Chapter 2. The floristic, environmental data, the relevé tables and the relationship between variables

by Guy BOUXIN°

Contents

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

Introduction

A data table consists of two sets: the individuals and the characters related to these individuals (BOUROCHE & SAPORTA, 1980). In the vegetation studies, the data tables are called relevés tables in which the lines generally correspond to the species or to the mesological factors (characters or variables) and columns corresponding to the relevés (individuals). Presence or abundance data for species occur in many forms and diversity is even greater with environmental data. The data are sometimes used in the raw state or transformed in various ways.

It seems to us that the attention given to the nature of the data is often insufficient because the choices made in this step of the analysis are likely to greatly influence the sequence of operations. The adequacy between the data type and the multivariate analysis must be as perfect as possible and deviating from this rule inevitably leads to results that are difficult to interpret.

The data types in general

The observed characteristics are quantitative or qualitative. In the first case, the variables take their values on a numerical scale (BOUROCHE & SAPORTA, 1980). More precisely, a characteristic is quantitative when the set of values it takes on individuals is included in the set of real numbers (table 1); We can carry out the usual algebraic operations on these characters: addition, multiplication by a constant value, averaging, variance, etc.

Species\Relevés	1	2	3	4	5	6	7	8	9	10	11	12
Acacia senegal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.81	0.00
Acacia hockii	0.00	0.00	6.29	0.00	3.13	0.00	1.91	0.01	6.31	1.85	2.86	0.00
Acacia polyacantha	23.01	36.83	0.00	2.16	0.00	64.43	0.00	0.67	0.00	3.90	0.00	0.00
Acacia sieberana	14.60	0.00	0.00	0.00	0.00	0.00	3.47	0.00	0.00	0.00	0.00	0.00
Acacia brevispica	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.00
Dichrostachys cinerea	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28
Albizia amara	0.00	0.00	0.00	0.00	0.00	0.00	14.92	0.00	0.00	13.25	0.00	0.00
Lannea humilis	0.00	0.00	0.00	0.00	0.00	0.00	0.71	0.00	0.00	0.00	0.00	0.42
Lannea fulva	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.53
Rhus natalensis	0.00	0.00	0.00	0.00	0.00	1.93	0.00	0.00	7.96	0.00	0.00	2.32
Markhamia obtusifolia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86	3.13	0.00	0.44
Ximenia caffra	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.28
Canthium lactescens	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17
Pavetta gardeniifolia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.71
Teclea nobilis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55
Dombeya rotundifolia	0.00	0.00	0.00	0.00	1.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Grewia trichocarpa	0.00	0.00	0.00	0.00	0.00	0.76	0.00	0.00	0.00	0.00	0.00	8.74
Tricalysia ruandensis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72

Table 1. Basal areas 1,3 m (d.b.h.) of trees and shrubs in 12 savannah 250 m² relevés .

A characteristic is qualitative when taking non-numerical modalities: gender, profession, hierarchical level, etc. The modalities of a qualitative character can be ordered (hierarchical level, for example), then it is said to be qualitative ordinal. Otherwise, it is said to be qualitative nominal. Note that on a qualitative character represented by its modalities, algebraic operations are no longer lawful.

An example of ordinal qualitative data is given in table 2: In the same set of 12 savanna relevés, the variable "basal area" is transformed as follows:

- less than 5 dm² 1,
- from 5 to less than10 dm² 2,
- de 10 to less than15 dm² 3,
- de 15 to less than 20 dm² 4,
- 20 dm² and more 5.

Species \ Relevés	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Acacia senegal	0	0	0	0	0	0	0	0	0	0	3	0
Acacia hockii	0	0	2	0	1	0	1	1	2	1	1	0
Acacia polyacantha	5	5	0	1	0	5	0	1	0	1	0	0
Acacia sieberana	3	0	0	0	0	0	1	0	0	0	0	0
Acacia brevispica	0	0	0	0	0	0	0	0	0	0	0	2
Dichrostachys cinerea	0	0	0	0	0	0	0	0	0	0	0	1
Albizia amara	0	0	0	0	0	0	3	0	0	3	0	0
Lannea humilis	0	0	0	0	0	0	1	0	0	0	0	1
Lannea fulva	0	0	0	0	0	0	0	0	0	0	0	1
Rhus natalensis	0	0	0	0	0	1	0	0	2	0	0	1
Markhamia obtusifolia	0	0	0	0	0	0	0	0	1	1	0	1
Ximenia caffra	0	0	0	0	0	0	0	0	0	0	0	1
Canthium lactescens	0	0	0	0	0	0	0	0	0	0	0	1
Pavetta gardeniifolia	0	0	0	0	0	0	0	0	0	0	0	1
Teclea nobilis	0	0	0	0	0	0	0	0	0	0	0	1
Dombeya rotundifolia	0	0	0	0	1	0	0	0	0	0	0	0
Grewia trichocarpa	0	0	0	0	0	1	0	0	0	0	0	2
Tricalysia ruandensis	0	0	0	0	0	0	0	0	0	0	0	1

Table 2. Basal areas in table 1 transformed into an ordinal scale from 1 to 5.

With ordinal data, it is difficult to any significance to an average value, or to any other parameter, calculated on a set of relevés.

