CHAPTER 10

Analysis of ecological distance by ordination

Analysis of ecological distance by ordination

This chapter presents some alternative methods for analysing ecological distance (differences in species composition among sites) to those introduced in the previous chapter on clustering.

Ordination methods geometrically arrange sites so that distances between them in the graph represent their ecological distances. The results of ordination are typically viewed as 2-dimensional graphs. In these graphs, each site is presented. Sites that are close together in the graph are interpreted as being similar in species composition, whereas sites that are far apart in the graph are interpreted as containing different species. Various numerical summaries support interpretation of the graphs.

In this chapter we describe both unconstrained ordination methods, which only use ecological distance, and constrained methods, which use environmental variables to guide the ordination. Some guidelines are provided at the end on choice of methods to use.

What is ordination?

In an ordination graph, sites are plotted so that distances between them in the graph reflect the ecological differences between them. In Figure 10.1, site A and site B are placed closest together. This reflects the smaller distance between these two sites.

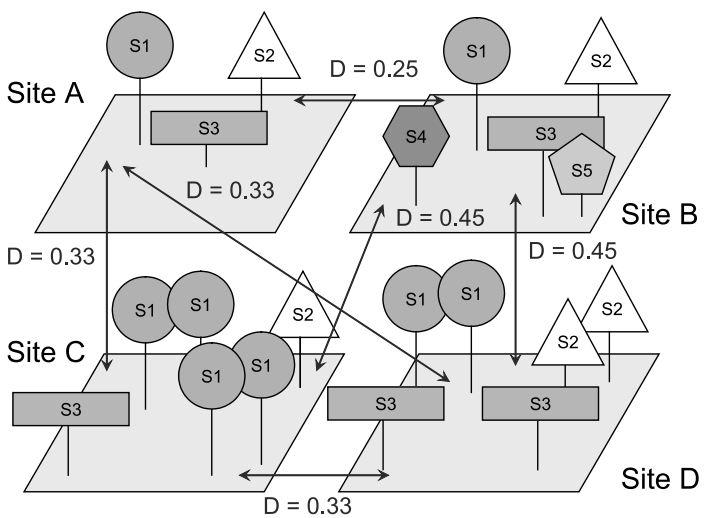

Figure 10.1(a) An ordination analysis will produce a graph that will reflect the ecological distances between sites. The same four sites and ecological distance (Bray-Curtis) were used as in Figure 8.1.

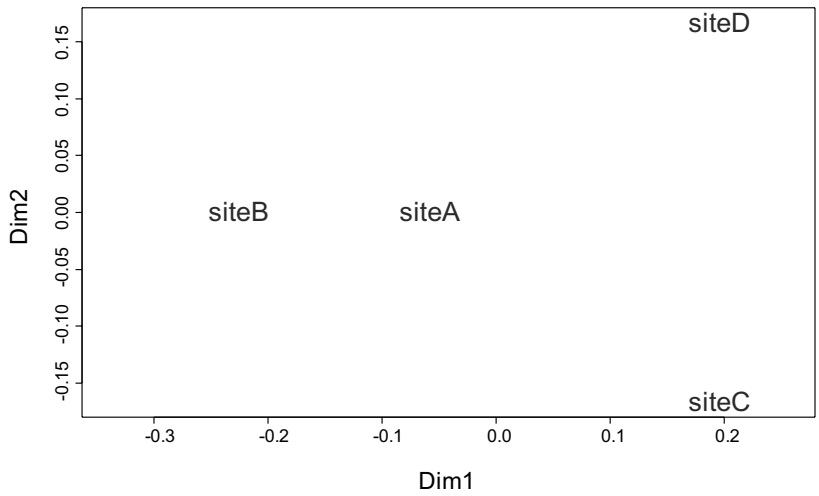

Figure 10.1(b) Sites that are closer together in the graph are more similar in species composition.

Constrained and unconstrained ordination techniques

The ordination techniques can be divided into two groups: unconstrained and constrained ordination techniques. Unconstrained ordination techniques are only based on the species matrix. Constrained ordination techniques use information from both the species and the environmental matrices. The constrained ordination techniques attempt to explain differences in species composition between sites by differences in environmental variables. You can thus examine the relationship between environmental variables and species composition.

Principal components analysis

Principal components analysis (PCA) is one of the oldest ordination techniques. It provides graphs that show the Euclidean distance between sites. No other ecological distances can be investigated with PCA. This ordination method is not ideal for analysis of information on species abundances because of the limitations of the Euclidean distance for describing community differences (see chapter 8 on ecological distance). However, when the original species matrix is suitably transformed, then the output of a PCA can become more meaningful (see below: principal components analysis on transformed species matrices).

If you calculate the PCA for the dune meadow dataset, then you will obtain the following result:

```
Partitioning of variance:

Total         84.12
Unconstrained 84.12

Eigenvalues, and their contribution to the variance

              PC1     PC2    PC3    PC4    PC5    PC6    PC7    PC8   PC9  PC10
lambda    24.7953 18.1466 7.6291 7.1528 5.6950 4.3333 3.1994 2.7819 2.482 1.854
accounted  0.2947  0.5105 0.6012 0.6862 0.7539 0.8054 0.8434 0.8765 0.906 0.928
             PC11   PC12   PC13   PC14   PC15   PC16   PC17   PC18   PC19
lambda    1.7471 1.3136 0.9905 0.6378 0.5508 0.3506 0.1996 0.1488 0.1158
accounted 0.9488 0.9644 0.9762 0.9838 0.9903 0.9945 0.9969 0.9986 1.0000

Scaling 1 for species and site scores
-- Sites are scaled proportional to eigenvalues
-- Species are unscaled: weighted dispersion equal on all dimensions


Species scores

             PC1       PC2       PC3       PC4       PC5        PC6
Achmil -1.112134  0.26681  0.02811  0.54725  1.57082  0.563344
Agrsto  2.530733 -2.07559  0.55423  0.91383 -0.33687  0.208704
…
Trirep -1.071640 -0.04554 -0.55554  0.67400  0.71926  4.092156
Viclat -0.196552  0.24913  0.30824 -0.19065 -0.59315  0.571611


Site scores (weighted sums of species scores)

          PC1       PC2       PC3       PC4       PC5        PC6
X1  -0.465152 -0.08008  0.78537 -0.32939  0.11728 -0.565387
X2  -0.892957 -0.57121  0.26702 -0.28747  0.52939  0.410929
…
X19  0.152525  1.01721 -0.59824 -0.94304 -0.11906 -0.005422
X20  1.270780  0.60335  0.27190  0.20930 -0.01970 -0.219946
```

The interpretation of the various parts of the output is as follows.

The total variance is the total variance of the species between sites. It is the sum of the individual variances of each species (column) of the species matrix. Since PCA is an unconstrained ordination method, the unconstrained variance equals the total variance. When you calculate the variance for separate species, then you will get the following results:

```
Achmil 1.5368421
Agrsto 7.2000000
Airpra 0.6184211
Alogen 6.9052632
Antodo 2.8921053
Belper 1.0815789
Brarut 3.6289474
Brohor 1.9868421
Calcus 1.5263158
Chealb 0.0500000
Cirarv 0.2000000
Elepal 5.5657895
Elyrep 4.3263158
Empnig 0.2000000
Hyprad 1.5236842
Junart 2.6210526
Junbuf 1.9236842
Leoaut 2.4315789
Lolper 7.9894737
Plalan 3.8000000
Poapra 3.4105263
Poatri 7.9236842
Potpal 0.3789474
Ranfla 1.3789474
Rumace 3.2526316
Sagpro 2.4210526
Salrep 1.9447368
Tripra 1.5236842
Trirep 3.6078947
Viclat 0.2736842
```

You can check that the sum of these variances equals the total variance in the PCA output.

Next, the eigenvalues are given. PCA is a technique that will create a new matrix from the species matrix. This new matrix has the same total variance as the original species matrix. The rows of this new matrix still correspond to the sites of the species matrix. The columns are new: these are the principal components. The eigenvalues show how much variance is found in each of the principal components.

We could have analysed differences in species composition by using each species as a separate axis in a graph. For example, Figure 10.2 shows a graph for the dune meadow dataset with axes for the two first species, using the values of the species matrix as coordinates to plot each site (see also Figure 8.2). You can notice for instance that site X4 has a value of 8 for *Agrostis stolonifera* and a value of 5 for *Alopecurus geniculatus*. Other sites are plotted in different positions, with some sites having the same position at the origin as these sites do not have either of the two species. This is an ordination graph, since sites that are close together are more similar in their composition of these two species. However, we only see information for two species and not for the remaining 28 species. Unfortunately, we can not produce graphs to show 30 axes at the same time. We are limited to graphs of 2 dimensions (3-dimensional graphs can be produced, but they are often difficult to read).

You could see PCA as a technique that creates new axes (or a matrix with new columns). The distances between the sites will remain the same in the new matrix (and thus the total variance remains the same). The advantage of the creation of new principal component axes is that more variance will be shown for the first two new axes, than if we plotted two original species axis. We can thus see a larger fraction of the total distance between sites.

The eigenvalues are listed in decreasing size. The first eigenvalue accounts for nearly 30% of the total variance. The first two eigenvalues account for more than 51% of all variance (the accounted cumulative variance of 0.5105 is provided below the second eigenvalue), and the first three eigenvalues account for more than 60% of variance. Thus, if we plot the positions on the first two new column variables (similar to the creation of Figure 10.3), we will see 51% of the variance in distances between sites. The maximum percentage of variance that could be shown by choosing original species is only 16% (Lolper + Poatri). Thus, the advantage of PCA is that axes are created that allow more variance to be shown in an ordination graph with a few axes. It will rarely be possible to show 100% of the variance in two or three axes. If too little of the variance can be shown, it is not very useful to provide an ordination graph.

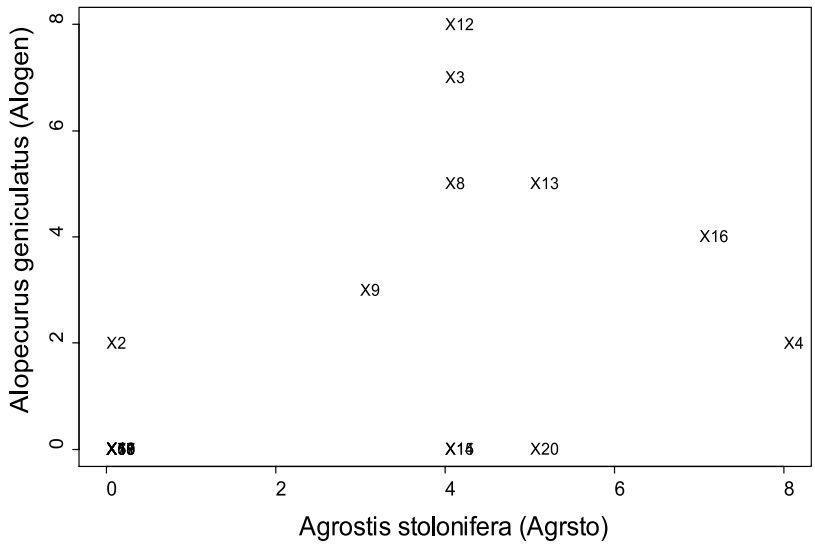

Figure 10.2 An ordination for the dune meadow dataset by using two species.

Next in the output, we get information on scaling of species and site scores. In PCA, different types of scaling of sites and species can be made. In graphs with scaling method 1, the distances between sites on the graph are approximations of the Euclidean distance between sites. For this reason, a synonym for scaling method 1 is distance scaling. If you use scaling method 2, then the distances between sites are not represented as accurately as with scaling method 1. Scaling method 2 is used to show the correlations among species better (see Figure 10.5).

The species scores provided next are coordinates that allow the plotting of species on a graph. The site scores provided next are the plotting coordinates for the sites. You can see that both scores are listed for columns named PC1 through PC6. There are actually 19 PC scores, one for each eigenvalue – the same as the number of species in the original matrix, unless there is something unusual in the data. The output was limited to 6 columns to save some space.

Plotting the site scores gives an ordination graph. Adding the species scores helps in its interpretation, as explained below (Figure 10.4).

Figure 10.3 shows the ordination graph for axes PC1 and PC2. You can interpret this ordination graph as follows. Since scaling method 1 was used, the distances between the sites reflect the Euclidean distance among them. Sites that are closer together have a small Euclidean distance in species composition – remembering, however, that the 2 axes do not show all the variance in distance. If most of the variance is captured in 2 dimensions, the correspondence will be very good, and anyway it is 'as good as it can be', as the 2 dimensions show as much of the variance as possible.

Various methods exist of investigating the goodness-of-fit of the first two axis of the PCA ordination. One method is to calculate the proportion of total variance that is explained for each site in the ordination graph, as shown in the following results:

	PC1	PC2
X1	0.13	0.14
X2	0.38	0.54
X3	0.03	0.69
X4	0.00	0.42
X5	0.40	0.41
X6	0.41	0.50
X7	0.65	0.67
X8	0.20	0.43
X9	0.00	0.38
X10	0.55	0.58
X11	0.20	0.39
X12	0.13	0.34
X13	0.07	0.55
X14	0.31	0.41
X15	0.62	0.78
X16	0.77	0.78
X17	0.00	0.39
X18	0.01	0.44
X19	0.01	0.40
X20	0.61	0.75

These results show that sites X15 and X16 are the best represented in Figure 10.3 (78% of variance), whereas site X1 is the least well represented (14%).

It is distances on the ordination graphs that are important. These are not changed if the signs (+ or -) of the scores are changed. Some software will yield scores that are the negative of the scores that you obtain with your software. The interpretation of the ordination graphs remains the same. This is true for any type of ordination graph.

The species scores show the direction from the origin (the point with coordinates (0,0) shown in the middle of Figure 10.3) where sites occur that have a larger than average value for the particular species. For example, sites X2 and X3 are expected to have larger than average values for *Poa trivialis* since this species and the two sites occur in the same direction (lower-left) from the center.

