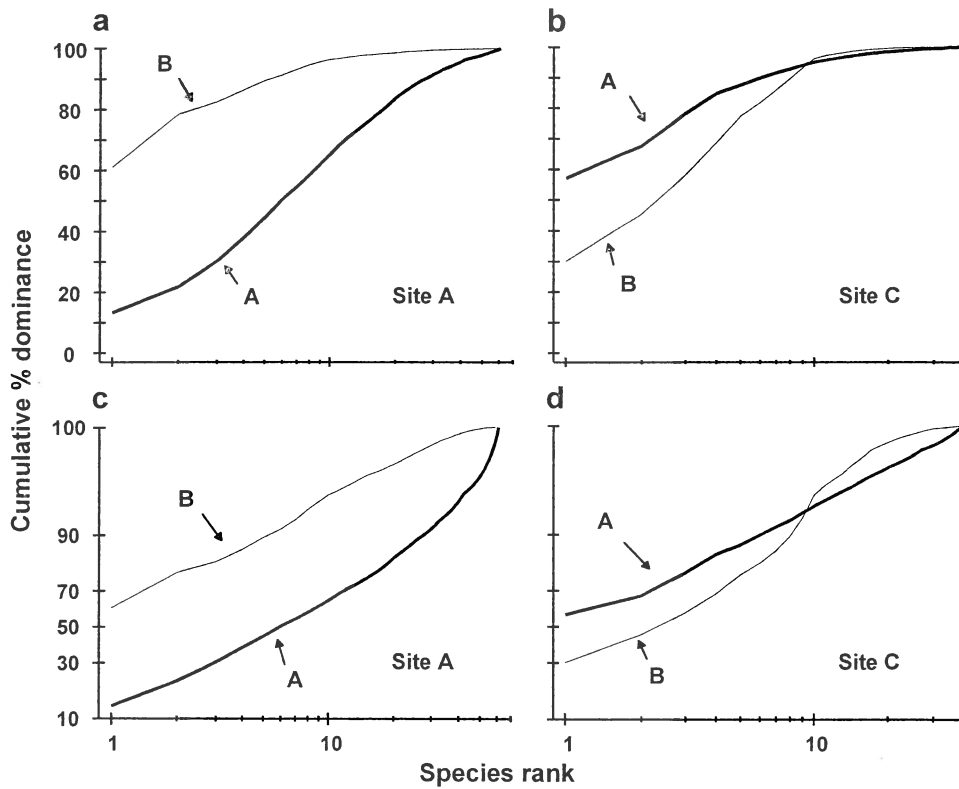


Fig. 8.9. *Frierfjord macrofauna {F}*. a), b) Standard ABC plots for sites A (reference) and C (potentially impacted). c), d) ABC plots for sites A and C with the y-axis subjected to modified logistic transformation. Abundance = thick line, biomass = thin line.

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 (see 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.

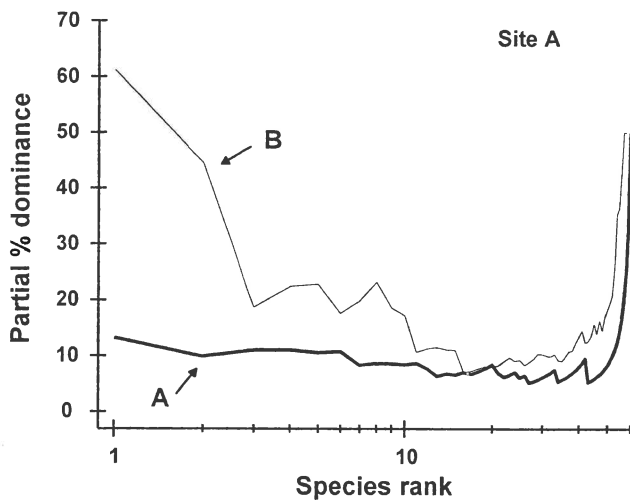


Fig. 8.10. *Frierfjord macrofauna {F}*. Partial dominance curves (abundance/biomass comparison) for reference site A (c.f. Figs 8.9a,c for corresponding standard and transformed ABC plots).

Phyletic role in ABC method

Warwick and Clarke (1994) have 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 (Mollusca, Crustacea, Echinodermata). The shift within polychaetes reflects the substitution of larger-bodied by smaller-bodied species, and not a change in the average size of individuals within a species. In most instances the phyletic changes reinforce the trend in species substitutions within the polychaetes, to produce the overall ABC

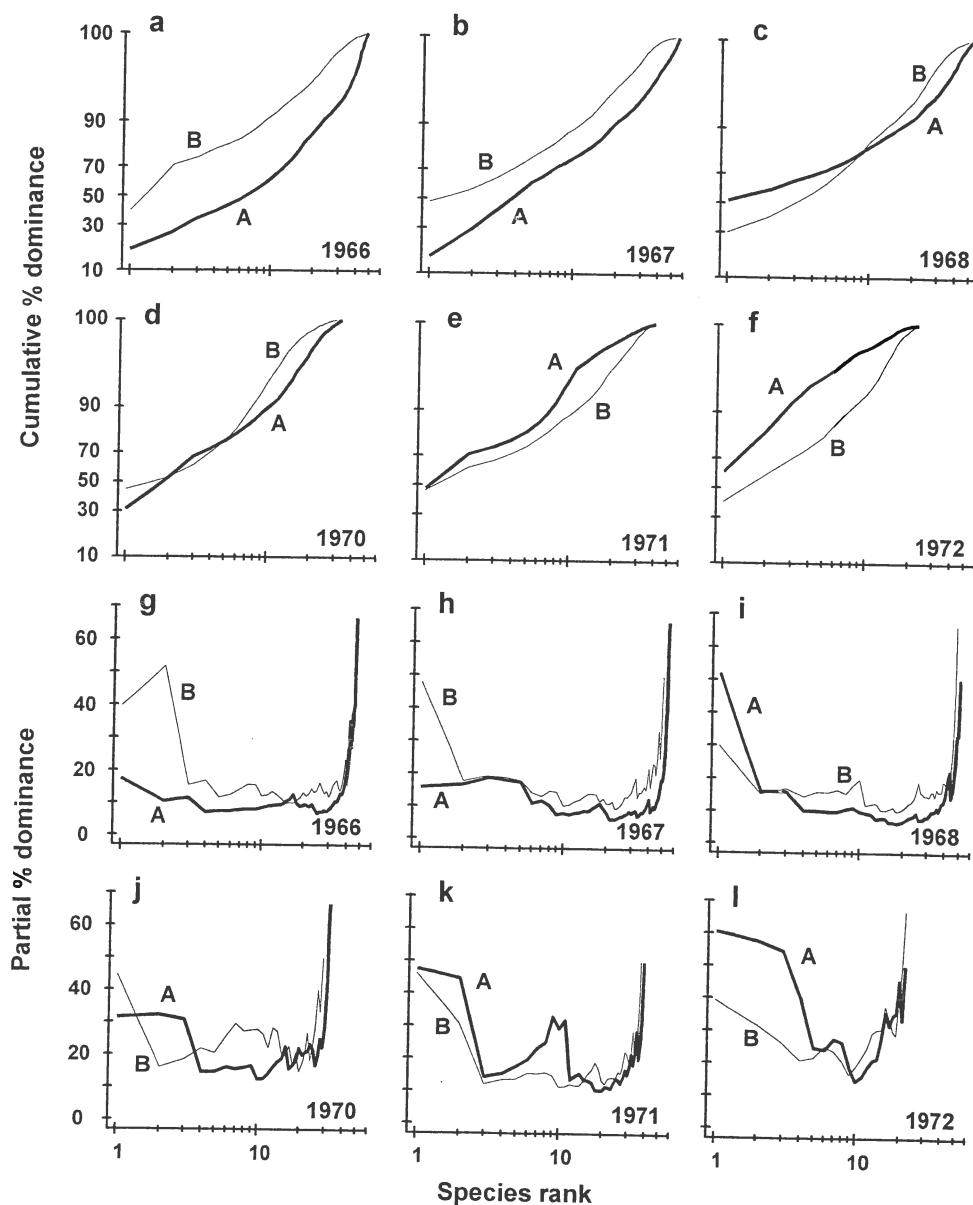


Fig. 8.11. *Loch Linnhe macrofauna {L}*. Selected years 1966-68 and 1970-72. a-f) ABC curves (logistic transform). g-l) Partial dominance curves for abundance (thick line) and biomass (thin line) for the same years.

response, but in some cases they may work against each other. In cases where the ABC method has not succeeded as a measure of the pollution status of marine macrobenthic communities, it is because small non-polychaete species have been dominant. Prior to the Amoco Cadiz oil-spill, small ampeliscid amphipods (crustacea) were present at the Pierre Noire station in relatively high abundance (Dauvin, 1984), and their disappearance after the spill confounded the ABC plots (Ibanez and Dauvin, 1988). It was the erratic presence of large numbers of small amphipods (*Corophium*) or molluscs (*Hydrobia*) which confounded these plots in the Wadden Sea (Beukema, 1988). These small non-polychaetous species are not indicative of polluted conditions, as Beukema points out. Indications of pollution or disturbance detected by this method should therefore be viewed with caution if the species responsible for the polluted configurations are not polychaetes.

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: cumulative dominance curves are plotted, 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

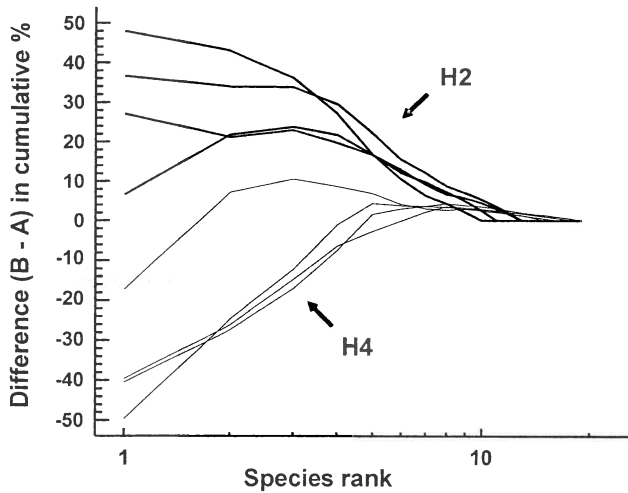


Fig. 8.12. Hamilton Harbour macrobenthos {H}. Difference (B-A) between cumulative dominance curves for biomass and abundance for four replicate samples at stations H2 (thick line) and H4 (thin line).

(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.[¶] 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 $\sum(B_i - A_i)$ 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:

$$W = \sum_{i=1}^S (B_i - A_i) / [50(S - 1)] \quad (8.11)$$

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 dump-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 four replicates at H2 and -0.082, 0.053, -0.081 and -0.068 for the four H4 samples. These data can be input into a standard univariate ANOVA (equivalent in the case of two groups to a standard 2-sample t -test), showing that there is indeed a clearly established difference in abundance-biomass patterns between these two sites ($F = 45.3, p < 0.1\%$).

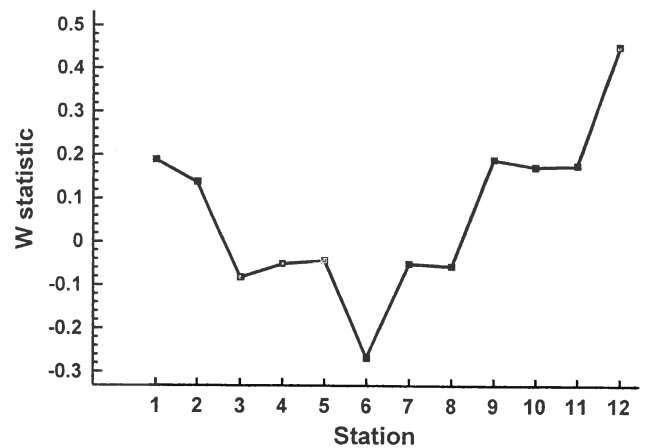


Fig. 8.13. Garroch Head macrofauna {G}. W values corresponding to the 12 ABC curves of Fig. 8.8, plotted against station number; station 6 is the centre of the dump ground (Fig. 8.3).

More general forms of hypothesis testing are possible, likely to be particularly relevant to the comparison of k -dominance curves calculated for replicates at a number of sites, times or conditions (or in some two-way layout, as discussed in Chapter 6). A measure of “dissimilarity” could be constructed between any pair of k -dominance (or $B-A$) curves, for example based on their absolute distance apart, summed across the species ranks. When computed for all pairs of samples in a study this provides a (ranked) triangular dissimilarity matrix, essentially similar in structure to that from a multivariate analysis; thus the 1-way and 2-way ANOSIM tests (Chapter 6) can be used in exactly the same way to test hypotheses about differences between *a priori* specified groups of samples. Clarke (1990) discusses some appropriate definitions of “dissimilarity” for use with dominance curves in such tests.

[¶] Note that, as always with an ABC curve, B_i and A_i do not necessarily refer to values for the same 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	8	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

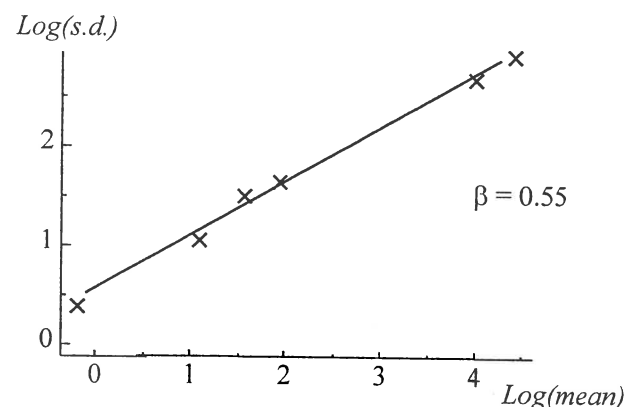
Two features are apparent:

- 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 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 $\log(\text{s.d.})$ against $\log(\text{mean})$ 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 $1 - \beta$, 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

deviations of the root-transformed abundances: the s.d. is now remarkably constant in spite of the order of magnitude difference in mean values across sites. An ANOVA would now be a valid and effective testing procedure for the hypothesis of “no site-to-site differences”, and the means and 95% confidence intervals for each site can be back-transformed to the original measurement scales for a more visually helpful plot.

Table 9.2. Frierfford macrofauna {F}. Mean and standard deviation over the four replicates at each site, for root-transformed abundances of *Thyasira sp.*

Site:	A	B	C	D	E	G
Mean(y^*)	2.01	1.45	0.43	2.42	9.00	7.40
S.d.(y^*)	0.97	1.10	0.87	1.10	1.04	1.04

Like all illustrations, though genuine enough, this one works out too well to be typical! In practice, there is usually a good deal of scatter in the *log s.d.* versus *log mean* plots; more importantly, most species will have many more zero entries than in this example and it is *impossible* to “transform these away”: species abundance data are simply not normally distributed and can only rarely be made so. Another important point to note here is that it is never valid to “snoop” in a data matrix of, perhaps, several hundred species for one or two species that display apparent differences between sites (or times), and then test the “significance” of these groups for that species. This is the problem of **multiple comparisons** referred to in Chapter 6; a purely random abundance matrix will contain *some* species which fallaciously appear to show differences between groups in a standard 5% significance level ANOVA (even were the ANOVA assumptions to be valid). The best that such “snooping” can do, in hypothesis testing terms, is identify one or two potential key or indicator species that can be tested with an entirely independent set of samples.

These two difficulties between them motivate the only satisfactory approach to most community data sets: a properly multivariate one in which all species are considered in combination in non-parametric methods of display and testing, which make no distributional assumptions at all about the individual counts.

MULTIVARIATE CASE

There being no necessity to transform to attain distributional properties, transformations play an entirely separate (but equally important) role in the clustering and ordination methods of the previous chapters, that

of defining the balance between contributions from common and rarer species in the measure of similarity of two samples.

Returning to the simple example of Chapter 2, a subset of the Loch Linnhe macrofauna data, Table 9.3 shows the effect of a 4th root transformation of these abundances on the Bray-Curtis similarities. The rank order of the similarity values is certainly changed from the untransformed case, and one way of demonstrating how dominated the latter is by the single most numerous species (*Capitella capitata*) is shown in Table 9.4. Leaving out each of the species in turn, the Bray-Curtis similarity between samples 2 and 4 fluctuates wildly when *Capitella* is omitted in the untransformed case, though changes much less dramatically under 4th root transformation, which downweights the effect of single species.

Table 9.3. Loch Linnhe macrofauna {L} subset. Untransformed and 4th root-transformed abundances for some selected species and samples (years), and the resulting Bray-Curtis similarities between samples.

Untransformed									
Sample:	1	2	3	4	Sample	1	2	3	4
Species									
<i>Echinoca.</i>	9	0	0	0	1	–			
<i>Myrioche.</i>	19	0	0	3	2	8	–		
<i>Labidopl.</i>	9	37	0	10	3	0	42	–	
<i>Amaeana</i>	0	12	144	9	4	39	21	4	–
<i>Capitella</i>	0	128	344	2					
<i>Mytilus</i>	0	0	0	0					
$\sqrt[4]{}$ -transformed									
Sample:	1	2	3	4	Sample	1	2	3	4
Species									
<i>Echinoca.</i>	1.7	0	0	0	1	–			
<i>Myrioche.</i>	2.1	0	0	1.3	2	26	–		
<i>Labidopl.</i>	1.7	2.5	0	1.8	3	0	68	–	
<i>Amaeana</i>	0	1.9	3.5	1.7	4	52	68	42	–
<i>Capitella</i>	0	3.4	4.3	1.2					
<i>Mytilus</i>	0	0	0	0					

Transformation sequence

The previous remarks about the family of power transformations apply equally here: they provide a continuum of effect from $\lambda = 1$ (no transform), for which only the common species contribute to the similarity, through $\lambda = 0.5$ (square root), which allows the intermediate abundance species to play a part, to $\lambda = 0.25$ (4th root), which takes some account also of rarer species. As noted earlier, $\lambda \rightarrow 0$ can be thought of as equivalent

to the $\log_e(y)$ transformation and the latter would therefore be more severe than the 4th root 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.[†]

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	54	21
$\sqrt[4]{}$ -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 (i.e. producing a 1 in all columns) clearly cannot discriminate

[†] 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.

between the samples in any way, and therefore 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	Sample	1	2	3	4
Species					1	–			
<i>Echinoca.</i>	1	0	0	0	2	33	–		
<i>Myrioche.</i>	1	0	0	1	3	0	80	–	
<i>Labidopl.</i>	1	1	0	1	4	57	86	67	–
<i>Amaeana</i>	0	1	1	1					
<i>Capitella</i>	0	1	1	1					
<i>Mytilus</i>	0	0	0	0					

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 the 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 biomass) vary between samples: in that case an ordination on untransformed data will be little different from a simple scatter plot of the counts for the two main species.

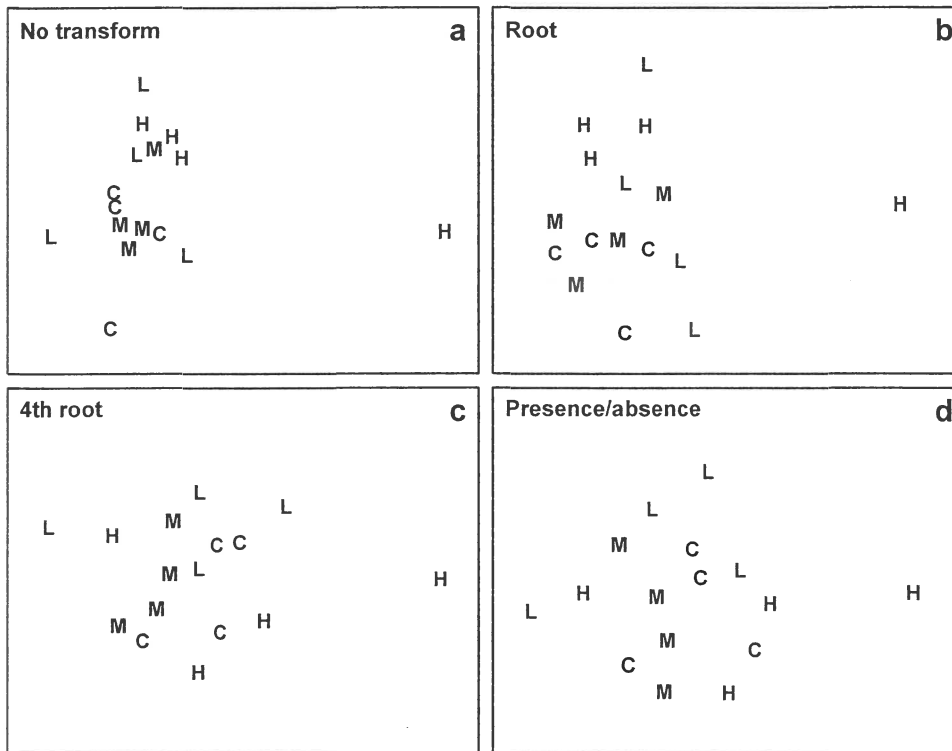
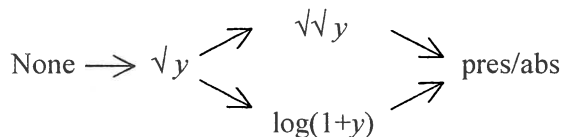


Fig 9.1 Dosing experiment, Solbergstrand {D}. MDS of nematode communities in four replicates from each of four treatments (C = control, L = low, M = medium, H = high dose of a hydrocarbon/copper contaminant mixture dosed to mesocosm basins), based on Bray-Curtis similarities from transformed data: a) no transform (stress = 0.08), b) $\sqrt{\text{stress}} = 0.14$, c) $\sqrt[4]{\text{stress}} = 0.18$, d) presence/absence (stress = 0.19).

CONCLUSIONS

- 1) The transformation sequence in a multivariate analysis, corresponding to a progressive downweighting of the common species, is effectively:



The choice of transformation from this sequence *can* affect the conclusions of an analysis (in fact, it may sometimes have more effect than choice of similarity coefficient or ordination method). In many respects the choice of transformation is more a *biological* than a statistical question: which “view” of the community do we wish to take (*shallow* or *deep*), given that there are potentially many different 2-dimensional summaries of this high-dimensional data?

- 2) Statistical considerations do enter, however, particularly in relation to the reliability of sampling. The numerical (and especially the biomass) dominants may be highly spatially variable. With untransformed data, a biomass MDS can be distorted by a single chance capture of a very large-bodied species, as can an abundance MDS by the capture of larvae or opportunist colonisers with a strong degree of spatial

clustering, such that replicate samples at the same location give counts varying from 10’s to 1000’s. At the other extreme, an analysis which places a lot of weight on species which are typically only ever found as single individuals in one sample is highly susceptible to the arbitrary “noise” introduced by the chance capture of rarer species.

- 3) The practical choice is therefore often between a moderate ($\sqrt{\text{y}}$) and rather severe ($\sqrt[4]{\text{y}}$ or \log) transform, retaining the hard-won quantitative information but downplaying the species dominants. Note that the 4th root transform is largely equivalent to reducing the original data to about a 6 or 7 point scale: 0 = absent, 1 = one individual, 2 = handful, 3 = sizeable number, 4 = abundant, ≥ 5 = very abundant. Rounding the transformed counts to this discrete scale will, in many cases, make rather little difference to the multivariate ordination (though this would not be the case for some of the univariate and graphical methods of Chapter 8). The scale may appear crude but is not unrealistic; individual species counts are often highly variable in (genuine) replicate sampling at one location and, if the primary requirement is a multivariate description, effort expended in deriving precise counts from a single sample could often be better spent in analysing more samples, to a less exacting level of detail. This is also a central theme of the following chapter.

CHAPTER 10: SPECIES REMOVAL AND AGGREGATION

SPECIES REMOVAL

For some univariate and graphical 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) ordination.* The number of species will sometimes need to be reduced, to avoid problems with computing eigenvalues (see Chapter 4).[¶]
- b) *Species analysis (by clustering or MDS ordination).* Although it is not difficult to construct dendrograms and also possible to compute MDS configurations for all species, the rarer species, whose occurrence at a particular station may largely be due to chance, must be excluded for any chance of an interpretable outcome (see Chapter 7).

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 between 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

[¶] As discussed in Chapter 4, PCA is not normally recommended for species data though, if needed, PRIMER can perform it straightforwardly. The main use of the PCA routine in PRIMER, however, is for multivariate analysis of environmental variables, see the next chapter. It is important to re-emphasize that sample MDS ordinations do not require the arbitrary initial step of eliminating some species, and this is one of the strengths of MDS: it down-weights the contributions from rare species in a natural way, without the need for initial removals.

to reduce species to the required number; typically $p = 3$ or 4).[†]

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 groups of 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 an arbitrarily 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 arbitrarily 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, it appears likely that there will be considerable *redundancy* in the species which characterise the composition of the community, that is, 'explain' the observed multivariate pattern of samples. For example, not only might it be possible to find a small subset of species which reproduces nearly all the information contained in the full analysis but there might be several such (mutually exclusive) subsets. Detection of such *structural redundancy* at the species level is a topic pursued in Chapter 16, but another way in which this idea can be exploited is considered here, 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-

[†] Or automatically in the PRIMER package, under the Edit option that selects variables, by specifying the number of 'most important' species (n) to retain, and letting the routine compute an appropriate p to achieve this.

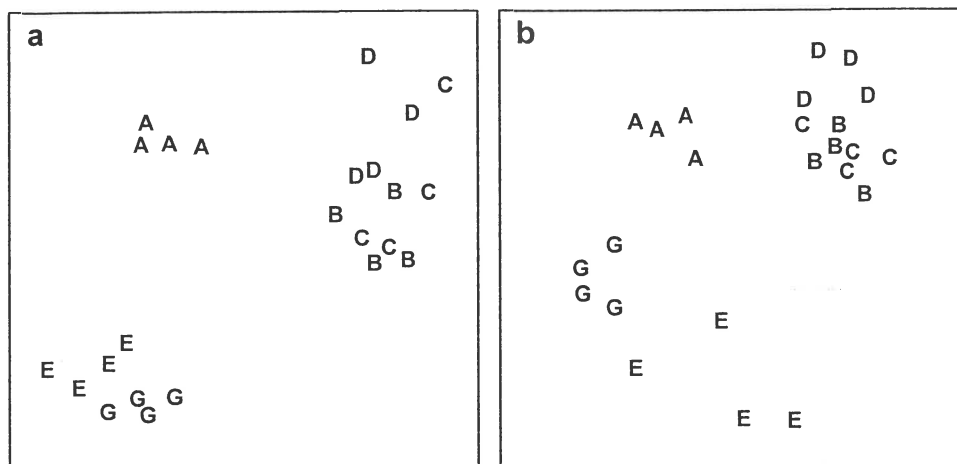


Fig. 10.1 Frierfjord macrofauna {F}. Sample MDS using Bray-Curtis similarities on \sqrt{x} transformed counts for: a) all 110 species, b) 19 arbitrarily selected species (stress = 0.14, 0.13 respectively).

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, rather than to species. If results from identifications to higher taxonomic levels are comparable to a full species analysis, this means that:

- a) A great deal of *labour can be saved*. Several groups of marine organisms are taxonomically difficult, for example (in the macrobenthos) several families of polychaetes and amphipods; as much time can be spent in separating a few of these difficult groups into species as the entire remainder of the sample, even in Northern Europe where taxonomic keys for identification are most readily available.
- b) *Less taxonomic expertise* is needed. Many taxa really require the skills of specialists to separate them into species, and this is especially true in parts of the world where fauna is poorly described. For certain groups of marine organisms, e.g. the meiobenthos, the necessary expertise required to identify even the major taxa (nematodes and copepods) to species is lacking in most laboratories which are concerned with the monitoring of marine pollution, so that these components of the biota are rarely used in such studies, despite their many inherent advantages (see Chapter 13).

For the marine macro- and meiobenthos, aggregations of the species data to higher taxonomic levels have been made[¶] and the resultant data matrices have been subjected to several forms of statistical analysis to see

[¶] This can be performed in PRIMER using the Aggregate routine, which pools the species counts in a data matrix to genus or family level, say, using an aggregation worksheet. The latter specifies the genus, family, order, class etc designations for each species, and such aggregation files are also of fundamental importance in computing biodiversity measures based on the taxonomic relatedness of species in each sample, see Chapter 17.

how much information has been lost compared with the full species-level analysis. Although such experiments have not often been done for other components of the marine biota (e.g. plankton), results from the benthic studies are remarkably clear in that very little information appears to be lost after a moderate degree of aggregation.

Methods amenable to aggregation

1) *Multivariate methods*. Although taxonomic levels higher than that of species can be used to some degree for all types of statistical analysis of community data, it is probably for multivariate methods that this is most appropriate, in view of the redundancy discussed above. All ordination/clustering techniques are amenable to aggregation, and there is now substantial empirical evidence that identification only to the family level for macrobenthos, and the genus level for meiobenthos, makes very little difference to the results (see, for example, Figs. 10.2–10.6, and more recent results described in Chapter 16). There are also certain possible theoretical advantages to conducting multivariate analyses at a high taxonomic level for pollution impact studies. Natural environmental variables which also affect community structure are rarely constant in surveys designed to detect pollution effects over relatively large geographical areas. In the case of the benthos, these 'nuisance' variables include water depth and sediment granulometry. However, it is a tenable hypothesis that these variables usually influence the fauna more by species replacement than by changes in the proportions of the major taxa present. Each major group, in its adaptive radiation, has evolved species which are suited to rather narrow ranges of natural environmental conditions, whereas the advent of pollution by man has been too recent for the evolution of suitably adapted species. Ordinations of abundance or biomass data of these major taxa are therefore

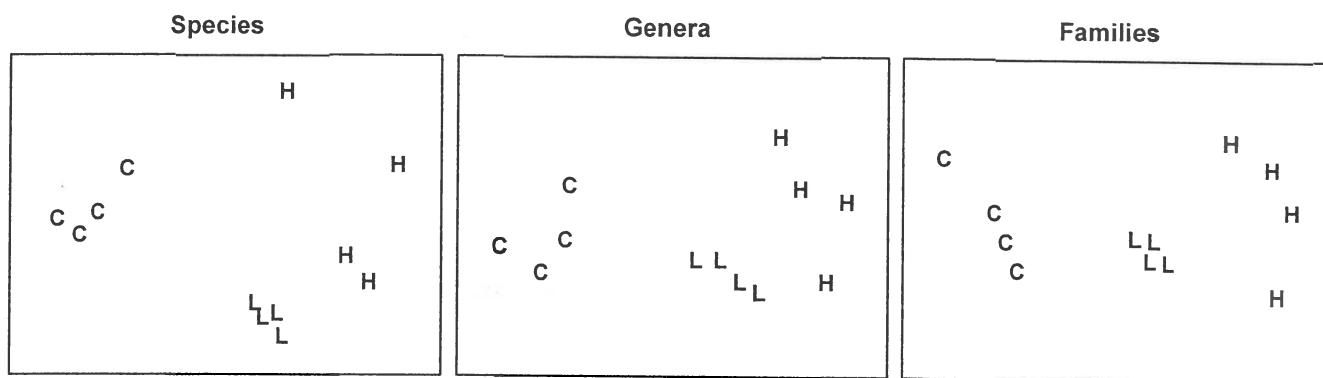


Fig. 10.2. Nutrient-enrichment experiment, Solbergstrand {N}. MDS plot of copepod abundances ($\sqrt[3]{x}$ -transformed, Bray-Curtis similarities) for four replicates from each of three treatments; species data aggregated into genera and families (stress = 0.09, 0.09, 0.08).

more likely to correlate with a contamination gradient than are species ordinations, the latter being more complicated by the effects of natural environmental variables when large heterogeneous geographical areas are considered. In short, higher taxa may well reflect well-defined pollution gradients more closely than species.

- 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.

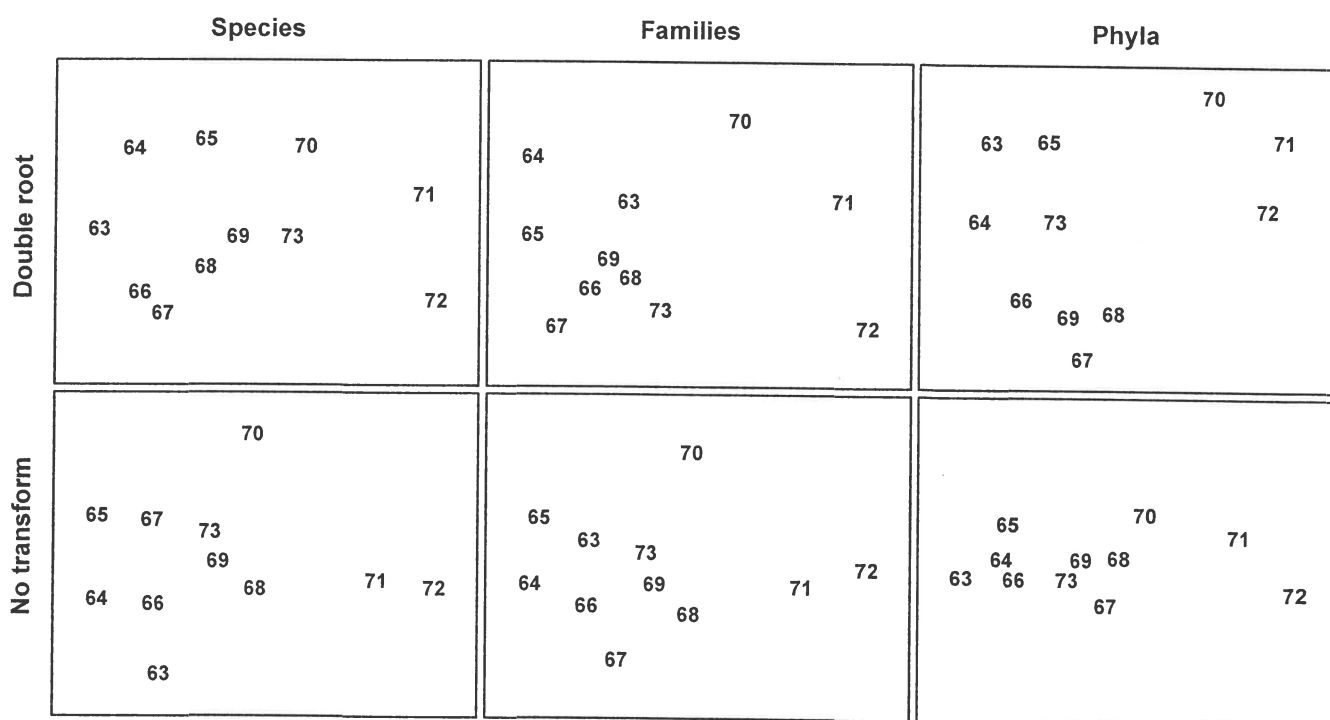


Fig. 10.3. Loch Linnhe macrofauna {L}. MDS (using Bray-Curtis similarities) of samples from 11 years. Abundances are $\sqrt[3]{x}$ -transformed (top) and untransformed (bottom), with 115 species (left), aggregated into 45 families (middle) and 9 phyla (right). (Reading across rows, stress = 0.09, 0.09, 0.10, 0.09, 0.09, 0.02).

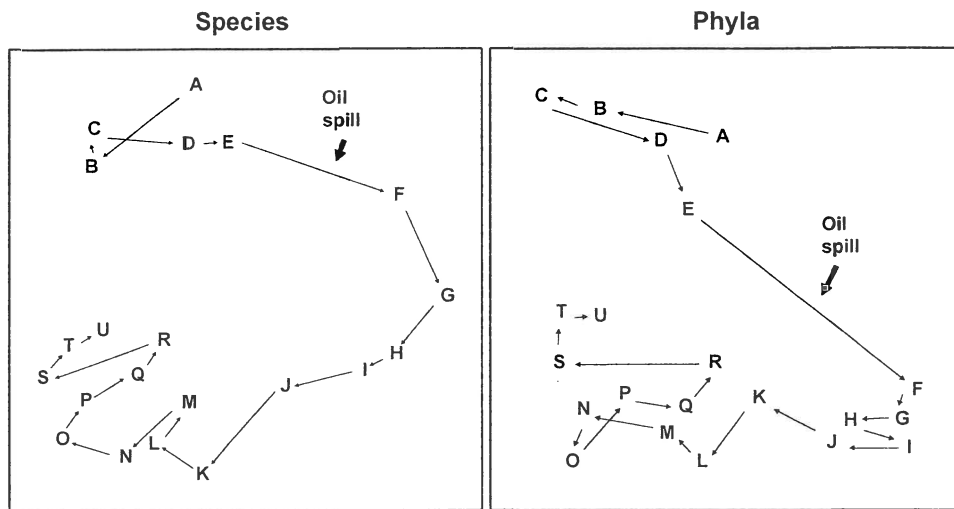


Fig. 10.4. Amoco-Cadiz oil spill {A}. MDS for macrobenthos at station "Pierre Noire" in the Bay of Morlaix. Species data (left) aggregated into phyla (right). Sampling months are A:4/77, B:8/77, C:9/77, D:12/77, E:2/78, F:4/78, G:8/78, H:11/78, I:2/79, J:5/79, K:7/79, L:10/79, M:2/80, N:4/80, O:8/80, P:10/80, Q:1/81, R:4/81, S:8/81, T:11/81, U:2/82. The oil-spill was during 3/78, (stress = 0.09, 0.07).

MULTIVARIATE EXAMPLES

Nutrient-enrichment experiment

In the soft-bottom mesocosms at Solbergstrand, Norway {N}, box-cores of sublittoral sediment were subjected to three levels of particulate organic enrichment (L = low dose, H = high dose and C = control), there being four replicates from each treatment. After 56 days the meiobenthic communities were analysed. Fig 10.2 shows that, for the copepods, there were clear differences in community structure between treatments at the species level, which were equally evident when the species data were aggregated into genera and families.

(Indeed, at the family level the configuration is arguably more linearly related to the pollution gradient than at the species level).

Loch Linnhe macrofauna

MDS ordinations of the Loch Linnhe macrobenthos are given in Fig. 10.3, using both double square root and untransformed abundance data. Information on the time-course of pollution events and changes in diversity are given in Fig. 10.7 (top left). The ordinations have been performed separately using all 115 species, the 45 families and the 9 phyla. In all ordinations there is a separation to the right of the years 1970, 1971 and 1972 associated with increasing pollution levels and

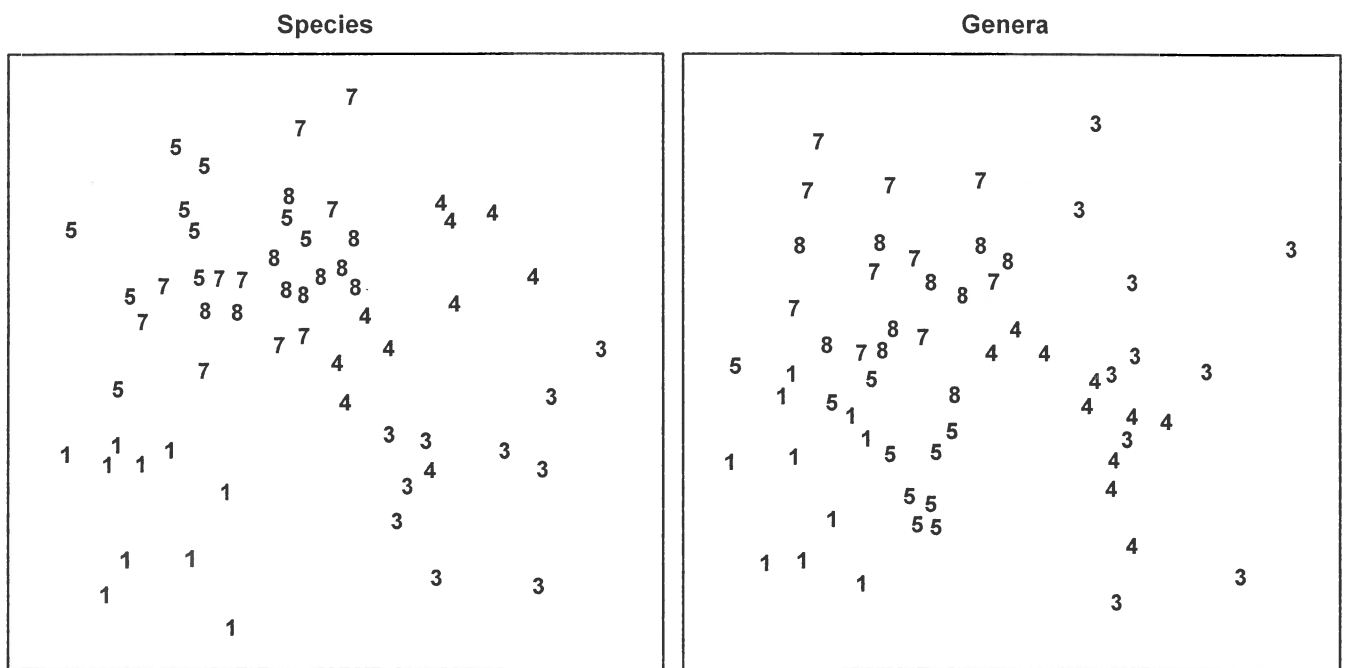


Fig. 10.5. Indonesian reef corals {I}. MDS for species (p=75) and genus (p=24) data at South Pari Island (Bray-Curtis similarities on untransformed % cover). The El Niño occurred in 1982-3. 1=1981, 3=1983 etc. (stress = 0.25).

community stress, and a return to the left in 1973 associated with reduced pollution levels and community stress. 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 February 1982, spanning the period of the wreck of the 'Amoco-Cadiz' in March 1978. The sampling site was some 40km from the initial tanker disaster but substantial coastal oil slicks resulted.

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 response at the phylum level is remarkably clear.

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 return towards the pre-El Niño condition. This is closely reflected in the genus level analysis.

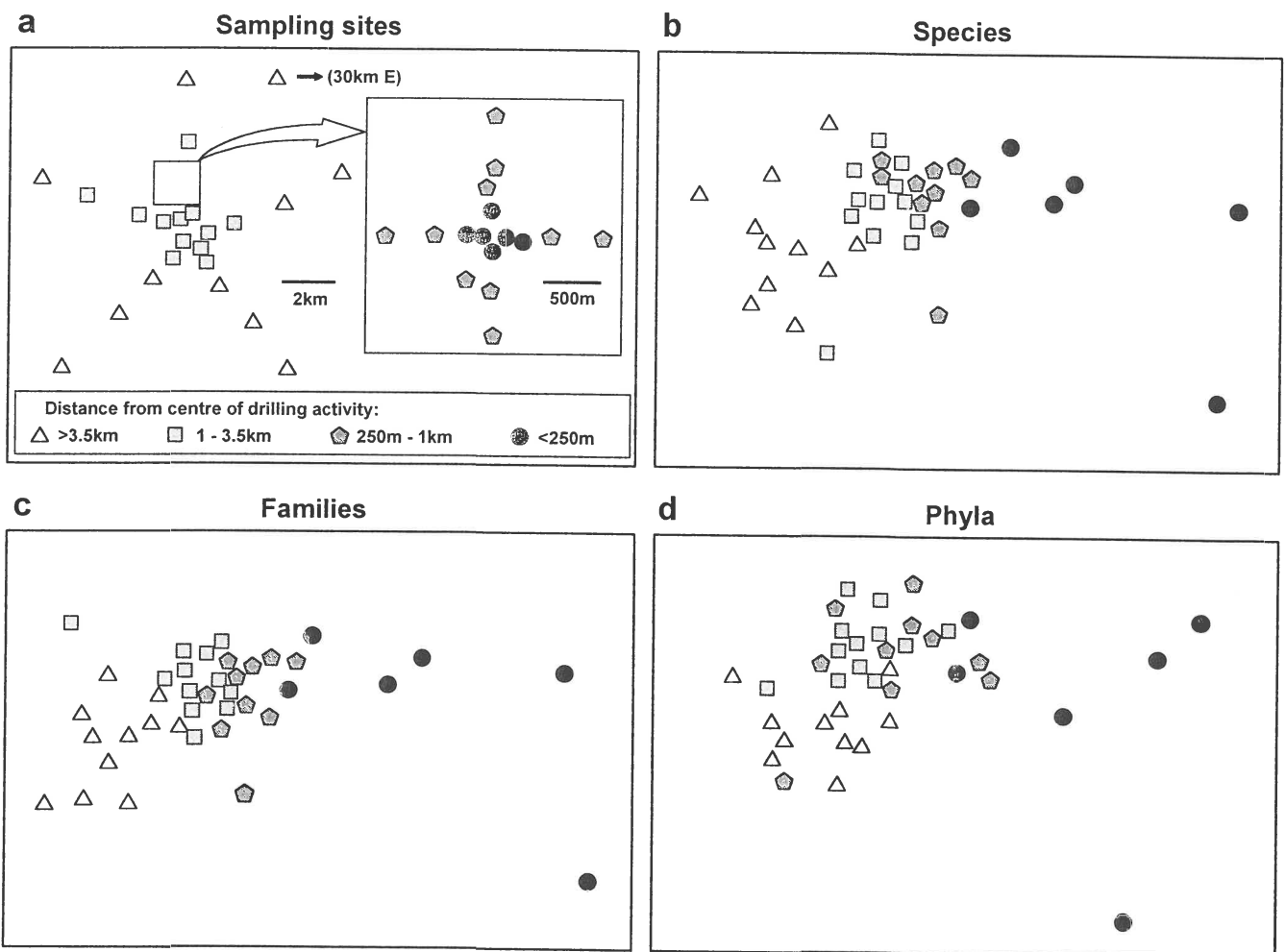


Fig. 10.6. Ekofisk oil-platform macrobenthos {E}. a) Map of station positions, indicating symbol/shading conventions for distance zones from the centre of drilling activity; b)-d) MDS for root-transformed species, family and phyla abundances respectively (stress = 0.12, 0.11, 0.13).

