Chapitre 2. Les données floristiques, mésologiques, les tableaux de relevés et les relations entre variables

par Guy BOUXIN°

Sommaire

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

Introduction

Un tableau de données est constitué de deux ensembles : les individus et les caractères relatifs à ces individus (BOUROCHE & SAPORTA, 1980). Dans les études de végétation, les tableaux de données s'appellent des tableaux de relevés dans lesquels les lignes correspondant généralement aux espèces ou aux facteurs mésologiques (caractères ou variables) et de colonnes correspondant aux relevés (individus). Les données portant sur la présence ou l'abondance des espèces se présentent sous de nombreuses formes et la diversité est encore plus grande avec les données mésologiques. Les données sont parfois utilisées à l'état brut ou transformées de diverses manières.

Dans l'ensemble, il nous paraît que l'attention portée à la nature des données est souvent insuffisante car les choix que l'on fait dans cette étape de l'analyse sont susceptibles d'influencer grandement la suite des opérations. L'adéquation entre le type de données et l'analyse multivariée doit être aussi parfaite que possible et s'écarter de cette règle conduit inévitablement à des résultats difficiles à interpréter.

Les types de données en général

Les caractères observés sont quantitatifs ou qualitatifs. Dans le premier cas, les variables prennent leurs valeurs sur une échelle numérique (BOUROCHE & SAPORTA, 1980). Plus précisément, un caractère est quantitatif lorsque l'ensemble des valeurs qu'il prend sur les individus est inclus dans l'ensemble des nombres réels (tableau 1) ; on peut effectuer sur ces caractères les opérations algébriques habituelles : addition, multiplication par une valeur constante, calcul de moyenne, variance, *etc*..

Espèces\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

Tableau 1. Surface terrière à 1,3 m d'arbres et d'arbustes dans 12 relevés de 250 m² de savane.

Un caractère est qualitatif lorsqu'il prend des modalités non numériques : sexe, profession, niveau hiérarchique, *etc*.. Les modalités d'un caractère qualitatif peuvent être ordonnées (niveau hiérarchique, par ex,), on dit alors qu'il est qualitatif ordinal. Sinon, on dit qu'il est qualitatif nominal. Remarquons que sur un caractère qualitatif représenté par ses modalités, certaines opérations algébriques ne sont plus licites.

Exemple de données qualitatives ordinales (tableau 2) : dans le même ensemble de relevés de savane, la variable « surface terrière » est transformée de la manière suivante :

- moins de 5 dm² 1,
- de 5 à moins de10 dm² 2,
- de 10 à moins de15 dm² 3,
- de 15 à moins de 20 dm² 4,
- 20 dm² et plus 5.

Espèces \ 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

Tableau 2. Surfaces terrières du tableau 1 transformées suivant une échelle ordinale de 1 à 5.

Il est difficile de donner une signification à une valeur moyenne, ou à tout autre paramètre, calculé sur un ensemble de relevés.

Si on remplace, dans le tableau 2 toutes les données supérieures à 1 par la valeur 1, on manipule alors des données de présence-absence (ou qualitatives nominales) et on perd une partie importante de la variabilité des données.

Les données floristiques et mésologiques

Les données floristiques

Nous présentons ici les données récoltées dans des surfaces finies.

Les données les plus simples sont évidemment les données de **présence-absence**, Dans chaque relevé, on dresse la liste des espèces et on leur attribue la cote **1** ; les autres espèces du tableau reçoivent la cote **0**.

Dans beaucoup d'études phytosociologiques, le coefficient d'**abondance-dominance** est utilisé selon l'échelle suivante (ROYER, 2009) :

5 : recouvrement de l'espèce compris entre 75 et 100 % de la surface totale,

4 : recouvrement de l'espèce compris entre 50 et 75 % de la surface totale,

3 : recouvrement de l'espèce compris entre 25 et 50 % de la surface totale,

2 : recouvrement de l'espèce compris entre 5 et 25 % de la surface totale, ou espèce très abondante, mais de recouvrement faible ;

1 : recouvrement de l'espèce inférieur à 5 % de la surface totale, ou plante abondante, mais de recouvrement très faible ;

+ : espèce peu abondante, à recouvrement très faible.

Deux autres symboles sont parfois utilisés :

r : espèce très rare ;

i : espèce représentée par un individu isolé.

D'autres symboles sont parfois aussi rencontrés.

En phytosociologie synusiale intégrée (GILLET, 2000), les dominances sont estimées en proportion de la surface effectivement recouverte par l'ensemble des végétaux de la synusie relevée et non en proportion de la surface totale du relevé, contrairement à l'usage classique présenté ci-dessus.

Plusieurs considérations viennent à l'esprit :

- ce coefficient intègre deux mesures ou estimations à savoir l'abondance qui correspond au nombre d'individus par unité de surface, et la dominance qui est le recouvrement total des individus de l'espèce considérée,
- dans le cas de la dominance, il s'agit d'une estimation visuelle,
- l'amplitude des classes n'est pas constante.

Ce coefficient, par son caractère très pratique et adaptable à de nombreuses situations, est largement utilisé. En plus, un coefficient de sociabilité est souvent associé à l'abondance-dominance. Il n'est tout simplement pas pris en considération dans les analyses statistiques.

Voici, dans le tableau 3, un exemple de données phytosociologiques (FERREZ, 2009). Onze espèces, présentes une seule fois, ne sont pas citées. Seulement 28,21 % des cellules du tableau sont occupés, ce qui est habituel dans ce genre de tableau.