If all data greater than 1 are replaced by a value of 1, then presence-absence (or nominal qualitative) data are manipulated and a large part of the variability of the data is lost.

The floristic and mesological data

The floristic data

The kinds of data collected in finished surfaces are now presented.

The simplest data are, of course, the presence-absence data. In each relevé, the species are listed and assigned a rating of 1; the other species in the table are rated 0.

In many phytosociological studies, the coefficient of abundance-dominance is used according to the following scale (ROYER, 2009):

5: covering of the species between 75 and 100% of the total area,

4: covering of the species between 50 and 75% of the total area,

3: covering of the species between 25 and 50% of the total area,

2: covering of the species between 5 and 25% of the total area, or very abundant species, but of weak cover;

1: covering of the species less than 5 of the total surface, or abundant plant, but of very weak covering;

+: Scarce species, very weak covering,

Two other symbols are sometimes used:

r: very rare species ;

i: species represented by an isolated individual.

In integrated synusial phytosociology (GILLET, 2000), the dominance is estimated in proportion to the surface actually covered by all the plants of the considered synusia and not in proportion to the total surface of the relevé, contrary to the classical use presented above.

Several considerations come to mind:

- this coefficient incorporates two measures or estimates, namely the abundance corresponding to the number of individuals per unit area, and the dominance which is the total covering of the individuals of the species concerned,
- in the case of dominance, it is a visual estimate,
- the amplitude of the classes is not constant.

This coefficient, being very practical and adaptable to many situations, is widely used. In addition, a sociability coefficient is often associated with abundance-dominance. It is simply not taken into account in statistical analyses.

Table 3 presents an example of phytosociological data (FERREZ, 2009). Eleven species, present only once, are not cited. Only 28.21% of the cells in the table are occupied, which is usual in this kind of table.

Species\Relevés	8	13	12	6	7	10	9	1	5	4	3	2	11
Characteristic combination													
Polypodium vulgare	+	+	+	+	2	2	1	+	2	+	1	1	+
Asplenium scolopendrium	1	1	1	1	2	1	2	1	1	1	2	1	.
Moehringia trinervia	+	1	+	+	+	+	+	.	.	.	.	.	.
Cardamine impatiens	+	+	1	.	.	.	.	.	.	.	.	.	.
Species of the higher units													
Asplenium trichomanes subsp, *quadrivalens*	2	+	+	1	1	1	1	2	1	1	1	.	2
Geranium robertianum subsp, *robertianum*	2	2	2	2	.	1	1	+	1	+	1	1	+
Cardaminopsis arenosa subsp, *borbasii*	1	.	.	1	+	2	.	1	+	.	.	.	1
Cystopteris fragilis	.	.	.	+	.	.	.	1	.	.	.	+	.
Mycelis muralis	.	.	.	.	.	1	+	.	.	.	.	.	
Other species													
Hedera helix	2	2	2	+	1	.	.	1	.	2	1	2	1
Moehringia muscosa	.	1	.	.	.	.	.	+	.	.	.	2	.
Carex digitata	.	+	.	.	.	.	.	.	.	.	.	.	+
Arabis turrita	.	.	.	.	.	1	.	.	.	.	.	.	+

Table 3. *Moehringio trinervae – Geranietum robertiani* Gillet ass. nov. *hoc loco*.

In quantitative studies, several measures exist (GOUNOT, 1969):

The **density** is the number of individuals per unit area. This measure is suitable for well-individualized species such as many woody plants or annual herbaceous plants.

The phytomass is the plant mass present in the community. It is expressed in kg/m². Its measure, if it wants to be precise, is time consuming and non recommandable because it involves the destruction of vegetation.

The cover

One distinguishes (FEHMI, 2010):

The **aerial cover** is the uppermost surface of the vegetation and is expressed as a percentage area occupied by each species. The sum of the aerial cover of all species plus the unoccupied area (exposed bare ground) equals 100 %.

The **species cover** is the cover for each plant species evaluated independently from the cover of other species. It is the independent aerial cover of each species, as if no others occurred in the sample area, expressed as a percentage of the area occupied by each species. The sum of the combined species covers can exceed 100 %, but each individual species cannot have a cover that exceeds 100 %.

The **basal cover** (called basal area in forestry) is the cover where the stem meets the surface of the ground and is in common use in areas dominated by bunch grasses. In forestry, the stem area is measured at 1.3 m above the ground (or d.b.h. = diameter at breast height); the area of the corresponding disc is then calculated. For a given species, abundance is the sum of the basal areas in the relevé; it is expressed in dm² in the relevé or in m² per hectare. This measure is easy and convenient enough to be used in a fairly large study.

Other types of cover exist, but are hardly used in vegetation studies; they are given by ANDERSON (1986) and FEHMI (2010).

The **frequency** is the percentage of plots containing a species in relation to the total number of plots studied. The number, shape or dispersion of the plots in the relevé (with the exception of extensive observation) affects the amplitude of the data. The frequency in a relevé can also be estimated by the technique of the quadrat point which is most often materialized by a needle. A species is present at a point if there is contact between a vertical needle sliding vertically in a frame and a given species. This technique is practical for species forming tufts such as the *Poaceae* and *Cyperaceae*, which is impractical in other cases.

Further information can be found in van der MAAREL (2005).

