

Chapter 4. Techniques of factor analysis of vegetation data

by Guy BOUXIN[°]

Contents

Introduction	2
The main techniques of factor analysis and ordination	4
Techniques of factor analysis	5
Factor analysis <i>s.s.</i> (FA).....	7
Principal component analysis (PCA)	7
Correspondence analysis (CA).....	8
Detrended correspondence analysis (DCA)	12
Principal coordinate analysis (PCoA ou NMS).....	12
The constrained ordinations	12
Multiple factor analysis.....	14
Co-inertie analysis (MCoA).....	17
Other techniques	17
Bray & Curtis ordination (Polar ordination)	17
Non-metric multidimensional scaling, NMDS	18
Use of techniques	19
Several conclusions	24
References	24
Annexe: calculation of eigenvalues and eigenvectors	32
Used data	32
Examples of principal component analysis.....	32
Matrix of centered reduced data.....	32
Correlation matrix	33
Eigenvalues	33

[°] rue des Sorbiers 33 à B. 5101 Erpent mail : guy.bouxin@proximus.be

Eigenvectors.....	33
Example of calculation	33
Example of correspondence analysis	34
Weighted matrix	34
Covariance matrix.....	35
Eigenvalues.....	35
Eigenvectors.....	35
Example of calculation	35
Example of non-symmetric correspondence analysis.....	36
Weighted matrix	36
Eigenvalues.....	37
Eigenvectors.....	37

Introduction

The analysis of the data, *sensu* BOUROUCHE & SAPORTA (1980), groups numerous and very different methods of statistical analysis. We distinguish two main approaches: automatic classification, which consists in classifying statistical units and variables using previously established algorithms, and factor analysis, which uses the properties of Euclidean vector spaces to describe individuals and variables (FOUCARD, 1982). According to PALM (1993), factor analysis methods are fundamental tools for analyzing data tables that have no particular structure. They have essentially a descriptive purpose by condensing the information contained in a table, often consisting of a large number of rows and columns, into a few two-dimensional graphical representations, accompanied by tables containing the numerical values of characteristics intended to help the user during the interpretation.

These analytical techniques therefore seem well adapted to the analysis of the vegetation relevé tables. They have been used for several decades in the study of vegetation and are referred to as "ordination". Strictly speaking, ordination seeks to place vegetation in relation to one or more environmental gradients associated with the axes of analysis (GOODALL, 1954). It starts from the concept that vegetation is a *continuum*. It consists in placing each relevé in relation to one or more axes in such a way that its relative position, with respect to the axes, provides maximum information about its composition. In ordination techniques, the synthetic approach is not always considered as an objective in itself, but becomes a tool to put the floristic data in relation with the environmental factors (or others) likely to affect the composition of the vegetation (GREIG-SMITH, 1964). Ordination, following WHITTAKER (1973), is generally considered as a technique

of indirect gradient analysis; in this respect the author wrote: "Gradient analysis studies the vegetation in terms of gradients (or more or less continuous variations) of the environment, specific populations and communities". The arrangement of samples according to environmental gradients as the basis of the analysis is called ordination. The ordination is either direct (arrangement of the samples according to their position along one or more environmental gradients) or indirect (from the comparison of the samples, and after elaborate calculations, the samples are arranged along abstract or directions of community variation). These abstract axes may or may not correspond to environmental gradients. A gradient in the composition of vegetation can be characterized by the amount of turnover from one end to the other, which WHITTAKER (1967) called "diversity beta". According to PALMER (available at: <http://ordination.okstate.edu/glossary.htm>), beta diversity is a measure of how different samples are from each other and / or how apart there are on gradients of species. Alternatively, it is a measure of the length of an ecological gradient or an ordination axis. An axis or gradient with high beta diversity has completely different species compositions (i.e. share no species) at opposite ends. On the contrary, an axis or gradient with low beta diversity has specific compositions quite similar at its ends.

According to GAUCH (1982b), the data of a sample-by-species matrix can be conceptualized as a *species space* in which the species are axes of multidimensional space and the samples are points located by their abundances for each species. Samples with similar species composition occupy nearby positions in species space. There is a converse *samples space* in which samples are axes and species are points placed along the axes according to the species importance in each sample. A dissimilarity matrix likewise may be conceived geometrically. *Samples dissimilarity space* uses sample dissimilarities as axes and samples as points. The corresponding *species dissimilarity space* uses species dissimilarities as axes and species as points. The fifth and final conceptual space is *ecological space*, having environmental gradients as axes. Species space is defined by a large number of dimensions (one per species). The sample space can be easily reduced in its number of dimensions as the species of a table often have more or less redundant dispersions. In contrast, the ecological space is represented by a small number of dimensions. Ecological space is the expected result of ordination.

Ordination, in its broad sense, has three objectives (GAUCH, 1982b): (1) summarizing community data, (2) relating community variation to environmental gradients and (3) understanding community structure. The classical analysis involves two steps: (1) producing a synthesis of the structure of the community, bringing it back to a space typically comprising one to three dimensions; (2) compare the structure thus revealed with the information available on the environment.

However, according to AUSTIN (1976), the properties of indirect ordination to organize and allow the interpretation of plant communities in relation to environmental gradients are disrupted by the nonlinearity of vegetation / environment relations.

The concept of ordination is also developed by WILDI (2013), in relation with several multivariate analyse techniques. The notions of arch effect, gradient and non-linearity in the response of vegetation to environmental factors dominate the author's scientific thinking. In chapter 6, we explain that those notions are false problems.

Factorial analysis in the statistical sense and ordination therefore have very different conceptual bases since factor analysis, in a spirit of data analysis, has no preconceived idea as to the meaning of the axes resulting from the analysis. The manner to exploit the results is greatly changed.

Several multivariate analysis techniques exist for both direct constrained ordination and indirect gradient analysis.

Some confusion in the use of the techniques presented comes from the fact that the same techniques are used either as data analysis or as an ordination technique. Through the presentation of examples, we try to show the respective contribution of each of the two ways of using these techniques. If ordination is very often used as an indirect gradient analysis technique, the reader will soon understand that we favor a much more statistical approach. We use factor analysis as one of the tools leading to a data analysis as extensive as possible. We caution the reader against the automatic and unthinking use of statistical techniques, often and probably unconsciously used to reinforce preconceived notions of vegetation or species-environment relationships. The softwares are numerous, easily accessible and too often misused.

In this chapter, we succinctly present the most common factorial and ordination techniques. The applications are described in the following chapters. A few brief conclusions are finally drawn.

The main techniques of factor analysis and ordination

Our intention is not to review all the methods developed since the 1950s but to present the main ones, to show the spirit in which they were used and above all to make us think about how to use them from now on.

We begin by recalling some basic notions of factor analysis, while advising readers to learn more about the subject.

From a historical point of view, let us simply mention the technique called "weighted average ordination (WA)" illustrated by WHITTAKER (1960). A coordinate is calculated for each relevé, based on the species composition and the weighting assigned to each species. It is effective if there is a predominantly simple gradient and if the distribution of species along an environmental or other gradient is known (*e.g.* position of a species along a pioneer-climatic gradient). With DEL MORAL & WATSON (1978), the distribution of species along a gradient is analysed by a reciprocal averaging ordination. WA was also used by OLSVIG (1979). This technique can be used with the "scores" attributed by ELLENBERG (1974) to many species (light figure, temperature figure, continentality figure, moisture figure, pH figure, nitrogen figure, salt content). However, these data should be handled with great caution (WAMELINK *et al.*, 2002).

The techniques presented now all belong to multivariate analysis techniques.

Techniques of factor analysis

This paragraph is devoted to factor analysis techniques. Both techniques are based on matrix computation. Most of the elements of this paragraph are taken from PALM (1993).

Let a matrix X of dimensions $n \times p$, such that $p \leq n$, containing arbitrary numbers. Consider first the square matrix $\mathbf{X}'\mathbf{X}$, of dimensions $p \times p$ and of rank r ($r \leq p$). This matrix admits r positive eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_r,$$

to which are associated r eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$.

Consider now the square matrix $\mathbf{X}\mathbf{X}'$, of dimensions $n \times n$. This matrix is also of rank r and has r positive eigenvalues:

$$\mu_1 \geq \mu_2 \geq \dots \mu_r,$$

with which r are eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$.

The eigenvectors have an arbitrary sign related to the algorithm of calculation used and the sign of all the elements of an eigenvector can be changed.

The graphical representation of the results often plays an important role in the interpretation of the results. Consider that one represents the matrix X by n line vectors in the space of the p columns or by p vectors-columns in the space of the n lines. From a geometrical point of view, this corresponds to the replacement of the space with p (or n) dimensions by a q -dimensional subspace and considering that the n (or p) points lie exactly in this space. The quality of this approximation is given by the ratio:

$$\frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^r \lambda_k}$$

which measures the ratio between the sum of the squares of the distances of the projected points in the q -dimensional space and the sum of the squares of the distances of points in the space at p (or n) dimensions. However, this report gives only a general idea of the quality of representation for all the points, and there may be important differences from one point to another, for example, a point which may be very close of the q -dimensional subspace and the other on the contrary being able to be further away from this subspace. To calculate the quality of the representation of a given point, one calculates, for example, the ratio between the square of the distance at the origin of the projection of this point in the q -dimensional subspace and the square of the distance at the origin of this point in the space with p or with n dimensions. The quality of the approximation of a point by a given axis in fact measures the importance of the angle formed by the segment linking the origin of the axes with the given point and the projection of this segment on the axis in question. More precisely, this is the square of the cosine of this angle. A similar geometric interpretation is given to the quality of the approximation of a point by a set of axes: this is the square of the cosine of the angle formed by the segment connecting the origin of the axes to the given point and the projection of this segment in the space of the considered axes. This value is, moreover, equal to the sum of the squares of the cosines of the projection on each of the axes.

A schematic representation of the quality of the approximations is given, in the case of two axes, in figure 1: the quality of the approximation of the point P by the axis 1 is $\cos^2\alpha_1$; the quality of the approximation of this point by the axis 2 is equal to $\cos^2\alpha_2$ and the quality of the approximation of the point P by the plane formed by the axes 1 and 2 is equal to:

$$\cos^2\alpha = \cos^2\alpha_1 + \cos^2\alpha_2.$$

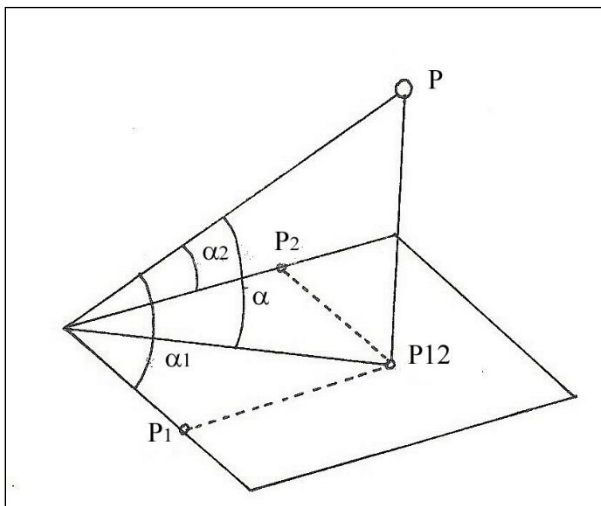


Figure 1. Quality of the approximation of a point by an axis or a plane (P_1 = projection of the point P on the axis 1, P_2 = projection of the point P on the axis 2, P_{12} = projection of the point P on the plane formed by the axes 1 and 2). Redrawn from PALM (1993).

