CHAPTER 8

Analysis of differences in species composition

Analysis of differences in species composition

This chapter describes how the difference in species composition can be investigated by calculating the ecological distance between two sites. The methods can be applied to an entire species matrix by calculating ecological distances between all pairs of sites.

The results can be presented as a distance or dissimilarity matrix. In later chapters on clustering and ordination, some methods are shown for analysing such distance matrices.

How can differences in species composition be investigated?

This chapter describes some methods for measuring the difference in species composition between two sites. Rather than stating the difference in abundance between each and every species, a single statistic is calculated that expresses the difference in species composition. Consider Figure 8.1 as an example. You could report that site A and B share the same species S1, S2 and S3 with the same abundance, whereas species S4 and S5 only occur on site B. An ecological distance will summarize these differences in a single distance statistic. In the case of the Bray-Curtis distance (there are many different methods of calculating a distance), you would calculate that the difference between site A and B is 0.25.

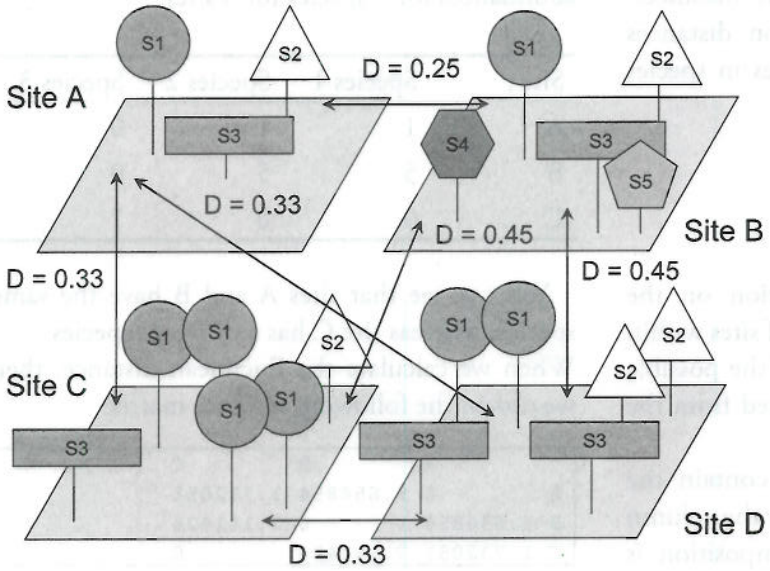

Figure 8.1 Four sites with different species composition. The differences in species composition between each subset of two sites can be expressed by a single ecological distance such as the Bray-Curtis distance. The Bray-Curtis distance between A and B is 0.25, and between A and C it is 0.33.

The advantage of using an ecological distance is that differences in species composition can be summarized with a single statistic. The disadvantage of using an ecological distance is that information on the identities of the species is not available any longer. For some research objectives, loosing information on species identities does not pose problems, whereas other equally valid objectives require that information is available on the species identities. We will see in chapter 10 that species identities can be added to an ordination diagram, although many ordination methods use distance matrices. If you are interested in fully exploring how abundance or presence-absence of particular species changes in between sites, then you need to conduct a separate analysis for each species (chapter 6 and 7).

Ecological distance

A good ecological distance describes the difference in species composition. For sites that share most of their species, the ecological distance should be small. When sites have few species in common, the ecological distance should be large. There is no single way to define ecological distance. The literature lists a large number of distances. We only present a subset of the possible distances. These are among the most common distances that are used for analysing differences in species composition.

Distance matrices

Distance matrices provide information on the ecological distance between all pairs of sites within your data. Table 8.1 provides one of the possible distance matrices that can be calculated from the dune meadow dataset.

The cells in the distance matrix contain the distance between the sites indicated by the column and row names. As the species composition is exactly the same when you compare a site with itself, its ecological distance is zero. Also the order in which you make the comparison does not matter: the distance between X1 and X2 is the same as that between X2 and X1.

Euclidean distance

The Euclidean distance is calculated by using each species as a different axis to plot each site and then measuring the distance between the sites (see Figure 8.2). Formulae for calculating the Euclidean distance and other distances are provided in Box 8.1 (page 129).

The Euclidean distance is not a good ecological distance if it is used on raw species matrices – matrices that contain the abundance of each species on each site. When the species matrix is modified by a particular transformation (see below: standardizations of the species data before calculating the distance matrix), then the Euclidean distance becomes better at expressing ecological distance.

