

Chapitre 4. Les techniques d'analyse factorielle des données de végétation

par Guy BOUXIN[°]

Sommaire

Introduction	2
Les principales techniques d'analyse factorielle et d'ordination	5
Techniques d'analyse factorielle	5
L'analyse factorielle s.s. (FA)	7
L'analyse en composantes principales (ACP ou PCA).....	8
Les analyses des correspondances (CA)	9
L'analyse des correspondances détendancée (DCA)	13
L'analyse des coordonnées principales (PCoA ou NMS).....	13
Les ordinations sous contraintes	13
L'analyse factorielle multiple	15
Analyse de co-inertie (MCoA).....	17
Autres techniques.....	18
La technique de Bray & Curtis (Polar ordination)	18
Gradation non métrique multidimensionnelle (Non-metric multidimensional scaling, NMDS)	19
Utilisation des techniques.....	20
Quelques conclusions	26
Références	27
Annexe : calcul des valeurs et vecteurs propres	34
Données d'origine.....	34
Exemples d'analyse en composantes principales.....	34
Matrice des données centrées réduites	34
Matrice de corrélation	35

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

Valeurs propres	35
Vecteurs propres	35
Exemple de calcul	35
Exemple d'analyse des correspondances	36
Matrice pondérée	36
Matrice de covariance	37
Valeurs propres	37
Vecteurs propres	37
Exemple de calcul	38
Exemple d'analyse non-symétrique des correspondances	38
Matrice pondérée	38
Valeurs propres	39
Vecteurs propres	39

Introduction

L'analyse des données, *sensu* BOUROUCHE & SAPORTA, (1980) regroupe des méthodes très nombreuses et très différentes d'analyse statistique. On distingue deux démarches principales : la classification automatique, qui consiste à classer les unités statistiques et les variables à l'aide d'algorithmes préalablement établis, et l'analyse factorielle, qui utilise les propriétés des espaces vectoriels euclidiens pour décrire les individus et les variables (FOUCART, 1982). D'après PALM (1993), les méthodes d'analyse factorielle sont des outils fondamentaux de l'analyse des tableaux de données qui ne présentent pas de structure particulière. Elles visent essentiellement un but descriptif, en condensant l'information contenue dans un tableau, constitué souvent d'un nombre élevé de lignes et de colonnes, en quelques représentations graphiques à deux dimensions, accompagnées de tableaux reprenant les valeurs numériques de caractéristiques destinées à aider l'utilisateur lors de l'interprétation.

Ces techniques d'analyse semblent donc bien adaptées à l'analyse des tableaux de relevés de végétation. Elles ont été utilisées depuis quelques dizaines d'années dans l'étude de la végétation et reprises sous le vocable « ordination ». Au sens strict, l'ordination cherche à mettre la végétation en relation avec un ou plusieurs gradients environnementaux associés aux axes de l'analyse (GOODALL, 1954). Elle part du concept que la végétation est un *continuum*. Elle consiste à placer chaque relevé en relation avec un ou plusieurs axes de telle manière que sa position relative par rapport aux axes apporte un maximum

d'informations sur sa composition. Dans les techniques d'ordination, la démarche synthétique n'est pas toujours considérée comme une fin en soi mais devient un outil pour mettre les données floristiques en relation avec les facteurs environnementaux (ou autres) susceptibles d'affecter la composition de la végétation (GREIG-SMITH, 1964). L'ordination, à la suite de WHITTAKER (1973), est généralement considérée comme une technique d'analyse indirecte de gradient ; à ce sujet l'auteur écrivait : « L'analyse des gradients étudie la végétation en termes de gradients (ou variations plus ou moins continues) de l'environnement, de populations spécifiques et de communautés ». L'arrangement d'échantillons en fonction de gradients de l'environnement comme base de l'analyse s'appelle ordination. L'ordination est soit directe (arrangement des échantillons suivant leur position le long d'un ou plusieurs gradients environnementaux) soit indirecte (de la comparaison des échantillons, sont dérivés, par des calculs élaborés, des axes qui indiquent des changements dans la composition des communautés, le long desquels les échantillons sont arrangés). Un gradient dans la composition de la végétation peut être caractérisé par la quantité de « turnover » d'une extrémité à l'autre, ce que WHITTAKER (1967) a appelé « diversité bêta ». D'après PALMER (disponible à l'adresse : <http://ordination.okstate.edu/glossary.htm>), la diversité bêta est une mesure de la manière dont des échantillons diffèrent les uns des autres et/ou de leur degré d'éloignement sur des gradients de composition spécifique. Alternativement, c'est une mesure de la longueur d'un gradient écologique ou d'un axe d'ordination. Un axe ou un gradient avec une forte diversité bêta a des compositions spécifiques complètement différentes (sans aucune espèce en commun) à ses extrémités. Inversement, un axe ou gradient avec une faible diversité bêta a des compositions spécifiques assez semblables à ses extrémités.