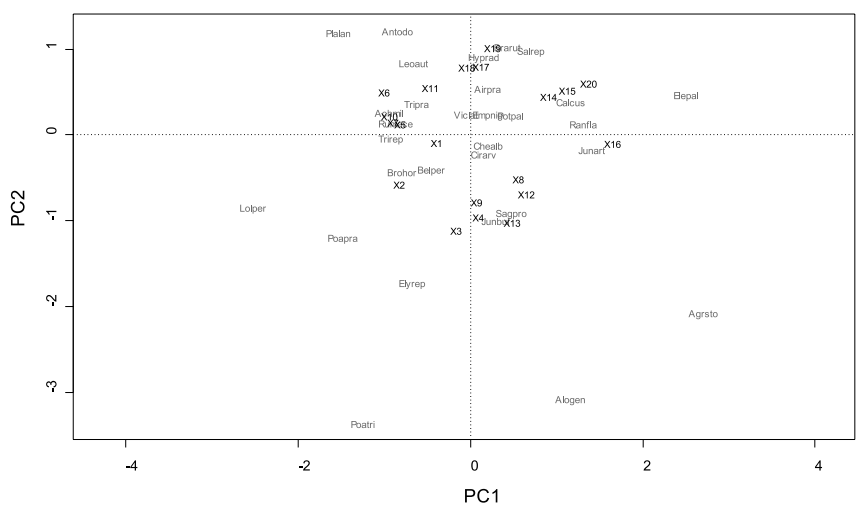

Figure 10.3 PCA ordination graph for the first two axes for the dune meadow dataset using scaling method 1.

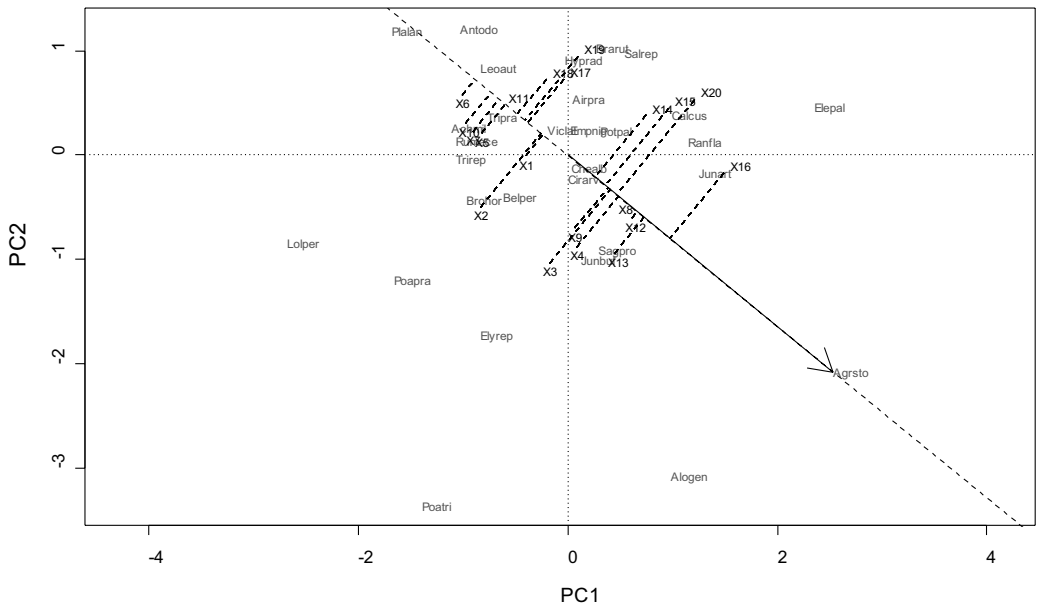

Figure 10.4 Interpretation of species scores for an ordination graph. The vector is drawn for species *Agrostis stolonifera*. The projections for the sites indicate a ranking of sites from low (lowest X6) to high (highest X16) abundance for the species.

Figure 10.4 shows a more formal method of investigating the species scores. The arrow is drawn for species *Agrostis stolonifera*. This arrow shows the direction from the origin for which sites have larger abundances for this species. By constructing perpendicular lines for each site, showing their projection onto this arrow, we get an indication of differences in abundance between sites. Sites X13 and X16 are projected farthest from the origin in the direction of the species vector. We could expect that these sites have larger abundances for the particular species than other sites. Sites X6 and X10 are projected at the opposite side of the species vector. We can expect lower than average abundances for these sites. When we check the original species matrix (chapter 2), we can see that this interpretation is a good approximation of the actual situation, with X3 and X16 having large abundance and X6 and X10 having low abundance. However, the original species matrix shows that site X4 has the largest abundance for *Agrostis stolonifera* whereas it does not have the largest position on the species vector. Since we do not see all variance in a graph, the positions will not completely reflect the exact rankings. The value of the ordination lies more in providing a good summary of distances among sites and their overall relationship with species. If you are interested in the information about a particular species, such as *Agrostis stolonifera*, then you would be better looking at that species alone, perhaps with regression analysis (chapter 6).

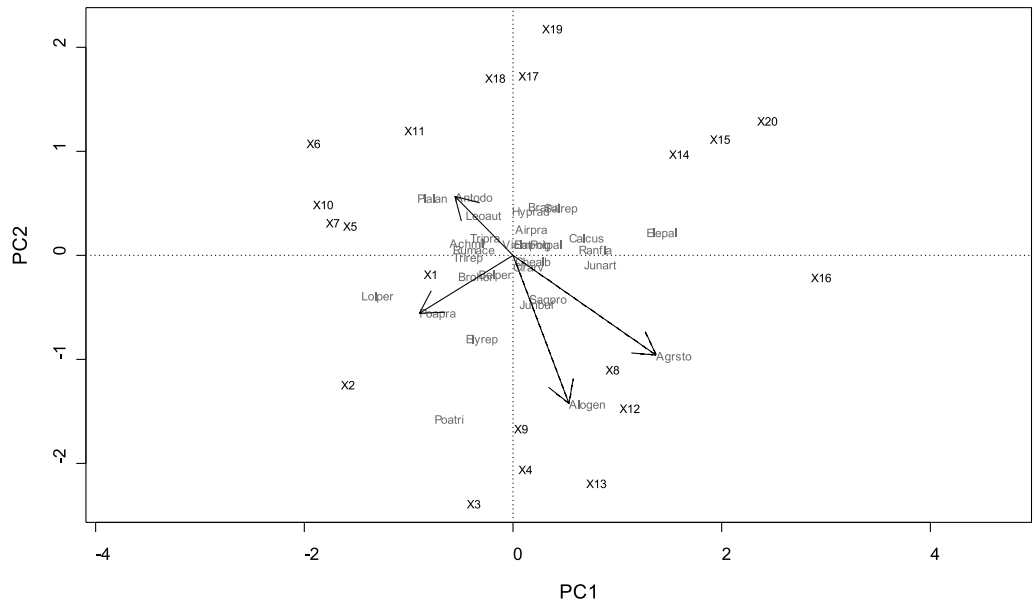

Figure 10.5 Interpretation of species scores for an ordination graph with scaling method 2. The angles between the species vectors indicate correlation among species.

As we saw earlier, scaling method 2 is used to reflect the correlations among species better. Species that have a small angle between their vectors are expected to be strongly positively correlated. Species with angles between vectors at 90 or 270 degrees are expected to not to be correlated and species with angles of 180 degrees are expected to be strongly negatively correlated. We can thus see from Figure 10.5 that *Agrostis stolonifera* is positively correlated with *Alopecurus genitalicus*, not correlated with *Poa pratensis*, and negatively correlated with *Anthoxanthum odoratum*. Correlation between species refers to similarity in abundances for the sites of two species. The respective correlations are 0.54, 0.07 and -0.45.

Note that different software packages will provide different coordinates (scores) for species and sites, even if the same scoring type was used. These differences are caused by different approaches to multiplying species and site scores in attempts to provide better graphs of sites and species. The interpretation of the graphs will remain the same, despite the differences in scales.

Some research has been done on how many principal component axes should be analysed to get a good ecological picture of the total variance of a dataset. One of the better methods is based on the broken-stick distribution. The name comes from the idea that if you chopped up the total variance into pieces randomly, in much the same way you could randomly break a stick into pieces, some of the parts would be larger than others. We are interested in parts of the variance which reflect real structure, not just random 'breaking'. So a test for the importance of principal components can be based on comparison of what you would get with random breaking and what you actually see. If you use the test based on the broken-stick distribution for the dune meadow dataset, then you obtain the following result:

	1	2	3	4
eigenvalue	24.79532	18.14662	7.629135	7.152772
percentage of variance	29.47484	21.57136	9.068950	8.502685
cumulative percentage of variance	29.47484	51.04620	60.115145	68.617831
broken-stick percentage	18.67231	13.40916	10.777577	9.023191
broken-stick cumulative %	18.67231	32.08147	42.859047	51.882238
% > bs%	1.00000	1.00000	0.000000	0.000000
cum% > bs cum%	1.00000	1.00000	1.000000	1.000000

Two criteria are available to select the number of significant axes. The first criterion is to select axes for which the percentage of variance is larger than the corresponding percentage of variance of the broken-stick distribution. For the PCA results reported above, only two axes are significant: axis 1 with 29.5% (>18.7% of the broken-stick distribution) and axis 2 with 21.6% (>13.4%). Another criterion is to select the axes for which the cumulative percentage of variance is larger than the corresponding cumulative percentage of variance of the broken-stick distribution. For the dune meadow dataset, this criterion would result in all axes to be significant.

If you used scaling method 1 (distance scaling), then there is a technique to select those species that significantly contributed to the axes shown in an ordination graph. This technique is based on an equilibrium circle. Species that significantly contribute to the ordination will have vectors outside of the equilibrium circle (Legendre and Legendre 1998). Figure 10.6 shows the vectors for the significant species for axis 1 and 2 of the PCA of the dune meadow dataset.

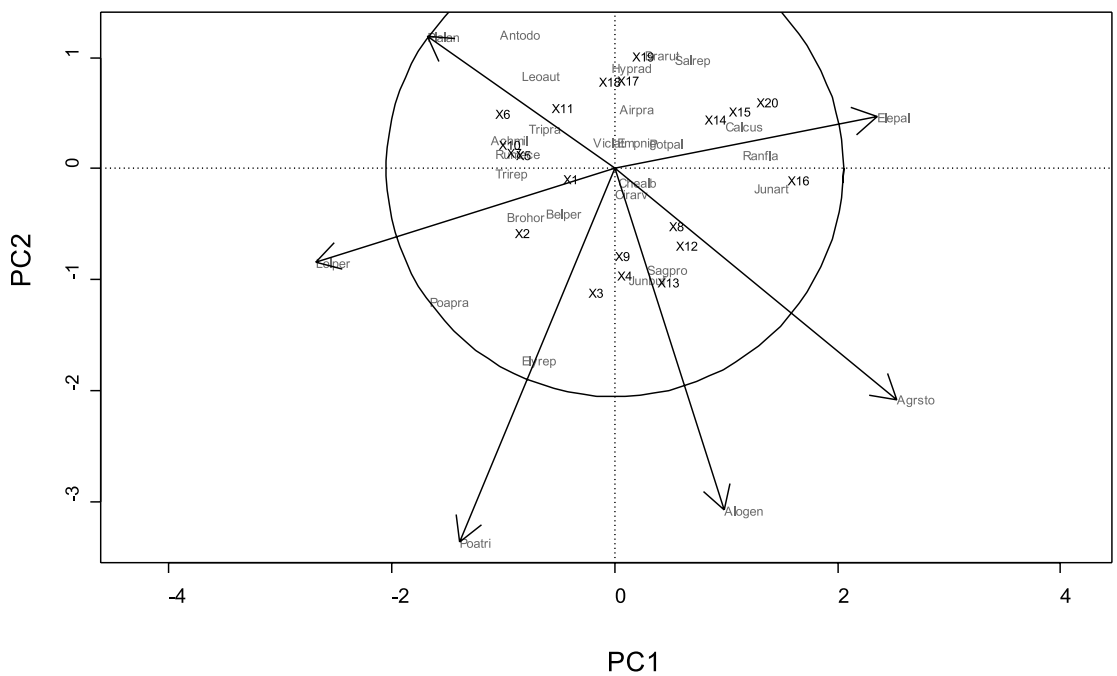

Figure 10.6 Equilibrium circle for the first two PCA axes for the dune meadow dataset. Vectors are drawn for the species that significantly contributed to the ordination graph.

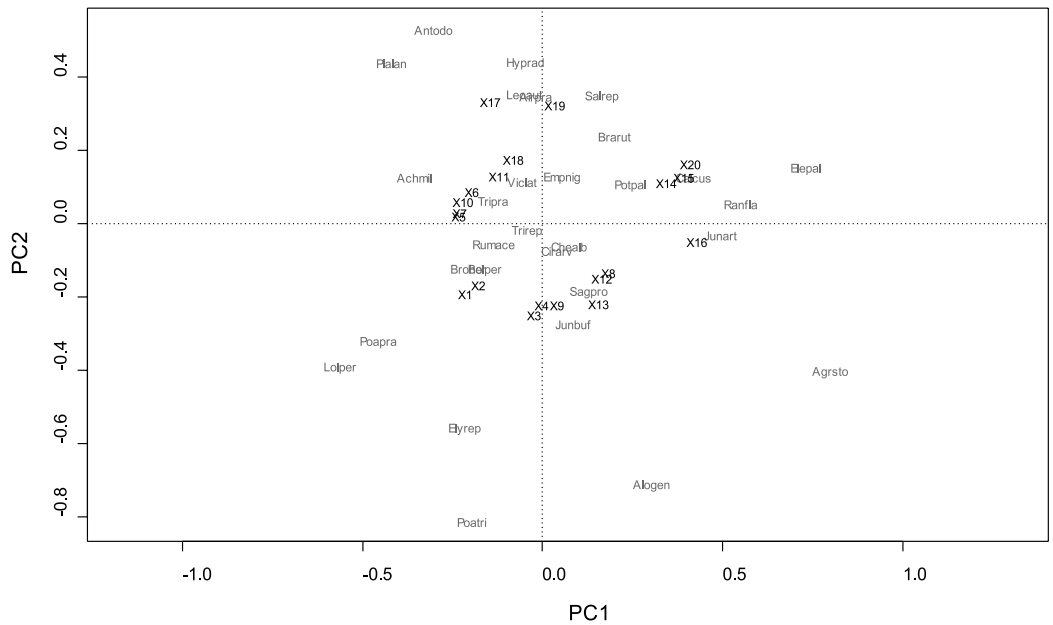

Figure 10.7 The first two PCA axes for the dune meadow dataset, after the species matrix was transformed so that distances among sites reflect their Hellinger distance.