Ekofisk oil-platform macrobenthos

Changes in community structure of the soft-bottom benthic macrofauna in relation to oil drilling activity at the Ekofisk platform in the North Sea {E} have been studied by Gray *et al* (1990), Warwick and Clarke (1991). The positions of the 39 sampling stations around the rig are coded by different symbol and shading conventions in Fig. 10.6a, according to their distance from the centre of drilling activity at that time. In the MDS species abundance analysis (Fig. 10.6b), community composition in all of the zones is distinct, and there is a clear gradation of change from the (black circle) inner to the (open triangle) outer zones. Formal significance testing (using the methods

of Chapter 6) confirms statistically the differences between all zones. The MDS has been repeated with the species data aggregated into families (Fig. 10.6c) and phyla (Fig. 10.6d). The separation of sites is still clear, and pairwise comparisons confirm the statistical significance of differences between all zones, even at the phylum level, which does show some deterioration of the pattern. This is in contrast to (species-level) univariate and graphical/distributional measures, in which only the inner zone (less than 250m from the rig) was significantly different from the other three zones (see Chapter 14). Thus, phylum level analyses are again shown to be surprisingly sensitive in detecting pollution-induced community change, and little information at all is lost by working at the family level.

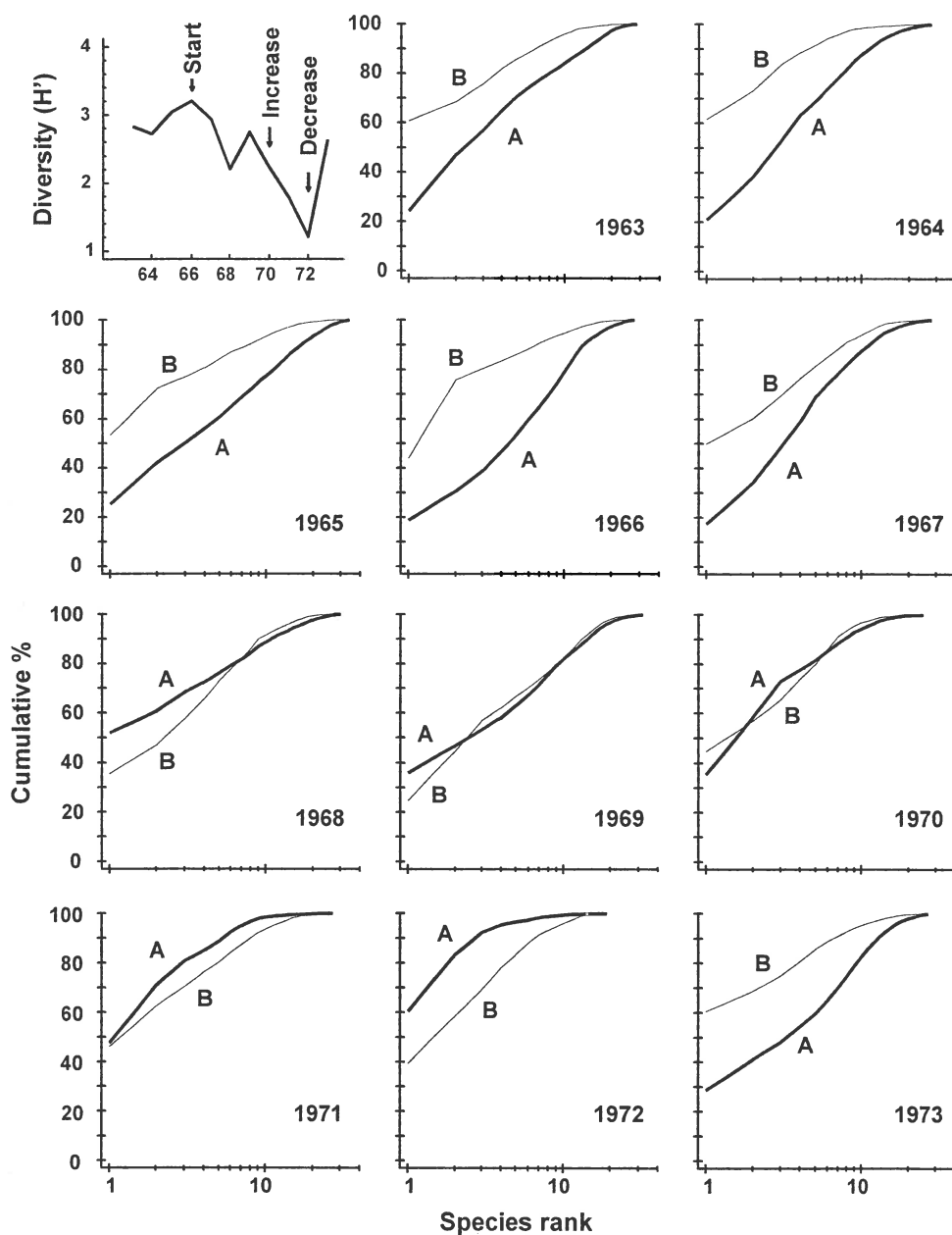


Fig. 10.7. Loch Linnhe macrofauna {L}. Shannon diversity (H') and ABC plots over the 11 years, 1963 to 1973, for data aggregated to family level (c.f. Fig. 8.7). Abundance = thick line, biomass = thin line.

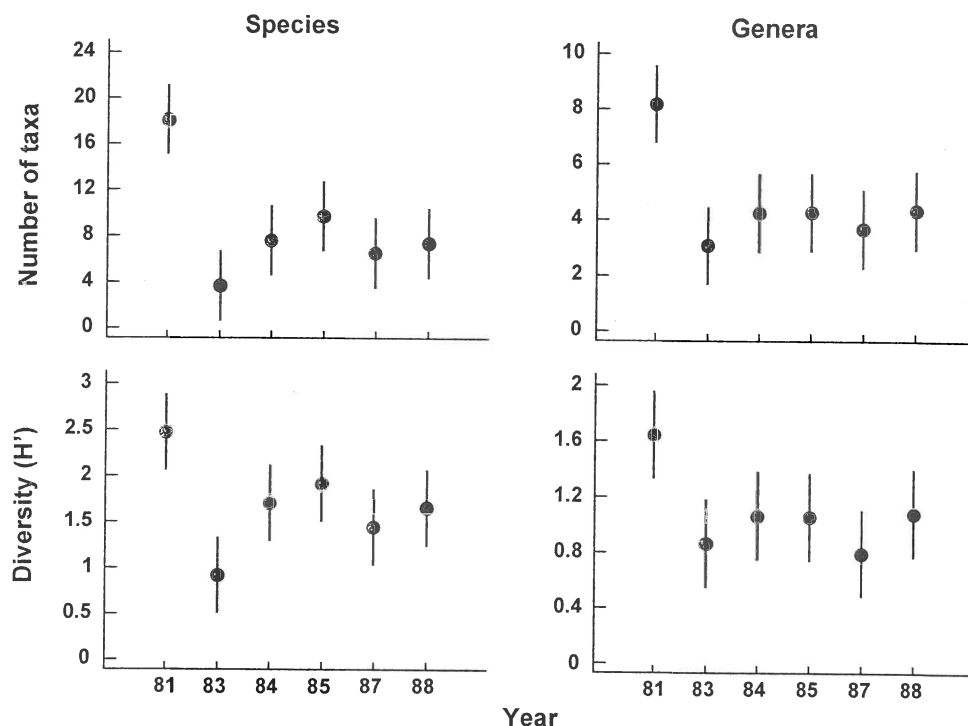


Fig. 10.8. Indonesian reef corals {I}. Means and 95% confidence intervals for number of taxa and Shannon diversity at South Tikus Island, showing the impact and partial recovery from the 1982-3 El Niño. Species data (left) have been aggregated into genera (right).

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 general recommendations and each case must be treated on its individual merits. However, for routine monitoring of organic enrichment situations using macrobenthos, one can by now be rather certain that family level analysis will be perfectly adequate. Similarly, for meiofaunal taxa, there are by now many examples where multivariate analysis of genus-level information is indistinguishable from that for species.[¶] For other components of the marine fauna, and observational studies not concerned with detection of organic enrichment impacts, the body of evidence supporting a particular choice of taxonomic level is much smaller.

[¶] See also Chapter 16, which extends the ideas of this chapter by providing the tools for quantifying relative change in the information a multivariate analysis provides, as the taxonomic level (or indeed the transformation) is varied.

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.[¶] 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

[¶] 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.

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.[†] The appropriate (rank) dissimilarity matrix would use normalised 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 variables appearing as columns and the samples as rows.[‡]

[†] 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' *A* (e.g. Mardia *et al.*, 1979) are valid.

[‡] This is in contrast with abundance matrices which, because of their generally larger number of variables (species) are usually transposed, i.e. the samples are displayed as columns. The PRIMER software package can handle data entered either way round, though in reading in earlier “.pm1” PRIMER files it assumes biotic matrices are species-by-samples and environmental ones are samples-by-variables.

Table 11.1. Garroch Head dump ground {G}. Sediment metal concentrations (ppm), water depth at the site (m) and organic loading of the sediment (% carbon and nitrogen), for the transect of 12 stations across the sewage-sludge dump site (centre at station 6), see Fig. 8.3.

Station	Cu	Mn	Co	Ni	Zn	Cd	Pb	Cr	Dep	%C	%N
1	26	2470	14	34	160	0	70	53	144	3	0.53
2	30	1170	15	32	156	0.2	59	15	152	3	0.46
3	37	394	12	38	182	0.2	81	77	140	2.9	0.36
4	74	349	12	41	227	0.5	97	113	106	3.7	0.46
5	115	317	10	37	329	2.2	137	177	112	5.6	0.69
6	344	221	10	37	652	5.7	319	314	82	11.2	1.07
7	194	257	11	34	425	3.7	175	227	74	7.1	0.72
8	127	246	10	33	292	2.2	130	182	70	6.8	0.58
9	36	194	6	16	89	0.4	42	57	64	1.9	0.29
10	30	326	11	26	108	0.1	44	52	80	3.2	0.38
11	24	439	12	34	119	0.1	58	36	83	2.1	0.35
12	22	801	12	33	118	0	52	51	83	2.3	0.45

No replication is available for the 12 stations so the variance-to-mean plots suggested in Chapter 9 are not possible, but simple scatter plots of all pairwise combinations of variables (*draftsman plots*, see the later Fig. 11.9) suggest that log transformations are appropriate for the concentration variables, though not for water depth. The criteria here are that variables should not show marked skewness across the samples, enabling meaningful normalisation, and that the relationships between them should be approximately linear; the standard product-moment correlations between variables and Euclidean distances between samples are then satisfactory summaries. In pursuit of this, note that whilst each variable could in theory be subjected to a different transformation it is more logical to apply the same transformation to all variables of the same type. Thus the decision to log all the metal data stems not just from the draftsman plots but also from previous experience that such concentration variables often have standard deviations proportional to their means; i.e. a roughly constant *percentage* variation is log transformed to a stable *absolute* variance.

Fig. 11.1 displays the first two axes (PC1 and PC2) of a PCA ordination on the transformed data of Table 11.1. In fact, the first component accounts for much of the variability (61%) in the full matrix, the first two components accounting for 88%, so the 2-d plot provides an accurate summary of the sample relationships. Broadly speaking, PC1 represents an axis of increasing contaminant load:

$$\begin{aligned}
 PC1 = & 0.38 Cu' - 0.22 Mn' - 0.08 Co' + 0.15Ni' \\
 & + 0.37 Zn' + 0.33 Cd' + 0.37 Pb' + 0.35 Cr' \\
 & - 0.12 Dep' + 0.37 C' + 0.33 N' \quad (11.1),
 \end{aligned}$$

since most of the sizeable coefficients are positive. The dashes denote that the variables have been log transformed (excepting Dep) and normalised to zero mean and unit standard deviation. Fig. 11.1 shows a

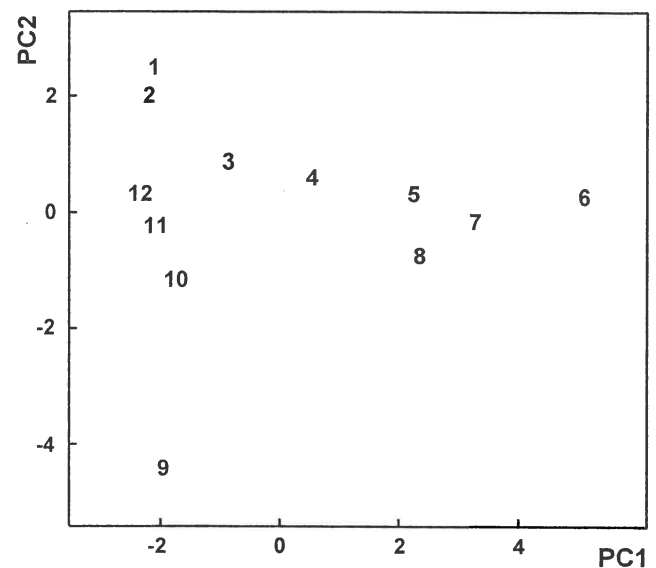


Fig. 11.1. Garroch Head dump ground {G}. Two-dimensional PCA ordination of the 11 environmental variables of Table 11.1 (transformed and normalised), for the stations (1–12) across the sewage-sludge dump site centred at station 6 (% variance explained = 88%).

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 ($\beta = -0.29$, $p < 0.1\%$).

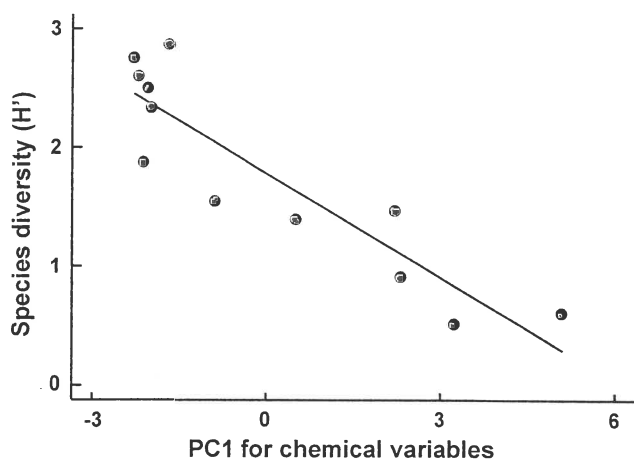


Fig. 11.2. Garroch Head macrofauna {G}. Linear regression of Shannon diversity (H'), at the 12 sampling stations, against the first PC axis score from the environmental PCA of Fig. 11.1, which broadly represents an axis of increasing contaminant load (equation 11.1).

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 in *bubble plots* by representing the values of this variable as a symbol of differing sizes and superimposing these symbols on the biotic ordination of the corresponding samples. This, or the alternative superimposition of coded values for the variable, can be an effective means of noting consistent differences in the environmental variable between biotic clusters or observing a smooth relationship with ordination gradients (Field *et al.*, 1982).[¶]

[¶] Bubble plots, superimposing environmental data onto an ordination, are a basic feature provided in the PRIMER MDS plotting routine. 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 useful insight into the role of individual taxa in shaping the biotic picture (see section 4 of the PRIMER User Manual/Tutorial), especially when the number of taxa is small, as is the case for the phylum-level "meta-analysis" of Chapter 15.

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 the application of a *reverse logarithmic* transformation, $\log(36 - s)$, or more precisely:

$$s^* = a - b \log_e(36 - s) \quad (11.2)$$

where $a = 8.33$, $b = 3$ are simple constants chosen for this data to constrain the transformed variable s^* to lie, when rounded to the nearest integer, in the range 1 (low) to 9 (high salinity).[†] Fig. 11.4 then clearly displays the strong correlation of the zooplankton community structure with the salinity gradient. It also helps to focus attention on sites which appear slightly anomalous in this respect, and raises questions about whether there are secondary environmental variables which might explain the biological differentiation of samples at similar salinities.[‡]

[†] In the PRIMER TRANSFORM routine the expression for the salinity variable is therefore: $\text{int}(0.5 + 8.33 - 3 * \log(36 - V))$, and these bubble values can then be used to label the MDS plot.

[‡] Note the *horseshoe effect* (more properly termed the *arch effect*), which is a common feature of the ordination from single, strong environmental gradients. Both theoretically and empirically, non-metric MDS would seem to be less susceptible to this than metric ordination methods, but without the drastic (and somewhat arbitrary) intervention in the plot that a technique like detrended correspondence analysis uses (specifically to "cut and paste" such ordinations to a straight line), some degree of curvature is unavoidable and natural. Where samples towards opposite ends of the environmental gradient have few species in common (thus giving dissimilarities near 100%), samples which are even further apart on the gradient have little scope to increase their dissimilarity further. To some extent, non-metric MDS can compensate for this by the flexibility of its monotonic regression of distance on dissimilarity (Chapter 5), but arching of the tails of the plot is clearly inevitable after dissimilarities of 100% are reached.

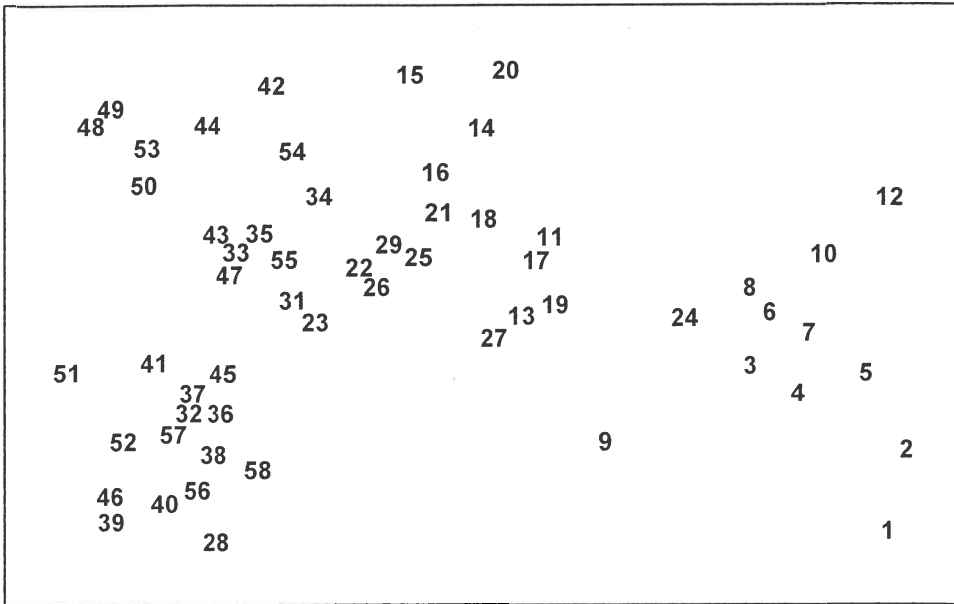


Fig. 11.3. Bristol Channel zooplankton {B}. Biotic MDS for the 57 sampling sites (1-29, 31-58) mapped in Fig. 3.2, from the same Bray-Curtis similarities on \sqrt{V} -transformed abundances used for the cluster analysis of Fig. 3.3 (stress = 0.11).

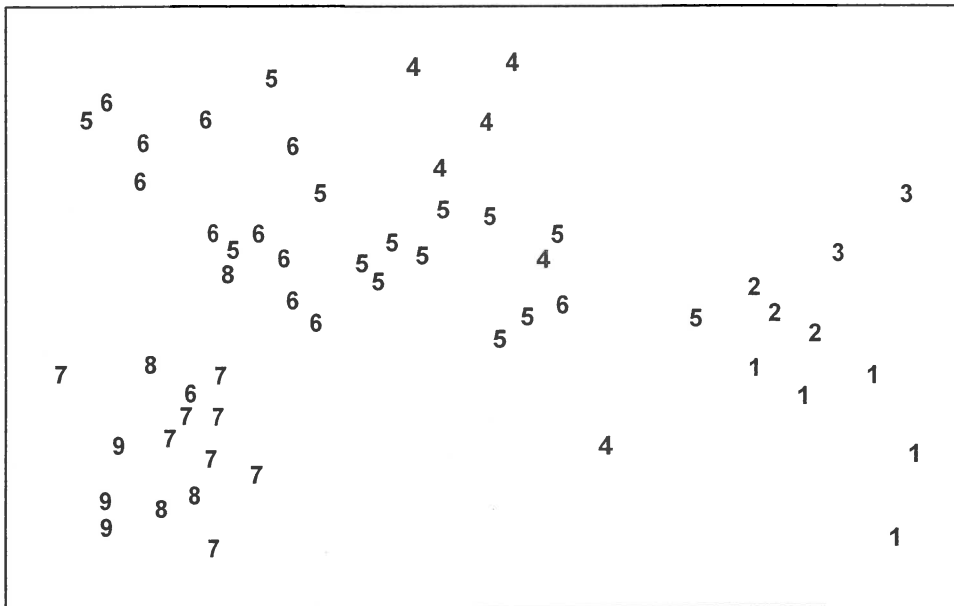


Fig. 11.4. Bristol Channel zooplankton {B}. MDS of Fig. 11.3, with superimposed codes representing increasing salinity levels at the 57 sites; 1: <26.3, 2: (26.3, 29.0), 3: (29.0, 31.0), ..., 8: (34.7, 35.1), 9: >35.1 ppt.

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).[¶] 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 plot 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 discussed above: Fig. 11.5b displays the values of % carbon in the sediment (Tables 11.1) as circles of varying diameter, which confirms the main axis of the biotic MDS as one of the increasing organic enrichment. Several of the metal concentrations from Table 11.1 show a similar

[¶] 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).

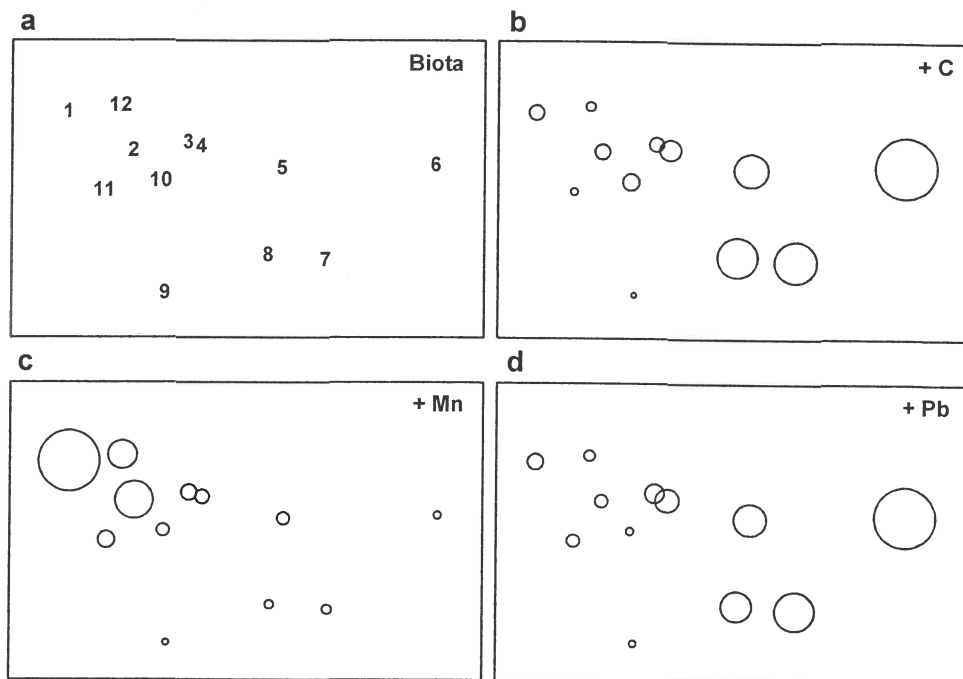


Fig. 11.5. Garroch Head macrofauna {G}. a) MDS of Bray-Curtis similarities from $\sqrt{\text{transformed}}$ species biomass data at the 12 stations (Fig. 8.3); b)-d) the same MDS but with superimposed circles of increasing size with increasing sediment concentration of C, Mn and Pb, from Table 11.1.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

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*).

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 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, as bubble plots, some instructive patterns emerge. MPD (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. Some other variables, such as the height up the shore (Fig. 11.6d), appear to bear little relation to the overall biotic structure, in that samples within the same faunal groups are frequently at opposite extremes of the intertidal range.

These plots, however, make clear the limitations in relating the community structure to a single environmental variable at a time: there is no basis for answering questions such as “how well does the full set of abiotic data jointly explain the observed biotic pattern” and “is there a subset of the environmental variables that explains the pattern equally well, or better?” These questions are answered in classical multivariate statistics by techniques such as canonical correlation (e.g. Mardia *et al*, 1979) but, as discussed in earlier chapters, this requires assumptions which are unrealistic for species abundance or biomass data (correlation and Euclidean distance as measures of similarity for biotic data, linear relationships between abundance and environmental gradients etc).

Instead, the need is to relate community structure to multivariate descriptions of the abiotic variables, using the type of non-parametric, similarity-based methods of previous chapters.

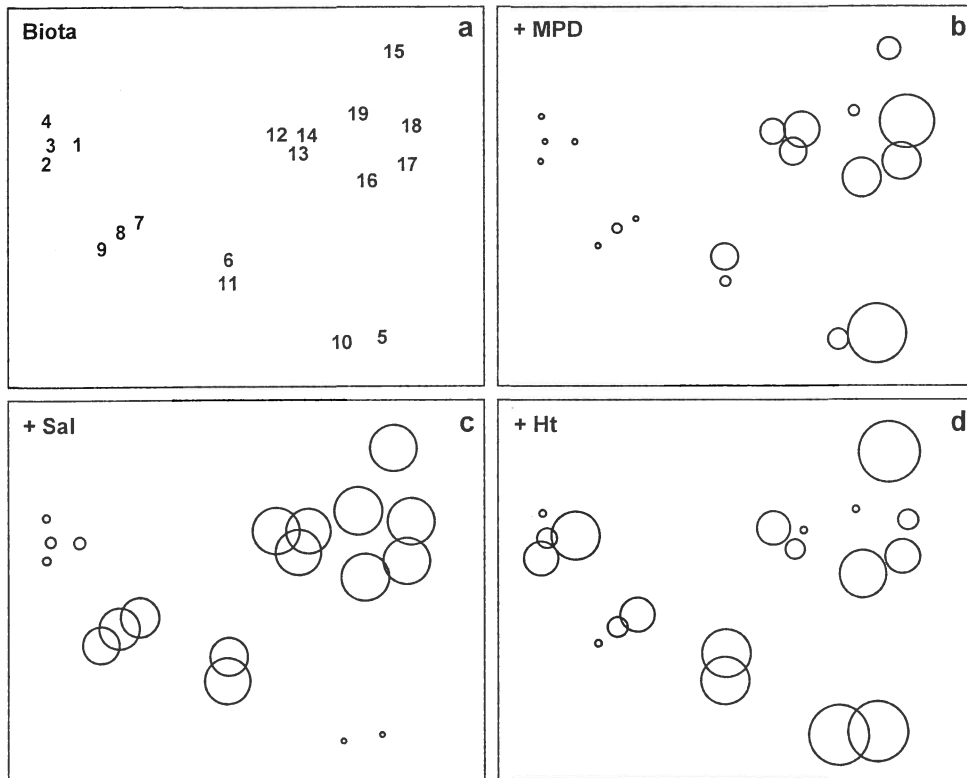


Fig. 11.6. Exe estuary nematodes {X}. a) MDS of species abundances at the 19 sites, as in Fig. 5.1; b)-d) the same MDS but with superimposed circles representing, respectively, median particle diameter of the sediment, its interstitial salinity and height up the shore of the sampling locations. (Stress = 0.05).

LINKING BIOTA TO MULTIVARIATE ENVIRONMENTAL PATTERNS

The intuitive premise adopted here is that if the suite of environmental variables responsible for structuring the community were known[¶], then samples having rather similar values for these variables would be expected to have rather similar species composition, and an ordination based on this abiotic information would group sites in the same way as for the biotic plot. If key environmental variables are omitted, the match between the two plots will deteriorate. By the same token, the match will also worsen if abiotic data which are irrelevant to the community structure are included.

The Exe estuary nematode data {X} again provides an appropriate example. Fig. 11.7a repeats the species MDS for the 19 sites seen in Fig. 11.6a. The remaining

[¶] These might sometimes include biotic as well as abiotic data, e.g. when assessing how sediment meiofaunal communities might be affected by densities of a structuring (critical) macrofaunal species. There is an implicit assumption here, of course, that the observed sample patterns are not dominated by internal stochastic forces, e.g. competitive interactions within the assemblage constituting the biotic data matrix. If they are, the procedure will fail to “explain” the community structure in terms of the provided set of environmental variables, naturally.

plots in Fig. 11.7 are of specific combinations of the six sediment variables: H₂S, Sal, MPD, %Org, WT and Ht, as defined above. For consistency of presentation, these plots are also MDS ordinations but based on an appropriate dissimilarity matrix (Euclidean distance on the normalised abiotic variables). In practice, since the number of variables is small, and the distance measures the same, the MDS plots will be largely indistinguishable from PCA configurations (note that 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, especially 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.

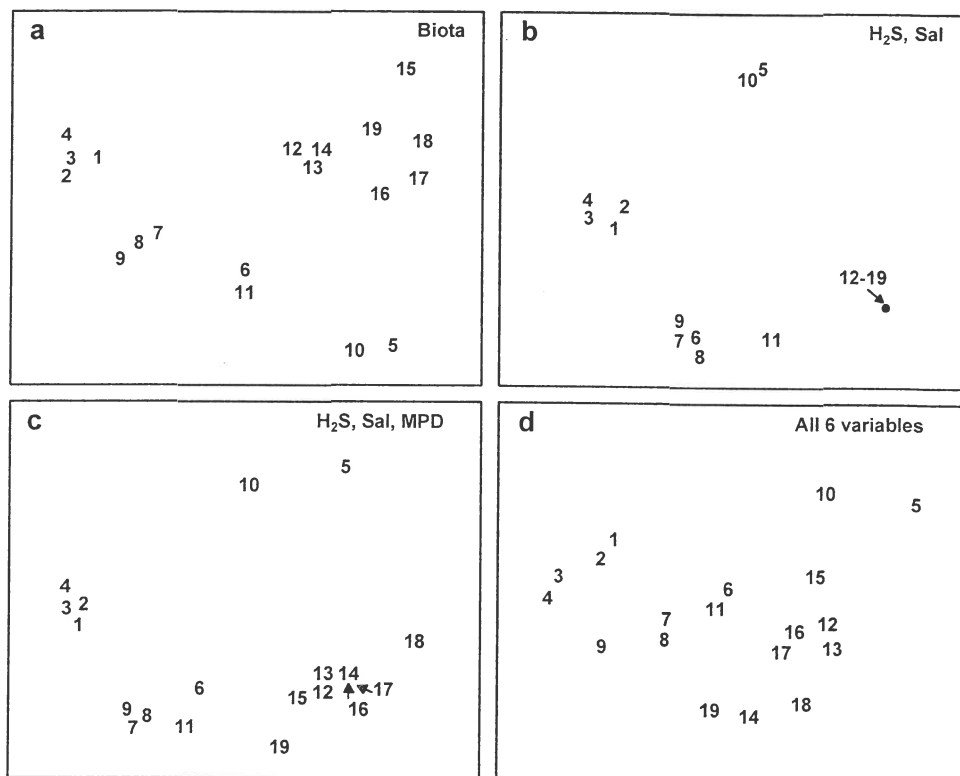


Fig. 11.7. *Exe estuary nematodes* {X}. MDS ordinations of the 19 sites, based on: a) species abundances, as in Fig. 5.1; b) two sediment variables, depth of the H₂S layer and interstitial salinity; c) the environmental combination “best matching” the biotic pattern: H₂S, salinity and median particle diameter; d) all six abiotic variables. (Stress = 0.05, 0, 0.04, 0.06).

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.[†] These are chosen differently to match the respective form of the data (i.e. 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.

[†] For example, in spite of the very low stress in Fig. 11.7, a 2-d Procrustes fit of 11.7a with 11.7c 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). This match will probably be better in 3-d but will be fully expressed, without arbitrary dimensionality constraints, in the underlying similarity matrices.

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

$$\rho_s = 1 - \frac{6}{N(N^2 - 1)} \sum_{i=1}^N (r_i - s_i)^2 \quad (11.3)$$

A standard alternative is Kendall's τ (Kendall, 1970) which, in practice, tends to give rather similar results to ρ_s . The third possibility is a modified form of Spearman, the **weighted Spearman** (or **harmonic**[‡]) rank correlation:

$$\rho_w = 1 - \frac{6}{N(N-1)} \sum_{i=1}^N \frac{(r_i - s_i)^2}{r_i + s_i} \quad (11.4)$$

[‡] This is so defined by Clarke and Ainsworth (1993) because it is algebraically related to the average of the harmonic mean of each (r_i, s_i) pair. The denominator term, $r_i + s_i$, down-weights the contribution of large ranks; these are the low similarities, the highest similarity corresponding to the lowest value of rank similarity (1), as usual. Note that ρ_w and τ tend to give consistently lower values than ρ_s for the same match; nothing should therefore be inferred from a comparison of absolute values of ρ_s , τ and ρ_w .

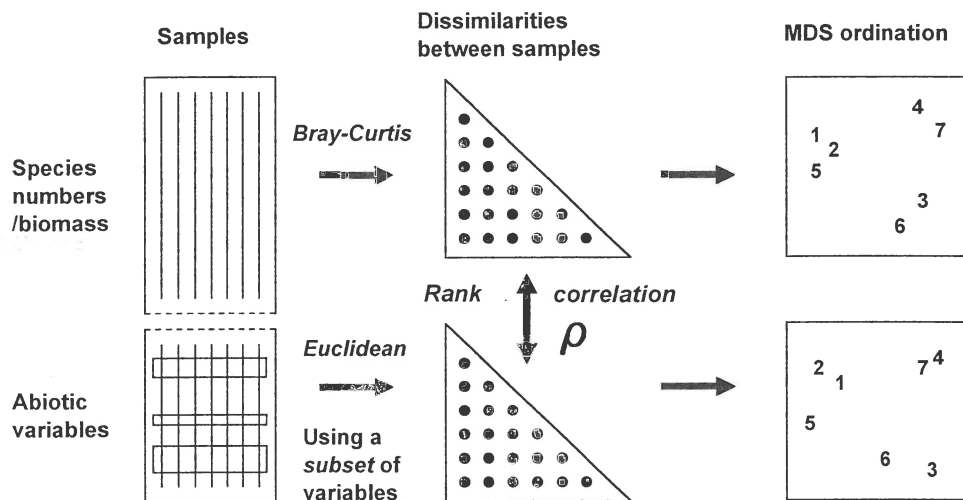


Fig. 11.8. Schematic diagram of the BIO-ENV procedure: selection of the abiotic variable subset maximising rank correlation (ρ) between biotic and abiotic (dis)similarity matrices.

The constant terms are defined such that, in both (11.3) and (11.4), ρ lies in the range $(-1, 1)$, with the extremes of $\rho = -1$ and $+1$ corresponding to the cases where the two sets of ranks are in complete opposition or complete agreement, though the former is unlikely to be attainable in practice because of the constraints inherent in a similarity matrix. Values of ρ around zero correspond to the absence of *any* match between the two patterns, but typically ρ will be positive. It is tempting, but wholly wrong, to refer ρ_s to standard statistical tables of Spearman's rank correlation, to assess whether two patterns are "significantly" matched ($\rho > 0$). This is invalid because the ranks $\{r_i\}$ (or $\{s_i\}$) are not mutually independent variables, since they are based on a large number (N) of strongly interdependent similarity calculations.

In itself, this does not compromise the use of ρ_s as an index of agreement of the two triangular matrices. However, it could be less than ideal because few of the equally-weighted difference terms in equation (11.3) involve "nearby" samples. In contrast, the premise at the beginning of this section makes it clear that we are seeking a combination of environmental variables which attains a good match of the *high* similarities (low ranks) in the biotic and abiotic matrices. The value of ρ_s , when computed from triangular similarity matrices, will tend to be swamped by the larger number of terms involving distant pairs of samples, contributing large squared differences in (11.3). This motivates the down-weighting denominator term in (11.4). However, experience suggests that, typically, this modification affects the outcome only marginally and, in the interests of simplicity of explanation, the well-known Spearman coefficient may be preferred.

The BIO-ENV procedure

The matching of biotic to environmental patterns can now take place[¶], as outlined schematically in Fig. 11.8. Combinations of the environmental variables are considered at steadily increasing levels of complexity, i.e. k variables at a time ($k = 1, 2, 3, \dots, v$). Table 11.2 displays the outcome for the Exe estuary nematodes.

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 ₂ S (.62)	%Org (.54)	Sal (.53)	...
2	H ₂ S, Sal (.76)	H ₂ S, MPD (.67)	H ₂ S, %Org (.65)	Sal, %Org ... (.61)
3	H₂S, Sal, MPD (.80)	H ₂ S, Sal, %Org (.75)	H ₂ S, Sal, WT ... (.73)	
4	H ₂ S, Sal, MPD, %Org (.79)	H ₂ S, Sal, MPD, Ht ... (.78)		
5	H ₂ S, Sal, MPD, %Org, Ht ... (.79)			
6	H ₂ S, Sal, MPD, %Org, Ht, WT (.77)			

The *single* abiotic variable which best groups the sites, in a manner consistent with the faunal patterns, is the depth of the H₂S layer ($\rho_w = 0.62$); next best is the organic content ($\rho_w = 0.54$), etc. Of course, since the faunal ordination is not essentially 1-dimensional

[¶] This is implemented in the PRIMER BIOENV routine, which performs a full search, up to a fixed number of variable combinations.

(Fig. 11.7a), it would not be expected that a single environmental variable would provide a very successful match, though knowledge of the H₂S variable alone does distinguish points to the left and right of Fig. 11.7a (samples 1 to 4 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 H₂S 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.[†] 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.

An exhaustive search over v variables involves

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

combinations, i.e. 63 for the Exe estuary study, though this number quickly becomes prohibitive when v is larger than about 15. 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 (this is implemented in the BVSTEP procedure of Chapter 16). In practice though, 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.

[†] 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 on the interpretation. Experience suggests that combinations giving the same value of ρ to two decimal places do not give rise to ordinations which are distinguishable in any practically important way, thus it is recommended that ρ is quoted only to this accuracy, as in Table 11.2.

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). As indicated earlier, choice of transformations is aided by a *draftsman plot*, i.e. scatter plots of all pairwise combinations of variables, Fig. 11.9. Here, this is after all the concentration variables, but not water depth, have been log transformed[‡], in line with the recommendations on page 11-2.

The draftsman plot, and the associated correlation matrix between all pairs of variables, can then be examined for evidence of collinearity (page 11-5), indicated by straight-line relationships, with little scatter, in Fig. 11.9. A further rule-of-thumb would be to reduce all subsets of (transformed) variables which have mutual correlations averaging more than about 0.95 to a single representative. This suggests that C, Cu, Zn and Pb are so highly inter-correlated that it would serve no useful purpose to leave them all in the BIO-ENV analysis. For every good match that included C, there would be equally good matches including Cu, Zn or Pb, leading to a plethora of effectively identical solutions. Here, the organic carbon load (C) is retained and the other three excluded, leaving 8 abiotic variables in the full BIO-ENV search. This results in an optimal match of the biotic pattern with C, N and Cd ($\rho_w = 0.78$). The corresponding ordination plots are seen in Fig. 11.10. The biotic MDS of Fig. 11.10a, though structured mainly by a single strong gradient towards the dump centre (e.g. the organic enrichment gradient seen in Fig. 11.10b), is not wholly 1-dimensional. Additional information, on a heavy metal, appears to improve the “explanation”.

CONCLUDING REMARKS

Further examples of the BIO-ENV procedure are given in Clarke and Ainsworth (1993), Clarke (1993), Somerfield *et al* (1994) and many subsequent applications. For a series of data sets on impacts on benthic macrofauna around N Sea oil rigs, Olsgard *et al* (1997, 1998) use the BIO-ENV procedure in a particularly interesting way. They examine which transformations (Chapter 9) and what level of taxonomic aggregation (Chapter 10) tend to maximise the BIO-ENV correlation, ρ . The hypotheses examined are that certain

[‡] This actually uses a $\log(c+x)$ transformation where c is a constant such as 1 or 0.1. The necessity for this, rather than a simple $\log(x)$ transform, comes from the zero values for the Cd concentrations in Table 11.1, $\log(0)$ being undefined. A useful rule-of-thumb here is to set the constant c to the lowest non-zero measurement, or the concentration detection limit.

