1.1 Introduction

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. A different type of test is required for groups identified *a posteriori*.

3) *Determining levels of stress* or disturbance, by attempting to construct biological measures from the community data which are indicative of disturbed conditions. These may be absolute measures (“this observed structural feature is indicative of pollution”) or relative criteria (‘under impact, this coefficient is expected to decrease in comparison with control levels’). Note the contrast with the previous stage, which is restricted to demonstrating differences between groups of samples, not ascribing directional 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 (species richness), perhaps for a given number of individuals, 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* that 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* (Lambshead, Platt & Shaw (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 *SAD curves* or 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 & 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* 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.

Methods of this type in the manual include: *hierarchical agglomerative clustering* (see Everitt (1980)) in which samples are successively fused into larger groups; *binary divisive clustering*, in which groups are successively split; and two types of ordination method, *principal components*

analysis (PCA, e.g. Chatfield & Collins (1980)) and *non-metric/metric multi-dimensional scaling* (nMDS/mMDS, the former often shortened to MDS, Kruskal & Wish (1978)).

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

¶ *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.*

§ *These terms tend to be used interchangeably by ecologists, so we will do that also, but in statistical language the methods given here are all clustering techniques, classification usually being reserved for classifying unknown new samples into known prior group structures.*

Revision #20
Created 9 February 2022 09:37:52 by Arden
Updated 16 October 2024 02:08:00 by Marti