Principal components analysis on transformed species matrices

A principal components analysis (PCA) will produce ordination graphs that portray the Euclidean distance among sites. This ordination technique is therefore limited to investigating the Euclidean distance among sites. There is one method of using PCA to investigate some other ecological distances among sites, however. This method is based on transforming the species matrix first, followed by PCA on the transformed matrix. By proper transformations, investigations are thus possible of the chi-square distance, Hellinger distance and distances between species profiles, which are better ecological distances than the Euclidean distance (see Chapter 8; Legendre and Gallagher 2001).

Figure 10.7 gives an example of the dune meadow dataset after a transformation that leaves distances representing the Hellinger distance.

Principal coordinates analysis

Principal coordinates analysis (PCoA) is an ordination technique that is similar to PCA. The technique has the advantage over PCA that any ecological distance can be investigated. In fact, when you calculate a PCoA with the Euclidean distance, then you will obtain the same result as with a PCA. A synonym for principal coordinates analysis is metric multidimensional scaling. The idea of PCoA is to start with a distance matrix. Then try to find an arrangement of sites such that that the distances between the sites in the graph match as closely as possible those in the distance matrix.

A PCoA for the dune meadow dataset based on the Bray-Curtis distance, gives the following result:

```
$points
             Dim1         Dim2
X1  -0.237339546 -0.17173114
X2  -0.197122801 -0.12450970
…
X19 -0.003609141  0.31387032
X20  0.340688509  0.10539824

$eig
[1]  9.032980e-02  5.381042e-02

$ac
[1] 0

$GOF
[1] 0.5714973 0.5961463

$eigen.total
[1] 4.990226

$R.constant
[1] 1.494617

$Rscale
[1] TRUE

$scaling
[1] 1

$cproj
               Dim1          Dim2
Achmil  -4.01695503   0.88822361
Agrsto   9.92187299  -7.98344244
…
Trirep  -2.19199995  -0.53783022
Viclat  -0.73957850   0.84489420
```

The results are similar to those for a PCA. First, the scores of the sites are given. These scores correspond to the different axes calculated by the PCoA, analogous to the PCA. Figure 10.8 shows a graph of the first two axes of the PCoA. You can see, for example, that site X1 is expected to have a small Bray-Curtis distance from site X2, and a large distance from site X20.

Next in the results, the eigenvalues are given. The eigenvalues can be interpreted in the same way as the eigenvalues of PCA: they express how much variance is shown on each axis.

Lower in the output, some parameters are provided that can be used to interpret the goodness-of-fit of the PCoA – that is, the extent to which the ordination graph reflects the distance matrix accurately. The GOF provides two ways of assessing the fit with the extracted axes. The sum of the eigenvalues for the 2 extracted axes in our results may be expressed as a ratio over the

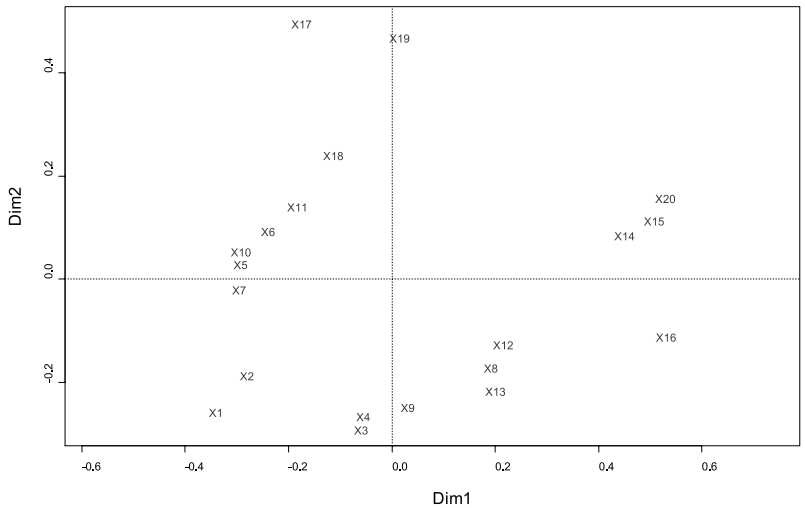

Figure 10.8 Ordination graph for the first two axes of a PCoA of the dune meadow dataset based on the Bray-Curtis distance.

total of absolute values of all eigenvalues, or over the total of all positive eigenvalues. This method of calculating the GOF may seem strange in comparison to PCA, but results from the fact that some eigenvalues may be negative in PCoA.

If all 19 axes are calculated for the PCoA, then you obtain the eigenvalues that are shown below.

You can see that eigenvalues are provided in a sequence of largest to the smallest value. This means, as for PCA, that most variance will be shown by making an ordination graph with the first two axes. When you take a closer look, you will see that the last 5 eigenvalues are negative.

The GOF gives the percentage of the first two axes when the absolute value is taken for the negative eigenvalues or when the negative eigenvalues are not considered. You can see that calculated by these methods, the first two axes show 57% and 60% of variance.

Negative eigenvalues are a result of the fact that it is impossible to calculate site scores so that the distance among the sites will equal those in the distance matrix. This will occur when distances are not metric (see chapter 8). When the distances are not metric, some eigenvalues will be negative. It is not possible to calculate site scores for negative eigenvalues. It is therefore not possible to obtain an ordination of sites that will exactly reproduce the distances of the distance matrix, even if all 19 axes are used.

The Bray-Curtis and Kulczynski distances are known to produce negative eigenvalues for some datasets. One solution (called the Caillez correction method) that will not produce negative eigenvalues is to add some constant to the elements of the distance matrix. This solution will now allow calculation of site scores for each eigenvalue. However, the distance matrix is modified and total variance was artificially increased in a rather arbitrary way. If your main interest is investigating

```
$eig
 [1]  9.032980e-02  5.381042e-02  2.428758e-02  2.011838e-02  1.469129e-02
 [6]  1.245426e-02  8.901072e-03  5.065536e-03  3.920619e-03  3.248000e-03
[11]  2.891603e-03  1.009173e-03  8.483670e-04  2.105743e-04 -5.112869e-18
[16] -1.390796e-03 -2.255631e-03 -2.880746e-03 -3.901214e-03
```

ecological distance, then you may opt not to correct for negative eigenvalues and only to analyse the first axes. Another solution would be to analyse the square-roots of the distance (the square root of the Bray-Curtis distance is expected to be metric).

A standard PCoA does not calculate scores for species. There is one way of calculating the same species scores as for PCA when the Euclidean distance is investigated. This approach is based on the correlations between species and axes, and weighting these scores by the percentage of variance expressed on each axis. The site and species scores provided in the PCoA output will be exactly the same as for PCA when you investigate the Euclidean distance. These species scores are labelled as `cproj` in the output. There is no fixed technique to calculate species scores for PCoA, however. Another technique is to calculate weighted average scores (see below: non-metric multidimensional scaling).

The species scores can be plotted in an ordination graph. Figure 10.9 shows the graph for the dune meadow dataset. The species scores should be interpreted similarly as the species scores of a PCA (see also Figure 10.4): sites that are plotted at the same direction from the origin as species are expected to have higher abundance of the species. We expect therefore that site X1 has higher abundance of *Lolium perenne* and *Poa pratensis*.

Non-metric multidimensional scaling

Non-metric multidimensional scaling (NMS or NMDS) is an ordination technique that is related to PCoA. The calculations are also done on a distance matrix. However, the positions of sites in the ordination are chosen so that rank order only of intersite distances is represented. If sites 1 and 2 are the closest in the distance matrix, they will be in the ordination graph. If site 1 is further

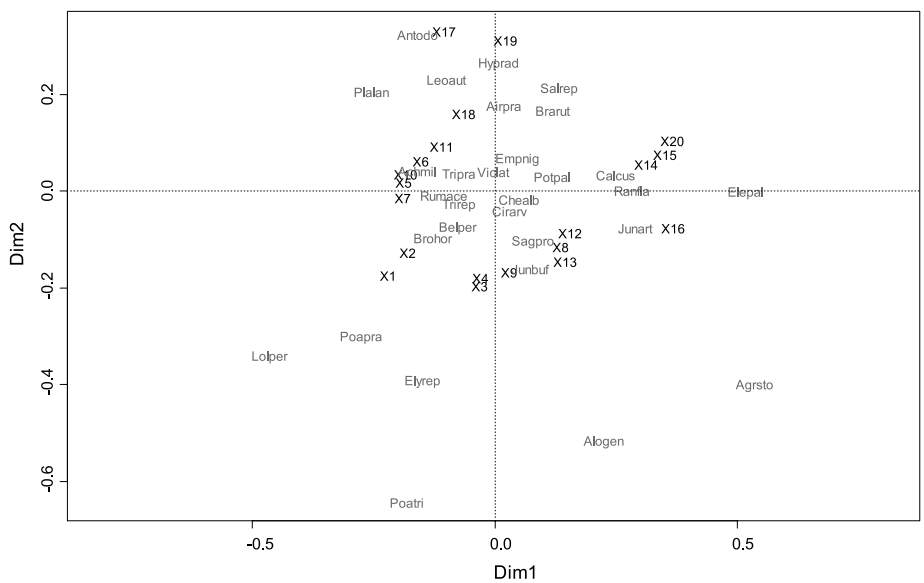

Figure 10.9 Ordination graph for the first two axes of a PCoA of the dune meadow dataset based on the Bray-Curtis distance. The species coordinates indicate the correlation of site scores with each species.

from 3 than 2, then this will also be true in the graph. However, the graph could look the same for distance matrices A and B:

Distance matrix A

	Site 1	Site 2	Site 3
Site 1	0.0	0.1	0.2
Site 2	0.1	0.0	0.3
Site 3	0.2	0.3	0.0

Distance matrix B

	Site 1	Site 2	Site 3
Site 1	0.0	0.1	0.8
Site 2	0.1	0.0	0.9
Site 3	0.8	0.9	0.0

Unless you have no faith in the quantitative nature of the distances calculated in your distance matrix, this does not seem a desirable feature.

Another difference between PCoA and NMS is that the final result is obtained through some random process. When you repeat the analysis, then you may obtain different results. You can repeat the analysis several times and then represent the best result. The best result is the one that reflects the rank-order of distances in the original distance matrix the best. The statistic that reflects how well the configuration represents the distances is called 'stress'. A smaller stress means that a better NMS ordination was calculated. Final stress values should ideally be smaller than 10% and not larger than 30% to represent species abundance data accurately.

When you calculate a NMS ordination for the dune meadow dataset based on the Bray-Curtis distance, and show the best ordination out of 100 for two axes, then you will obtain a similar result to:

```
$points
           [,1]       [,2]
X1    3.3829673 -4.7738116
X2    2.0565435 -2.6775501
X3    2.6890856 -0.4458917
X4    3.1468039 -0.7250620
X5    0.1917841 -2.9803663
X6   -0.7784914 -2.2831423
X7    0.1469503 -2.4612942
X8    1.6818537  1.8604024
X9    2.9114147  0.5966568
X10 -0.2924911 -2.2601592
X11 -1.7602236 -0.9072029
X12  2.9598205  2.4678441
X13  4.0871950  2.0987138
X14 -1.0003700  6.0644557
X15  0.3415692  5.6201219
X16  2.6892989  6.1292587
X17 -5.8442737 -2.9432844
X18 -2.5547698  0.1139317
X19 -5.1139348  1.3438575
X20  0.2423127  6.4558772

$stress
[1] 11.91173
```

You can see that this output provides the scores for the sites, and an indication of the stress. Figure 10.10 provides the ordination graph with the calculated scores.

The NMS does not provide species scores. As in PCoA, you could add the correlation or weighted average scores to the ordination graph. The interpretation of weighted average scores is as follows. The sites that are closest to a species are expected to have the highest abundance, whereas sites that are farther away are expected to have lower abundance. Figure 10.13 gives a more explicit example for the interpretation of weighted average scores. Note this is not the same as the interpretation in PCA, illustrated in Figure 10.4.

Figure 10.11 shows a graph where the weighted average scores of species were added for the NMS ordination of Figure 10.10. The species scores suggest that the abundance of *Vicia lathyroides* is highest for site X11, for example.

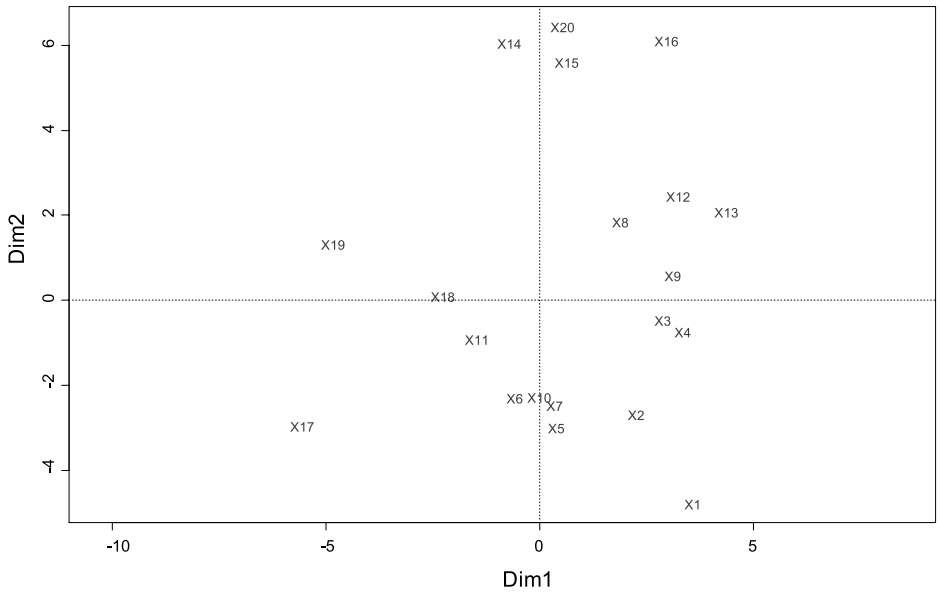

Figure 10.10 Ordination graph for a two-dimensional NMS of the dune meadow dataset based on the Bray-Curtis distance. The best configuration out of 100 is shown.