Espèces\Relevés	8	13	12	6	7	10	9	1	5	4	3	2	11
Combinaison caractéristique													
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	.	.	.	.	.	.	.	.	.	.
Espèces des unités supérieures													
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	+	.	.	.	.	.	
Autres espèces													
Hedera helix	2	2	2	+	1	.	.	1	.	2	1	2	1
Moehringia muscosa	.	1	.	.	.	.	.	+	.	.	.	2	.
Carex digitata	.	+	.	.	.	.	.	.	.	.	.	.	+
Arabis turrita	.	.	.	.	.	1	.	.	.	.	.	.	+

Tableau 3. *Moehringio trinervae – Geranietum robertiani* Gillet ass. Nov. *hoc loco*.

Dans les études quantitatives, plusieurs mesures sont utilisées (GOUNOT, 1969):

La **densité** qui est le nombre d'individus par unité de surface. Cette mesure convient pour les espèces bien individualisées comme beaucoup de plantes ligneuses ou des herbacées annuelles.

La **phytomasse** est la masse végétale présente dans la communauté. Elle s'exprime en kg/m². Sa mesure, si elle se veut précise, est longue et peu recommandable car elle entraîne la destruction de la végétation.

Le **recouvrement**

On distingue (FEHMI, 2010):

Le **recouvrement aérien** est la proportion de chaque espèce vue de la surface la plus élevée de la végétation, ce qui correspond à une vue aérienne. Il s'exprime par le pourcentage d'aire occupé par chaque espèce. La somme des recouvrements aériens de toutes les espèces plus le sol inoccupé est égale à 100 %.

Le **recouvrement spécifique** est le recouvrement de la partie supérieure de chaque espèce, indépendamment de la couverture surplombante des autres espèces. La somme des recouvrements des espèces peut dépasser 100 % mais chaque espèce ne peut avoir de recouvrement dépassant 100 %.

Le **recouvrement foliaire** est le recouvrement de toutes les couches de chaque espèce depuis la surface la plus élevée jusqu'au niveau du sol. La somme des recouvrements spécifiques, comme le recouvrement de chaque espèce peut dépasser 100 %.

Le **recouvrement basal** (appelé surface terrière dans les ouvrages de dendrométrie forestière) est la surface occupée par les parties aériennes des individus de l'espèce au niveau du sol ou, dans le cas des arbres, à hauteur de poitrine (diameter breast height = d.b.h. des auteurs anglo-saxons). En général, on mesure la circonférence à 1,3 m du sol et on calcule ensuite l'aire du disque correspondant. Pour une espèce donnée, l'abondance est la somme des surfaces terrières dans le relevé ; on l'exprime en dm² dans le relevé ou en m² par hectare. Cette mesure est aisée et assez pratique que pour être utilisée dans une étude assez importante.

D'autres types de recouvrement existent mais sont peu utilisés dans des études de végétations ; ils sont présentés par ANDERSON (1986) et FEHMI (2010).

La **fréquence** est le pourcentage de placettes contenant une espèce par rapport au nombre total de placettes étudiées. Le nombre, la forme ou le mode de dispersion des placettes dans le relevé (à l'exception d'une observation exhaustive) ont une incidence sur l'amplitude des données. La fréquence dans un relevé peut aussi être estimée par la technique du point quadrat qui est le plus souvent matérialisé par une aiguille. Une espèce est présente en un point s'il y a contact entre une aiguille verticale glissant verticalement dans un bâti et une espèce donnée ; cette technique est pratique pour les espèces formant des touffes comme les *Poaceae*, peu pratique dans les autres cas.

D'autres informations se trouvent dans van der MAAREL (2005).

Dans le cas de relevés phytosociologiques, le coefficient d'abondance-dominance est utilisé pour toutes les espèces. Dans le cas d'études quantitatives (BOUXIN, 1975 & 1976), plusieurs critères d'abondances

sont utilisés suivant les espèces : surface terrière pour les arbres, fréquence pour les arbres et arbustes de moins de 1,5 m de haut, fréquence (points quadrats) pour les espèces en touffe (*Poaceae* et *Cyperaceae*), simple présence pour les autres. Cela ne facilite pas l'analyse statistique.

Les données mésologiques sensu lato

Les données mésologiques traduisent les caractéristiques de paramètres très différents portant à la fois sur des caractéristiques de l'environnement général d'un relevé ou des caractéristiques observées au sein d'un relevé. Des caractéristiques physionomiques, géologiques, pédologiques sont ainsi décrites au moyen de variables qualitatives, quantitatives ou ordinales. Les caractéristiques physiques et chimiques d'un sol ou d'une eau portant ou entourant la végétation sont mesurées sur le terrain ou en laboratoire et exprimées dans des unités différentes, avec des amplitudes de variations très diverses. Dans un même tableau, on peut trouver (BOUXIN, 1975):

