4. Ordination in reduced space

4.1. Generalities

Contrary to most clustering techniques, which aim at revealing discontinuities in the data, ordination is particularly good at displaying **gradients**. A detailed account on how to compute an ordination goes beyond this short introduction. Here we shall present an overview of the most useful methods available, with an intuitive explanation about the way they work.

Suppose you have a series of observations (objects) characterised by two variables. The objects could be represented in a two-dimensional space, each dimension being one of the variables (Figure 17):

	Var.1	Var.2
Obj.1	1	5
Obj.2	4	4
Obj.3	3	3
Obj.4	1	3
Obj.5	2	2
Obj.6	4	1

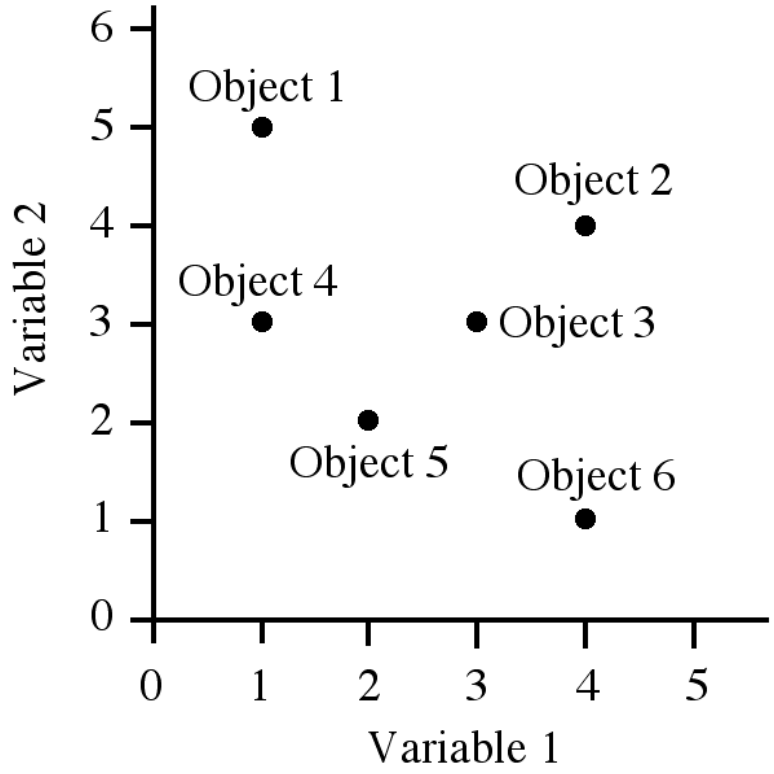

Figure 17 - Ordination of six objects in the space of two variables.

A matrix of raw data (for instance objects × physico-chemical measurements) generally contains more than two variables, often many of them. In this case it becomes difficult, cumbersome and not very informative to draw the objects in a series of planes defined by all possible pairs of descriptors. For instance, if the matrix contains 10 descriptors, the number of planes to draw would be equal to (10 × 9)/2 = 45. Such a series of scatterplots would allow neither to bring out the most important structures of the data, nor to visualise the relationships among descriptors (which, in general, are not linearly independent from one another anyway).

The aim of the ordination methods is to represent the data in a **reduced number of orthogonal axes**, constructed in such a way that they represent, in decreasing order, the main trends of the data. Here we shall mention four basic techniques: principal component analysis (PCA), correspondence analysis (CA), principal coordinate analysis (PCoA) and nonmetric multidimensional scaling (NMDS).

4.2. Principal component analysis (PCA)

Imagine, again, a data set made of *n* objects by *p* variables. The *n* objects can be represented as a cluster of points in a *p*-dimensional space. Now, this cluster is generally not completely spheroidal: it is elongated in several directions, flattened in others. These directions are not necessarily aligned with one single dimension (= one single variable) of the multidimensional space. The direction where the cluster is most elongated corresponds to the direction of *largest variance* of the cluster.

PCA realises a rigid rotation of the original system of axes, such as the successive new axes (called principal components) are orthogonal to one another, and correspond to the successive dimensions of maximum variance of the scatter of points. The principal components

give the positions of the objects in the new system of coordinates (Figure 18):