The quality of the approximation of a point is often called "square cosine", because of the geometric interpretation that has just been given, or "contribution of an axis to the explanation of a point" or "relative contribution of an axis to the position of a point". In the examples, we will retain "relative contribution".

Factor analysis s.s. (FA)

The main elements of the summary of the article by DAGNELIE (1973) are reproduced here.

Factor analysis consists of a set of methods developed in experimental psychology and used mainly in the statistical interpretation of the results of psychological tests. The main notions of factor analysis are introduced from a simple example borrowed from the psychological field: the results of three tests (problems of reasoning, problems of arithmetic and addition calculations), to which different people can be subjected, are related to two basic abilities of these people (reasonableness and numeracy). In such a situation, the aim is to explain the correlations observed between the test results, highlighting various underlying fundamental attitudes and identifying as far as possible these abilities.

The transposition of the factor analysis into phytosociology is carried out by considering that plant species take the place of tests, that vegetation relevés replace people and that ecological factors are substituted for fundamental skills. From the observed correlations between species, the objective is to highlight and identify as much as possible the main underlying ecological factors. Alternatively, other problems may be addressed, such as studying the differential power of species, defining sociological groups of closely related species, establishing a floristic classification of relevés, comparing two or several classifications of the relevés and the study of the distribution of species and groups of species in the set of the studied relevés.

The first application of factor analysis in phytosociology is due to GOODALL (1954). The contribution of factor analysis to the study of plant communities was developed by DAGNELIE (1960). This technique has gone out of fashion, under the roller of correspondence analysis.

We must distinguish analytical techniques from a single table, such as principal component analysis, correspondence analysis (see CHESSEL, DUFOUR & THIOULOUSE, 2004) and analyses of several tables (see DRAY, DUFOUR & CHESSEL, 2007).

Principal component analysis (PCA)

Consider now that we have the observations relative to p variables and carried out on n individuals. These data are in the form of a matrix, Y , of dimension $n \times p$. The p variables correspond to the species (coefficient of abundance-dominance, biomass for example) or factors of the environment. The initial

variables are replaced by the reduced centered variables. The matrix $X'X$ then corresponds to the correlation matrix. In the phytosociological tables, the rows correspond to the species (variables) and the columns to the relevés (individuals).

The non-centered principal component analysis simultaneously orders the sites and the species. The weight of sites and species in the analysis is proportional to their total abundance. This technique is not susceptible to distortion by rare species. Examples are given by EZCURRA (1987) and EZCURRA *et al.* (1987).

When species have a unimodal distribution along environmental gradients, PCA produces representations of the first two axes of the factorial analysis, that forms an arc or a horseshoe, which is considered a distortion by GAUCH (1982b) *inter alia*. PCA, as an ordination, is considered to be the most effective when species present monotonous responses along environmental gradients. This way of interpreting the results shows all the difference that there can be between an ordination and an analysis of data that does not involve any element outside the data table.

The principal component analysis introduces complementary variables, which are not taken into account in the calculations (FOUCARD, 1982). For this purpose it is sufficient to calculate the covariances or correlations with the main components and to represent them in the basis of the main components already obtained. HUSSON *et al.* (2009) distinguish active variables, which are included in PCA calculations and additional or illustrative variables.

It is also possible to introduce complementary individuals.

Correspondence analysis (CA)

The starting point of **correspondence analysis** is a two-input frequency table. The frequencies constitute the elements of a matrix Y , of dimensions $n \times p$, n and p representing the numbers of modalities relating to the two criteria taken into consideration. The rows and columns of matrix Y are of the same nature, unlike the matrix of data for a principal component analysis, where rows correspond to variables and columns to individuals. Once again, one must repeat that CA is part of a set of statistical techniques whose propose is to approach the tables independently of any mathematical hypothesis. They should therefore not be used to answer hypotheses or to bring a set of data into a paradigm such as the indirect gradient.

However, this analysis is applicable to the standard relevé tables, with the rows corresponding to the species and the columns to the relevés. It is very often used with coefficients of abundance-dominance or presence data. It is not suitable for biomass or continuous variables of the environment at all.

The matrix of frequencies (or presences or abundances-dominances) Y is first transformed into a matrix of relative frequencies, F , obtained by dividing each element of Y by the sum of all the elements :

$$f_{ij} = y_{ij}/N$$

$$(i = 1, \dots, n; j = 1, \dots, p)$$

This matrix F is, in turn, transformed into a matrix X by the following transformation:

$$x_{ij} = (f_{ij} - f_{i.}f_{.j})/\sqrt{f_{i.}f_{.j}},$$

$f_{i.}$ et $f_{.j}$ being the marginal relative frequencies of rows and columns :

$$f_{i.} = \sum_{j=1}^p f_{ij}$$

$$f_{.j} = \sum_{i=1}^n f_{ij}$$

In this transformation, rows and columns play a symmetric role. We also note that the deviations:

$$f_{ij} - f_{i.}f_{.j}$$

are in fact the deviations between the observed relative frequencies observed and the expected relative frequencies under the assumption of independence between the rows and the columns. On the other hand, except to the constant $1/\sqrt{N}$, the quantities x_{ij} are equal to the square roots of the quantities calculated for each of the np cells of the contingency table and sum to obtain the value χ_{obs}^2 when carrying out the test χ_2^2 of independence between rows and columns.

After calculating the eigenvalues and the eigenvectors, the representation of the line-points and that of the column-points in spaces with two or more dimensions are important but not unique tools of the interpretation of the starting data. In these graphs, the proximity of two line points or two column points reflects the similarity of the "profiles", that is to say, the conditional distributions, relative to these two lines or to the two columns. However, as with principal component analysis, it should be borne in mind that two points (rows or columns) may be close in a subspace (on an axis or on a factor plane) without being necessarily closed in the complete space, that is, the space of the set of factors. We cannot therefore conclude from the similarity of the profiles only if the corresponding points are well represented, that is to say if the sum of the square cosines for the subspace is sufficiently large.

In practice, line points and point-columns are first identified which have a strong contribution to the factors used for graphic representation and which, at the same time, have a satisfactory representation quality. For these points, one examines the projections on the axes and more particularly the sign of these projections, so as to highlight any conjunctions or oppositions (line points or column points with projections of the same sign or opposite signs). Examination of the initial data or, more precisely, the comparison of the profiles of

the rows and columns with the average profile of the rows and columns also makes it possible to verify these conjunctions or oppositions. If we use the algorithm of HILL (1973), the analysis is called "Reciprocal averaging (or RA)".

According to GAUCH (1982b) and many authors thereafter, CA proved superior to principal corresponding analysis analysis. CA presents, as PCA the problem of arc configuration but with PCA, the ends of the arc sometimes meet. When treating long gradients of vegetation, with high beta diversity, CA is generally superior.

Non-symmetric correspondence analysis (NSCA) (GIMARET-CARPENTIER *et al.*, 1998, KROONENBERG & LOMBARDO, 1999) starts from the observation of how the data are collected in the field. Species occurrences are observed among relevés. The relevés (number, position, for example) are chosen by the observer, not the species. The result is a fundamental asymmetry that is explored by NSCA. The difference between CA and NSCA is precisely the metric. Species have a uniform weight equal to unity in NSCA whereas CA is based on the metric which implies a double "averaging" calculated from the marginal totals of the rows and columns of the tables. This latter technique is thus called dual. In NSCA, high importance is given to abundant species and the effect of rare species is severely limited. In a table of occurrences, therefore, a large number of rare taxa will not be a limiting factor. A very low weight automatically returns a taxon to the origin.

To clarify the difference between CA and NSCA :

In the data transformation,

- For CA, one calculates : $don1(I, J) = don(I, J) * \sum / (hc(J) * hr(I)) - 1$,
- For NSCA, one calculates : $don1(I, J) = don(I, J) / hc(J) - hr(I) / \sum$.

The covariance matrix is then calculated as follows :

- for CA, $y1(J1, J2) = y1(J1, J2) + don1(I, J1) * don1(I, J2) * hr(I) * Sqr(hc(J1) * hc(J2))$
- for NSCA, $y1(J1, J2) = y1(J1, J2) + don1(I, J1) * don1(I, J2) * Sqr(hc(J1) * hc(J2))$.

With don() : original data (i.e., presence of a species in a relevé), don1() : transformed data before calculation of the covariance matrix, sum : sum on all the lines and columns, hc() : sum on one column, hr() : sum on one line (or raw), y1() : covariance.

The eigenvalues are then calculated from that covariance matrix.

Another variant of CA called "**correspondence analysis of lines of data**" was proposed by GREENACRE (2010). It starts from the observation that vegetation or wildlife relevés have the same area or volume and asks whether abundances should be transformed according to the total abundance of each relevé. This CA analyses non-relativized lines.

Multiple correspondence analysis (MCA) (PALM, 1993) is a CA applied to complete disjunctive tables; the two-criteria contingency table is replaced by a multi-criteria table. The typical example is the analysis of a table with individuals having to answer at a set of multiple-choice questions. An example is provided by COQUILLARD & GUEUGNOT (1991); it is an ecological study of 27 relevés whose belonging to four syntaxons is known and characterized by 12 variables. Each variable is represented by two or more modalities (for example, the topographic situation is defined by four modalities: front of hill, summit of slope, middle of slope, bottom of slope). Such a technique, little used, offers interesting prospects when analyzing non-monotonic variables; species-related variables such as abundance-dominance could also be coded into categories. Applications are rare (FRIED *et al.*, 2009).

The analysis called "fuzzy correspondence analysis" also associates several modalities with the variables (DOLÉDEC & CHEVENET, 1997). It is considered a form of MCA. It uses a "fuzzy coded" table with t species constituting the rows and m modalities constituting the columns. V is the number of variables. The variable 1 has $m(1)$ modalities, the variable 2 has $m(2)$ modalities, ..., and the variable v a $m(v)$ modalities.

$$m = \sum_{j=1}^v m(j)$$

the intersection of the i^e species with the k^e modality of the j^e variable is the value a_{ij}^k with

$1 \leq i \leq t$, $1 \leq j \leq v$, et $1 \leq k \leq m(j)$. Soit a_{ij} et p_{ij}^k qui valent respectivement

$$a_{ij} = \sum_{k=1}^{m(j)} a_{ij}^k \text{ et } p_{ij}^k = \frac{a_{ij}^k}{a_{ij}}$$

Let $\mathbf{P} = \left| p_{ij}^k \right|$ a table with t species in rows and m modalities in columns. Also

$$p_{ij} = \sum_{k=1}^{m(j)} p_{ij}^k = 1, p_{i.} = \sum_{j=1}^v p_{ij} = v, \text{ et } p_{..} = \sum_{i=1}^t p_{i.} = tv$$

The weights of the rows of table \mathbf{P} are uniformly equal to $r_i = \frac{v}{tv} = \frac{1}{t}$ and the weights of the columns

of the same table are equal to :

$$c_j^k = \frac{\sum_{i=1}^t p_{ij}^k}{tv} = \frac{1}{v} \left(\frac{1}{t} \sum_{i=1}^t p_{ij}^k \right) = \frac{1}{v} \overline{p_j^k}$$

If the affinity of a species for a variable is unknown, it is coded "0, 0, 0, 0". It is therefore implicitly replaced by the mean profile of the corresponding variable. The consequence is that a species with such a

value is not taken into account in the calculation of the weight of the column. The table **P** is then treated as a correspondence analysis.