- la pente générale du terrain, caractère quantitatif variant de 0 à 90°,
- l'exposition, représentée par quatre variable qualitatives (1 ou 0 respectivement pour les expositions nord, sud, est ou ouest),
- l'altitude, caractère quantitatif, en mètres,
- la topographie, représentée par trois variables qualitatives (1 ou 0, respectivement pour les topographies convexe, plan ou concave),
- le degré de fermeture du milieu, représenté par trois variables qualitatives (1 ou 0, respectivement ouvert, moyennement fermé ou fermé),
- la profondeur du sol, représentée par trois variables qualitatives (1 ou 0, respectivement pour un lithosol, un sol moyennement profond ou un sol profond),
- le type de sol, représenté par trois variables qualitatives (1 ou 0, respectivement pour un sol tropical récent, un ferralsol ou un ferrisol),
- l'importance des affleurements rocheux, caractère quantitatif sous forme d'un comptage de contacts sur un ensemble de 125 points régulièrement répartis,
- l'importance de sol nu, caractère quantitatif sous forme d'un comptage de contacts sur un ensemble de 125 points régulièrement répartis,
- l'épaisseur de l'horizon humifère, caractère quantitatif mesuré en centimètres (moyenne de plusieurs mesures),
- la structure et la texture en surface et en profondeur, représentées chaque fois plusieurs variables qualitatives (1 ou 0),
- la capacité de rétention en eau (capacité au champ) d'échantillons de sol prélevés dans l'horizon A0 et dans l'horizon A1, caractères quantitatifs exprimés en %,

- le pH des horizons A0 et A1, caractères quantitatifs mesurés sur le terrain, avec une précision d'1/10^{e} d'unité,
- la somme des cations totaux échangeables dans les horizons A0 et A1, caractères quantitatifs exprimés en méq/100 g de sol séché à l'air,
- le taux de saturation en bases des horizons A0 et A1, caractères quantitatifs exprimés en %,
- le pourcentage de matière organique des horizons A0 et A1 des horizons A0 et A1, caractères quantitatifs.

D'autres exemples sont présentés en annexe.

La standardisation et la transformation des données

Comme les variables réunies dans une même étude sont parfois exprimées dans des unités très différentes ou même si elles sont toutes de même nature, la variabilité est parfois énorme et rend l'analyse des données peu utile, tellement certaines variables ou certaines valeurs particulières prennent une importance démesurée par rapport aux autres.

Il y a deux manières de modifier les données (ORLÓCI & KENKEL, 1985 ; PODANI, 2000 et WILDI, 2010) : la standardisation et la transformation. La standardisation modifie les données en utilisant des statistiques calculées à partir des données elles-mêmes. Ce sont par exemple la variance, l'amplitude, la moyenne, le total ou simplement la valeur maximum. La standardisation est souvent utilisée pour compenser des différences de poids ou d'unités entre les données. La transformation des données, au sens strict, utilise des fonctions mathématiques dont les paramètres ne dépendent pas des données.

La standardisation des variables

Exemples :

Le **centrage** : la moyenne est soustraite de chaque valeur.

$$x_{ij}^{'} = x_{ij} - \overline{x_i}$$

La **standardisation linéaire** : la valeur de la variable *i* est multipliée par une constante, qui est dérivée de toutes les observations de la variable (par exemple : amplitude, écart-type)

La **standardisation par l'amplitude** : la variable est rééchelonnée dans l'intervalle [0,1].

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

La **standardisation par l'écart-type** :

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

$s_i = \left[\dfrac{\sum_{j=1}^{m} \left(x_{ij} - \overline{x_i} \right)^2}{m-1} \right]^{1/2}$ qui est l'écart-type.

Le numérateur est la somme des carrés des écarts par rapport à la moyenne. Cette transformation est conseillée quand les variables sont mesurées dans des unités très différentes (pH, concentration, température) ; elle intervient dans le calcul du coefficient de corrélation.

La **standardisation par le total** :

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

. Les variables ayant de grandes valeurs sont diminuées en importance et celles avec de faibles valeurs réévaluées en importance.

La **standardisation par le maximum** :

$x'_{ij} = x_{ij} / \max_j \left\{ x_{ij} \right\}$. Toutes les valeurs sont divisées par le maximum des variables de l'échantillon.

La **standardisation par la longueur unitaire des vecteurs** :

$x'_{ij} = x_{ij} / \left\{ \sum_{j=1}^{m} x_{ij}^2 \right\}^{1/2}$. En partant de l'esp ace des variables, les individus se trouvent au sommet de vecteurs partant de l'origine. La somme des carrés vaut 1 pour chaque variable.

La transformation

Exemples :

La **binarisation** :

C'est la conversion des données quantitatives en qualitatives (présences-absences pour la végétation),

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

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

p est souvent égal à 0.

La **transformation logarithmique** : chaque valeur est remplacée par son logarithme décimal (base 10) ou népérien (base *e*). Cette transformation diminue les très grandes différences absolues au sein des variables,

comme c'est le cas avec le pH, qui est le logarithme décimal d'une concentration ; elle est aussi utilisée pour linéariser certaines relations entre variables.

La **transformation Arc sinus :**

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

Cette fonction convertit les variables et donne une amplitude [0,1]. Elle est souvent associée à une transformation racine carrée.

La transformation des données phytosociologiques

La transformation de ces données est toutefois indispensable pour s'intégrer dans des opérations mathématiques. Dans les calculs des analyses multivariées, des sommes sont effectuées sur les lignes et parfois sur les colonnes des tableaux. Mais quelle signification peut-on attribuer, par exemple, à une somme sur une colonne d'un tableau comprenant, comme c'est souvent le cas,

- des espèces dont certaines sont représentées par une abondance,
- d'autres par une dominance
- et un ensemble d'espèces très faiblement représentées ?