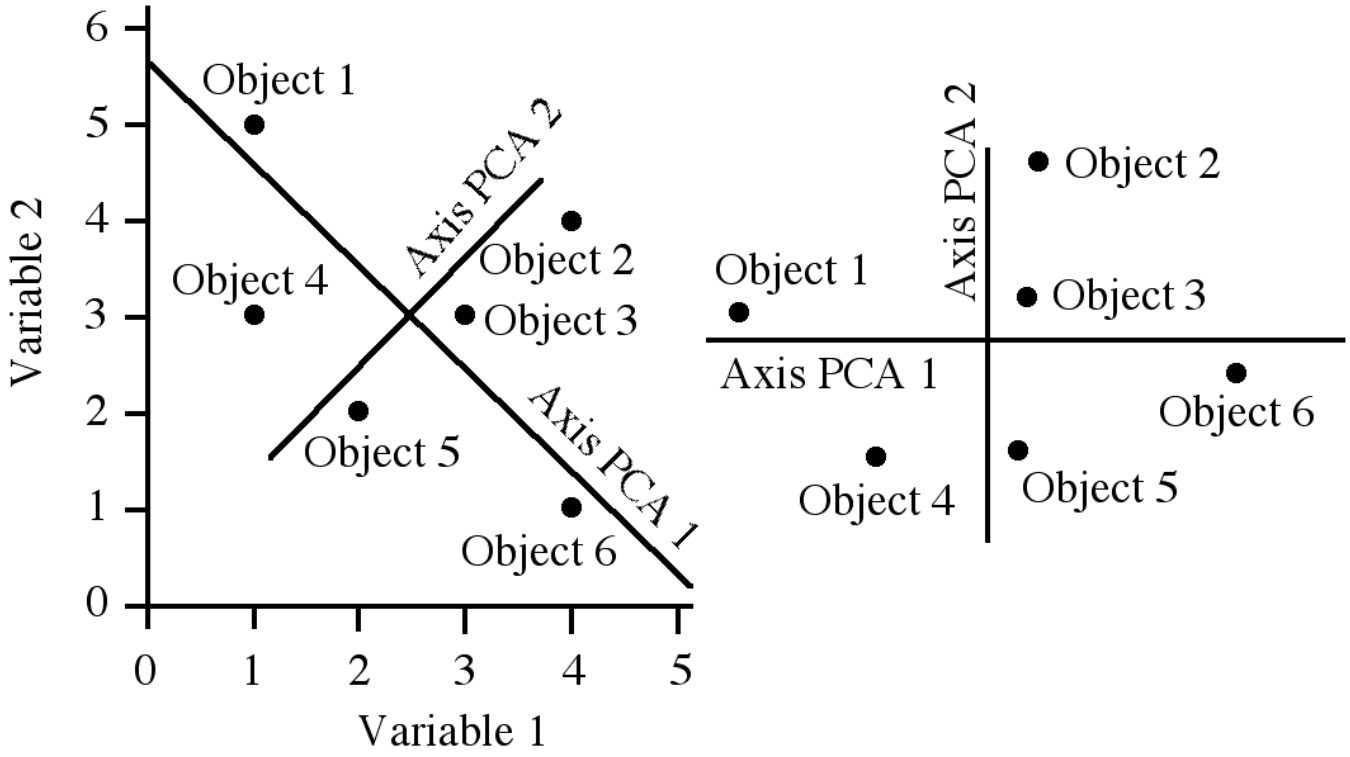

Figure 18 - PCA rotation of the 6 objets of Figure 17.

Each **principal component** is actually a **linear combination of the original variables**. Therefore, one can interpret the axes of a PCA by verifying which variable(s) contribute most to the few first principal components. One can also represent the variables on the PCA diagram representing the objects. The variables take the form of vectors (Figure 19). Note, however, that when one is specifically interested in the relationships among variables, another type of projection is preferable (see later on).

Each principal component is built on what is called an eigenvector, that has an associated **eigenvalue** λ_i. This eigenvalue gives the amount of variance that is represented on the axis. The eigenvalues always go in decreasing order, i.e. the first axis represents the largest part of the

variance, the second axis less than the first, and so on. There are as many principal components as there are variables in the original data set.

The total variation of the data is given in several programs by the total sum of squares (total SS, i.e. the variance without the division by degrees of freedom). In some programs (e.g. in Canoco) the ordination summary also presents the results with the total SS set equal to 1, so that the eigenvalues can readily be interpreted as proportions of variance: an eigenvalue of 0.705 means that the axis represents 70.5% of the total SS of the data.

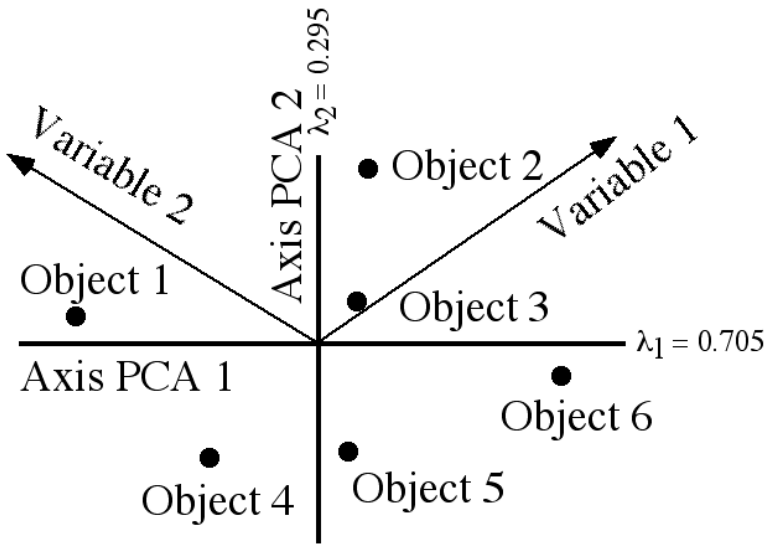

Figure 19 - PCA diagram of the data of Figures 17 and 18, with projection of the original variables. There were only two variables in the data, thus there are two PCA axes. Scaling type 1 (see below).

Technical parenthesis - According to a very important theorem in statistics called the *central limit theorem*, when a random variable results from several independent and additive effects, of which none has a dominant variance, then this variable tends towards a normal distribution even if the effects are not themselves normally distributed. This can be applied to many ecological variables. It follows that, taken together, the ecological variables tend to follow a multinormal distribution. Now, the first principal axis giving the orientation of the first principal component exposed above goes actually through the greatest dimension of the concentration ellipsoid describing the multinormal distribution. In the same way, the following principal axes (orthogonal to one another, i.e. at right angles to one another, and successively shorter) go through the following dimensions of the *p*-dimensional ellipsoid. A maximum of *p* principal axes can be derived from a data table containing *p* variables.

For practical purposes, it must be known that one can conduct a PCA and display its results in different ways. In its basic form, PCA (1) is computed on the raw (centred but otherwise untransformed) variables, and (2) the result respects the Euclidean distance among objects. One can act on these two properties, however.

Covariance or correlation? - Covariance or correlation are the association measures used to compare all the pairs of variables in PCA. Both are linear measures. One important decision to make is on which of these association matrices the PCA will be computed. This is because of the Euclidean property of the analysis: remember that Euclidean distance is very sensitive to the scales of the variables. Therefore, conducting a **PCA on the raw** (actually centred) **variables** (= PCA on a covariance matrix) is only **appropriate when these variables are dimensionally homogeneous**. Otherwise, it is advised to eliminate the effect of the differences in scale among the variables. This can be done by running a PCA on a correlation matrix, since correlation is a covariance computed on standardized variables.