ROBERTS (1986) showed that "fuzzy set ordination" is a general technique that includes techniques specific to ecologists such as direct gradient analysis (WHITTAKER, 1967), BRAY & CURTIS ordination (1957) as well as the ordination "environmental scalars" by LOUCKS (1962). Species responses are not limited to a precise mathematical function; they may be, for example, non-linear or discontinuous. Each axis of ordination is defined beforehand, so the basis of the interpretation is predetermined. The ecologist can therefore use the "fuzzy set ordination" to explicitly test hypotheses. Examples of "fuzzy set ordination" were given by BANYIKWA *et al.* (1990), FULTON (1996), BOYCE (1998) and FEOLI *et al.* (2003).

Detrended correspondence analysis (DCA)

This is a widely used variant of correspondence analysis (HILL & GAUCH *Jr*, 1980). It therefore treats species and relevés at the same time. This technique is supposed to correct the horseshoe effect, that is to say the arc or V, which appears classically when the points corresponding to the relevés or species are presented in the plane formed by the first two axes. DCA consists of flattening the "horseshoe". This is achieved by cutting the first axis into segments and setting the mean value on the second axis to zero in each segment. Another defect is also corrected, that is to say, the tendency to compress the extremities of the axes with respect to their environment; this is achieved by rescaling the axes so as to equalize as much as possible the variance within the segments of the values of the species along the axes of ordination. In this technique, eigenvalues can no longer be considered as explained proportions of the total variance.

Principal coordinate analysis (PCoA ou NMS)

The principal coordinate analysis (PCoA or NMS) is a generalization of the analysis of the principal component analysis. It was proposed by GOWER (1966). Unlike the PCA which calculates the distance between points in a space (relevés or species) using the Euclidean distance, PCoA allows the use of other distance measurements that better reflect the difference between relevés, for example. The distance between two relevés may be the number of species present in one relevé or in the other but not in both (PIELOU, 1977), or the coefficient of BRAY & CURTIS (1957).

The constrained ordinations

Constrained ordination techniques seek to automatically detect patterns of species variation that can be explained by observed environmental variables. It is therefore an analysis dealing simultaneously with two tables of very different data, species and environmental variables. The resulting diagrams express not only the patterns of variation in the specific composition but also the links between species and environmental

variables (JONGMAN *et al.*, 1987). As we shall see, the term "canonical analysis" frequently used is not correct and should no longer be used.

The "canonical" correspondence analysis or CCA (ter BRAAK, 1986, CHESSEL *et al.*, 1987, McCUNE & MEFFORD, 1999) is unique among ordination methods in that the correspondence analysis of the main table (vegetation relevés) is subjected to the constraint of the mesological variables of a second table by means of a multiple regression. This means that the ordering of relevés and species is constrained by relationships with the environmental variables. CCA is appropriate in cases where the responses of the species to the mesological variables are unimodal and the corresponding mesological variables have been measured. This is a direct gradient analysis. Scoring scores are linear combinations of environmental parameters.

The "canonical" analysis of correspondence (CCA) is a method closely linked to correspondence analysis. Basically, the table of floristic relevés is considered to be part of the correspondence analysis. This is a combination of regression and correspondence analysis. These authors also use the expression "analysis of correspondence on instrumental variables" or AFCVI. CCA is calculated from the double-centered floristic table of the analysis of the correspondences and the centered and normalized mesological table and then transformed on an orthonormal basis.

The "canonical" correspondence analysis (ter BRAAK, 1986 and 1987) selects the linear combination of environmental variables that maximizes the dispersion of coordinates associated with species. In other words, CCA chooses the best weights (c_j) for environmental variables. This gives the first axis of the analysis. The analysis uses the following linear regression equation:

$$x_i = c_0 + c_1 z_{1i} + c_2 z_{2i} + \dots + c_q z_{qi}$$

in which

z_{ji} is the value of environmental variable j at site i ,

c_j is the weight belonging to that variable,

x_i est is the value of the resulting compound environmental at site i .

The second and further CCA axes also select linear combinations of environmental variables that maximize the dispersion of the species scores, but subject to the constraint of being uncorrelated with previous CCA axes. As many axes can be extracted as there are environmental variables.

CCA is therefore a 'restricted correspondence analysis' in the sense that the site scores are restricted to be a linear combination of measured environmental variables. By incorporating the restriction in the two-way weighted averaging algorithm of CA, we obtain an algorithm to CCA. It is therefore not a canonical analysis in the strict sense of the term. However, the CCA acronym may be conserved, considering the three letters as the initials of the three words "Constrained Correspondence Analysis".

Species and sites are positioned as points in the CCA as in CA and interpreted in the same way; sites with high abundances of a given species tend to be close to this species. The environmental variables are represented by vectors and the points corresponding to the species must be projected on these axes. The "biplot scores" are the coordinates of the tops of the lines. The longer the line of a parameter of the environment, the narrower the relation of the vegetation with this parameter. The position of the points-species relative to the lines of the parameters is used to interpret their relationship with this parameter. By lowering a perpendicular from a point-species to the line of the parameter, one has an idea of the degree of response of the species to this parameter. It can also be deduced whether the parameter is favorable (species on the same side as the point of the arrow) or unfavorable to the species (contrary situation).

Constraints become less intense as the number of environmental variables increases. The greater the number of environmental variables, the more the CCA tends to become a simple CA. CCA is therefore a **direct gradient analysis**, unlike the correspondence analysis which is often used as an **indirect gradient analysis**.

Redundancy analysis (RDA) is the "constrained" form of PCA (JONGMAN *et al.*, 1987). RDA is chosen if the length of the gradient measured from a CA is less than 3 and the eigenvalues are higher than with CCA (CLARKE *et al.*, 2005). The DCCA is the application of the CCA algorithm to the "detrended correspondence analysis".

Multiple factor analysis

Multiple factor analysis (AFMULT) is a method of simultaneous analysis of K tables. The AFMULT is therefore a generalization of the techniques of coupling tables. It is due to ESCOFIER & PAGES (1984, 1986, 1989).

We have K arrays having in common n individual-lines (figure 2), each corresponding to a group of column-variables.

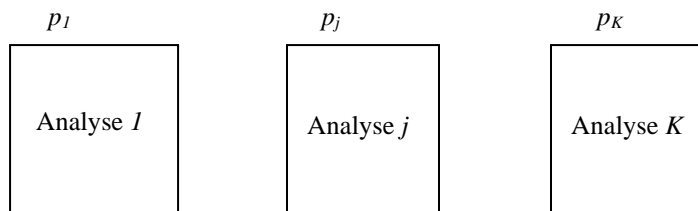


Figure 2. A set of K tables having n individuals in common.

These include, for instance, physico-chemical analyses of water samples taken at a number of sites along a river at different times of the year or for several years. They can also be tables with very different data such as a table of floristic relevés associated with a table of physiographic characteristics and a table of soil analyses; these three tables have in common the individuals (relevés) arranged in rows.

In order to standardize the role of the tables in the simultaneous analysis, the AFMULT consists in bringing to a common scale the inertias (or another parameter) projected on the first axis of the analysis of each table.

AFMULT, in order to standardize the role of the tables in the simultaneous analysis, consists of reducing to a common scale the inertia projected on the first axis of the analysis of each table. The originality of this technique is the introduction of a weighting of the variables, which at the origin is the first eigenvalue of each table of the set.

Multiple factor analysis is built on the principal component analysis or correspondence analysis (both classical and non-symmetric). It requires, for instance, an ACP of the table formed by the combination of the K tables and an ACP for each of the separate tables.

CENTOFANTI *et al.* (1989) give an example with 10 physico-chemical parameters of water surface, measured in 10 stations at 12 dates in a reservoir lake. HAURY *et al.* (1998) studied the ecological determinism of phytocoenose distribution, taking the environmental variables into account; five groups of variables were used: 15 for physical mesology, 10 for geology, 3 for physico-chemistry of water, 10 for vegetation characteristics and 92 for macrophytes. The results of the first five axes have been exploited and clearly show the relationships between various environmental factors and floristic groups. DEVINEAU (2001) presents a joint analysis of four tables, respectively a floristic table with indices of importance of trees and shrubs, a floristic table of frequencies of "regenerations and sub-shrubs", a table relating to soil profiles and a final table with the chemical and textural characteristics of the upper horizon. CHUA *et al.* (2016) conducted one set of analysis using all the data to explore the factors that influence overall seedling recruitment and a second analysis on secondary forests to explore the importance of these factors within recovering forests. They used multiple factor analysis to examine the correlation among the seedling community, adult community, regeneration environment and distance to potential seed sources.

This technique, by its various possibilities, is one of the most promising of recent years.

The steps of the analysis are:

- The calculation of the correlation matrix between the variables of the original table, and its eigenvalues;

- The calculation of the correlation matrices and the first eigenvalues of the sub-tables;
- In the original table, each sub-table is weighted by the first eigenvalue corresponding; for each line, the original variables are divided by the square root of the first corresponding eigenvalue;
- The correlation matrix of the weighted table is calculated, as well as its vectors and eigenvalues; the coordinates and relative contributions of the individuals are calculated as in a classical analysis, the coordinates of the species are obtained by multiplying the lines of the eigenvectors by the square root of the eigenvalues of the weighted table and the square root of the weighting parameter.

In this way, our results correspond to those produced by the Rcmdr software (chapter 13).

We adapted this technique to classical correspondence analysis and to non-symmetric correspondence analysis. For the CA and the ANSC, the sums are calculated on the rows and columns and the weighting by the square root of the first eigenvalue relates directly to the original variables. The sums and the grand total are subject to this weighting. The various stages of calculation are then carried out as in conventional analyzes. However, there is the problem of weighting the sub-tables. Let us take the example of principal component analysis: the first eigenvalue is, to a certain extent, a function of the trace of the matrix, which itself corresponds to the number of variables (in the precise case of a calculating a correlation matrix). A small array produces a smaller eigenvalue than a large one. Mixing sub-tables with a very different number of variables greatly favors sub-tables with a small number of variables. It is therefore not recommended to use this technique by mixing floristical data (often very numerous) and mesological, pedological, chemical, ... often fewer. A solution exists, however, by weighting with the quotient between the first eigenvalue and the trace of the correlation matrix, instead of weighting with the first eigenvalue itself; this avoids the pitfall of a weighting due solely to the size of the matrix.

In correspondence analysis, the trace of the matrix and therefore the eigenvalues, depend in part on the grand total of the data table. Take the example of presence table: if most of the cells in the table correspond to 1 (dense table) and therefore that few cells are empty, the trace of the matrix will be larger than if many cells are empty. The first eigenvalues are therefore larger when an table is rich in data. Thus, if we mix sub-tables with very different data densities, the sparse sub-tables are crushed by the dense sub-tables. We thus had the idea of weighting the floristic sub-tables comprising only "0" and "1" by the proportion of cells occupied by "1" (which we will call **density** thereafter) and thus rebalance the weight of each sub-table.