Pour bien comprendre l'esprit dans lequel les techniques d'ordination sont utilisées, nous présentons les cinq concepts spatiaux selon GAUCH (1982b). Il y a l'*espace des espèces* (1), dans lequel les espèces sont les axes d'un espace multidimensionnel et les placettes des points localisés par leurs abondances pour chaque espèce. Il y a l'*espace des placettes* (2) (ou échantillons, ou relevés de végétation) dans lequel les placettes sont des axes et les espèces sont des points placés le long d'axes en fonction des « quantités » (ou abondance) des espèces dans chaque placette. A partir de ces deux espaces, des matrices de dissimilitude ou de distance pour chaque paire d'espèces ou pour chaque paire de placettes sont calculées. *Les espaces des dissimilitudes entre espèces* (3) *ou entre placettes* (4) sont ainsi définis. Le cinquième et dernier espace conceptuel est l'*espace écologique* défini par des gradients environnementaux comme axes. L'espace des espèces est défini par un grand nombre de dimensions (une par espèce). L'espace des relevés peut être réduit facilement dans son nombre de dimensions comme les espèces d'un tableau présentent souvent des dispersions plus ou moins redondantes. Par contraste, l'espace écologique est représenté par un petit nombre de dimensions. L'espace écologique est le résultat attendu d'une ordination.

L'ordination, prise au sens large, se fixe trois objectifs (GAUCH, 1982b) : (1) synthétiser les données d'un tableau de relevés, (2) mettre les communautés en relation avec des gradients de l'environnement et (3)

comprendre la structure des communautés. Au sens strict, l'ordination a pour but de définir un espace écologique à partir des données originelles. L'analyse classique comprend deux étapes : (1) produire une synthèse de la structure de la communauté, en la ramenant à un espace comprenant typiquement de une à trois dimensions ; (2) comparer la structure ainsi révélée avec l'information disponible sur l'environnement.

Toutefois, d'après AUSTIN (1976), les propriétés de l'ordination indirecte d'organiser et de permettre l'interprétation des communautés végétales en relation avec des gradients environnementaux sont perturbées par la non-linéarité des relations végétation/environnement.

Le concept d'ordination est aussi développé par WILDI (2013), en relation avec les techniques d'analyse multivariées. Les notions d'effet fer à cheval, de gradient et de non-linéarité dans la réponse de la végétation aux facteurs de l'environnement dominant la pensée scientifique de cet auteur. Dans le chapitre 6, nous expliquons que ces notions sont de faux problèmes.

L'analyse factorielle au sens statistique et l'ordination ont donc des bases conceptuelles bien différentes puisque l'analyse factorielle, dans un esprit d'analyse des données, n'émet aucune idée préconçue en ce qui concerne la signification des axes issus de l'analyse. La manière d'exploiter les résultats s'en trouve fortement changée.

Plusieurs techniques d'analyse multivariée existent aussi bien pour l'analyse directe de gradient ("constrained ordination") que pour l'analyse indirecte de gradient.

Une certaine confusion dans l'utilisation des techniques présentées vient de ce que les mêmes techniques sont utilisées, soit comme analyse de données, soit comme technique d'ordination. Par la présentation d'exemples, nous espérons montrer l'apport respectif de chacune des deux façons d'utiliser ces techniques. Si l'ordination est très souvent utilisée comme technique d'analyse indirecte de gradient, le lecteur comprendra rapidement que nous privilégions une approche beaucoup plus statistique. Nous utilisons les analyses factorielles comme un des outils conduisant à une analyse de données aussi développée que possible. Nous mettons une nouvelle fois le lecteur en garde contre une utilisation automatique et peu réfléchie de techniques statistiques, trop souvent et sans doute inconsciemment retenues pour conforter les idées préconçues que l'on se fait de la végétation ou des relations espèce-milieu. Les logiciels sont nombreux, facilement accessibles et trop souvent mal utilisés.

Dans ce chapitre, nous présentons succinctement les techniques d'analyse factorielle et d'ordination les plus communes. Les applications sont données dans les chapitres suivants. Quelques brèves conclusions sont enfin tirées.

Les principales techniques d'analyse factorielle et d'ordination

Notre intention n'est pas de passer en revue toutes les méthodes développées depuis les années 1950 mais bien de présenter les principales, de montrer l'esprit dans lequel elles ont été utilisées et surtout de faire réfléchir à la manière de les utiliser dorénavant.

Nous commençons en rappelant quelques notions de base de l'analyse factorielle, tout en conseillant aux lecteurs de se documenter de manière plus approfondie sur le sujet.