Scaling - As mentioned above, both the objects and the variables can be represented on the same diagram, called a biplot. Two types of biplots can be used to represent PCA results:

• ***PCA Scaling 1 = Distance biplot:*** the eigenvectors are scaled to unit length; the main properties of the biplot are the following: (1) **Distances among objects in the biplot are approximations of their Euclidean distances in multidimensional space**. (2) Projecting an object at right angle on a descriptor approximates the position of the object along that descriptor. (3) Since descriptors have length 1 in the full-dimensional space, the length of the projection of a descriptor in the reduced space indicates how much it contributes to that space. *(4) The angles among descriptor vectors are meaningless.*

• ***PCA Scaling 2 = correlation biplot:*** the eigenvectors are scaled to the square root of their eigenvalue. The main properties of the biplot are the following: (1) *Distances among objects in the biplot are not*

approximations of their Euclidean distances in multidimensional space. (2) Projecting an object at right angle on a descriptor approximates the position of the object along that descriptor. (3a) PCA on a covariance matrix: descriptors have length s_j (= their standard deviation) in the full-dimensional space; therefore, the length of the projection of a descriptor in the reduced space is an approximation of its standard deviation. (3b) PCA on a correlation matrix: all the descriptors have unit variance (s = 1); the length of the projection of a descriptor in the reduced space reflects its contribution to that space. (4) **The angles between descriptors in the biplot reflect their correlations.**

Bottom line: if the main interest of the analysis is to interpret the relationships among objects, choose scaling 1. If the main interest focuses on the relationships among descriptors, choose scaling 2.

Equilibrium contribution circle - in one of the four options above, i.e., correlation PCA with scaling 2, it is possible to draw, on a plane made of two principal components, a circle representing the equilibrium contribution of the variables. Equilibrium contribution is the length that a descriptor-vector would have if it contributed equally to all the dimensions (principal axes) of the PCA. Variables that contribute little to a given reduced space (say, the 1×2 plane) have vectors that are shorter than the radius of the equilibrium contribution circle. Variables that contribute more have vectors whose lengths exceed the radius of that circle. The circle has a radius equal to $\sqrt{(d/p)}$, where d equals the number of dimensions of the reduced space considered (usually d=2) and p equals the total number of descriptors (and hence of principal components) in the analysis. In a covariance PCA+scaling 2, the equilibrium contribution must be computed separately for each descriptor, and is equal to $s_j\sqrt{(d/p)}$, where s_j is the standard deviation of the descriptor considered. Figure 20 shows an example.

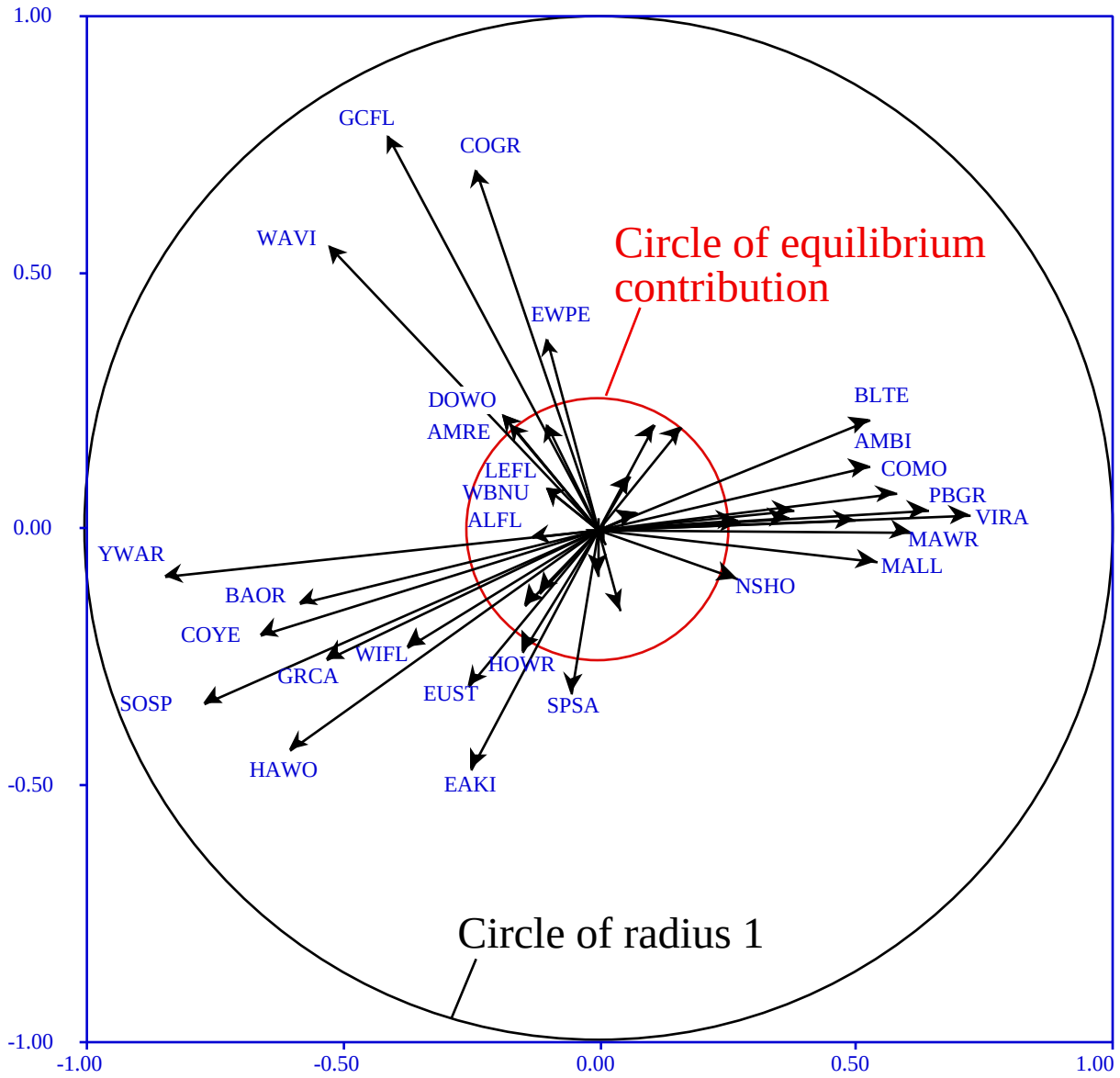

Figure 20 - PCA on a correlation matrix of Hellinger-transformed species data. Scaling type 2. Axes 1 × 2. Circle of equilibrium contribution (red). Circle of radius 1 (black): maximum length possible for a vector in a PCA on a correlation matrix.