For CA and ANSC, the sums are calculated on the rows and columns and the weighting by the square root of the adopted parameter relates directly to the original variables. The sums and the grand total are subject to this weighting. The various calculation steps are then carried out as in conventional analyzes.

Co-inertia analysis (MCoA)

The analysis of multiple co-inertia was explained mathematically by CHESSEL & HANAFI (1996). According to APONTE *et al.* (2011), MCoA is a multi-table multivariate analysis technique that allows the simultaneous analysis of several data tables. It maximizes the variance within each individual table and the correlation between the scores of each individual table (individual analysis) as well as the synthesis scores (analysis of the tables juxtaposed) so as to provide a reference structure. MCoA thus optimizes the covariance between several individual analyses and a reference analysis which can be seen as a common structure shared by the individual analyses.

It is also possible to process two sets of data that have not been collected at the same sites, by adding a neighborhood matrix between the sites of the two sampling plans (spatial RLQ analysis, DOLÉDEC *et al.* 1996, DRAY *et al.*, 2002). The technique was also used by DRAY *et al.*, 2003) which links ecological data tables. Two recent examples are given by JAMIL *et al.* (2013) and WESULS *et al.* (2012). They explain the technique as follows: RLQ analysis links a matrix of environmental variables relevés (table **R**) to a matrix of species by functional traits (Table **Q**) using species coverages in the relevés (Table **L**) as links. This is an extension of the co-inertia analysis. Before the RLQ analysis, the tables are subjected to separate multivariate analyses. Correspondence analysis was used for the species-relevés table (Table **L**). The environmental variables-relevés table (Table **R**) was analysed by a principal component analysis. For the species by functional traits table (Table **Q**), a principal component analysis, adapted to a mixture of quantitative and categorical data, was calculated. The AC relevés scores table **L** were used as row weights in the PCA of table **R** and the CA species scores of table **L** were used as row weights in the analysis in table **Q**. RLQ analysis is an extension of the co-inertia analysis that simultaneously takes into account the information contained in tables **R**, **L** and **Q**. There is also another form of analysis called partial RLQ.

Other techniques

The techniques that follow are fundamentally different from those we have just presented, in that they do not use the calculation of eigenvalues and eigenvectors derived from a matrix of covariance or distances.

Bray & Curtis ordination (Polar ordination)

Historically, a first important method is the polar ordination proposed by BRAY & CURTIS (1957, in COTTAM *et al.*, 1973). Let a set of relevés to be ordered. Measurements of similarity are calculated for each relevé with all others. Two extreme relevés are defined either on the basis of the greatest dissimilarity with the rest of the set of relevés and between them or because they represent the extremes of an environmental

gradient. Similarity measurements are used to arrange all other relevés in a sequence along an axis, between the pair of extreme points. A second pair of extreme points, different from the first, is selected and all relevés are arranged along the axis between the second pair. The two axes determine the coordinates of the relevés in a plane, each relevé corresponding to a point on which various characteristics can be superimposed as variables of the environment. The relevés can still be ordered relative to a third axis and so on. Numerous variants of this technique have been proposed. An important description of the vegetation of the State of Wisconsin was carried out with the help of this technique (CURTIS, 1959).

Non-metric multidimensional scaling, NMDS

This is an original technique, proposed by KRUSKAL (1964a and b). We use here the explanations provided by ORLÓCI & KENKEL (1985) and McCUNE & MEFFORD (1999).

Let us start with a table Y of relevés with p rows (species) and n relevés (columns). The objective is to provide a table X with n columns and $t \leq p$ lines, in which the relevés are placed in a different order, according to their resemblances (measured by a distance d), that order reflecting a vegetation gradient or the preponderant influence of one or another factor. The t axes are independent. The calculations begin with the calculation of a $n \times n$ distance matrix (d) computed from table Y . In this first table, the relevés (columns) are placed in the order defined by the observer. Another $n \times n$ distance matrix (D) is then calculated, based on the table X , containing n columns with coordinates attributed to the relevés and defined by the observer (random numbers or coordinates resulting from another ordination) and a small number of lines t also fixed by the observer (1, 2 or more, $\leq p$). X is modified during iterative cycles. The objective is to make the elements of X less and less random so that the order of the elements of D approaches closer and closer to the order of the corresponding elements of d . The degree to which the objective is achieved is measured in terms of a stress coefficient $s(D;d)$. If the value of the stress is too great, the iterations continue; the order of the elements of X is changed and a new table D is computed and then compared with d . When the stress is small enough or when no improvement can be made any more, the configuration of X is definitive. The elements of X are the coordinates of the n relevés on the t axes. ORLÓCI & KENKEL (1985) recommend choosing a random start configuration and not choosing a configuration based on another ordination. They use Euclidean distance. It is also possible to use the McCUNE & MEFFORD software (1999).

This particular technique calls for some remarks:

- The first dimension in a two-dimensional space is not the same as the first dimension in a three-dimensional space or in a single space. For a given number of dimensions, the solution for a particular axis is unique.
- The axis numbers are arbitrary, so that the percentage of variance associated with an axis does not necessarily form a descending series with the increasing number of axes.

- The appropriate number of axes should be determined by the final stress evolution as a function of the number of axes chosen. It is necessary to choose a number of axes beyond which stress reduction is small.

The solution is influenced by the initial order of the relevés in Table X. If one chooses the "random" option, several analyses provide different results.

There are many variants of this technique (PRENTICE, 1977 and 1980, MINCHIN, 1987). The distance measurements can also be made using Euclidean distance, SØRENSEN, JACCARD or correlation coefficients. The "chord" distance we will use is as follows:

$$\delta_{jk} = [2(1-q_{jk})]^{1/2} \text{ in which}$$

$$q_{jk} = \frac{\sum y_{ij} \cdot y_{ik}}{[\sum (y_{ij})^2 \cdot \sum (y_{ik})^2]^{1/2}}, i = 1, \dots, p.$$

ZIMOWSKI *et al.* (2014) give a very explicit example on how this technique is used.

Use of techniques

We consulted 394 papers dealing with these techniques, from the beginning of their use until 2004. To quote them all would take up too much space but they can be found almost by consulting journals like Journal of Vegetation Science, Vegetatio, Journal of Ecology, Ecology, Ecological Monographs, Plant Ecology, Acta Oecologia, Phytocoenologia among the most represented.

Factor analysis *sensu stricto* was practically the only one used before 1965, it is no longer used. The principal component analysis, widely used between 1965 and 1974, is only used at present for vegetation data but is still used to synthesize monotonic data tables such as environmental variables. The analysis of BRAY & CURTIS, little used, was not met after 1984. The principal coordinate analysis, also little used, always appears sporadically. Multiple factor analysis is very rarely used. With a higher frequency, non-metric multidimensional analysis, in its many variants, remains used, as well as redundancy analysis. Most currently used are CA (RA included), CCA and DCA (DCCA has not made a real breakthrough). The two most widely used techniques are therefore very different. CCA is a direct gradient analysis technique whereas DCA and CA are indirect gradient analysis techniques. It is not certain that researchers always choose one or the other technique in full knowledge of their main properties. As ØKLAND (1996) explains, ordination *sensu stricto*, and constrained ordination follow different objectives. The persistence of CCA shows the need for coupling floristic tables and mesological tables. Too often, the choice of techniques is linked to the search for a gradient in species composition (VAN LOOY *et al.*, 2003; NATTA *et al.*, 2003) without any preliminary study on the dispersion of species.

We then consulted all the articles published in the Journal of Vegetation Science from its beginning in 1990 to the last volume of 2016 and constructed the following figure. Only the most frequently used techniques have been used.

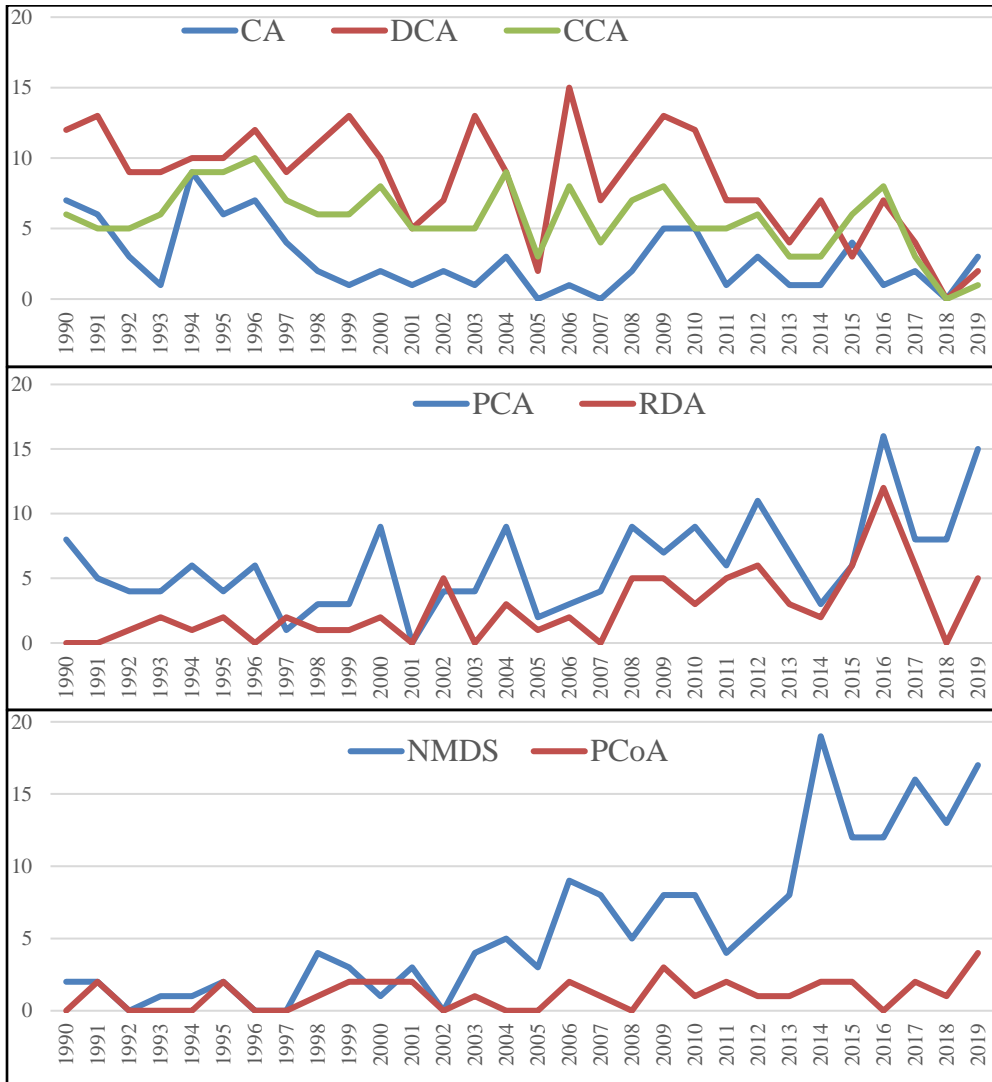


Figure 3. Annual use frequency of various techniques from 1990 to 2019.