In the case of phytosociological relevés, the coefficient of abundance-dominance is used for all species. In the case of quantitative studies (BOUXIN, 1975 & 1976), several abundance criteria are used depending on the species: basal area for trees, frequency for trees and shrubs less than 1.5 m high, frequency (quadrat points) for bunched species, a simple presence for the others. This does not facilitate statistical analysis.

The mesological data sensu lato

The mesological data reflect the characteristics of very different parameters that relate both to the general environmental characteristics of a relevé or the characteristics observed inside a relevé. Physiological, geological, pedological characteristics are thus described using qualitative, quantitative or ordinal variables. The physical and chemical characteristics of a soil or water bearing or surrounding vegetation are measured in the field or in laboratories and expressed in very different units with widely varying amplitudes of variation. In the same table, one can find (BOUXIN, 1975) :

- the general slope of the site, a quantitative character varying from 0 to 90 °,
- the exposure, represented by four qualitative variables (1 or 0 respectively for the north, south, east or west exposures),
- the altitude, quantitative character, in meters,
- the topography, represented by three qualitative variables (1 or 0, respectively for convex, plane or concave topographies),
- the degree of closure of the site, represented by three qualitative variables (1 or 0, respectively open, moderately closed or closed);
- the soil depth, represented by three qualitative variables (1 or 0, respectively, for lithosol, moderately deep soil or deep soil),
- the soil type, represented by three qualitative variables (1 or 0, respectively, for a recent tropical soil, ferralsol or ferrisol);

- the importance of the rock outcrops, a quantitative character in the form of a contact count on a set of 125 regularly distributed points,
- the importance of bare soil, a quantitative character in the form of a contact count on a set of 125 regularly distributed points,
- the thickness of the humiferous horizon, a quantitative character measured in centimeters (average of several measurements),
- surface and depth structure and texture, each time represented by several qualitative variables (1 or 0),
- the water holding capacity (field capacity) of soil samples taken in the A0 horizon and in the A1 horizon, quantitative characters expressed in %,
- the pH of the A0 and A1 horizons, quantitative characters measured in the field, with an accuracy of $1/10^{th}$ of a unit,
- the sum of the total exchangeable cations in the A0 and A1 horizons, quantitative characters expressed in meq/100 g of air-dried sol,
- the degree of saturation in bases of horizons A0 and A1, quantitative characters expressed in%,
- the percentage of organic matter in the horizons A0 and A1 of the horizons A0 and A1, quantitative characters.

Other examples are given in the appendix.

The standardization and the transformation of the data

Since the variables collected in one study are sometimes expressed in very different units or even if they are all of the same nature, the variability is sometimes enormous and makes the analysis of the data little useful, so certain variables or particular values take an excessive importance compared to the others.

.

There are two ways of modifying the data (PODANI, 2000): standardization and transformation. Standardization modifies the data by using statistics calculated from the data itself. These are, for example the variance, the amplitude, the average, the total or simply the maximum value. Standardization is often used to compensate for differences in weight or units between data. Data transformation, in the strict sense, uses mathematical functions whose parameters do not depend on the data.

The standardization of the variables

Examples:

Centering: the mean is subtracted from each value.

$$x'_{ij} = x_{ij} - \overline{x_i}$$

Linear standardization: the values of the variable *i* are multiplied by a constant, a statistic that is derived from all the observations for the variable (*e.g.*, range, standard error).

Standardization by range: the variable is rescaled to the interval [0,1].

$$\left[x_{ij} - \min_j \{x_{ij}\}\right] / \left[\max_j \{x_{ij}\} - \min_j \{x_{ij}\}\right]$$

S**tandardization by standard deviation**:

$$x'_{ij} = \left\{\left(x_{ij} - \overline{x_i}\right)^2\right\} / s_i \text{ with}$$

$$s_i = \left[\frac{\sum_{j=1}^{m}\left(x_{ij} - \overline{x_i}\right)^2}{m-1}\right]^{1/2}$$ that is the standard deviation.

The numerator is the sum of squared deviations from the mean. This approach is recommended if the variables are measured in very different units (pH, concentration, temperature). The correlation coefficient includes this operation.

Standardization with the total:

$$x'_{ij} = x_{ij} \Big/ \sum_{j=1}^{m} x_{ij}$$

The variables having large values are diminished, and those with small values are increased in importance.

Standardization by maximum:

$x'_{ij} = x_{ij} / \max_j \{x_{ij}\}$. All values are divided by the maximum of the variables found in the sample.

S**tandardization to unit vector length**:

$$x'_{ij} = x_{ij} \Big/ \left\{\sum_{j=1}^{m} x_{ij}^2\right\}^{1/2}$$. In the variable space, the objects are at the endpoints of vectors directed from the origin. The sum of squares becomes 1 for each variable.

Transformation

Examples:

Binarization:

Quantitative data are converted into qualitative data (presences-absences for vegetation),

$x'_{ij} = 1$, si $x_{ij} > p$,

$x'_{ij} = 0$, si $x_{ij} \leq p$,

p is often equal to 0.

Logarithmic transformation: each value is replaced by its logarithm (base 10 or *e*). This transformation diminishes large absolute differences. It is also used to linearize certain relations between variables.

Arcus sinus transformation:

$x'_{ij} = \arcsin x_{ij}$.