The following example illustrates that the Euclidean distance on the raw species matrix will not always describe ecological distance well. Imagine that you recorded the following abundances for 3 species for 3 sites:

Site	Species 1	Species 2	Species 3
A	1	1	0
B	5	5	0
C	0	0	1

You can see that sites A and B have the same species, whereas site C has a different species. When we calculate the Euclidean distance, then we obtain the following distance matrix:

```
          A        B        C
A         0 5.656854 1.732051
B 5.656854        0 7.141428
C 1.732051 7.141428        0
```

Table 8.1 Distance matrix for the dune meadow dataset using the Bray-Curtis distance

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
X1	0.000	0.467	0.448	0.524	0.639	0.636	0.552	0.655	0.600	0.574	0.560	0.925	0.843	1.000	1.000	0.922	0.879	0.778	1.000	1.000
X2	0.467	0.000	0.341	0.356	0.412	0.511	0.439	0.537	0.476	0.294	0.541	0.714	0.600	0.788	0.908	0.893	0.825	0.594	0.808	0.945
X3	0.448	0.341	0.000	0.271	0.470	0.568	0.475	0.325	0.341	0.470	0.556	0.440	0.425	0.750	0.714	0.671	0.891	0.612	0.831	0.775
X4	0.524	0.356	0.271	0.000	0.500	0.634	0.506	0.412	0.379	0.477	0.584	0.525	0.513	0.797	0.735	0.667	0.900	0.667	0.789	0.763
X5	0.639	0.412	0.470	0.500	0.000	0.297	0.229	0.639	0.506	0.349	0.627	0.692	0.684	0.881	0.848	0.895	0.621	0.543	0.703	0.892
X6	0.636	0.511	0.568	0.634	0.297	0.000	0.227	0.591	0.600	0.319	0.450	0.639	0.753	0.806	0.803	0.852	0.683	0.493	0.722	0.848
X7	0.552	0.439	0.475	0.506	0.229	0.227	0.000	0.525	0.488	0.277	0.444	0.627	0.644	0.875	0.841	0.890	0.673	0.552	0.746	0.887
X8	0.655	0.537	0.325	0.412	0.639	0.591	0.525	0.000	0.317	0.542	0.528	0.440	0.370	0.562	0.429	0.425	0.891	0.642	0.746	0.493
X9	0.600	0.476	0.341	0.379	0.506	0.600	0.488	0.317	0.000	0.600	0.595	0.351	0.413	0.758	0.662	0.653	0.895	0.681	0.781	0.699
X10	0.574	0.294	0.470	0.477	0.349	0.319	0.277	0.542	0.600	0.000	0.413	0.718	0.737	0.761	0.848	0.895	0.621	0.486	0.703	0.892
X11	0.560	0.541	0.556	0.584	0.627	0.450	0.444	0.528	0.595	0.413	0.000	0.672	0.754	0.821	0.745	0.877	0.702	0.322	0.556	0.810
X12	0.925	0.714	0.440	0.525	0.692	0.639	0.627	0.440	0.351	0.718	0.672	0.000	0.353	0.695	0.621	0.588	0.920	0.742	0.697	0.697
X13	0.843	0.600	0.425	0.513	0.684	0.753	0.644	0.370	0.413	0.737	0.754	0.353	0.000	0.649	0.679	0.606	0.875	0.800	0.812	0.719
X14	1.000	0.788	0.750	0.797	0.881	0.806	0.875	0.562	0.758	0.761	0.821	0.695	0.649	0.000	0.362	0.544	0.897	0.843	0.855	0.455
X15	1.000	0.908	0.714	0.735	0.848	0.803	0.841	0.429	0.662	0.848	0.745	0.621	0.679	0.362	0.000	0.357	0.895	0.720	0.778	0.296
X16	0.922	0.893	0.671	0.667	0.895	0.852	0.890	0.425	0.653	0.895	0.877	0.588	0.606	0.544	0.357	0.000	1.000	0.867	0.906	0.344
X17	0.879	0.825	0.891	0.900	0.621	0.683	0.673	0.891	0.895	0.621	0.702	0.920	0.875	0.897	0.895	1.000	0.000	0.762	0.565	0.913
X18	0.778	0.594	0.612	0.667	0.543	0.493	0.552	0.642	0.681	0.486	0.322	0.742	0.800	0.843	0.720	0.867	0.762	0.000	0.552	0.690
X19	1.000	0.808	0.831	0.789	0.703	0.722	0.746	0.746	0.781	0.703	0.556	0.697	0.812	0.855	0.778	0.906	0.565	0.552	0.000	0.742
X20	1.000	0.945	0.775	0.763	0.892	0.848	0.887	0.493	0.699	0.892	0.810	0.697	0.719	0.455	0.296	0.344	0.913	0.690	0.742	0.000

One feature of our distance matrix is that information about the original species is not longer present – the distance matrix does not mention Species 1, Species 2 or Species 3 anywhere. Some textbooks mention in this case that the analysis is in Q-mode rather than in R-mode. This simply means that differences in sites are being investigated (differences between the rows of the species matrix) rather than differences in species (differences between the columns of the species matrix) – but remember that the differences were actually derived from differences in species composition. Since ecological datasets often contain more species than sites, the distance matrix will often be of smaller size than the species matrix.