CA, DCA and CCA are obviously the most frequently encountered techniques until 2010 and then their use frequencies decrease, specially since 2017. PCA is frequently chosen for tables of mesological parameters or functional traits, rather than for usual vegetation tables; the use of PCA for functional traits is largely the cause of its more frequent use in recent years. The increase in RDA follows that of PCA, which is logical since RDA is a constrained PCA. The frequency of NMDS clearly increases. The other techniques are less

popular. We deduce that the concept of gradient dominates research in vegetal ecology. When the gradient is temporal, the study becomes a succession study (AUSTIN, 1977, DEBUSSCHE *et al.*, 1996). One may also wonder to what extent the choice of a technique is not influenced by the most known sets of software. PC-ORD (McCUNE & GRACE, 2002) and CANOCO (ter BRAAK & ŠMILAUER, 2002) are more and more supplanted by "R" programs. What is surprising is that non-symmetric correspondence analysis is totally ignored and that multiple factor analysis appears only recently (CHUA *et al.*, 2016).

Very close conclusions have been drawn by VON WEHRDEN *et al.* (2009) in a study also covering the years 1990-2007, in a large number of journals, notably by the information provided by the "Institute of Scientific Information" or ISI. The use of multivariate analysis techniques in the field of ecology has been increasing since 1990. CCA is the most used technique, followed by CA / RA. DCA, NMDS, PCA and RDA have recently gained in importance. The authors also stressed the importance of available software. Over the entire period used, CANOCO remains the most used set. Since 1998, PC-ORD has gained prominence and more recently, R-based software is easily accessible. VON WEHRDEN *et al.* (2009) highlight the weakness of several techniques such as DCA especially on the parameters influencing detrending or for other techniques the importance of the standardization of the data of the table of statements or the choice of an index in the calculation of distance matrices. Some mathematical information useful for interpretation, such as eigenvalues, are not always used. The performances of the techniques are also always evaluated in terms of ordination, that is, in their ability to reveal indirect gradients.

In most studies, the percentages of inertia associated with the axes are given. They make it possible to evaluate the number of axes to be taken into consideration. When the cumulative percentages of the first two or three axes are important (without it being possible to set a standard), multivariate analysis is generally considered successful.

It should be noted that, in a very large majority of studies, only the scores of the relevés or of the species on the axes of the multivariate analysis are taken into account in the interpretation of the results. Interpretive aids are rarely used, probably because they are not provided by most available programs. Relative contributions of species or relevés are, however, valuable aids in organizing interpretation. They are taken into account in particular by BONIN & ROUX (1978), BONIN *et al.* (1983), BOUXIN (1986, 1987a & b, 1991, 1995 & 1999).

Multivariate analyses are also used as a technique to reduce the "background noise" of a table of relevés (GAUCH, 1982a). Indeed, the relevé tables contain interesting variations, interpretable in historical or environmental terms, as well as random, unusable variations. So this is a technique of data reduction.

They are also used to study multi-species patterns in relevé or plots grids or simply when the relevés are placed in rows. In most studies, species, relevés or both sets are presented in a two-dimensional space (axes 1 and 2, 1 and 3, 2 and 3, etc.). In the study of a transect, ESTÈVE (1978) recommends that each axis be represented separately, taking into account the position of the relevés in space. The function relating the relevé scores on the considered axis (on the y-axis) and the position of the relevés (on the abscissa) highlights homogeneous zones, transition zones and zones with high floristic variability. The axes are placed in succession, parallel to each other. In any case, the ESTÈVE recommendation can be extended to any map of relevés. Examples are given by BACHACOU & CHESSEL (1979), BOUXIN (1983), BOUXIN & LE BOULENGÉ (1983), BOUXIN & GAUTIER (1979, 1982) and GAUTIER (1979). GALIANO (1983) uses the coordinates of relevés along the first axis of a RA as a basis for studying multispecific patterns. Examples of applications will be presented in this chapter as well as in the chapters on statistical analysis of vegetation tables.

Multivariate analyses were also used as the basis for important phyto-ecological or phytosociological studies. These include GUINOCHE (1973), BONIN & ROUX (1978), BONIN *et al.* (1983), MÄKIRINTA (1989).

Many authors do not analyse complete tables but eliminate rare species such as BARUCH (1984), BELSKY (1983), BRZEZIECKI (1987), BURKE (2001), COOPER (1984), HALL & SWAINE, HERMY & STIEPERAERE (1981) and NEWBERY (1991) among others. These simplifications are supposed to eliminate or reduce observed distortions, such as the horseshoe effect. Some authors do not hesitate to deal with very large tables (1117 records), which are then simplified by the elimination of rare and marginal species (BERGMEIER & DIMOPOULOS, 2001).

Multivariate analysis methods are also techniques of data reduction: they simply make it possible to transform a table with n rows (species) into a table with two, three or a few rows. This reduction in dimensionality is considered as one of the objectives of the analysis by DALE (1975) and AUSTIN (1976), insofar as it is associated with a gradient search. The simplified tables, whose redundancies have been eliminated, can also be used in classification algorithms, for example (BERTHET *et al.*, 1976; BOUXIN, 1978, 1983, 1986, 1987a, b, 1995 and 1999, BOUXIN & DEFLANDRE, 1988) or, as has been seen, as a basis for studying multispecies motifs.

The problem of the distortions that appear in the representation of transformed variables in a 2- or 3-dimensional space, generated a lot of discussion. ESTÈVE (1978) clearly attaches great importance in the definition of the calculated distance between relevés and the method of applying the configuration on a plane rather than the linearity of the model. It is absurd to question the lack of linearity between the performance

of species and the factors of the environment, since these factors are not taken into account, in indirect ordination at least. In the interpretation, therefore, attention must be paid to the mutual distance between the relevés taken into account. Distortions are created by the difference between the specific richness of the relevés (β diversity), but it is far from being the only cause. Diversity is another cause (DARGIE, 1986). DEL MORAL (1980) finds that PCA, RA or NMDS give very good results with low β diversity, whereas the technique of BRAY & CURTIS (1957) gives bad results.

Opinions are, however, strongly divided; thus GAUCH & SCRUGGS (1979) consider that RA is sufficiently tolerant to a high β diversity, at least when compared with PCA; RA is used to calculate this β diversity [according to WHITTAKER & WOODWELL (1978), it is the measurement of the change in the species composition along an environmental gradient; it is calculated by the degree of decrease of the logarithm of the similarity between relevés as the separation increases along the gradient determined by the first axis of RA; it is expressed in "half-changes" in the composition of species]. OLSVIG (1979) presents different measures of this diversity. Other calculation techniques have been proposed by WILSON & MOHLER (1983) who insist that the "background noise" present in the relevé tables can obscure the ecological patterns. Examples and calculations are also given by ITOW (1991) and OKSANEN & TONTERI (1995). DEL MORAL & WATSON (1978) consider that rare species contribute strongly to distortions. LLOYD *et al.* (1994) express ideas similar to those of ESTÈVE (1978) when they declare preferring indirect ordination (DCA) to constrained ordination (CCA and DCCA) because they wish to estimate the vegetation independently of the environmental variables.

Another problem related to correspondence analysis is the GUTTMAN effect (ESTÈVE, 1978) which is the appearance of artificial factors functions of the first; when this effect is marked, it is sometimes useful to consider only factors 1 and 3.

Multivariate analysis methods are most often used from a descriptive point of view. Inference is rarely used. Recent techniques facilitate the transition to statistical inference: these are the randomization, bootstrap and Monte Carlo tests (MANLY, 1997). The randomization tests were applied by BOUXIN (1999) to the principal component analyses and the correspondence analyses; they relate to the relative contributions of axes, records and species; the permutations relate to each line independently (here the species); it is thus possible to determine, with a risk $\alpha = 0.05$, the number of significant axes and the species or the relevés presenting significant relative contributions. The bootstrap technique is used by KNOX & PEET (1989) to retain the number of useful axes in a DCA and by PILLAR (1999) to determine the number of dimensions in the analysis of a sample of relevés. In a fuzzy set ordination, FULTON (1996) uses a Monte Carlo test to test the robustness of the set of data against the environmental preferences calculated for each species. Ter BRAAK (1987 and 1990) and ter BRAAK & ŠMILAUER (2002), in the canonical correspondence analysis, use the distribution obtained by randomly changing the environmental variables between the sites in order to

evaluate the statistical significance of the effects of environmental variables. ALLEN *et al.* (1991) use the bootstrap technique to evaluate the relationship between altitude and the first axis of CCA. REED *et al.* (1993) use a Monte Carlo test to evaluate the significance of axes produced by DCA and DCCA.

Several conclusions

Which technique to choose? GREIG-SMITH (1964) found that choosing a technique to use in a particular case was not easy to do. This depended on the type of data, the diversity of data gathered, and sometimes available software. The choice depended mainly on the objective sought, the type of information sought to be extracted from the relevé table. It was therefore difficult to give general recommendations. But the same author noted that, in cases where several techniques had been applied to the same set of data, they produced coherent results, which was encouraging. This finding has also been made recently by several authors, including HUEBNER & VANKAT (2003) for DCA, CCA and NMDS and SCHADE *et al.* (2003) with respect to DCA, PCA & NMDS. The robustness of the various techniques has been observed but the limits to this robustness are very poorly known (AUSTIN *et al.*, 1994); these authors emphasize the importance of well knowing the shape of the response curve of species along environmental gradients. KARADŽIĆ & POPOVIĆ (1994) increase the performances of PCA, in the search for gradients, by double standardization of the relevé tables, bearing on both rows and columns; in this case PCA is even greater than CA. A good way to evaluate the quality of the representation of the results of a technique is to use several of them and to compare their respective contributions (ØKLAND, 1996). We must remember the lack of aids in the interpretation of several techniques. Only PCA, CA, NSCA and CCA (in the ADE-4 software) provide these aids.

We will try to answer these questions in the following chapters.

january 2023

References

- ALLEN, T.F.H. & HOEKSTRA, T.W. (1991). The confusion between sacle-defined levels and conventional levels of organisation in ecology. *Journal of Vegetation Science* **1**: 5-12.
- APONTE, C., GARCIA, L. V., PÉREZ-RAMOS, I. M., GUTIÉRREZ, E. & MARAÑÓN, T. (2011). Oak tress and soill interactions in Mediterranean forests: a positive feedback model. *Journal of Vegetation Science* **22**: 856-867.
- AUSTIN, M.P. (1976). On non-linear species response models in ordination. *Vegetatio* **33**: 33-41.