Number of axes to interpret - *PCA is not a statistical test*, but a heuristic procedure: it aims at representing the major features of the data on a reduced number of axes (hence the often used expression "ordination in reduced space"). Usually, the user examines the eigenvalues, and decides how many axes are worth representing and

displaying on the basis of the amount of variance explained. The decision can be completely arbitrary (for instance, interpret the number of axes necessary to represent 75% of the variance of the data), or helped by one of several procedures proposed to set a limit between the axes that represent interesting features of the data and axes that merely display the remaining, essentially random variance. One of these procedures is to compute the average of all eigenvalues and interpret only the axes whose eigenvalues are larger than that average. Another is to compute a model called the *broken stick model*, which randomly divides a stick of unit length into the same number of pieces as there are PCA axes. The pieces are then put in order of decreasing length and compared to the eigenvalues. One interprets only the axes whose eigenvalues are larger than the length of the corresponding piece of stick, or, alternately, one may compare the sum of eigenvalues, from 1 to *k*, to the sum of the values from 1 to *k* predicted by the broken stick model.

Uses and misuses of PCA. Principal component analysis is a very powerful technique, but it has its limits. The main applications and limits, as well as several pitfalls, are discussed here.

The main application of PCA in ecology is the ordination of sites on the basis of *quantitative environmental variables*. This includes physical and chemical descriptors, but excludes untransformed species abundances in most cases (see Chapter 2, double zero problem).

PCA has originally been defined for data with *multinormal distributions*. In its use in ecology, however, PCA is not very sensitive to departure from multinormality, as long as the distributions are not exaggerately skewed. If some are, then the few first PCA axes will display several objects with extreme values on these variables instead of showing the main trends of the data.

PCA is computed from a matrix of dispersion (*linear* covariances or correlations). This means that the method must be applied on a

dispersion matrix among *descriptors* that are *quantitative* and for which valid estimations of covariances or correlations may be obtained. These conditions are violated in the following cases:

(1) The eigenvalues and eigenvectors of a dispersion matrix cannot be estimated using a number of objects *n* smaller or equal to the number of descriptors *p*.

(2) The data matrix must *not* be transposed, since covariances or correlations among objects are meaningless.

(3) Covariances and correlations are defined for quantitative variables. However, PCA is very robust to variations in the *precision* of data. Since a Pearson correlation coefficient on semiquantitative data is equivalent to a Spearman's ρ, a PCA on such variables yields an ordination where the relationship among variables is estimated using that measure. Furthermore, Gower (1966, *in* Legendre & Legendre 1998) has shown that, with binary descriptors, PCA positions the objects, in the multidimensional space, at distances that are the square roots of complements of simple matching coefficients S_1 (i.e., $\sqrt{1-S_1}$).

(4) With data sets with many zeros, PCA produce inadequate estimates of the site positions.

(5) Avoid the (all too common) mistake of interpreting the relationships among variables with help of the proximities of the apices (points) of the vector-arrows instead of the angles (when appropriate) on the biplots!

4.3 Pre-transformation of species data

Principal component analysis is very useful for the ordination of matrices of environmental data. On the contrary, since it is a linear method working in a Euclidean space, it is not adapted to raw species abundance data (at least when long ecological gradients underly the species data set), since the zero is treated as any other value.

However, Legendre & Gallagher (2001)[1] have shown how to overcome this problem. The trick is to *pre-transform the species data* in such a way that, after PCA, the distance respected among objects is no more the Euclidean distance, but an ecologically meaningful one, i.e. a distance that does not take the double zeros into account in the computation of resemblances between objects. These transformations can be devised to obtain any distance measure that contains a Euclidean component. The transformations proposed in that paper are devised to obtain following distance coefficients: chord distance (D_3), χ^2 metric (D_{15}), χ^2 distance (D_{16}), distance between species profiles and Hellinger distance (D_{17}). Table VI gives the transformations to apply to species data so that a Euclidean distance applied to the sites respects the distance considered.

These transformations have been devised mainly with ordination in mind, but be aware that they can be applied to any species data set upon which one wishes to apply a linear method. For instance, it has been said (Chapter 3) that the Ward clustering and *K*-means partitioning methods were linear, and therefore not adapted to *untransformed* species data. This means that these methods can be applied to species data that have been transformed as proposed by Legendre & Gallagher.

Another, very powerful possibility is to apply ANOVA-related techniques to species abundance data. An ordination-related form of multivariate ANOVA will be presented later.

[1] Legendre, P. & E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271-280.

Table VI - Pre-transformation of species abundance data to respect ecologically meaningful distances among sites when using linear analytical methods like PCA, RDA, *K*-means clustering, and so on.

Distance to be respected	**Transformation**
Chord distance (D_3)	$y'_{ij} = \frac{y_{ij}}{\sqrt{\sum_{j=1}^{p} y_{ij}^2}}$
χ^2 metric (D_{15})	$y'_{ij} = \frac{y_{ij}}{y_{i+}\sqrt{y_{+j}}}$
χ^2 distance (D_{16})	$y'_{ij} = \sqrt{y_{++}}\,\frac{y_{ij}}{y_{i+}\sqrt{y_{+j}}}$
Distance between species profiles	$y'_{ij} = \frac{y_{ij}}{y_{i+}}$
Hellinger distance (D_{17})	$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}}$

where y'_{ij} is the transformed value of the *j*-th species in the *i*-th object; y_{ij} is the raw abundance of the *j*-th species in the *i*-th object; y_{i+} is the sum of abundances of all species in the *i*-th object; y_{+j} is the sum of abundances of the *j*-th species in all objects; y_{++} is the grand total, i.e. the sum of all abundances in the raw data table.

4.4 Correspondence analysis (CA)

CA is actually a PCA on species data table that has been transformed into a table of Pearson's χ^2 statistic. The raw data are first transformed into profiles of conditional probabilities weighted by the row and column sums, and the resulting table is submitted to a PCA. The result is an ordination where it is the χ^2 distance (D_{16}) that is preserved among sites instead of the Euclidean distance D_1. The χ^2 distance does not consider the double zeros. Therefore, CA is a method adapted to the analysis of species abundance data (without pre-transformation).

Note that the data submitted to a CA must be dimensionally homogeneous and equal to 0 or positive (which is the case of species counts or presence-absence data).