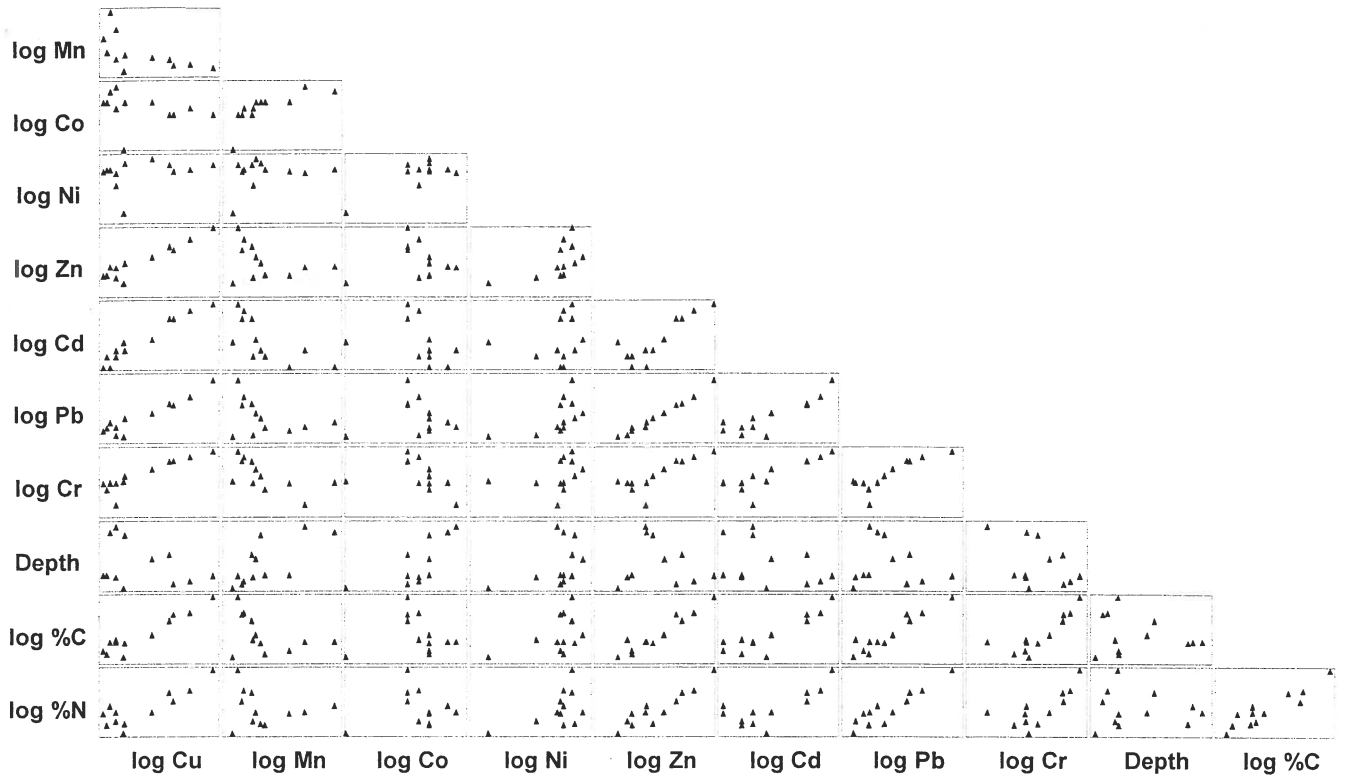


Fig. 11.9. Garroch Head macrofauna {G}. Draftsman plot (all possible pairwise scatter plots) for the 11 abiotic variables recorded at 12 sampling stations across the sewage sludge dumpsite. All variables except water depth have been log transformed.

parts of the community, on the spectrum of rare to common species, may delineate the underlying impact gradient more clearly (see page 9-4), as may some taxonomic levels, higher than species (see page 10-2). Another question which naturally arises is the extent to which the conclusions from BIO-ENV can be supported by significance tests. This is problematic given the lack of model assumptions underlying this

procedure, which can be seen as both a strength (generality, ease of understanding, simplicity of interpretation) and a weakness (lack of a structure for formal statistical inference). A simple RELATE test is available (see Chapter 15) of the null hypothesis that there is *no* relationship between the biotic information and a specific abiotic pattern, i.e. that ρ is effectively zero. This can be examined by a permutation or

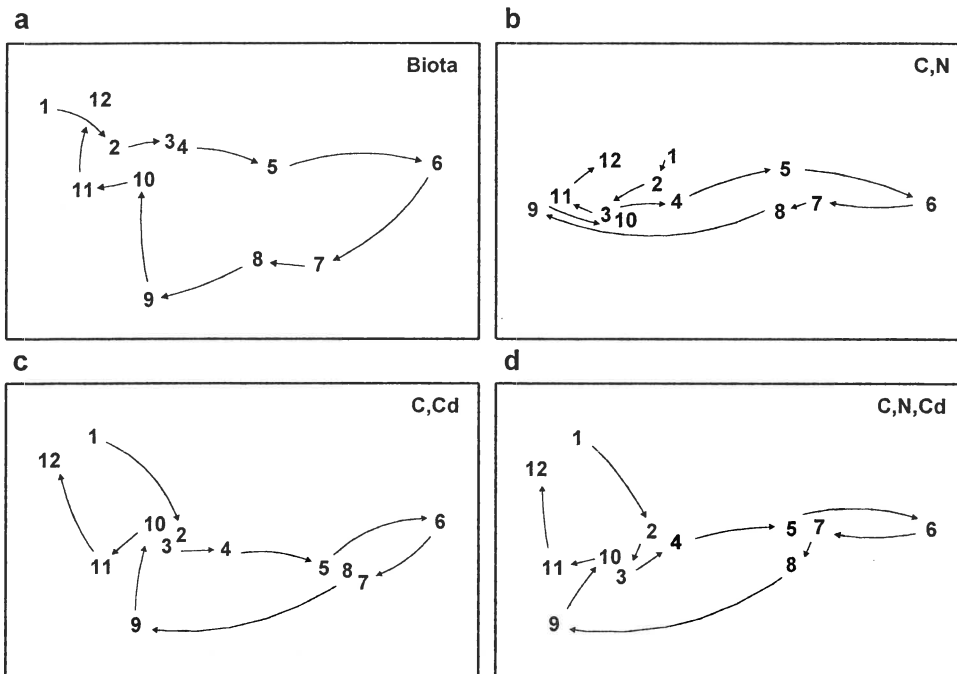


Fig 11.10. Garroch Head macrofauna {G}. MDS plots for the 12 sampling stations across the sewage-sludge dump site (Fig. 8.3), based on: a) species biomass, as in Fig. 11.5a; b)-d) three combinations of carbon, nitrogen and cadmium concentrations (log transformed) in the sediments, the best match with the biota over all combinations of the 8 variables being for C, N and Cd ($\rho_w = 0.78$). (Stress = 0.05, 0, 0.01, 0.01).

randomisation test, of a type met previously in Chapter 6 (Fig. 6.9), in which ρ is recomputed for all (or a random subset of) permutations of the sample labels in one of the two underlying similarity matrices. As usual, if the observed value of ρ exceeds that found in 95% of the simulations, which by definition correspond to unrelated ordinations, then the null hypothesis can be rejected at the 5% level.

Note however that this is *not* a valid procedure if the abiotic set being tested against the biotic pattern is the result of optimal selection by the BIO-ENV procedure, on the *same* data. For v variables, this is implicitly equivalent to carrying out $2^v - 1$ null hypothesis tests, each of which potentially runs a 5% risk of *Type I error* (rejecting the null hypothesis when it is really true). This rapidly becomes a very large number of tests as v increases, and a naïve RELATE test on the optimal combination is almost certain to indicate a ‘significant’ biotic-abiotic relation, even with entirely random data sets![¶] However, the null hypothesis of ‘no relation’ is often not tenable from the start and testing is then rather irrelevant: the BIO-ENV procedure is best thought of as an *exploratory* tool. A more convincing *confirmatory* strategy is to use an initial set of data to suggest an optimal combination of abiotic variables, and an *independent* data set, utilising *only* that reduced number of variables, in a RELATE test or a second BIO-ENV analysis. If there were any variables featuring marginally and arbitrarily in the first run of BIOENV, they would be unlikely to do so again on the second run.

Design

Two final points can be made about the sampling design. The general subject of experimental and field survey design is an immense one, which cannot be covered here.[†] 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 member of the set but, in general, one of them is chosen arbitrarily as a *proxy* for the rest (e.g. in the Garroch Head data, C was a proxy for the highly inter-correlated set C, Cu, Zn, Pb). 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[‡] 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.

[¶] What is needed here, of course, is a randomisation test which incorporates the fitting stage and thus allows for the selection bias in the optimal solution; this is planned for a future version of PRIMER.

[†] Green (1979) provides some useful guidelines, mainly on field observational studies, and Underwood (1998) concentrates on design of field manipulative experiments; both books are largely concerned with univariate data but many of the fundamental issues are common to all analyses.

[‡] 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 the 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 marine biota have mainly been concerned with populations of individual species, rather than communities.

In all cases care should be taken to avoid *pseudo-replication*, 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).

It should be made clear at the outset that the treatment of experiments in this chapter is somewhat cursory.

The subject of ecological experiments requires a book of its own, indeed it gets an excellent one in Underwood (1998). The latter, however, in common with most biologically-oriented texts on experiments, is almost entirely concerned with *univariate* analysis of single attributes (a population abundance, a diversity measure, etc). Experiments with multiple outcomes which are analysed by multivariate methods are still far from commonplace, though becoming more evident in recent literature (for papers with a methodological bent see, for example, Anderson 2001a,b, Chapman and Underwood, 1999, Krzanowski (in press), Legendre and Anderson, 1999, McArdle and Anderson, 2001, Underwood and Chapman, 1998).

'NATURAL EXPERIMENTS'

It is doubtful whether so called *natural experiments* deserve to be called '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), Underwood (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* 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 (macro-benthos) have planktonic larval development and dispersal, there being no room for compromise (Warwick, 1984).

To test these hypotheses we can compare situations where the causal mechanisms differ and therefore give rise to different predictions about pattern. For example, the reproductive dichotomy noted above between macrobenthos and meiobenthos breaks down in the deep-sea, in polar latitudes and in fresh water, although the physical sediment constraints in these situations will be the same as in temperate shelf seas. The evolutionary hypothesis therefore predicts that there should be a unimodal pattern in these situations, whereas the ecological hypothesis predicts that it should remain bimodal. Using these situations as natural experiments, we can therefore falsify one or the other (or both) of these hypotheses.

However, natural experiments of this kind extend outside the purpose of this manual, and our chosen example concerns the ecological effects of disturbance on community structure.

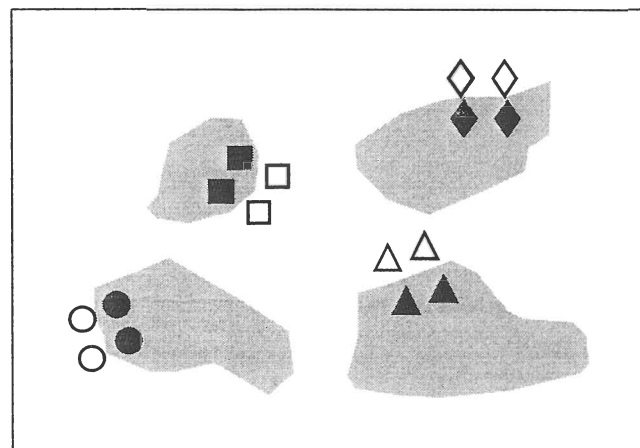


Fig. 12.1. Tasmania, Eaglehawk Neck {T}. Sketch showing the type of sample design. Sample positions (same symbols as in Fig. 12.3) in relation to disturbed sediment patches (shaded).

The effects of disturbance by soldier crabs on meiobenthic community structure {T}

On a sheltered intertidal sandflat at Eaglehawk Neck, on the Tasman Peninsula in S.E. Tasmania, the burrowing and feeding activities of the soldier crab *Mictyris platycheles* are evidenced as intensely disturbed areas of sediment which form discrete patches interspersed with smooth undisturbed areas. The crabs feed by manipulating sand grains in their mandibles and removing fine particulate material, but they are not predators on the meiofauna, though their feeding and burrowing activity results in intense sediment disturbance. This situation was used as a 'natural experiment' on the effects of disturbance on meiobenthic community structure. Meiofauna samples were collected in a spatially blocked design, such that each block comprised two disturbed and two undisturbed samples, each 5m apart (Fig. 12.1).

Table 12.1. Tasmania, Eaglehawk Neck {T}. Mean values per core sample of univariate measures for nematodes, copepods and total meiofauna (nematodes + copepods) in the disturbed and undisturbed areas. The significance levels for differences are from a two-way ANOVA, i.e. they allow for differences between blocks, although these were not significant at the 5% level.

	Total individuals (N)	Total species (S)	Species richness (d)	Shannon diversity (H')	Species evenness (J')
Nematodes					
Disturbed	205	14.4	2.6	1.6	0.58
Undisturbed	200	20.1	3.7	2.2	0.74
Significance (%)	91	1	0.3	0.1	1
Copepods					
Disturbed	94	5.4	1.0	0.96	0.59
Undisturbed	146	5.7	1.0	0.84	0.49
Significance (%)	11	52	99	52	38
Total meiofauna					
Disturbed	299	19.8	3.4	2.0	0.66
Undisturbed	346	25.9	4.4	2.3	0.69
Significance (%)	48	1	3	3	16

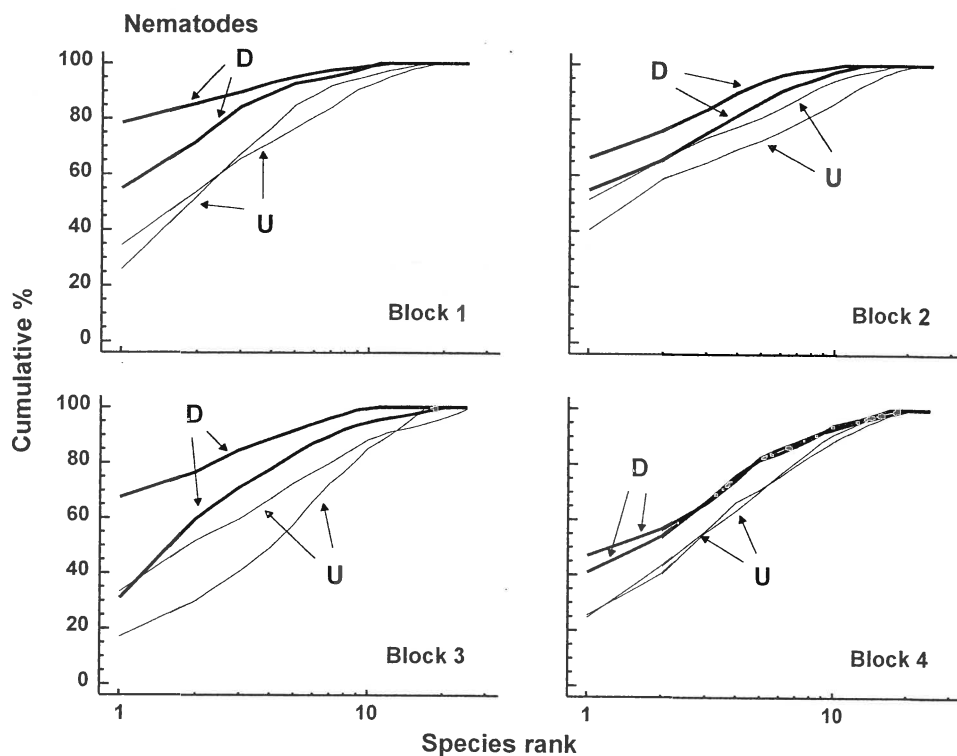


Fig. 12.2. Tasmania, Eaglehawk Neck {T}. Replicate *k*-dominance curves for nematode abundance in each sampling block. D = disturbed, U = undisturbed.

Univariate indices. The significance of differences between disturbed and undisturbed samples (treatments) was tested with two-way ANOVA (blocks/treatments), Table 12.1. For the nematodes, species richness, Shannon 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 of the undisturbed samples (except for one block for the

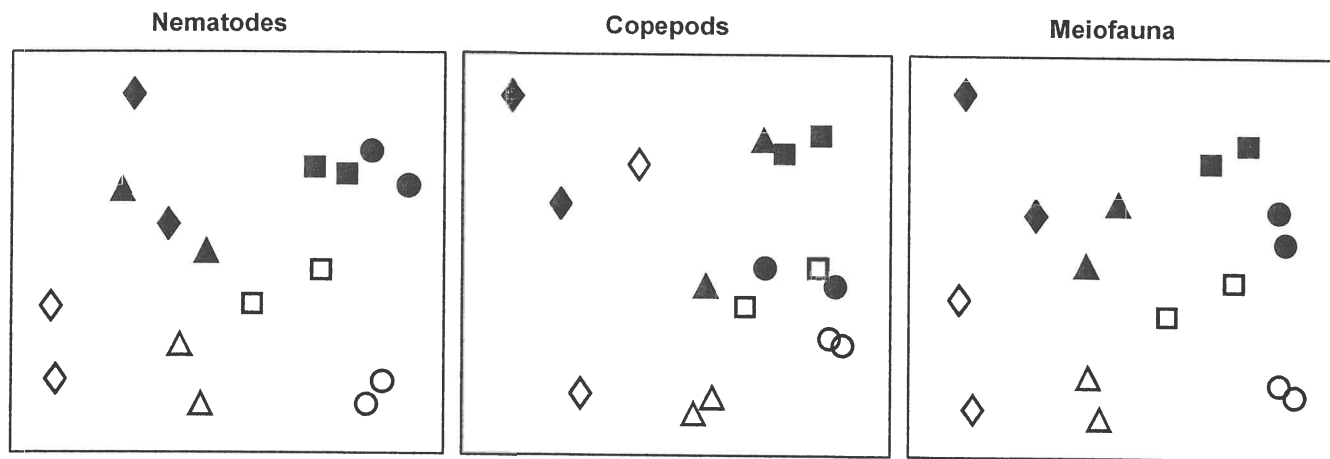


Fig. 12.3. Tasmania, Eaglehawk Neck {T}. MDS configurations for nematode, copepod and 'meiofauna' (nematode + copepod) abundance (root-transformed). Different shapes represent the four blocks of samples. Open symbols = undisturbed, filled = disturbed (stress = 0.12, 0.09, 0.11 respectively).

Table 12.2. Tasmania, Eaglehawk Neck {T}. Results of the two-way ANOSIM test for treatment (disturbance/no disturbance) and block effects.

	Disturbance		Blocks	
	R	Sig.(%)	R	Sig.(%)
Nematodes	1.0	1.2	0.85	0.0
Copepods	0.56	3.7	0.62	0.0
Meiofauna	0.94	1.2	0.85	0.0

copepods), and the blocks are arranged in the same sequence across the plot. For both nematodes and copepods, two-way ANOSIM shows a significant effect of both treatment (disturbance) and blocks, Table 12.2, but the differences are more marked for the nematodes (with higher values of the R statistic).

Conclusions. Univariate indices and graphical/distributional plots were only significantly affected by crab disturbance for the nematodes. Multivariate analysis revealed a similar response for nematodes and copepods (i.e. it seems to be a more sensitive measure of community change). In multivariate analyses, natural variations in species composition across the beach (i.e. between blocks) were about as great as those between treatments within blocks, and the disturbance effect would not have been clearly evidenced without this blocked sampling design.

FIELD EXPERIMENTS

Field manipulative experiments include, for example, caging experiments to exclude or include predators, controlled pollution of experimental plots, and big-bag experiments with plankton. Their use has been predominantly for (univariate) *population* rather than *community* studies, although multivariate analysis of manipulative experiments is becoming more widespread (see, for example, Anderson and Underwood, 1997, Morrissey *et al.*, 1996, Gee and Somerfield, 1997, Austen and Thrush, 2001). The following example is one in which univariate, graphical and multivariate statistical analyses have been applied to meiobenthic communities.

Azoic sediment recolonisation experiment with predator exclusion {Z}

Olafsson and Moore (1992) studied meiofaunal colonisation of azoic sediment in a variety of cages designed to exclude epibenthic macrofauna to varying degrees: A – 1 mm mesh cages designed to exclude all macrofauna; B – 1 mm control cages with two ends left open; C – 10 mm mesh cages to exclude only larger macro-

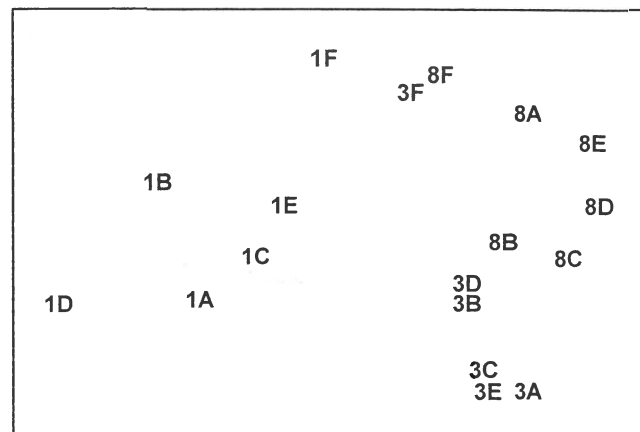


Fig. 12.4. Azoic sediment recolonisation experiment {Z}. MDS configuration for harpacticoid copepods (4th root transformed abundances) after 1, 3 and 8 months, with 6 different treatments (A–F), see text (stress = 0.07).

fauna; D – 10 mm control cages with two ends left open; E – open unmeshed cages; F – uncaged background controls. Three replicates of each treatment were sampled after 1 month, 3 months and 8 months and analysed for nematode and harpacticoid copepod species composition.

Univariate indices. The presence of cages had a more pronounced impact on copepod diversity than nematode diversity. For example, after 8 months, H' and J' (but not S) for copepods had significantly higher values inside the exclusion cages than in the control cages with the ends left open, but for the nematodes, differences in H' were of borderline significance ($p = 5.3\%$).

Graphical/distributional plots. No significant treatment effect for either nematodes or copepods could be detected between k -dominance curves for all sampling dates, using the ANOSIM test referred to at the end of Chapter 8.

Multivariate analysis. For the harpacticoid copepods there was a clear successional pattern of change in community composition over time (Fig. 12.4), but no such pattern was obvious for the nematodes. Fig. 12.4 uses data from Table 2 in Olafsson and Moore's paper, which are for the 15 most abundant harpacticoid species in all treatments and for the mean abundances of all replicates within a treatment on each sampling date. On the basis of these data, there is no significant treatment effect using the 2-way ANOSIM test for one replicate per cell[¶] (see page 6-10), but the fuller replicated data may have been more revealing.

[¶] This is the PRIMER ANOSIM2 test, which will be uninformative in the presence of sizeable treatment/time interactions, a likely possibility here.

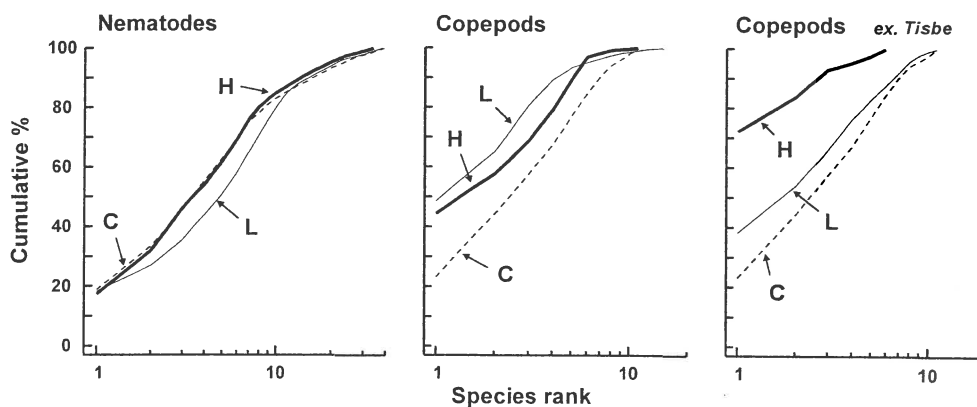


Fig. 12.5. Nutrient enrichment experiment {N}. *k*-dominance curves for nematodes, total copepods and copepods omitting the 'weed' species of Tisbe, for summed replicates of each treatment, C = control, L = low and H = high dose.

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 m³) to mesocosms (1–1000 m³). Recent examples of microcosm experiments analysed by multivariate means can be found, for example, in Austen and McEvoy (1997), Schratzberger and Warwick (1998a, 1999), and mesocosm experiments in Austen *et al* (1998), Widdicombe and Austen (1998, 2001). Macrocosms (larger than 10³ m³), usually involving the artificial enclosure of natural areas in the field, have also been used, for pelagic studies.

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 produced organic enrichment by the addition of powdered *Ascophyllum nodosum* to the surface of the cores, 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 Shannon diversity.

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 (<i>d</i>)	Shannon diversity (<i>H'</i>)	Species evenness (<i>J'</i>)
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
<i>F</i> ratio	0.04	1.39	5.13
<i>Sig</i> level (<i>p</i>)	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
<i>F</i> ratio	17.72	2.65	4.56
<i>Sig</i> level (<i>p</i>)	<0.1%	ns	<5%

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 notable feature of the copepod

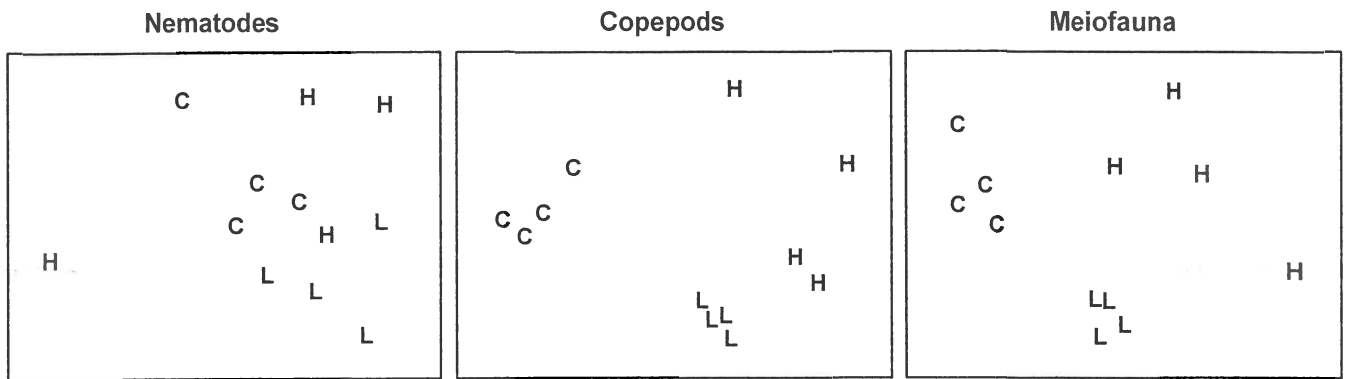


Fig. 12.6. Nutrient enrichment experiment {N}. MDS of $\sqrt{\sqrt{\cdot}}$ -transformed abundances of nematodes, copepods and total meiofauna (nematodes + copepods). C = control, L = low dose, H = high dose (stress = 0.18, 0.09, 0.12).

Table 12.4. Nutrient enrichment experiment {N}. Values of the *R* statistic from the ANOSIM test, in pairwise comparisons between treatments, together with significance levels. C = control, L = low dose, H = high dose.

	Treatment	Statistic value (<i>R</i>)	% Sig. level
Nematodes	(L,C)	0.27	2.9
	(H,C)	0.22	5.7
	(H,L)	0.28	8.6
Copepods	(L,C)	1.00	2.9
	(H,C)	0.97	2.9
	(H,L)	0.59	2.9

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.

Multivariate analysis. Fig. 12.6 shows that, in an MDS of $\sqrt{\sqrt{\cdot}}$ -transformed species abundance data, there is no obvious discrimination between treatments for the *nematodes*. In the ANOSIM test (Table 12.4) the values of the *R* statistic in pairwise comparisons between

treatments are low (0.2–0.3), but there is a significant difference between the low dose treatment and the control, at the 5% level. For the *copepods*, there is a clear separation of treatments on the MDS, the *R* statistic values are much higher (0.6–1.0), and there are significant differences in community structure between all treatments.

Conclusions. The univariate and graphical/distributional techniques show lowered diversity with increasing dose for copepods, but no effect on nematodes. The multivariate techniques clearly discriminate between treatments for copepods, and still have some discriminating power for nematodes. Clearly the copepods have been much more strongly affected by the treatments in all these analyses, but changes in the nematode community may not have been detectable because of the great variability in abundance of nematodes in the high dose boxes. The responses observed in the mesocosm were similar to those sometimes observed in the field where organic enrichment occurs. For example, there was an *increase* in abundance of epibenthic copepods (particularly *Tisbe* spp.) resulting in a *decrease* in the nematode/copepod ratio. In this experiment, however, the *causal link* is closer to being established, though the possible constraints and artefacts inherent in any laboratory mesocosm study should always be borne in mind (see, for example, the discussion in Underwood and Peterson, 1988).

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

COMPONENTS

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

Pelagos

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

Benthos (soft-bottom)

- macrobenthos
- meiobenthos
- microbenthos, not much used in community studies

Benthos (hard-bottom)

- epifauna (encrusting forms, e.g. 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 (e.g. 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; e.g. Colebrook (1986) describes 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.
- c) Fish are of immediate commercial and public interest, and so studies of fish communities are more directly related to the needs of environmental managers than, for example, the meiobenthos (despite the fact, of course, that the latter are vitally important to the early life-stages of fish!).

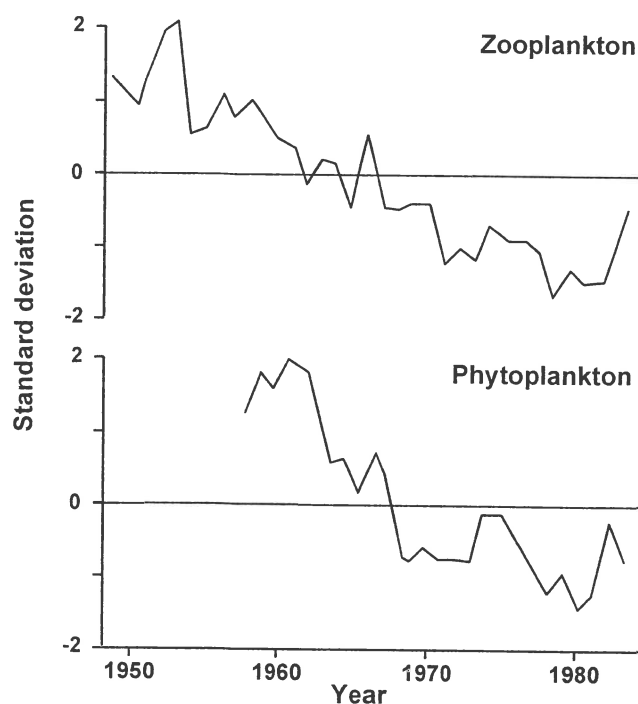


Fig. 13.1. Continuous Plankton Recorder Survey of the NE Atlantic {P}. First principal components for zooplankton and phytoplankton over the first 35 years of the survey (from Colebrook, 1986). Graphs scaled to zero mean and unit variance.

The *disadvantages* of fish are that:

- Strictly quantitative sampling which is equally representative of all the species in the community is difficult. The overall catching efficiency of nets, traps etc. is often unknown, as are the differing abilities of species to evade capture or their susceptibility to be attracted to traps. Visual census methods are also not free from bias, since some species will be more conspicuous in colouration or behaviour than other dull secretive species.
- Uncertainty about site fidelity is usually, but not always, a problem.

Example: Maldives coral reef-fish

For a study in the Maldivian islands, Dawson-Shepherd *et al* (1992) used visual census methods to compare reef-fish assemblages at 23 coral reef-flat sites, 11 of which had been subjected to coral mining for the construction industry and 12 were non-mined controls. The MDS (Fig. 13.2) clearly distinguished mined from non-mined sites.

MACROBENTHOS

The *advantages* of soft-bottom macrobenthos are that:

- They are relatively non-mobile and are therefore useful for studying the local effects of pollutants.
- Their taxonomy is relatively easy.
- Quantitative sampling is relatively easy.
- There is an extensive research literature on the effects of pollution, particularly organic enrichment, on macrobenthic communities, against which specific case-histories can be evaluated.

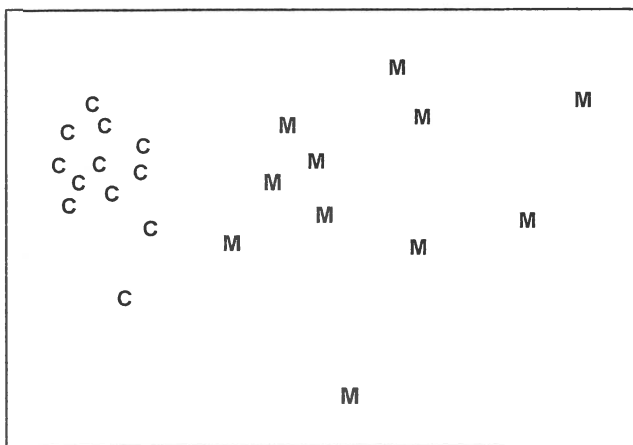


Fig. 13.2. Maldivian Islands, coral-reef fish {M}. MDS ordination of fish species abundance data from mined (M) and control (C) reef-flats (stress = 0.08).

This combination of advantages has resulted in the soft-bottom macrobenthos being probably the most widely used component of the marine biota in environmental impact studies. Despite this, they do have several *disadvantages*:

- Relatively large-volume sediment samples must be collected, so that sampling requires relatively large research ships.
- Because it is generally not practicable to bring large volumes of sediment back to the laboratory for processing, sieving must be done at sea and is rather labour intensive and time consuming (therefore expensive).
- The potential response time of the macrobenthos to a pollution event is slow. Their generation times are measured in years, so that although losses of species due to pollution may take immediate effect, the colonisation of new species which may take advantage of the changed conditions is slow. Thus, the full establishment of a community characterising the new environmental conditions may take several years.
- The macrobenthos are generally unsuitable for causality experiments in mesocosms, because such experiments can rarely be run long enough for fully representative community changes to occur, and recruitment of species to mesocosm systems is often a problem because of their planktonic larval stages (see Chapter 12).

Example: Amoco Cadiz oil-spill

The sensitivity of macrobenthic community structure to pollution events, when using multivariate methods of data analysis, is discussed in Chapter 14. The response of the macrobenthos in the Bay of Morlaix to the Amoco Cadiz oil-spill some 40 km away, described in Chapter 10, is a good example of this (Fig. 13.3).

MEIOBENTHOS

Apart from sharing the advantage of non-mobility, and therefore usefulness for local effects studies, the relative advantages and disadvantages of the meio-benthos are exactly the reverse of the macrobenthos. Their advantages are:

- Because of their small size and high density in marine sediments, quantitative sampling of the meio-benthos is easy from small ships, open boats etc.
- The small volume of the samples means that they

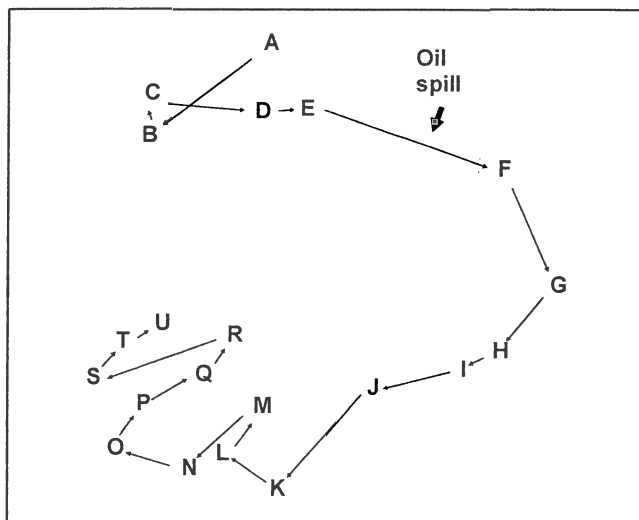


Fig. 13.3. *Amoco-Cadiz oil spill, Bay of Morlaix {A}*. MDS for macrobenthos at station "Pierre Noire", at approximately three-monthly sampling intervals (stress = 0.09).

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 N America, and in many parts of the world the fauna is almost completely unknown. However, to a considerable degree, three factors mitigate 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 as well documented as for the macrobenthos, and there is only a modest 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 of perturbation affecting the macrobenthos but not the meiobenthos, and it was suggested by Warwick *et al* (1990c) that this is more likely to be physical disturbance of the sediment resulting from the regular passage of large cruise liners within the harbour, rather than pollution. This is because the macrobenthos are much more dependent on sediment stability to maintain diversity than are the meiobenthos.

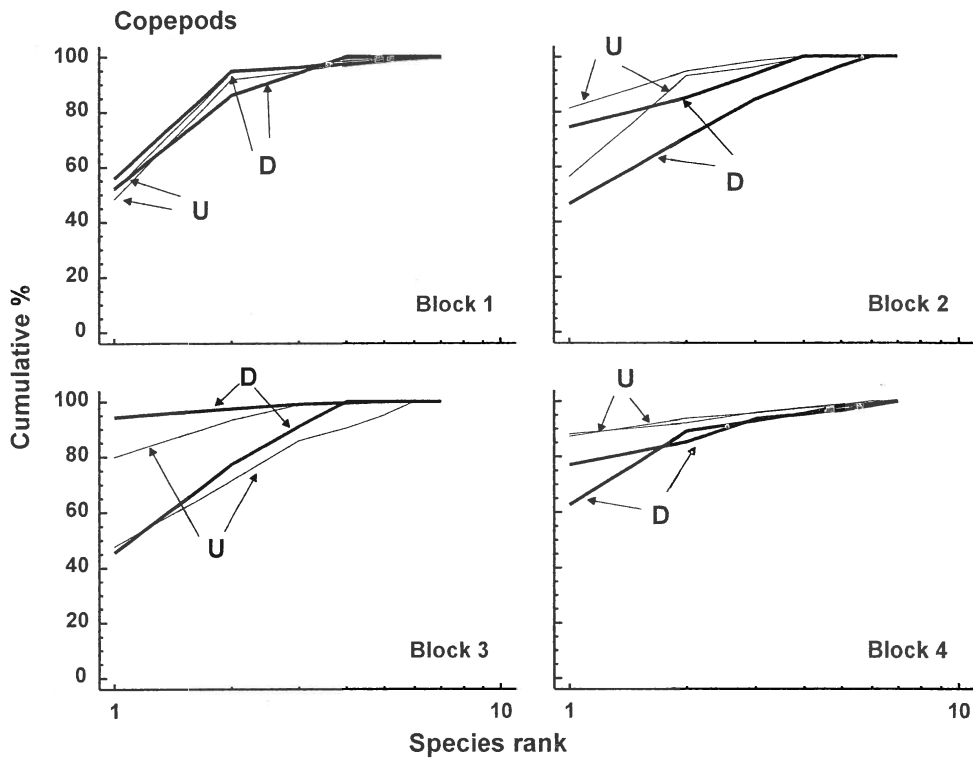


Fig. 13.4. Tasmania, Eaglehawk Neck {T}. *k*-dominance curves for disturbed (D) and undisturbed (U) copepod samples in 4 separate sampling blocks.

HARD-BOTTOM EPIFAUNA

The *advantages* of using hard-bottom encrusting faunas, reef-corals etc. are:

- a) They are immobile and therefore good for local effects studies.
- b) A major advantage over sedimentary faunas is that non-destructive (visual) sampling is possible.

The *disadvantages* are:

- a) Remote sampling is difficult. Intertidal or shallow subtidal sites can be surveyed (the latter by divers), but use of remote cameras requires a greater level of technical sophistication.

- b) Enumeration of colonial organisms is difficult, so that abundance units such as number of colonies or percentage cover must be used; biomass measurements are difficult to make.

Example: Indonesian reef corals

The example shown in Fig. 13.6, of the effects of the 1982-3 El Niño on reef coral communities at South Pari Island, was described in Chapter 10. A clear difference 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.

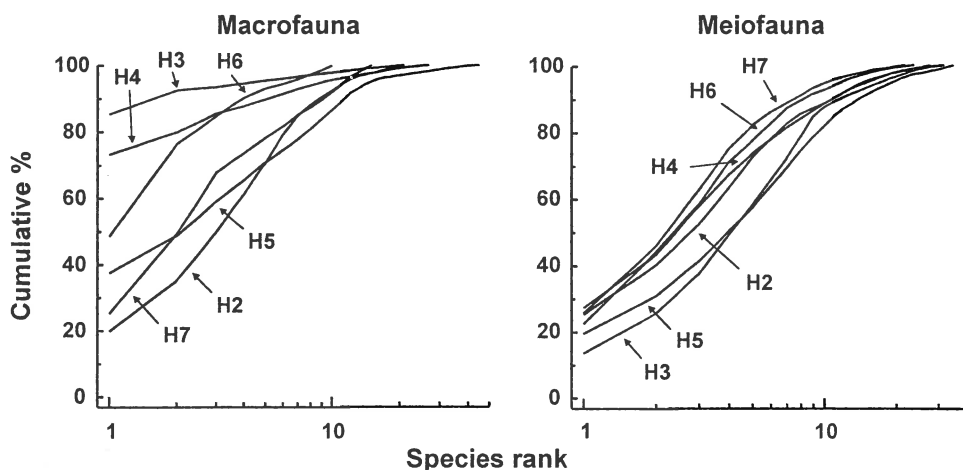


Fig. 13.5. Hamilton Harbour, Bermuda {H}. *k*-dominance curves for macrobenthos (left) and meiobenthic nematodes (right) at six stations (H2-H7).

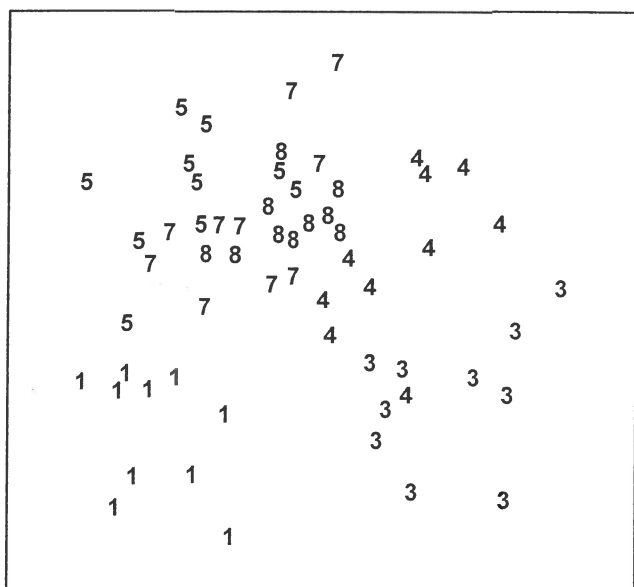


Fig. 13.6. Indonesian reef-corals {I}. MDS for coral species percentage cover data for South Pari Island (10 replicate transects in each year). 1=1981, 3=1983 etc. (stress = 0.25).

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*:

- Remote sampling is difficult.
- Quantitative extraction from the substrate, and comparative quantification of abundances between different substrate types, are difficult.
- Responses to perturbation are largely unknown.
- A suitable habitat (e.g. algae) is not always available. A solution to this problem, and also problem (b),

that has sometimes been tried in practice, is 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 macroalgae (*Chondrus*, *Laurencia*, *Lomentaria*, *Cladophora*, *Polysiphonia*) each collected at eight sites near low water from rocky shores on the Isles of Scilly, UK (Gee and Warwick, 1994). The MDS plots for meio-benthos 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 levels than species may be sufficient for environmental impact analyses. So, when planning a survey, consideration should be given not only to the number of stations and number of replicates to be sampled, but also to the level of taxonomic discrimination which will be used, and which measure(s) of the relative importance of these taxa will be made.

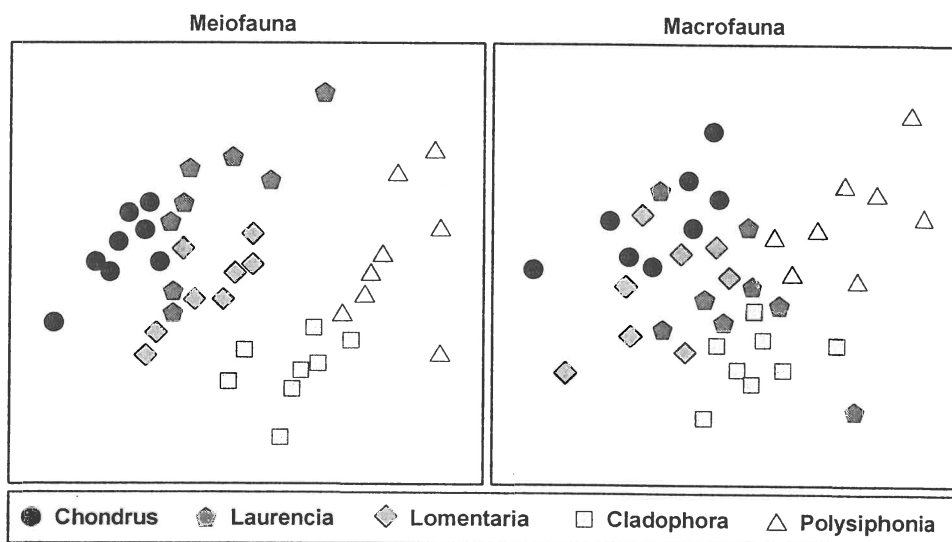


Fig. 13.7. Isles of Scilly seaweed fauna {S}. MDS of standardised $\sqrt{4}$ -transformed meiofauna and macrofauna species abundance data. The five seaweed species are indicated by different symbol and shading conventions (stress = 0.19, 0.18).

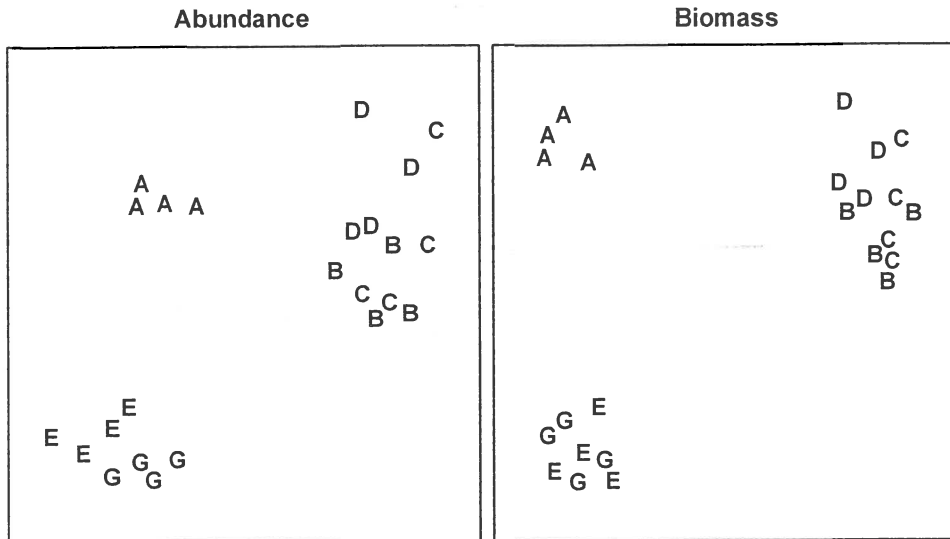


Fig. 13.8. Frierfjord macrofauna {F}. MDS ordinations for abundance and biomass of the 4 replicates at each of the 6 sites (stress = 0.10, 0.08).