This function converts variables with a range [0,1]. It is often associated with a square root transformation.

The transformation of phytosociological data

However, the transformation of these data is essential to integrate mathematical operations. In the calculations of the multivariate analyses, sums are made on the rows and sometimes on the columns of the tables. But what significance can be attributed, for example, to a sum on a column of a table comprising, as is often the case,

- species, some of which are represented by abundance,
- others by dominance
- and a group of very poorly represented species?

Since the coefficients 1 and 2 do not have the same range as the coefficients 3, 4 and 5, can we conclude that 1 + 2 + 3 = 6 ?

The following transformation was proposed by GILLET (2000): from the code r which takes the value 0.1, from the code + which takes the value 0.5 and from the other codes 1, 2, 3, 4 and 5, mean cover is calculated, which gives table 4.

AD code	Digital AD	Mean cover
r	0,1	0,03%
+	0,5	0,30%
1	1	3%
2	2	14%
3	3	32%
4	4	57%
5	5	90%

Table 4. Table of correspondence between the abundance-dominance code (AD code), the quantitative abundance-dominance index (Digital AD), and the mean cover.

Code 2 is sometimes also subdivided into three:

- 2m: very abundant elements, less than 5% cover,
- 2a: cover between 5 and 12.5%, any abundance,
- 2b: cover between 12.5 and 25%, any abundance.

From all these codes, van der MAAREL created the following scale:

Scale of abundance-dominance	van der MAAREL's ordinal scale
r	**1**
+	**2**
1	**3**
2m	**4**
2a	**5**
2b	**6**
3	**7**
4	**8**
5	**9.**

Standardization by the individuals

Various standardizations comparable to those of variables are sometimes applied to individuals. For more details, see PODANI (2000).

Double standardization

Each value is divided by both the sum on the columns and the sum by the rows. This standardization is used in correspondence analysis and we will inevitably return to the subject.

Calculations

The various standardizations and transformations are facilitated by softwares like VegAna (see chapter 13)

Types of multidimensional data tables

We call multidimensional data the set of values of a certain number of variables on an individual (FOUCART, 1982). A multidimensional data table consists of a set of variables measured over a set of individuals.

The tables individuals x quantitative characters

This kind of table is one of the simplest: the characters of individuals are quantitative variables with real values, not necessarily continuous. The term x_{ij} is therefore a real number, representing the measure of the variable x_i on the individual j.

The ordinal data tables

In this type of table, the characters of individuals are ordinal variables on which it is much more delicate to perform mathematical operations. The abundance-dominance scale is an ordinal scale.

The presence tables

Characters of individuals are represented by qualitative variables of type 0-1. This type of table is common with vegetation data when abundance is impossible to estimate. Only the presence of species, noted 1, is registrable. Tables generally include a very large number of zeros, which makes their analysis tricky.

The contingency tables

In a contingency table, the term $n_{i,j}$ of the i^{th} row and the j^{th} column is the number of individuals possessing the modality i of the character 1 and the modality j of the character 2. In principle, the modalities are exclusive and exhaustive: an individual of the population cannot possess more than one modality of the same character, it possesses one and only one. Lines and columns play a similar role.

The logical tables

The logical tables indicate, for each individual, the membership of a particular group or, what is equivalent, the modality of a qualitative variable that it possesses. The coding used is the logical coding: the membership is represented by 1, the non-belonging by zero. The term x_{ij} is equal to 1 or 0 according to whether the individual j belongs to the group i (or takes the modality i to the qualitative character) or not. In each column, only one term is equal to 1.

Let us consider, for example, the concentrations of ammonium, nitrite and nitrate in 10 sites in a stream (table 5). Let us transform the first line into a logical table (table 6): the ammonium concentration is represented by two modalities depending on whether it is lower or higher than or equal to the average of the measured concentrations. In the first line, the concentrations below or equal to the mean take the value 1, the

others the value 0. In the second, the value 1 is attributed to the above average concentrations, the value 0 to the others. In each column, only one term is equal to 1. In other situations, the concentrations could be represented by a greater number of modalities, but always with the modality 1 represented only once in each column.

	1	2	3	4	5	6	7	8	9	10
ammonium mg/l	0	0	0	0	0,4	0,2	0	0	0	0
nitrite mg/l	0,0125	0,15	0,15	0,1	0,3	0,3	0,3	0,3	0,3	0,15
nitrate mg/l	0	5	5	10	10	5	10	10	17,5	10

Table 5. Mesurements of three chemical parameters of the water from ten Bocq sites.

NH_4^+ < average	1	1	1	1	0	0	1	1	1	1
NH_4^+ ≥ average	0	0	0	0	1	1	0	0	0	0

Table 6. Logical table of the parameter ammonium.

The complete disjunctive tables

A complete disjunctive table is formed by the juxtaposition of several logical tables. Each logical table corresponds to a partition of the set of individuals: the term $x_{i1,j}$ is equal to 1 or 0 depending on whether the individual belongs to group 1 or not, the terms x $_{i2,j}$, $x_{i3,j}$ being defined in a similar manner. Each column of the complete disjunctive table contains as many times the value 1 as there are logical tables.