For technical reasons not developed here, CA ordination produces one axis less than min[*n*,*p*]. As in PCA, the orthogonal axes are ranked in decreasing order of variation represented, but instead of the total SS of the data, the variation is measured as a quantity called the *total inertia*. Individual eigenvalues are always smaller than 1. To know the amount of variation represented on an axis, one must divide the eigenvalue of this axis by the total inertia of the species data matrix.

In CA, both the objects and the species are generally represented as points on the same **joint plot**. Two **scalings** are most useful in ecology. They are explained here for data matrices where objects are rows and species are columns:

• ***CA scaling type 1:*** rows are at the centroids of columns. This scaling is the most appropriate if one is primarily interested in the ordination of **objects** (sites). In the multidimensional space, χ^2 distance is preserved among objects. See Figure 20 below. Interpretation:

(1) The distances among objects in the reduced space approximate their χ^2 distance. Thus, object points that are close to one another are likely to be relatively similar in their species relative frequencies.

(2) Any object found near the point representing a species is likely to have a high contribution of that species. For presence-absence data, the object is more likely to possess the state "1" for that species.

• ***CA scaling type 2:*** columns are at the centroids of rows. This scaling is the most appropriate if one is primarily interested in the ordination of **species**. In the multidimensional space, χ^2 distance is preserved among species. Interpretation:

(1) The distances among species in the reduced space approximate their χ^2 distance. Thus, species points that are close to one another are likely to have relatively similar relative frequencies in the objects.

(2) Any species that lies close to the point representing an object is more likely to be found in that object, or to have a higher frequency there than in objects that are further away in the joint plot.

The following example (Table VII) will be submitted to a correspondence analysis:

Table VII - Artificial data for CA

	Spec.1	Spec.2	Spec.3
Obj.1	1	5	2
Obj.2	4	4	6
Obj.3	3	3	0
Obj.4	1	3	5
Obj.5	2	2	4
Obj.6	4	1	0

Since there are 6 objects and 3 species, the number of CA axes will be min(6,3)–1 = 2.

Using scaling 1, one obtains the following joint plot (Figure 21):

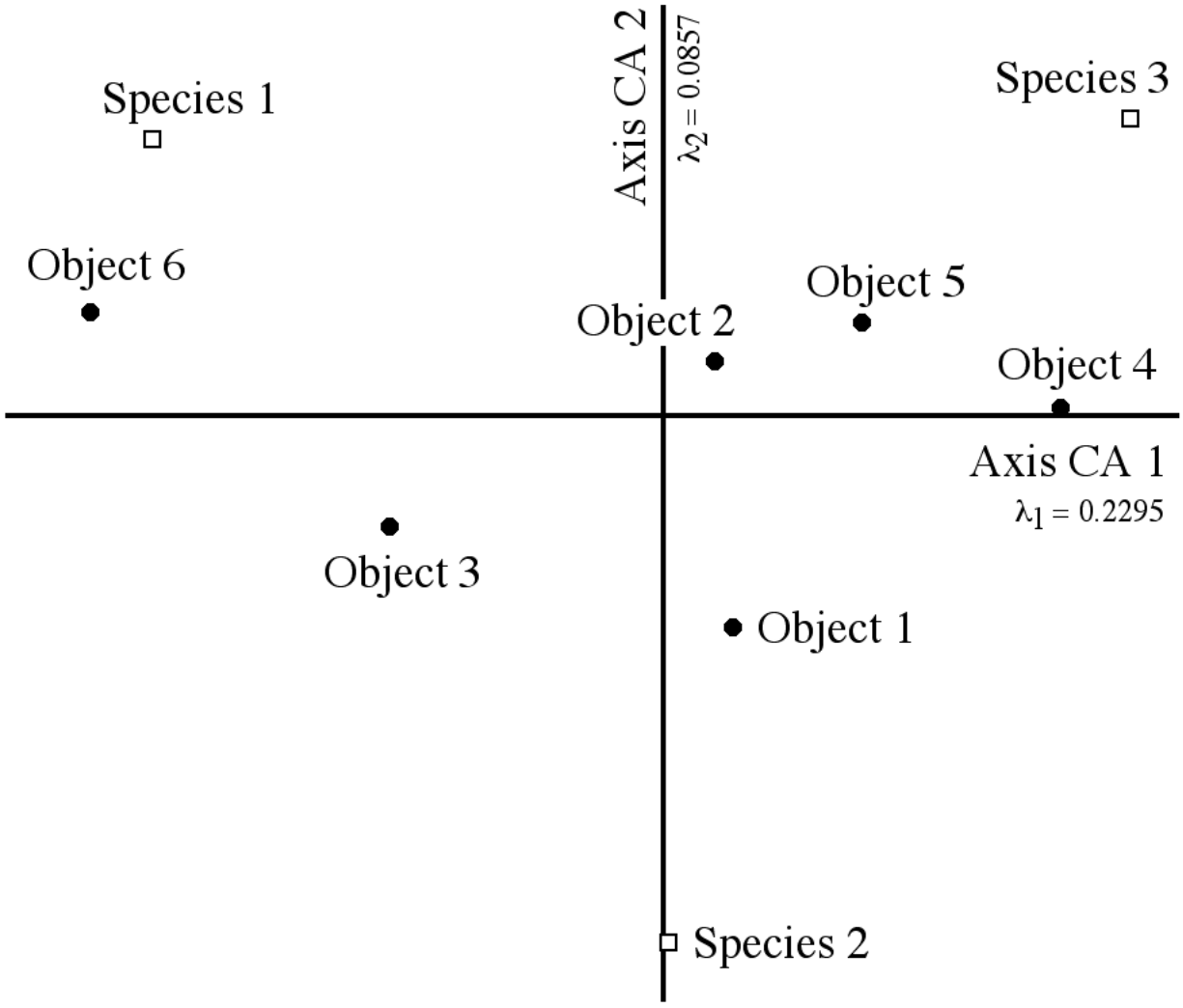

Figure 21 - 1 × 2 plane of the CA of the data shown in Table VII. Scaling type 1.

Note that in scaling 1, the objects are at the barycentre (centre of mass) of the variables. Note, also, that the variables are displayed as points (contrary to PCA).