Abundance, biomass and production

As a measure of the relative ecological importance for soft-sediment and water-column sampling of species, biomass is better than abundance, and production in turn is better than biomass. However, the determination of annual production of all species within a community over a number of sites or times would be so time consuming as to be completely impracticable.[¶] We are therefore left with the alternatives of studying abundances, biomasses, or both. Abundances are marginally easier to measure, biomass may be a better reflection of ecological importance, and measurement of both abundance and biomass opens the possibility of comparing species-by-sites matrices based on these two different measures (e.g. by the ABC method discussed in Chapter 8).

In practice, multivariate analyses of abundance and biomass data often give remarkably similar results, despite that fact that the species mainly responsible for discriminating between stations are usually different. In Fig. 13.8, for example, the Frierfjord macrobenthos MDS configurations for abundance and biomass are very similar but it is small polychaete species which are mainly responsible for discriminating between sites on the basis of abundance, and species such as the large echinoid *Echinocardium cordatum* which discriminate the sites on the basis of biomass.

[¶] Although relative "production" of species can be approximated using empirical relationships between biomass, abundance and production, and these "production" matrices subjected to multivariate analysis, see Chapter 15.

Species or higher taxa

We have already seen in Chapter 10 that, in many pollution-impact studies, it has been found for both graphical and multivariate analyses that there is surprisingly little loss of information when the species data are aggregated to higher taxa, e.g. genera, families or sometimes even phyla. For the detection of pollution impact, initial collection of data at the level of higher taxa would result in a considerable saving of time (and cost) in the analysis of samples. Such a strategy would, of course, be quite inappropriate if the objective were to be differently defined, for example, the quantification of biodiversity properties.

RECOMMENDATIONS

It is difficult to give firm recommendations as to which components or attributes of the biota should be studied, since this depends on the problem in hand and the expertise and funds available. In general, however, the wider the variety of components and attributes studied, the easier the results will be to interpret. A broad approach at the level of higher taxa is often preferable to a painstakingly detailed analysis of species abundances. If only one component of the fauna is to be studied, then consideration should be given to working up a larger number of stations/replicates at the level of higher taxa in preference to a small number of stations at the species level. Of course, a large number of replicated stations at which both abundance and biomass are determined at the species level is always the ideal!

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).

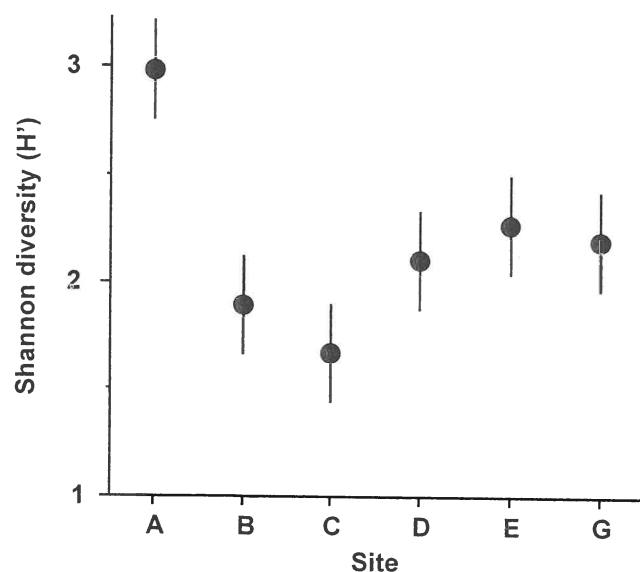


Fig. 14.1. Frierfjord macrobenthos {*F*}. Shannon diversity (mean and 95% confidence intervals) for each station.

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.

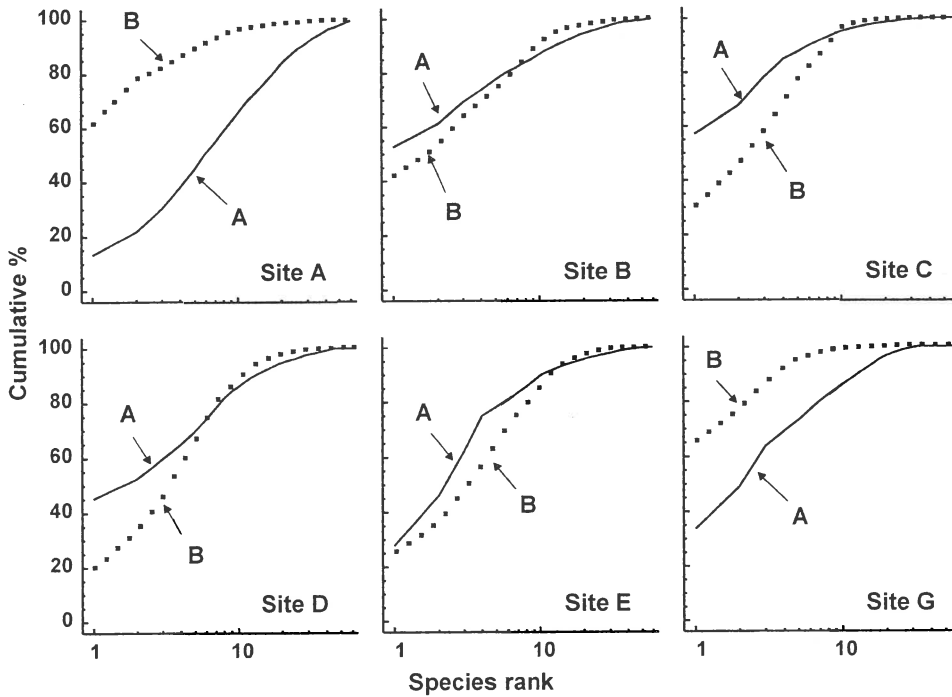


Fig. 14.2. Frierfjord macrobenthos {F}. ABC plots based on the totals from 4 replicates at each of the 6 sites. Solid lines: abundances; dotted lines: biomass.

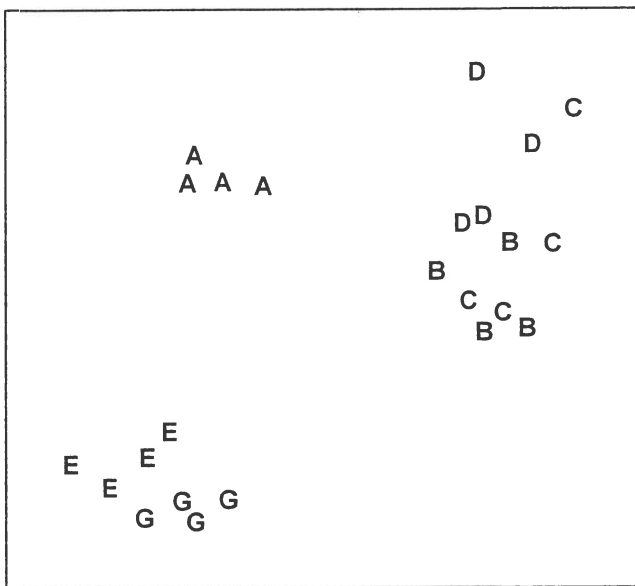


Fig. 14.3. Frierfjord macrobenthos {F}. MDS of 4 replicates at each of sites A-E, G, from Bray-Curtis similarities on 4th root-transformed counts (stress = 0.10).

EXAMPLE 2: Macrobenthos from the Ekofisk oilfield, N. Sea

Changes in community structure of the soft-bottom benthic macrofauna in relation to oil drilling activity at the Ekofisk platform in the North Sea {E} have been described by Gray *et al* (1990). The positions of the sampling stations around the rig are coded by shading and symbol conventions in Fig. 14.4a, according to their distance from the active centre of drilling activity at the time of sampling.

Univariate indices

It can be seen from Fig. 14.4b that species diversity was only significantly reduced in the zone closer than 250m from the rig, and that the three outer zones did not differ from each other in terms of Shannon diversity (this conclusion extends to the other standard measures such as species richness and evenness).

Graphical/distributional plots

The *k*-dominance curves (Fig. 14.4c) also only indicate a significant effect within the inner zone, the curves for the three outer zones being closely coincident.

Multivariate analysis

In the MDS analysis (Fig. 14.4d) community composition in all of the zones was distinct, and there was a clear gradation of change from the inner to outer zones. Formal significance testing (using ANOSIM) confirmed statistically the differences between all zones. It will be recalled from Chapter 10 that there was also a clear distinction between all zones at higher taxonomic levels than species (e.g. family), even at the phylum level for some zones.

Conclusions

Univariate and graphical methods of data analysis suggest that the effects on the benthic fauna are rather localised. The MDS is clearly more sensitive, and can detect differences in community structure up to 3 km away from the centre of activity.

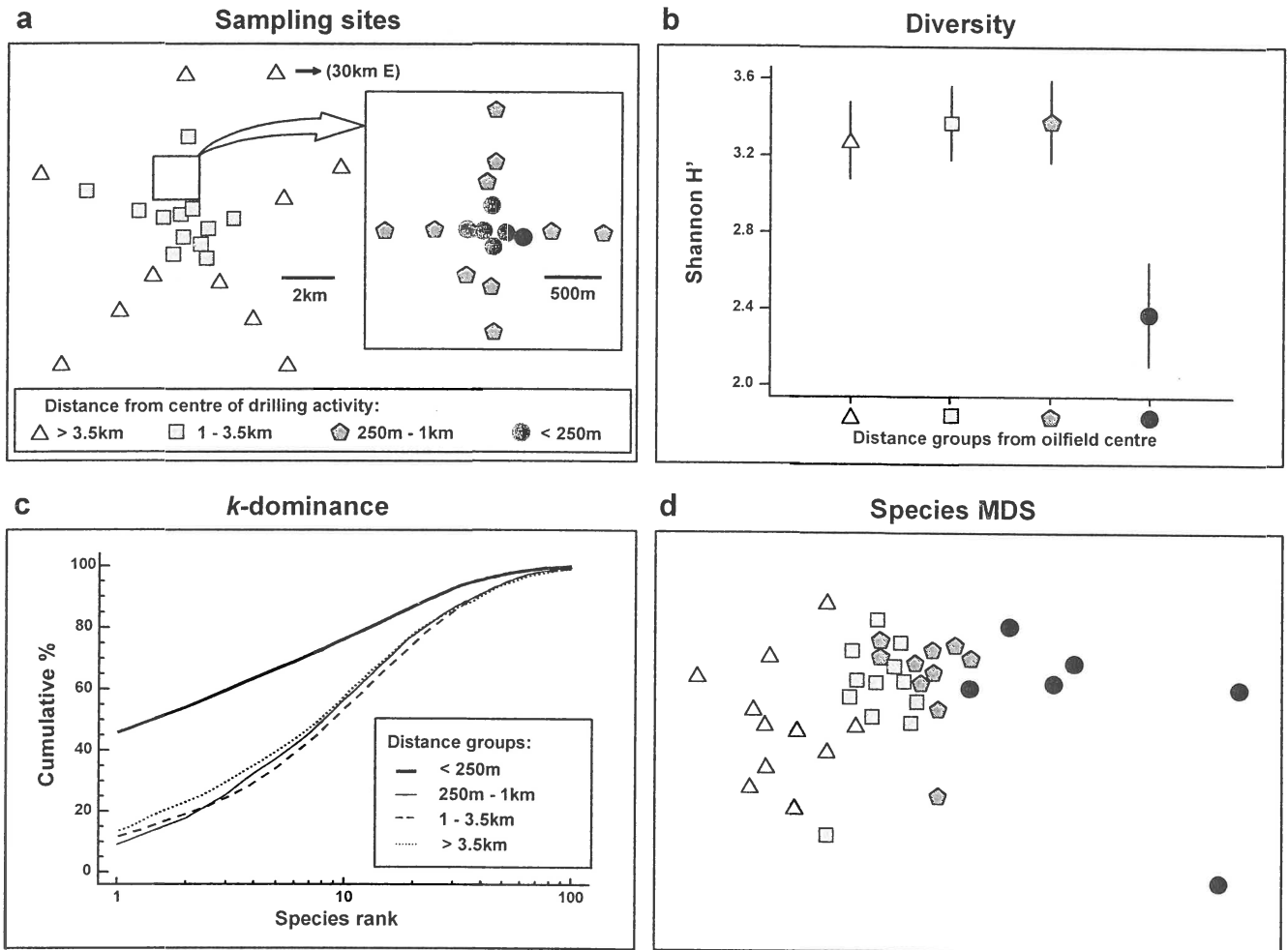


Fig. 14.4. *Ekofisk macrobenthos* {E}. a) Map of sampling sites, represented by different symbol and shading conventions according to their distance from the 2/4K rig at the centre of drilling activity; b) Shannon diversity (mean and 95% confidence intervals) in these distance zones; c) mean k -dominance 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 reef 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).

Graphical/distributional plots

The mean k -dominance curves were similar in 1981 and 1985, with the curves for 1983, 1984, 1987 and 1988 more elevated (Fig. 14.6). Tests on the replicate curves (using the method briefly referred to at the end

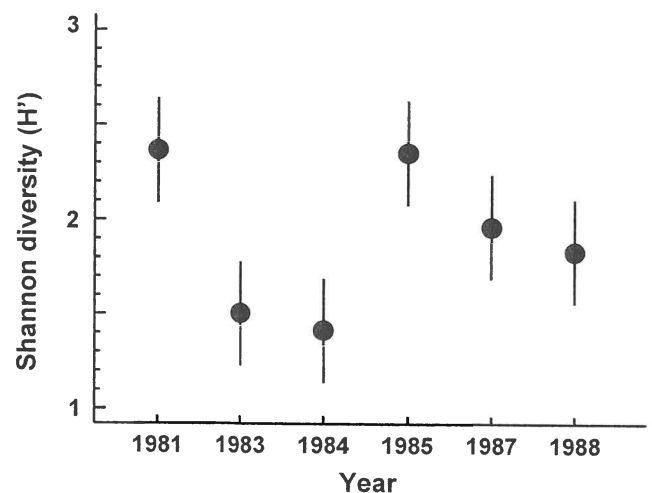


Fig. 14.5. *Indonesian reef corals, Pari Island* {I}. Shannon diversity (means and 95% confidence intervals) of the species coral cover from 10 transects in each year.

of Chapter 8) confirmed the significance of differences between 1981, 1985 and the other years, but the latter were not distinguishable from each other.

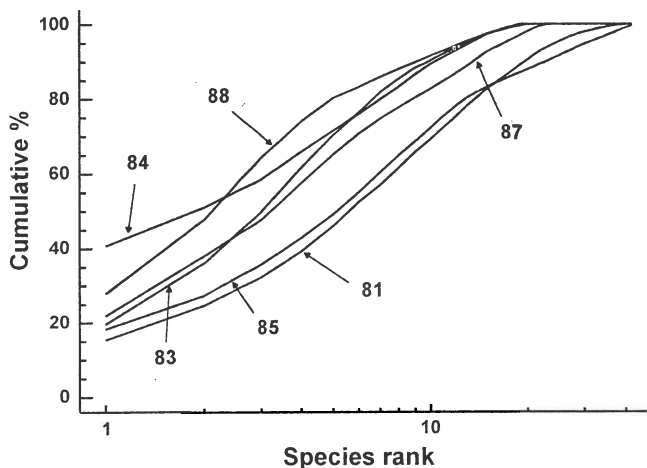


Fig. 14.6. Indonesian reef-corals, Pari Island {I}. *k*-dominance curves for totals of all ten replicates in each year.

Multivariate analysis

Though the MDS has rather a high stress it nonetheless shows an immediate location shift in community composition at the ten replicate sites between 1981 and 1983, and ANOSIM indicates significant differences between all pairs of years. Recovery proceeded in the pre-El Niño direction but was not complete by 1988 (Fig. 14.7).

Conclusions

All methods of data analysis demonstrated the dramatic post El Niño decline in species, though the multivariate techniques were seen to be more sensitive in monitoring the recovery phase in later years.

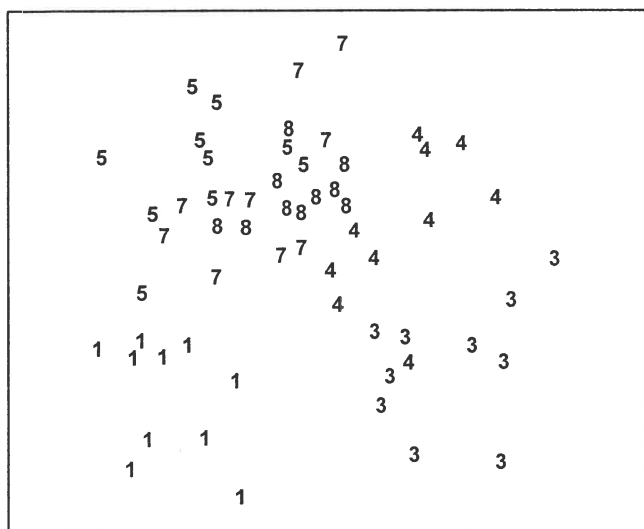


Fig. 14.7. Indonesian reef-corals, Pari Island {I}. MDS for coral species percentage cover data (1 = 1981, 3 = 1983 etc).

EXAMPLE 4: Fish communities from coral reefs in the Maldives

In the Maldivian islands, Dawson-Shepherd *et al* (1992) compared reef-fish assemblages at 23 coral reef-flat sites {M}, 11 of which had been subjected to coral mining for the construction industry and 12 were non-mined controls. The reef-slopes adjacent to these flats were also surveyed.

Univariate indices

Using ANOVA, no significant differences in diversity (Fig. 14.8) were observed between mined and control sites, with no differences either between reef flats and slopes.

Graphical/distributional plots

No significant differences could be detected between mined and control sites, in *k*-dominance curves for either species abundance or biomass. Fig. 14.9 displays the mean curves for reef-flat data pooled across the replicates for each condition.

Multivariate analysis

The MDS (Fig.14.10) clearly distinguished mined from control sites on the reef-flats, and also to a lesser degree even on the slopes adjacent to these flats, where ANOSIM confirmed the significance of this difference.

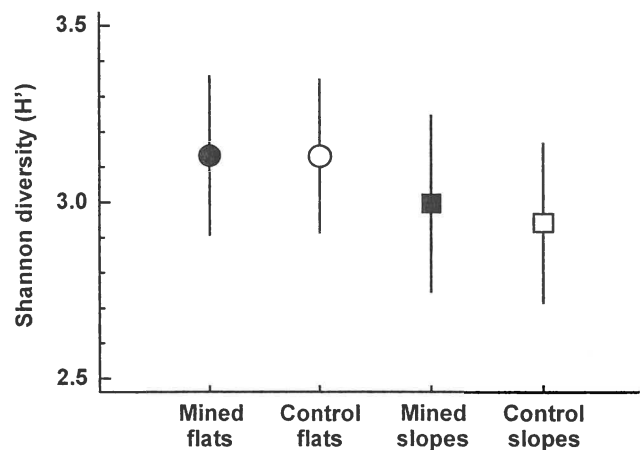


Fig. 14.8. Maldivian Islands, coral-reef fish {M}. Shannon species diversity (means and 95% confidence intervals) at mined (closed symbols) and control (open symbols) sites, for both reef flats (circles) and reef slopes (squares).

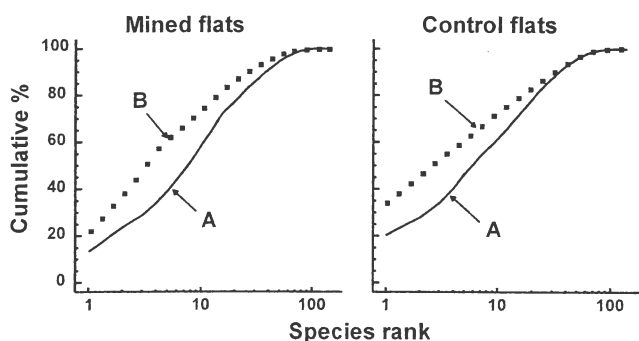


Fig. 14.9. Maldives Islands, coral-reef fish {M}. Average *k*-dominance curves for abundance and biomass at mined and control reef-flat sites.

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 macroalgae (*Chondrus*, *Laurencia*, *Lomentaria*, *Cladophora*, *Polysiphonia*) each collected at eight sites near low water from rocky shores on the Isles of Scilly {S} (Fig. 14.11).

Univariate indices

The meiofauna and macrofauna showed clearly different

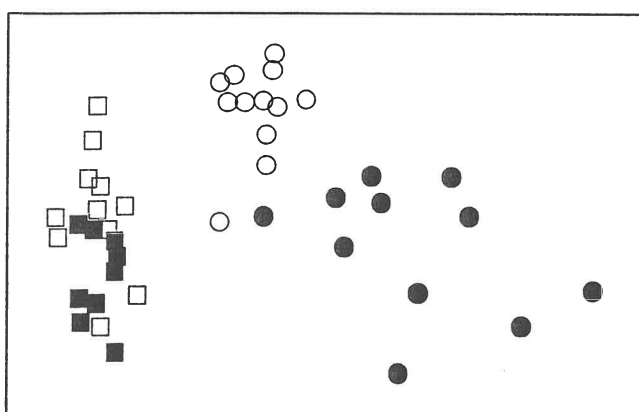


Fig. 14.10. Maldives Islands, coral-reef fish {M}. MDS of 4th root-transformed species abundance data. Symbols as in Fig. 14.8, i.e. circles = reef-flat, squares = slope, solid = mined, open = control (stress = 0.09).

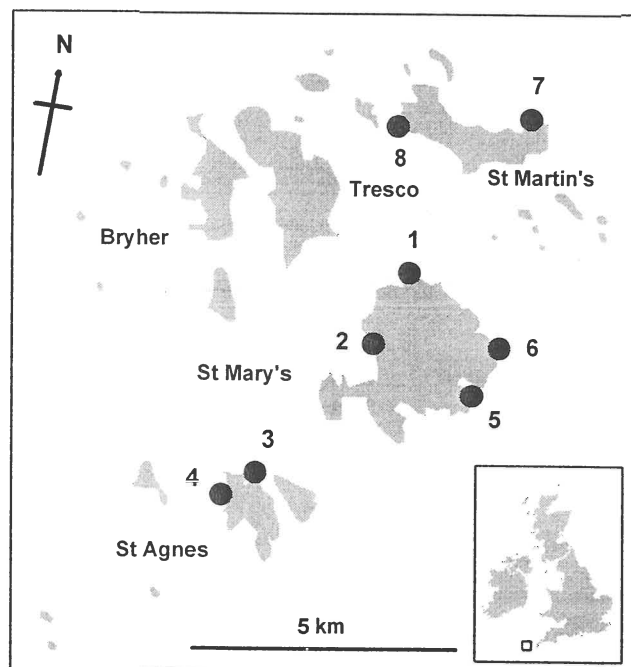


Fig. 14.11. Isles of Scilly {S}. Map of the 8 sites from each of which 5 seaweed species were collected.

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 species 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.

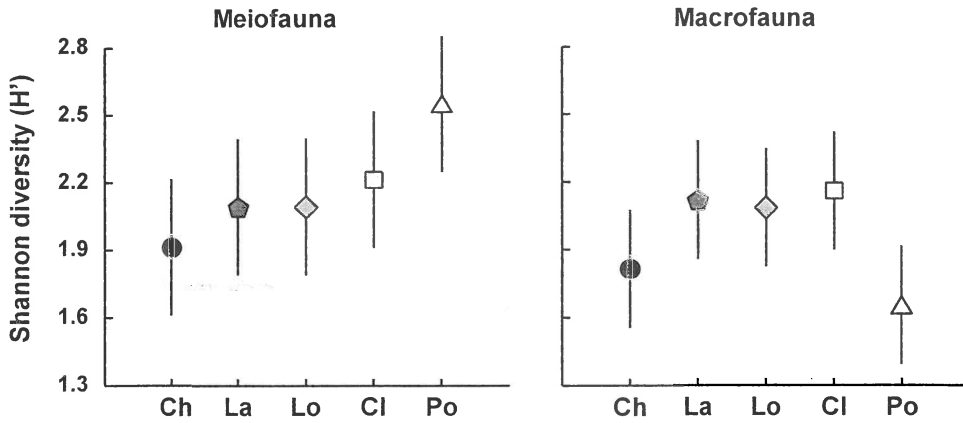


Fig. 14.12. Isles of Scilly seaweed fauna {S}. Shannon diversity (mean and 95% confidence intervals) for the meiofauna and macrofauna of different weed species: Ch = Chondrus, La = Laurencia, Lo = Lomentaria, Cl = Cladophora, Po = Polysiphonia.

Conclusions

The MDS was more sensitive than the univariate or graphical methods for discriminating between weed species. Univariate and graphical methods gave different results for macrobenthos and meiobenthos, whereas for the multivariate methods the results were similar for both.

EXAMPLE 6: Meiobenthos from the Tamar Estuary, S.W. England

Austen and Warwick (1989) compared the structure of the two major taxonomic components of the meiobenthos, nematodes and harpacticoid copepods, in the Tamar estuary {R}. Six replicate samples were taken at a series of ten intertidal soft-sediment sites (Fig. 14.14).

Graphical/distributional plots

The average *k*-dominance curves showed no clear sequencing of sites for the nematodes, for example the curve for site 1 was closely coincident with that for

for site 10 (Fig. 14.15). For the copepods, however, the curves became increasingly elevated from the mouth to the head of the estuary. However, for both nematodes and copepods, many of the curves were not distinguishable from each other.

Multivariate analysis

In the MDS, both nematodes and copepods showed a similar (arched) sequencing of sites from the mouth to the head of the estuary (Fig. 14.16). ANOSIM showed that the copepod assemblages were significantly different in all pairs of sites, and the nematodes in all pairs except 6/7 and 8/9.

Conclusions

The multivariate technique was more sensitive in discriminating between sites, and gave similar patterns for nematodes and copepods, whereas graphical methods gave different patterns for the two taxa. For nematodes, factors other than salinity seemed to be more important in determining diversity profiles, but for copepods salinity correlated well with diversity.

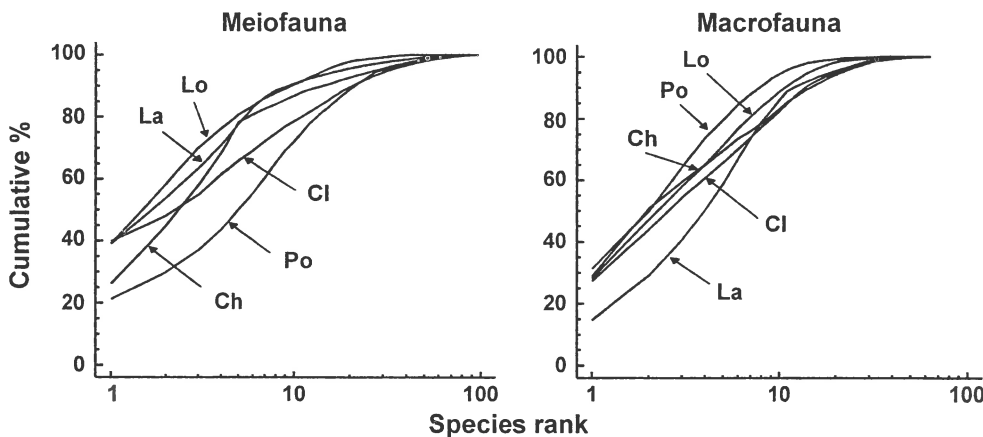


Fig. 14.13. Isles of Scilly seaweed fauna {S}. *k*-dominance curves for meiofauna (left) and macrofauna (right). Ch = Chondrus, La = Laurencia, Lo = Lomentaria, Cl = Cladophora, Po = Polysiphonia.

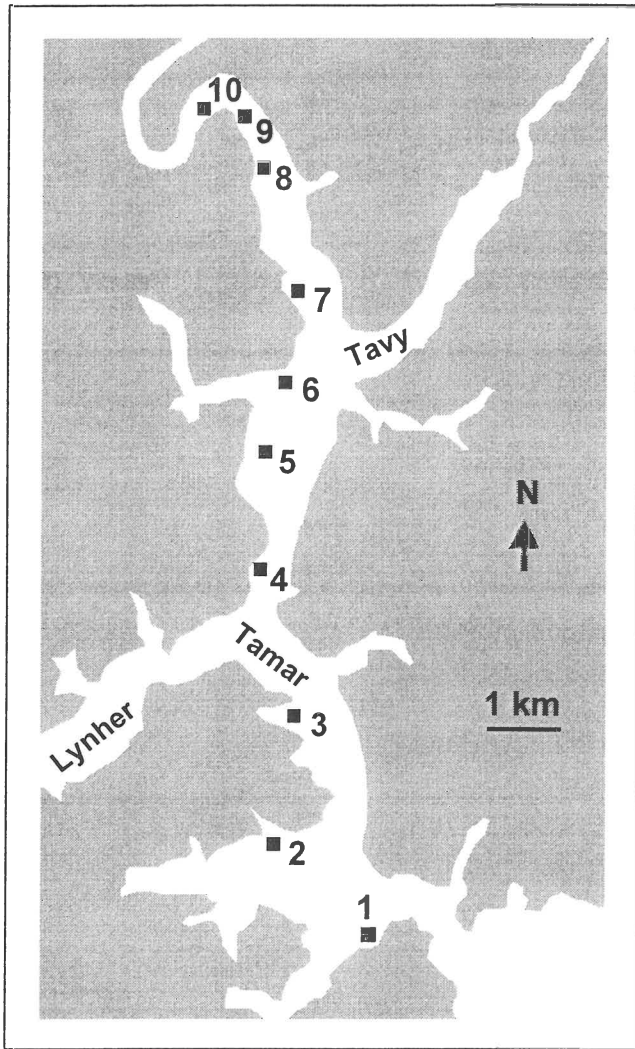


Fig. 14.14. Tamar estuary meiobenthos {R}. Map showing locations of 10 intertidal mud-flat sites.

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 *deleterious* change which can be used in value judgements. Only the species-independent methods of data analysis lend themselves to the determination of deleterious responses although, as we have seen in Chapter 8, even the interpretation of changes in diversity is not always straightforward in these terms. There is a need to employ sensitive techniques for determining stress which utilise the full multivariate information contained in a species/sites matrix, and three such possibilities form the subject of the next chapter.

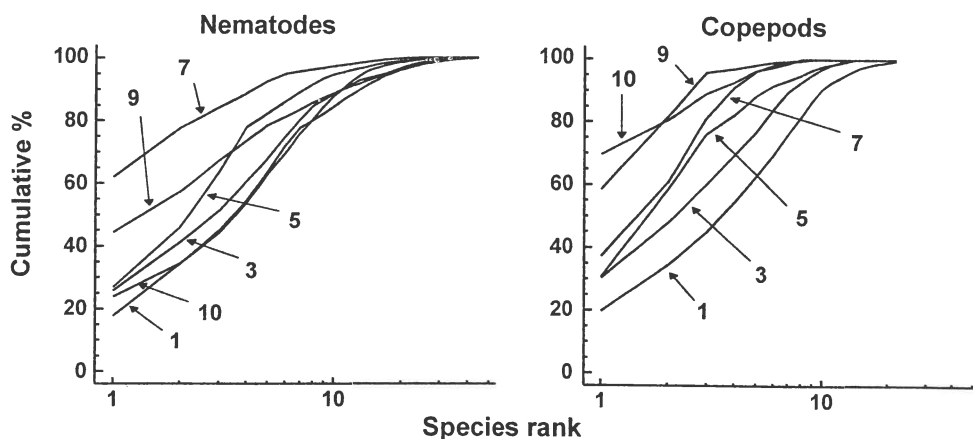


Fig. 14.15. Tamar estuary meiobenthos {R}. *k*-dominance curves for amalgamated data from 6 replicate cores for nematode and copepod species abundances. For clarity of presentation, some sites have been omitted.

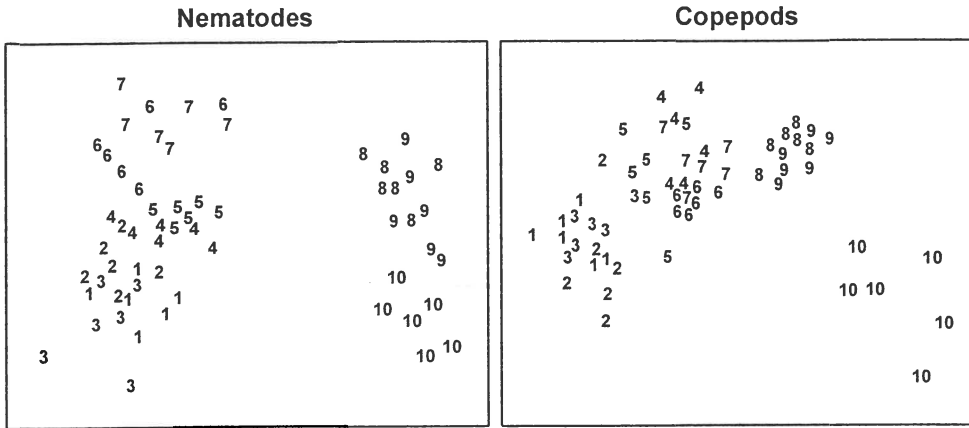


Fig. 14.16. *Tamar estuary meiobenthos {R}*. MDS of 4th root-transformed nematode and copepod species abundance data for six replicate cores at each of 10 stations.

RECOMMENDATIONS

It is important to apply a wide variety of classes of data analysis, as each will give different information and this will aid interpretation. Sensitive multivariate methods will give an 'early warning' that community changes are occurring, but indications that these changes

are deleterious are required by environmental managers, and the less sensitive taxa-independent methods will also play a role. Amongst the latter are the newly-devised biodiversity measures based on taxonomic (or phylogenetic) distinctness of the species making up a sample – see the discussion in Chapter 17 of their advantages over classical diversity indices.

CHAPTER 15: MULTIVARIATE MEASURES OF COMMUNITY STRESS

We have seen in Chapter 14 that multivariate methods of analysis are very sensitive for detecting differences in community structure between samples in space, or changes over time. Generally, however, these methods are 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) are employed. 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 theory (Connell, 1978; Huston, 1979) and empirical observation (e.g. Dauvin, 1984; Widdicombe and Austen, 1998) suggest that increasing levels of disturbance may either decrease or increase diversity, or 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 sometimes 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, 1994). 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 a 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 potentially 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 often 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 Ekofisk $\{E\}$ study.

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:

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

where B/A is simply the mean body-weight, 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 this regression is irrelevant. For each data set the abundance and biomass data were

first aggregated to phyla, following the classification of Howson (1987); 14 phyla were encountered overall (see the later Table 15.1). Abundance and biomass were then combined to form a *production matrix* using the above formula. All data sets were then merged into a single production matrix and an MDS performed on the standardised, 4th root-transformed data using the Bray-Curtis similarity measure. All macrobenthic studies from a single region (the NE Atlantic shelf) for which both abundance and biomass data were available were used, as follows:

- 1) A transect of 12 stations sampled in 1983 on a west-east transect (Fig. 1.5) across a sewage sludge dump-ground at Garroch Head, Firth of Clyde, Scotland {G}. Stations in the middle of the transect show clear signs of gross pollution.
- 2) A time series of samples from 1963–1973 at two stations (sites 34 and 2, Fig. 1.3) in West Scottish sea-lochs, L. Linnhe and L. Eil {L}, covering the period of commissioning of a pulp-mill. The later years show increasing pollution effects on the macrofauna, except that in 1973 a recovery was noted in L. Linnhe following a decrease in pollution loading.
- 3) Samples collected at six stations in Frierfjord (Oslofjord), Norway {F}. The stations (Fig. 1.1) were ranked in order of increasing stress A–G–E–D–B–C, based on thirteen different criteria. The macrofauna at stations B, C and D were considered to be influenced by seasonal anoxia in the deeper basins of the fjord.
- 4) Amoco-Cadiz oil spill, Bay of Morlaix {A}. In view of the large number of observations, the 21 sampling occasions have been aggregated into five years for the meta-analysis: 1977 = pre-spill, 1978 = immediate post-spill and 1979-81 = recovery period.

- 5) Two stations in the Skagerrak at depths of 100 and 300m. The 300m station showed signs of disturbance attributable to the dominance of the sediment reworking bivalve *Abra nitida*.
- 6) An undisturbed station off the coast of Northumberland, NE England.
- 7) An undisturbed station in Carmarthen Bay, S Wales.
- 8) An undisturbed station in Kiel Bay; mean of 22 sets of samples.

In all, this gave a total of 50 samples, the disturbance status of which has been assessed by a variety of different methods including univariate indices, dominance plots, ABC curves, measured contaminant levels etc. The MDS for all samples (Fig.15.1) takes the form of a *wedge* with the pointed end to the right and the wide end to the left. It is immediately apparent that the long axis of the configuration represents a scale of disturbance, with the most disturbed samples to the right and the undisturbed samples to the left. (The reason for the spread of sites on the vertical axis is less obvious). The relative positions of samples on the horizontal axis can thus be used as a measure of the relative severity of disturbance. Another gratifying feature of this plot is that in all cases increasing levels of disturbance result in a shift in the same direction, i.e. to the right. For visual clarity, the samples from individual case studies are plotted in Fig. 15.2, with the remaining samples represented as dots.

- 1) *Garroch Head (Clyde) sludge dump-ground {G}*. Samples taken along this transect span the full scale of the long axis of the configuration (Fig. 15.2a). 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.

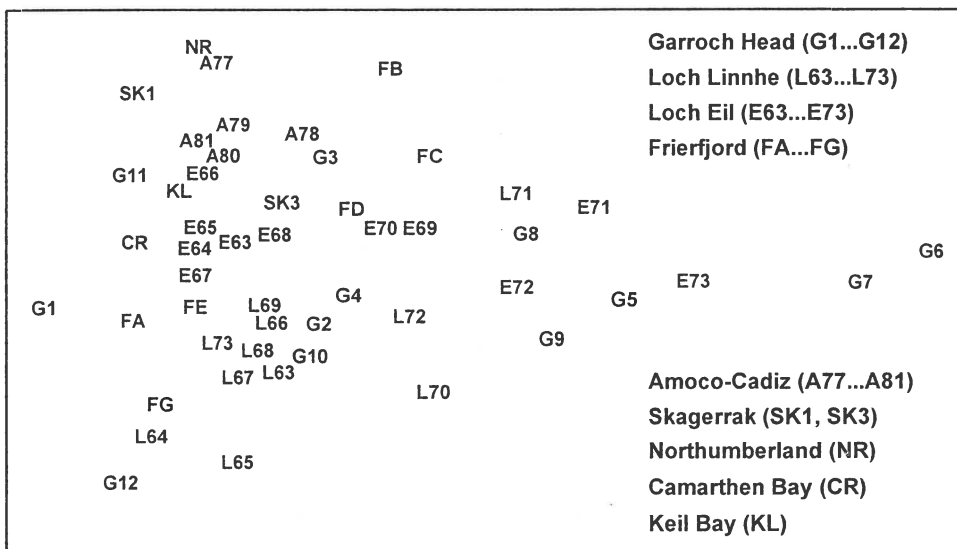


Fig. 15.1. Joint NE Atlantic shelf studies ("meta-analysis") {J}. Two dimensional MDS ordination of phylum level "production" data (stress = 0.16).

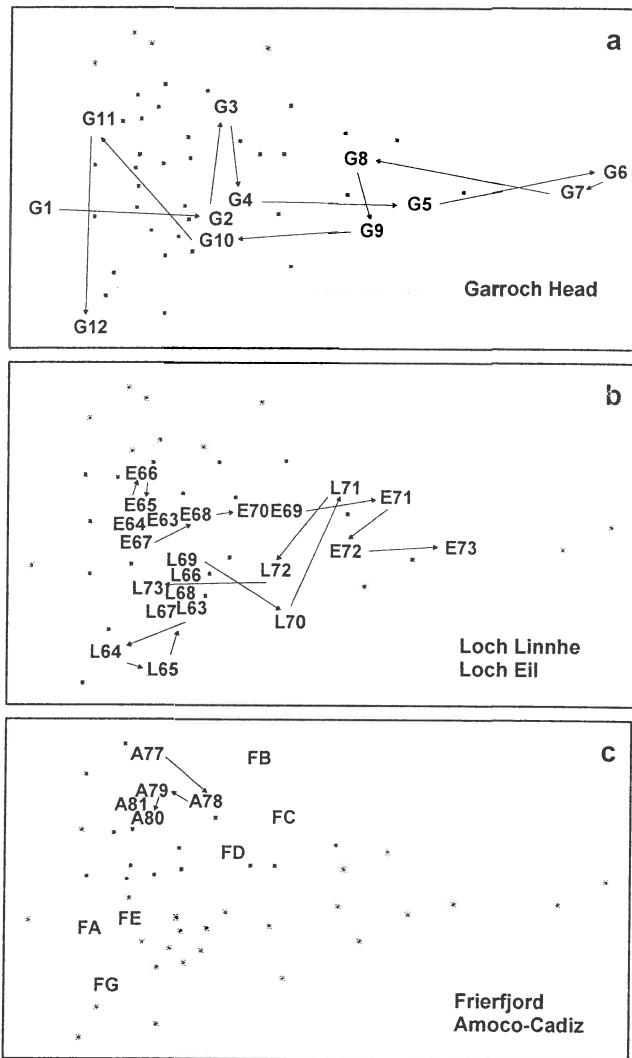


Fig. 15.2. Joint NE Atlantic shelf studies ("meta-analysis") {J}. As Fig. 15.1 but with individual studies highlighted: a) Garroch Head (Clyde) dump-ground; b) Loch Linnhe and Loch Eil; c) Frierfjord and Amoco-Cadiz spill (Morlaix).

- 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 order 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 examine this, Warwick and Clarke (1993a) super-imposed 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 (a technique described in Chapter 11). Both variables had high and low values scattered arbitrarily across the configuration, 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 entirely 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 retained on the macrofaunal ecologists' sieves. Other less important phyla show no clear distribution pattern, except that most are absent from the extreme right-hand samples.

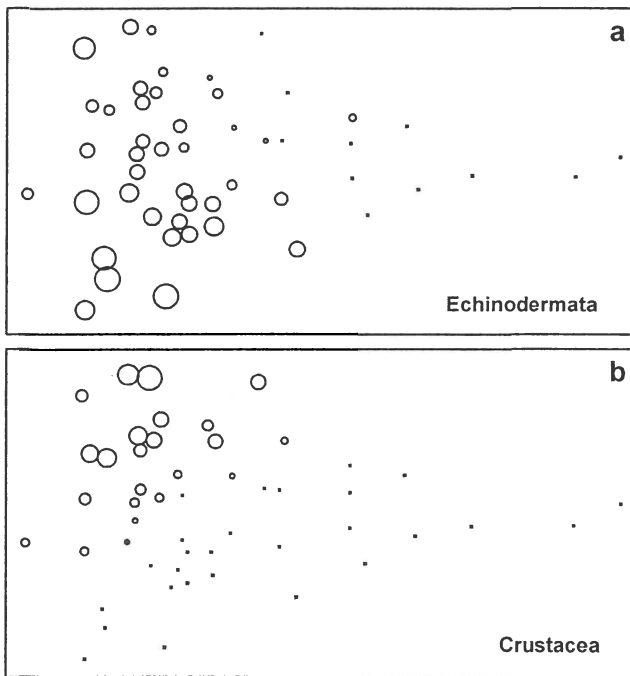


Fig. 15.3. Joint NE Atlantic shelf studies (“meta-analysis”) {J}. As Fig. 15.1 but highlighting the role of specific phyla in shaping the MDS; symbol size represents % production in each sample from: a) echinoderms, b) crustaceans.

This multivariate approach to the comparative scaling of benthic community responses to environmental stress seems to be more satisfactory than taxon-independent methods, having both generality and consistency of behaviour. It is difficult to assess the sensitivity of the technique because data on abundance and biomass of phyla are not available for any really low-level or subtle perturbations. However, its ability to detect the deleterious effect of the Amoco-Cadiz oil spill, where diversity was not impaired, and to rank the Frierfjord samples correctly with respect to levels of stress which had been determined by a wide variety of more time-consuming species-level techniques, suggests that this approach may retain much of the sensitivity of multivariate methods. It certainly seems, at least, that there is a high signal/noise ratio in the sense that natural environmental variation does not affect the communities at this phyletic level to an extent which masks the response to perturbation. The fact that this meta-analysis “works” has a rather weak theoretical basis. Why should Mollusca as a *phylum* be more sensitive to perturbation than Annelida, for example? The answer to this is unlikely to be straightforward and would need to be addressed by considering a broad range of toxicological, physiological and ecological characteristics which are more consistent within than between phyla.