Comme les coefficients **1** et **2** n'ont pas la même amplitude que les coefficients **3**, **4** et **5**, peut-on conclure que 1 + 2 + 3 = 6 ?

La transformation suivante a été proposée par GILLET (2000) : à partir du code *r* qui prend la valeur 0,1, du code + qui prend la valeur 0,5 et des autres codes *1, 2, 3, 4 et 5*, le recouvrement moyen est calculé, ce qui donne le tableau 4.

AD code	AD numérique	Recouvrement moyen
r	0,1	0,03%
+	0,5	0,30%
1	1	3%
2	2	14%
3	3	32%
4	4	57%
5	5	90%

Tableau 4. Table de correspondance entre le code d'abondance-dominance (AD code), l'indice quantitatif d'abondance-dominance (AD numérique) et le recouvrement moyen.

Le code 2 est parfois aussi subdivisé en trois :

- *2m* : éléments très abondants, recouvrement inférieur à 5 %,
- *2a* : recouvrement compris entre 5 et 12,5 %, abondance quelconque,
- *2b* : recouvrement compris entre 12,5 et 25 %, abondance quelconque.

A partir de tous ces codes, van der MAAREL a créé l'échelle suivante :

Echelle d'abondance-dominance	Echelle ordinale de van der MAAREL
r	**1**
+	**2**
1	**3**
2m	**4**
2a	**5**
2b	**6**
3	**7**
4	**8**
5	**9.**

La standardisation par les individus

Diverses standardisations comparables à celles de variables sont parfois appliquées sur les individus. Pour plus de détails, voir PODANI (2000).

La double standardisation

Chaque valeur est divisée à la fois par la somme sur les colonnes et la somme par les lignes. Cette standardisation est utilisée dans l'analyse des correspondances et nous reviendrons inévitablement sur le sujet.

Calculs

Les diverses standardisations et transformations sont facilitées par des programmes comme Vegana (voir chapitre 13).

Les types de tableaux de données multidimensionnelles

On appelle donnée multidimensionnelle l'ensemble des valeurs d'un certain nombre de variables sur un individu (FOUCART, 1982). Un tableau de données multidimensionnelles est constitué d'un ensemble de variables mesurées sur un ensemble d'individus.

Les tableaux individus x caractères quantitatifs

Ce type de tableau est l'un des plus simples : les caractères des individus sont des variables quantitatives à valeurs réelles, non nécessairement continues. Le terme x_{ij} est donc un nombre réel, représentant la mesure de la variable x_i sur l'individu j.

Les tableaux de données ordinales

Dans ce type de tableau, les caractères des individus sont des variables ordinales sur lesquelles il est beaucoup plus délicat de faire des opérations mathématiques. L'échelle d'abondance-dominance est une échelle ordinale.

Les tableaux de présences

Les caractères des individus sont représentés par des variables qualitatives de type **0-1**. Ce type de tableau est fréquent avec des données de végétation quand l'abondance est impossible à estimer, Seule la présence des espèces, notée **1**, est enregistrable. Les tableaux comprennent généralement un très grand nombre de zéros, ce qui rend leur analyse délicate.

Les tableaux de contingence

Dans un tableau de contingence, le terme $n_{i,j}$ de la i^e ligne et de la j^e colonne est le nombre d'individus possédant la modalité i du caractère 1 et la modalité j du caractère 2. En principe, les modalités sont exclusives et exhaustives : un individu de la population ne peut posséder plus d'une modalité d'un même caractère, il en possède une et une seule. Lignes et colonnes jouent un rôle similaire.

Les tableaux logiques

Les tableaux logiques indiquent, pour chaque individu, l'appartenance à un groupe particulier ou, ce qui est équivalent, la modalité d'une variable qualitative qu'il possède. Le codage utilisé est le codage logique : l'appartenance est représentée par 1, la non-appartenance par zéro. Le terme x_{ij} est égal à **1** ou **0** suivant que l'individu j appartient au groupe i (ou s'il prend la modalité i au caractère qualitatif) ou non. Dans chaque colonne, un terme et un seul est égal à **1**.

Prenons exemple simple, les concentrations en ammonium, nitrite et nitrate dans 10 sites d'un ruisseau (tableau 5). Transformons la première ligne en tableau logique (tableau 6): la concentration en ammonium est représentée par deux modalités suivant qu'elle est - soit inférieure - soit supérieure ou égale à la moyenne des concentrations mesurées. Dans la première ligne, les concentrations inférieures ou égales à la moyenne prennent la valeur **1**, les autres la valeur **0**, Dans la seconde, la valeur **1** est attribuée aux concentrations supérieures à la moyenne, la valeur **0** aux autres. Dans chaque colonne, un terme et un seul est égal à **1**. Dans d'autres situations, les concentrations pourraient être représentées par un plus grand nombre de modalités, mais avec toujours la modalité 1 représentée une seule fois dans chaque colonne.

Paramètres	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

Tableau 5. Mesures de trois paramètres chimiques dans l'eau de dix relevés du Bocq.

NH_4^+ < moyenne	1	1	1	1	0	0	1	1	1	1
NH_4^+ ≥ moyenne	0	0	0	0	1	1	0	0	0	0

Tableau 6. Tableau logique du paramètre ammonium.