D'un point de vue historique, citons simplement la technique appelée « weighted average ordination (WA) » illustrée par WHITTAKER (1960). Une coordonnée est calculée pour chaque relevé, en fonction de la composition en espèces et la pondération attribuée à chaque espèce. Elle est efficace s'il y a un gradient simple prédominant et si on connaît quelque peu la distribution des espèces le long d'un gradient environnemental ou autre (par exemple, position d'une espèce le long d'un gradient pionnier-climacique). Avec DEL MORAL & WATSON (1978), la distribution des espèces le long d'un gradient est analysée par une ordination de type « reciprocal averaging ». WA a également été utilisée par OLSVIG (1979). Cette technique est utilisable avec les « score » attribués par ELLENBERG (1974) à de nombreuses espèces (éclairage, température, caractère continental, humidité, pH, ammonium ou nitrate, teneur en sel). Ces données doivent toutefois être manipulées avec beaucoup de prudence (WAMELINK *et al.*, 2002).

Les techniques présentées maintenant appartiennent toutes aux techniques d'analyses multivariées.

Techniques d'analyse factorielle

Ce paragraphe est consacré aux techniques d'analyse factorielle. Les deux techniques reprises font appel au calcul matriciel. La plupart des éléments de ce paragraphe sont tirés de PALM (1993).

Soit une matrice X de dimensions $n \times p$, telles que $p \leq n$, contenant des nombres quelconques. Considérons d'abord la matrice carrée $X'X$, de dimensions $p \times p$ et de rang r ($r \leq p$). Cette matrice admet r valeurs propres positives :

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

auxquelles sont associés r vecteurs propres u_1, u_2, \dots, u_r .

Considérons maintenant la matrice carrée XX' , de dimensions $n \times n$. Cette matrice est également de rang r et possède r valeurs propres positives :

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

auxquelles sont associés r vecteurs propres v_1, v_2, \dots, v_r .

Les vecteurs propres ont un signe arbitraire lié à l'algorithme de calcul utilisé et on peut changer le signe de tous les éléments d'un vecteur propre.

La représentation graphique des résultats joue souvent un rôle important dans l'interprétation des résultats. Considérons que l'on représente la matrice X par n vecteurs-lignes dans l'espace des p colonnes ou par p vecteurs-colonnes dans l'espace des n lignes. D'un point de vue géométrique, cela revient à remplacer l'espace à p (ou n) dimensions par un sous-espace à q dimensions et à considérer que les n (ou p) points se situent exactement dans cet espace. La qualité de cette approximation est donnée par le rapport :

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

qui mesure le rapport entre la somme des carrés des distances des points projetés dans l'espace à q dimensions et la somme des carrés des distances des points dans l'espace à p (ou n) dimensions. Ce rapport ne donne cependant qu'une idée globale de la qualité de la représentation pour l'ensemble des points, et il peut y avoir des différences importantes d'un point à l'autre, un point pouvant, par exemple, être très près du sous-espace à q dimensions et l'autre au contraire pouvant être plus éloigné de ce sous-espace. Pour chiffrer la qualité de la représentation d'un point donné, on calcule, par exemple, le rapport entre le carré de la distance à l'origine de la projection de ce point dans le sous-espace à q dimensions et le carré de la distance à l'origine de ce point dans l'espace à p ou à n dimensions. La qualité de l'approximation d'un point par un axe donné mesure en fait l'importance de l'angle formé par le segment reliant l'origine des axes au point donné et la projection de ce segment sur l'axe en question. De façon plus précise, il s'agit du carré du cosinus de cet angle. Une interprétation géométrique analogue est donnée à la qualité de l'approximation d'un point par un ensemble d'axes : il s'agit alors du carré du cosinus de l'angle formé par le segment reliant l'origine des axes au point donné et la projection de ce segment dans l'espace des axes pris en considération. Cette valeur est d'ailleurs égale à la somme des carrés des cosinus de la projection sur chacun des axes.

Une représentation schématique de la qualité des approximations est donnée, dans le cas de deux axes, à la figure 1 : la qualité de l'approximation du point P par l'axe 1 est égale à $\cos^2\alpha_1$; la qualité de l'approximation de ce point par l'axe 2 est égale à $\cos^2\alpha_2$ et la qualité de l'approximation du point P par le plan formé par les axes 1 et 2 est égale à :

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

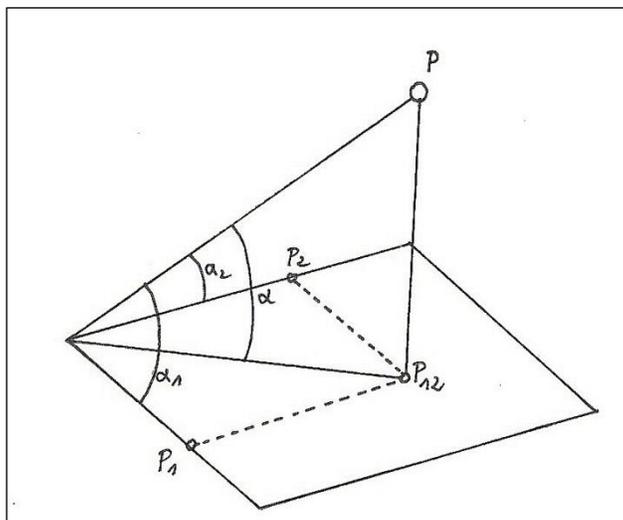


Figure 1. Qualité de l'approximation d'un point par un axe ou un plan (P_1 = projection du point P sur l'axe 1; P_2 = projection du point P sur l'axe 2; P_{12} = projection du point P sur le plan formé par les axes 1 et 2). Redessiné d'après PALM (1993).