The application of these findings to the evaluation of data from new situations requires that both abundance

and biomass data are available. The scale of perturbation is determined by the 50 samples present in the meta-analysis. These can be regarded as the training set against which the status of new samples can be judged. The best way to achieve this would be to merge the new data with the training set to generate a single production matrix for a re-run of the MDS analysis. The positions of the new data in the two dimensional configuration, especially their location on the principal axis, can then be noted. Of course the positions of the samples in the training set may then be altered relative to each other, though such re-adjustments would be expected to be small. It is also natural, at least in some cases, that each new data set should add to the body of knowledge represented in the meta-analysis, by becoming part of an expanded training set against which further data are assessed. This approach would preserve the theoretical superiority and practical robustness of applying MDS (Chapter 5) in preference to ordination methods such as PCA.

However, there are circumstances in which more approximate methods might be appropriate, such as when it is preferable to leave the training data set unmodified. Fortunately, because of the relatively low dimensionality of the multivariate space (14 phyla, of which only half are of significance), a two-dimensional PCA of the “production” data gives a plot which is rather close to the MDS solution. The eigenvectors for the first three principal components, which explain 72% of the total variation, are given in Table 15.1. The value of the PC1 score for any existing or new sample can then easily be calculated from the first column of this table, without the need to re-analyse the full data set. This score could, with certain caveats (see below), be interpreted as a disturbance index. This index is on a continuous scale but, on the basis of the training data set given here, samples with a score of $>+1$ can be regarded as grossly disturbed, those with a value between -0.2 and $+1$ as showing some evidence of disturbance and those with values <-0.2 as not signalling disturbance with this methodology. A more robust, though less incisive, interpretation would place less reliance on the *absolute* 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.

It should be noted that the training data is unlikely to be fully representative of all types of perturbation that could be encountered. For example, in Fig. 15.1, all

Table 15.1. Joint NE Atlantic shelf studies (“meta-analysis”) {J}. Eigenvectors for the 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

the grossly polluted samples 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). Other studies have looked at specific impacts in relation to the above training set, and reinforced the same pattern, though it must be said that there are disappointingly few examples in the soft-sediment macrobenthic literature to date. It is unclear whether this represents a paucity of data of the right type (biomass measurements are still uncommon, in spite of the relative ease with which they can be made, given the faunal sorting necessary for abundance quantification), or reflects a failure of the analysis to generalise; there appear to be no published data with a contradictory trend, however.

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, total PAHs 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 scattered 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.

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 km 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 outlined above, 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 i th group ($i = 1, \dots, g$), 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 ($N = \sum_i N_i$). Then the *dispersion sequence*

$$\bar{r}_1/k, \bar{r}_2/k, \dots, \bar{r}_g/k \quad (15.4)$$

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.)

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

As an example, the relative dispersion values given by equation (15.4) have been computed[¶] 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 which follow 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.

These four examples all involve either experimental or spatial replication but a similar phenomenon can also be seen with temporal replication. Warwick *et al* (in press) report a study of macrobenthos in Tees Bay, UK, for annual samples (taken at the same two times each year) over the period 1973–96. This straddled a significant, and widely reported, phase shift in planktonic communities in the N Sea, in about 1987. The multivariate dispersion index (IMD), contrasting pre-1987 with post-1987, showed a consistent negative value (increase in inter-annual dispersion in later years) for each of six locations in Tees Bay, at each of the two sampling times (Table 15.4).

Table 15.4. Variation inter-annually, Tees Bay macrobenthos $\{V\}$. Index of Multivariate Dispersion (IMD) between pre- and post-1987 years, before/after a reported change in N Sea pelagic assemblages.

	March	September
Area 0	-0.15	-0.15
Area 1	-0.09	-0.60
Area 2	-0.33	-0.33
Area 3	-0.35	-0.36
Area 4	-0.28	-0.67
Area 6	-0.46	-0.15

[¶] Both the IMD and the relative dispersion values are computed by the PRIMER MVDISP routine.

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 seriation are likely to be modified by disturbances of various kinds. For example, dredging may affect the turbidity and sedimentation regimes and major engineering works may alter the wave climate. Elimination of a particular predator may affect patterns which are due to differential mortality of species caused by that predator. Increased disturbance may also result in the relaxation of interspecific competition, which may in turn result in a breakdown of the pattern of seriation induced by this mechanism. Where a clear sequence of community change along transects is evident in the undisturbed situation, the degree of breakdown of this sequencing could provide an index of subsequent disturbance. Clarke *et al* (1993) have described a possible index of *multivariate seriation*, and applied it to a study of the impact of dredging on intertidal coral reefs at Ko Phuket, Thailand {K}.

In 1986, a deep-water port was constructed on the south-east coast of Ko Phuket, involving a 10-month dredging operation. Three transects were established across nearby coral reefs (Fig. 15.5), transect A being closest to the port and subject to the greatest sedimentation, partly through the escape of fine clay particles through the southern containing wall. Transect C was some 800 m away, situated on the edge of a channel where tidal currents carry sediment plumes away from the reef, and transect B was expected to receive an intermediate degree of sedimentation.

Data from surveys of these three transects, perpendicular to the shore, are presented here for 1983, 86, 87 and 88 (see Chapter 16 for later years). Line-samples 10m long were placed parallel to the shore at 10m intervals along the perpendicular transect from the inner reef flat to the outer reef edge, 12 lines along each of transects A and C and 17 along transect B. The same

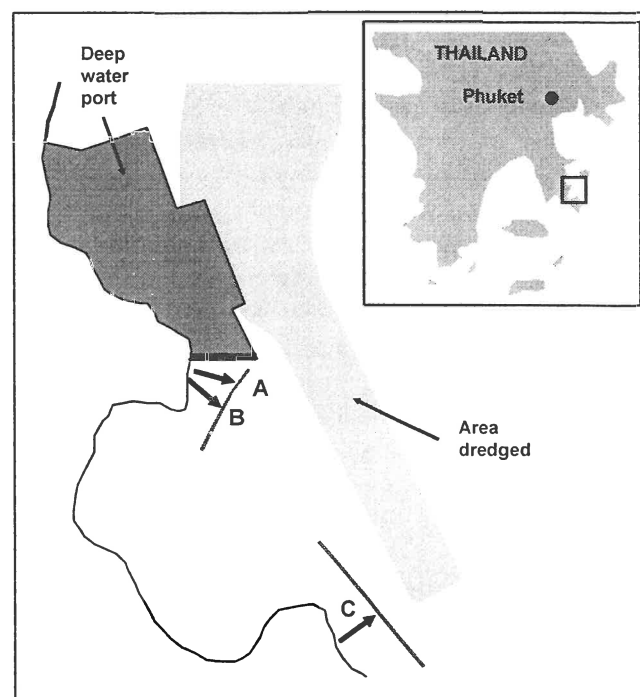


Fig. 15.5. Ko Phuket corals {K}. Map of study site showing locations of transects, A, B and C.

transects were relocated each year and living coral cover (m) of each species recorded. Transect C was not surveyed in 1986.

The basic data were root-transformed and Bray-Curtis similarities calculated between every pair of samples *within* each year/transect combination; the resulting triangular similarity matrices were then input to non-metric MDS (Fig. 15.6). By joining the points in an MDS, in the order of the samples along the inshore-to-offshore transect, one can visualise the degree of seriation, that is, the extent to which the community changes in a smooth and regular fashion, departing ever further from its state at the start of the transect. A measure of linearity of the resulting sequence could be constructed directly from the location of the points in the MDS. However, this could be misleading when the stress is not zero, so that the pattern of relationships between the samples cannot be perfectly represented in 2 dimensions; this will often be the case, as with some of the component plots in Fig. 15.6. (Even where the stress is low, the well-known *horseshoe effect*, Seber, 1984, will mitigate against a genuinely linear sequence appearing in a 2-d ordination as a straight line; see the footnote on page 11-4). Again, a more satisfactory approach is to work with the fundamental similarity matrix that underlies the MDS plots, of whatever dimension.

The index of multivariate seriation (IMS) proposed is therefore defined as a Spearman correlation coefficient (ρ_s , e.g. Kendall, 1970, see also equation 6.3) computed

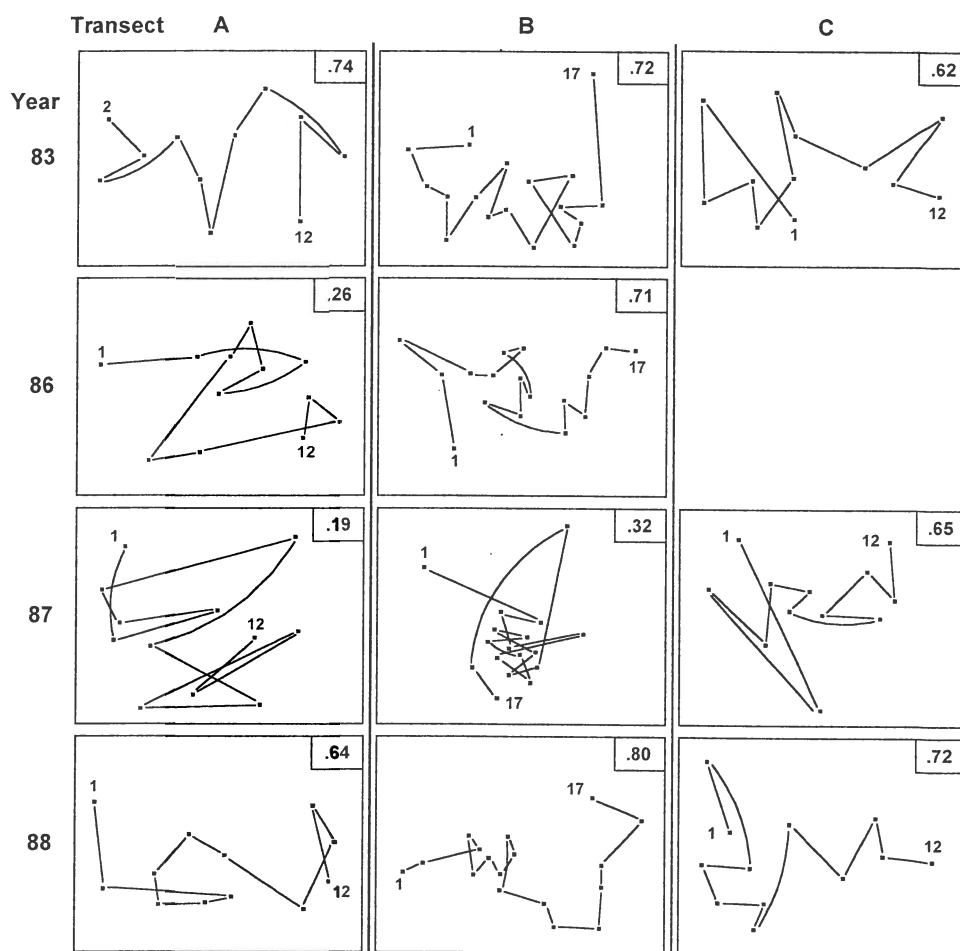


Fig. 15.6. *Ko Phuket corals* {K}. MDS ordination of the changing coral communities (species cover data) along three transects (A to C) at four times (1983 to 1988). The lines indicate the degree of seriation by linking successive points along a transect, from onshore (1) to offshore samples (12 or 17); 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).

between the corresponding elements of two triangular matrices of rank “dissimilarities”. The first is that of Bray-Curtis coefficients calculated for all pairs from the n coral community samples ($n = 12$ or 17 in this case). The second is formed from the inter-point distances of n points laid out, equally-spaced, along a line. If the community changes exactly match this linear sequence (for example, sample 1 is close in species composition to sample 2, samples 1 and 3 are less similar; 1 and 4 less similar still, up to 1 and 12 having the greatest dissimilarity) then the IMS takes the value 1. If, on the other hand, there is no discernible biotic pattern along the transect, or if the relationship between the community structure and distance offshore is very 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 *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 routine ANOSIM2), and again referred to briefly in Chapter 11 in the context of the BIO-ENV procedure.

[¶] The calculations for the tests were carried out using the PRIMER RELATE routine. In exactly the same way, community change could be related to a temporal trend (equally-spaced points in time or around a circle, such as a seasonal cycle) or to the sampling positions in a 2-dimensional spatial layout. There are null hypothesis tests specifically for seriation and cyclicity in RELATE, and a general test for lack of relationship between any two supplied similarity matrices with the same label sets (independently derived). The relation to spatial layout can be tested by first calculating Euclidean distances on an “environmental” data matrix consisting simply of the co-ordinates of the sample positions, then feeding this and the biotic triangular matrix into the RELATE routine.

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, eqt 3.7) gives an appropriate adjustment to ρ_s , and this form is used in the analysis below.

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 clusters separated by large gaps); the community change follows a quite gradual pattern. The values of the IMS are consequently high (Table 15.5), ranging from 0.62 (transect C) to 0.72 (transect B).

The correlation with a linear sequence is highly significant in all three cases. Note that in the 1983 MDS for transect A, the furthest inshore sample has been omitted; it had very little coral cover and was an outlier on the plot, resulting in an unhelpfully condensed display of the remaining points. (This is to be expected in MDS analyses where one sample has a higher dissimilarity to *all* other samples than any other dissimilarity in the matrix, and the MDS needs to be replotted with this point removed). There is no similar technical need, however, to remove this sample from the IMS calculation; this was not done in Table 15.5 though doing so would increase the ρ_s value from 0.65 to 0.74 (as indicated in Fig. 15.6).

Table 15.5. *Ko Phuket corals* {K}. Index of Multivariate Seriation (IMS) along the three transects, for four sampling occasions. Figures in parentheses are the % significance levels in a permutation test for absence of seriation ($T = 999$ simulations).

Year	Transect A	Transect B	Transect C
1983	0.65 (0.1%)	0.72 (0.1%)	0.62 (0.1%)
1986	0.26 (3.8%)	0.71 (0.1%)	–
1987	0.19 (6.4%)	0.32 (0.2%)	0.65 (0.1%)
1988	0.64 (0.1%)	0.80 (0.1%)	0.72 (0.1%)

On transect A, subjected to the highest sedimentation, visual inspection of the MDS gives a clear impression of the breakdown of the linear sequence for the next two sampling occasions. The IMS is dramatically reduced to 0.26 in 1986, when the dredging operations commenced, although the correlation with a linear sequence is still just significant ($p=3.8\%$). By 1987 the IMS on this transect is further reduced to 0.19 and the correlation with a linear sequence is no longer significant. On transect B, further away from the dredging activity, the loss of seriation is not evident until 1987, when the sequencing of points on the MDS configuration breaks down and the IMS is reduced to 0.32, although the latter is still significant ($p=0.2\%$). Note that the MDS plots of Fig. 15.6 may not tell the whole story; the stress values lie between 0.07 and 0.14, indicating that the 2-dimensional pictures are not perfect representations (though unlikely seriously to mislead, see Chapter 5). The largest stress is, in fact, that for transect B in 1987, so that the seriation that is still detectable by the test is only imperfectly seen in the 2-dimensional plot. It is also true that the increased number of points (17) on transect B, in comparison with A and C (12), will lead to a more sensitive test. On transect C there is no evidence of the breakdown of seriation at all, either from the IMS values or from inspection of the MDS plot. By 1988 transects A and B had completely recovered their seriation pattern, with IMS values equal to or higher than their 1983 values, highly significant correlations with a linear sequence ($p<0.1\%$) and clear sequencing evident on the MDS plots. There was clearly a graded response, with a greater breakdown of seriation occurring earlier on the most impacted transect, some breakdown on the middle transect but no breakdown at all on the least impacted transect.

Overall, the breakdown in the pattern of seriation was due to the increase in distributional range of species which were previously confined to distinct sections of the shore. This is commensurate with the disruption of almost all the types of mechanism which have been invoked to explain patterns of seriation, and gives us no clue as to which of these is the likely cause.

CHAPTER 16: FURTHER COMPARISON OF MULTIVARIATE PATTERNS

To motivate the methods of this chapter consider again the analysis of macrobenthic samples from the Bay of Morlaix {A}, before and after the Amoco-Cadiz oil spill. The MDS of Fig. 16.1 shows a clear signal of community change through time, a combination of cyclical seasonal fluctuations (the samples are approximately quarterly) with the major perturbation of the oil spill after approximately a year, and a partial recovery over the next four years. The intricate and informative picture is based on a matrix of 257 species but the question naturally arises as to whether all these species are influential in forming the temporal pattern. This cannot be the case, of course, because many species are very uncommon. The later Fig. 16.3a shows an identical MDS plot based on only 125 species, the omitted 'least important' 132 species accounting for only 0.2% of the total abundance and, on average, being absent from all 5 replicate samples on 90% of the 21 sampling times. However, the question still remains: do all the 125 species contribute to the MDS or is the pattern largely determined by a small number of highly influential species? If the latter, an MDS of that small species subset should generate an ordination that looks very like Fig. 16.1, and this suggests the following approach (Clarke and Warwick, 1998a).

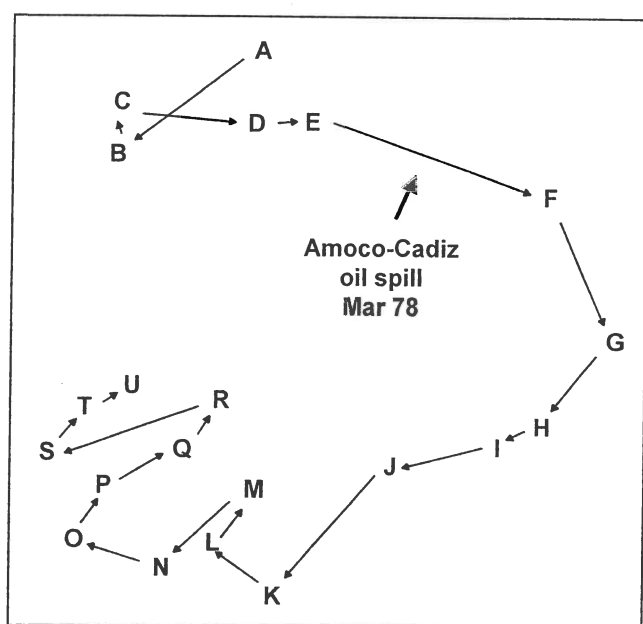


Fig. 16.1. Amoco-Cadiz oil spill {A}. MDS for 257 macrobenthic species in the Bay of Morlaix, for 21 sampling times (A, B, C, ..., U; see legend to Fig. 10.4 for precise dates). The ordination is based on Bray-Curtis similarities from fourth root-transformed abundances and the samples were taken at approximately quarterly intervals over 5 years, reflecting normal seasonal cycles and the perturbation of the oil spill (stress = 0.09).

MATCHING OF ORDINATIONS

The BIO-ENV technique of Chapter 11 can be generalised in a natural way, to the selection of species rather than environmental variables. The procedure is shown schematically in Fig. 16.2. Here the two starting data sets are not: 1) biotic, and 2) abiotic descriptions of the same set of samples, but: 1) the faunal matrix, and 2) a copy of that *same* faunal matrix. Variable sets (species) are selected from the second matrix such that their sample ordination matches, 'as near as makes no difference', the ordination of samples from the first matrix, the full species set. This matching process, as seen in Chapter 11, best takes place by optimising the correlation between the elements of the underlying similarity matrices, rather than matching the respective ordinations, because of the approximation inherent in viewing inter-sample relationships in only 2-dimensions, say. The appropriate correlation coefficient could be Spearman or Kendall, or some weighted form of Spearman, but there is little to be gained in this context from using anything other than the simplest form, the standard Spearman coefficient (ρ).

A definition of a 'near-perfect' match is needed, and this is (somewhat arbitrarily) deemed to be when ρ exceeds 0.95. Certainly two ordinations from similarity matrices that are correlated at this level will be virtually indistinguishable and could not lead to different interpretation of the patterns. The requirement is therefore to find the smallest possible species subset whose Bray-Curtis similarity matrix correlates at least at $\rho = 0.95$ with the (fixed) similarity matrix for the full set of species.

There is a major snag, however, to carrying over the BIO-ENV approach to this context. A search through all possible subsets of 125 species involves: 125 possibilities for a single species, ${}_{125}C_2 (= 125.124/2)$ pairs of species, ${}_{125}C_3 (= 125.124.123/6)$ triples, etc., and this number clearly gets rapidly out of control. In fact a full search would need to look at $2^{125} - 1$ possible combinations, an exceedingly large number!

Stepwise procedure

One way round the problem is to search not over every possible combination but some more limited space, and the natural choice here is a *stepwise* algorithm which operates sequentially and involves both forward

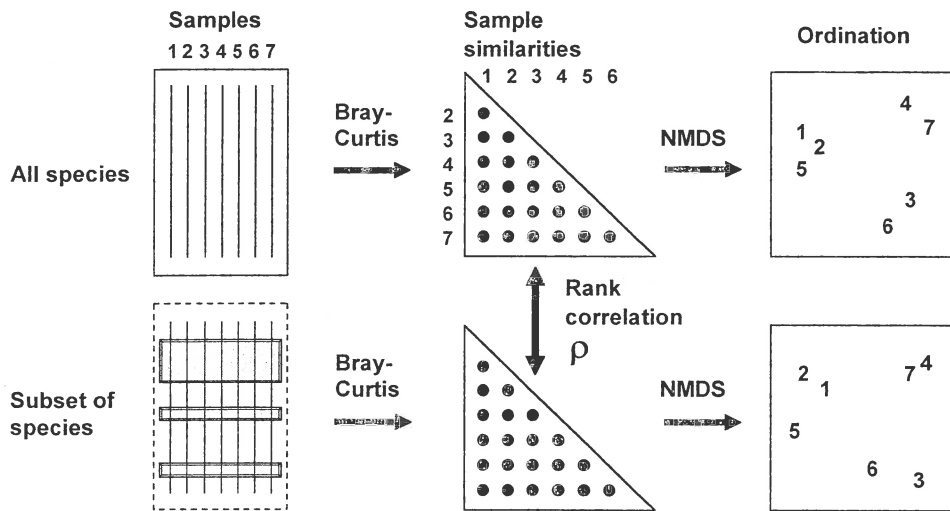


Fig. 16.2. Schematic diagram of selection of a subset of species whose multivariate sample pattern matches that for the full set of species. The search is either over all subsets of species (generalised BIO-ENV routine) or, more practically, stepwise selection of species (BVSTEP routine), with the aim of finding the smallest subset of species giving rank correlation between the similarity matrices of $\rho \geq 0.95$.

and backward-stepping phases.[¶] At each stage, a selection is made of the best single species to *add to* or *drop from* the existing selected set. Typically, the procedure will start with a null set, picking the best single variable (maximising ρ), then adding a second variable which gives the best combination with the first, then adding a third to the existing pair. The backward elimination phase then intervenes, to check whether the first selected variable can now be dropped, the combination of second and third selections alone not having been considered before. The forward selection phase returns and the algorithm proceeds in this fashion until no further improvement is possible by the addition of a single variable to the existing set or, more likely here, the stopping criterion is met (ρ exceeds 0.95). In order fully to clarify the alternation of forward and backward stepping phases, Table 16.1 describes a purely hypothetical (and unrealistically convoluted) search over 6 variables. Analogously to the MDS algorithm of Chapter 6, it is quite possible that such an iterative search procedure will get trapped in a local optimum and miss the true best solution; only a minute fraction of the vast search space is ever examined. Thus, it may be helpful to begin the search at several, different, random starting points, i.e. to start sequential addition or deletion from an existing, randomly selected set of 25% (say) of the species.[†]

[¶] This concept may be familiar from stepwise multiple regression in univariate statistics, which tackles a similar problem of selecting a subset of explanatory variables which account for as much as possible of the variance in a single response variable.

[†] The PRIMER BVSTEP routine carries out this stepwise approach on the active worksheet (the faunal data matrix), for a separately specified, fixed similarity matrix (Bray-Curtis on the faunal data matrix here). There are options always to exclude, or always to include, certain variables (species) in the selection, to start the algorithm either with none, all or a random set of variables in the initial selection, and to output results of the iteration at various levels of detail.

EXAMPLE: Amoco-Cadiz oil spill

Applying this (BVSTEP) procedure to the 125-species set from the Bay of Morlaix, a smallest subset of only 9 species can be found, whose similarity matrix across the 21 samples correlates with that for the full species set, at $\rho \geq 0.95$. The MDS plot for the 21 samples based only on these 9 species is shown in Fig. 16.3b and is seen to be largely indistinguishable from 16.3a. The make-up of this *influential species* set is discussed later but it is important to realise, as often with stepwise procedures, that this may be far from a unique solution. There are likely to be other sets of species, a little larger in number or giving a slightly lower ρ value, that would do a (nearly) equally good job of 'explaining' the full pattern.

One interesting way of seeing this is to discard the initial selection of 9 species, and search again for a further subset that produces a near-perfect match ($\rho \geq 0.95$) to the pattern for the full set of 125 species. Fig. 16.3c shows that a second such set *can* be found, this time of 11 species. If the two sets are discarded, a third (of 14 species), then a fourth (of 18 species) can also be identified, and Fig. 16.3d and e again show the high level of concordance with the full set, Fig. 16.3a. There are now 73 species left and a fifth set can just about be pulled out of them (Fig. 16.3f), though now the algorithm terminates at a genuine maximum of ρ ; a match better than $\rho = 0.91$ cannot be found by the stepwise procedure, even after several attempts with different random starting positions. If these (27) species are also discarded, the ability of the remaining 46 species to reconstruct the initial pattern degrades slowly (Fig. 16.3g) then rapidly (Fig. 16.3h and i), i.e. little of the original 'signal' remains.

Table 16.1. Hypothetical illustration of stages in a stepwise algorithm (F: forward selection, B: backward elimination steps) to select a subset of species which match the multivariate sample pattern for a full set (here, 6 species). Bold underlined type indicates the subset with the highest ρ at each stage, and italics denote a backward elimination step that decreases ρ and is therefore ignored. The procedure ends when ρ attains a certain threshold ($\rho \geq 0.95$), or when forward selection does not increase ρ .

Step	Direction	Species sets						Best ρ
1	F	1	<u>2</u>	3	4	5	6	0.6
2	F	2+1	2+3	<u>2+4</u>	2+5	2+6		0.65
3	B	<u>2</u>	4					0.6
4	F	2+4+1	2+4+3	<u>2+4+5</u>	2+4+6			0.7
5	B	2+4	2+5	<u>4+5</u>				0.8
6	B	<u>4</u>	5					0.55
7	F	<u>4+5+1</u>	4+5+2	4+5+3	4+5+6			0.85
8	B	<u>4+5</u>	4+1	5+1				0.8
9	F	4+5+1+2	4+5+1+3	<u>4+5+1+6</u>				0.9
10	B	4+5+1	4+5+6	4+1+6	<u>5+1+6</u>			0.92
11	B	5+1	<u>5+6</u>	1+6				0.93
12	B	<u>5</u>	6					0.55
13	F	5+6+1	<u>5+6+2</u>	5+6+3	5+6+4			0.94
14	B	<u>5+6</u>	5+2	6+2				0.93
15	F	5+6+2+1	5+6+2+3	<u>5+6+2+4</u>				<u>0.95</u>
16	B	<u>5+6+2</u>	5+6+4	5+2+4	6+2+4			0.94
17	STOP	$\rho = 0.95$ threshold reached, for species subset <u>2+4+5+6</u>						

Clarke and Warwick (1998a) discuss the implication of these plots for concepts of *structural redundancy* in assemblages (and, arguably, for functional redundancy, or at least compensation capacity). They investigate whether the various sets of species ‘peeled’ out from the matrix have a similar taxonomic structure. For example, Table 16.2 displays the first and second ‘peeled’ species lists and defines a *taxonomic mapping* coefficient, used to measure the degree to which the first set has taxonomically closely-related counterparts

in the second set, and vice-versa. (Note that such taxonomic-relatedness concepts are the basis of several biodiversity indices proposed in Chapter 17.) A permutation test can be constructed which leads to the conclusion that the peeled subsets are more taxonomically similar (i.e. have greater *taxonomic coherence*) than would be expected by chance. The number of such coherent subsets that can be ‘peeled out’ from the matrix is clearly some measure of redundancy of information content.

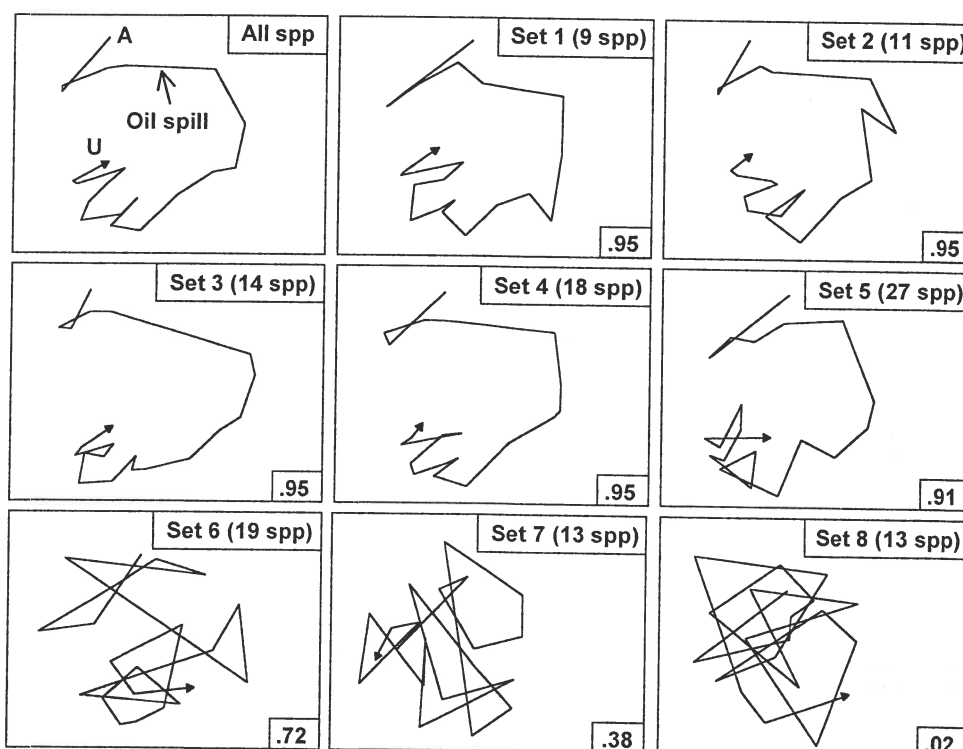
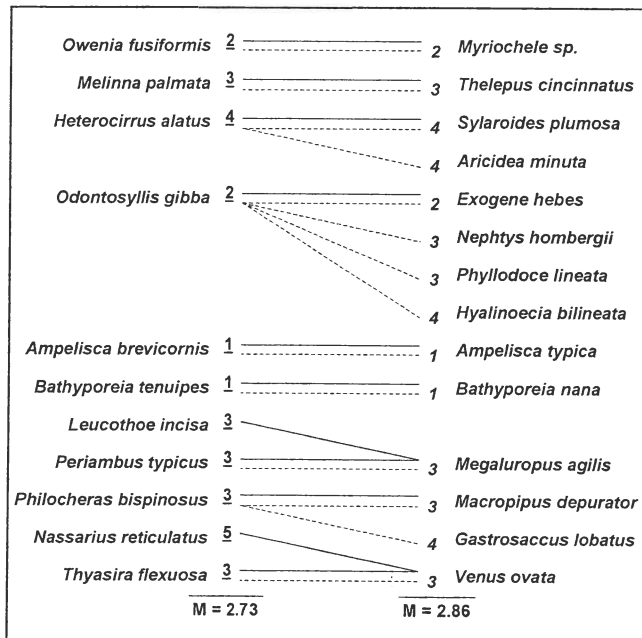


Fig. 16.3. Amoco-Cadiz oil spill {A}. MDS plots from 21 samples (approximately quarterly) of macrobenthos in the Bay of Morlaix (Bray-Curtis on 4th-root transformed abundances). a) As Fig. 16.1 but discarding the rare species, leaving 125; b)–f) based on a succession of five, small, mutually exclusive subsets of species, generated by the BVSTEP procedure, showing the high level of matching with the full data (ρ values in bottom right of plots, and number of species in top right); g)–i) after successive removal of the species in previous plots, the ability to match the original pattern by selecting from the remaining species rapidly degrades (stress = 0.09, 0.08, 0.08, 0.08, 0.12, 0.12, 0.21, 0.24, 0.24 respectively).

Table 16.2. Amoco-Cadiz oil spill {A}. Illustration of taxonomic mapping of the second and third 'peeled' species subsets (i.e. those underlying Fig. 16.3c,d), from the successive application of BVSTEP, highlighting the (closer than random) taxonomic parallels between the species sets which are capable of 'explaining' the full pattern of Fig. 16.3a. Continuous lines represent the closest relatives in the right-hand set to each species in the left-hand set (underlined values are the number of steps distant through the taxonomic tree, see Chapter 17 for examples). Dashed lines map the right-hand set to the left-hand (non-underlined values are again the taxonomic distances). The taxonomic mapping similarity coefficient, M , averages the displayed mean taxonomic distances.



Viewed at a pragmatic level, the message of Fig. 16.3 is therefore clear. It is not a single, small set of species which is responsible for generating the observed sample patterns of Fig. 16.1, of disturbance and (partial) recovery superimposed on a seasonal cycle. Instead, the same temporal patterns are imprinted several times in the full species matrix. The steady increase in size of successive 'peeled' sets reflects the different *signal-to-noise ratios* for different species, or groups of species. The signal can be reproduced by only a few species initially but, as these are sequentially removed, the remaining species have increasingly higher 'noise' levels, requiring an ever greater number of them to generate the same strength of 'signal'. Clarke and Warwick (1998a) give further macrobenthic examples, of time series from Northumberland subtidal sites, whose structural redundancy is at a similar level (4–5 peeled subsets), though this is by no means a universal phenomenon (M G Chapman, pers. comm., for rocky shore assemblages; Clarke and Gorley, 2001, for zooplankton communities).

FURTHER EXTENSIONS

Both BIO-ENV and BVSTEP procedures can be generalised to accommodate possibilities other than their 'defaults' of selecting abiotic variables to optimise a match with fixed biotic similarities, and selecting subsets of species to link to the sample patterns of the full species set. In fact, the main distinction between the two concepts is simply one of whether a full search is performed, as in BIO-ENV, or, for large scale problems, a stepwise search approach is adopted, as in BVSTEP.

The first (fixed) similarity matrix can be from species (e.g. Bray-Curtis), environmental variables (e.g. Euclidean), or even a *model matrix*, such as the equally-spaced inter-point distances in the seriation matrix of Chapter 15. The second matrix, whose variables are to be selected from, can also be of biotic or abiotic form. Some possible applications involve searching for:[¶]

- 1) species within one faunal group that 'best explain' the pattern of a different faunal group ('BIO-BIO'), e.g. key macrofaunal species which are structuring (or are correlated with environmental variables that are structuring) the full meiofaunal assemblages;
- 2) species subsets which best respond to (characterise) a given gradient of one or more observed contaminants ('ENV-BIO');
- 3) species subsets which match a given spatial or temporal pattern ('MODEL-BIO'), e.g. the model might be the geographic layout of samples, expressed literally as inter-sample distances, or a linear time-trend (equal-spaced steps, as with seriation), or a circular pattern appropriate to a single seasonal cycle, etc;
- 4) subsets of environmental variables which best characterise an *a priori* categorisation of samples ('MODEL-ENV'), e.g. selecting quantitative beach morphology variables which best delineate a given classification of beach types (F. Valesini, pers. comm.).

[¶] All these combinations are possible in both the PRIMER BIO-ENV (full search) and BVSTEP (stepwise search) routines. The fixed (dis)similarity/distance matrix can be of any form (biotic, abiotic or model, created within PRIMER or imported from outside) and the data matrix from which variables are to be selected has basically two options for (dis)similarity calculation, which the user has to specify (Bray-Curtis for species, or normalised/non-normalised Euclidean for environment). Often, of course, it will not make sense to run a full search for a species data matrix; conversely, it will not be optimal to run a stepwise search when there are only a small number of variables in an environmental array.

SECOND-STAGE MDS

It is not normally a viable sampling strategy, for soft-sediment benthos at least, to use the BVSTEP procedure to identify a subset of species as the only ones whose abundance is recorded in future. Savings of monitoring effort at the identification analysis stage can sometimes be made, however, by working at a higher taxonomic level than species (see Chapter 10). Where full species-based information is available, MDS plots can be generated at different levels of taxonomic aggregation (i.e. using species, genera, families, etc) and the configurations visually compared. Another axis of choice for the biologist is that of the transformation applied to the original counts (or biomass, or cover). Chapter 9 shows that different transformations pick out different components of the assemblage, from only the dominant species (no transform), through increasing contributions from mid-abundance and less-common species ($\sqrt{\quad}$, $\sqrt[4]{\quad}$, log) to a weighting placing substantial attention on less-common species (presence/absence). The environmental impact, or other spatial or temporal 'signal', may be clearer to discern from the 'noise' under some transformations than it is for others.

Amoco-Cadiz oil spill

The difficulty quickly arises that so many MDS plots are generated that visual comparison is no longer easy, and it is always subjective, relying only on the 2-d approximation inherent in an MDS plot, rather than the full high-dimensional information. For example, Fig. 16.4 displays the MDS plots for the Morlaix study at only two taxonomic levels: data at species and aggregated to family level, for each of the full range of transformations, but it is already difficult to form a clear summary of the relative effects of the different choices. However, part of the solution to this problem has already been met earlier in the chapter. For every pair of MDS plots – or rather the similarity matrices that underlie them – it is easy to define a measure of how closely the sample patterns match: it is the Spearman rank correlation (ρ) applied to the elements of the similarity matrices. Different transformations and aggregation levels will affect the absolute range of calculated Bray-Curtis similarities but, as always, it is their relative values that matter. If all statements of the form 'sample A is closer to B than it is to C' are identical for the two similarity matrices then the conclusions of the analyses will be identical, the MDS plots will match perfectly and ρ will take the value 1.

Table 16.3 shows the results of calculating the rank correlations (ρ) between every pair of analysis options

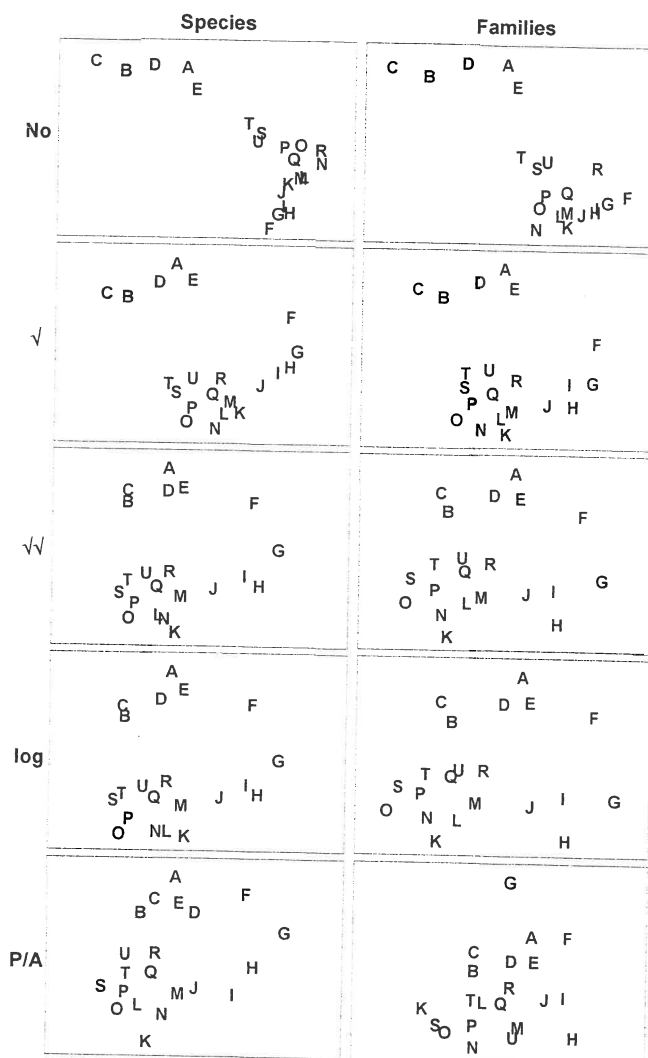


Fig. 16.4. Amoco-Cadiz oil spill {A}. MDS plots of the 21 sampling occasions (A, B, C, ...) in the Bay of Morlaix, for all macrobenthic species (left) and aggregated into families (right), and for different transformations of the abundances (in top to bottom order: no transform, root, 4th-root, $\log(1+x)$, presence/absence). For precise dates see the legend to Fig. 10.4; the oil-spill occurred between E and F (stress, reading left to right: 0.06, 0.07; 0.07, 0.08; 0.09, 0.10; 0.09, 0.09; 0.14, 0.18).

represented in Fig. 16.4. For example, the largest correlation is 0.996 for untransformed species and family-level analyses, the smallest is 0.639 between untransformed and presence/absence family-level analyses, etc. Though Table 16.3 is clearly a more quantitatively objective description of the pairwise comparisons between analyses, the plethora of coefficients still make it difficult to extract the overall message. Looking at the triangular form of the table, however, the reader can perhaps guess what the next stage is! Spearman correlations are themselves a type of similarity measure: two analyses telling essentially the same story have a higher ρ (high similarity) than two analyses giving very different pictures (low ρ , low similarity). All that needs adjustment is the similarity scale, since correlations can potentially take values

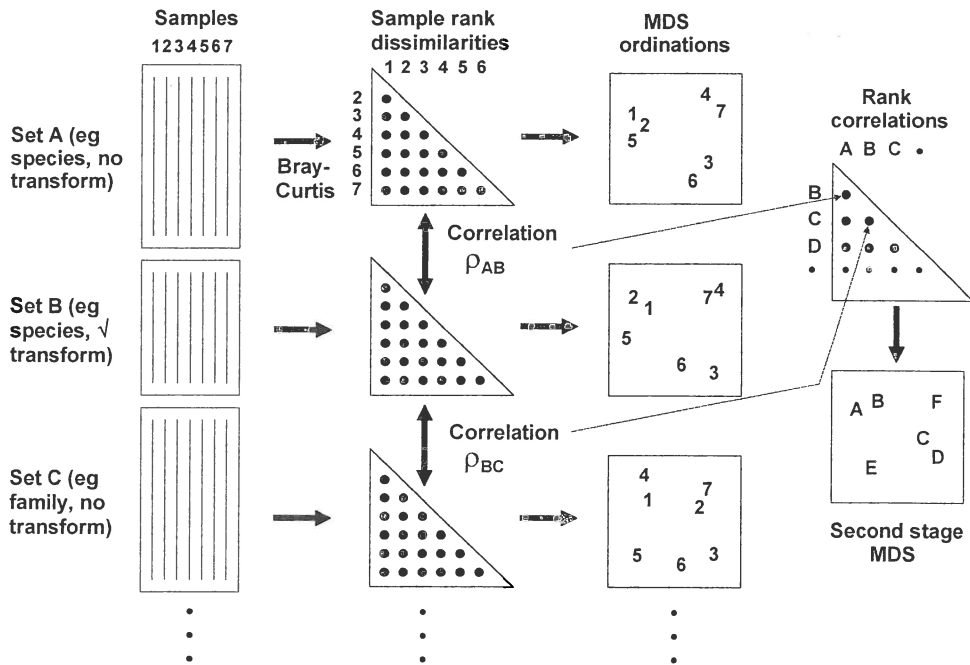


Fig. 16.5. Schematic diagram of the stages in quantifying and displaying agreement, by second-stage MDS, of different multivariate analyses of a corresponding set of samples.

Table 16.3. Amoco-Cadiz oil spill {A}. Spearman correlation matrix between every pair of similarity matrices underlying the 10 plots of Fig. 16.4, measuring the extent to which they 'tell the same story' about the 21 Morlaix samples. These correlations (rank ordered) are treated like a similarity matrix and input to a second-stage MDS. Key: s = species-level analysis, f = family-level; 0 = no transform, 1 = root, 2 = 4th root, 3 = log(1+x), 4 = presence/absence.

	s0	s1	s2	s3	s4	f0	f1	f2	f3
s1	.970								
s2	.862	.949							
s3	.852	.942	.995						
s4	.736	.847	.961	.946					
f0	.996	.965	.855	.845	.726				
f1	.949	.993	.961	.958	.865	.947			
f2	.791	.893	.972	.974	.953	.785	.924		
f3	.760	.869	.962	.971	.946	.753	.904	.993	
f4	.645	.756	.877	.870	.923	.639	.792	.946	.929

in (-1, 1) rather than (0,100) say. In practice, negative correlations in this context will be rare (but if they arise they indicate even less similarity of the two pictures) but the problem is entirely solved anyway by working, as usual, with the ranks of the ρ values, i.e. rank (dis)similarities. It is then natural to input these into an MDS ordination, as shown schematically in Fig. 16.5.