Let us form table 7. The first two lines of table 6 are reproduced. Then, for the nitrite, two lines are created: one for concentrations lower than or equal to the average (**1** for these values and **0** for the others) and a second for concentrations higher than the average (**1** for these values and **0** for the others). For nitrate, three lines are created: one for non-measurable concentrations (**1** for non-measurable values and **0** for others), one for measurable concentrations and below or equal to mean (**1** for these values and **0** for the others) and a third for above-average concentrations (**1** for these values and **0** for the others).

NH_4^+ < average	1	1	1	1	0	0	1	1	1	1
NH_4^+ ≥ average	0	0	0	0	1	1	0	0	0	0
NO_2^- < average	1	1	1	1	0	0	0	0	0	1
NO_2^- ≥ average	0	0	0	0	1	1	1	1	1	0
NO_3^- = 0	1	0	0	0	0	0	0	0	0	0
NO_3^- > 0 et < average	0	1	1	0	0	1	0	0	0	0
NO_3^- ≥ average	0	0	0	1	1	0	1	1	1	1

Table 7. Complete disjunctive table constructed from table 5.

In each column, the sum is equal to three. This table is thus formed by the juxtaposition of three logical tables.

Particular case of phytosociological tables

With abundance-dominance, one can cut each line into as many lines as there are levels of abundance: one line for the +, one for the 1, and so on up to 5, without considering the abundances 0 (species not present in a relevé). Indeed, in a phytosociological table, the number of absences is generally clearly greater than those present (whatever the abundance). If a complete disjunctive table is subjected to a non-symmetric correspondence analysis, it will be largely dominated by absences, which is not the objective of the analysis. When a species is absent from a relevé, the sum on the rows pertaining to that species is zero and not 1 as in a logical table. This type table is called a simple disjunctive table.

The joined tables

Many floristic and mesological tables are mixed tables, containing continuous quantitative variables (i.e. basal areas), numbers of individuals, qualitative variables (simple presence of poorly represented species), mesological tables are often mixed and include quantitative variables but qualitative or ordinal variables, which makes the analysis difficult; some variables are initially transformed variables such as the pH which is a cologarithm.

In phytosociological tables, there is often a significant proportion of species represented by their simple presence (with a score of 1 or 0 if it is absent) creating a large number of empty cells.

Distance and proximity tables

These are square tables constructed from indices of distance or proximity. We also speak of similarity or dissimilarity.

An index of distance (or dissimilarity) is a symmetric function with real and positive values defined between two individuals: the more the individuals i and i' resemble each other, the lower the value of this index. With an index of proximity (or similarity), the more similar the individuals i and i', the higher the value of this index. A proximity index can take negative values.

The relationships between variables

Many methods rely on the analysis of linear dependencies between the observed characters. In the example of BOUROCHE & SAPORTA (1980), we find here a relation in the form of a narrow and elongated cloud along a straight line, between the price of apartments and their area (figure 1).

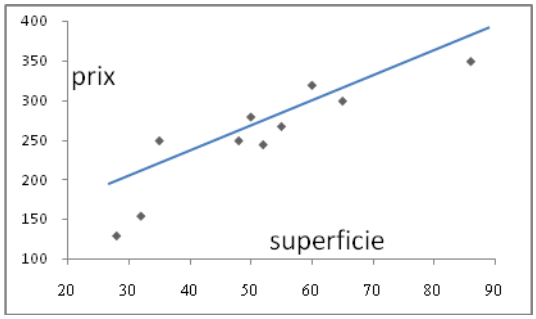


Figure 1. Relation between area and the price of apartments.

However, relationships between parameters recorded under natural conditions do not always occur in this form. Let us take first the case of relations between mesological parameters.

The first example involves chemical analyses of soil in savanna (Akagera Park, Rwanda) in the A1 horizon of 70 samples (figure 2).

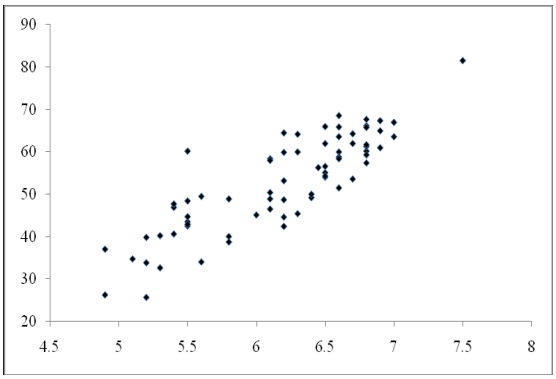


Figure 2. Relationship between the pH of a soil and the degree of saturation in bases. On the abscissa, the pH, on the ordinate, the degree of base saturation.

This relationship is clear, there is a fairly clear relationship between the two parameters and a fit by linear regression is possible.

The second example relates to chemical analyses of water samples taken once a month for one year in a river. These are two usual parameters for defining the biological quality of water, namely ammoniacal nitrogen and orthophosphate (figure 3).