La qualité de l'approximation d'un point est souvent appelée "cosinus carré", du fait de l'interprétation géométrique qui vient d'en être donnée, ou encore "contribution d'un axe à l'explication d'un point" ou "contribution relative d'un axe à la position d'un point". Dans les exemples, nous retiendrons "contribution relative".

L'analyse factorielle s.s. (FA)

Nous reprenons ici les principaux éléments du résumé de l'article de DAGNELIE (1973).

L'analyse factorielle est constituée d'un ensemble de méthodes mises au point en psychologie expérimentale et utilisées surtout dans l'interprétation statistique des résultats de tests psychologiques. Les principales notions d'analyse factorielle sont introduites en partant d'un exemple simple, emprunté au domaine psychologique : les résultats de trois tests (problèmes de raisonnement, problèmes d'arithmétique et calculs d'addition), auxquels peuvent être soumises différentes personnes, sont mis en rapport avec deux aptitudes fondamentales de ces personnes (aptitude au raisonnement et aptitude au calcul). Dans une telle situation, le but poursuivi est d'expliquer les corrélations observées entre les résultats des tests, en mettant en évidence diverses attitudes fondamentales sous-jacentes et en identifiant autant que possible ces aptitudes.

La transposition de l'analyse factorielle en phytosociologie se fait en considérant que les espèces végétales tiennent lieu de tests, que les relevés de végétation remplacent les personnes et que des facteurs écologiques se substituent aux aptitudes fondamentales. L'objectif est alors, à partir des corrélations observées entre espèces, de mettre en évidence et d'identifier autant que possible les principaux facteurs

écologiques sous-jacents. Subsidairement, d'autres problèmes peuvent être abordés, tels que l'étude du pouvoir différentiel des espèces, la définition de groupes sociologiques d'espèces étroitement liées entre elles, l'établissement d'une classification floristique des relevés, la comparaison de deux ou plusieurs classifications des relevés et l'étude de la répartition des espèces et des groupes d'espèces dans l'ensemble des relevés étudiés.

La première application de l'analyse factorielle en phytosociologie est due à GOODALL (1954). La contribution de l'analyse factorielle à l'étude des communautés végétales a été développée par DAGNELIE (1960). Cette technique est passée de mode, sous le rouleau compresseur de l'analyse des correspondances.

Il faut distinguer les techniques d'analyse d'un tableau unique, comme l'analyse en composantes principales, les analyses des correspondances (voir CHESSEL, DUFOUR & THIOULOUSE, 2004) et les analyses de plusieurs tableaux (voir DRAY, DUFOUR & CHESSEL, 2007).

L'analyse en composantes principales (ACP ou PCA)

Considérons maintenant qu'on dispose des observations relatives à p variables et effectuées sur n individus. Ces données se présentent sous la forme d'une matrice, Y , de dimension $n \times p$. Les p variables correspondent aux espèces (coefficient d'abondance-dominance, biomasse par ex.) ou des facteurs de l'environnement. Les variables initiales sont remplacées par les variables centrées réduites. La matrice $X'X$ correspond alors à la matrice de corrélation. Dans les tableaux phytosociologiques, les lignes correspondent aux espèces (variables) et les colonnes aux relevés (individus).

L'analyse non centrée des composantes principales ordonne simultanément les sites et les espèces. Le poids des sites et espèces dans l'analyse est proportionnel à leur abondance totale. Cette technique n'est pas susceptible de distorsion par des espèces rares. Des exemples sont donnés par EZCURRA (1987) et EZCURRA *et al.* (1987).

Quand des espèces présentent une distribution unimodale le long de gradients environnementaux, PCA produit des représentations, sur le plan des deux premiers axes de l'analyse factorielle, en arc ou en fer à cheval, ce qui est considéré comme une distorsion par GAUCH (1982b) *inter alia*. PCA, en tant qu'ordination, est considérée comme la plus performante quand les espèces présentent des réponses monotones le long des gradients environnementaux. Cette manière de considérer les résultats montre toute la différence qu'il peut y avoir entre une ordination et une analyse des données qui ne fait intervenir aucun élément extérieur au tableau de données.

L'analyse en composantes principales permet d'introduire des variables complémentaires, dont on ne veut pas tenir compte dans les calculs (FOUCART, 1982) ; pour cela, il suffit de calculer les covariances ou les corrélations avec les composantes principales et de les représenter dans la base des composantes principales déjà obtenues. HUSSON *et al.* (2009) distinguent les variables actives, qui entrent dans les calculs de l'ACP et les variables supplémentaires ou illustratives.