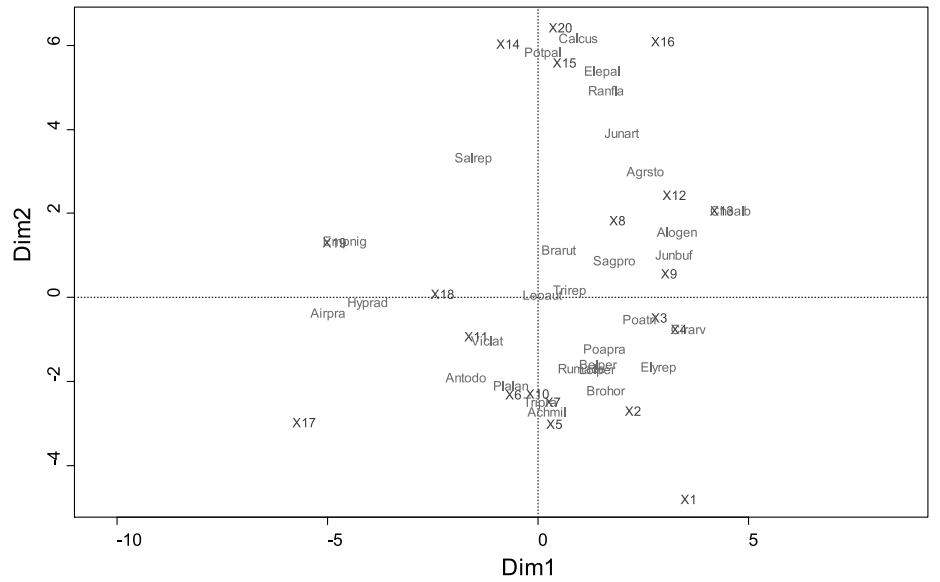

Figure 10.11 Ordination graph for a two-dimensional NMS of the dune meadow dataset based on the Bray-Curtis distance, with weighted-average scores for species.

Correspondence analysis

Correspondence analysis (CA) or reciprocal averaging (RA) is an ordination technique that shows the chi-square distance among sites. This technique was developed independently by several authors that gave it different names. One of the algorithms calculates site and species scores by consecutive steps of calculating site scores as the weighted average of the species scores, followed by calculating species scores as the weighted average of the site scores, until the results converge. This technique has an advantage over PCA because the chi-square distance is a better ecological distance than the Euclidean distance that is shown in PCA. The PCA on transformed species matrices is an alternative method that allows using better ecological distances than the chi-square distance (see above: Principal component analysis on transformed species matrices).

A correspondence analysis of the dune meadow dataset gives the result shown on the next page.

The interpretation of the different parts of the output is quite similar to the output of the PCA provided above. The interpretation is as follows.

The partitioning of the mean squared contingency coefficient reflects how much of the variance in chi-square distances was calculated. As CA is an unconstrained ordination technique, all variance is calculated.

Next the eigenvalues are given. As for PCA and PCoA, the eigenvalues are listed from highest to lowest, and their sum equals the total variance. Again as for PCA and PCoA, by plotting the first axes, more variance will be shown. When plotting the first two axes, 44% of variance will be shown (the accounted values indicate the cumulated proportion of variance).

The scaling method 1 is the same as for PCA. This scaling method means that the distances between sites reflect the chi-square distance. Scaling method 2 means that the distances between the species reflect the chi-square distance among species.

The species and site scores are the coordinates that are used in ordination graphs of the CA. Figure 10.12 shows the ordination graph for the first two axes of the CA for the dune meadow dataset.

```
Partitioning of mean squared contingency coefficient:

Total         2.115
Unconstrained 2.115

Eigenvalues, and their contribution to the mean squared contingency coefficient

             CA1    CA2    CA3    CA4    CA5    CA6     CA7     CA8     CA9   CA10
lambda    0.5360 0.4001 0.2598 0.1760 0.1448 0.1079 0.09247 0.08091 0.07332 0.0563
accounted 0.2534 0.4426 0.5654 0.6486 0.7170 0.7680 0.81175 0.85000 0.88467 0.9113
             CA11    CA12    CA13    CA14    CA15     CA16     CA17     CA18
lambda    0.04826 0.04125 0.03523 0.02053 0.01491 0.009074 0.007938 0.007002
accounted 0.93410 0.95360 0.97025 0.97995 0.98700 0.991293 0.995046 0.998356
             CA19
lambda    0.003477
accounted 1.000000

Scaling 1 for species and site scores
-- Sites are scaled proportional to eigenvalues
-- Species are unscaled: weighted dispersion equal on all dimensions


Species scores

             CA1      CA2      CA3       CA4      CA5      CA6
Achmil -1.241039  0.13375 -1.15041 -0.021261 -1.73513 -0.57465
Agrsto  1.275444 -0.32647  0.55258  0.057909 -0.36618 -0.06869
…
Trirep -0.104710 -0.03213 -0.40404  0.063081 -0.49082  1.64252
Viclat -0.845393  0.58712 -0.90362 -2.384693  3.05575  4.41315


Site scores (weighted averages of species scores)

         CA1        CA2      CA3      CA4      CA5      CA6
X1  -0.59425 -0.6848644 -0.07380 -0.88373 -0.14948 -0.60267
X2  -0.46320 -0.4401641 -0.04948 -0.49792 -0.37167  0.02160
…
X19 -0.50536  2.0648336  0.99756 -0.07422 -0.02797 -0.05283
X20  1.42353  0.6761410 -0.33944 -0.23205  0.60727 -0.55941
```

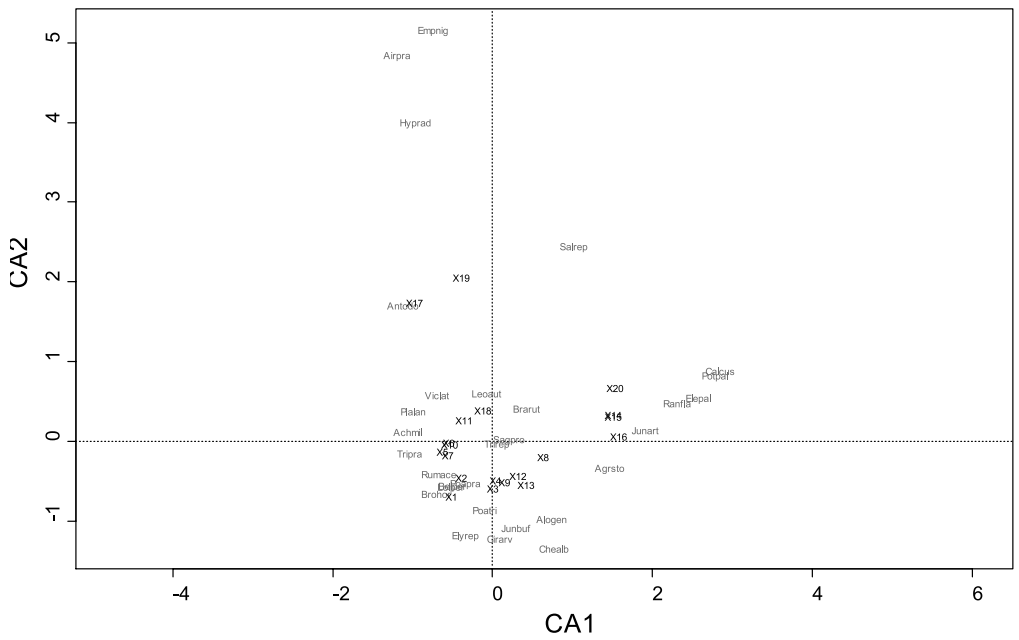

Figure 10.12 Ordination graph for the first two axes of correspondence analysis of the dune meadow dataset, using scaling method 1.

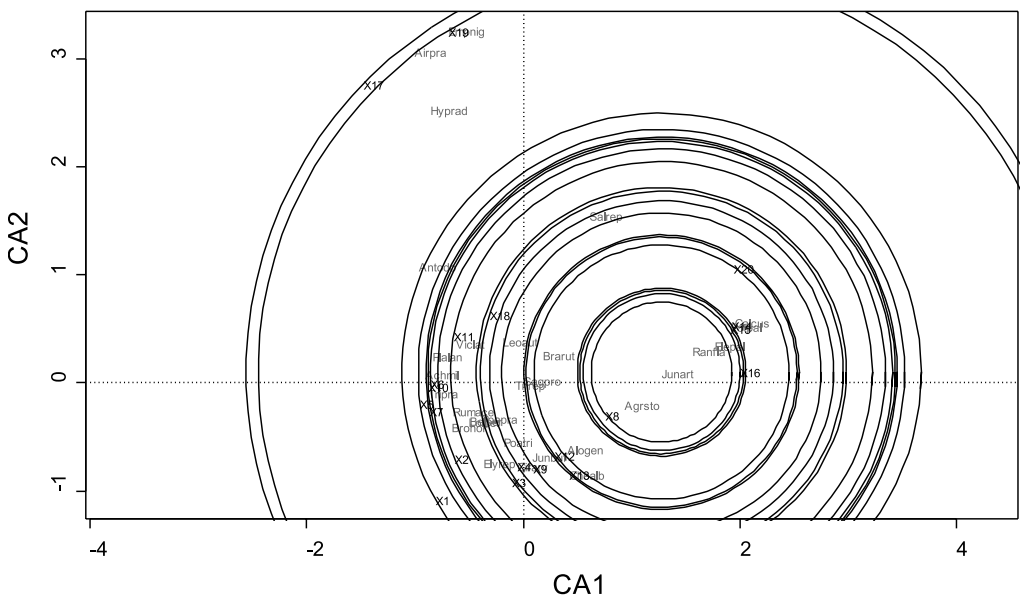

Figure 10.13 Interpretation of species scores for a correspondence analysis. Larger circles indicate a smaller expectation for the abundance of *Juncus articulatus* for the correspondence analysis of the dune meadow dataset (scaling method 2).

Species scores should be interpreted as weighted average scores. Sites that have positions in the graph close to the species are expected to have high abundance for the species, whereas sites that are farther are expected to have low abundance. Figure 10.13 shows how the position of *Juncus articulatus* should be interpreted. A smaller circle – or smaller distance between the species and site – means that we expect higher abundance for a site. We thus expect high abundance for sites X8 and X16, and low abundance for site X17. Because we do not show all variance in a two dimensional graph, the actual abundance may differ. Species that are drawn at intermediate distance from the origin are often best portrayed, not species at the edges or close to the origin.

The site scores can be thought of as a measure of 'suitability' of the site. CA assumes a unimodal distribution of species on this suitability gradient – either too high or too low results in low abundance of a species. PCA, on the other hand, assumes a linear relationship with 'suitability'. This difference in the underlying assumptions may be another advantage of CA over PCA, because the unimodal distribution is more common in nature than the linear distribution.

Simulation studies have shown that the ordination provided by a CA will often not reconstruct the known structure of the data. Based on this information, some modifications were made in CA resulting in the ordination technique of detrended correspondence analysis (DCA). However, some other simulations have shown that DCA is often not successful either in reconstructing the known structure of data. Although DCA has often been used in the past, we prefer to use other methods for ordination.

Redundancy analysis

Redundancy analyis (RDA) is the first of the constrained ordination techniques that will be discussed in this chapter. RDA is an ordination technique that is related to PCA. Both techniques show the Euclidean distance between sites in the ordination graphs. The difference is that in RDA the ordination is constrained by the environmental variables, shown in the environmental matrix. The new axes for displaying the species matrix are constrained to be linear combinations of the columns of the environment matrix, similar to the fitted values you would get by regressing PCA scores on the environmental variables.