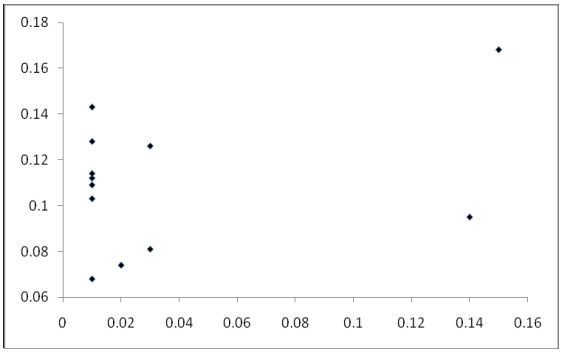
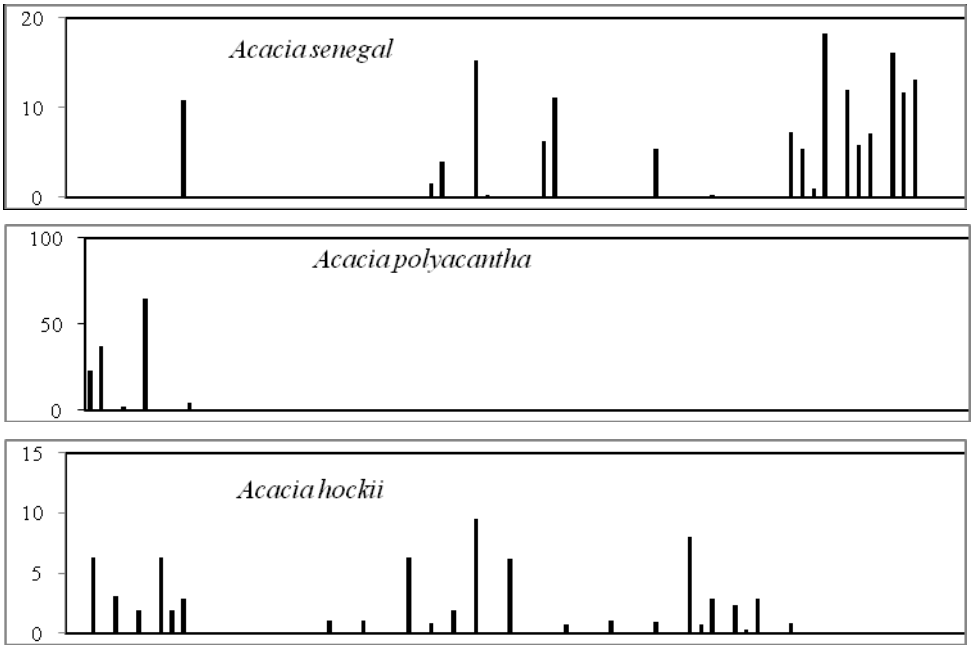
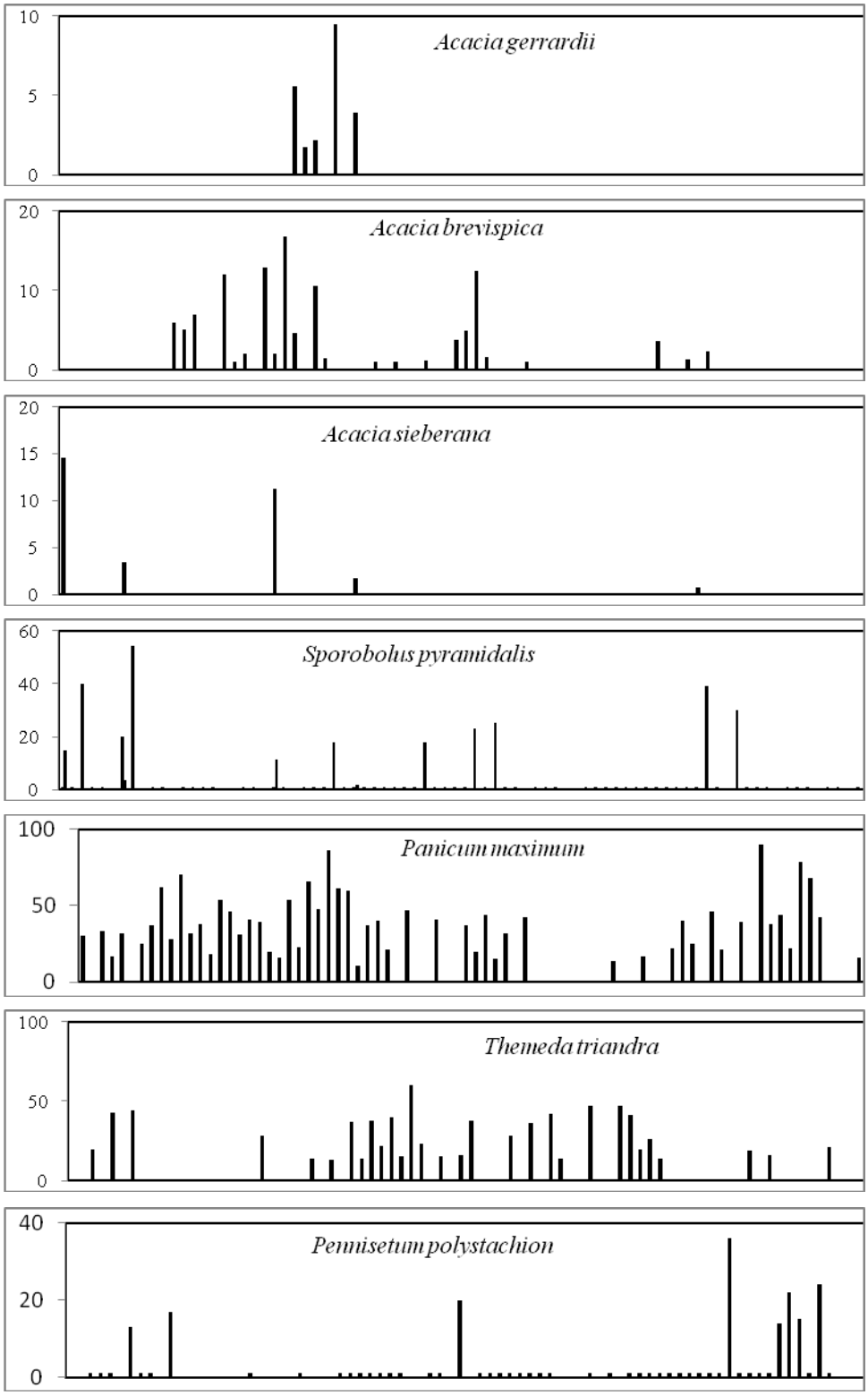