Les tableaux disjonctifs complets

Un tableau disjonctif complet est formé par la juxtaposition de plusieurs tableaux logiques. Chaque tableau logique correspond à une partition de l'ensemble des individus : le terme $x_{i1,j}$ est égal à **1** ou **0** suivant que l'individu appartient au groupe 1, ou non, les termes $x_{i2,j}$, $x_{i3,j}$ étant définis de façon analogue. Chaque colonne du tableau disjonctif complet contient autant de fois la valeur **1** qu'il y a de tableaux logiques.

Formons le tableau 7. On reprend d'abord les deux lignes du tableau 6. Ensuite, pour le nitrite, on crée deux lignes : une pour les concentrations inférieures ou égales à la moyenne (**1** pour ces valeurs et **0** pour les autres) et une deuxième pour les concentrations supérieures à la moyenne (**1** pour ces valeurs et **0** pour les autres). Pour le nitrate, on crée trois lignes : une pour les concentrations non mesurables (**1** pour les valeurs non mesurables et **0** pour les autres), une seconde pour les concentrations mesurables et inférieures ou égales à la moyenne (**1** pour ces valeurs et **0** pour les autres) et une troisième pour les concentrations supérieures à la moyenne (**1** pour ces valeurs et **0** pour les autres).

NH_4^+ < moyenne	1	1	1	1	0	0	1	1	1	1
NH_4^+ ≥ moyenne	0	0	0	0	1	1	0	0	0	0

NO_2^- < moyenne	1	1	1	1	0	0	0	0	0	1
NO_2^- ≥ moyenne	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 < moyenne	0	1	1	0	0	1	0	0	0	0
NO_3^- ≥ moyenne	0	0	0	1	1	0	1	1	1	1

Tableau 7. Tableau disjonctif complet construit à partir du tableau 5.

Dans chaque colonne, la somme est égale à trois. Ce tableau est donc formé par la juxtaposition de trois tableaux logiques.

Cas particulier des tableaux phytosociologiques

Avec les abondances-dominances, on peut découper chaque ligne en autant de lignes qu'il y a de niveaux d'abondance : une ligne pour les +, une pour r, une pour les i, une pour les 1, et ainsi de suite jusqu'à 5, sans considérer les abondances 0 (espèce absente dans un relevé). En effet, dans un tableau phytosociologique, la somme des absences est généralement nettement supérieure à celles présences (quelle que soit l'abondance). Si l'on soumet un tableau disjonctif complet à une analyse non symétrique des correspondances, celle-ci sera largement dominée par les absences, ce qui n'est pas l'objectif de l'analyse. Quand une espèce est absente d'un relevé, la somme sur les lignes se rapportant à cette espèce est nulle et non unitaire comme dans un tableau logique. Ce type de tableau s'appellera tableau disjonctif simple.

Les tableaux mixtes

Beaucoup de tableaux floristiques et mésologiques sont des tableaux mélangés, contenant des variables quantitatives continues (surfaces terrières), des nombres d'individus, des variables qualitatives (simple présence d'espèces faiblement représentées), Les tableaux mésologiques sont aussi souvent mélangés et comprennent des variables quantitatives mais des variables qualitatives ou ordinales, ce qui rend l'analyse délicate ; certaines variables sont au départ des variables transformées comme le pH qui est un cologarithme.

Dans les tableaux phytosociologiques, il y a souvent une proportion importante d'espèces représentées par leur seule présence (avec la cote **1** ou **0** si elle est absente) ce qui crée un grand nombre de cellules vides.

Les tableaux de distance, de proximité

Il s'agit de tableaux carrés construits à partir d'indices de distance, de proximité. On parle aussi de similitude (*similarity* en anglais) ou de dissimilitudes (*dissimilarity* en anglais).

Un indice de distance (ou de dissimilarité) est une fonction symétrique à valeurs réelles et positives définie entre deux individus : plus les individus *i* et *i'* se ressemblent, plus la valeur de cet indice est faible. Avec un indice de proximité (ou de similitude), plus les individus *i* et *i'* se ressemblent, plus la valeur de cet indice est élévée. Un indice de proximité peut prendre des valeurs négatives.

Les liaisons entre variables

Beaucoup de méthodes reposent sur l'analyse des liens linéaires entre les caractères observés, En reprenant l'exemple de BOUROCHE & SAPORTA (1980), on trouve ici une relation sous forme d'un nuage étroit et allongé le long d'une droite, entre le prix d'appartements et leur superficie (figure 1).

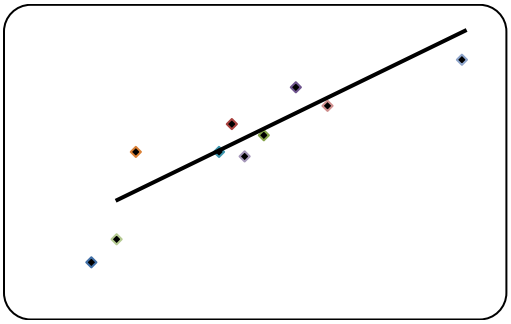


Figure 1. Relation entre la superficie et le prix d'achat d'appartements.

Toutefois, les relations entre paramètres enregistrés dans des conditions naturelles, ne se présentent pas toujours sous cette forme. Prenons tout d'abord le cas des relations entre paramètres mésologiques.

Le premier exemple porte sur des analyses chimiques de sol en savane (Parc de l'Akagera, Rwanda) dans l'horizon A1 de 70 échantillons (figure 2).