A RDA of the dune meadow dataset, using the type of management as a constraining variables gives the following result:

```
Partitioning of variance:

Total         84.12
Constrained   29.23
Unconstrained 54.89

Eigenvalues, and their contribution to the variance

             RDA1    RDA2   RDA3     PC1    PC2    PC3    PC4    PC5    PC6    PC7
lambda    14.8654 10.6904 3.6750 15.2700 8.4275 6.8989 5.6749 3.9884 3.1212 2.5875
accounted  0.1767  0.3038 0.3475  0.1815 0.2817 0.3637 0.4312 0.4786 0.5157 0.5464
             PC8    PC9   PC10   PC11   PC12   PC13   PC14   PC15   PC16
lambda    2.3802 1.8181 1.3762 0.9951 0.7853 0.6610 0.4666 0.2827 0.1594
accounted 0.5747 0.5963 0.6127 0.6245 0.6339 0.6417 0.6473 0.6506 0.6525
```

```
Scaling 1 for species and site scores
-- Sites are scaled proportional to eigenvalues
-- Species are unscaled: weighted dispersion equal on all dimensions

Species scores

           RDA1     RDA2     RDA3      PC1      PC2      PC3
Achmil  0.71492  1.19801  0.61663 -0.79733  0.03573  0.02448
Agrsto -0.29018 -3.22454  0.08531  2.85395 -0.50083 -0.52967
…
Trirep  0.87258  1.85598  1.07994 -0.13477 -0.84708  1.12297
Viclat  0.09410  0.44974  0.85491 -0.09270  0.38100  0.10988

Site scores (weighted sums of species scores)

       RDA1     RDA2      RDA3      PC1     PC2      PC3
X1   0.3169  0.10573  0.312124 -0.88490  0.4845 -0.83687
X2   0.9030  0.39920  0.667488  0.10346 -0.7506 -0.26163
…
X19 -0.9860  0.53097  0.103590 -0.60726 -0.2858  0.50996
X20 -1.2599 -0.33664 -0.270509  0.81481  0.2488 -0.22649

Site constraints (linear combinations of constraining variables)

       RDA1    RDA2     RDA3
X1   0.2150 -0.7191  0.12068
X2   0.5179  0.7110  0.50471
…
X19 -0.8890  0.1518 -0.02099
X20 -0.8890  0.1518 -0.02099

Biplot scores for constraining variables

                RDA1    RDA2     RDA3
ManagementHF  0.4838  0.2911 -0.82535
ManagementNM -0.9793  0.1972 -0.04651
ManagementSF  0.2369 -0.9340  0.26735

Centroids for factor constraints

                RDA1    RDA2     RDA3
ManagementBF  0.5179  0.7110  0.50471
ManagementHF  0.4980  0.2541 -0.42245
ManagementNM -0.8890  0.1518 -0.02099
ManagementSF  0.2150 -0.7191  0.12068

Test for significance of all constrained eigenvalues
Pseudo-F:         2.840018
Significance:     < 0.001
Based on 1000 permutations under reduced model.
```

When you analyse the results, you could notice that the format is similar to the output of a PCA and CA. Note that there are 4 management practices, so 3 degrees of freedom or dimensions for this variable.

First the total, constrained and unconstrained variances are provided. The RDA provides an ordination that is constrained by some environmental variables. The RDA shows 29.2 from the total 84.1 variance, or 34.7% of variance is shown. This means that not all differences in Euclidean distance among sites are shown, but only those differences that can be related to differences in environmental variables.

Next the eigenvalues are provided. As for methods that were discussed earlier, the eigenvalues show how much variance is expressed on each axis. You can see two types of axes. The RDA axes are calculated for the RDA ordination. The PCA axes are calculated by a PCA on the variance that was not explained by the RDA. You can see that the RDA axes show the 34.7% of the total variance that can be explained by constrained axes (the accounted values indicate the cumulated proportion of variance). Again eigenvalues are listed from highest to lowest, thus making a graph of the first axes will result in most variance being shown.

The output follows with information on the scaling method used. As for PCA, scaling method 1 means that the distances among sites in the ordination graph corresponds to Euclidean distance among sites. Scaling method 2 would produce a correlation scaling as for PCA.

Next the species scores and site scores are provided. These scores are used in an ordination graph. These scores depend on the scaling method, and on differences in scaling of axes between software packages. The species scores should be interpreted as vectors indicating the direction of larger abundance for the species (see Figure 10.4). The biplot scores for constraining variables tell you how to interpret the RDA axes. RDA1 is roughly NM versus the rest, and RDA2 is roughly SF versus the rest (Figure 10.14).

RDA is a regression-type model that predicts where sites will occur in an ordination graph based on the environmental variables. The site constraints show the predicted value for each site. Note that the same value is predicted for each site with the same type of management. This is what we expect, as the predicted position will depend only on the environmental variables included, in this case management.

The biplot scores and the centroid scores are two methods of plotting an environmental variable in the ordination graph. The biplot scores are scores for vectors of each of the environmental variables used in the analysis. The direction from the origin to the biplot score shows sites with higher values for the environmental variable. The interpretation is thus similar to the species scores (see above and Figure 10.4). The centroid scores show the average position that is predicted for sites of the same category. The biplot and centroid scores are further discussed below in this section.

The output finishes with a significance test. The value of $P < 0.001$ shows that it is not very likely that the pattern that we observed was just random. We can thus be confident that we described an actual pattern of our data. The significance value was calculated from a permutation test. Recommended numbers of permutations are 1000 for a significance level of 0.05 and 5000 for a significance level of 0.01.

When we plot the site, species and centroid scores for the first RDA axes, then we obtain the ordination graph shown in Figure 10.14 (next page).

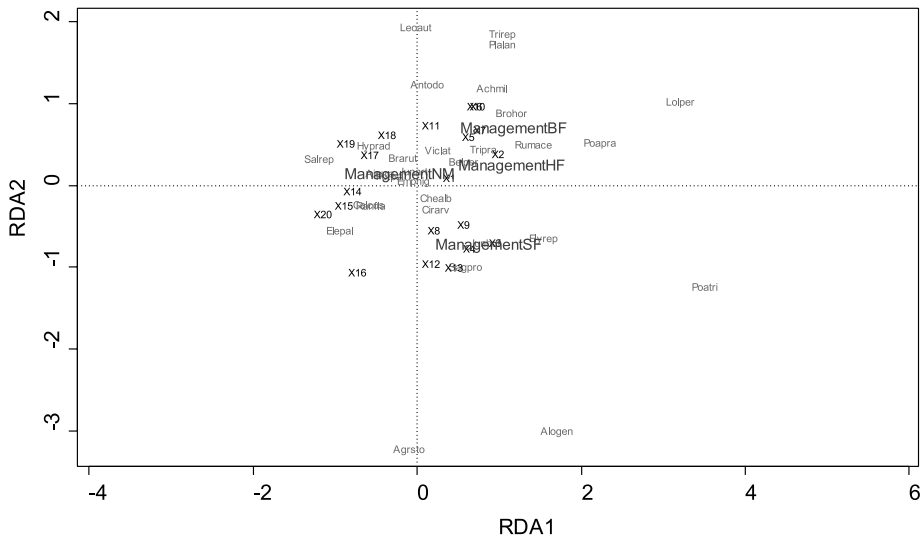

Figure 10.14 RDA ordination graph for the first two axes for the dune meadow dataset using scaling method 1 and the type of management as a constraining variable.

As discussed earlier, the species scores show the direction of higher abundance of a particular species. We thus expect higher abundance of *Alopecurus genitalicus* for sites X13 and X12, and lower abundance for X17, X18 and X19. The centroid scores show where sites of the same category of management are expected in the graph. These scores can be interpreted in a similar way as a site score. For instance, we expect sites with standard farming to contain more *Poa trivialis and Alopecurus genitalicus.*

As for regression analysis, it is possible to include several environmental variables. Figure 10.15 shows the ordination graph for the first two RDA axes when all environmental variables were used as explanatory variables. You can see that the ordination graph has become very busy with all the site, species, biplot and centroid scores. A method of producing graphs that can be interpreted more easily is to produce separate graphs for different sets of scores, for instance one graph with site and centroid scores, and one graph with species scores.

A slightly more complicated modification of RDA is partial redundancy analysis. This method is based on first removing the variance that can be explained by one subset of data, and then only analysing the residual variance. By this technique, it is for instance possible to remove the influence of spatial variables in the data first, and then analyse the influence by other variables. It is most useful for removing known environmental effects so that possible further effects can be explored (Borcard et al. 1992).

Similar to PCA, it is possible to transform the species matrix before the RDA, so that ecological distances other than the Euclidean distance are shown, such as the Hellinger distance or the distance between species profiles (see above: Principal components analysis on transformed species matrices). This transformation approach can substantially increase the usefulness of RDA.

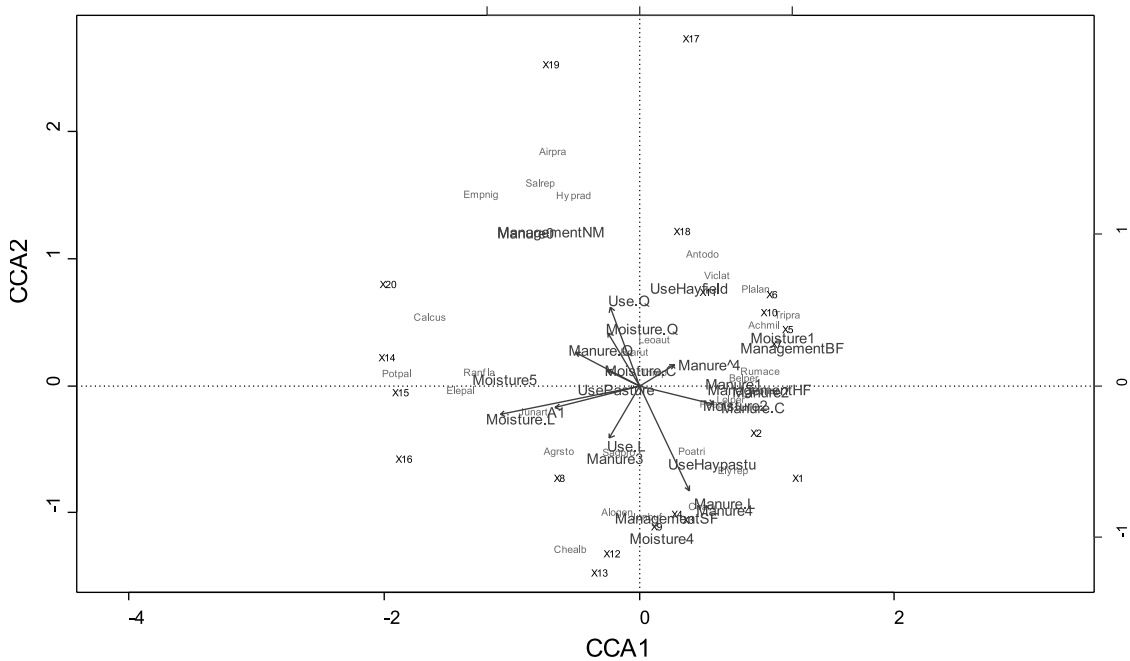

Figure 10.15 RDA ordination graph for the first two axes for the dune meadow dataset using scaling method 1 and all the environmental variables as constraining variables.

Canonical correspondence analysis

Canonical correspondence analyis (CCA) is an ordination technique that is related to CA, as suggested by its name. Both techniques show the chi-square distance among sites in the ordination graphs. As discussed in chapter 8, this distance is not the best ecological distance measure, although it is better than the Euclidean distance. In CCA the ordination is constrained by the environmental variables, shown in the environmental matrix. The approach of CCA is similar to RDA, with the CCA axes constrained to be linear combinations of environmental variables.

A CCA of the dune meadow dataset, using the type of management as a constraining variable gives the following result (next page):

```
Partitioning of mean squared contingency coefficient:

Total         2.1153
Constrained   0.6038
Unconstrained 1.5114

Eigenvalues, and their contribution to the mean squared contingency coefficient

            CCA1   CCA2   CCA3    CA1    CA2    CA3    CA4    CA5     CA6     CA7
lambda    0.3186 0.1825 0.1027 0.4474 0.2030 0.1630 0.1346 0.1294 0.09494 0.07904
accounted 0.1506 0.2369 0.2855 0.2115 0.3075 0.3845 0.4482 0.5093 0.55421 0.59158
              CA8     CA9    CA10    CA11    CA12    CA13     CA14     CA15
lambda    0.06526 0.05004 0.04321 0.03870 0.02385 0.01773 0.009172 0.007959
accounted 0.62243 0.64609 0.66651 0.68481 0.69609 0.70447 0.708805 0.712568
              CA16
lambda    0.004157
accounted 0.714533

Scaling 2 for species and site scores
-- Species are scaled proportional to eigenvalues
-- Sites are unscaled: weighted dispersion equal on all dimensions

Species scores

           CCA1     CCA2      CCA3       CA1       CA2      CA3
Achmil  0.19825  0.72622  0.300280  0.536978 -0.435913  0.64609
Agrsto -0.10058 -0.73046  0.007614 -0.681142  0.217728  0.03744
…
Trirep -0.04953  0.33644  0.186947 -0.090075 -0.056136  0.12455
Viclat -0.14617  0.98179  1.261557  0.241368 -0.211654 -0.51063


Site scores (weighted averages of species scores)

        CCA1    CCA2     CCA3       CA1      CA2      CA3
X1   1.35561  0.4747  1.06564  0.823027 -2.26986 -0.78064
X2   0.86419  0.8142  1.87245 -0.198170 -0.21094 -0.45453
…
X19 -2.61902  0.6190  0.42849  2.309910  2.21260 -0.08448
X20 -2.32115 -1.3309 -1.00044 -1.476716 -0.09132 -0.90420


Site constraints (linear combinations of constraining variables)

       CCA1     CCA2     CCA3
X1   0.5601 -1.38599  0.35086
X2   0.4313  1.32739  1.70492
…
X19 -1.8785 -0.05503 -0.06852
X20 -1.8785 -0.05503 -0.06852
```

```
Biplot scores for constraining variables

                CCA1     CCA2     CCA3
ManagementHF  0.3716  0.42702 -0.82287
ManagementNM -0.9989 -0.02624 -0.03819
ManagementSF  0.3665 -0.90307  0.22847

Centroids for factor constraints

                CCA1     CCA2     CCA3
ManagementBF  0.4313  1.32739  1.70492
ManagementHF  0.5583  0.63731 -1.22396
ManagementNM -1.8785 -0.05503 -0.06852
ManagementSF  0.5601 -1.38599  0.35086

Test for significance of all constrained eigenvalues
Pseudo-F:         2.130750
Significance:     0.001
Based on 1000 permutations under reduced model.
```

The output provided above is similar to the output for CA and for RDA.

First the total and constrained variance is given, followed by the variance shown on each axes by the eigenvalues. We can see that 28.5% of variance is accounted in the CCA, with 23.7% on the first two axes (the accounted values indicate the cumulated proportion).

The output is followed by an indication of the scaling method that was used. The scaling method is the same as for CA, scaling method 1 meaning that the chi-square distances among sites are shown.

The species, site, biplot, and centroid scores are the coordinates for species, sites and environmental variables in an ordination graph. The interpretation of these scores is similar to the interpretation for RDA. Differences with RDA are that the species scores are weighted average scores (as for CA, see Figure 10.13), and that the distance among sites reflects the chi-square distance.