The resulting picture is termed a *second-stage MDS* and is displayed in Fig. 16.6 for the Morlaix analyses of Fig. 16.4. The relationship between the various analysis options is now summarised in a clear and straightforward fashion (with near-zero stress). The different transformations form the main (left to right) axis, in steady progression through: no transform, $\sqrt{\quad}$, $\sqrt[4]{\quad}$ and $\log(1+x)$, to pres/abs. The difference between

species and family level analyses largely forms the other (bottom to top) axis. Three important points are immediately clear:

- 1) Log and $\sqrt[4]{\quad}$ transforms are virtually identical in their effect on the data, with differences between these transformations being much smaller than that between species and family-level analyses in that case.
- 2) With the exception of these two, the transformations generally have a much more marked effect on the outcome than the aggregation level (the relative distance apart on the MDS of the points representing different transformations, but the same taxonomic level, is much greater than the distance apart of species and family-level analyses, for the same transformation).

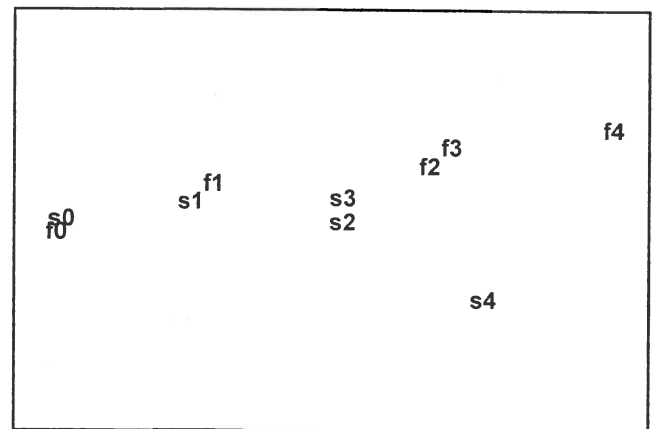


Fig. 16.6. Amoco-Cadiz oil spill {A}. Second-stage MDS of the 10 analyses of Fig. 16.4. The proximity of the points indicates the extent to which different analysis options capture the same information. s = species-level analysis, f = family-level; 0 = no transform, 1 = root, 2 = 4th root, 3 = log(1+x), 4 = presence/absence. Stress = 0.01, so the 2-d picture tells the whole story, e.g. that choice of aggregation level has less effect here than transformation.

- 3) The effect of taxonomic aggregation becomes greater as the transformation becomes more severe, so that for presence/absence data the difference between species and family-level is much more important than it is for untransformed or mildly transformed counts. Whilst this is not unexpected, it does indicate the necessity to think about analysis choices in combination, when designing a study.

Other applications

The concept of a *second-stage MDS* used on rank correlations between similarity matrices – from different taxonomic aggregation levels (species, genus, family, trophic group) and, in the same analysis, different faunal groups (nematodes, macrofauna) recorded for the same set of sites – was introduced by Somerfield and Clarke (1995), for studies in Liverpool Bay and the Fal estuary, UK. Olsgard *et al* (1997, 1998) expanded the scope to include the effects of different transformation, simultaneously with differing aggregation levels, for data from N Sea oilfield studies.[¶] Other interesting applications include Kendall and Widdicombe (1999) who examined different body-size components of the fauna as well as different faunal groups, from a hierarchical spatial sampling design (spacings of 50cm, 5m, 50m, 500m) in Plymouth subtidal waters. They used a second-stage MDS to display the effects of different combinations of body-sizes, faunal groups and transformation. Olsgard and Somerfield (2000) introduced the pattern from environmental variables as an additional point on a second-stage MDS, together with biotic analyses from different faunal components (polychaetes, molluscs, crustacea, echinoderms) at another N Sea oilfield. The idea is that biotic subsets whose multivariate pattern links well to the environmental data will be represented by points on the second-stage MDS which lie close to the environmental point. The converse operation can also be envisaged, as a visual counterpart to the BIO-ENV procedure. For small numbers of environmental variables, the abiotic patterns from subsets of these can be represented as points on the second-stage MDS, in which the (fixed) biotic similarity matrix is also shown. The best environmental combinations should then ‘converge’ on the (single) biotic point.

[¶] They also carried out another interesting analysis, assessing BIO-ENV results in the light of analysis choices. It was hypothesised earlier (p9-4 and 10-2), that a contaminant impact may manifest itself more clearly in the assemblage pattern for intermediate transform and aggregation choices. Olsgard *et al* (1997) do indeed show, for the Valhall oilfield, that the BIO-ENV matching of sediment macrobenthos to the degree of disturbance from drilling muds disposal (measured by sediment THC, Ba concentrations etc), was optimised by intermediate transform ($\sqrt{}$) and aggregation level (family).

EXAMPLE: Phuket coral-reef time series

A rather different application of second-stage MDS[¶] is motivated by considering the two-way layout from a time-series of coral-reef assemblages, along an onshore-offshore transect in Ko Phuket, Thailand $\{K\}$. These data were previously met in Chapter 15, where only samples from the earlier years 1983, 86, 87, 88 were considered (as available to Clarke *et al*, 1993). The time series has been subsequently expanded to the 13 years 1983–2000, omitting 1984, 85, 89, 90 and 96, on transect A (Brown *et al*, in press). This transect consisted of 12 equally-spaced positions along the onshore-offshore gradient, and was subject to sedimentation disturbance from dredging for a new deep-water port in 1986 and 87. For 10 months during late 1997 and 98 there was also a wide scale sea-level depression in the Indian Ocean, leading to significantly greater irradiance exposures at mid-day low tides. Elevated sea temperatures were also observed (in 1991, 95, 97, 98), sometimes giving rise to coral bleaching events, but these generally resulted in only short-term partial mortalities.

The two (crossed) factors here are the years and the positions down the transect (1–12, at the same spacing each year). Separate MDS plots of the onshore-offshore (seriation) pattern for each year show some visual differences which can be summarised in Spearman rank correlations (ρ) between their underlying similarity matrices. These correlations in turn are entered into the second-stage MDS to produce Fig. 16.7; it demonstrates both the sedimentation-based disruption to the gradient in 1986 and 87, and the negative sea-level anomaly of 1998. Interestingly, these are on opposite sides of the MDS plot, suggesting that the departures from the ‘normal’ onshore-offshore gradient are of a different type.

Note the subtlety of what this analysis tries to isolate: the compositions of the transect over the different years are not directly compared, as they would be if all were analysed together on the same (first-stage) MDS, for example. There may (and will) be natural year-to-year fluctuations in abundances which would separate the transects on a combined MDS plot but which do not disrupt the serial change in assemblage down the transect. The second-stage procedure will not be sensitive to such natural fluctuations. In fact, it eliminates them

[¶] Both types of problem are catered for in the PRIMER 2STAGE routine, the inputs either being a series of similarity matrices (which can be taken from any source provided they refer to the same set of sample labels), or a single similarity matrix, from a 2-way layout with appropriate factors defined.

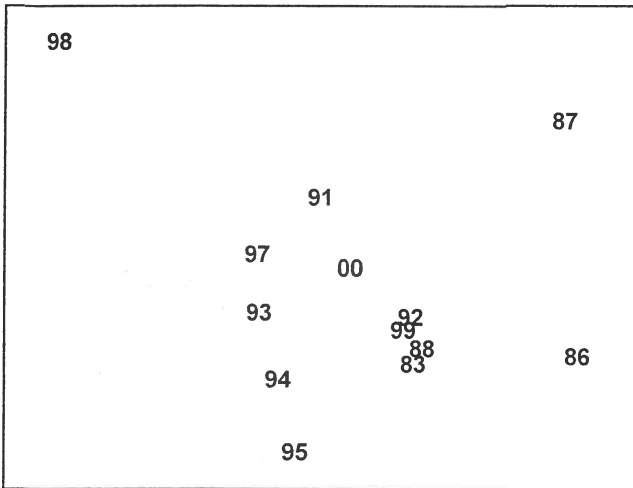


Fig. 16.7. *Ko Phuket corals {K}. Second-stage MDS plot of the 13 different years in the period 1983 to 2000, based on comparing the multivariate analyses, for each year, of the 12 transect positions down the shore (transect A). Note the differences in the onshore-offshore seriation pattern in 1986/7 and again in 1998, evidenced by the separation of these years on the plot (stress = 0.11).*

by concentrating only on whether the (gradient) *pattern* is the same each year – assemblage similarities between the same transect points in different years do not enter the calculations (as seen in Fig. 16.5). Disruptions to the gradient pattern in certain years are, in a sense, *interactions* between gradient and year, removing year-to-year *main effects*, and this is what the second-stage MDS sets out to display in such a context.[†]

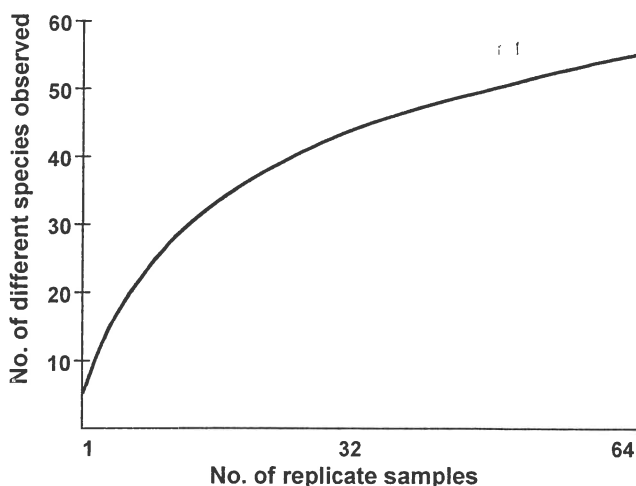
[†] The idea has parallels with the ANOSIM2 procedure at the end of Chapter 6, which sets out to remove the main effects of factor A in a two-way layout, by concentrating only on the patterns of factor B, in separate MDS plots for each level of factor A.

CHAPTER 17: BIODIVERSITY MEASURES BASED ON RELATEDNESS OF SPECIES

SPECIES RICHNESS DISADVANTAGES

Chapter 8 discussed a range of diversity indices based on species richness and the species abundance distribution. Richness (S) is widely used as the preferred measure of biological diversity (*biodiversity*) but it has some major drawbacks, many of which apply equally to other diversity indices such as H' , H , J' , etc.

1) **Observed richness is heavily dependent on sample size/effort.** In nearly all marine contexts, it is not possible to collect exhaustive census data. The assemblages are sampled using sediment cores, trawls etc, and the 'true' species richness of a station is rarely fully represented in such samples. For example, Gage and Coghill (1977) describe a set of contiguous core samples taken for macrobenthic species in a Scottish sea-loch. A *species-area plot* (or accumulation curve) which illustrates how the number of different species detected increases as the samples are accumulated[†], shows that, even after 64 replicate samples are taken at this single locality, the observed number of species is still rising.



“The harder you look, the more species you find” is fundamental to much biological sampling and the asymptote of accumulation curves is rarely reached. Observed species richness S is therefore

[†] This uses the Species-Area plot routine in PRIMER, which has the option of plotting the curve in the presented sample order or (as here) randomising that order a large number of times. In the latter case, the resulting curves are averaged to obtain the smoothed relationship of average number of species for each number of replicates (sample size).

highly sensitive to sample size and totally non-comparable across studies involving unknown, uncontrolled or simply differing degrees of sampling effort. The same is true, to a lesser extent, of many other standard diversity indices. Fig. 17.1 shows the effect of increasing numbers of individuals on the values of some of the diversity indices defined in Chapter 8. This is a sub-sampling study, selecting different numbers of individuals at random from a single, large community sample. The only index to demonstrate a lack of bias in mean value is Simpson diversity, given here in the form $1-\lambda'$, see equation (8.4). Comparison of richness, Shannon, evenness, Brillouin etc values for differing sample sizes is clearly problematic.

- 2) **Species richness does not directly reflect phylogenetic diversity.** “A measure of biodiversity of a site ought ideally to say something about how different the inhabitants are from each other” (Harper and Hawksworth, 1994). It is clear that a sample consisting of 10 species from the same genus should be seen as much less *biodiverse* than another sample of 10 species, all of which are from different families: genetic, phylogenetic or, at least, taxonomic relatedness of the individuals in a sample is the key concept which is developed in this chapter, into practical indices which genuinely reflect biodiversity and are robust to sampling effort variations.
- 3) **No statistical framework exists for departure of S from ‘expectation’.** Whilst observed species richness measures can be compared across sites (or times) which are subject to strictly controlled and equivalent sampling designs, there is no sense in which the values of S can be compared with some absolute standard, i.e. we cannot generally answer the question “what do we expect the richness to be at this site?”, in the absence of anthropogenic impact, say.
- 4) **The response of S to environmental degradation is not monotonic.** Chapter 8 discusses the well-established paradigm (see Wilkinson, 1999, and references therein) that, under moderate levels of disturbance, species richness may actually increase, before decreasing again at higher impact levels. It would be preferable to work with a biological index whose relation to the degree of perturbation was purely monotonic (increasing or decreasing, but not both).

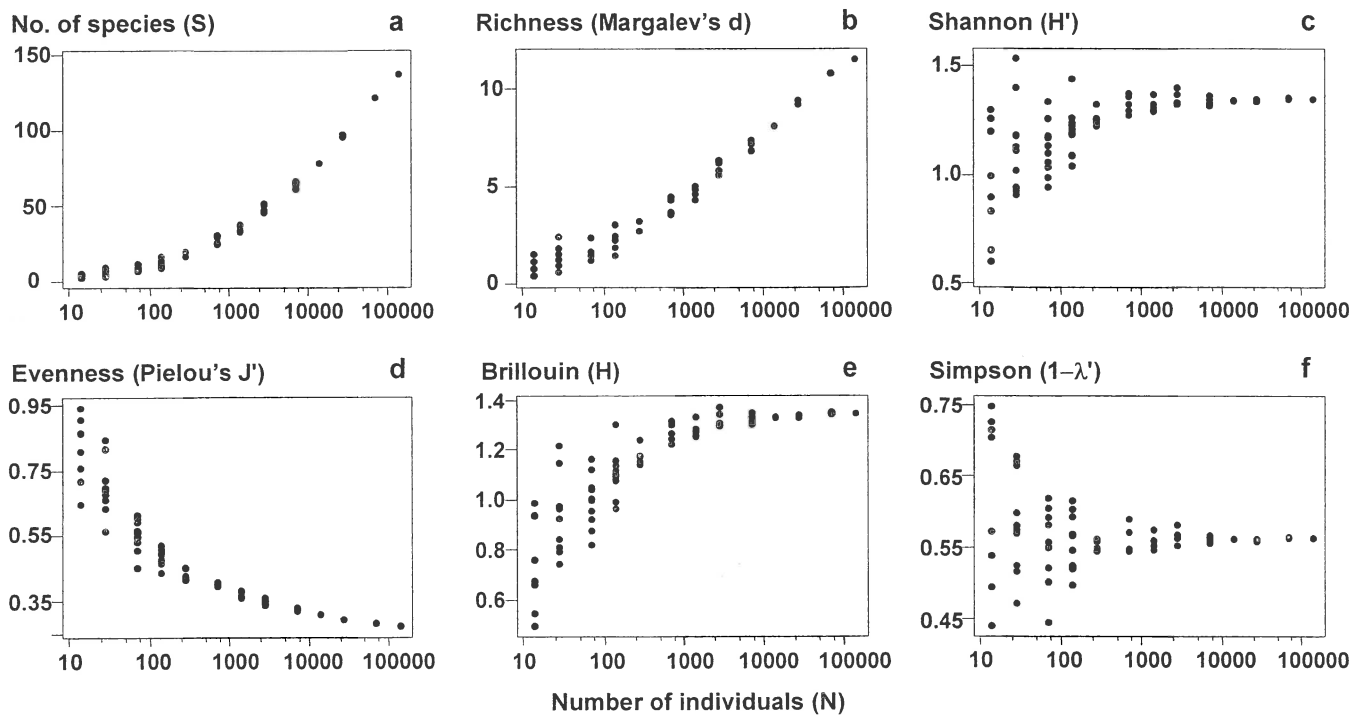


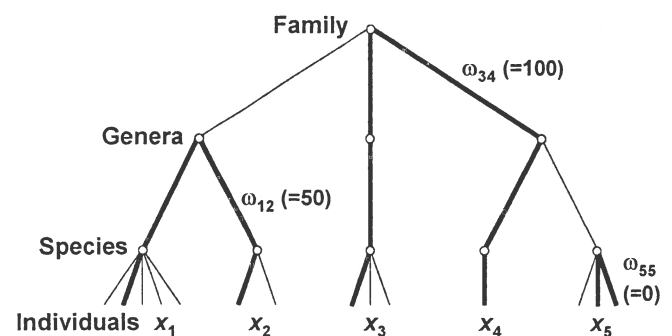
Fig. 17.1. Amoco-Cadiz oil spill {A}, pooled pre-impact data. Values of 6 standard diversity indices (y-axis, see Chapter 8 for definitions), for simulated samples of increasing numbers of individuals (x-axis, log scaled), drawn randomly without replacement from the full set of 140,344 macrobenthic organisms.

5) **Richness can vary markedly with differing habitat type.** Again, the ideal would be a measure which is less sensitive to differences in natural environmental variables but is responsive to anthropogenic disturbance.

AVERAGE TAXONOMIC DIVERSITY AND DISTINCTNESS

Two measures, which address some of the problems identified with species richness and the other diversity indices, are defined by Warwick and Clarke (1995a). They are based not just on the species abundances (denoted by x_i , the number of individuals of species i in the sample) but also the **taxonomic distances** (ω_{ij}), through the classification tree, between every pair of individuals (the first from species i and the second from species j). For a standard Linnean classification, these are discrete distances, the simple tree below illustrating path lengths of zero steps (individuals from the same species), one step (same genus but different species) and two steps (different genera)[¶]. Clarke and Warwick (1999) advocate a simple linear scaling whereby the largest number of steps in the

tree (two species at greatest taxonomic distance apart) is set to $\omega = 100$. Thus, for a sample consisting only of the 5 species shown, the path between individuals in species 3 and 4 is $\omega_{34} = 100$, between species 1 and 2 is $\omega_{12} = 50$, between two individuals of species 5 is $\omega_{55} = 0$, etc.



Average taxonomic diversity of a sample is then defined (Warwick and Clarke, 1995a) as:

$$\Delta = [\sum \sum_{i < j} \omega_{ij} x_i x_j] / [N(N-1)/2] \quad (17.1)$$

where the double summation is over all pairs of species i and j ($i, j = 1, 2, \dots, S; i < j$), and $N = \sum_i x_i$, the total number of individuals in the sample. Δ has a simple interpretation: it is the average 'taxonomic distance apart' of every pair of individuals in the sample or, to put it another way, the expected path length between any two individuals chosen at random.

[¶] The principle extends naturally to a phylogeny with continuously varying branch lengths and even, ultimately, to a molecular-based genetic distance between individuals (of the same or different species), see Clarke and Warwick (2001), Fig. 1.

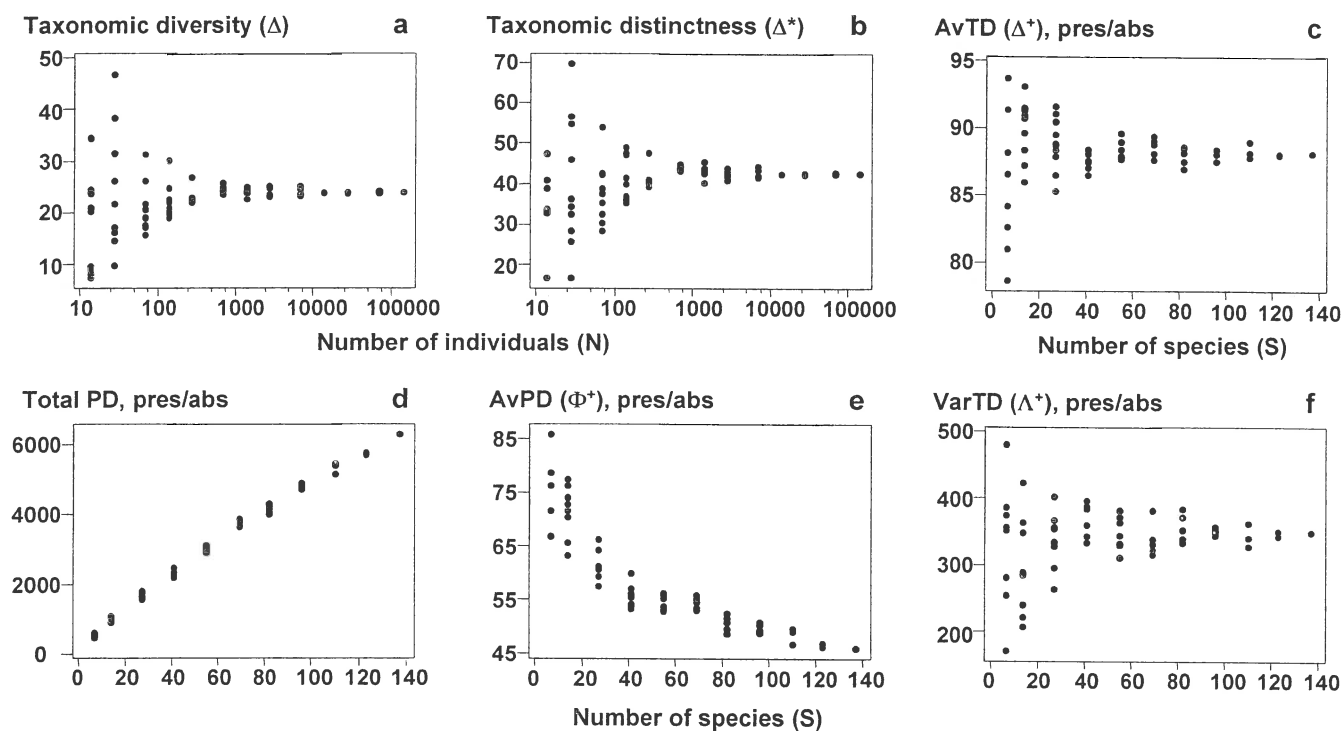


Fig. 17.2. *Amoco-Cadiz oil spill* {A}, pooled pre-impact data. a), b) Quantitative indices (y-axis): Average taxonomic diversity (Δ) and distinctness (Δ^*) for random subsets of fixed numbers of individuals (x-axis, logged), drawn randomly from the pooled sample, as in Fig. 17.1. c)–f) List-based (presence/absence) indices (y-axis): Average taxonomic distinctness (Δ^+), total phylogenetic diversity (PD), average phylogenetic diversity (Φ^+) and Variation in taxonomic distinctness (Λ^+), for random subsets of fixed numbers of species (x-axis) drawn from the full species list for the pooled sample. The sample-size independence of TD-based indices is clear, contrasting with PD and most standard diversity measures (Fig. 17.1).

Note also that when the taxonomic tree collapses to a single-level hierarchy (all species in the same genus, say), Δ becomes

$$\begin{aligned} \Delta^\circ &= [2 \sum_{i < j} p_i p_j] / (1 - N^{-1}), \text{ where } p_i = x_i / N, \\ &= (1 - \sum_i p_i^2) / (1 - N^{-1}) \end{aligned} \quad (17.2)$$

which is a form of Simpson diversity. The Simpson index is actually defined from the probability that any two individuals selected at random from a sample belong to the same species (Simpson, 1949). Δ is therefore seen to be a natural extension of Simpson, from the case where the path length between individuals is either 0 (same species) or 100 (different species) to a more refined scale of intervening relatedness values (0 = same species, 20 = different species in the same genera, 40 = different genera but same family, etc).[†] It follows that Δ will often track Simpson diversity fairly closely. To remove the dominating effect of the species abundance distribution $\{x_i\}$, leaving a measure which is more nearly a pure reflection of the taxonomic hierarchy, Warwick and Clarke (1995a) proposed dividing Δ by the Simpson index Δ° , to give *average taxonomic distinctness*

[†] In addition, there is a relationship between Δ and Simpson indices computed at all higher taxonomic levels, as recently reported by Shimatani (2001)

$$\Delta^* = [\sum \sum_{i < j} \omega_{ij} x_i x_j] / [\sum \sum_{i < j} x_i x_j] \quad (17.3)$$

Another way of thinking of this is as the expected taxonomic distance apart of any two individuals chosen at random from the sample, provided those two individuals are not from the same species.

A further form of the index, exploited greatly in what follows, takes the special case where quantitative data is not available and the sample consists simply of a species list (presence/absence data). Both Δ and Δ^* reduce to the same coefficient

$$\Delta^+ = [\sum \sum_{i < j} \omega_{ij}] / [S(S-1)/2] \quad (17.4)$$

where S , as usual, is the observed number of species in the sample and the double summation ranges over all pairs i and j of these species ($i < j$). Put simply, the average taxonomic distinctness (*AvTD*) Δ^+ of a species list is the average taxonomic distance apart of all its pairs of species. This is a very intuitive definition of biodiversity, as average *taxonomic breadth* of a sample.

Sampling properties

For quantitative data, repeating the pairwise exercise (Fig. 17.1) of random subsampling of individuals from a single, large sample, Fig. 17.2a and b show that both

taxonomic diversity (Δ) and average taxonomic distinctness (Δ^*) inherit the sample-size independence seen in the Simpson index, from which they are generalised. Clarke and Warwick (1998b) formalise this result by showing that, whatever the hierarchy or subsample size, Δ is exactly unbiased and Δ^* is close to being so (except for very small subsamples). For non-quantitative data (a species list), the corresponding question is to ask what happens to the values of Δ^+ for random subsamples of a fixed number of *species* drawn from the full list. Fig. 17.2c demonstrates that the mean value of Δ^+ is unchanged, its exact unbiasedness in all cases again being demonstrated in Clarke and Warwick (1998b). This lack of dependence of Δ^+ (in mean value) on the number of species in the sample has far-reaching consequences for its use in comparing historic data sets and other studies for which sampling effort is uncontrolled, unknown or unequal.

EXAMPLES: Ekofisk oil-field and Tees Bay soft-sediment macrobenthos

The earlier Fig. 14.4 demonstrated a change in the sediment macrofaunal communities around the Ekofisk oil-field {E}, out to a distance of about 3 km from the centre of drilling activity. This was only evident, however, from the multivariate (MDS and ANOSIM) analyses, not from univariate diversity measures such as Shannon H' , where reduced diversity was only apparent up to a few hundred metres from the centre (Fig. 17.3a). The implication is that the observed community change resulted in no overall loss of diversity but this is not the conclusion that would have been drawn from calculating the quantitative average taxonomic distinctness index, Δ^* . Fig. 17.3b shows a clear linear trend of increase in Δ^* with (log) distance from the centre, the relationship only breaking

down into a highly variable response for the strongly impacted sites, within 100m of the drilling activity.

A further example, from the coastal N Sea, is given by a time-series of macrobenthic samples, with data averaged over 6 locations in Tees Bay, UK, (Warwick *et al*, in press). Samples were taken in March and September for each of the years 1973 to 1996, and Fig. 17.4 shows the September inter-annual patterns for four (bio)diversity measures. Notable is the clear increase in Shannon diversity at around 1987/88 (Fig. 17.4b), coinciding with significant widescale changes in the N Sea planktonic system which have been reported elsewhere (e.g. Reid *et al*, in press). However, Shannon diversity is very influenced here by the high numbers of a single abundance dominant (*Spiophanes bombyx*), whose decline after 1987 led to greater equitability in the quantitative species diversity measures. A more far-reaching change, representative of what was happening to the community as a whole, is indicated by looking at the taxonomic relatedness statistics based only on presence/absence data. Use of simple species lists has the advantage here of ensuring that no one species can dominate the contributions to the index. Average taxonomic distinctness (Δ^+) is seen to show a marked decline at about the time of this N Sea regime shift (Fig. 17.4c), indicating a biodiversity loss, a very different (and more robust) conclusion than that drawn from Shannon diversity.

OTHER RELATEDNESS MEASURES

The remainder of this chapter deals only with data in the form of a species list for a locality (presence/absence data). There is a substantial literature on measures incorporating, primarily, phylogenetic relationships amongst species (see references in the review-type papers of Faith, 1994, Humphries *et al* 1995). The

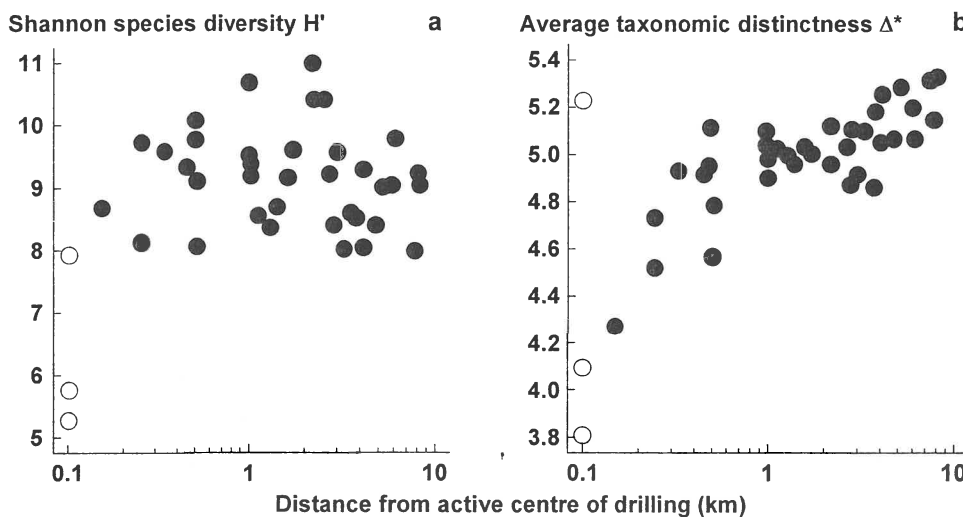


Fig. 17.3. Ekofisk macrobenthos {E}. a) Shannon diversity (H') for the 39 sites (y-axis), plotted against distance from centre of drilling activity (x-axis, log scale). b) Quantitative average taxonomic distinctness Δ^* for the 39 sites, indicating a response trend not present for standard diversity indices.

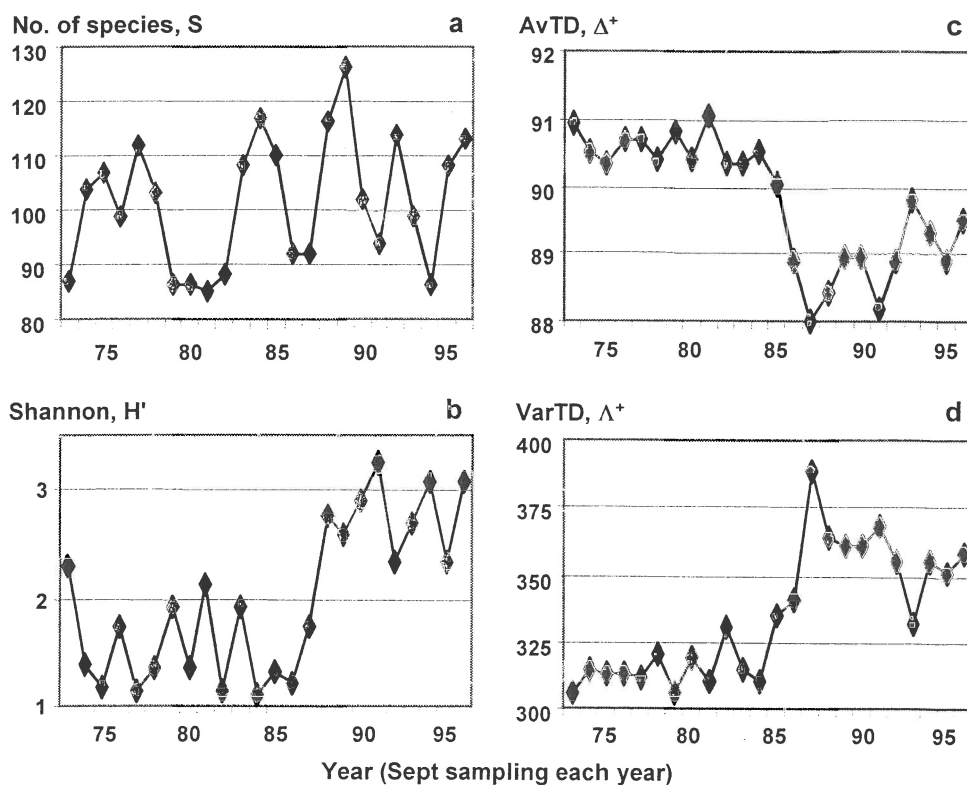


Fig. 17.4. Variations inter-annually in Tees Bay macrobenthos $\{V\}$. (Bio)diversity indices for Tees Bay areas combined, from sediment samples in September each year, over the period 1973–96, straddling a major regime-shift in N Sea ecosystems, about 1987. a) Richness, S ; b) Shannon, H' ; c) Average taxonomic distinctness, Δ^+ , based on presence/absence and reflecting the mean taxonomic breadth of the species lists; d) Variation in taxonomic distinctness, Λ^+ (also pres/abs), reflecting unevenness in the taxonomic hierarchy.

context is conservation biology, with the motivation being the selection of individual species, or sets of species (or reserves), with the highest conservation priority, based on the unique evolutionary history they represent, or their complementarity to existing well-conserved species (or reserves). Warwick and Clarke (2001) draw a potentially useful distinction of terminology between this individual species-focused conservation context and the use, as in this chapter, of relatedness information to monitor differences in community-wide patterns in relation to changing environmental conditions. They suggest that the term taxonomic/phylogenetic *distinctiveness* (of a species) is reserved for weights assigned to individual species, reflecting their priority for conservation; whereas taxonomic/phylogenetic *distinctness* (of a community) summarises features of the overall hierarchical structure of an assemblage (the spread, unevenness etc. of the classification tree).

Phylogenetic diversity (PD)

In the *distinctiveness* context, Vane-Wright *et al* (1991), Williams *et al* (1991) and May (1990) introduced measures based only on the topology ('elastic shape') of a phylogenetic tree, appropriate when branch lengths are entirely unknown, and Faith (1992, 1994) defined a *phylogenetic diversity (PD)* measure based on known branch lengths: PD is simply the cumulative branch length of the full tree. Whether this is thought of as representing the total evolutionary history, the genetic

turnover or morphological richness, it is an appealingly simple statistic. Unfortunately, Fig. 17.5 demonstrates some of the disadvantages of using these measures in a *distinctness* context. The figure compares only samples (lists) with the same number of species (7), at four hierarchical levels (say, species within genera within families, all in one order), so that each step length is set to 33.3. Fig. 17.5b and c have the same tree topology, yet we should not consider them to have the same average (or total) distinctness, since each species is more taxonomically similar to its neighbours in b than c (reflected in Δ^+ values of 33.3 and 66.6 respectively). Similarly, contrasting Fig. 17.5d and e, the total PD is clearly identical, the sum of all the branch lengths being 333 in both cases, but this does not reflect the more equitable distribution of species amongst higher taxa in d than e (Δ^+ does, however, capture this intuitive element of biodiversity, with respective values of 52 and 43).

Average PD

More importantly, there is another clear reason why phylogenetic diversity PD is unsuitable for monitoring purposes. Firstly, note that PD itself is a *total* rather than average property; as new species are added to the list it always increases. This makes PD highly dependent on species richness S and thus sampling effort, a demonstration of which can be seen in Fig. 17.2d (and the later Fig. 17.9a), a near straight line relationship of PD with S . This is to be expected, and

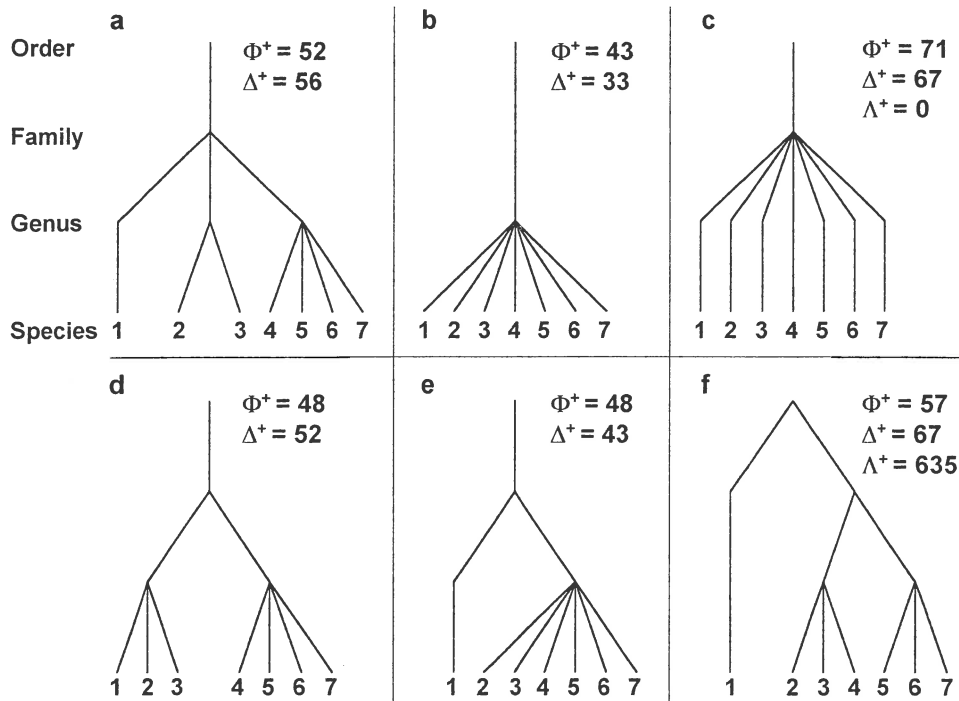


Fig. 17.5. a)-f) Example taxonomic hierarchies for presence/absence data on 7 species (i.e. of fixed species richness), with 4 levels and 3 step lengths (thus each of 33.3, though the third step only comes into play for plot f). Φ^+ : average phylogenetic diversity, Δ^+ : average taxonomic distinctness, Λ^+ : variation in TD. The plots show, inter alia: the expected 'biodiversity' decrease from a) to d) and e) to b) (in both Δ^+ and Φ^+), and from d) to e) (but only in Δ^+ , not in Φ^+); unevenness of f) in relation to c), reflected in increased Λ^+ though unchanged Δ^+ .

a better equivalent to average taxonomic distinctness (AvTD , Δ^+) would be *average phylogenetic diversity* (AvPD), defined as the ratio:

$$\Phi^+ = \text{PD} / S \quad (17.5)$$

This is a very intuitive summary of average distinctness, being the contribution that each species makes on average to the total tree length, but unfortunately it does not have the same lack of dependence on sampling effort that characterises Δ^+ . Fig. 17.2e (and the later Fig. 17.9b) show that its value decreases markedly as the number of species (S) increases, making it misleading to compare AvPD values across studies with differing levels of sampling effort.

'Total' versus 'average' measures

Note the distinction here between *total* and *average* distinctness measures. AvPD (Φ^+) is the analogue of AvTD (Δ^+), both being ways of measuring the *average* taxonomic breadth of an assemblage (a species list), for a given number of species. Δ^+ will give the same value (on average) whatever that number of species; Φ^+ will not. Total PD measures the *total* taxonomic breadth of the assemblage and has a direct analogue in *total taxonomic distinctness*:

$$\text{TTD} = S \cdot \Delta^+ = \sum_i [(\sum_{j \neq i} \omega_{ij}) / (S-1)] \quad (17.6)$$

Explained in words, this is the average taxonomic distance from species i to every other species, *summed* over all species, $i = 1, 2, \dots, S$. (Taking an average rather than a sum gets you back to AvTD , Δ^+ .) TTD

may well be a useful measure of total taxonomic breadth of an assemblage, as a modification of species richness which allows for the species inter-relatedness, so that it would be possible, for example, for an assemblage of 20 closely-related species to be deemed less 'rich' than one of 10 distantly-related species. In general, however, like total PD, total TD will tend to track species richness rather closely, and will only therefore be useful for tightly controlled designs in which effort is identical for the samples being compared, or sampling is sufficiently exhaustive for the asymptote of the species-area curve to have been reached (i.e. comparison of censuses rather than samples).

Variation in TD

Finally, a comparison of Fig. 17.5c and f shows that the scope for extracting meaningful biodiversity indices (unrelated to richness) from simple species lists has not yet been exhausted. Average taxonomic distinctness is the same in both cases ($\Delta^+ = 66.6$) but the tree constructions are very different, the former having consistent, intermediate taxonomic distances between pairs of species, in comparison with the latter's disparate range of small and large values. This can be conveniently summarised in a further statistic, the variance of the taxonomic distances $\{\omega_{ij}\}$ between each pair of species i and j , about their mean value Δ^+ :

$$\Lambda^+ = [\sum \sum_{i < j} (\omega_{ij} - \Delta^+)^2] / [S(S-1)/2] \quad (17.7)$$

termed the *variation in taxonomic distinctness*, VarTD . Its behaviour in a practical application will be examined

later in the chapter[¶], but note for the moment that it, too, appears to have the desirable sampling property of (approximate) lack of dependence of its mean value on sampling effort (see Fig. 17.2f).

‘EXPECTED DISTINCTNESS’ TESTS

Species master list

The construction of taxonomic distinctness indices from simple species lists makes it possible to address another of the ‘desirable features’ listed at the beginning of the chapter: there *is* a potential framework within which TD measures can be tested for departure from ‘expectation’. This envisages a *master list* or *inventory* of species, within defined taxonomic boundaries and encompassing the appropriate region/biogeographic area, from which the species found at one locality can be thought of as drawn. For example, the next illustration uses the entire British faunal list of free-living marine nematodes, a total of 395 species identified to date. The species complement at any particular locality and/or historic period (e.g. putatively impacted areas such as Liverpool Bay or the Firth of Clyde) can be compared with the master list, to ask whether the observed subset of species is representative of the biodiversity expressed in the full species inventory. Clearly, such a comparison is impossible for species richness S , or *total* TD or PD, since the list at one location is automatically shorter than the master list. Also, comparison of S between different localities (or historic periods) is invalidated by the inevitable differences in sampling effort in constructing the lists for different places (or times). However, the key observation here (Clarke and Warwick, 1998b) is that average taxonomic distinctness (Δ^+) of a randomly selected sublist does not differ, in mean value, from AvTD for the master list. So, localities that have attracted differing degrees of sampling effort are potentially directly comparable, with each other and with Δ^+ for the full inventory. The latter is the ‘expected value’ for average distinctness from a defined faunal group,

and reductions from this level, at one place or time, *can* potentially be interpreted as loss of biodiversity.

Testing framework

Furthermore, there is a natural testing framework for how large a decrease (or increase) from expectation needs to be, in order to be deemed statistically ‘significant’. For an observed set of m species at one location, sublists of size m are drawn at random from the master inventory, and their AvTD values computed. From, say, 999 such simulated sublists, a *histogram* can be constructed of the expected range of Δ^+ values, for sublists of that size, against which the true Δ^+ for that locality can be compared. If the observed Δ^+ falls outside the central 95% of the simulated Δ^+ values, it is considered to have departed significantly from expectation: a two-sided test is probably appropriate since departure could theoretically be in the direction of enhanced as well as reduced distinctness.

The next stage is to repeat the construction of these 95% probability intervals for a range of sublist sizes ($m = 10, 15, 20, \dots$) and plot the resulting upper and lower limits on a graph of Δ^+ against m . When these limit points are connected across the range of m values, the effect is to produce a *funnel plot* (such as seen in Fig. 17.8). The real Δ^+ values for a range of observational studies are now added to this plot, allowing simultaneous comparison to be made of distinctness values with each other and with the ‘expected’ limits.[†]

EXAMPLE: UK free-living nematodes

Warwick and Clarke (1998) examined 14 species lists from a range of different habitats and impacted/undisturbed UK areas ($\{U\}$, Fig. 17.6), referring them to a 6-level classification of free-living, marine nematodes (Lorenzen, 1994), based on cladistic principles. The taxonomic groupings were: species, genus, family, suborder, order and subclass, all within one class, thus giving equal step lengths between adjacent taxonomic levels of 16.67 (species within different subclasses then being at a taxonomic distance of $\omega = 100$). The relatively comprehensive British master list (compiled from Platt and Warwick, 1983, 1988, Warwick *et al* 1998) consisted of 395 species, the individual area/habitat sublists ranging in size from 27 to 164 species.