Il est également possible d'introduire des individus complémentaires.

Les analyses des correspondances (CA)

Le point de départ de l'**analyse des correspondances** est un tableau de fréquences à deux entrées. Ces fréquences constituent les éléments d'une matrice Y , de dimensions $n \times p$, n et p représentant les nombres de modalités relatives aux deux critères pris en considération. Les lignes et les colonnes de la matrice Y sont de même nature, contrairement à la matrice des données pour une analyse en composantes principales, où les lignes correspondent aux variables et les colonnes aux individus. Rappelons, encore une fois, que CA fait partie d'un ensemble de techniques statistiques qui se proposent d'aborder les tableaux indépendamment de toute hypothèse mathématique. Il ne faudrait donc pas les utiliser pour répondre à des hypothèses ou pour faire entrer un ensemble de données dans un paradigme comme le gradient indirect.

Toutefois, cette analyse est applicable aux tableaux de relevés classiques, avec les lignes correspondant aux espèces et les colonnes aux relevés. Elle est très souvent utilisée avec des coefficients d'abondance-dominance ou des données de présence. Elle ne convient pas du tout pour des biomasses ou des variables continues de l'environnement.

La matrice des fréquences (ou présences ou abondances-dominances) Y est d'abord transformée en une matrice de fréquences relatives, F , obtenue en divisant chaque élément de Y par la somme des éléments, c'est-à-dire par l'effectif total :

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

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

Cette matrice F est, à son tour, transformée en une matrice X par la transformation suivante :

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

$f_{i.}$ et $f_{.j}$ étant les fréquences relatives marginales des lignes et des colonnes :

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

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

Dans cette transformation, les lignes et les colonnes jouent un rôle symétrique. On remarque aussi que les écarts :

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

sont en fait les écarts entre les fréquences relatives observées et les fréquences relatives attendues sous l'hypothèse d'indépendance entre les lignes et les colonnes. D'autre part, à la constante $1/\sqrt{N}$ près, les quantités x_{ij} sont égales aux racines carrées des quantités qu'on calcule pour chacune des np cellules du tableau de contingence et qu'on somme pour obtenir la valeur χ_{obs}^2 lors de la réalisation du test χ_2^2 d'indépendance entre lignes et colonnes.

Après le calcul des valeurs propres et des vecteurs propres, la représentation des points-lignes et des points-colonnes dans des espaces à deux ou plusieurs dimensions sont des outils importants mais pas uniques de l'interprétation des données de départ. Dans ces graphiques, la proximité de deux points-lignes ou de deux points-colonnes traduit la similitude des "profils", c'est-à-dire des distributions conditionnelles, relatifs à ces deux lignes ou à ces deux colonnes. Toutefois, comme pour l'analyse en composantes principales, on ne perdra pas de vue que deux points (lignes ou colonnes) peuvent être proches dans un sous-espace (sur un axe ou sur un plan factoriel) sans nécessairement être proches dans l'espace complet, c'est-à-dire l'espace de l'ensemble des facteurs. On ne pourra donc conclure à la similitude des profils que si les points correspondants sont bien représentés, c'est-à-dire si la somme des cosinus carrés pour le sous-espace est suffisamment grande.

En pratique, on repère en premier lieu les points-lignes et les points-colonnes qui ont une forte contribution aux facteurs utilisés pour la représentation graphique et qui ont, en même temps, une qualité de représentation satisfaisante. Pour ces points, on examine alors les projections sur les axes et plus particulièrement le signe de ces projections, de manière à mettre en évidence les éventuelles conjonctions ou oppositions (points-lignes ou points colonnes avec projections de même signe ou de signes contraires). L'examen des données initiales ou, plus exactement, la comparaison des profils des lignes et des colonnes au profil moyen des lignes et des colonnes permet également de vérifier ces conjonctions ou oppositions. Si on utilise l'algorithme de HILL (1973), l'analyse s'appelle alors "Reciprocal averaging (ou RA)".

Selon GAUCH (1982b) et beaucoup d'auteurs par la suite, CA s'est montré supérieure à l'analyse des composantes principales. CA présente, comme PCA le problème de la configuration en arc mais avec PCA, les extrémités de l'arc se rejoignent parfois. Lorsqu'il s'agit de traiter de longs gradients de végétation, avec une diversité bêta élevée, CA apparaît généralement supérieure.

L'**analyse non-symétrique des correspondances** (NSCA) (GIMARET-CARPENTIER *et al.*, 1998 ; KROONENBERG & LOMBARDO, 1999) part de l'observation de la manière dont les données sont récoltées sur le terrain. Les présences des espèces sont observées parmi des relevés. Les relevés (nombre, position, par exemple) sont choisis par l'observateur, pas les espèces. Il en résulte une dissymétrie fondamentale qui est explorée par NSCA. La différence entre CA et NSCA concerne précisément la métrique. Les espèces ont un poids uniforme égal à l'unité dans NSCA alors que CA est basée sur la métrique χ_2 qui implique un double "averaging" calculé à partir des totaux marginaux des lignes et colonnes des tableaux. Cette dernière technique est ainsi qualifiée de duale. Dans NSCA, une forte importance est accordée aux espèces abondantes et l'effet des espèces rares est fortement limité. Dans un tableau d'occurrences, un grand nombre de taxons rares ne sera donc pas un facteur limitant. Un poids très faible ramène automatiquement un taxon vers l'origine.