- AUSTIN, M.P. (1977). Use of ordination and other multivariate descriptive methods to study succession. *Vegetatio* **35**: 165-175.
- AUSTIN, M.P., NICHOLLS, A.O., DOHERTY, M.D. & MEYERS, J.A. (1994). Determining species response functions to an environmental gradient by means of a β -function. *Journal of Vegetation Science* **5**: 215-228.
- BACHACOU, J. & CHESSEL, D. (1979). Étude des structures spatiales en forêt alluviale rhénane. III. Dispersion interspécifique et analyse des correspondances. *Oecol. Plant.* **14**: 371-388.
- BANYIKWA, F.F., FEOLI, E. & ZUCCARELLO, V. (1990). Fuzzy set ordination and classification of Serenget short grasslands, Tanzania. *Journal of Vegetation Science* **1**: 97-104.
- BARUCH, Z. (1984). Ordination and classification of vegetation along an altitudinal gradient in the Venezuelan páramos. *Vegetatio* **55**: 115-126.
- BELSKY, A. J. (1983). Small-scale pattern in grassland communities in the Serengeti National Park, Tanzania. *Vegetatio* **55**: 141-151.
- BERGMEIER, E. & DIMOPOULOS, P. (2001). *Fagus sylvatica* forest vegetation in Greece: Syntaxonomy and gradient analysis. *Journal of Vegetation Science* **12**: 109-126.
- BERTHET, P., FEYTMANS, E., STEVENS, D. & GENETTE, A. (1976). A new divisive method of classification illustrated by its applications to ecological problems. *Proc. Ninth Int. Biom. Conf., invited papers. Vol. II*: 366-382.
- BONIN, G., AUBERT, G., BARBERO, M., GAMISANS, J., GRUBER, M., LOISEL, R., QUEZEL, P., SANDOZ, H., THINON, M., VERDENNE, G. (1983). Mise en évidence de la dynamique de quelques écosystèmes forestiers et préforestiers provençaux aux étages méditerranéens S.L. à l'aide des taxons indicateurs. *Vegetatio* **54** : 79-96.
- BONIN, G. & ROUX, M. (1978). Utilisation de l'analyse factorielle des correspondances dans l'étude phytocologique de quelques pelouses de l'Apennin lucano-calabrais. *Oecol. Plant.* **13**: 121-138.
- BOUROUCHE, J.-M. & SAPORTA, G. (1980). *L'analyse des données*. P.U.F. Collection Que sais-je ? 125 pp.
- BOUXIN, G. (1978). Etude comparative d'une nouvelle méthode de classification descendante polythétique avec réallocations. *Biométrie-Praximétrie* **18**: 20-48.
- BOUXIN, G. (1983). Multi-scaled pattern analysis: an example with savanna vegetation and a proposal for a sampling design. *Vegetatio* **52** : 161-169.
- BOUXIN, G. (1986). Le traitement statistique des tableaux de végétation. I. Les petits tableaux. *Biom. Praximétrie* **26** : 49-72.
- BOUXIN, G. (1987a). Le traitement statistique des tableaux de végétation. II. Les ensembles de tableaux et les grands tableaux. *Biom. Praximétrie* **27** : 65-97.
- BOUXIN, G. (1987b). La végétation herbacée des ruisselets d'un massif boisé sur roches éodévonienne, en Belgique. *Coll. Phytosociol.* **XV**: 93-106.

- BOUXIN, G. (1991). La végétation aquatique et du bord de l'eau dans le bassin versant du Bocq (Condroz, Belgique). *Revue des Sciences de l'Eau* **4** : 185-210.
- BOUXIN, G. (1999). Description de la végétation aquatique et du bord de l'eau dans le bassin hydrographique de la Molinee (Condroz, Belgique) par l'analyse multiscalaire des motifs. *Ecologie* **30** : 139-163.
- BOUXIN, G. & DEFLANDRE, G. (1988). Les groupements végétaux dans la pelouse calcaire de Belvaux. *Colloques phytosociologiques*. **XVI**: 619-628.
- BOUXIN, G. & GAUTIER, N. (1979). Structure de la strate herbacée dans deux pelouses calcaires du district mosan. *Oecol. Plant.* **14**: 219-231.
- BOUXIN, G. & GAUTIER, N. (1982). Pattern analysis in Belgian limestone grasslands. *Vegetatio* **49**, 65-83.
- BOUXIN, G. & LE BOULENGE, E. (1983). A phytosociological system based on multi-scaled pattern analysis: a first example. *Vegetatio* **54**: 3-16.
- BOYCE, R.L. (1998). Fuzzy set ordination along an elevation gradient on a mountain in Vermont, USA. *Journal of Vegetation Science* **9**: 191-200.
- BRAY, J.R. & CURTIS, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**: 325-349.
- BRZEZIECKI, B. (1987). Analysis of vegetation-environment relationships using a simultaneous equations model. *Vegetatio* **71**: 175-184.
- BURKE, A. (2001). Classification and ordination of plant communities of the Naukluft Mountains, Namibia. *Journal of Vegetation Science* **12**: 53-60.
- CENTOFANTI, M., CHESSEL, D & DOLEDEC, S. (1989). Stabilité d'une structure spatiale et compromis d'une analyse statistique multi-tableaux : application à la physico-chimie d'un lac réservoir. *Revue des Sciences de l'Eau* **2**: 71-93.
- CHESSEL, D., DUFOUR, A.B. & THIOULOUSE, J. (2004). The ade4 package- I : One-table method. *R News* **4**: 5-10.
- CHESSEL, D. & HANAFI, M. (1996). Analyses de la co-inertie de K nuages de points. *Revue de Statistique appliquée* **44**: 35-60.
- CHESSEL, D., LEBRETON, J. D. & YOCCOZ, N. (1987). Propriétés de l'analyse canonique des correspondances. Une utilisation en hydrobiologie. *Revue de Statistique Appliquée* **35**: 55-72
- CLARKE, P.J., LATZ, P.K. & ALBRECHT, D.E. (2005). Long-term changes in semi-arid vegetation: Invasion of an exotic perennial grass has larger effects than rainfall variability. *Journal of Vegetation Science* **16**: 237-248.
- CHUA, S.C., RAMAGE, B.S. & POTTS, M.D. (2016). Soil degradation and feedback processes affect long-term recovery of tropical secondary forests. *Journal of Vegetation Science* **27**: 800-811.
- COOPER, A. (1984). An application of multivariate methods to a study of community composition and structure in an escarpment woodland in northeast Ireland. *Vegetatio* **55**: 93-104.

- COQUILLARD, P. & GUEUGNOT, J. (1991). Regressive dynamic and denudation processes of vegetation on volcanoes in the 'Chaîne des Puys' (French Massif Central): assay of interpretation. *Vegetatio* **93**: 131-141.
- COTTAM, G., GLENN GOFF, F. & WHITTAKER, R.H. (1973). Wisconsin comparative ordination. In Ordination and Classification of Communities. *Handbook of Vegetation Science* **5**: 195-221.
- CURTIS, J.T. (1959). *The vegetation of Wisconsin. An ordination of plant communities*. Univ. Wisconsin Press, Madison. Xi + 657 pp.
- DAGNELIE, P. (1960). Contribution à l'étude des communautés végétales par l'analyse factorielle. *Bulletin du Service de la Carte Phytogéographique. Série B. Carte des Groupements végétaux*. **V**: 7-71.
- DAGNELIE, P. (1973). L'analyse factorielle. In Ordination and Classification of Communities. *Handbook of Vegetation Science* **5**: 225-248.
- DALE, M.B. (1975). On objectives of methods of ordination. *Vegetatio* **30**: 15-32.
- DARGIE, T.C.D. (1986). Species richness and distortion in reciprocal averaging and detrended correspondence analysis. *Vegetatio* **65**: 95-98.
- DEBUSSCHE, M., ESCARRÉ, J., LEPART, J., HOUSSARD, C. & LAVOREL, S. (1996). Changes in Mediterranean plant succession: old-fields revisited. *Journal of Vegetation Science* **7**: 519-526.
- DEL MORAL, R. (1980). On selecting indirect ordination methods. *Vegetatio* **42** : 75-84.
- DEL MORAL, R. & WATSON, A.F. (1978). Gradient structure of forest vegetation in the Central Washington Cascades. *Vegetatio* **38** : 29-48.
- DEVINEAU, J.-L. (2001). Les espèces ligneuses indicatrices des sols dans des savanes et jachères de l'Ouest du Burkina Faso. *Phytocoenologia* **31**: 325-351.
- DOLÉDEC, S., CHESSEL, D., TER BRAAK, C.J.F. & CHAMPELY, S. (1996). Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* **3**: 143-166.
- DRAY, S., DUFOUR, A.B. & CHESSEL, D. (2007). The ade4 package- I : Two-table and K-table method. *R News* **7**: 47-52
- DRAY, S., PETTORELLI, N. & CHESSEL, D. (2002). Matching data sets from two different spatial samples. *Journal of Vegetation Science* **13**: 867-874.
- DRAY, S., CHESSEL, D. & THIOULOUSE, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology* **84**: 3078-3089.
- ELLENBERG, H. (1974). Zeigewerte der Gefäßpflanzen Mitteleuropas. *Scripta Geobotanica* **IX**: 1-122.
- ESCOFIER, B. & PAGES, J. (1984). *L'analyse factorielle multiple : une méthode de comparaison de groupes de variables*. In: *Data Analysis and Informatics III*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 41-55.

- ESCOFIER, B. & PAGES, J. (1986). *Le traitement des variables qualitatives et des tableaux mixtes par analyse factorielle multiple*. In: Data Analysis and Informatics **IV**. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 179-191.
- ESCOFIER, B. & PAGES, J. (1989). *Multiple factor analysis: results of a three-year utilization*. In: *Multiway data analysis*. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 277-285.
- ESTÈVE, J. (1978). Les méthodes d'ordination : éléments pour une discussion. *Biométrie et Ecologie* **1** : 223-250.
- EZCURRA, E. (1987). A comparison of reciprocal averaging and non-centred principal component analysis. *Vegetatio* **71**: 41-47.
- EZCURRA, E., EQUIHUA, M. & LÓPEZ-PORTILLO, J. (1987). The desert vegetation of El Pinacate, Sonora, Mexico. *Vegetatio* **71**: 49-60.
- FEOLI, E., GANIS, P., ORTOLAN, I., SITONI, D. & ZERIHUN, W. (2003). Measuring the effects of human impact on vegetation by integrating phytosociology and remote sensing in a fuzzy set approach. *Journal of Vegetation Science* **14**: 751-760.
- FOUCART, T. (1982). *Analyse factorielle. Programmation sur micro-ordinateurs*. Masson, Paris. 246 pp.
- FRIED, G., CHAUVEL, B. & REBOUD, X. (2009). A functional analysis of large-scale temporal shifts from 1970 to 2000 in weed assemblages of sunflower crops in France. *Journal of Vegetation Science* **20** : 49-58.
- FULTON, M.R. (1996). An application of fuzzy set ordination to determine tree habitat suitability of sites from a regional data set. *Journal of Vegetation Science* **7**: 739-746.
- GALIANO E.F. (1983). Detection of multispecies patterns in plant populations. *Vegetatio* **53**: 129-138.
- GAUCH, H., G. Jr. (1982a). Noise reduction by eigenvector ordinations. *Ecology* **63**: 1643-1649.
- GAUCH, H.G., Jr. (1982b). *Multivariate analysis in community ecology*. Cambridge University Press. 298 pp.
- GAUCH, H.G., Jr. & SCRUGGS, W.M. (1979). Variants of polar ordination. *Vegetatio* **40**: 147-153.
- GAUTIER, C. (1979). Analyse des grilles en présence-absence, cas mono- et multispécifique. *Oecol. Plant.* **14**: 251-264.
- GIMARET-CARPENTIER, C, CHESSEL, D. & PASCAL, J.-P. (1998). Non-symmetric correspondence analysis: an alternative for species occurrences data. *Plant Ecology* **138**: 97-112.
- GOODALL, D.W. (1954). Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Austral. J. Bot.* **2**: 304-324.
- GOWER, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**: 325-338.
- GREENACRE, M. (2010). Correspondence analysis of raw data. *Ecology* **91**: 958-963.
- GREIG-SMITH, P. (1964). *Quantitative plant ecology (second edition)*. Butterworths, London. 256 pp.
- GUINOCHET, M. (1973). *Phytosociologie*. Masson et Cie. 227 pp.