[¶] The PRIMER DIVERSE routine has options to compute the full range of relatedness-based biodiversity measures discussed in this chapter: Δ , Δ^* , Δ^- , TTD, Λ^- , PD, Φ^- , simultaneously for all the samples in a species matrix. It returns the values to a worksheet that can be displayed as scatter plots, via the Draftsman plot routine, analysed in a multivariate way (e.g. by PCA, with indices as variables) or exported for conventional univariate analysis (e.g. ANOVA) elsewhere. These DIVERSE options require the availability of an **aggregation file**, detailing which species map to which genus, families etc, in exactly the same format needed for the Aggregate routine which was used to perform higher taxonomic level analyses in Chapter 10.

[†] Histogram and funnel plots of the ‘expected’ spread of Δ^+ values for a given subsample size (or size range), drawn from a master species list, are plotted in the PRIMER TAXDTEST routine, accessible when the active sheet is the aggregation file for the master list. An option is given to superimpose a real data value on the simulated histogram, or a set of real values on the funnel plot.

They included two studies of the same (generally impacted) area, the Firth of Clyde, carried out by different workers and resulting in very disparate sublist sizes (53 and 112).

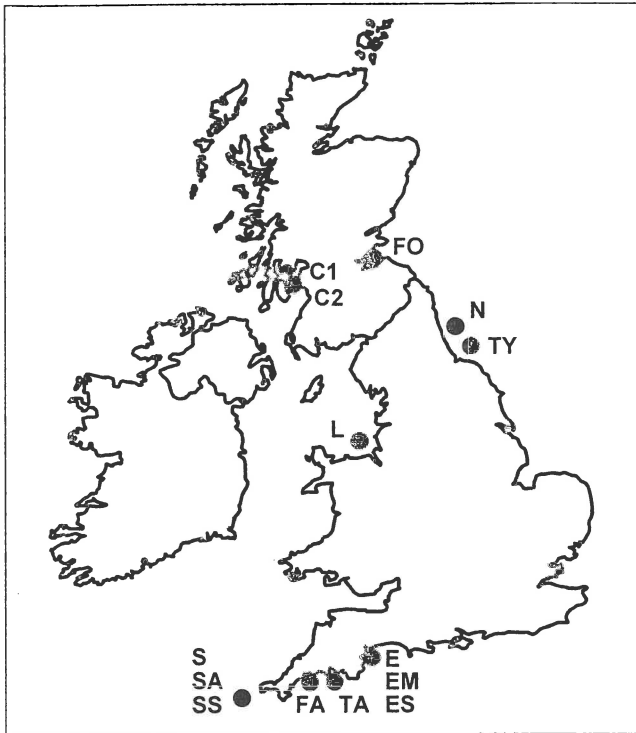


Fig. 17.6. UK regional study, free-living nematodes {U}. The location/habitat combinations for the 14 species sublists whose taxonomic distinctness structure is to be compared. Sublittoral offshore sediments at N: Northumberland (Warwick and Buchanan 1970); TY: Tyne (Sommerfield et al, 1993); L: Liverpool Bay (Sommerfield et al, 1995). Intertidal sand beaches at ES: Exe (Warwick, 1971); C1: Clyde (Lamshead, 1986); C2: Clyde (Jayasree, 1976); FO: Forth (Jayasree, 1976); SS: Scilly (Warwick and Coles, 1977). Estuarine intertidal mudflats at EM: Exe (Warwick, 1971); TA: Tamar (Austen and Warwick, 1989); FA: Fal (Sommerfield et al, 1994a,b). Algal habitats in SA: Scilly (Gee and Warwick, 1994 a,b). Also mixed habitats at E: Exe, S: Scilly.

Histograms

Species richness levels of the 14 lists are clearly not comparable since sampling effort is unequal. However, the studies have been rationalised to a common taxonomy and AvTD values may be meaningfully compared. Fig. 17.7 contrasts two of the studies, which have similar-length species lists: sandy sites in the Exe estuary (ES, 122 species) and the Firth of Clyde (C1, 112 species). Fig. 17.7a displays the histogram of Δ^+ values for 999 random subsamples of size $m = 122$, drawn from the full inventory of 395 species, and this is seen to be centred around the master AvTD of 78.7, with a (characteristic) left-skewness to the Δ^+ distribution. The observed Δ^+ of 79.1 for the Exe data falls very close to this mean, in the body of the distribution, and therefore suggests no evidence of reduced taxonomic distinctness. Fig. 17.7b shows the histogram of simulated Δ^+ values in subsets of size $m = 112$, having (of course) the same mean Δ^+ of 78.7 but, in contrast, the observed Δ^+ of 74.1 for the Clyde data now falls well below its value for any of the randomly selected subsets, demonstrating a significantly reduced average distinctness.

Funnel plots

Fig. 17.8 displays the funnel plot, catering for all sublist sizes. The simulated 95% probability limits are again based on 999 random selections for each of $m = 10, 15, 20, \dots, 250$ species from the 395. The mean Δ^+ is constant for all m (at 78.7) but the limits become increasingly wide as the sample size decreases, reducing the likelihood of being able to detect a change in distinctness (i.e. reducing the *power* of the test). The probability limits also demonstrate the left-skewness of the Δ^+ distribution about its mean throughout, though especially

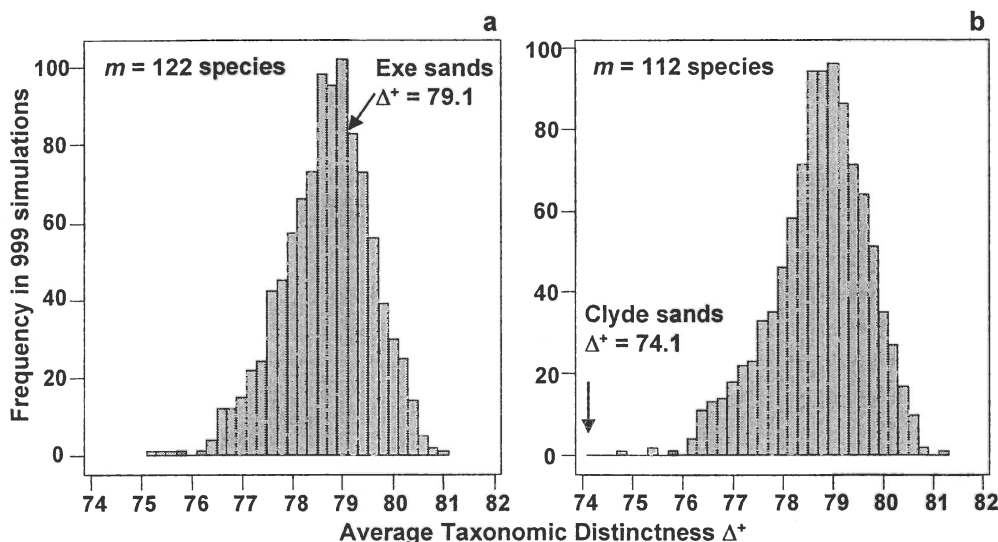


Fig. 17.7. UK regional study, free-living nematodes {U}. Histograms of simulated AvTD, from 999 sublists drawn randomly from a UK master list of 395 species. Sublist sizes of a) $m=122$, b) $m=112$, corresponding to the observed number of species in the Exe (ES) and Clyde (C1) surveys. True Δ^- also indicated: the Exe value is central but the null hypothesis that AvTD for the Clyde equates to that for the UK list as a whole is clearly rejected ($p < 0.001$ or 0.1%)

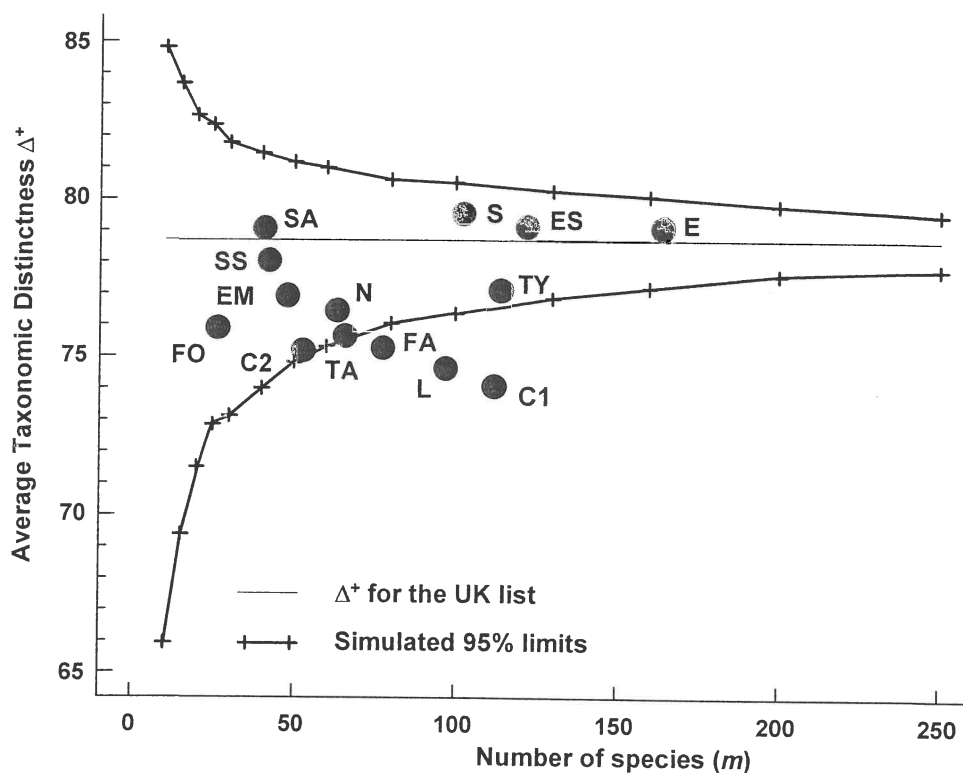


Fig. 17.8. UK regional study, free-living nematodes {U}. Funnel plot for simulated AvTD, as in Fig. 17.7, but for a range of sublist sizes $m=10, 15, 20, \dots, 250$ (x -axis). Crosses, and thick lines, indicate limits within which 95% of simulated Δ^+ values lie; the thin line indicates mean Δ^+ (the AvTD for the master list), which is not a function of m . Points are the true AvTD (y -axis) for the 14 location/habitat studies (see Fig. 17.6 for codes), plotted against their sublist size (x -axis).

for low numbers of species. Superimposing the real Δ^+ values for the 14 habitat/location combinations, five features are apparent:

- 1) The impacted areas of Clyde, Liverpool Bay, Fal and, to a lesser extent, Tamar, are all seen to have significantly reduced average distinctness, whereas pristine locations in the Exe and Scilly have Δ^+ values close to that of the UK master list.
- 2) Unlike species richness (and in keeping with the 'desirability criteria' stated earlier), Δ^+ does not appear to be strongly dependent on habitat type: Exe sand and mud habitats have very different numbers of species but rather centrally-placed distinctness; Scilly algal and sand habitats have near-identical Δ^+ values. Warwick and Clarke (1998) also demonstrate a lack of habitat dependence in Δ^+ from a survey of Chilean nematodes (data of W Wieser).
- 3) There is apparent monotonicity of response of the index to environmental degradation (also in keeping with another initial criterion). To date, there is no evidence of average taxonomic distinctness increasing in response to stress.
- 4) In spite of the widely differing lengths of their species lists, it is notable that the two Clyde studies (C1, C2) return rather similar (depressed) values for Δ^+ .
- 5) There is no evidence of any empirical relation in the (Δ^+ , S) scatter plot. We know from the sampling theory that the mechanics of calculating Δ^+ does

not lead to an *intrinsic* relationship between the two but that does not prevent there being an *observed* correlation; the latter would imply some genuine assemblage structuring which predisposed large communities to be more (or less) 'averagely distinct' than small communities. The lack of an intrinsic, mechanistic correlation greatly aids the search for such interesting observational relationships (see also the later discussion on AvTD, VarTD correlations). The same cannot be said for phylogenetic diversity, PD. Fig. 17.9a shows the expected near-linear relation between total PD and S for these meiofaunal studies (total TD and S would have given a similar picture) but, more significantly, Fig. 17.9b bears out the previous statements about the dependence also of *average* PD (Φ^+) on S . This *intrinsic* relationship, shown by the declining curve for the expected value of Φ^+ as a function of the number of species in the list, contrasts markedly with the constant mean line for Δ^+ in Fig. 17.8. Nothing can therefore be read into an observed negative correlation of Φ^+ and S in a practical study: such a relationship would be likely, as here, to be purely mechanistic, i.e. artefactual.

AvTD is therefore seen to possess many of the features listed at the beginning of the chapter as desirable in a biodiversity index – a function, in part, of its attractive mathematical sampling properties (for formal statistical results on unbiasedness and variance structure see Clarke and Warwick, 1998b, 2001). Many questions

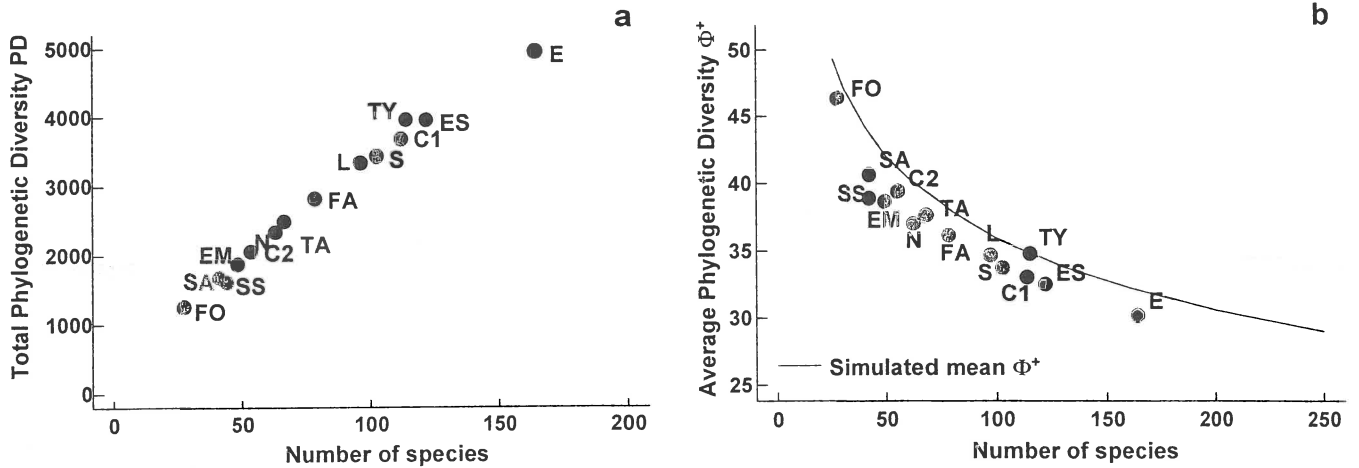


Fig. 17.9. UK regional study, free-living nematodes {U}. Scatter plots for the 14 location/habitat studies (Fig. 17.6) of: a) total PD, b) AvPD against list size m , the latter also showing the declining 'expected' mean Φ^+ with m , simulated from sublists of the UK master list.

remain, however – from theoretical issues of its dependence (or lack of it) on essentially arbitrary assumptions about relative weighting of step lengths through the taxonomic tree, to further practical demonstration of its performance (or lack of it) for other faunal groups and environmental impacts. The following example addresses these two questions in particular.

EXAMPLE: N Europe groundfish surveys

An investigation of the taxonomic structure of demersal fish assemblages in the North Sea, English Channel and Irish Sea, motivated by concerns over the impacts of beam trawling, is reported by Rogers *et al* (1999). A total of 277 ICES Quarter-rectangles were sampled {Q}, for 93 species of groundfish, by research vessels from different N European countries. Sampling effort per rectangle was not constant. For the purposes of display, quarter-rectangles were grouped into 9 larger sea-areas: 1–Bristol Channel, ..., 9–Eastern Central N Sea (Fig. 17.10, see legend for area definitions).

There is a wealth of taxonomic detail to exploit in this case. The analysis uses a 14-level classification (Fig. 17.11), based on phylogenetic information, compiled by J.D. Reynolds (Univ E Anglia), primarily from Nelson (1994) and McEachran and Miyake (1990). The distinctness structure of this master list, and its AvTD of $\Delta^+ = 80.1$, for all groundfish species that could be reliably sampled and identified, becomes the standard against which the species lists from the various quarter-rectangles are assessed.

Funnel plot

Fig. 17.12 displays the resulting funnel plot of the range of Δ^+ values expected from sublists of size 5 to 35,

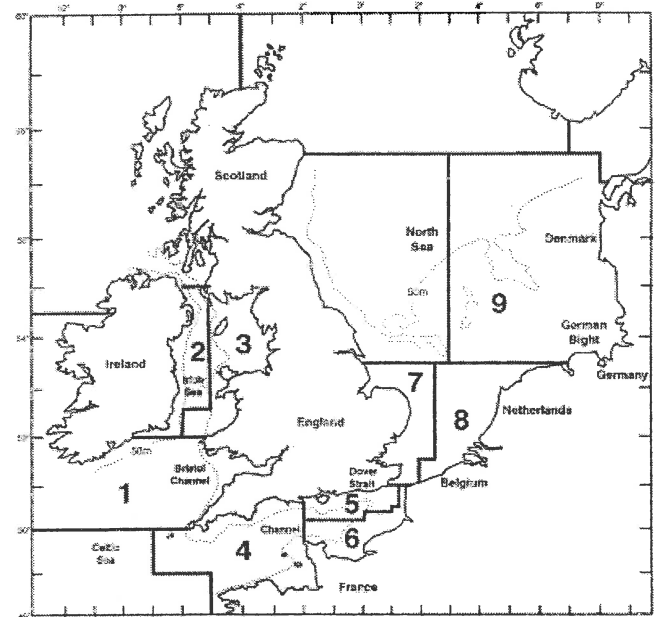


Fig. 17.10. Quarter ICES rectangles, groundfish surveys {Q}. 277 rectangles from 9 sea areas. 1: Bristol Channel, 2: W Irish Sea, 3: E Irish Sea, 4: W Channel, 5: NE Channel, 6: SE Channel, 7: SW North Sea, 8: SE North Sea, 9: E Central North Sea.

repeating the mean, lower and upper limits in sub-plots of observed Δ^+ values for the 9 sea areas. Δ^+ is clearly seen to be reduced in some areas, particularly 6, 8 and 9, whilst remaining at 'expected' levels in others. Rogers *et al* (1999) discuss possible explanations for this, noting the contribution made by the spatial pattern of elasmobranchs, a taxonomic group they argue may be particularly susceptible to disturbance by commercial trawling, because of their life history traits.

Weighting of step lengths

Many of the fine-scale phylogenetic groupings in Fig. 17.11 are utilised comparatively rarely (e.g. subgenera only within *Raja*, tribe only within the Pleuronectidae

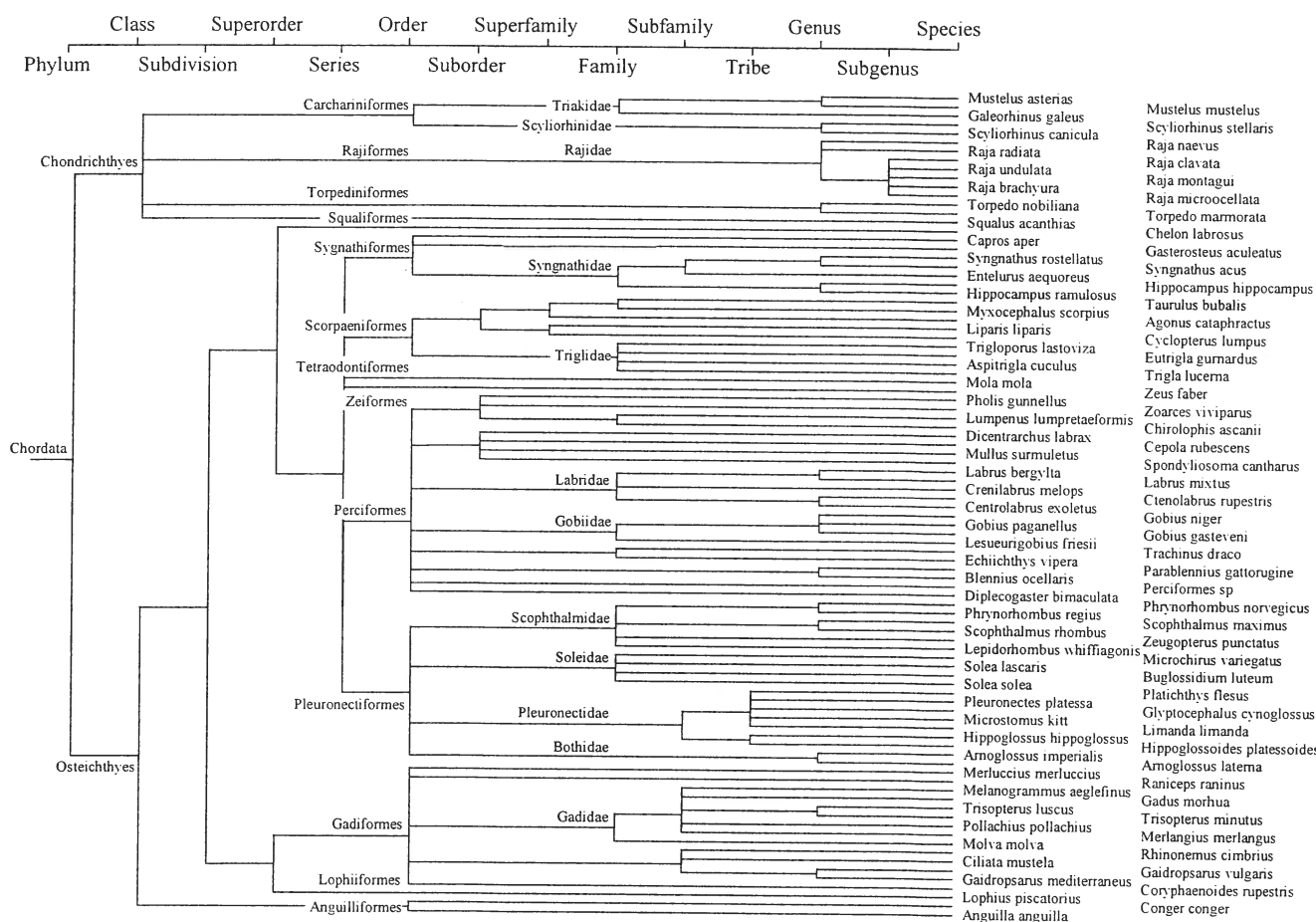


Fig. 17.11. Quarter ICES rectangles, groundfish surveys {Q}. 14-level classification (phylogenetically-based) used for the construction of taxonomic distances between 93 demersal fish species, those that could be reliably sampled and identified for the 277 rectangles in this N European study.

Table 17.1. Quarter ICES rectangles, groundfish surveys {Q}. The 13 taxonomic/phylogenetic categories (k) used in the groundfish study, the standard taxonomic distances $\{\omega_k\}$ and an alternative formulation $\{\omega_k^{(0)}\}$ based on taxon richness $\{s_k\}$ at each level. ω_k (or $\omega_k^{(0)}$) is the path length between species from different taxon group k but the same group $k+1$.

k	Taxon	s_k	ω_k	$\omega_k^{(0)}$
1	Species	93	7.7	1.3
2	Sub-genus	89	15.4	6.9
3	Genus	72	23.1	8.9
4	Tribe	67	30.8	12.5
5	Sub-family	59	38.5	21.4
6	Family	41	46.2	22.9
7	Super-family	39	53.8	27.4
8	Sub-order	33	61.5	44.4
9	Order	14	69.2	54.9
10	Series	9	76.9	61.4
11	Super-order	7	84.6	65.6
12	Sub-division	6	92.3	85.3
13	Class	2	100.0	100.0

etc), and the standard assumption that all step lengths between taxonomic levels are given equal weight (7.69, in this case) may appear arbitrary. For example, if a new category is defined which is not actually used, then the resulting change in all the step lengths, in order to accommodate it, seems unwarranted. The natural alternative here is to make the step lengths proportional to the extent of group melding that takes place, larger steps corresponding to larger decreases in taxon richness. A null category would then add no additional step length. Table 17.1 shows the resulting taxonomic distances $\{\omega^{(0)}\}$ between species connected at the differing levels, contrasted with the standard, equal-stepped, distances $\{\omega\}$. Obviously, both are standardised so that the largest distance in the tree (between species in the different classes Chondrichthyes and Osteichthyes) is set to 100.

Fig. 17.13 demonstrates the minimal effect these revised weights have on the calculation of average taxonomic distinctness, Δ^+ . It is a scatter plot of $\Delta^{+(0)}$ (revised weights) against Δ^+ (standard, equal-stepped, distances) for the 277 quarter-rectangle species lists. The relation

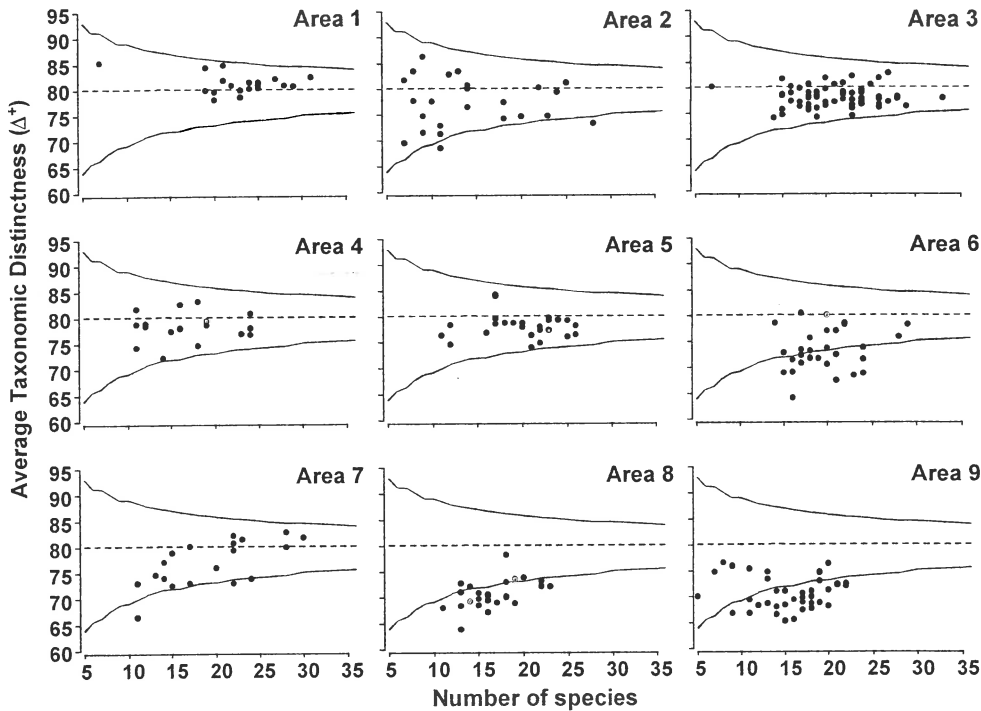


Fig. 17.12. Quarter ICES rectangles, groundfish surveys {Q}. AvTD (presence/absence data) against observed number of species, in each of 274 rectangles, grouped into 9 sea areas (Fig. 17.10). Dashed line indicates mean of 5000 simulated sublists for each size $m = 5, 6, 7, \dots, 35$, confirming the theoretical unbiasedness and therefore comparability of Δ^+ for widely differing degrees of sampling effort. Continuous lines denote 95% probability limits for Δ^+ from a single sublist of specified size from the master list (of 93 species).

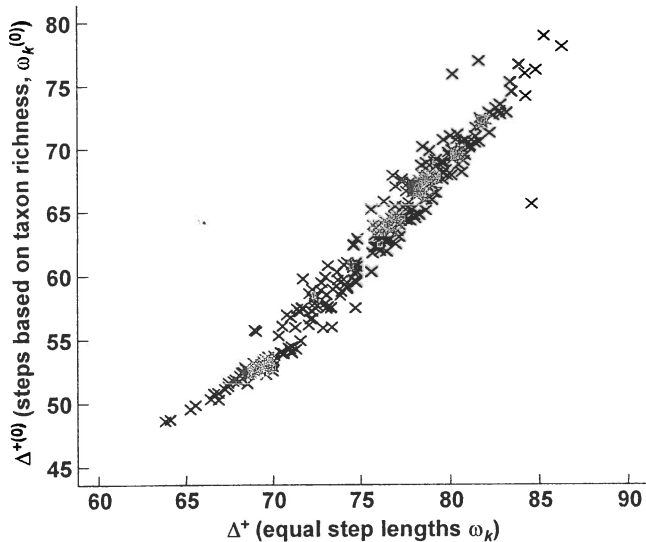


Fig. 17.13. Quarter ICES rectangles, groundfish surveys {Q}. Comparison of observed Δ^+ , for each of 277 rectangles, between two weighting options for taxonomic distance between species: equal step-lengths between hierarchical levels (x-axis), and lengths proportional to change in taxonomic richness at that step (y-axis).

is seen to be very tight, with only about 3 samples departing from near-linearity. (These are outliers of very low species richness – in one case as few as 2 species – and have been removed from Fig. 17.12.) Clearly, the relative values of Δ^+ are robust in this case to the precise definition of the step-length weights, a reassuring conclusion which is also borne out for the UK nematode study {U}. For the data of Fig. 17.8, Clarke and Warwick (1999) consider the effects of various alternative step-length definitions, consistently increasing or decreasing the weights at higher taxonomic levels as well as weighting them by changes in taxon richness.

The only alteration to the conclusions came from decreasing the step lengths at the higher (coarsest) taxonomic levels, especially suppressing the highest level altogether (so that species within different subclasses were considered no more taxonomically distant than those within different orders). The Scilly data sets then showed a clear change in their average distinctness in comparison with the other 11 Δ^+ values.[¶]

The unusual structure of the Scillies sublists is also exemplified, in a more elegant way, by considering not just *average* but *variation* in taxonomic distinctness.

VARIATION IN TAXONOMIC DISTINCTNESS, Δ^+

VarTD was defined in equation (17.7), as the variance of the taxonomic distances $\{\omega_{ij}\}$ between each pair of species i and j , about their mean distance Δ^+ . It has the potential to distinguish differences in taxonomic structure resulting, for example, in assemblages with some genera becoming highly species-rich whilst a range of other higher taxa are represented by only one (or a very few) species. In that case, average TD may be unchanged but variation in TD will be greatly increased, and Clarke and Warwick (2001) argue (on a sample of one!) that this might be expected to be characteristic of island fauna, such as that for the Isles of Scilly.

[¶] Both the PRIMER DIVERSE and TAXDTEST routines allow such compression of taxonomic levels, either at the top or bottom of the tree (or both), and also permit automatic computation of step-length weights based on changes in taxon richness and, indeed, any user-specified weighting.

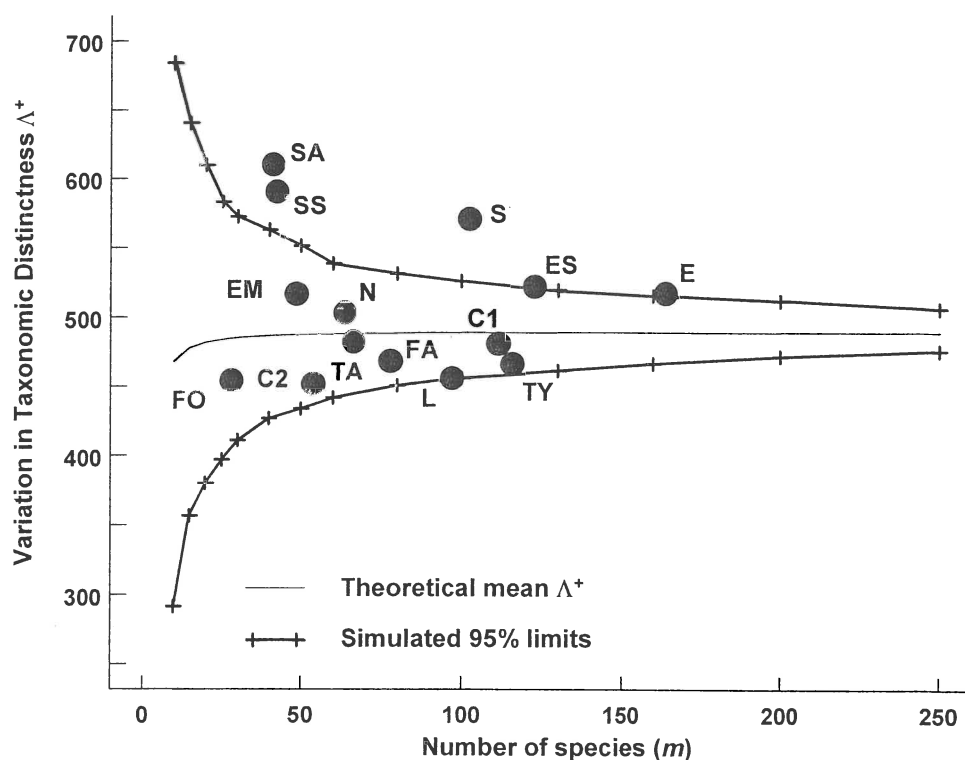


Fig. 17.14. UK regional study, free-living nematodes $\{U\}$. Funnel plot, as in Fig. 17.8, but for simulated VarTD (Λ^+), against sublist sizes $m=10, 15, 20, \dots, 250$ (x -axis), drawn from the 395-species master list. Thin line denotes the theoretical (and simulated) mean Λ^+ , which is no longer entirely constant, declining very slightly for small values of m . The bias is clearly negligible, however, showing that (like Δ^+) Λ^+ is comparable across studies with differing sampling effort (as here). Superimposed observed Λ^+ values for the 14 location/habitat combinations (Fig. 17.6) show a significantly larger than expected VarTD for the Scilly datasets.

For the UK nematode study $\{U\}$, Fig. 17.14 displays the funnel plot for VarTD (Λ^+) which is the companion to Fig. 17.8 (for AvTD, Δ^+). It is constructed in the same way, by many random selections of sublists of a fixed size m from the UK master list of 395 nematode species, and recomputation of Λ^+ for each subset. The resulting histograms are typically more symmetric than for Δ^+ , as seen by the 95% probability limits for ‘expected’ Λ^+ values, across the full range of sublist sizes: $m = 10, 15, 20, 25, \dots$, shown in Fig. 17.14. Three features are noteworthy:

- 1) The simulated mean Λ^+ (thin line in Fig. 17.14) is again largely independent of sublist size, only declining slightly for very short lists (and the slight bias is dwarfed by the large uncertainty at these low sizes). Clarke and Warwick (2001) derive an exact formula for the sampling bias of Λ^+ and show, generally, that it will be negligible. This again has important practical implications because it allows Λ^+ to be meaningfully compared across (historic) studies in which sampling effort is uncontrolled.
- 2) The various UK habitat/location combinations all fall within ‘expected’ ranges, with the interesting exception of the Scilly data sets. These have significantly higher VarTD values, as discussed above.
- 3) Λ^+ therefore appears to be extracting independent information, separately interpretable from Δ^+ , about the taxonomic structure of individual data sets. This assertion is testable by a *bivariate* approach.

JOINT (AvTD, VarTD) ANALYSES

The histogram and funnel plots of Figs. 17.7 and 17.8 are univariate analyses, concentrating on only one index at a time. Also possible is a bivariate approach in which (Δ^+ , Λ^+) values are considered jointly, both in respect of the observed outcomes from real data sets and their expected values under subsampling from a master species inventory. Fig. 17.15 shows the results of a large number of random selections of $m = 100$ species from the 395 in the UK nematode list $\{U\}$; each selection gives rise to an (AvTD, VarTD) pair and these are graphed in a scatter plot (Fig. 17.15a). Their spread defines the ‘expected’ region (rather than *range*) of distinctness behaviour, for a sublist of 100 species. Superimposed on the same plot are the observed (Δ^+ , Λ^+) pairs for three of the studies with list sizes of about that order: all three (Clyde, Liverpool Bay and Scilly) are seen to fall outside the expected structure, though in different ways, as previously discussed.

‘Ellipse’ plots

It aids interpretation to construct the bivariate equivalent of the univariate 95% probability limits in the histogram or funnel plots, namely a 95% probability region, within which (approximately) 95% of the simulated values fall. An adequate description here is provided by the ellipse from a fitted bivariate normal distribution to separately transformed scales for Δ^+ and Λ^+ .

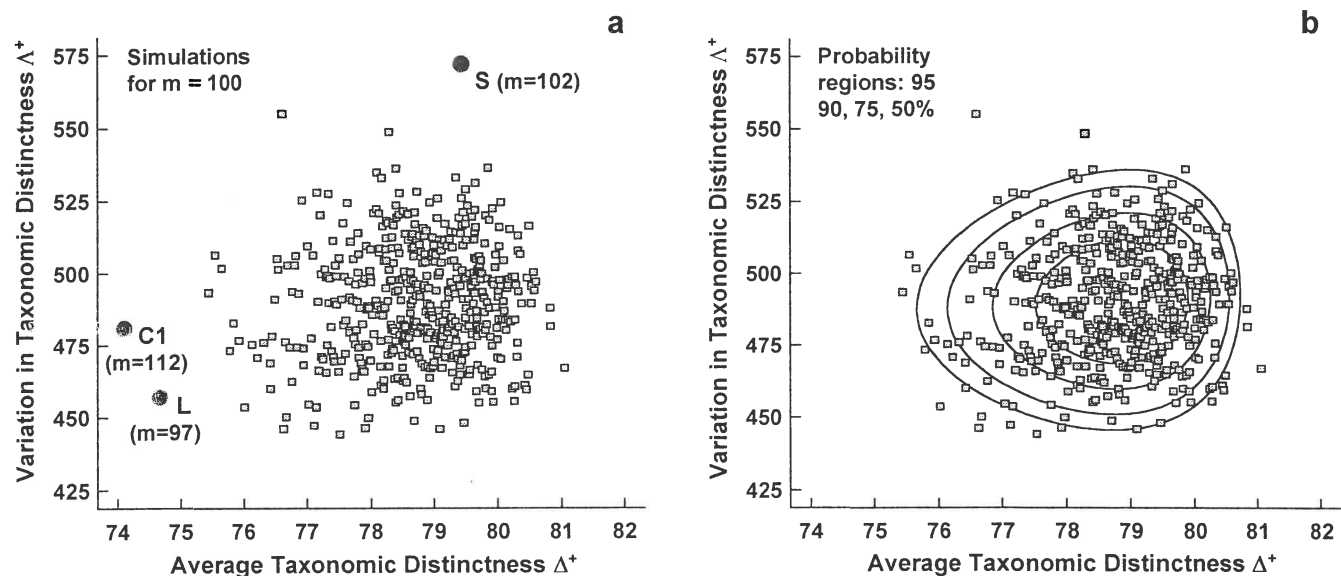


Fig. 17.15. UK regional study, free-living nematodes {U}. a) Scatter plot of (AvTD, VarTD) pairs from random selections of $m = 100$ species from the UK nematode list of 395; also superimposed are three observed points: Clyde (C1), Liverpool Bay (L) and Scilly (S), all falling outside 'expectation'. b) Probability contours (back-transformed ellipses) containing approximately 95, 90, 75 and 50% of the simulated values. Both plots are based on 1000 simulations though only 500 points are displayed, for clarity.

AvTD in particular needs a reverse power transform to eliminate the left-skewness though, as previously noted, any transformation of VarTD can be relatively mild, if needed at all. Clarke and Warwick (2001) discuss the fitting procedure in detail[¶] and Fig. 17.15b shows its success in generating convincing probability contours, containing very close to the nominal levels of 50, 75, 90 and 95% of simulated data points. In the normal convention, the 'expected region' is taken as the outer (95%) contour, which is an ellipse on the transformed scales, though typically 'egg-shaped' when back-transformed to the original (Δ^+ , Λ^+) plot.

A different region needs to be constructed for each sublist size or, in practice, for a range of values of m straddling the observed sizes. Though these could all be displayed on a single plot, it may improve clarity to separate them in groups of two or three, as in Fig. 17.16. The conclusions are largely unchanged from those drawn earlier, for the separate funnel plots, and it would be reasonable to query the advantage of a bivariate approach here. However, it has at least three merits:

1) The procedure automatically compensates for the repeated testing inherent in two separate, univariate tests.

2) The 'failure to reject' region of the null hypothesis, inside the simulated 95% probability contour, is not rectangular, as it would be for two separate tests. This opens the possibility for other faunal groups, where simulated Δ^+ and Λ^+ values may be negatively correlated (as appears to happen for components of the macrobenthos, Clarke and Warwick, 2001), that significance could follow from the combination of moderately low AvTD and VarTD values, where neither of them on their own would indicate rejection.

3) It aids interpretation of spatial biodiversity patterns to know whether there is any intrinsic, artefactual correlation to be expected between the two indices, resulting from the fact that they are both calculated from the same set of data. Here, Fig. 17.15 shows emphatically that no such internal correlation is to be expected (though, as just commented, the independence of Δ^+ and Λ^+ is not a universal result, and needs to be examined by simulation for each new master list). Yet the *empirical* correlation between Δ^+ and Λ^+ for the 14 studies is *not* zero but large and positive (Fig. 17.17). This implies a genuine correlation from location to location in these two assemblage features, which it is legitimate to interpret. The suggestion (Clarke and Warwick, 2001) is that pollution may be connected with a loss both of the normal wide spread of higher taxa (reduced Δ^+), *and* that the higher taxa lost are those with a simple subsidiary structure, represented only by one or two species, genera or families, leaving a more balanced tree (reduced Λ^+).

[¶] Accomplished by the PRIMER TAXDTEST routine, which automatically carries out the simulations and transformation/fitting of bivariate probability regions to obtain (transformed) 'ellipse' plots, for specified sublist sizes, on which real data pairs (Δ^+ , Λ^+) may be superimposed

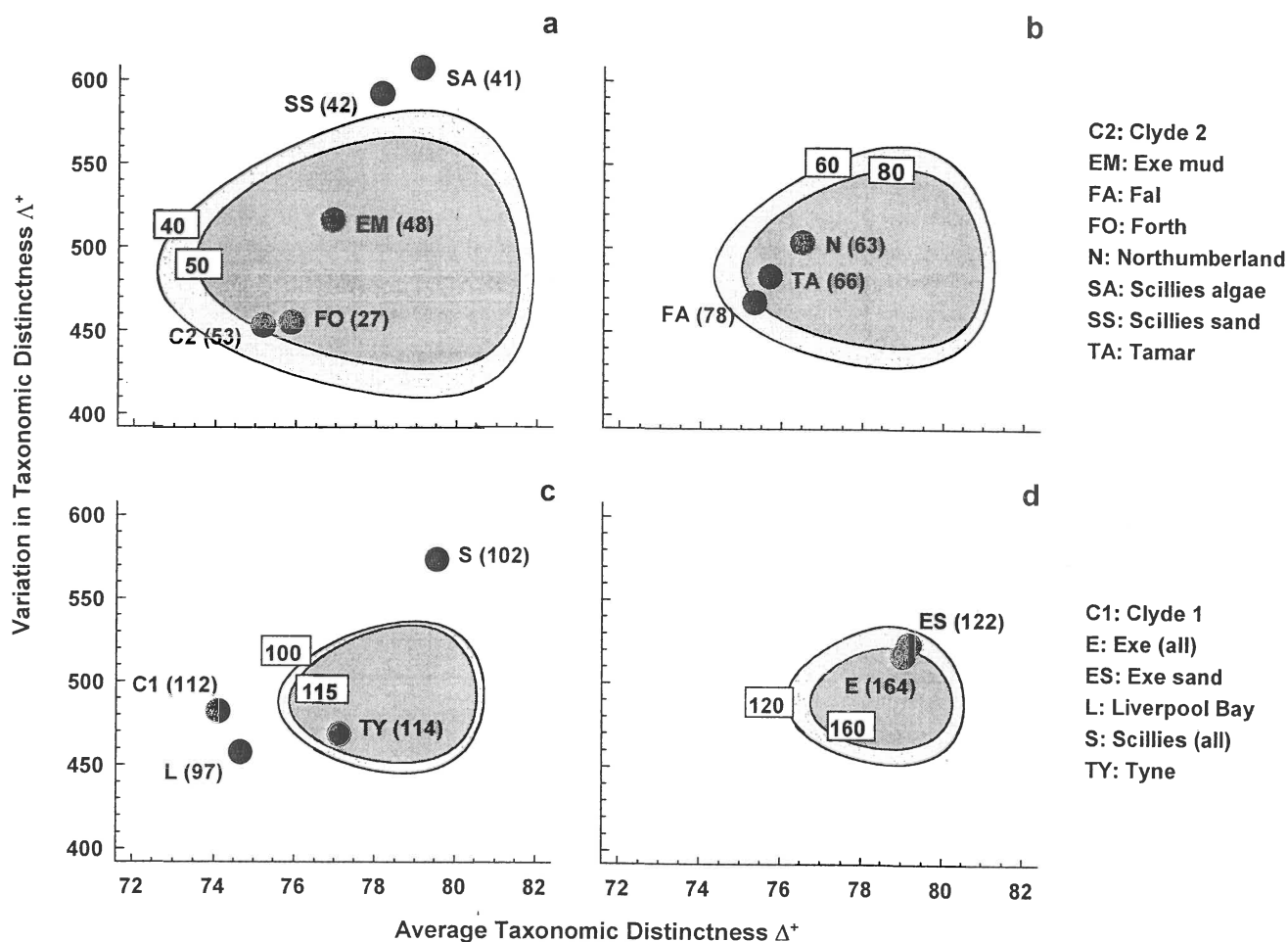


Fig. 17.16. UK regional study, free-living nematodes {U}. 'Ellipse' plots of 95% probability regions for (AvTD, VarTD) pairs, as for Fig. 17.15 but for a range of sublist sizes: a) $m = 40, 50$; b) $m = 60, 80$; c) $m = 100, 115$; d) $m = 120, 160$. The observed (Δ^+ , Λ^+) values for the 14 location/habitat studies are superimposed on the appropriate plot for their particular species list size (given in brackets). As seen in the separate funnel plots (Figs. 17.8 and 17.14), Clyde, Liverpool Bay, Fal (borderline) and all the Isles of Scilly data sets depart significantly from expectation.