In this example, the eigenvalues are equal to 0.2295 and 0.0857. Since there are only two axes, the total inertia (sum of all eigenvalues) equals 0.2295+0.0857=0.3152, which, for each axis, amounts to following proportions of variation:

0.2295/0.3152 = 72.8% for axis 1

0.0857/0.3152 = 27.2 % for axis 2

If scaling type 2 was used, the biplot would be the following (Fig. 22):

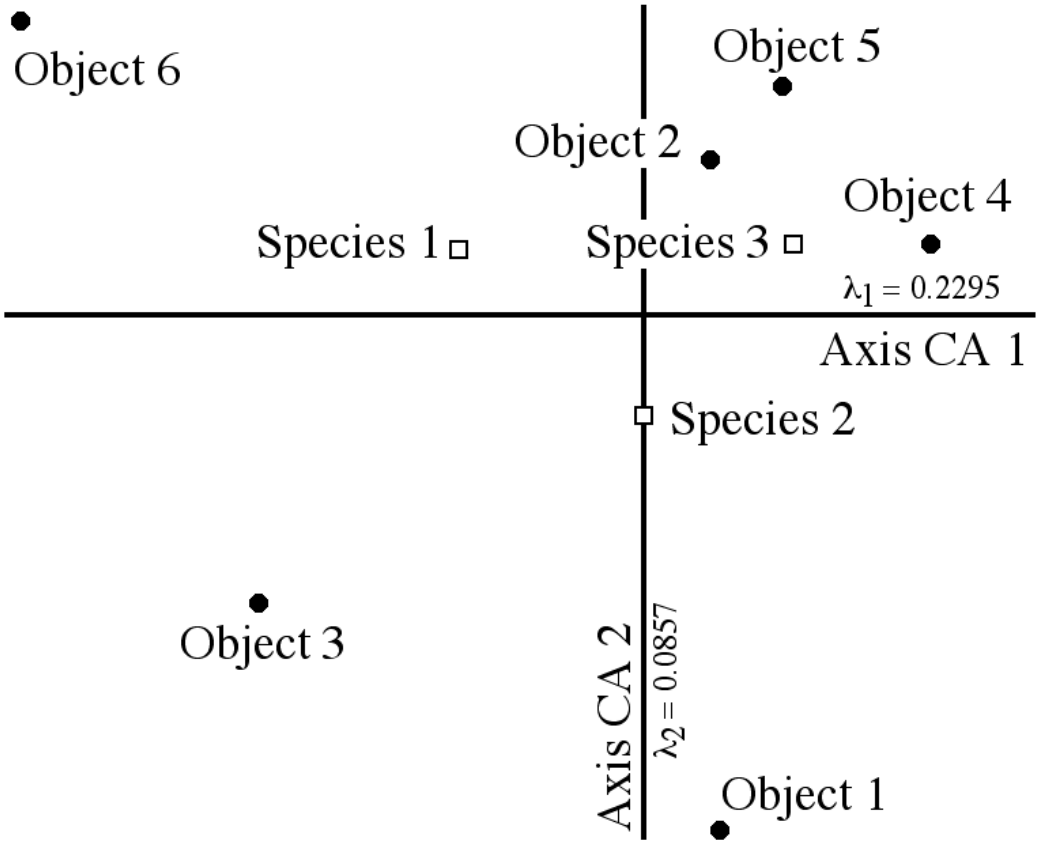

Figure 22 - 1 × 2 plane of the CA of the data shown in Table VII. Scaling type 2.

Note that in scaling 2 the species are at the barycentre (centre of mass) of sites.

Words of caution

Correspondence analysis has first been described to analyse contingency tables. Therefore, it tends to overemphasise extreme values. In certain cases, it seems to be very sensitive to rare species, which tend to be located at extreme positions in the ordination diagram. The very rare species, however, do not much influence the overall ranking and structure of the main ordination axes; only the diagrams are affected badly. Therefore, although it may be advisable to eliminate the rarest species from the data table if there are several of them, it is not necessary to do so. If the data set is full of them, anyway, the use of CA is questionable.

Arch and horseshoe effects - Long environmental gradients often support a succession of species (Figure 23). Since the species that are controlled by environmental factors tend to have unimodal distributions, a long gradient may encompass sites that, at both ends of the gradient, have no species in common; thus, their distance reaches a maximum value (or their similarity is 0). But if one looks at either end of the succession, the sites still represent a continuation of the ecological succession, so contiguous sites continue to grow more different from each other. Therefore, instead of a linear trend, the gradient is represented on a pair of CA axes as an **arch** (Figure 24A). Several *detrending* techniques have been proposed to counter this effect, leading to ***detrended correspondence analysis (DCA)***:

- detrending by segments: axis I is divided into a number of segments, and, within each one, the mean of the object scores along axis 2 is made equal to zero. This methods has been strongly rejected by many authors. Actually, the scores on the second axis are essentially meaningless;

- detrending by polynomials: another line of reasoning about the origin of the arch effect leads to the observation that when an arch occurs, the second axis can be seen as quadratically related to the first (i.e. it is a second-order polynomial of the first). This makes up for the

parabolic shape of the scatter of points. Hence, a solution is to make the second axis not only linearly, but also quadratically independent from the first. Although intuitively attractive, this method of detrending has to be applied with caution because it actually imposes a more constraining model on the data.

Note that the arch-like pattern is even stronger in PCA. There the extreme sites tend to be actually *closer* to one another as the number of nonoverlapping species increases, because the double zeros implied are considered in the Euclidean space as a resemblance between the sites. Thus, the extreme sites become closer as the number of double zeros increases. One can clearly see that this is an ecological nonsense. This pattern is called the **horseshoe effect (Figure 24B)**, because the extremities of the arch bend inwards.