Pour expliciter la différence entre CA et NSCA :

Dans la transformation des données,

- Pour CA, on calcule : $\text{don1}(I, J) = \text{don}(I, J) * \text{somme} / (\text{hc}(J) * \text{hr}(I)) - 1$,

- Pour NSCA, on calcule : $\text{don1}(I, J) = \text{don}(I, J) / \text{hc}(J) - \text{hr}(I) / \text{somme}$.

On calcule ensuite la matrice de covariance comme suit :

- Pour CA, $y1(J1, J2) = y1(J1, J2) + \text{don1}(I, J1) * \text{don1}(I, J2) * \text{hr}(I) * \text{Sqr}(\text{hc}(J1) * \text{hc}(J2))$

- Pour NSCA, $y1(J1, J2) = y1(J1, J2) + \text{don1}(I, J1) * \text{don1}(I, J2) * \text{Sqr}(\text{hc}(J1) * \text{hc}(J2))$.

Avec $\text{don}()$: données de départ (par exemple, présence d'une espèce dans un relevés), $\text{don1}()$: données transformée avant calcul de la matrice de covariance, somme : somme sur toutes les lignes et colonnes, $\text{hc}()$: somme sur une colonne, $\text{hr}()$: somme sur un ligne (ou rangée), $y1()$: covariance.

Les valeurs propres sont ensuite calculées à partir de cette matrice de covariance.

Une autre variante de CA appelée « **analyse des correspondances des lignes de données** » a été proposée par GREENACRE (2010). Elle part de la constatation que les relevés de végétation ou de faune ont même aire ou même volume et pose la question de savoir s'il faut alors transformer les abondances en fonction de l'abondance totale de chaque relevé. Cette CA analyse les lignes non « relativisées ».

L'**analyse des correspondances multiples** (MCA) (PALM, 1993) est une CA appliquée à des tableaux disjonctifs complets ; le tableau de contingence à deux critères est remplacé par un tableau à plusieurs critères. L'exemple typique est l'analyse d'un tableau présentant des personnes devant répondre à un ensemble de questions à choix multiples. Un exemple est fourni par COQUILLARD & GUEUGNOT (1991); il s'agit d'une étude écologique de 27 relevés dont l'appartenance à quatre syntaxons est connue et caractérisés par 12 variables. Chaque variable est représentée par deux ou plusieurs modalités (par exemple, la situation topographique est définie par quatre modalités : front de colline, sommet de versant, milieu de versant, bas de versant). Une telle technique, peu utilisée, offre pourtant des perspectives intéressantes lorsqu'il s'agit

d'analyser des variables non monotones ; des variables liées aux espèces comme l'abondance-dominance pourraient aussi être codifiées en catégories. Les applications sont rares (FRIED *et al.*, 2009).

L'analyse appelée "**fuzzy correspondence analysis**" ("fuzzy" signifie "flou") associe aussi plusieurs modalités aux variables (DOLÉDEC & CHEVENET, 1997). Elle est considérée comme une forme de MCA. Elle utilise une table "fuzzy coded" avec t espèces constituant les lignes et m modalités constituant les colonnes. v est le nombre de variables. La variable 1 a $m(1)$ modalités, la variable 2 a $m(2)$ modalités, ..., et la variable v a $m(v)$ modalités.

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

l'intersection de la i^e espèce avec la k^e modalité de la j^e variable se trouve la valeur a_{ij}^k avec $1 \leq i \leq t$, $1 \leq j \leq v$, et $1 \leq k \leq m(j)$. Soit a_{ij} et p_{ij}^k qui valent respectivement

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

Soit $\mathbf{P} = \left| p_{ij}^k \right|$ un tableau avec t espèces en lignes et m modalités en colonnes. Soit aussi

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

Les poids des lignes du tableau \mathbf{P} sont uniformément égaux à $r_i = \frac{v}{tv} = \frac{1}{t}$ et les poids des colonnes du même tableau valent

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

Si l'affinité d'une espèce pour une variable est inconnue, elle est codée "0, 0, 0, 0". Elle est donc implicitement remplacée par le profil moyen de la variable correspondante. La conséquence est qu'une espèce avec une telle valeur n'est pas prise en considération dans le calcul du poids de la colonne. La table \mathbf{P} est alors traitée comme une analyse des correspondances.