Because of the properties of distance matrices that the distance between the same sites is zero and that the order of calculating the distance does not matter, a distance matrix can be summarized with no loss of information as:

```
          A          B
B 5.656854
C 1.732051 7.141428
```

You can see in the distance matrix that the distance between A and C is about 1.7, whereas the distance between A and B is roughly 5.7. But remember now that A and C do not share any species, whereas A and B have the same species. The Euclidean distance depends greatly on the abundances of each species, not just which species are shared. A and B are far apart using Euclidean distance because A has 2 plants and B has 10. However in most applications we would like to give more emphasis to the extent to which species are shared and give more weight to differences in composition than abundances of the same species.

Although the Euclidean distance is not a very good distance for investigating how species are shared between sites, it is still used in some ordination and clustering techniques as these only allow for this distance.

The fact that the Euclidean distance is not a good distance for all situations also constrains one graphical method of representing differences in species composition. Each species can be represented by one axis. Sites can then be plotted by using the abundance of each site as coordinates. For our example, we could graphically represent our sites as in Figure 8.2.

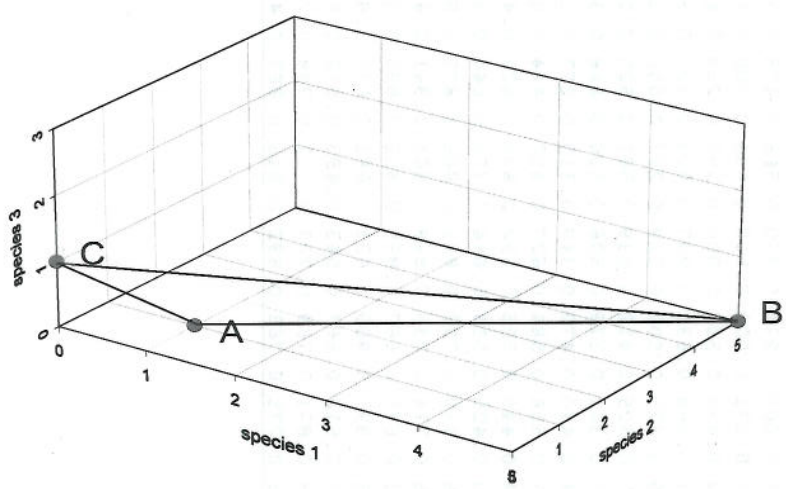

Figure 8.2 By using each species as an axis, you can position sites on a plot that reveals their ecological distance.

When you measure the straight line distance between the sites in Figure 8.2, you will obtain the Euclidean distance. You can see again that site A is closer to C than to B.

One of the solutions to give greater weight to differences in composition and still to use the Euclidean distance is to calculate species proportions first and then calculate the Euclidean distance (see below: standardizations of the species data before calculating the distance matrix). Another method would be to use presence-absence, transforming all non-zero abundance values to 1. The Euclidean distance is then the same as a count of the number of species that occur in one site but not both. The solution that is more commonly used is to adopt another method of calculating distance as shown in the next section.

Other distances

Distances can be classified in various ways. One classification method can be by the range of output values that can be expected.

Some distances are restricted to be within the range of zero to one. When the distance is zero, two sites are completely similar for every species. When the distance equals 1 they are completely dissimilar, which means that they do not share any species. The **Bray-Curtis distance** (or Odum distance, or the one-complement of the Steinhaus similarity) and **Kulczynski distance** fall within this category.

When we calculate the Bray-Curtis distance for the dataset that we described earlier, we obtain:

```
          A         B
B 0.6666667
C 1.0000000 1.0000000
```

Similarly, for the Kulczynski distance, we obtain:

```
    A   B
B 0.4
C 1.0 1.0
```

You can see now that the distance of site C from the other sites is 1. This indicates that site C does not share any species with the other sites. If site C shared any species with the other sites, then the distance would be smaller than 1. The distance between A and B is smaller, which is what we wanted since we know that these sites share some species.

Another thing that you can observe is that you obtain different values for the distance between A and B when using the Bray-Curtis distance and when using the Kulczynski distance. You are thus faced with having to make a choice of a particular distance as different results are obtained. Although there are some methods for comparing distances (see below: choice of a distance), you will in general need to make a prior choice of the distance that you want to use.

The Bray-Curtis and Kulczynski distances are calculated from differences in abundance of each species. Because of this calculation method, the final distance will be influenced more by species with largest differences in abundances. When some species are dominant in your dataset (for example that one species has 90% of differences in abundance among sites), the Bray-Curtis and Kulczynski distances will mainly reflect differences for those species only. For this reason, some researchers prefer to transform the species matrix by a square-root, double square-root (fourth-root) or logarithmic transformation (see chapter 2), so that dominant species will influence the analysis less.