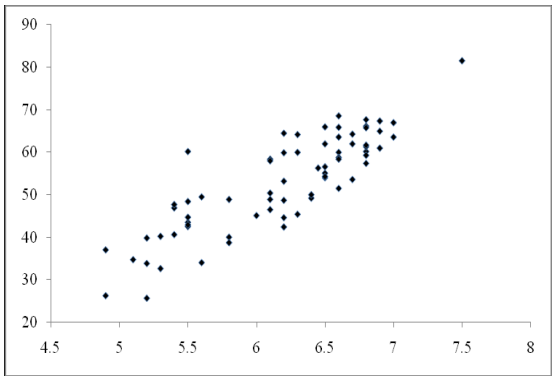


Figure 2. Relation entre le pH d'un sol et le taux de saturation en bases. En abscisse, le pH, en ordonnée, en ordonnée, le taux de saturation en bases.

Cette relation est claire, il y a une relation assez nette entre les deux paramètres et un ajustement par régression linéaire est possible.

Le second exemple se rapporte à des analyses chimiques d'échantillons d'eau prélevés une fois par mois pendant une année dans une rivière. Il s'agit de deux paramètres habituels pour définir la qualité biologique de l'eau, à savoir l'azote ammoniacal et l'orthophosphate (figure 3).

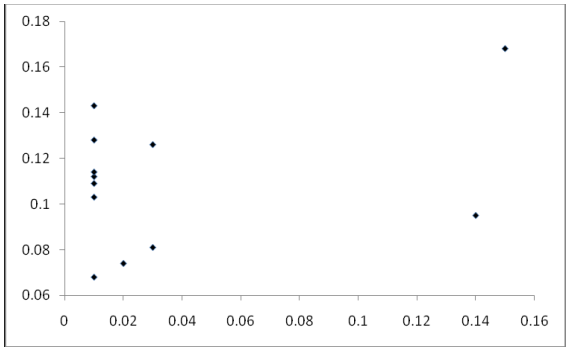
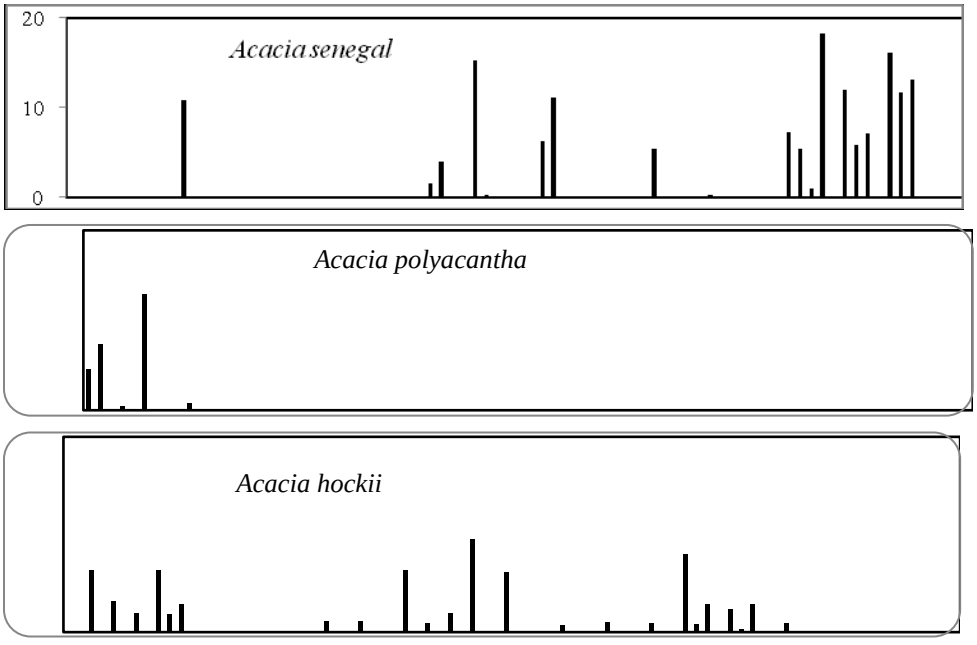
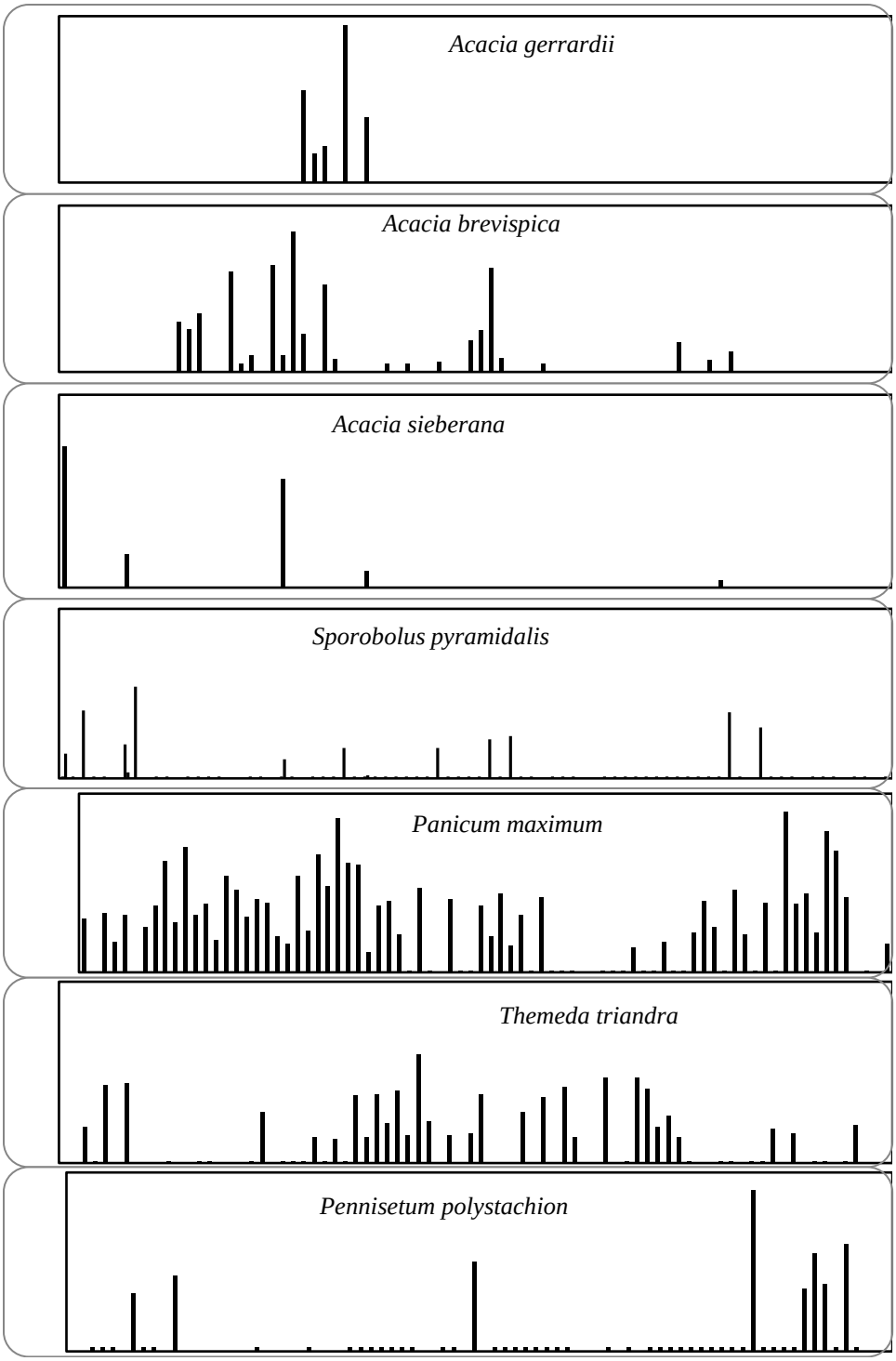