Figure 10.16 (next page) shows the ordination graph for the dune meadow dataset, using the scores provided above.

The significance test is also based on permutation, as for RDA. The $P < 0.001$ indicates that the observed relationship between environmental variables and ecological distance is not due to chance.

As for RDA, there is a partial CCA method. In this method, it is possible to remove the effect of one subset of variables first, and then analyse the effects of other variables.

Distance-based redundancy analysis and canonical analysis of principal coordinates

Distance-based redundancy analyis (db-RDA) and canonical analysis of principal coordinates (CAP) are constrained ordination techniques that analyse results of PCoA further. The advantage of these techniques is that any type of ecological distance can be analysed, including distances that are better for investigating differences in species composition as seen in chapter 8. Since db-RDA and CAP are constrained ordination techniques, the influence of environmental variables on differences in the selected good ecological distance can be investigated, and they therefore offer more advanced ways of constrained analysis than RDA and CCA. Although both db-RDA and CAP analyse the results of PCoA, the methods that they

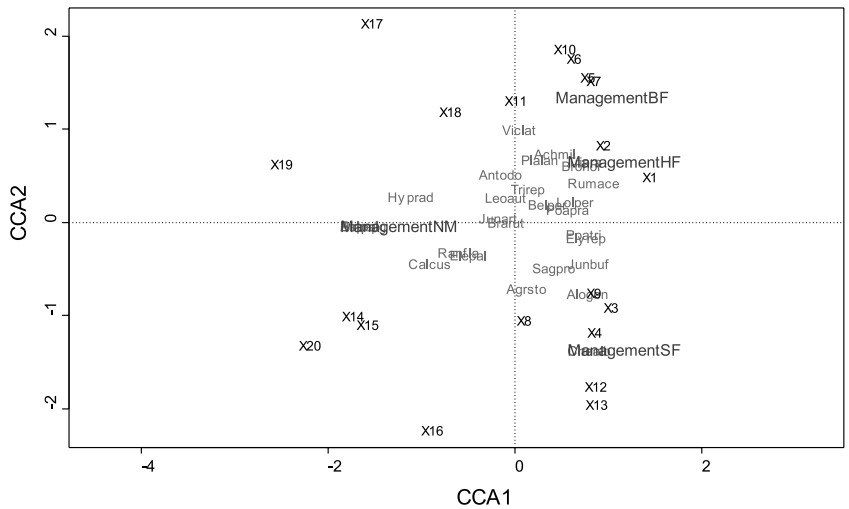

Figure 10.16 CCA ordination graph for the first two axes for the dune meadow dataset using scaling method 1 and management as constraining variable.

use are different: db-RDA uses RDA, whereas CAP uses linear discriminant analysis or canonical correlation analysis. A detailed discussion of the differences between both methods is beyond the scope of this chapter (see Legendre and Anderson 1999; Anderson and Willis 2003).

A db-RDA of the dune meadow dataset, using the type of management as a constraining variable, and the Bray-Curtis distance to express ecological distance gives the following result:

```
Partitioning of squared Bray distance:

Total         87.28
Constrained   28.50
Unconstrained 58.78

Eigenvalues, and their contribution to the squared Bray distance

             CAP1    CAP2    CAP3      PC1     PC2     PC3    PC4     PC5     PC6     PC7
lambda    17.0957 8.6126 2.7943 24.1862 9.2599 7.1894 5.891 3.9894 2.8640 1.6239
accounted  0.1959 0.2945 0.3265  0.2771 0.3832 0.4656 0.533 0.5787 0.6116 0.6302
             PC8     PC9    PC10    PC11    PC12     PC13     PC14
lambda    1.385 1.1412 0.6116 0.3267 0.1858 0.08386 0.04491
accounted 0.646 0.6591 0.6661 0.6698 0.6720 0.67294 0.67345
```

```
Scaling 1 for species and site scores
-- Sites are scaled proportional to eigenvalues
-- Species are unscaled: weighted dispersion equal on all dimensions

Species scores

           CAP1       CAP2      CAP3      PC1      PC2      PC3
Achmil  0.48210 -1.5325659 -0.459148 -0.64885 -0.19187 -1.38213
Agrsto  0.23882  3.6158066 -0.582905  2.26057 -0.05501 -0.12210
…
Trirep  0.53227 -2.3206841 -0.882848  0.27147  0.43253 -0.78951
Viclat  0.03078 -0.5633329 -0.911828 -0.07164  0.03791  0.53027

Site scores (weighted averages of species scores)

        CAP1    CAP2     CAP3      PC1      PC2      PC3
X1   1.10024 -0.3275 -0.50105 -1.16792 -1.16579  0.02871
X2   0.85167 -0.5197 -0.51581  0.05581 -0.23602 -0.26113
…
X19 -1.07753 -0.5148 -0.12207 -0.74980  0.89546  0.40619
X20 -1.29531  0.8673  0.17204  1.00331 -0.28929  0.22866

Site constraints (linear combinations of constraining variables)

       CAP1       CAP2      CAP3
X1   0.4100  0.5796560 -0.124936
X2   0.4034 -0.7544929 -0.397613
…
X19 -0.9646 -0.0001609  0.004098
X20 -0.9646 -0.0001609  0.004098

Biplot scores for constraining variables

                CAP1       CAP2     CAP3
ManagementHF  0.3872 -0.3126081  0.86739
ManagementNM -0.9999 -0.0002350  0.01051
ManagementSF  0.4250  0.8465954 -0.32035

Centroids for factor constraints

                CAP1       CAP2      CAP3
ManagementBF  0.4034 -0.7544929 -0.397613
ManagementHF  0.4235 -0.2426984  0.383574
ManagementNM -0.9646 -0.0001609  0.004098
ManagementSF  0.4100  0.5796560 -0.124936

Test for significance of all constrained eigenvalues
Pseudo-F:        2.26279
Significance:    0.006
Based on 1000 permutations under reduced model.
```

The output is very similar to the output of a RDA. The site, species, biplot and centroid scores should be analysed in a similar way. We can see that db-RDA expresses 32.6% of total squared distance. The first two axes express 29.4% of squared distance (the accounted values indicate the cumulated proportion).

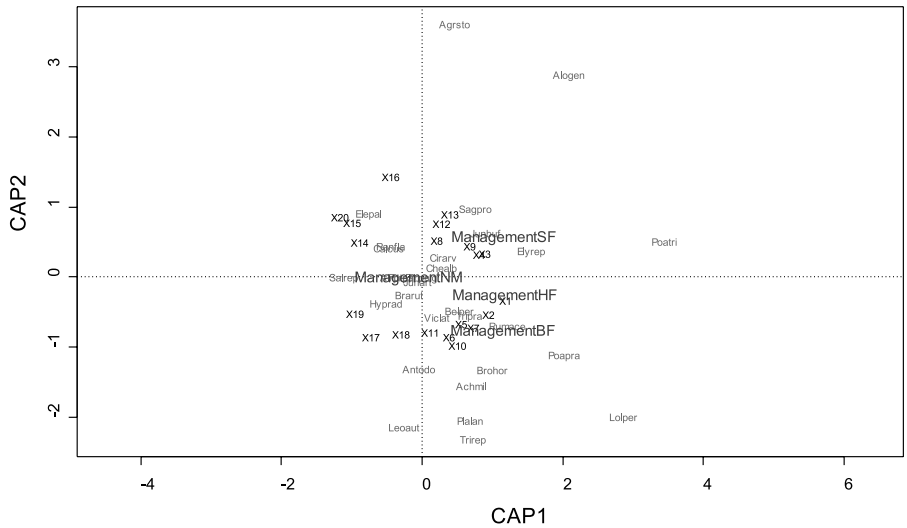

Figure 10.17 Db-RDA ordination graph for the first two axes for the dune meadow dataset using scaling method 1, the Bray-Curtis distance and the type of management as a constraining variable.

When we plot the site, species and centroid scores for the first db-RDA axes, then we obtain the ordination graph shown in Figure 10.17

As discussed for RDA, the species scores show the direction of higher abundance of a particular species. We thus expect higher abundance of *Alopecurus genitalicus* for sites X13 and X12, and lower abundance for X17, X18 and X19. The centroid scores show where sites of the same category of management are expected in the graph. These scores can be interpreted in a similar way as a site score. For instance, we expect that sites with standard farming will contain more *Agrostis stolonifera* and *Alopecurus genitalicus*.

Whereas we investigated the influence of a categorical variable on the differences in species composition above, you can also include continuous variables as explanatory variables. A db-RDA of the dune meadow dataset using the depth of the A1 horizon as the constraining variable and the Bray-Curtis distance to express ecological distance gives the result shown on the next page.

The results are very similar to those for management shown earlier. Because only one continuous explanatory variable was used, only one constrained ordination axis was obtained accounting for 15.8% of total squared Bray-Curtis distance.

Because only one constrained axis was obtained, an ordination graph was constructed that also used the first residual axis as shown in Figure 10.18. The vector for the depth of the A1 horizon indicates the direction in the graph where sites are expected that have deeper A1 horizon: these sites occur on the right-hand side of the graph. We therefore infer from the graph that sites that have a deeper A1 horizon are expected to contain more *Agrostis stolonifera* and *Eleocharis palustris*, whereas sites that have a more shallow A1 horizon would have more *Lolium perenne* and *Poa pratensis*.

```
Call:
capscale(formula = dune ~ A1, data = dune.env, distance = "bray")

Partitioning of squared Bray distance:

Total         87.28
Constrained   13.80
Unconstrained 73.49

             CAP1     PC1     PC2    PC3    PC4    PC5   PC6    PC7    PC8
lambda    13.7975 21.6515 18.8817 8.7678 6.8894 4.5809 4.083 3.2127 1.8245
accounted  0.1581  0.2481  0.4644 0.5648 0.6438 0.6962 0.743 0.7798 0.8007
            PC9   PC10   PC11   PC12   PC13    PC14
lambda    1.414 1.1725 0.5201 0.3357 0.1033 0.05102
accounted 0.817 0.8304 0.8363 0.8402 0.8413 0.84193

Scaling 1 for species and site scores
-- Sites are scaled proportional to eigenvalues
-- Species are unscaled: weighted dispersion equal on all dimensions

Species scores

           CAP1      PC1      PC2      PC3      PC4      PC5
Achmil -0.67633 -1.04282  0.27838 -0.28142  0.95411 -0.65458
Agrsto  1.64726  2.89575  0.65578  0.51099  0.68881  0.23853
…
Trirep  0.21369 -0.86349  0.61722  0.31855 -0.82554  0.60036
Viclat -0.18672 -0.18742 -0.13515 -0.27223 -0.66189  0.35546

Site scores (weighted averages of species scores)

        CAP1      PC1      PC2      PC3      PC4      PC5
X1  -1.19851 -0.13951  0.83107 -0.92749  0.30446  0.18771
X2  -0.80470 -0.33601  0.73210 -0.09861  0.08395  0.49669
…
X19 -0.06976 -0.23731 -1.47115  0.64386 -0.22138  0.47097
X20  1.36948  1.49536 -1.20732 -0.43657  0.05829 -0.27214

Site constraints (linear combinations of constraining variables)

        CAP1
X1  -0.54756
X2  -0.36059
…
X19 -0.30717
X20 -0.36059

Biplot scores for constraining variables

   CAP1
A1    1

Test for significance of all constrained eigenvalues
Pseudo-F:        2.26279
Significance:    0.006
Based on 1000 permutations under reduced model.
```

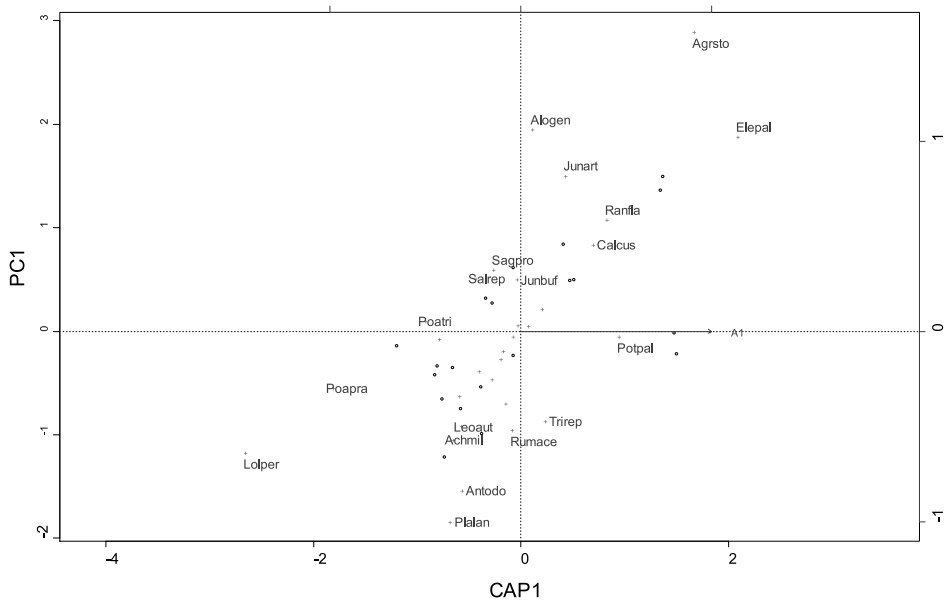

Figure 10.18 Db-RDA ordination graph for the first constrained and first residual axis for the dune meadow dataset using scaling method 1, the Bray-Curtis distance and the depth of the A1 horizon as a constraining variable. Sites are plotted as circles and species as crosses.

Choice of ordination method

As we saw in chapter 8, different ecological distance measures have different properties. The properties of the ordination method you use therefore depend on the properties of the distance measure on which it is based. The first rule should be that you use a distance measure that is a useful ecological distance measure. A good analysis practice is also to repeat the analysis with several good distance measures and investigate whether all these analyses lead to the same conclusion.