Some other distances will not be constrained to a maximum value of one, but can have larger distances. The Euclidean distance is one of those measures. Other distances of this category that are better in representing differences in species composition include the **Hellinger** and the **Chi-square** distance. Both depend on differences in proportions of species between the two sites. The Chi-square distance is actually not that good for species data, but it is the distance that is used in some common ordination methods (see chapter 10). The Hellinger distance will

normally perform better (be a better reflection of ecological distance) than the Chi-square distance. Both the Chi-square and Hellinger distance will be influenced differently by the species with smaller abundances than by the species with larger abundances. Some researchers find this a feature that is not desirable, since species with smaller abundances are usually not well sampled and species with smaller abundances often contribute more to the Chi-square and Hellinger distances. We would expect that species with larger abundances are better sampled, and express differences among sites better. This could be a reason to opt for the Bray-Curtis or Kulczynski distance.

The results for the example that we used earlier are for the Hellinger distance:

```
          A        B
B 0.000000
C 1.414214 1.414214
```

For the Chi-square distance, we obtain:

```
             A        B
B 1.570092e-16
C     3.752777 3.752777
```

One thing that we can observe immediately is that the distance between A and B is 0 (with a small calculation difference for the Chi-square distance). This means that they share exactly the same species, and that the proportions of each species are the same (0.5 in this example). Distances with C are also exactly the same for sites A and B. The Chi-square and Hellinger distances calculate different values, so this means again that you will need to choose which measure to use (but see below: choice of a distance measure).

Box 8.1 How to calculate ecological distance?

We will calculate the ecological distance between sites A and C of Figure 8.1. Site A has three species (S1, S2 and S3) each with 1 tree, whereas site C has 4 trees of species S1, 1 tree of S2 and 1 tree of S3. In the formulae, the abundances of the species of site A are indicated by a_1, a_2 and a_3, and the abundances of the species of site C are indicated by c_1, c_2 and c_3. If a species does not occur on a specific site, it should be listed for the site with abundance = 0.

The formulae for calculating the ecological distances are:

Bray-Curtis: $$D = 1 - 2\frac{\sum_{i=1}^{S}\min(a_i, c_i)}{\sum_{i=1}^{S}(a_i + c_i)}$$

Kulczynski: $$D = 1 - \frac{1}{2}\left(\frac{\sum_{i=1}^{S}\min(a_i, c_i)}{\sum_{i=1}^{S}a_i} + \frac{\sum_{i=1}^{S}\min(a_i, c_i)}{\sum_{i=1}^{S}c_i}\right)$$

Euclidean: $$D = \sqrt{\sum_{i=1}^{S}(a_i - c_i)^2}$$

Chi-square: $$D = \sqrt{\sum_{i=1}^{S}\frac{(a_+ + c_+)}{(a_i + c_i)}\left(\frac{a_i}{a_+} - \frac{c_i}{c_+}\right)^2} \text{ with } a_+ = \sum_{i=1}^{S}a_i$$

Hellinger: $$D = \sqrt{\sum_{i=1}^{S}\left(\sqrt{\frac{a_i}{a_+}} - \sqrt{\frac{c_i}{c_+}}\right)^2} \text{ with } a_+ = \sum_{i=1}^{S}a_i$$

For the Bray-Curtis distance, inserting the abundances for each species in the formula gives:

$$D = 1 - 2\frac{\min(1,4) + \min(1,1) + \min(1,1)}{(1+4) + (1+1) + (1+1)} = 1 - 2\frac{1+1+1}{5+2+2} = 1 - 2\frac{3}{9} = 1 - \frac{6}{9} = \frac{3}{9}$$

Note that use of the functions within Biodiversity.R will directly calculate the distance.

Some distances only look at differences in presence or absence of a species. Some of these distances measures are the Jaccard and Sorenson distance. These also fall into the category of the indices that are constrained within the 0-1 interval. The Sorenson distance is actually the Bray-Curtis when the species matrix has been transformed to presence-absence or 1/0. These distances are not influenced by differences in species abundance between samples, so that species that have larger abundances will not carry a larger weight in the analysis. Often however, this is not a characteristic that you want in your analysis – observing just one individual may not mean very much, whereas observing many individuals on one site and none on another site may be much less likely to be by random chance only. As calculation time with modern computers can be short, you could actually test whether abundance and presence-absence data result in the same patterns when you analyse your data.

