11.1 Introduction

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 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, Clarke & Warwick (1982)) or in combination (Clarke & 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.

Environmental data analysis

An analogous range of multivariate methods is available for display and testing of environmental samples as has been described for biotic data: species are simply replaced by physical/chemical variables. However, the matrix entries are now of a rather different type and lead to different analysis choices. No longer do zeros predominate; the readings are usually more nearly continuous and, though their distributions are often right-skewed (with variability increasing with the mean), it is often possible to transform them to approximate normality (and stabilise the variance) by a simple root or logarithmic transformation, see Chapter 9. Under these conditions, Euclidean distance is an appropriate measure of dissimilarity and PCA (Chapter 4) is an effective ordination technique, though note that this will need to be performed on the correlation rather than the covariance matrix, i.e. the variables will usually have different units of measurement and need normalising to a common scale (see the discussion on page 4.4).

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.

[¶] *Methods such as canonical correlation (e.g. Mardia, Kent & Bibby (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.*

[§] *The ANOSIM tests in the PRIMER package are not now the only possibility; the data will have been transformed to approximate normality so classical multivariate (MANOVA) tests such as Wilks' Λ (e.g. Mardia, Kent & Bibby (1979)) may be valid, but only if the number of variables is small in relation to the number of samples.*

Revision #7
Created 4 March 2022 06:08:44 by Arden
Updated 6 November 2024 01:55:13 by Abby Miller