Figure 3. Relations entre l'azote ammoniacal et l'orthophosphate, en mg/l d'azote et de phosphore.

En abscisse, la concentration en azote ammoniacal (mg/l d'azote), en ordonnée, la concentration en orthophosphate (en mg/l de phosphore). Cette relation est moins nette et aucun ajustement ne semble possible avec de telles données. Les relations entre paramètres mésologiques sont loin d'être aussi nettes que dans la figure 2 et des relations moins marquées comme dans la figure 3 sont fréquentes.

Les relations entre abondances d'espèces montrent aussi une multitude de motifs possibles (figures 4 et 5):




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

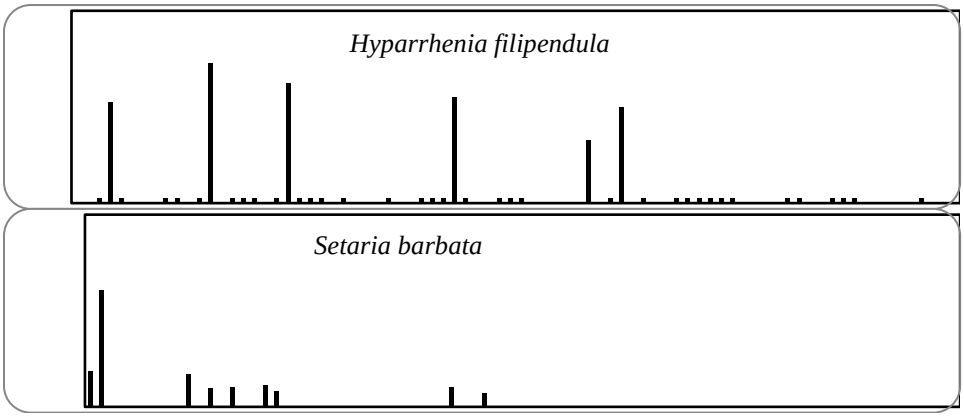


Figure 4. Variation de l'abondance le long d'un transect, dans un ensemble de 80 relevés de 250 m², entre le lac Ihéma et la colline Kionja (Parc national de l'Akagera). En ordonnée : les surfaces terrières, en dm² pour les *Acacia* et la fréquence de contacts avec 125 points pour les graminées.