Standardizations of the species data before calculating the distance matrix

Except for presence-absence data, differences in abundances among species will influence the calculation of the distances. This is a desirable characteristic of some distances, but for some datasets that are strongly dominated by a few species this may cause a constraint. When most of the abundance is taken by a few species, the distance may only express the differences of sites for this small subset of species. In such cases, you could diminish the influence of strongly dominant species by first taking logarithms, taking square-root, whereas some scientists have even advocated taking 4th roots of all the values in your data. As the typical species matrix contains many zeroes, you would need to take a $\log(n+1)$ as described in chapter 2.

You can have a similar effect of a strong dominance in your dataset by sites that contain a larger number of individuals. An approach that standardizes each site to the same abundance is to divide cell abundance by the total abundance for each site (so that species sum up to 1 for each site). When you use this method, you will compare differences in **species proportions** for each site. For example, for our original dataset, the species matrix with the proportions will be:

Site	Species 1	Species 2	Species 3	Total
A	0.5	0.5	0	1
B	0.5	0.5	0	1
C	0	0	1	1

You can now see that site A and B have exactly the same values as differences in total abundance are not longer there. The data of the species proportions for each site have been described as **species profiles** in some texts. By investigating species profiles, you are sure that differences in total abundance among sites will not influence the results. When looking at the results shown in Figure 8.1, you could have noticed that an ecological distance of 0.33 was calculated between sites A and D by the Bray-Curtis distance, although both sites contain the same species and the same proportions of each species. If the species matrix would have been standardized, then the ecological distance between sites A and D would have become 0.

It is your choice to determine whether ecological distance should reflect the raw abundances (and thus also differences in site totals) or species proportions. When you opt to standardize the species matrix, you could investigate differences in total abundance among sites by the regression techniques described in an earlier chapter as an additional analysis. Dividing by the site total has the added advantage that some distances will now provide you with similar outcomes. For example,

the results for the Bray-Curtis and Kulczynski distances will be the same for species profiles.

Some standardizations of the species data have the feature that when the Euclidean distance is calculated from the standardized data, this distance will be a more suitable ecological distance **for the original dataset**. Such standardization approaches have been described for the distance between species profiles, the Hellinger distance, the Chi-square distance and the chord distance (Legendre and Gallagher 2001). For example, if we calculate the Euclidean distance for the matrix with proportions (species profiles), then we obtain the following distance matrix:

```
          A        B
B 0.000000
C 1.224745 1.224745
```

You can verify that site A is closer to B than to C. Remember that the Euclidean distance matrix for the original data shown in the beginning of this chapter indicated that site A was closer to C than to B, which was not a preferred method of indicating differences in species composition.

Choice of a distance measure

One desirable characteristic of an ecological distance could be that sites that do not share any species are all given the same maximum distance. The distances that we described here that have this property are the Bray-Curtis and the Kulczynski distances. They should thus be preferred for analysing differences in species composition according to this criterion.

The behaviour of a distance can be analysed with **artificial datasets**. Imagine for instance that you have 10 sites with the following species abundances:

	Species 1	Species 2	Species 3	Species 4	Species 5	Site index
Site 1	7	1	0	0	0	1
Site 2	4	2	0	1	0	2
Site 3	2	4	0	1	0	3
Site 4	1	7	0	0	0	4
Site 5	0	8	0	0	0	5
Site 6	0	7	1	0	0	6
Site 7	0	4	2	0	2	7
Site 8	0	2	4	0	1	8
Site 9	0	1	7	0	0	9
Site 10	0	0	8	0	0	10

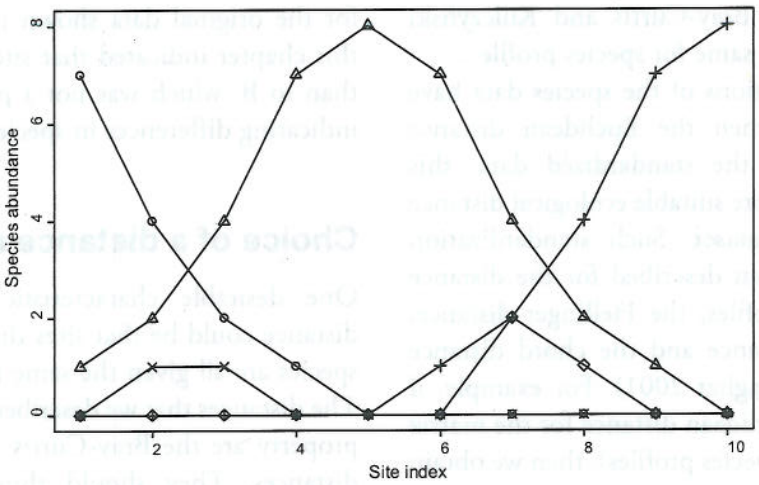

Figure 8.3 Artificial dataset with the abundance of 5 species for 10 sites.