Figure 3. Relationship between ammoniacal nitrogen and orthophosphate, in mg / l of nitrogen and phosphorus.

On the abscissa, the concentration of ammoniacal nitrogen (mg/l of nitrogen), on the ordinate, the concentration of orthophosphate (in mg/l of phosphorus). This relationship is less clear-cut and no adjustment seems possible with such data. The relations between the mesological parameters are far from being as clear as in figure 2 and less marked relationships as in figure 3 are frequent.

The relationships between species abundances also show a multitude of possible figures (4 and 5):




Acacia gerrardii
Acacia brevispica
Acacia sieberana
Sporobolus pyramidalis
Panicum maximum
Themeda triandra
Pennisetum polystachion

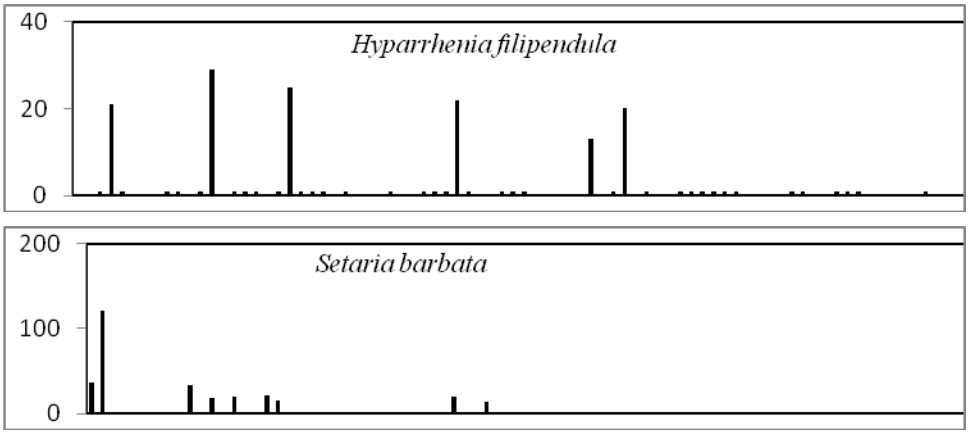


Figure 4. Variation of abundance along a transect, in a set of 80 250 m² relevés, between Lake Ihéma and Kionja Hill (Akagera National Park). On the ordinate: the basal areas, in dm² for the *Acacia* and the frequency of contact with 125 points for the grasses.

These data are illustrated as follows.

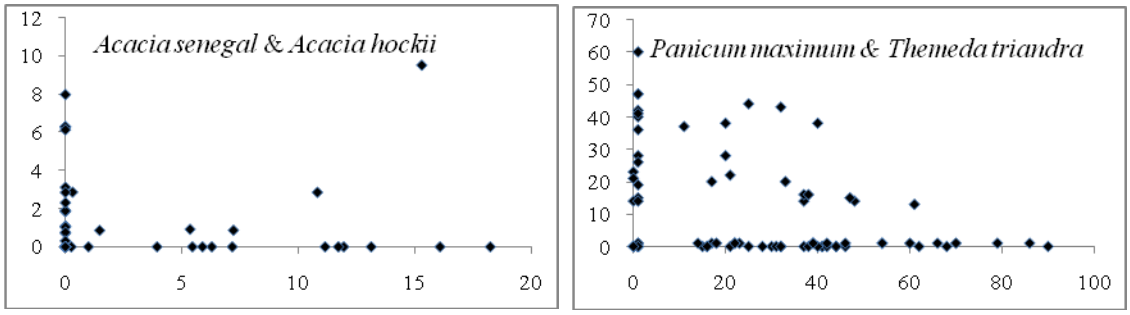


Figure 5. Relationship between the basal area of the tree *Acacia senegal* and the abundance of *Panicum maximum* (tufted grass), estimated by counting in 125 contact points with a needle.

In conclusion, it seems very difficult to associate a simple distribution model for species or imagine that we will find ourselves with clear relationships between specific abundances. Exploring the links between dispersion of species and environmental factors is not simplified.

Continuum and discontinuum

In studies in plant ecology, one can not go further in the study of the relationships between vegetation and the environment without taking into account the dependence between certain developments in concepts on vegetation, methods of mathematical analysis and knowledge of environmental processes.

