11 Ordination — First encounter

This chapter introduces readers with no experience of ordination methods before to its underlying concepts. If you are at all familiar with ordination, we suggest you may want to go straight to Chapter 12, as the method we use here, Bray–Curtis ordination, is rarely used. However, it is an excellent tool to explain the underlying idea of ordination.

In Chapter 10 several measures of association were applied on the Argentinean zoobenthic data. These gave a 6-by-6 matrix, and as long as this matrix is reasonably small, we can directly compare the numbers by looking at them to see which species are the most similar. But what do you do if the matrix has dimensions 10-by-10, or 500-by-500? Unless you have an excellent memory for numbers interpretation becomes difficult, if not impossible, by just looking at the matrix. In this chapter, and the next four chapters, we look at techniques that provide a graphical representation of either, the N-by-N association matrix, or the original data matrix. There are different ways of doing this, and the methods can be split into ordination or clustering. We only look at ordination. The aim of ordination is twofold. First, it tries to reduce a large number of variables into a smaller number of easier to interpret variables. And, secondly, it can be used to reveal patterns in multivariate data that would not be identified in univariate analyses.

Ordination methods give easy-to-read graphical outputs from relatively easy-to-use software, and this may partly explain their popularity.

We start by explaining the underlying idea of ordination. Probably, the best way to do this is to use one of the oldest techniques available: Bray–Curtis ordination. The mathematical calculations required for this method are so simple that they can be done with pen and paper.

11.1 Bray–Curtis ordination

Euclidean distances among the six sites for an Argentinean zoobenthic data set were presented in Table 10.7. We have reproduced these distances in Table 11.1, and these will be used to illustrate Bray–Curtis ordination. The choice for the Euclidean distance function is purely didactical; it provides higher contrast among sites. It is by no means the best choice for these data. Our goal is to visualise the distances along a single line. We start with a 6-by-6 matrix with dissimilarities, and then consider how we can graph them in a way that the positions of the sites on the map will say something about the relationships among the sites.

Table 11.1. Euclidean distances among the six sites for the Argentinean data used in Table 10.1. The larger the value, the more dissimilar are the two sites.

	1	2	3	4	5	6
1	0	377.11	504.73	248.70	220.96	412.07
2		0	785.96	147.50	213.83	718.32
3			0	689.66	582.31	116.98
4				0	165.02	613.87
5					0	512.54
6						0

The way Bray–Curtis ordination (or Polar ordination) does this, is as follows. Imagine sitting in a room with 10 other people. There is one person you like (person A), and one you don't like at all (person B). Place them at either side of the room, and draw a straight line between them. Now, place all other eight people along this line in such a way that distances between each person, and persons A and B, reflect how much you like, or dislike, them. So, all we do is identify the two extremes, and use these as a reference point for all other subjects. It sounds simple, and indeed it is.

For the zoobenthic data, Bray–Curtis ordination identifies the two sites that are the most dissimilar. These will have the highest value in Table 11.1, which is 785.96, and between sites 2 and 3. So, these are the two most dissimilar sites. The method places these two sites at the ends of a line (Figure 11.1). The length of this line is 1 unit. We then need to place the four remaining sites along this line. To determine the position of site along the line, say site 1, we have to estimate the distance between this site and each of the ones at the ends of the line. This can be done by drawing two circles. The first circle has its centre at the left end point, and the second circle has the right end point as its centre. The radius of the first circle is given by the Euclidean distance between sites 3 and 1, which is 504.73, but expressed as a fraction of 785.96 (the line has unit length, but represents the distance of 785.96). The radius of the second circle is determined by the Euclidean distance between sites 2 and 1, which is 377.11 (again expressed as a fraction of 785.96). The position of site 1 on the line is determined by vertical projection of the interception of the two circles on the line (Figure 11.1). So it represents a compromise between two distances to both sites at the end of the line. The same process is repeated for the other sites. Further axes can be calculated; see McCune and Grace (2002), and Beals (1984) for details.

Instead of the Euclidean distances, various other measures of similarity can be used (Chapter 10).