For this artificial dataset, we arranged sites in the sequence of 1-10 (site index) based on ecological similarities between the sites, an ordering that reflects the 'smooth' changes in species abundance from site 1 as displayed in Figure 8.3. These changes are related to unimodal patterns in species abundances that are often observed in surveys. You can see that the abundance of one species (O) decreases compared to site 1, whereas abundance increases and then diminishes for another species (Δ). The particular sequence of the sites that is shown (the site index) is the only sequence that provides such smooth changes. The resulting smooth patterns show that the site index is an intuitively appealing ordering - every ecological distance provides a particular rank order that reflects certain assumptions. When we calculate the ecological distance from site 1 to the other sites, then we obtain the results that are shown in Figure 8.4. The horizontal axis shows

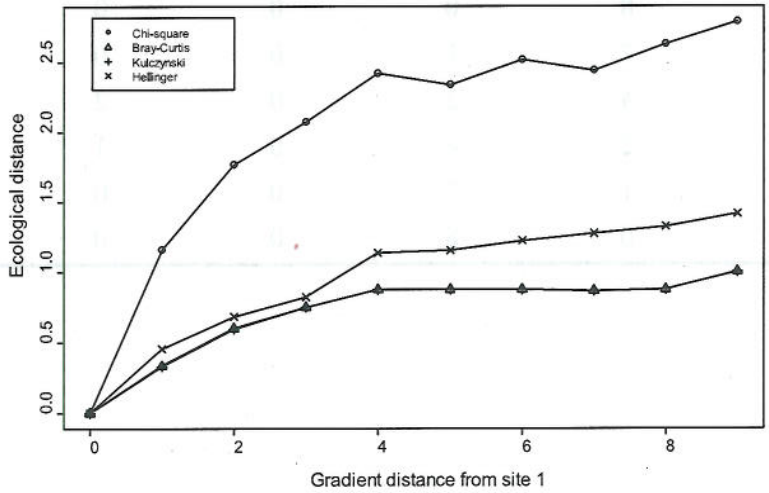

Figure 8.4 Ecological distances from the first site with the other sites of Figure 8.3.

the differences in position on the horizontal axis of Figure 8.3. This distance is the distance on a hypothetical gradient that is best related to changes in species composition.

You can see that the increments are monotonic (never decreasing) for the Bray-Curtis, Kulczynski and Hellinger distances. For the chi-square distance, the pattern fluctuates from the fourth site onwards. Although our impression of distance on the horizontal axis keeps increasing, the ecological distance does not always increase. This is the reason why we do not prefer the chi-square distance since we prefer distances that respect the order in which we arranged the sites. In this example, we prefer the other distances as they show monotonic relationship with the preferred arrangement of sites.

Another way in which the artificial dataset can be analysed is by comparing the ecological distance matrix with a **gradient distance matrix.** The gradient distance matrix is calculated from all the differences between the sites on the x axis (the gradient axis). In this matrix, the difference between site 1 and 5 equals 4, and the difference between site 3 and 5 equals 2. A summary statistic that can be calculated is the **correlation** between the values of the ecological distance matrix and the values of the gradient distance matrix. The significance level for the test that the real correlation could equal zero is calculated with a Mantel test. When you conduct a **Mantel test** for the Bray-Curtis distance for the artificial dataset, then you will obtain the following result:

```
Mantel statistic based on Pearson's product-moment correlation

Mantel statistic r: 0.833
      Significance: < 0.001

Empirical upper confidence limits of r:
  90%   95% 97.5%   99%
0.219 0.281 0.338 0.406

Based on 1000 permutations
```

The results show that correlation is high (r = 0.833) and that the significance level for the test of zero correlation is low ($P < 0.001$). This means that we have evidence for large correlation among the values of the gradient and ecological distance matrices, reflecting what we see in Figure 8.3.

Although analysis of artificial datasets (such as the one provided above) have shown that the Bray-Curtis and Kulczynski distances are good at reflecting the intuitive ordering of sites, these distances suffer from a problem that they are not metric, which can cause some problems in subsequent analyses. Being metric means that the distance between sites A and C is smaller or equal to the sum of the distance between A and B and the distance between B and C. When distances are metric, you could construct a triangle that represents the distances between A, B and C. Being metric is one requirement for the use of some ordination methods. Sometimes a transformation could help – you could for instance calculate the square-root of the Bray-Curtis distance if the Bray-Curtis itself are not metric for your data. Often the square-root of distances will be metric. Another choice that you have is to use a metric distance that also performs well for artificial datasets, such as the Hellinger distance.