In the second half of the twentieth century, there was a controversy between proponents of vegetation representation in the form of continuums rather than discrete communities (AUSTIN *in* van der MAAREL 2005). The concept of the continuum is based on the principle of specific individuality, with each species being distributed in relation to all environmental factors, including interactions with other species. There are not two species that exhibit the same dispersion. From this we deduce the principle of community continuity,

which evolves along continuous environmental gradients, with gradual changes in species populations along the gradients.

From this concept, was born the direct ordination method also called direct gradient analysis by WHITTAKER (1956). It is the analysis of dispersions of species and collective properties (such as specific richness) in relation to environmental variables conventionally considered as environmental gradients.

The next step was the use of multivariate analyses to determine the major gradients present in the vegetation data themselves. The resulting graphical representations summarize the main axes of variation present in a similarity matrix from the data table. On the axes of the graphs, one associates either environmental gradients, or stages of succession over time, or different grazing regimes.

There is no real consensus on how to conceptualise the representation of species or plant communities along gradients. The performance of species along a simple environmental gradient can take very diverse forms. The models of nonlinear responses of species along gradients were studied by AUSTIN (1976, 1980 and 2005, *in* van der MAAREL p.72) and AUSTIN *et al*. (1984). Several models have been proposed such as the GAUSS curve, the polynomial functions, the β or γ functions or the ecological response curve incorporating the competitions between species.

This makes statistical analysis very difficult because, in addition, we must treat very large variabilities, much larger than those found in the field of experimentation. Multivariate analyses must address three levels of complexity:

- the multidimensional nature of environmental gradients and species responses,
- the curvilinear and non-monotonic character of the responses of species,
- the significant differences between species distribution patterns and idealized models.

In addition, an important limitation is the small proportion of space sampled, even in large datasets (AUSTIN *et al*., 1984).

According to Austin (1980), progress in the development of an explicit model of vegetation/environment relations is necessary, even more important than in the development of multivariate analysis techniques.

Discussions around this concept of gradient inevitably fall into several chapters.

Conclusions

We are therefore confronted with an analysis of tables that are very difficult to carry out, presenting enormous variabilities, sometimes with types of variables making certain illicit mathematical operations and particular properties such as the very large number of empty cells, mainly in phytosociological tables. The used techniques must take into account the particular properties of the relevés tables. The ease of calculation, access to a large number of transformations and standardizations, should not lead us to undertake any analysis. In the chapters that follow, we will see how the data types and results of the analyses are intimately linked.

January 2024

References

ANDERSON, E.W. (1986). A guide for estimating cover. *Rangelands* **8**: 236-238.

AUSTIN, M.P. (1976). On non-linear species response models in ordinations. *Vegetatio* **33**: 33-41.

AUSTIN, M.P. (1980). Searching for a model for use in vegetation analysis. *Vegetatio* **42**: 11-21.

AUSTIN, M.P. (2005). *Vegetation and environment: discontinuities and continuities*. In van der MAAREL, E. *Vegetation Ecology*. Blackwell Publishing. 52 -105.

AUSTIN, M.P., CUNNINGHAM, R. B. & FLEMING, P. M. (1984). New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* **55**: 11-27.

BOUROCHE, J.–M. & G. SAPORTA (1980). *L'analyse des données*. Presses Universitaires de France. Collection Que sais-je ? 127 pp.

BOUXIN, G. (1975). Ordination and classification in the savanna vegetation of the Akagera park (Rwanda, Central Africa). *Vegetatio* **29**: 155-167.

BOUXIN, G. (1976). Ordination and classification in the upland Rugege forest (Rwanda, Central Africa). *Vegetatio* **32**: 97-115.

FEHMI, J. (2010). Confusion among three common plant cover definitions may result in data unsuited for comparison. *Journal of Vegetation Science* **21**: 273-279.

GILLET, F. (2000). *La phytosociologie synusiale intégrée. Guide méthodologique*. Université de Neuchâtel, Laboratoire d'écologie végétale et de phytosociologie. 68 pp.

FERREZ, Y. (2009). Contribution à l'étude phytosociologique des groupements végétaux des parois calcaires (classe des *Asplenietea trichomanis* (Br.-Bl. *in* Meier & Br.-Bl. 1934) Oberdorfer 1977) du massif jurassien et de la Franche-Comté. *Les Nouvelles Archives de la Flore jurassienne* 7: 123-158.

FOUCART, T. (1982). *Analyse factorielle. Programmation sur ordinateur*. Masson, Paris, 243 pp.

GOUNOT, M. (1969). *Étude quantitative de la végétation*. Masson et Cie. 314 pp.

PODANI, J. (2000). *Introduction to the exploration of multivariate data*. Blackhuys Publishers, Leiden. 407 pp.

ROYER, J.-M. (2009). Petit précis de phytosociologie sigmatiste. *Bulletin de la Société Botanique du Centre-Ouest. Nouvelle série. Numéro spécial* **33** : 1-86.

van der MAAREL, E. (2005). *Vegetation Ecology*. Blackwell Publishing. 395 pp.

WHITTAKER, R. (1956). Vegetation of the Great Smoky Mountains. *Ecological Monographs,* **26**: 1-80.