- HALL, J.B. & SWAINE, M.D. (1976). Classification and ecology of closed-canopy forest in Ghana. *Journal of Ecology* **64**: 913-951.
- HAURY, J., JAFFRÉ, M., DUTARTRE, A., PELTRE, M.-C., BARBE, J., TRÉMOLIÈRES, M., GUERLESQUIN, M. & MULLER, S. (1998). Application de la méthode « Milieu Et Végétaux fixés » à 12 rivières françaises : typologie floristique préliminaire. *Annls Limnol.* **34**: 129-138
- HILL, M.O. (1973). The intensity of spatial pattern in plant communities. *Journal of Ecology* **61**: 225-235.
- HERMY, M. & STIEPERAERE, H. (1981). An indirect gradient analysis of the ecological relationships between ancient and recent riverine woodlands to the South of Bruges (Flanders, Belgium). *Vegetatio* **44**: 43-49.
- HILL, MO. & GAUCH, Jr, H.G. (1980). Detrended correspondence analysis, an improved ordination technique. *Vegetatio* **42**: 47-58.
- HUEBNER, C.D. & VANKAT, J.L. (2003). The importance of environment vs. disturbance in the vegetation mosaic of Central Arizona. *Journal of Vegetation Science* **14**: 25-34.
- HUSSON, F., LÊ, S. & PAGÈS, J. (2009). *Analyse de données avec R*. Presses Universitaires de Rennes. 224 pp.
- ITOW, S. (1991). Species turnover and diversity patterns along an evergreen broad-leaved forest coenocline. *Journal of Vegetation Science* **2**: 477-484.
- JAMIL, T., OZINGA, W.A., KLEYER, M. & ter BRAAK, C. (2013). Selecting traits that explain species-environment relationships: a generalized linear mixed model approach. *Journal of Vegetation Science* **24**: 988-1000.
- JONGMAN, R.H.G., ter BRAAK, C.J.F. & van TONGEREN, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc Wageningen. 299 pp.
- KARADŽIĆ, B. & POPOVIĆ, R. (1994). Generalized Standardization Procedure in ecological ordination. *Journal of Vegetation Science* **5**: 259-262.
- KNOX, R.G. & PEET, R.K. (1989). Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. *Vegetatio* **80**: 153-165.
- KROONENBERG, P. M. & LOMBARDO R. (1999). Nonsymmetric correspondence analysis: a tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research* **34**: 367-396.
- KRUSKAL, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**: 1-27.
- KRUSKAL, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**: 115-129.
- LLOYD, A.H., ARMBRUSTER, W.S. & EDWARDS, M.E. (1994). Ecology of a steppe-tundra gradient in interior Alaska. *Journal of Vegetation Science*: 897-912.
- LOUCKS, O.L. (1962). Ordinating forest communities by means of environmental scalars and phytosociological indices. *Ecol. Monogr.* **32**: 137-166.

- MÄKIRINTA, U. (1989). Classification of South Swedish Isoetid vegetation with the help of numerical methods. *Vegetatio* **81**: 145-157.
- MANLY, B.F.J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. Second edition. Chapman & Hall. 399 pp.
- McCUNE, B & MEFFORD, M.J. (1999). *PC-Ord. Multivariate analysis of ecological data, version 4*. MjM Software Design, Gleneden Beach, OR, US.
- McCUNE, B. & GRACE, J.B. (2002). *Analysis of ecological communities*. MjM Software, Gleneden Beach, OR, US.
- MINCHIN, P.R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, **69**: 89-107.
- NEWBERRY, D. McC. (1991). Floristic variation within kerangas (heath) forest: re-evaluation of data from Sarawak and Brunei. *Vegetatio* **96**: 43-86.
- NATTA, A.K., SINSIN, B. & VAN DER MAESEN, G. (2003). A phytosociological study of riparian forests in Benin (West Africa). *Belgian Journal of Botany* **136**: 109-128.
- ØKLAND, R.H. (1996). Are ordination and constrained ordination alternative or complementary strategies in general ecological studies ? *Journal of Vegetation Science* **7**: 289-292.
- OKSANEN, J. & TONTERI, T. (1995). Rate of compositional turnover along gradients and total gradient length. *Journal of Vegetation Science* **6**: 815-884.
- OLSVIG, L.S. (1979). Pattern and diversity analysis of the irradiated oak-pine forest, Brookhaven, New York. *Vegetatio* **40**: 65-78.
- ORLÓCI, L. & KENKEL, N.C. (1985). *Statistical Ecology Monographs. Vol. I. Introduction to data analysis*. International Co-operative Publishing House : 339 pp.
- PALM, R. (1993). Les méthodes d'analyse factorielle : principes et applications. *Notes Stat. Inform.* (Gembloux) 93/1, 38 pp.
- PIELOU, E.C. (1977). *Mathematical Ecology*. Wiley-Interscience. 385 pp.
- PILLAR, V.D.P. (1999). The bootstrapped ordination re-examined. *Journal of Vegetation Science* **10**: 895-902.
- PRENTICE, I.C. (1977). Non-metric ordination methods in ecology. *Journal of Ecology* **65**: 85-94.
- PRENTICE, I.C. (1980). Vegetation analysis and order invariant gradient models. *Vegetatio* **42**: 27-34.
- REED, R.A., PEET, R.K., PALMER, M.W. & WHITE, P.S. (1993). Scale dependence of vegetation-environment correlations: A case study of a North Carolina piedmont woodland. *Journal of Vegetation Science* **4**: 329-340.
- ROBERTS, D. W. (1986). Ordination on the basis of fuzzy set theory. *Vegetatio* **66**: 123-143.
- SCHADE, J.D, SPONSELLER, R., COLLINS, S.L. & STILES, A. (2003). The influence of Mesquite canopies on understory vegetation: effects of landscape position. *Journal of Vegetation Science* **14**: 743-750.

- ter BRAAK, C.J.F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**: 1167-1179.
- ter BRAAK, C.J.F. (1987). Ordination. In "Data analysis in community and landscape ecology". JONGMAN, R.H.G., ter BRAAK, C.J.F. and van TONGEREN, O.F.R. editors. Pudoc, Wageningen. 91-173.
- ter BRAAK, C.J.F. (1990). *Updates notes: CANOCO version 3.10*. Agricultural Mathematics Group, Wageningen.
- ter BRAAK, C.J.F. & ŠMILAUER, P. (2002). *Canoco for windows 4.5*. Biometris, Wageningen-UR, NL.
- VAN LOOY, K., HONNAY, O., BOSSUYT, B. & HERMY, M. (2003). The effects of river embankment and forest fragmentation of the plant species richness and composition of floodplain forests in the Meuse valley, Belgium. *Belgian Journal of Botany* **136**: 97-108.
- VON WEHRDEN, H., HANSPACH, J., BRUELHEIDE, H. & WESCHE, K. (2009). Pluralism and diversity: trends in the use and application of ordination methods 1990-2007. *Journal of Vegetation Science* **20** : 695-705.
- WAMELINK, G.W.W., JOOSTEN, V., VAN DOBBEN, H.F. & BERENDSE, F. (2002). Validity of Ellenberg indicator values judged from physico-chemical field measurements. *Journal of Vegetation Science* **13**: 269-278.
- WESULS, D., OLDELAND, J. & DRAY, S. (2013). Disentangling plant trait responses to livestock grazing from spatio-temporal variation: the partial RLQ approach. *Journal of Vegetation Science* **23**: 98-113.
- WHITTAKER, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**: 279-338.
- WHITTAKER, R.H. (1967). Gradient analysis of vegetation. *Biol. Rev.* **42**: 207-264.
- WHITTAKER, R.H. (1973). *Ordination and classification of communities*. Handbook of vegetation science. Part V. Dr. W. Junk b.v. - Publishers - The Hague. 737 pp.
- WHITTAKER, R.H. & WOODWELL, G.M. (1978). *Retgression and coenocline distance*. pp 51-70 in R.H. WHITTAKER, editor. Ordination of plants communities, Dr. W. Junk, The Hage, The Netherlands.
- WILDI, O. (2013) *Data Analysis in Vegetation Ecology* (second edition). Wiley-Blackwel. 311 pp.
- WILSON, M.V. & MOHLER, C.L. (1983). Measuring compositional change along gradients. *Vegetatio* **54** : 129-141.
- ZIMOWSKI, M., LEUSCHNER, H. H., GÄRTNER, H. & BERRGMEIER, E. (2014). Age and diversity of Mediterranean dwarf shrublands: a dendrochronological approach along an altitudinal gradient on Crete. *Journal of Vegetation Science* **25**: 122-134.

Annexe: calculation of eigenvalues and eigenvectors

Used data

We present a simple example with three species, taken from the Tailfer5 table used in this chapter. Compared to this table, several empty relevés have been removed.

Species/Relevés	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	19	23	24	25	26	27	28	29	Sum
<i>Carex pendula</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	9
<i>Carex remota</i>	0	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	12
<i>Deschampsia cespitosa</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1	1	18
Sum	1	2	2	2	2	2	2	2	1	2	2	1	2	1	1	2	1	2	1	1	1	1	3	2	39

Examples of principal component analysis

Matrix of centered reduced data

Relevés	<i>Carex pendula</i>	<i>Carex remota</i>	<i>Deschampsia cespitosa</i>	Sum
1	-0.7746	-1	0.57735	-1.19725
2	-0.7746	1	0.57735	0.80275
3	-0.7746	1	0.57735	0.80275
4	-0.7746	1	0.57735	0.80275
5	-0.7746	1	0.57735	0.80275
6	-0.7746	1	0.57735	0.80275
8	-0.7746	1	0.57735	0.80275
9	-0.7746	1	0.57735	0.80275
10	-0.7746	-1	0.57735	-1.19725
11	-0.7746	1	0.57735	0.80275
12	-0.7746	1	0.57735	0.80275
13	-0.7746	-1	0.57735	-1.19725
14	-0.7746	1	0.57735	0.80275
15	-0.7746	-1	0.57735	-1.19725
16	1.290994	-1	-1.732051	-1.441057
17	1.290994	-1	0.57735	0.868344
19	-0.7746	1	-1.732051	-1.506651
23	1.290994	-1	0.57735	0.868344
24	1.290994	-1	-1.732051	-1.441057
25	1.290994	-1	-1.732051	-1.441057
26	1.290994	-1	-1.732051	-1.441057
27	1.290994	-1	-1.732051	-1.441057
28	1.290994	1	0.57735	2.868344
29	1.290994	-1	0.57735	0.868344
Sum	0	0	0	0

Note: null totals are rounded off.