Comparing a distance matrix with differences for an environmental variable

The relationship between a distance matrix and a quantitative environmental variable can be analysed with a Mantel test or by a graph as shown above for the artificial dataset. The ecological distance matrix can be based on any distance measure discussed. Then also calculate an **environmental distance matrix**, which defines the distance between pairs of sites based on the environmental variables measured. If these are quantitative the Euclidean distance may be appropriate. As in many other situations, the graph gives much insight into the relationships. The Mantel test serves to ensure that we are not mislead by patterns arising by chance. Remember it is based on correlation, only describing linear relationships between the ecological and environmental distances.

For example, if we were interested in the relationship between species composition as expressed by the Bray-Curtis distance and the depth of the A1 horizon for the dune meadow dataset, then we would obtain the following result for the Mantel test:

```
Mantel statistic based on Pearson's
product-moment correlation

Mantel statistic r: 0.2379
      Significance: 0.045

Empirical upper confidence limits of r:
  90%   95% 97.5%   99%
0.182 0.229 0.267 0.310

Based on 1000 permutations
```

Simultaneously with the Mantel test, we can plot the ecological distance against the environmental distance as in Figure 8.5.

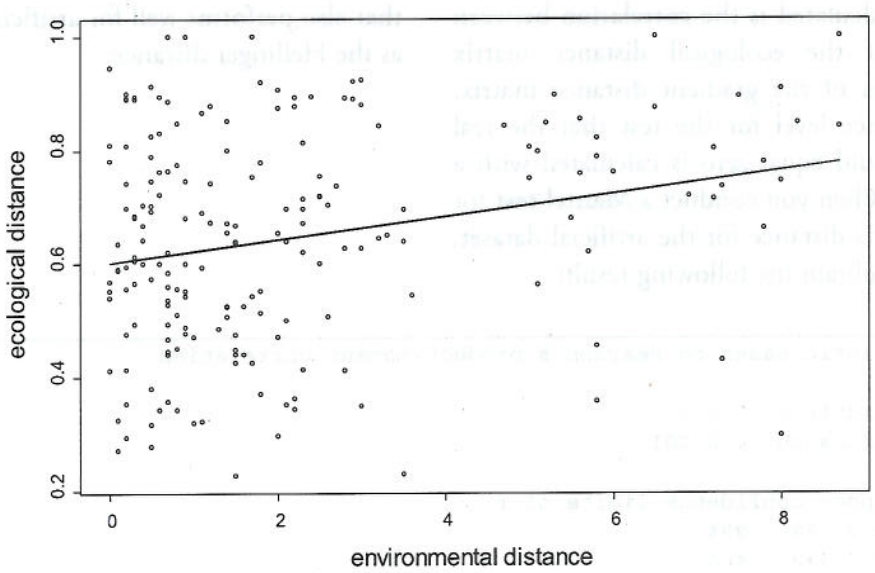

Figure 8.5 Bray-Curtis distances in relationship with differences in depth of the A1 horizon for the dune meadow dataset. The line indicates the fitted relationship between the distances by GAM.

The results show that although the significance level of the correlation is quite small (P =0.045), there is a large scatter of observations and correlation is low. This means that the correspondence between the differences in depth of the A1 horizon and the ecological distance is not very good.

A similar method to the Mantel test is to use the ANOSIM (Analysis of Similarity) test. When sites are classified by a categorical environmental variable, the method examines whether sites within categories are more similar than sites in different categories. A significance level for a test of no difference between categories is calculated. The R statistic that is calculated can be interpreted as a correlation coefficient: values close to 0 indicate little correlation with the groups, and values close to 1 or close to -1 indicate strong correlation. This statistic is unlikely to be smaller than 0 since that would indicate that similarities within categories are systematically lower than similarities among categories.

For example, the ANOSIM for the distance matrix of the dune meadow dataset based on the Kulczynski distance and for management gives the following result:

```
anosim(dis = ecology.distance, grouping
= Management)
Dissimilarity: kulczynski

ANOSIM statistic R: 0.2397
      Significance: 0.016

Based on 1000 permutations
```

Also similar to the procedure for a quantitative variable, we can plot the ecological distance against the environmental distance as in Figure 8.6. To express environmental distance, we calculate a distance of 0 if both sites have the same type of management (for example if both sites are of category hobby farming), and a distance of 1 if both sites have a different type of management (for example, one with hobby farming and one with standard farming). This method of expressing distance for categorical variables is the **Gower distance**. When the environmental variable is a categorical variable, you can also conduct the Mantel test by using the Gower distance to calculate the environmental distance matrix.