There are some methods to investigate how well the distances in the ordination graph represent the total distances as provided in a distance matrix. The first method calculates the percentage of variance that is displayed in the graph. The second method compares the ecological distance between sites with the distance between the positions of the sites in the ordination graph. The distances can be plotted against one another and so can you can check how good the correlation is. Figure 10.19 shows the correspondence between the Bray-Curtis distance on the first two axes of a PCoA (as shown in Figure 10.8), and the total Bray-Curtis distance among sites. You can see that the overall correlation is quite good, but that some sites are plotted relatively closer than other sites despite the same total distance.

A third method of investigating how well the distances in the ordination graph represent the total distance combines clustering results with ordination results. It is recommended to plot the results of single linkage (see Chapter 9) on top of an ordination graph. The results of single linkage produce a minimum spanning tree that shows how sites and clusters can be joined together by the smallest possible total length of segments among them. Figure 10.20 gives an example for the first two axes of PCoA. For example, sites X20 and X15 are plotted close together in the graph and were also joined by the minimum spanning tree. This is an indication that the graph provides a good representation of their ecological distance.

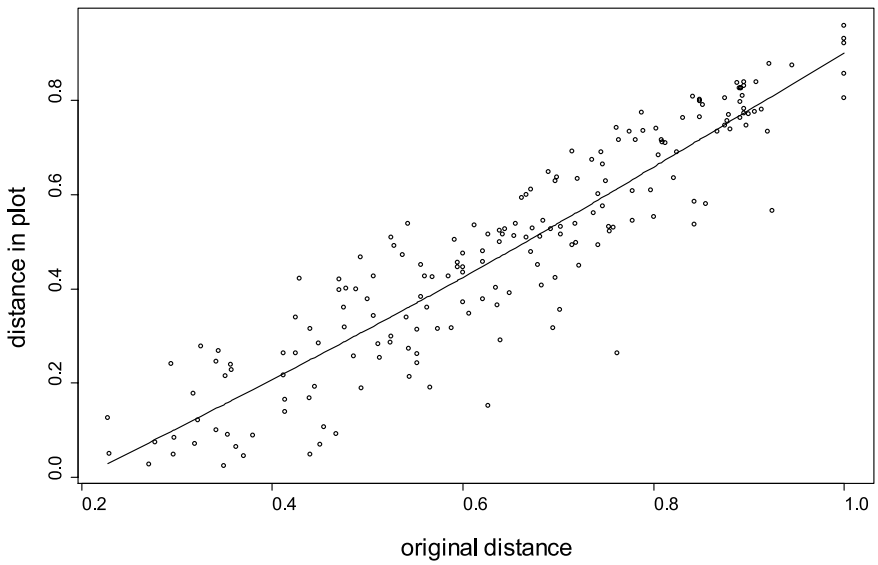

Figure 10.19 Relationship between distances between site positions in an ordination graph (Figure 10.8) and total Bray-Curtis distance between sites. The line shows the fit of a GAM (see chapters 6 and 7) between the original distances and the distances in the ordination graph.

If sites that had a small distance in the graph were joined through the minimum spanning tree at far distance in the graph, such as sites X1 and X2, this means that they are only joined much later in the process. The distance between X1 and X2 is thus not well presented in the graph. A better ordination graph will have a shorter length of the minimum spanning tree.

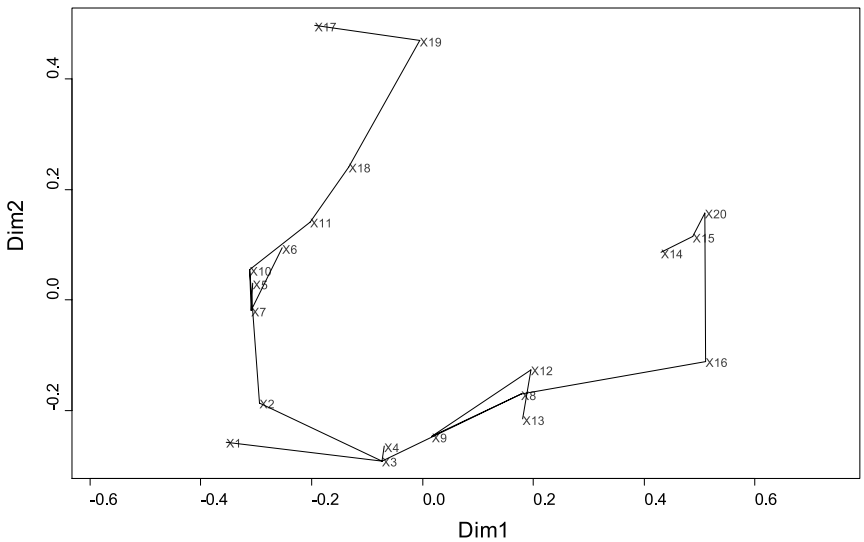

Figure 10.20 Plotting the cluster structure on top of an ordination graph (Figure 10.8) by a minimum spanning tree to investigate how well ecological distance is represented in the ordination graph.

Further interpretation of ordination graphs by indirect gradient analysis

As with constrained ordination, indirect gradient analysis methods also seek to understand the relationship between environmental variables of a site and their species composition. They are applied *after* an unconstrained ordination analysis. The key idea is to try to relate the pattern of sites in the ordination graph to environmental variables.

There are three methods of investigating quantitative environmental variables.

The first method calculates a fitted vector of an environmental variable with the ordination configuration. This vector shows the direction in the ordination graph where sites are expected with values that are higher than average for the environmental variable. This is a similar approach as the one shown earlier for calculating correlation scores for a species for a PCoA or NMS. The interpretation is also similar (see also Figure 10.4).

When you calculate the vector scores for the depth of the A1 horizon and the first two axes of a PCoA based on the Bray-Curtis distance for the dune meadow dataset, then you obtain the result shown below.

The scores for the head of the vector (listed in the result for Dim1 and Dim2) can be used to plot a vector for the environmental variable onto the ordination graph. Figure 10.21 shows how the results presented above can be presented graphically. You can see that we expect greater depth of the A1 horizon on sites X14, X15, X16 and X20.

The second method is to plot the values of the environmental variable as a bubble graph. This approach is more general than fitting a vector, as this method does not assume that the values will increase linearly on an axis. Figure 10.22 shows the bubble graph of the depth of the A1 horizon. Large bubbles indicate a larger value for the A1 horizon. The graph indicates that in general depth of the A1 horizon increases with Dim1, but there are exceptions. We can see large values for X14 and X16, but the value of X20 is not so large. There is no sign of depth of A1 changing with Dim2 except for X14 and X15. This example shows how the vector (Figure 10.21) picks out trends but does not show any detail of deviation from the trend.

The third method of investigating a quantitative environmental variable is to fit a surface that models how the environmental variable changes over the ordination graph. Rather than simply plotting the bubbles, we can try to describe the pattern in bubble size by a smooth surface, using a GAM as described in chapter 7.

When calculating the surface for the depth of the A1 horizon for the PCoA based on the Bray-Curtis distance for the dune meadow dataset, you obtain the result presented in Figure 10.23. This figure shows a similar picture as Figure 10.22, with increasing values from left to right in the ordination graph (the GAM approach can reveal more complex patterns, but in this case the algorithm fitted a linear trend). The lower value for site X20 is not reflected. If the residuals from the fitted surface were plotted (figure not included), then it would be clear that X20 is not represented well in the Figure 10.23. It is a good statistical practice to investigate residuals.

For categorical environmental variables, there are some other methods of indirect gradient analysis.

The first method is to use a different symbol for each category of the environmental variable. Figure 10.24 gives an example for the type of management for the PCoA ordination for the dune meadow dataset. We can see that all sites with nature management are plotted at the top

```
      Dim1    Dim2     r2 Pr(>r)
A1 0.98806 0.15404 0.3845 < 0.01 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
P values based on 100 permutations.
```

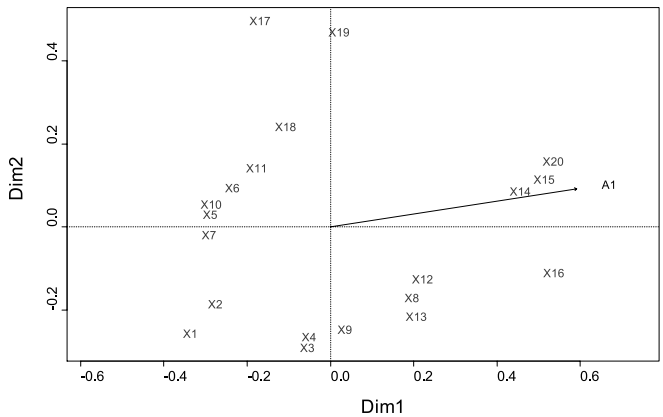

Figure 10.21 Plotting a vector for a quantitative environmental variable onto an ordination graph (Figure 10.8).

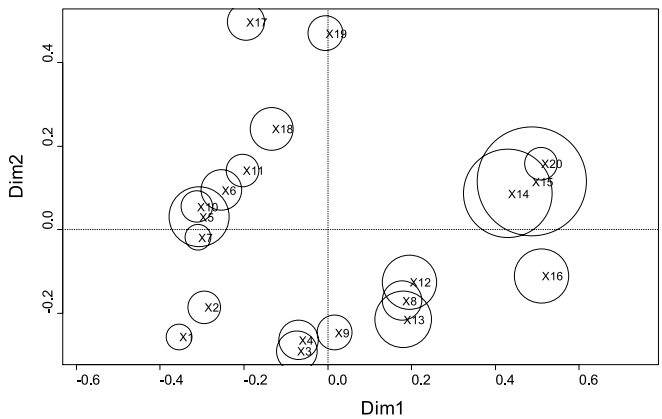

Figure 10.22 Plotting a bubble graph for a quantitative environmental variable onto an ordination graph (Figure 10.8).

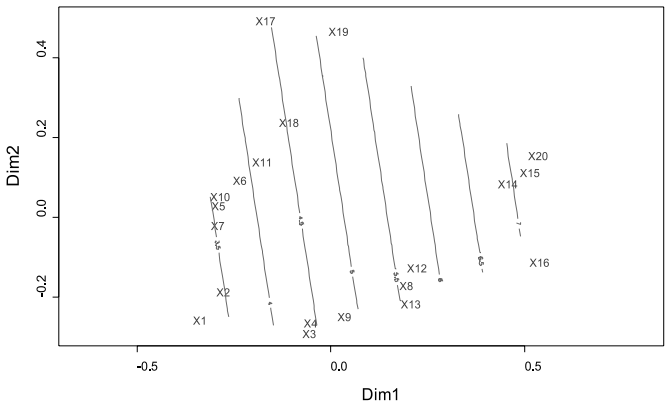

Figure 10.23 Plotting a contour for a quantitative environmental variable onto an ordination graph (Figure 10.8).

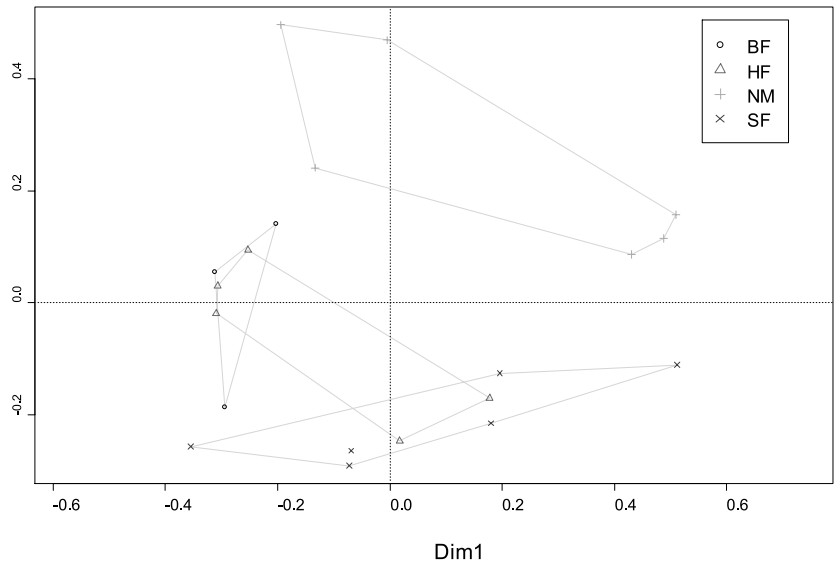

Figure 10.24 Plotting different symbols for different categories of a categorical environmental variable onto an ordination graph (Figure 10.8). A convex hull encloses all sites of the same category.

part of the graph. We can also see that some sites with hobby farming are more similar in species composition to sites with biological farming, whereas other sites with hobby farming are more similar in species composition to sites with standard farming.

The second method for categorical environmental variables is to calculate the average plotting position for each category (the centroids). When you calculate the average positions for management for the dune meadow dataset and the PCoA graph, then you obtain the following result:

```
Centroids:
       Dim1    Dim2
PBF -0.2694  0.0032
PHF -0.1350 -0.0623
PNM  0.1822  0.2609
PSF  0.0650 -0.2106

Goodness of fit:
      r2 Pr(>r)
P 0.4482   0.01 **
```

These centroid positions can be plotted onto the ordination graph (Figure 10.25). By connecting each site with the centroid of the same category (a spiderplot), you can investigate whether some sites are outliers. We can observe again that sites with nature management have a different species composition, and that sites with hobby farming are either more similar to sites with standard farming or sites with biological farming. Since the convex hulls (Figure 10.24) are more sensitive to outliers, you need to be careful in making conclusions that species composition is similar when convex hulls overlap. When convex hulls do not overlap, this provides evidence that species composition is dissimilar.

A third method is to calculate confidence ellipses that predict where sites of a certain category will occur. This method estimates a confidence interval for sites of each category, using the positions of the sites on the horizontal and vertical axes as input variables. This approach is thus more sophisticated than the previous methods for categorical variables.