ROBERTS (1986) a montré que la "fuzzy set ordination" est une technique générale qui inclut des techniques spécifiques aux écologues comme l'analyse directe de gradient (WHITTAKER, 1967), l'ordination de BRAY & CURTIS (1957) ainsi que l'ordination "environmental scalars" de LOUCKS (1962). Les réponses des espèces ne sont pas limitées à une fonction mathématique précise ; elles peuvent être, par

exemple, non linéaires ou discontinues. Chaque axe de l'ordination est défini au préalable, ainsi la base de l'interprétation est prédéterminée. L'écologue peut donc utiliser la "fuzzy set ordination" pour tester explicitement des hypothèses. Des exemples de "fuzzy set ordination" ont été donnés par BANYIKWA *et al.* (1990), FULTON (1996), BOYCE (1998) et FEOLI *et al.* (2003).

L'analyse des correspondances détendancée" (DCA)

Il s'agit ici d'une variante très utilisée de l'analyse des correspondances (HILL & GAUCH Jr, 1980). Elle traite donc en même temps relevés et espèces. Cette technique est censée corriger l'effet "fer à cheval", c'est-à-dire l'arc ou le V, qui apparaît classiquement lorsque les points correspondant aux relevés ou espèces sont présentés dans le plan formé par les deux premiers axes. DCA consiste à aplatir le « fer à cheval ». Cela s'obtient en découpant le premier axe en segments et en mettant à zéro, dans chaque segment, la valeur moyenne sur le second axe. Un autre défaut est aussi corrigé, c'est-à-dire la tendance à comprimer les extrémités des axes par rapport à leur milieu ; cela s'obtient en rééchelonnant les axes de manière à égaliser autant que possible la variance, au sein des segments, des valeurs des espèces le long des axes de l'ordination. Dans cette technique, les valeurs propres ne peuvent plus être considérées comme des proportions expliquées de la variance totale.

L'analyse des coordonnées principales (PCoA ou NMS)

L'analyse des coordonnées principales (PCoA ou NMS) est une généralisation de l'analyse des composantes principales. Elle fut proposée par GOWER (1966). A la différence de l'ACP qui calcule la distance entre des points d'un espace (relevés ou espèces) au moyen de la distance euclidienne, PCoA autorise l'utilisation d'autres mesures de distance susceptibles de mieux traduire la différence entre relevés, par exemple. La distance entre deux relevés peut être le nombre d'espèces présentes dans un relevé ou dans l'autre mais pas dans les deux (PIELOU, 1977), ou encore le coefficient de BRAY & CURTIS (1957).

Les ordinations sous contraintes

Les techniques d'ordination sous contraintes cherchent à détecter automatiquement les motifs de variations des espèces qui peuvent être expliqués par des variables observées de l'environnement. Il s'agit donc d'analyses traitant simultanément deux tableaux de données bien différentes, espèces et variables environnementales. Les diagrammes qui en résultent expriment non seulement les motifs de variation dans la composition spécifique mais aussi les liens entre espèces et variables de l'environnement (JONGMAN *et al.* 1987). Comme nous le verrons, l'appellation « analyse canonique » fréquemment utilisée n'est pas correcte et ne devrait plus être utilisée.

L'analyse « canonique » des correspondances (ter BRAAK, 1986 ; CHESSEL *et al.* 1987 ; McCUNE & MEFFORD, 1999) est unique parmi les méthodes d'ordination en ce sens que l'analyse des correspondances du tableau principal (relevés de végétation) est soumise à la contrainte des variables mésologiques d'un second tableau au moyen d'une régression multiple. Cela signifie que l'ordination des relevés et des espèces est soumise à une contrainte par les relations avec les variables mésologiques. CCA convient dans les cas où les réponses des espèces aux variables mésologiques sont unimodales et où les variables mésologiques correspondantes ont été mesurées. Il s'agit donc ici d'une analyse directe de gradient. Les scores des relevés sont des combinaisons linéaires des paramètres environnementaux.

L'analyse « canonique » des correspondances (CCA) est une méthode très liée à l'analyse des correspondances. Fondamentalement, le tableau des relevés floristiques est considéré comme relevant de l'analyse des correspondances. Il s'agit d'une combinaison de la régression et de l'analyse des correspondances. Ces auteurs utilisent aussi l'expression "analyse des correspondances sur variables instrumentales" ou AFCVI. CCA est calculée à partir du tableau floristique doublement centré de l'analyse des correspondances et du tableau mésologique centré et normé puis transformé sur base orthonormale.

L'analyse « canonique » des correspondances (ter BRAAK, 1986 et 1987) sélectionne la combinaison linéaire de variables de l'environnement qui maximise la dispersion des coordonnées associées aux espèces. En d'autres mots, CCA choisit les meilleurs poids (c_j) pour les variables de l'environnement. Cela donne le premier axe de l'analyse. L'analyse utilise l'équation de régression linéaire suivante :

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

dans laquelle

z_{ji} est la valeur de la variable de l'environnement j du site i

c_j est le poids de cette variable (positif ou négatif)

x_i est la valeur de la variable environnementale composée en résultant, pour le site i .