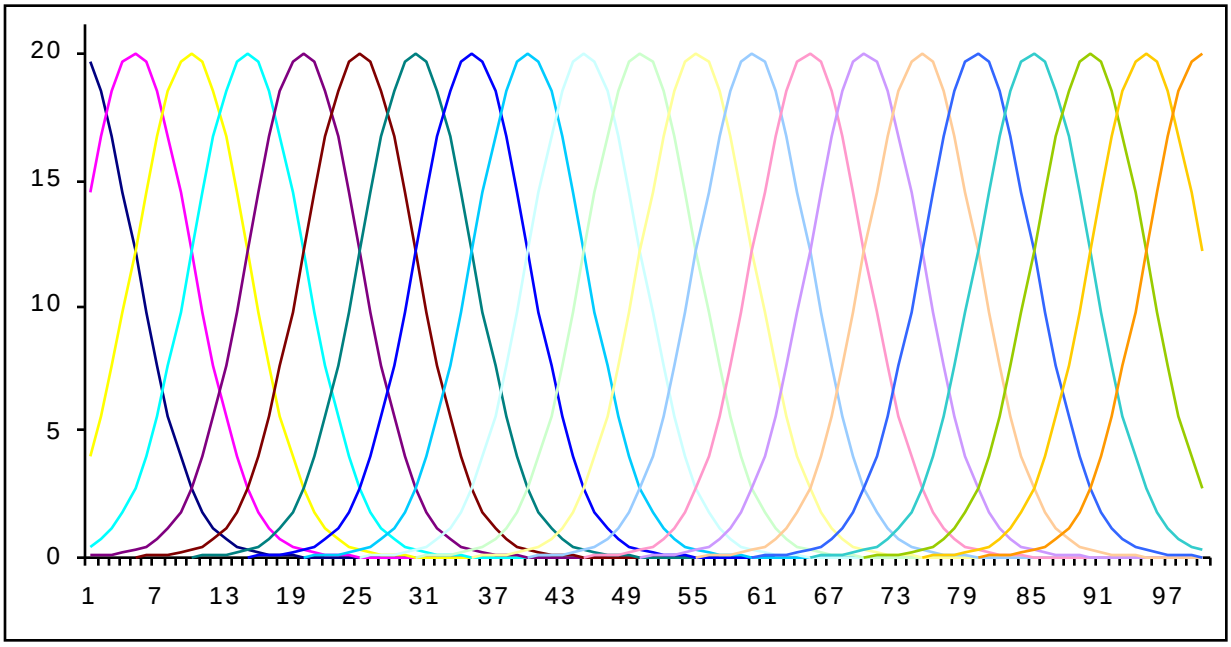

Figure 23 - Succession of species along an ideal gradient (species packing model).

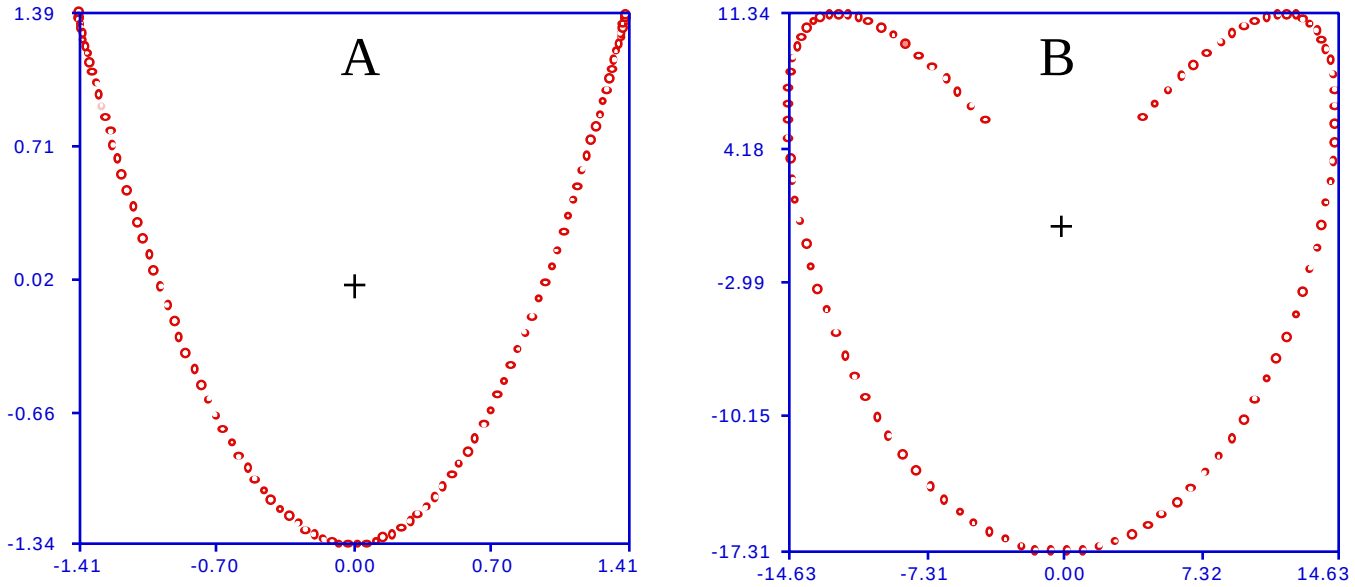

Figure 24 - CA and PCA on the data of Figure 23. 1×2 plane, objects, CA and PCA type 1 scalings.

CA has long been the "queen" of ordination methods for species abundance data. It can be applied to presence-absence data as well. However, since the Legendre & Gallagher transformations have been proposed to overcome the "barrier" between the species data and the linear methods, PCA and its constrained version RDA (see further down) regain interest among ecologists.

4.5 Principal coordinate analysis (PCoA)

PCA as well as CA impose the distance preserved among objects: Euclidean distance (and several others with pre-transformations) for PCA and χ^2 distance for CA. But if one would like to ordinate objects on the basis of yet another distance measure, more appropriate to the problem at hand, then PCoA is the method to apply. It allows to obtain a Euclidean representation of a set of objects whose relationships are

measured by any similarity or distance coefficient chosen by the user. For example, if the coefficient is S_{16}, which can combine descriptors of many mathematical types into a single measure of resemblance, then the ordination will represent the relationships among the objects based upon these many different variables. This would not be possible with PCA or CA.

Like PCA and CA, PCoA produces a set of orthogonal axes whose importance is measured by eigenvalues. Since it is based on an association matrix, it can represent the relationships either among objects (if the association matrix was in Q mode) or variables (if the association matrix was in R mode). If one wants both at the same time, one can compute species scores after a site ordination, on the basis of the site scores (using weighted averages or correlations).