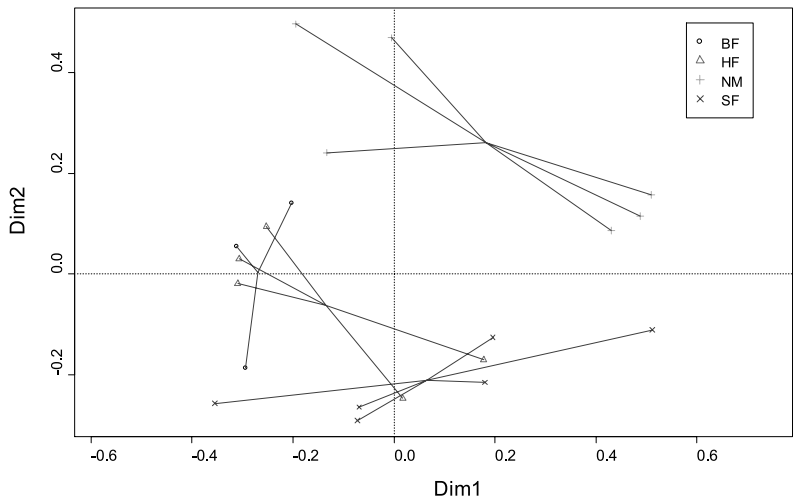

Figure 10.25 Connecting sites to the centroid of each category onto an ordination graph (Figure 10.8).

Figure 10.26 gives an example for the type of management for the PCoA ordination for the dune meadow dataset. The ellipses indicate where 95% of sites of the same category are expected to occur. Different symbols were also used for the different categories. You can see again that sites with nature management occur at the top of the graph, and that there is an overlap for sites with hobby farming and the other two categories of standard farming and biological farming.

Further interpretation of ordination graphs for individual species

By analogy with indirect gradient analysis methods, patterns of some individual species can be analysed *after* an ordination analysis. You can use the ordination method to check whether sites are different in species composition, and then check for the species that contribute most to the differences.

Figure 10.27 shows the results of the CAP

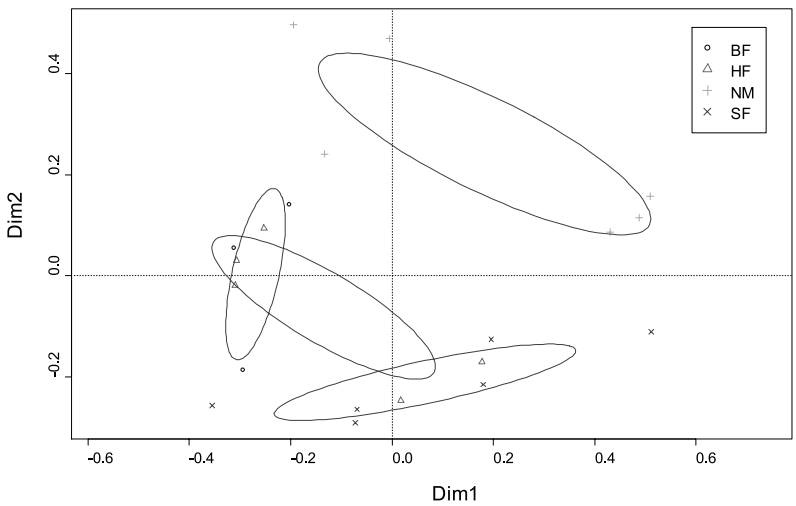

Figure 10.26 Drawing confidence ellipses for each category onto an ordination graph (Figure 10.8).

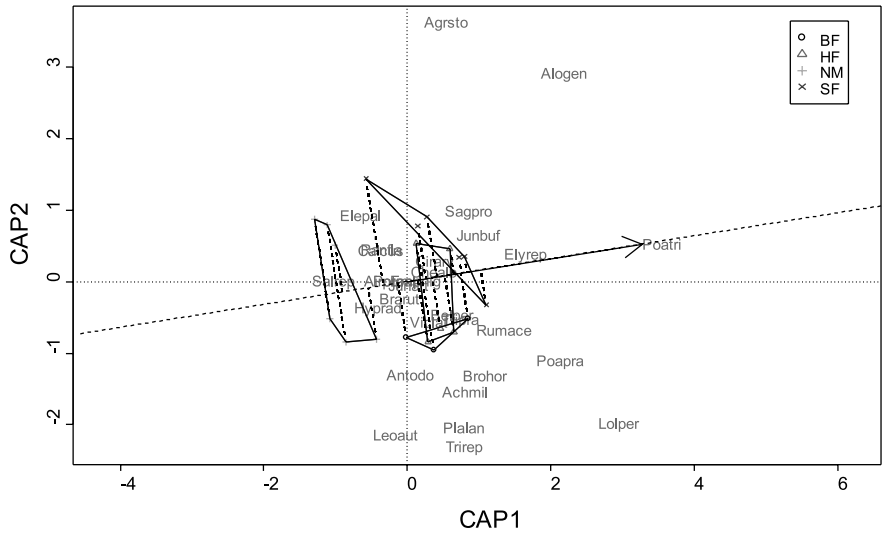

Figure 10.27 Investigating for important species that contribute to the differences in species composition. The first axes of a CAP analysis of the dune meadow dataset based on the Bray-Curtis distance and with type of management as explanatory factor.

analysis for the dune meadow dataset (based on the Bray-Curtis distance) for differences in species composition related to differences in management. The results of this analysis were provided above (including Figure 10.17). Added to the earlier results are the hulls for the different management categories, symbols for the different categories and the interpretation for species *Poa trivialis*. Since this species has a long species vector, we expect that it contributes to differences in species composition between types of management. We can also expect that abundances for this species will be lower for nature management, since the projected scores for this species are lower for this type of management.

The formal way of testing for the differences for *Poa trivialis* for the different types of management is a regression analysis, as seen in chapter 7. A GLM regression with log link and quasipoisson variance functions gives the result shown on the next page.

We can see from the regression coefficients that fewer individuals of *Poa trivialis* are predicted for nature management (checking the dune meadow datasets reveals that the species does not occur on the six quadrats with nature management). The ANOVA provides evidence for differences among categories for management. The standard errors and significance levels for the regression coefficients are large, however. What is happening? The reason is that the sample size is quite small in comparison to the four categories of management to investigate differences for individual species. In this case, we can observe actual differences in abundance, but we can not confirm that these differences were not observed by chance. This means that we could demonstrate that the sites are different in composition, but could not check for an individual species. This result is not entirely surprising, since the occurrences of species are correlated with each other. As we investigate several species at the same time in an ordination analysis, the investigation becomes more powerful.

```
glm(formula = Poatri ~ Management, family = quasipoisson(link = log),
    data = dune.env, na.action = na.exclude)

Deviance Residuals:
      Min         1Q     Median         3Q        Max
-2.708013  -0.331340  -0.000091   0.157273   1.776335

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.2993     0.2908   4.468 0.000388 ***
Management[T.HF]    0.2693     0.3512   0.767 0.454260
Management[T.NM]  -20.6019  3711.5165  -0.006 0.995640
Management[T.SF]    0.2412     0.3432   0.703 0.492330

(Dispersion parameter for quasipoisson family taken to be 0.9301272)

    Null deviance: 63.412  on 19  degrees of freedom
Residual deviance: 17.842  on 16  degrees of freedom


Analysis of Deviance Table

           Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
NULL                         19     63.412
Management  3   45.570        16     17.842 16.331 3.98e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

References

Anderson MJ and Willis TJ. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84, 511–525.

Borcard D, Legendre P and Drapeau P 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045-1055.

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.

Legendre P and Anderson MJ. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69: 1-24.

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)

Makarenkov V and Legendre P. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology* 83: 1156-1161.

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

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.

ter Braak CJF. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167-1179.

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 meadow dataset

Calculating a principal component analysis (PCA):

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCA (or PCA (prcomp))

→scaling: 1

→Plot method: ordiplot

→Plot method: text sites

→Plot method: text species

→Plot method: equilibrium circle

Calculating a PCA on a transformed matrix:

Biodiversity > Community matrix > Transform community matrix…

→Method: Hellinger

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCA

Conducting a principal coordinates analysis (PCoA)

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCoA (or PCoA (Caillez))

→Distance: bray

Calculating a non-metric multidimensional scaling (NMS)

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: NMS (or NMS (standard))

→Distance: bray

→NMS axes: 2

→NMS permutations: 100

Calculating a correspondence analysis (CA or WA)

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: CA

→scaling: 1

Calculating a redundancy analysis (RDA):

Biodiversity > Analysis of ecological distance > Constrained ordination…

→Ordination method: RDA

→scaling: 1

→permutations: 100

→Explanatory: Management

Calculating a canonical correspondence analysis (CCA)

Biodiversity > Analysis of ecological distance > Constrained ordination…

→Ordination method: CCA

→scaling: 2

→permutations: 100

→Explanatory: Management

Calculating distance-based redundancy analysis (db-RDA)

Biodiversity > Analysis of ecological distance > Constrained ordination…

→Ordination method: capscale

→distance: bray

→permutations: 100

→Explanatory: Management

Calculating the correlation between distance in an ordination graph and total distance

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCoA

→Distance: bray

→Plot method: ordiplot

→Plot method: distance displayed

Plotting clustering results onto an ordination graph:

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCoA

→Distance: bray

→Plot method: ordiplot

→Plot method: ordicluster

Plotting quantitative environmental variables onto an ordination graph:

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCoA

→Distance: bray

→Plot variable: A1

→Plot method: ordiplot

→Plot method: vectorfit

→Plot method: ordibubble

→Plot method: ordisurf

Plotting categorical environmental variables onto an ordination graph:

Biodiversity > Analysis of ecological distance > Unconstrained ordination…

→Ordination method: PCoA

→Distance: bray

→Plot variable: Management

→Plot method: ordiplot

→Plot method: factorfit

→Plot method: ordihull

→Plot method: ordispider

→Plot method: ordiellipse

→Plot method: ordisymbol

Doing the analyses with the command options of Biodiversity.R

Calculating a principal component analysis (PCA)

```
Ordination.model1 <- rda(dune)
summary(Ordination.model1, scaling=1)
plot1 <- ordiplot(Ordination.model1, scaling=1, type="text")
plot2 <- ordiplot(Ordination.model1, scaling=2, type="text")
```

Calculating the variance of each species of the species matrix

```
inertcomp(Ordination.model1, display='species',
  statistic='explained', proportional=F)
```

Calculating the proportion of variance explained for an ordination graph

```
goodness(Ordination.model1, display='sites', choices=c(1:2),
  statistic='explained')
```

Adding a vector and perpendicular lines for a particular species to an ordination plot

```
ordivector(plot1,"Agrsto",lty=2)
```

Calculating correlations among vectors:

```
cor.test(dune[,"Alogen"],dune[,"Agrsto"])
```

Calculating the number of ecologically meaningful principal components:

```
PCAsignificance(Ordination.model1,axes=30)
```

Drawing an equilibrium circle

```
ordiequilibriumcircle(Ordination.model1,plot1)
```

Calculating a PCA on a transformed matrix

```
Community.1 <- disttransform(dune, method='Hellinger')
Ordination.model2 <- rda(Community.1)
summary(Ordination.model2, scaling=1)
plot3 <- ordiplot(Ordination.model2, scaling=1, type="text")
```

Calculating a principal coordinates analysis (PCoA)

```
distmatrix <- vegdist(dune,method='bray')
Ordination.model3 <- cmdscale(distmatrix, k=nrow(dune)-1,
  eig=T, add=F)
Ordination.model3 <- add.spec.scores( Ordination.model3, dune,
  method='pcoa.scores', Rscale=T, scaling=1, multi=1)
plot4 <- ordiplot(Ordination.model3, type='text')
```

Calculating a non-metric multidimensional scaling (NMS)

```
distmatrix <- vegdist(dune, method='bray')
initNMS <- NMSrandom(distmatrix, perm=100, k=2)
Ordination.model4 <- postMDS(initNMS, distmatrix)
Ordination.model4 <- add.spec.scores( Ordination.model4, dune,
  method='wa.scores')
Ordination.model4
plot5 <- ordiplot(Ordination.model4)
```

Calculating a correspondence analysis (CA or WA)

```
Ordination.model5 <- cca(dune)
summary(Ordination.model5, scaling=1)
plot6 <- ordiplot(Ordination.model5, type='text', scaling=1)
```

Calculating a redundancy analysis (RDA)

```
Ordination.model6 <- rda(dune ~ Management, dune.env)
summary(Ordination.model6, scaling=1)
permutest.cca(Ordination.model6, permutations=1000)
plot7 <- ordiplot(Ordination.model1, type='text', scaling=1)
```

Calculating a canonical correspondence analysis (CCA)

```
Ordination.model7 <- cca(dune ~ Management, dune.env)
summary(Ordination.model7, scaling=2)
permutest.cca(Ordination.model7, permutations=1000)
plot8 <- ordiplot(Ordination.model7, type='text', scaling=1)
```

Calculating distance-based redundancy analysis (db-RDA)

```
Ordination.model8 <- capscale(dune ~ Management, dune.env)
summary(Ordination.model8, scaling=1)
permutest.cca(Ordination.model8, permutations=1000)
plot9 <- ordiplot(Ordination.model8, type='text', scaling=1)
```

Calculating the correlation between distance in an ordination graph and total distance

```
distdisplayed(dune, plot4, distx='bray')
```

Plotting clustering results onto an ordination graph

```
distmatrix <- vegdist(dune, method='bray')
cluster <- hclust(distmatrix, method='single')
ordicluster(plot4, cluster)
```

Plotting quantitative environmental variables onto an ordination graph

```
fitted <- envfit(plot4, A1, permutations=100)
plot(fitted)
fitted <- vectorfit(plot4, A1, permutations=100)
ordibubble(plot4, A1)
ordisurf(plot4, A1)
```

Plotting categorical environmental variables onto an ordination graph

```
ordisymbol(plot4, dune.env, 'Management', legend=T)
fitted <- envfit(plot4, Management, permutations=100)
plot(fitted)
fitted <- factorfit(plot4, Management)
ordihull(plot4, Management)
ordispider(plot4, Management)
ordiellipse(plot4, Management)
```