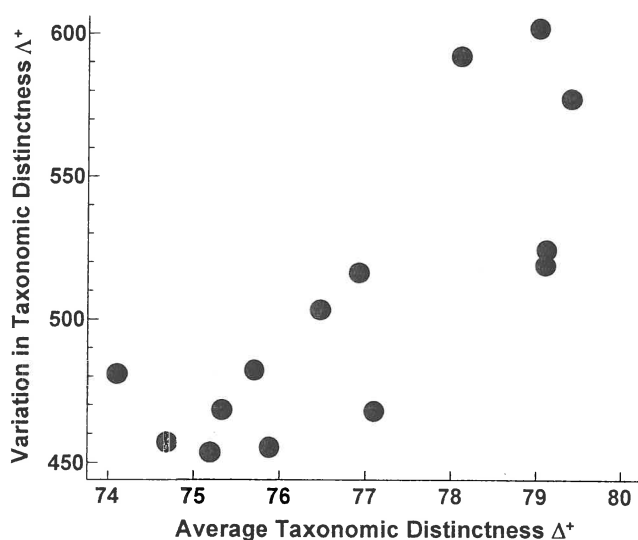


Fig. 17.17. UK regional study, free-living nematodes {U}. Simple scatter plot of observed (AvTD, VarTD) values for the 14 location/habitat studies, showing the strongly positive empirical correlation (Pearson $r = 0.79$), which persists even if the three Scilly values are excluded ($r = 0.75$).

CONCLUDING REMARKS

Other applications

These taxonomic distinctness indices detailed above are relatively recent and, so far, have been applied in few studies in the published literature. Some marine exceptions, not illustrated earlier, are Hall and Greenstreet (1998) for demersal fish, Piepenburg *et al* (1997) for starfish and brittle-stars in polar regions, Price *et al* (1999) for Atlantic starfish assemblages[¶], Somerfield *et al* (1997) for further sediment macrobenthos around

[¶] Iszak and Price (2001) also consider the extension of taxonomic distinctness ideas to the construction of a similarity coefficient (for presence/absence data) between two assemblages, taking into account the taxonomic relatedness of their species. Effectively, they use what is defined by Clarke and Warwick (1998a) as a **taxonomic mapping** similarity between two species lists, used there to examine the taxonomic relatedness of the species sets successively 'peeled' from the full list, in the structural redundancy analysis of influential groups of species, discussed in Chapter 16.

oilfields, and Ward *et al* (submitted) for a latitudinal study of pelagic copepods. Two non-marine examples are D. Danielopol (pers. comm.) for groundwater copepods and Shimatani (2001) for forest stands. A bivariate biodiversity example is given by Warwick and Light (in press), who use ‘ellipse’ plots of expected (Δ^+ , Λ^+) values, from live faunal records of the Isles of Scilly, to examine whether easily sampled bivalve and gastropod death assemblages could be considered representative of the taxonomic distinctness structure of the live fauna.

Caveats

In view of its limited application to date, too much should not be claimed for this methodology![¶] It *is* surprising that anything sensible can be said about diversity at all, for data consisting simply of species lists, and arising from unknown or uncontrolled sampling effort (which usually renders it impossible to read anything into the relative size of these lists). Yet, much of the later part of this chapter suggests that not only can we find one index (AvTD) which may be validly compared across such studies, and which captures an intuitive sense of what biodiversity means, but we can also find a second one (VarTD), with similarly attractive statistical properties, and which seems to capture (in some cases, at least) an entirely independent attribute of biodiversity structure.

Nonetheless, it is clear that controlled sampling designs, carried out in a strictly uniform way across different spatial, temporal or experimental conditions, must provide additional, meaningful, comparative diversity information (on richness, primarily) that Δ^+ and Λ^+ are designed to ignore. Even here, though, concepts of taxonomic relatedness can expand the relevance of richness indices: rather than use S , or one of its variants (see Chapter 8), total taxonomic distinctness (TTD) or total phylogenetic diversity (PD), see pages 17-5 and 17-6, capture the richness of an assemblage in terms of its number of species *and* whether they are closely or distantly related.

Sensitivity and robustness

Returning for a moment to the quantitative form of average distinctness, equation (17.3), the Ekofisk oil-field study suggested that such relatedness measures may have a greater sensitivity to disturbance events than is seen with species-level richness or evenness

indices (Warwick and Clarke, 1995). This suggestion was not borne out by subsequent oil-field studies (Sommerfield *et al*, 1997), particularly where the impact was less sustained, the data collection at a less extensive level and hence the gradients more subtly entwined with natural variability. But it would be a mistake to claim sensitivity as a rationale for this approach: there is much empirical evidence that the best way of detecting subtle community shifts arising from environmental impacts is not through univariate indices at all, but by non-parametric multivariate display and testing (Chapter 14). The difficulty with the multivariate techniques is that, since they match precise species identities through the construction of similarity coefficients, they can be sensitive to wide scale differences in habitat type, geographic location (and thus species pool) etc.

Though independent of particular species identities, many of the traditional univariate indices have their own sensitivities, to habitat type, dominant species and sampling effort differences, as we have discussed. The general point here is that robustness (to sampling details) and sensitivity (to impact) are usually conflicting criteria. What is properly claimed for average taxonomic distinctness is not sensitivity but:

- a) **relevance** – it is a genuine reflection of biodiversity loss, gain, or neither (rather than recording simply a change of assemblage composition), and one that appears to respond in a monotonic way to impact;
- b) **robustness** – it can be meaningfully compared across studies from widely separated locations, with few (or even no) species in common, from different habitats, using data in presence/absence form (and thus not sensitive to dominant species), and with different levels of sampling effort. This makes its natural use the comparison of regional/global studies and/or historic data sets.

Taxonomic artefacts

A natural question is the extent to which relatedness indices are subject to taxonomic artefacts. Linnean hierarchies can be inconsistent in the way they define taxonomic units across different phyla, for example. This concern can be addressed on a number of levels. As suggested earlier, the concept of mutual distinctness of a set of species is not constrained to a Linnean classification. The natural metric may be one of genetic distance (e.g. Nei, 1996) or that from a soundly-based phylogeny combining molecular approaches with more traditional morphology. The Linnean classification clearly gives a discrete approximation to a more continuous distinctness measure, and this is why it is important to establish that the precise weightings given to the

[¶] A quote from Dr Johnson seems apt: “It is not done well; but you are surprised to find it done at all”! (Boswell’s ‘Life of Johnson’, 1763)

step lengths between taxonomic levels are not critical to the relative values that the index takes, across the studies being compared. Nonetheless, it is a legitimate concern that a cross-phyletic distinctness analysis could represent a simple shift in the balance of two major phyla as a decrease in biodiversity, not because the phylum whose presences are increasing is genuinely less (phylogenetically) diverse but because its taxonomic sub-units have been arbitrarily set at a lower level. Such taxonomic artefacts could be readily examined by computing, for example, the (AvTD, VarTD) structure across different phyla in a standard species catalogue (e.g. Howson, 1987). The pragmatic approach, as illustrated here for the UK nematode study and the groundfish data, is to work within a well-characterised, reasonably taxonomically coherent group.

Master species list

Concerns about the precise definition of the master list (e.g. its biogeographic range or habitat specificity) also naturally arise. Note, however, that the existence of such a wide-scale inventory is not a central requirement, more of a secondary refinement. It is not used in constructing and contrasting the values of Δ^+ for individual samples, and only features in two ways in these analyses:

- 1) In the funnel plots (Figs. 17.8, 17.12, 17.14), location of the points does not require a master species list, the latter being used only to display the background reference of the mean value and limits that would be expected for samples drawn at random from such an inventory. In Fig. 17.12, in fact, the limits are not even that relevant since they apply to *single* samples rather than, for example, to the mean of the tens of samples plotted for each sea area. The most useful plot for interpretation here is simply a standard *means plot* of the observed mean Δ^+ and its 95% confidence interval, calculated from the replicates for each sea area (see Rogers *et al.*, 1999 and Warwick and Clarke, 2001).
- 2) In Table 17.1 and Fig. 17.3, the master species list is employed to calculate step lengths in a revised

form of Δ^+ – weighting by taxon richness at the different hierarchical levels. The existence of a master inventory makes this procedure more appealing, since if the taxon richness weighting was determined only by the samples to hand, the index would need to be adjusted as each new sample (containing further species) was added. The message of this chapter, however, is that the additional complication of adjusting the weights in Δ^+ for differences in taxon richness is unnecessary. Constant step lengths appear to be adequate.

The inventory is therefore only used for setting a background context, the theoretical mean and funnel limits. Various lists could sensibly be employed: global, local geographic, biogeographic provinces, or simply the combined species list of all the studies being analysed. The addition of a small number of newly-discovered species to the master inventory is unlikely to have a detectable effect on the overall mean and funnel for Δ^+ . If these are located in the taxonomic tree at random with respect to the existing taxa (rather than all belonging to the same high or low order group) they will have little or no effect on the theoretical mean Δ^+ . This, of course, is one of the advantages of using an index of *average* rather than *total* taxonomic distinctness.

It also makes clear what the limitations are to the validity of Δ^+ comparisons. Whilst many marine community studies seem to consist of the low-level (species or genera) identifications which are necessary for meaningful computation of Δ^+ , there are always some taxa that cannot be identified to this level. There is no real difficulty here, since Δ^+ is always used in a relative manner, *provided* these taxa are treated in the same way in all samples (e.g. omitted). The ability to impose *taxonomic consistency*, by suitable omissions or re-groupings, is clearly an important caveat on the use of taxonomic distinctness for historic or widely-sourced data sets. Where such conditions can be met, however, we believe that these (and possibly other) measures based on taxonomic relatedness, have a promising future role in biodiversity assessment in relation to global change and at global scales.



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). The entries give all pages on which each set is analysed and also its source reference (see also Appendix 3). These are not always the appropriate references for the analyses of 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, 16-1 to 16-6, 17-2, 17-3
- B – Bristol Channel, England. *Zooplankton*. (Collins and Williams, 1982).
p 3-5, 3-6, 7-2 to 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, 5-9, 9-4
- E – Ekofisk oil platform, N.Sea. *Macrofauna*. (Gray *et al*, 1990).
p 8-5, 8-6, 10-5, 10-6, 14-2, 14-3, 15-5 to 15-7, 17-4
- 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 to 6-5, 8-8, 8-10, 10-1, 10-2, 13-6, 14-1, 14-2, 15-2, 15-3
- G – Garroch Head, sludge dump-ground, Scotland. *Macrofauna*. (Pearson and Blackstock, 1984).
p 1-6, 1-8, 1-11, 1-12, 8-5 to 8-7, 8-9, 8-12, 11-1 to 11-5, 11-9, 11-10, 15-2, 15-3
- H – Hamilton Harbour, Bermuda (GEEP Workshop). *Macrofauna, nematodes*. (Warwick *et al*, 1990c).
p 8-3, 8-4, 8-12, 13-3, 13-4
- I – Indonesian reef corals, S. Pari and S. Tikus Islands. *Coral % cover*. (Warwick *et al*, 1990b).
p 6-6, 6-7, 8-3, 8-4, 10-4, 10-5, 10-7, 13-4, 13-5, 14-3, 14-4, 15-5 to 15-7
- J – Joint NE Atlantic shelf studies (“meta-analysis”). *Macrofauna “production”*. (Warwick and Clarke, 1993a).
p 15-2 to 15-5
- K – Ko Phuket coral reefs, Thailand. *Coral species cover*. (Clarke *et al*, 1993; Brown *et al*, in press).
p 15-8 to 15-10, 16-7, 16-8
- L – Loch Linnhe and Loch Eil, Scotland, pulp-mill effluent. *Macrofauna*. (Pearson, 1975).
p 1-6, 1-7, 1-10, 4-5, 8-7, 8-8, 8-10, 8-11, 10-3, 10-4, 10-6, 10-7, 15-2, 15-3
- M – Maldiv Islands mining. *Coral reef fish*. (Dawson-Shepherd *et al*, 1992).
p 13-2, 14-4, 14-5, 15-5 to 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 to 15-7
- P – Plankton survey (Continuous Plankton Recorder), N.E. Atlantic. *Zooplankton, phytoplankton*. (Colebrook, 1986).
p 13-1
- Q – Quarter-ICES rectangles, beam-trawl surveys, N. Europe. *Groundfish*. (Rogers *et al*, 1998).
p 17-10 to 17-12
- R – Tamar estuary mud-flat, S.W. England. *Nematodes, copepods*. (Austen and Warwick, 1989).
p 14-6 to 14-8
- S – Scilly Isles, UK. *Seaweed metazoa*. (Gee and Warwick, 1994).
p 13-5, 14-5, 14-6
- T – Tasmania, Eaglehawk Neck. *Nematodes, copepods*. (Warwick *et al*, 1990a).
p 6-9, 12-2 to 12-4, 13-3, 13-4, 14-7
- U – UK regional studies. *Nematodes*. (Warwick and Clarke, 1998).
p 17-7 to 17-10, 17-13 to 17-15
- V – Variation inter-annually, Tees Bay, UK. *Macrobenthos*. (Warwick *et al*, in press).
p 15-7, 17-5
- W – Westerschelde estuary cores, Netherlands; mesocosm experiment on food supply. *Nematodes*. (Austen and Warwick, 1995).
p 6-10, 6-12
- X – Exe estuary, S.W. England. *Nematodes*. (Warwick, 1971).
p 5-2 to 5-4, 5-7, 6-12, 6-13, 7-1, 7-2, 11-5 to 11-9
- Y – Clyde, Scotland. *Nematodes*. (Lamshead, 1986).
p 6-7, 6-8
- Z – Azotic 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 community analyses that has been adopted and developed at the Plymouth Marine Laboratory (PML) for well over a decade, and has benefited from experience at numerous IOC/UNEP and commercial training courses. Methods papers from work at PML covered in this manual include: Field *et al* (1982), Warwick (1986), Clarke and Green (1988), Clarke (1990), Warwick and Clarke (1993a & b, 1995a), Clarke and Ainsworth (1993), Clarke *et al* (1993), Clarke and Warwick (1994, 1998a & b, 1999, 2001) and Somerfield and Clarke (1995, 1997). Clarke (1993, 1999), Warwick (1993) and Warwick and Clarke (2001) give general reviews, and a large number of papers from PML and authors worldwide exemplify their use via the PRIMER package (some are listed in Appendix 3 but there are currently over a thousand in the SCI).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 14: Relative sensitivities. This parallels the earlier sections of Warwick and Clarke (1991), from which all these figures (except Figs. 14.11 & 14.14) have been redrawn. Primary source versions of the

figures can be found as follows: Figs. 14.1–14.3, Gray *et al* (1988); Figs. 14.5–14.7, Warwick *et al* (1990b); Figs 14.9–14.10, Dawson-Shepherd *et al* (1992); Figs. 14.11–14.12, Gee and Warwick (1994); Figs. 14.14–14.16, Austen and Warwick (1989).

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

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

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

APPENDIX 3: REFERENCES CITED

- Addison, R.F., Clarke, K.R. (eds.) (1990). *Biological effects of pollutants in a subtropical environment*. *J. exp. mar. Biol. Ecol.* 138
- Agard, J.B.R., Gobin, J., Warwick, R.M. (1993). Analysis of marine macrobenthic community structure in relation to pollution, natural oil seepage and seasonal disturbance in a tropical environment (Trinidad, West Indies). *Mar. Ecol. Prog. Ser.* 92: 233-243
- Anderson, M.J. (2001a). Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* 58: 626-639
- Anderson, M.J. (2001b). A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26: 32-46
- Anderson, M.J., Underwood, A.J. (1997). Effects of gastropod grazers on recruitment and succession of an estuarine assemblage: a multivariate and univariate approach. *Oecologia* 109: 442-453
- Austen, M.C., McEvoy, A.J. (1997). The use of offshore meiobenthic communities in laboratory microcosm experiments: response to heavy metal contamination. *J. exp. Mar. Biol. Ecol.* 211: 247-261
- Austen, M.C., Somerfield, P.J. (1997). A community level sediment bioassay applied to an estuarine heavy metal gradient. *Mar. Envir. Res.* 43: 315-328
- Austen, M.C., Thrush, S.F. (2001). Experimental evidence suggesting slow or weak response of nematode community structure to a large suspension-feeder. *J. Sea Res.* 46: 69-84
- Austen, M.C., Warwick, R.M. (1989). Comparison of univariate and multivariate aspects of estuarine meiobenthic community structure. *Est. cstl. shelf Sci.* 29: 23-42
- Austen, M.C., Warwick, R.M. (1995). Effects of manipulation of food supply on estuarine meiobenthos. *Hydrobiologia* 311: 175-184
- Austen, M.C., Widdicombe, S., Villano-Pitacco, N. (1998). Effects of biological disturbance on diversity and structure of meiobenthic nematode communities. *Mar. Ecol. Prog. Ser.* 174: 233-246
- Bayne, B.L., Clarke, K.R., Gray, J.S. (eds) (1988). *Biological effects of pollutants: results of a practical workshop*. *Mar. Ecol. Prog. Ser.* 46
- Beukema, J.J. (1988). An evaluation of the ABC-method (abundance/biomass comparison) as applied to macrozoobenthic communities living on tidal flats in the Dutch Wadden Sea. *Mar. Biol.* 99: 425-433
- Box, G.E.P., Cox, D.R. (1964). An analysis of transformations. *J. R. Statist. Soc. Ser. B* 26: 211-243
- Bray, J.R., Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27: 325-349
- Brown, B.E., Clarke, K.R., Warwick, R.M. (in press). Serial patterns of biodiversity change in corals across shallow reef flats in Ko Phuket, Thailand, due to the effects of local (sedimentation) and regional (climatic) perturbations. *Mar. Biol.*
- Buchanan, J.B. (1993). Evidence of benthic pelagic coupling at a station off the Northumberland coast. *J. exp. Mar. Biol. Ecol.* 172: 1-10
- Caswell, H. (1976). Community structure: a neutral model analysis. *Ecol. Monogr.* 46: 327-354
- Chapman, M.G., Underwood, A.J. (1999). Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests. *Mar. Ecol. Prog. Ser.* 180: 257-265
- Chatfield, C., Collins, A.J. (1980). *Introduction to multivariate analysis*. Chapman and Hall, London
- Clarke, K.R. (1988). Detecting change in benthic community structure. Proceedings XIVth International Biometric Conference, Namur: Invited Papers. Societe Adolphe Quetelet, Gembloux, Belgium
- Clarke, K.R. (1990). Comparisons of dominance curves. *J. exp. mar. Biol. Ecol.* 138: 143-157
- Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18: 117-143
- Clarke, K.R. (1999). Non-metric multivariate analysis in community-level ecotoxicology. *Environ. Toxicol. Chem.* 18: 118-127
- Clarke, K.R., Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92: 205-219
- Clarke, K.R., Green, R.H. (1988). Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Prog. Ser.* 46: 213-226
- Clarke, K.R., Warwick, R.M. (1994). Similarity-based testing for community pattern: the 2-way layout with no replication. *Mar. Biol.* 118: 167-176
- Clarke, K.R., Warwick, R.M. (1998a). Quantifying structural redundancy in ecological communities. *Oecologia* 113: 278-289
- Clarke, K.R., Warwick, R.M. (1998b). A taxonomic distinctness index and its statistical properties. *J. appl. Ecol.* 35: 523-531
- Clarke, K.R., Warwick, R.M. (1999). The taxonomic distinctness measure of biodiversity: weighting of step lengths between hierarchical levels. *Mar. Ecol. Prog. Ser.* 184: 21-29
- Clarke, K.R., Warwick, R.M. (2001). A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar. Ecol. Prog. Ser.* 216: 265-278
- Clarke, K.R., Warwick, R.M., Brown, B.E. (1993). An index showing breakdown of seriation, related to disturbance, in a coral-reef assemblage. *Mar. Ecol. Prog. Ser.* 102: 153-160
- Clifford, D.H.T., Stephenson, W. (1975). *An introduction to numerical classification*. Academic Press, New York
- Colebrook, J.M. (1986). Environmental influences on long-term variability in marine plankton. *Hydrobiologia* 142: 309-325
- Collins, N.R., Williams, R. (1982). Zooplankton communities in the Bristol Channel and Severn Estuary. *Mar. Ecol. Prog. Ser.* 9: 1-11
- Connell, J.H. (1974). Field experiments in marine ecology. In: Mariscal, R. (ed.) *Experimental marine biology*. Academic Press, New York
- Connell, J.H. (1978). Diversity in tropical rain forests and coral reefs. *Science N.Y.* 199: 1302-1310

- Cormack, R.M. (1971). A review of classification. *J. R. Statist. Soc. Ser. A* 134: 321–367
- Dauvin, J.-C. (1984). Dynamique d'écosystèmes macrobenthiques des fonds sédimentaires de la Baie de Morlaix et leur perturbation par les hydrocarbures de l'Amoco-Cadiz. Doctoral thesis, Univ. Pierre et Marie-Curie, Paris
- Dawson-Shepherd, A., Warwick, R.M., Clarke, K.R., Brown, B.E. (1992). An analysis of fish community responses to coral mining in the Maldives. *Environ. Biol. Fish.* 33: 367–380
- Everitt, B. (1978). *Graphical techniques for multivariate data*. Heinemann, London
- Everitt, B. (1980). *Cluster analysis, 2nd edn*. Heinemann, London
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61: 1–10
- Faith, D.P. (1994). Phylogenetic pattern and the quantification of organismal biodiversity. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 345: 45–58
- Faith, D.P., Minchin, P.R., Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69: 57–68
- Field, J.G., Clarke, K.R., Warwick, R.M. (1982). A practical strategy for analysing multispecies distribution patterns. *Mar. Ecol. Prog. Ser.* 8: 37–52
- Fisher, R.A., Corbet, A.S., Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. anim. Ecol.* 12: 42–58
- Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Math.* 2: 282–285
- Gage, J.D., Coghill, G.G. (1977). Studies on the dispersion patterns of Scottish sea-loch benthos from contiguous core transects. In: Coull, B. (ed) *Ecology of marine benthos*. University of South Carolina Press, Columbia.
- Gee, J.M., Somerfield, P.J. (1997). Do mangrove diversity and leaf litter decay promote meiofaunal diversity? *J. exp. mar. Ecol. Biol.* 218: 13–33
- Gee, J.M., Warwick, R.M. (1994a). Metazoan community structure in relation to the fractal dimensions of marine microalgae. *Mar. Ecol. Prog. Ser.* 103: 141–150
- Gee, J.M., Warwick, R.M. (1994b). Body-size distribution in a marine metazoan community and the fractal dimensions of macroalgae. *J. exp. mar. Biol. Ecol.* 178: 247–259
- Gee, J.M., Warwick, R., Schaanning, M., Berge, J.A., Ambrose Jr, W.G. (1985). Effects of organic enrichment on meiofaunal abundance and community structure in sublittoral soft sediments. *J. exp. mar. Biol. Ecol.* 91: 247–262
- Goldman, N., Lamshead, P.J.D. (1989). Optimization of the Ewens/Caswell neutral model program for community diversity analysis. *Mar. Ecol. Prog. Ser.* 50: 255–261
- Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–328
- Gower, J.C. (1971). Statistical methods of comparing different multivariate analyses of the same data. pp. 138–149 in F.R. Hodson, D.G. Kendall and P. Tautu (eds.), *Mathematics in the archaeological and historical sciences*. Edinburgh University Press, Edinburgh
- Gower, J.C., Ross, G.J.S. (1969). Minimum spanning trees and single linkage cluster analysis. *Appl. Statist.* 18: 54–64
- Gray, J.S., Aschan, M., Carr, M.R., Clarke, K.R., Green, R.H., Pearson, T.H., Rosenberg, R., Warwick, R.M. (1988). Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.* 46: 151–165
- Gray, J.S., Clarke, K.R., Warwick, R.M., Hobbs, G. (1990). Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar. Ecol. Prog. Ser.* 66: 285–299
- Gray, J.S., Pearson, T.H. (1982). Objective selection of sensitive species indicative of pollution-induced change in benthic communities. I. Comparative methodology. *Mar. Ecol. Prog. Ser.* 9: 111–119
- Green, R.H. (1979). *Sampling design and statistical methods for environmental biologists*. Wiley, New York
- Greenacre, M.J. (1984). *Theory and applications of correspondence analysis*. Academic Press, London
- Hall, S.J., Greenstreet, S.P. (1998). Taxonomic distinctness and diversity measures: responses in marine fish communities. *Mar. Ecol. Prog. Ser.* 166: 227–229
- Harper, J.L., Hawksworth, D.L. (1994). Biodiversity: measurement and estimation. Preface. *Phil. Trans. Roy. Soc. Lond. Ser. B* 345: 5–12.
- Heip, C., Herman, P.M.J., Soetaert, K. (1988). Data processing, evaluation, and analysis. pp. 197–231 in R.P. Higgins and H. Thiel (eds.), *Introduction to the study of meiofauna*. Smithsonian Institution, Washington DC
- Hill, M.O. (1973a). Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61: 237–249
- Hill, M.O. (1973b). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427–432
- Hill, M.O. (1979a). *DECORANA – A FORTRAN program for detrended correspondence analysis and reciprocal averaging*. Cornell University, Ithaca, New York
- Hill, M.O. (1979b). *TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of individuals and attributes*. Cornell University, Ithaca, New York
- Hill, M.O., Gauch, H.G. (1980). Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42: 47–48
- Hope, A.C.A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. Ser. B* 30: 582–598
- Howson, C.M. (ed.) (1987). *Directory of the British marine fauna and flora*. Marine Conservation Society, Ross-on-Wye, Hertfordshire
- Humphries, C.J., Williams, P.H., Vane-Wright, R.I. (1995). Measuring biodiversity value for conservation. *Ann. Rev. Ecol. Syst.* 26: 93–111
- Hurlbert, S.H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577–586
- Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54: 187–211

- Huston, M. (1979). A general hypothesis of species diversity. *Am. Nat.* 113: 81–101
- Ibanez, F., Dauvin, J.-C. (1988). Long-term changes (1977–1987) in a muddy fine sand *Abra alba* – *Melinna palmata* community from the Western English Channel: multivariate time-series analysis. *Mar. Ecol. Prog. Ser.* 49: 65–81
- Izsak, C., Price, A.R.G. (2001). Measuring β -diversity using a taxonomic similarity index, and its relation to spatial scale. *Mar. Ecol. Prog. Ser.* 215: 69–77
- Jayasree, K. (1976). Systematics and ecology of free-living marine nematodes from polluted intertidal sand in Scotland. Ph.D. thesis, University of Aberdeen
- Jongman, R.H.G., ter Braak, C.F.J., van Tongeren, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc, Wageningen
- Kendall, M.A., Widdicombe, S. (1999). Small scale patterns in the structure of macrofaunal assemblages of shallow soft sediments. *J. exp. mar. Biol. Ecol.* 237: 127–140
- Kendall, M.G. (1970). *Rank correlation methods*. Griffin, London
- Kenkel, N.C., Orloci, L. (1986). Applying metric and nonmetric multidimensional scaling to some ecological studies: some new results. *Ecology* 67: 919–928
- Krzanowski, W.J. (in press). Multifactorial analysis of distance in studies of ecological community structure. *J. Agric. Biol. Environ. Sci.*
- Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27
- Kruskal, J.B., Wish, M. (1978). *Multidimensional scaling*. Sage Publications, Beverly Hills, California
- Kulczynski, S. (1928). Die Pflanzenassoziationen der Pieninen. *Bull. Int. Acad. Pol. Sci. Lett. Cl. Sci. Math. Nat. Ser. B*, Suppl II: 57–203
- Lambshead, P.J.D. (1986). Sub-catastrophic sewage and industrial waste contamination as revealed by marine nematode faunal analysis. *Mar. Ecol. Prog. Ser.* 29: 247–260
- Lambshead, P.J.D., Platt, H.M., Shaw, K.M. (1983). The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. *J. nat. Hist.* 17: 859–874
- Lance, G.N., Williams, W.T. (1967). A general theory of classificatory sorting strategies: I Hierarchical Systems. *Comp. J.* 9: 373–380
- Legendre, P., Anderson, M.J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69: 1–24
- Legendre, P., Legendre, L. (1998). *Numerical ecology, 2nd Engl. edn*. Elsevier, Amsterdam
- Lorenzen, S. (1994). *The phylogenetic systematics of free-living nematodes*. Ray Society, London
- Magurran, A.E. (1991). *Ecological diversity and its measurement*. Chapman and Hall, London
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209–220
- Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). *Multivariate analysis*. Academic Press, London
- May, R.M. (1990). Taxonomy as destiny. *Nature* 347: 129–130
- McArdle, B.H., Anderson, M.J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82: 290–297
- McEachran, J.D., Miyake, T. (1990). Phylogenetic interrelationships of skates: a working hypothesis (Chondrichthyes, Rajoidea). In: *Elasmobranchs as living resources: advances in the biology, ecology, systematics, and the status of the fisheries*, H.L. Pratt et al (eds). NOAA Technical Report NMFS 90: 285–304
- Morrisey, D.J., Underwood, A.J., Howitt, L. (1996). Effects of copper on the faunas of marine soft-sediments: an experimental field study. *Mar. Biol.* 125: 199–213
- Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Ann. Rev. Genet.* 30: 371–403
- Nelson, J.S. (1994). *Fishes of the world, 3rd edn*. Wiley, New York
- Olafsson, E., Moore, C.G. (1992). Effects of macroepifauna on developing nematode and harpacticoid assemblages in a subtidal muddy habitat. *Mar. Ecol. Prog. Ser.* 84: 161–171
- Olsgard, F., Somerfield, P.J. (2000). Surrogates in marine benthic investigations – which taxonomic unit to target? *J. Aquat. Ecosyst. Stress Recov.* 7: 25–42
- Olsgard, F., Somerfield, P.J., Carr, M.R. (1997). Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. *Mar. Ecol. Prog. Ser.* 149: 173–181
- Olsgard, F., Somerfield, P.J., Carr, M.R. (1998). Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. *Mar. Ecol. Prog. Ser.* 172: 25–36
- Pearson, T.H. (1975). The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. IV. Changes in the benthic fauna attributable to organic enrichment. *J. exp. mar. Biol. Ecol.* 20: 1–41
- Pearson, T.H., Blackstock, J. (1984). Garroch Head sludge dumping ground survey, final report. Dunstaffnage Marine Research Laboratory (unpublished)
- Pearson, T.H., Gray, J.S., Johannessen, P.J. (1983). Objective selection of sensitive species indicative of pollution-induced change in benthic communities. 2. Data analyses. *Mar. Ecol. Prog. Ser.* 12: 237–255
- Pielou, E.C. (1975). *Ecological diversity*. Wiley, New York
- Pielou, E.C. (1984). *The interpretation of ecological data. A primer on classification and ordination*. Wiley, New York
- Piepenburg, D., Voss, J., Gutt, J. (1997). Assemblages of sea stars (Echinodermata: Asteroidea) and brittle stars (Echinodermata: Ophiuroidea) in the Weddell Sea (Antarctica) and off Northeast Greenland (Arctic): a comparison of diversity and abundance. *Polar Biol.* 17: 305–322
- Platt, H.M., Warwick, R.M. (1983). *Freeliving marine nematodes. Part I. British Enoplids*. Synopses of the British Fauna no 28. Cambridge University Press, Cambridge
- Platt, H.M., Warwick, R.M. (1988). *Freeliving marine nematodes. Part II. British chromadorids*. E.J. Brill, Leiden
- Platell, M.E., Potter, I.C., Clarke, K.R. (1998). Resource partitioning by four species of elasmobranchs (Batoidea: Urolophidae) in coastal waters of temperate Australia. *Mar. Biol.* 131: 719–734

- Potter, I.C., Claridge, P.N., Hyndes, G.A., Clarke, K.R. (1997). Seasonal, annual and regional variations in ichthyofaunal composition in the inner Severn Estuary and inner Bristol Channel. *J. mar. biol. Ass. U.K.* 77: 507–525
- Price, A.R.G., Keeling, M.J., O'Callaghan, C.J. (1999). Ocean-scale patterns of 'biodiversity' of Atlantic asteroids determined from taxonomic distinctness and other measures. *Biol. J. Linn. Soc.* 66: 187–203
- Raffaelli, D., Mason, C.F. (1981). Pollution monitoring with meiofauna using the ratio of nematodes to copepods. *Mar. Poll. Bull.* 12: 158–163
- Reid, P.C., Barges, M.dF., Svendsen, E. (in press). A regime shift in the North Sea circa 1988 linked to changes in the North Sea fishery. *Fish. Res.*
- Rogers, S.I., Clarke, K.R., Reynolds, J.D. (1999). The taxonomic distinctness of coastal bottom-dwelling fish communities of the North-east Atlantic. *J. Anim. Ecol.* 68: 769–782
- Sanders, H.L. (1968). Marine benthic diversity: a comparative study. *Am. Nat.* 102: 243–282
- Scheffe, H. (1959). *The analysis of variance*. Wiley, New York
- Schiffman, S.S., Reynolds, M.L., Young, F.W. (1981). *Introduction to multi-dimensional scaling. Theory, methods and applications*. Academic Press, London
- Schratzberger, M., Warwick, R.M. (1998a). Effects of the intensity and frequency of organic enrichment on two estuarine nematode communities. *Mar. Ecol. Prog. Ser.* 164: 83–94
- Schratzberger, M., Warwick, R.M. (1998b). Effects of physical disturbance on nematode communities in sand and mud: a microcosm experiment. *Mar. Biol.* 130: 643–650
- Schratzberger, M., Warwick, R.M. (1999). Differential effects of various types of disturbances on the structure of nematode assemblages: an experimental approach. *Mar. Ecol. Prog. Ser.* 181: 227–236
- Schwinghamer, P. (1981). Characteristic size distributions of integral benthic communities. *Can. J. Fish. aquat. Sci.* 38: 1255–1263
- Seber, G.A.F. (1984). *Multivariate observations*. Wiley, New York
- Shimatani, K. (2001). On the measurement of species diversity incorporating species differences. *Oikos* 93: 135–147
- Shephard, R.N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika* 27: 125–140
- Simpson, E.H. (1949). Measurement of diversity. *Nature* 163: 688
- Sneath, P.H.A. (1957). The application of computers to taxonomy. *J. gen. Microbiol.* 17: 201–226
- Sneath, P.H.A., Sokal, R.R. (1973). *Numerical taxonomy*. Freeman, San Francisco
- Sokal, R.R., Rohlf, F.J. (1981). *Biometry*. Freeman, San Francisco
- Somerfield, P.J., Gage, J.D. (2000). Community structure of the benthos in Scottish sea-lochs. IV. Multivariate spatial pattern. *Mar. Biol.* 136: 1133–1145
- Somerfield, P.J., Gee, J.M., Warwick, R.M. (1994a). Benthic community structure in relation to an instantaneous discharge of waste water from a tin mine. *Mar. Pollut. Bull.* 28: 363–369
- Somerfield, P.J., Gee, J.M., Warwick, R.M. (1994b). Soft sediment meiofaunal community structure in relation to a long-term heavy metal gradient in the Fal estuary system. *Mar. Ecol. Prog. Ser.* 105: 79–88
- Somerfield, P.J., Gee, J.M., Widdicombe, S. (1993). *The use of meiobenthos in marine pollution monitoring programmes*. Plymouth Marine Laboratory Miscellaneous Publication LIB-33A,B, Plymouth
- Somerfield, P.J., Clarke, K.R. (1995). Taxonomic levels, in marine community studies, revisited. *Mar. Ecol. Prog. Ser.* 127: 113–119
- Somerfield, P.J., Clarke, K.R. (1997). A comparison of some methods commonly used for the collection of sublittoral sediments and their associated fauna. *Mar. Environ. Res.* 43: 143–156
- Somerfield, P.J., Olsford, F., Carr, M.R. (1997). A further examination of two new taxonomic distinctness measures. *Mar. Ecol. Prog. Ser.* 154: 303–306
- Somerfield, P.J., Rees, H.L., Warwick, R.M. (1995). Interrelationships in community structure between shallow-water marine meiofauna and macrofauna in relation to dredgings disposal. *Mar. Ecol. Prog. Ser.* 127: 103–112
- ter Braak, C.F.J. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167–1179
- Underwood, A.J. (1981). Techniques of analysis of variance in experimental marine biology and ecology. *Oceanogr. Mar. Biol. Ann. Rev.* 19: 513–605
- Underwood, A.J. (1992). Beyond BACI: the detection of environmental impact on populations in the real, but variable, world. *J. exp. Mar. Biol. Ecol.* 161: 145–178
- Underwood, A.J. (1997). *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, Cambridge
- Underwood, A.J., Chapman, M.G. (1998). A method for analysing spatial scales of variation in composition of assemblages. *Oecologia* 117: 570–578
- Underwood, A.J., Peterson, C.H. (1988). Towards an ecological framework for investigating pollution. *Mar. Ecol. Prog. Ser.* 46: 227–234
- Vane-Wright, R.I., Humphries, C.J., Williams, P.H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.* 55: 235–254
- Ward, P., Woodd-Walker, R., Clarke, A. (submitted). Seasonality rather than temperature regulates zooplankton diversity. *Nature*
- Warwick, R.M. (1971). Nematode associations in the Exe estuary. *J. mar. Biol. Ass. U.K.* 51: 439–454
- Warwick, R.M. (1984). Species size distributions in marine benthic communities. *Oecologia (Berlin)* 61: 32–41
- Warwick, R.M. (1986). A new method for detecting pollution effects on marine macrobenthic communities. *Mar. Biol.* 92: 557–562
- Warwick, R.M. (1988a). Effects on community structure of a pollutant gradient – summary. *Mar. Ecol. Prog. Ser.* 46: 207–211
- Warwick, R.M. (1988b). The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar. Pollut. Bull.* 19: 259–268

- Warwick, R.M. (1993). Environmental impact studies on marine communities: pragmatical considerations. *Aust. J. Ecol.* 18: 63-80
- Warwick, R.M., Ashman, C.M., Brown, A.R., Clarke, K.R., Dowell, B., Hart, B., Lewis, R.E., Shillabeer, N., Somerfield, P.J., Tapp, J.F. (in press). Inter-annual changes in the biodiversity and community structure of the macrobenthos in Tees Bay and the Tees estuary, UK, associated with local and regional environmental events. *Mar. Ecol. Prog. Ser.*
- Warwick, R.M., Buchanan, J.B. (1970). The meiofauna off the coast of Northumberland. I: The structure of the nematode population. *J. mar. biol. Assoc. UK* 50: 129-146
- Warwick, R.M., Carr, M.R., Clarke, K.R., Gee, J.M., Green, R.H. (1988). A mesocosm experiment on the effects of hydrocarbon and copper pollution on a sublittoral soft-sediment meiobenthic community. *Mar. Ecol. Prog. Ser.* 46: 181-191
- Warwick, R.M., Clarke, K.R. (1991). A comparison of methods for analysing changes in benthic community structure. *J. mar. Biol. Ass. U.K.* 71: 225-244
- Warwick, R.M., Clarke, K.R. (1993a). Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. *Mar. Ecol. Prog. Ser.* 92: 221-231
- Warwick, R.M., Clarke, K.R. (1993b). Increased variability as a symptom of stress in marine communities. *J. exp. mar. Biol. Ecol.* 172: 215-226
- Warwick, R.M., Clarke, K.R. (1994). Relearning the ABC: taxonomic changes and abundance/biomass relationships in disturbed benthic communities. *Mar. Biol.* 118: 739-744
- Warwick, R.M., Clarke, K.R. (1995a). New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Mar. Ecol. Prog. Ser.* 129: 301-305
- Warwick, R.M., Clarke, K.R. (1995b). Multivariate measures of community stress and their application to marine pollution studies in the East Asian region. *Phuket mar. biol. Cent. Res. Bull.* 60: 99-113
- Warwick, R.M., Clarke, K.R. (1998). Taxonomic distinctness and environmental assessment. *J. appl. Ecol.* 35: 532-543
- Warwick, R.M., Clarke, K.R. (2001). Practical measures of marine biodiversity based on relatedness of species. *Oceanogr. Mar. Biol. Ann. Rev.* 39: 207-231
- Warwick, R.M., Clarke, K.R., Gee, J.M. (1990a). The effects of disturbance by soldier crabs, *Mictyris platycheles* H. Milne Edwards, on meiobenthic community structure. *J. exp. Mar. Biol. Ecol.* 135: 19-33
- Warwick, R.M., Clarke, K.R., Suharsono. (1990b). A statistical analysis of coral community responses to the 1982-3 El Nino in the Thousand Islands, Indonesia. *Coral Reefs* 8: 171-179
- Warwick, R.M., Coles, J.W. (1977). The marine flora and fauna of the Isles of Scilly. Free-living Nematoda. *J. Nat. Hist.* 11: 393-407
- Warwick, R.M., Light, J. (in press). Death assemblages of molluscs on St. Martin's Flats, Isles of Scilly: a surrogate for regional biodiversity? *Biodivers. Conserv.*
- Warwick, R.M., Pearson, T.H., Ruswahyuni (1987). Detection of pollution effects on marine macrobenthos: further evaluation of the species abundance/biomass method. *Mar. Biol.* 95: 193-200
- Warwick, R.M., Platt, H.M., Clarke, K.R., Agard, J., Gobin, J. (1990c). Analysis of macrobenthic and meiobenthic community structure in relation to pollution and disturbance in Hamilton Harbour, Bermuda. *J. exp. Mar. Biol. Ecol.* 138: 119-142
- Warwick, R.M., Platt, H.M., Somerfield, P.J. (1998). *Freeliving marine nematodes. Part III. British Monhysterida*. Synopses of the British Fauna no 53. Field Studies Council, Shrewsbury
- Widdicombe, S., Austen, M.C. (1998). Experimental evidence for the role of *Brissopsis lyrifera* (Forbes, 1841) as a critical species in the maintenance of benthic diversity and the modification of sediment chemistry. *J. exp. Mar. Biol. Ecol.* 228: 241-255
- Widdicombe, S., Austen, M.C. (2001). The interaction between physical disturbance and organic enrichment: an important element in structuring benthic communities. *Limnol. Oceanog.* 46: 1720-1733
- Wilkinson, D.M. (1999). The disturbing history of intermediate disturbance. *Oikos* 84: 145-147
- Williams, P.H., Humphries, C.J., Vane-Wright, R.I. (1991). Measuring biodiversity: taxonomic relatedness for conservation priorities. *Aust. Syst. Bot.* 4: 665-679

ACKNOWLEDGEMENTS

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

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

K R Clarke
R M Warwick
2001