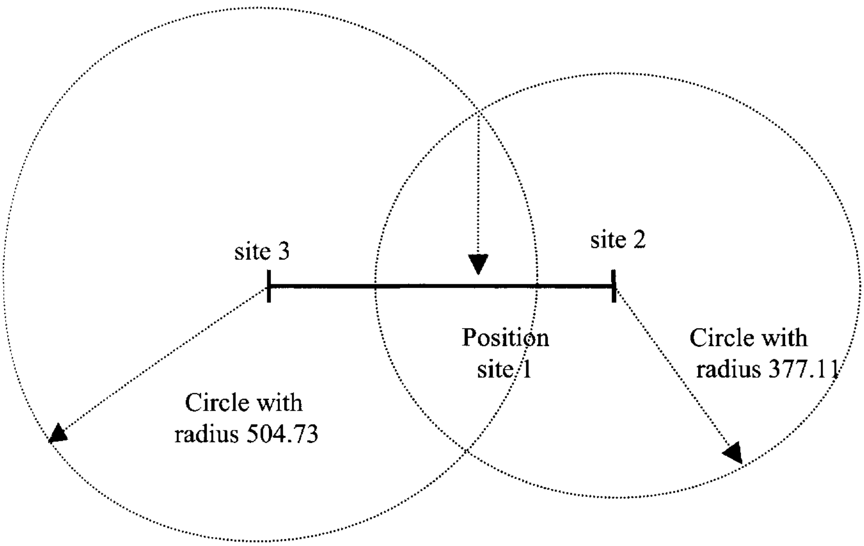

Figure 11.1. Underlying principle of Bray–Curtis ordination. The sites that are the most dissimilar are 3 and 2 and are plotted at both ends of a line with unit length. The other sites are placed on this line as compromise between distances to these two sites. An illustration for site 1 is presented.

There are different ways of selecting the end points of the line (or axis). The original method is shown in Figure 11.1, where the pair of variables that had the highest dissimilarity were selected as end points. However, this approach tends to give an axis that has one or two sites at one end of the axis, and all other sites at the other end of the axis, with isolated sites likely to be outliers. Indeed, Bray–Curtis ordination with the original method to select end points is a useful tool to identify outliers (provided these outliers have a low similarity with the other sites). An alternative is to select your own end points. This is called subjective end point selection. It allows you to test whether certain sites are outliers. The third option was developed by Beals (1984) and is called 'variance-regression'. The first end point calculated by this method is the point that has the highest variance of distances to all other points. An outlier will have large distances to most other points, and therefore the variation in these distances is small. Getting the second end point is slightly more complicated. Suppose that site 1 is the first end point. The algorithm will in turn consider each of the other sites as alternative end points. In the first step, it will take site 2 as the second end point. Let D_{1i} be the distance of site 1 to all other sites (except for sites 1 and 2). Furthermore, let D_{2i} be the distance of site 2 to all other sites. Then regress D_{1i} on D_{2i} and store the slope. Repeat this process for all other trial end points. The second end point will be the site that has the most negative slope. McCune and Grace (2002) justify this approach by saying that the second end point is at the edge of the main cloud of sites, opposite to end point 1.

The Bray–Curtis ordination using the zoobenthic data and the Euclidean distances in Table 11.1 is presented in Figure 11.2. The positions of the sites along the axis are represented by dots. The site names are plotted with an angle of 45 degrees above and below the points. Results indicate that sites 3 and 6 are similar, but dissimilar from the other 4 sites. The variance explained by the axis is 98%. Further technical details can be found in McCune and Grace (2002).

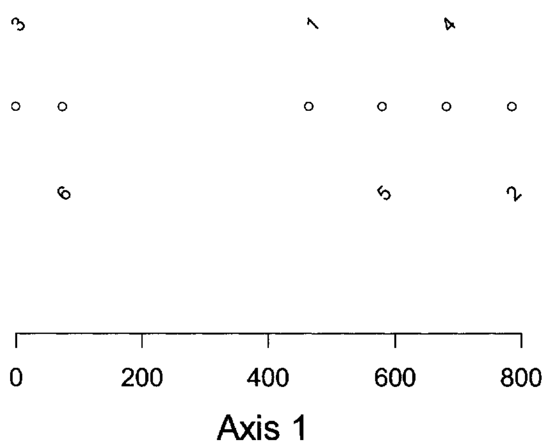

Figure 11.2. Bray–Curtis ordination using the Euclidean distance, applied to the zoobenthic data from Argentina. The original method to select end point was used, and 1 axis was calculated.

Bray–Curtis ordination is one of the oldest ordination methods in ecology and is a good didactical tool to introduce the concept of ordination. However, it might be difficult to get a scientific paper published that relies on a Bray–Curtis ordination. Since the 1950s, more complicated multivariate techniques have been developed and applied in ecology, e.g., principal component analysis, correspondence analysis, redundancy analysis (RDA), canonical correspondence analysis (CCA), canonical correlation analysis, partial RDA, partial CCA, variance partitioning or discriminant analysis, among many others. These techniques are discussed in the next four chapters.