Ces données sont illustrées comme suit,

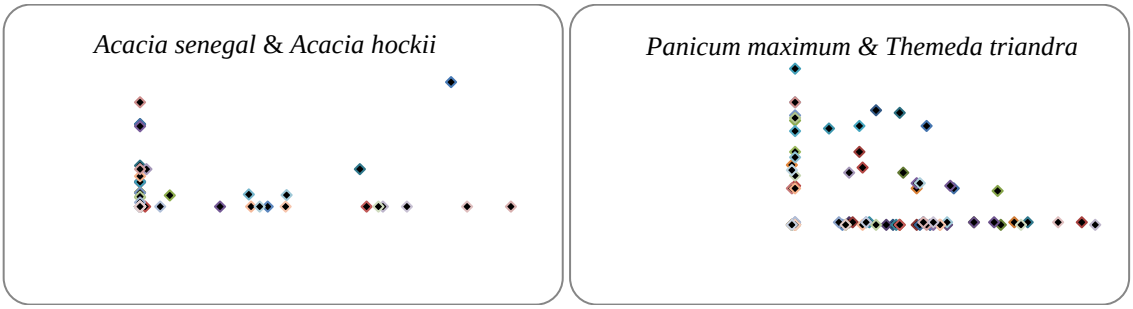


Figure 5. Relation entre les surfaces terrières d'*Acacia senegal* et *A. hockii* et les abondances de *Panicum maximum* et *Themeda triandra* (graminées en touffe) estimées par des comptages dans 125 points de contacts avec une aiguille.

En conclusion, il apparaît très difficile d'associer un modèle simple de distribution pour les espèces ou imaginer que l'on va se trouver avec des liens clairs entre abondances spécifiques. L'exploration de liens entre dispersion d'espèces et facteurs mésologiques ne s'en trouve pas simplifiée.

Continuum et discontinuum

En écologie végétale, on ne peut aller plus loin dans l'étude des relations entre la végétation et l'environnement sans prendre en compte la dépendance entre certains développements dans les concepts sur la végétation, les méthodes d'analyse mathématique et la connaissance des processus environnementaux.

Dans la seconde moitié du XX^{e} siècle, une controverse est née entre partisans de la représentation de la végétation sous forme de continuums plutôt que sous forme de communautés discrètes (AUSTIN in van der MAAREL 2005). Le concept de continuum s'appuie sur le principe de l'individualité spécifique, chaque

espèce étant distribuée en relation avec l'ensemble des facteurs environnementaux, incluant les interactions avec les autres espèces. Il n'y a pas deux espèces qui présentent la même dispersion. On en déduit le principe de continuité des communautés, qui évoluent le long de gradients environnementaux continus, avec des changements graduels dans les populations d'espèces.

De ce concept est née la méthode d'ordination directe appelée aussi analyse directe de gradient par WHITTAKER (1956). C'est l'analyse de dispersions d'espèces et de propriétés collectives (comme la richesse spécifique) en relation avec des variables environnementales conventionellement considérés comme des gradients environnementaux.

L'étape suivante fut l'utilisation d'analyses multivariées en vue de déterminer les gradients majeurs présents dans les données de végétation elles-mêmes. Les représentations graphiques qui en résultent résument les principaux axes de variations présents dans une matrice de similitudes issue du tableau de données. Aux axes des graphiques, on associe soit des gradients environnementaux, soit des étapes de succession dans le temps, soit différents régimes de pâturage.

Il n'y a pas de réel consensus sur la manère de conceptualiser la représentation des espèces ou des communautés végétales le long de gradients. La performance d'espèces le long d'un simple gradient environnemental peut prendre des formes très diverses. Les modèles de réponses non linéaires d'espèces le long de gradients ont été étudiés par AUSTIN (1976, 1980 et 2005, *in* van der MAAREL p.72) et AUSTIN *et al.* (1984). Plusieurs modèles ont été proposés comme la courbe de GAUSS, les fonctions polynomiales, les fonctions β ou γ ou encore la courbe de réponse écologique incorporant les compétitions entre espèces. Ceci rend l'analyse statistique très difficile car en plus, on doit traiter des variabilités très grandes, beaucoup plus grandes que celles que l'on rencontre dans le domaine de l'expérimentation. Les analyses multivariées doivent traiter trois niveaux de complexité :

- le caractère multidimensionel des gradients environnementaux et des réponses des espèces,
- le caractère curvilinéaire et non monotone des réponses des espèces,
- les écarts importants entre les formes de distribution des espèces et les modèles idéalisés.

En plus, une limitation importante vient de la faible proportion de l'espace échantillonné, même dans de grands ensembles de données (AUSTIN *et al*., 1984).

D'après AUSTIN(1980), des progrès dans le développement d'un modèle explicite des relations végétation/environnement sont nécessaires, même plus importants que dans le développement des techniques d'analyses multivariées.

Les discussions autour de ce concept de gradient reviennent inévitablement dans plusieurs chapitres.

Conclusions

Nous sommes donc confrontés à une analyse de tableaux très difficile à réaliser, présentant d'énormes variabilités, avec parfois des types de variables rendant certaines opérations mathématiques illicites et des propriétés particulières comme le très grand nombre de cases vides, principalement dans les tableaux phytosociologiques. Les techniques mises en œuvre doivent impérativement tenir compte des propriétés particulières des tableaux de relevés. La facilité des calculs, l'accès à un grand nombre de transformations et de standardisations ne doit pas nous conduire à entreprendre n'importe quelle analyse. Dans les chapitres qui suivent, nous verrons comment types de données et résulats des analyses sont intimement liés.

Janvier 2024

Références

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.

ORLÓCI, L. & KENKEL, N.C. (1985). *Statistical Ecology Monographs. Vol. I. Introduction to data analysis*. International Co-operative Publishing House: 339 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.

WILDI, O. (2010). *Data analysis in vegetation ecology*. Wiley-Blackwell. 211 pp.