Correlation matrix

	<i>Carex pendula</i>	<i>Carex remota</i>	<i>Deschampsia cespitosa</i>
<i>Carex pendula</i>	1		
<i>Carex remota</i>	-0.60247	1	
<i>Deschampsia cespitosa</i>	-0.54660	0.38490	1

Eigenvalues

$$\lambda_1 = 2.027863$$

$$\lambda_2 = 0.6118351$$

$$\lambda_3 = 0.353785$$

Eigenvectors

	v1	v2	v3
<i>Carex pendula</i>	-0.62046	-0.07933	0.780219
<i>Carex remota</i>	0.566674	0.642395	0.515952
<i>Deschampsia cespitosa</i>	0.542136	-0.76226	0.353624

Example of calculation

The three determinants

1 - 2.027863	-0.60247	-0.5466
-0.60247	1 - 2.027863	0.3849
-0.5466	-0.3849	1 - 2.027863

1 - 0.6118351	-0.60247	-0.5466
-0.60247	1 - 0.6118351	0.3849
-0.5466	-0.3849	1 - 0.6118351

1 - 0.353785	-0.60247	-0.5466
-0.60247	1 - 0.353785	0.3849
-0.5466	-0.3849	1 - 0.353785

are equal to zero.

The product of the matrix **A** and of the vector **v**₁ is equal to the product of the vector **v**₁ by the eigenvalue λ_1 (involving rounding errors).

$$\begin{bmatrix} 1 & -0.60247 & -0.54660 \\ -0.60247 & 1 & 0.38490 \\ -0.54660 & 0.38490 & 1 \end{bmatrix} * \begin{bmatrix} -0.62046 \\ 0.566674 \\ 0.542136 \end{bmatrix} = \begin{bmatrix} -0.62046 \\ 0.566674 \\ 0.542136 \end{bmatrix} * 2.027863$$

Similarly, the two other products involving respectively the second and third eigenvectors and the second and third eigenvalues lead to the same equalities (involving rounding errors).

$$\begin{bmatrix} 1 & -0.60247 & -0.54660 \\ -0.60247 & 1 & 0.38490 \\ -0.54660 & 0.38490 & 1 \end{bmatrix} * \begin{bmatrix} -0.07933 \\ 0.642395 \\ -0.76226 \end{bmatrix} = \begin{bmatrix} -0.07933 \\ 0.642395 \\ -0.76226 \end{bmatrix} * 0.6118351$$

$$\begin{bmatrix} 1 & -0.60247 & -0.54660 \\ -0.60247 & 1 & 0.38490 \\ -0.54660 & 0.38490 & 1 \end{bmatrix} * \begin{bmatrix} 0.780219 \\ 0.515952 \\ 0.353624 \end{bmatrix} = \begin{bmatrix} 0.780219 \\ 0.515952 \\ 0.353624 \end{bmatrix} * 0.353785$$

For example, the *Carex pendula* coordinate on the first axis is equal to:

$$-0.62046 * \sqrt{2.027863} = -0.883554.$$

The coordinates of the relevés on an axis are obtained by multiplying the elements of the columns of the table of the reduced centered data by the corresponding values of the eigenvectors of the axis in question.

For example, the coordinate of the first relevé on the first axis is equal to:

$$(-0.7746 * 0.62046) + (-1 * 0.566674) + (0.57735 * 0.42136) = 0.226936.$$

Example of correspondence analysis

As in this example, there are only three species for 24 relevés, the matrix is automatically permuted by the program and the columns correspond to the species.

Weighted matrix

Relevés	<i>Carex pendula</i>	<i>Carex remota</i>	<i>Deschampsia cespitosa</i>	Sum
1	-1	-1	1.166666	-0.833334
2	-1	0.625	0.083333	
3	-1	0.625	0.083333	-0.291667
4	-1	0.625	0.083333	-0.291667
5	-1	0.625	0.083333	-0.291667
6	-1	0.625	0.083333	-0.291667
8	-1	0.625	0.083333	-0.291667
9	-1	0.625	0.083333	-0.291667
10	-1	-1	1.166666	-0.833334
11	-1	0.625	0.083333	-0.291667

12	-1	0.625	0.083333	-0.291667
13	-1	-1	1.166666	-0.833334
14	-1	0.625	0.083333	-0.291667
15	-1	-1	1.166666	-0.833334
16	3.333333	-1	-1	1.333333
17	1.166666	-1	0.083333	0.083333
19	-1	2.25	-1	0.25
23	1.166666	-1	0.083333	0.083333
24	3.333333	-1	-1	1.333333
25	3.333333	-1	-1	1.333333
26	3.333333	-1	-1	1.333333
27	3.333333	-1	-1	1.333333
28	0.444444	0.083333	-0.27778	0.249993
29	1.166666	-1	0.083333	0.083333
Sum	5.611103	-3.416667	-0.527787	1.166651

Note: totals are rounded off.

Covariance matrix

	<i>Carex pendula</i>	<i>Carex remota</i>	<i>Deschampsia cespitosa</i>
<i>Carex pendula</i>	0.52849		
<i>Carex remota</i>	-0.2344	0.220085	
<i>Deschampsia cespitosa</i>	-0.18232	-0.01396	0.140313

Eigenvalues

$$\lambda_1 = 0.698526$$

$$\lambda_2 = 0.190362$$

Eigenvectors

<i>Carex pendula</i>	1.804984	-0.27453
<i>Carex remota</i>	-0.75147	-1.29819
<i>Deschampsia cespitosa</i>	-0.40151	1.002723

Example of calculation

The two determinants

$0.52849 - 0.698526$	-0.2344	-0.18232
-0.2344	$0.220085 - 0.698526$	-0.01396
-0.18232	-0.01396	$0.140313 - 0.698526$

$0.52849 - 0.190362$	-0.2344	-0.18232
-0.2344	$0.220085 - 0.190362$	-0.01396
-0.18232	-0.01396	$0.140313 - 0.190362$

are equal to zero.

The coordinates of the floristic variables are obtained by multiplying each eigenvector by the square root of the corresponding eigenvalue.

For example, the *Carex pendula* coordinate on the first axis is equal to:

$$1.804984 * \sqrt{0.698526} = 1.508567.$$

The coordinates for each relevé are obtained by summing the products of the scores of the species of the original table (here always equal to 1) by the corresponding values of the eigenvectors, all weighted by the marginal totals of each relevé.

For example, the coordinate of relevé 1 is equal to -0.40151, as it is the only present species. For relevé 2, we have $(1 * (-0.75147) + 1 * (-0.40151)) / 2$ or -0.57649.

Example of non-symmetric correspondence analysis

The calculations are done from the columns (here the relevés).

Weighted matrix

For presentation purposes, the matrix is permuted.

	<i>Carex remota</i>	<i>Carex pendula</i>	<i>Deschampsia cespitosa</i>	Somme
R1	-0.6923077	-0.92307693	1.61538461	0
R2	-0.6923077	0.57692307	0.11538461	0
R3	-0.6923077	0.57692307	0.11538461	0
R4	-0.6923077	0.57692307	0.11538461	0
R5	-0.6923077	0.57692307	0.11538461	0
R6	-0.6923077	0.57692307	0.11538461	0
R8	-0.6923077	0.57692307	0.11538461	0
R9	-0.6923077	0.57692307	0.11538461	0
R10	-0.6923077	-0.92307693	1.61538461	0
R11	-0.6923077	0.57692307	0.11538461	0
R12	-0.6923077	0.57692307	0.11538461	0
R13	-0.6923077	-0.92307693	1.61538461	0
R14	-0.6923077	0.57692307	0.11538461	0
R15	-0.6923077	-0.92307693	1.61538461	0
R16	2.3076923	-0.92307693	-1.38461539	0

R17	0.8076923	-0.92307693	0.11538461	0
R19	-0.6923077	2.07692307	-1.38461539	0
R23	0.8076923	-0.92307693	0.11538461	0
R24	2.3076923	-0.92307693	-1.38461539	0
R25	2.3076923	-0.92307693	-1.38461539	0
R26	2.3076923	-0.92307693	-1.38461539	0
R27	2.3076923	-0.92307693	-1.38461539	0
R28	0.3076923	0.07692307	-0.38461539	0
R29	0.8076923	-0.92307693	0.11538461	0

Note: null totals are rounded off.

Eigenvalues

$$\lambda_1 = 0.548992$$

$$\lambda_2 = 0.214318$$

Eigenvectors

- R1 -1.006226e-01 3.622473e-01
- R2 -1.509014e-01 -8.664303e-02
- R3 -1.509014e-01 -8.664303e-02
- R4 -1.509014e-01 -8.664303e-02
- R5 -1.509014e-01 -8.664303e-02
- R6 -1.509014e-01 -8.664303e-02
- R8 -1.509014e-01 -8.664303e-02
- R9 -1.509014e-01 -8.664303e-02
- R10 -1.006226e-01 3.622473e-01
- R11 -1.509014e-01 -8.664303e-02
- R12 -1.509014e-01 -8.664303e-02
- R13 -1.006226e-01 3.622473e-01
- R14 -1.509014e-01 -8.664303e-02
- R15 -1.006226e-01 3.622473e-01
- R16 3.516240e-01 -7.812254e-02
- R17 1.774848e-01 2.009065e-01
- R19 -1.127841e-01 -4.847791e-01
- R23 1.774848e-01 2.009065e-01
- R24 3.516240e-01 -7.812254e-02
- R25 3.516240e-01 -7.812254e-02

R26 3.516240e-01 -7.812254e-02
 R27 3.516240e-01 -7.812254e-02
 R28 7.979980e-02 -1.158478e-01
 R29 1.774848e-01 2.009065e-01

The calculation of the determinants is not presented here because it is too long.

The coordinates of the relevés are obtained by multiplying the value of the eigenvector on the corresponding axis, divided by the sum on the column of the relevé, all multiplied by the total sum of the table.

For example, the coordinate of relevé 1 on the first axis is equal to:

$$-0.1006226/\sqrt{1}*\sqrt{39} = -0.6283878.$$

The coordinates for each species are obtained as indicated on the following line:
 $\text{cosp} = \text{cosp} + \text{don}[i,j]*V[j, b]/\text{sqrt}(\text{hc}[j]*\text{sqrt}(\text{nbrsp}/(\text{Lamb}[b]*\text{somme})).$

We first calculate the constant $\text{sqrt}(\text{nbrsp} / (\text{Lamb}[b] * \text{sum}))$:

For the first axis, we have: 0.374321968

For instance, the coordinate of *Carex pendula* (species present in relevés 16, 17 and 23 to 29) on axis 1 is equal to:

For R16: $0.351624 * 0.3743219684$ (for $\text{hc} = 1$)
 For R17: $0.1744848 / \sqrt{2} * 0.3743219684$ (for $\text{hc} = 2$)
 For R23: $0.1774848 / \sqrt{2} * 0.3743219684$ (for $\text{hc} = 2$)
 For R24: $0.351624 * 0.3743219684$ (for $\text{hc} = 1$)
 For R25: $0.351624 * 0.3743219684$ (for $\text{hc} = 1$)
 For R26: $0.351624 * 0.3743219684$ (for $\text{hc} = 1$)
 For R27: $0.351624 * 0.3743219684$ (for $\text{hc} = 1$)
 For R28: $0.0797998 / \sqrt{3} * 0.3743219684$ (for $\text{hc} = 3$)
 For R29: $0.1744848 / \sqrt{2} * 0.3743219684$ (for $\text{hc} = 2$)

which is 0.816253