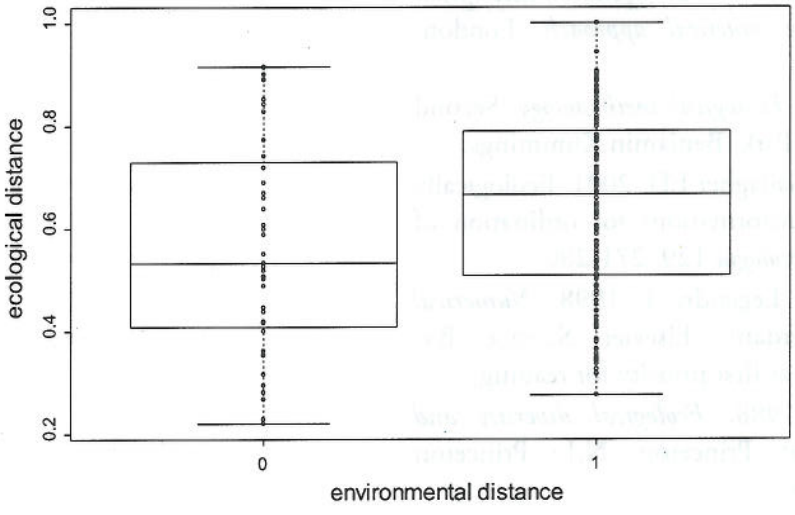

Figure 8.6 Kulczynski distances in relationship with differences in management for the dune meadow dataset.

The ANOSIM test and Figure 8.6 show that there is evidence for a relationship between the ecological distance and the type of management, but that the relationship is not very strong. We can observe several cases that have the same type of management but have a large ecological distance. We can also observe several cases that have a different type of management but have a small ecological distance. This pattern is summarized by the low ANOSIM statistic of 0.24.

In many situations it will be better to use a constrained ordination technique (see Chapter 10) to investigate the influence of environmental variables on species composition since these techniques provide a more comprehensive result.

References

Gotelli NJ and Ellison AM. 2004. *A primer of ecological statistics.* Sunderland: Sinauer Associates.

Jongman RH, ter Braak CJF and Van Tongeren OFR. 1995. *Data analysis in community and landscape ecology.* Cambridge: Cambridge University Press.

Kent M and Coker P. 1992. *Vegetation description and analysis: a practical approach.* London: Belhaven Press.

Krebs CJ. 1994. *Ecological methodology.* Second edition. Menlo Park: Benjamin Kummings.

Legendre P and Gallagher ED. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271-280.

Legendre P and Legendre L 1998. *Numerical ecology.* Amsterdam: Elsevier Science BV. (recommended as first priority for reading)

Magurran AE. 1988. *Ecological diversity and its measurement.* Princeton, N.J.: Princeton University Press.

McGarigal K, Cushman S and Stafford S. 2000. *Multivariate statistics for wildlife and ecology research.* New York: Springer-Verlag.

Pielou E C. 1969. *An introduction to mathematical ecology.* New York: Wiley - Interscience.

Quinn GP and Keough MJ. 2002. *Experimental design and data analysis for biologists.* Cambridge: Cambridge University Press,

Shaw PJA. 2003. *Multivariate statistics for the environmental sciences.* London: Hodder Arnold.

Doing the analyses with the menu options of Biodiversity.R

Select the species and environmental matrices:

Biodiversity > Environmental Matrix > Select environmental matrix

→Select the dune.env dataset

Biodiversity > Community matrix > Select community matrix

→Select the dune dataset

Calculating distance matrices:

Biodiversity > Analysis of ecological distance > Calculate distance matrix…

→Distance: bray

Transformations of the species data:

Biodiversity > Community matrix > Transform community matrix…

→Method: Hellinger

Biodiversity > Analysis of ecological distance > Calculate distance matrix…

→Distance: Euclidean

Calculating the rank-correlation with the mantel test

Biodiversity > Analysis of ecological distance > Compare distance matrices…

→Type of test: mantel

→Community distance: kulczynski

→Environmental distance: euclidean

→Environmental variable: A1

→Correlation: kendall

Calculating the ANOSIM test

Biodiversity > Analysis of ecological distance > Compare distance matrices…

→Type of test: anosim

→Community distance: kulczynski

→Environmental variable: Management

Doing the analyses with the command options of Biodiversity.R

Calculating distance matrices

```
euclidean.distance <- vegdist(dune,method="euclidean")
euclidean.distance
bray.distance <- vegdist(dune,method="bray")
bray.distance
```

Transformations of the species data

```
community.hel <- disttransform(dune, method='Hellinger')
hellinger.distance <- vegdist(community.hel,
  method="euclidean")
```

Calculating the rank-correlation with the mantel test

```
envir.distance <- vegdist(dune.env$A1, method="euclidean")
ecology.distance <- vegdist(dune, method="kul")
mantel(envir.distance, ecology.distance, "kendall")
plot(envir.distance, ecology.distance)
```

Calculating an ANOSIM test

```
ecology.distance <- vegdist(dune, method="kul")
anosim(ecology.distance, dune.env$Management)
```