In the case of Euclidean association measures, PCoA behaves in a Euclidean manner. For instance, computing a Euclidean distance among sites and running a PCoA will yield the same results as running a PCA on a covariance matrix and scaling 1 on the same data.

But if the association coefficient used is nonmetric, semimetric or has other problems of "non-Euclideanarity", then PCoA will react by producing several *negative* eigenvalues in addition to the positive ones (an a null one in-between). The negative eigenvalues can be seen as the representation of the non-Euclidean part of the structure of the association matrix and it is, of course, not representable on "real" ordination axes. In most cases this does not affect the representation of the objects on the several first principal axes, but in several applications this can lead to problems. There are technical solutions to this problem (e.g. Lingoes or Caillez correction), but they are not always recommendable, and go beyond the scope of this introduction.

The ordination axes of a PCoA can be interpreted like those of a CA: proximity of objects represent similarity in the sense of the association measure used (Figure 25).

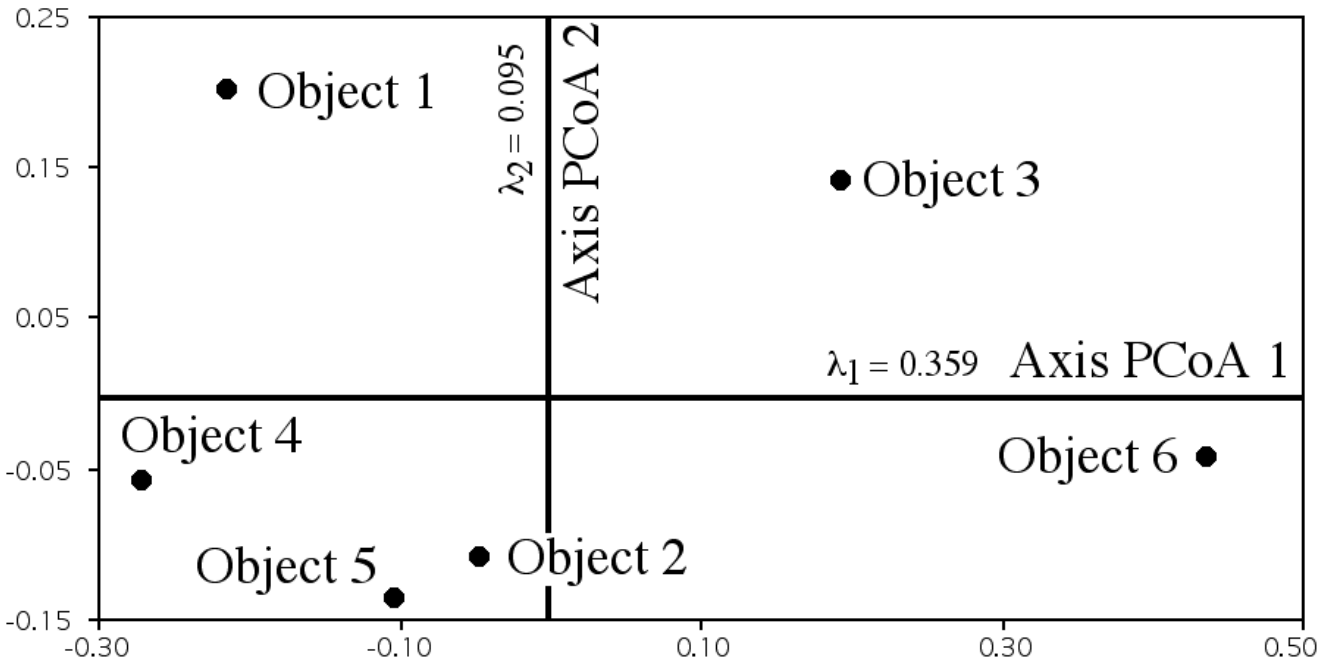

Figure 25 - 1 × 2 axis of a PCoA on a D_{14} (Bray-Curtis) matrix of the data of Table VII. The two first axes have eigenvalues of 0.359 and 0.095; they represent 59.5% and 20.2% variation respectively. This PCoA gives 4 positive, one zero and one negative eigenvalue.

PCoA is sometimes used as an intermediate, technical step in more complex analyses. For instance, if a method requires the objects to be represented by a set of quantitative and orthogonal variables, and the raw variables are of various mathematical types and non-orthogonal, then one can compute a PCoA of these data using a distance measure appropriate for the raw variables, and then use the PCoA axes as new, quantitative and orthogonal descriptors. PCoA is also involved in the construction of PCNM spatial descriptors (see Chapter 6).

4.6 Nonmetric multidimensional scaling (NMDS or MDS)

If the user's priority is not to preserve the exact distances among objects, but rather to represent as well as possible the **ordering relationships** among objects in a **small and specified number of**

axes, then NMDS may be the solution. Like PCoA, NMDS is not limited to Euclidean distance matrices. It can produce ordinations of objects from any distance matrix. The method can also proceed with missing distance estimates, as long as there are enough measures left to position an object with respect to a few others.

NMDS is *not* an eigenvalue technique, and it does *not* maximise the variability associated with individual axes of the ordination. As a result, plots may arbitrarily be rotated, centred, or inverted. The procedure goes as follows (very schematically; for details see Legendre & Legendre p. 445sq.):

1. Specify the number *m* of axes (dimensions) desired.

2. Construct an initial configuration of the objects in the *m* dimensions, to be used as a starting point of an iterative adjustment process. This is a tricky step, since the end-result may depend on the starting configuration. A PCoA ordination may be a good starting point.

3. An iterative procedure tries to position the objects in the desired number of dimensions in such a way as to minimize a *stress function* (scaled from 0 to 1), which measures how far the reduced-space configuration is from being monotonic to the original distances in the association matrix.

4. The adjustment goes on until the stress value can no more be diminished, or it attains a predefined value (tolerated lack-of-fit).

5. Most NMDS programs rotate the final solution using PCA for easier interpretation.

For a given and small number of axes (e.g. 2 or 3), NMDS often achieves a less deformed representation of the relationships among objects than a PCoA can show on the same number of axes. But NMDS remains a computer-intensive solution, exposed to the risk of suboptimal solutions in the iterative process (because the objective function to minimize has reached a *local minimum*).