Le second axe (et les suivants) sélectionne aussi des combinaisons de variables environnementales qui maximisent la dispersion des coordonnées des espèces, mais sous la contrainte d'être non corrélées à l'axes CCA précédent (ou aux autres axes CCA précédents). L'analyse extrait autant d'axes qu'il y a de variables environnementales.

CCA est une analyse des correspondances "sous contrainte" dans le sens que les coordonnées des sites sont contraintes à être une combinaison linéaire de variables environnementales mesurées. En introduisant la restriction dans l'algorithme CA où lignes et colonnes du tableau originel jouent un rôle symétrique, on obtient un algorithme pour CCA. Ce n'est donc pas une analyse canonique au sens strict du terme. Cela n'empêche pas de conserver le sigle CCA, en considérant les trois lettres comme les initiales des trois mots « Constrained Correspondence Analysis ».

Les espèces et sites sont positionnés sous forme de points dans le diagramme CCA comme dans CA et sont interprétés de la même manière ; les sites avec des abondances élevées d'une espèce donnée tendent à être proches de cette espèce. Les variables environnementales sont représentées par des vecteurs et les points correspondant aux espèces doivent être projetés sur ces axes. Les "biplot scores" sont les coordonnées des sommets des lignes. Plus longue est la ligne d'un paramètre de l'environnement, plus étroite est la relation de la végétation avec ce paramètre. La position des points-espèces par rapport aux lignes des paramètres est utilisée pour interpréter leur relation avec ce paramètre. En abaissant une perpendiculaire depuis un point-espèce jusqu'à la ligne du paramètre, on a une idée du degré de réponse de l'espèce à ce paramètre. On peut aussi en déduire si le paramètre est favorable (espèce du même côté que la pointe de la flèche) ou défavorable à l'espèce (situation contraire).

Les contraintes deviennent moins strictes quand le nombre de variables de l'environnement augmente. Plus le nombre de variables environnementales est grand, plus la CCA tend à devenir une simple CA. CCA est donc une **analyse directe** de gradient, contrairement à l'analyse des correspondances qui est souvent utilisée comme une **analyse indirecte** de gradient.

L'analyse des redondances (RDA) est la forme « sous contraintes » de PCA (JONGMAN *et al.* 1987). RDA est choisie si la longueur du gradient mesurée à partir d'une CA est inférieure à 3 et que les valeurs propres sont supérieures qu'avec CCA (CLARKE *et al.*, 2005). La DCCA est l'application de l'algorithme CCA à la "detrended correspondence analysis".

L'analyse factorielle multiple

L'analyse factorielle multiple (AFMULT) est une méthode d'analyse simultanée de tableaux. Elle est donc une généralisation des techniques de couplage de tableaux. Elle est due à ESCOFIER & PAGES (1984, 1986, 1989).

On dispose de K tableaux ayant en commun n lignes-individus (figure 2), chacun d'eux correspondant à un groupe de variables-colonnes.

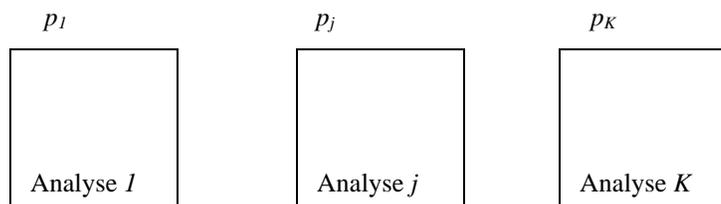


Figure 2. Ensemble de K tableaux ayant en commun n individus.

Il s'agit, par exemple, d'analyses physico-chimiques d'échantillons d'eau prélevés en un certain nombre de sites le long d'une rivière à différents moments de l'année ou pendant plusieurs années. Il peut s'agir aussi de tableaux présentant des données très différentes comme un tableau de relevés floristiques associé à un tableau de caractéristiques physiographiques et à un tableau d'analyses de sols ; ces trois tableaux ont en commun les individus (relevés) disposés ici en lignes.

L'AFMULT, pour uniformiser le rôle des tableaux dans l'analyse simultanée, consiste à ramener à une échelle commune les inerties projetées sur le premier axe de l'analyse de chaque tableau.

L'analyse factorielle multiple, au départ, est construite sur la technique d'analyse en composantes principales. Elle nécessite, par exemple, une ACP du tableau formé par la réunion des K tableaux de départ, une ACP pour chacun des tableaux séparés et une ACP d'un tableau pondéré.

CENTOFANTI *et al.* (1989) nous donnent un exemple avec 10 paramètres physico-chimiques mesurés en surface dans 10 stations à 12 dates dans un lac réservoir. HAURY *et al.* (1998) ont étudié le déterminisme écologique de distribution des phytocénoses, en prenant en considération les variables du milieu ; cinq groupes de variables ont été utilisés : 15 pour la mésologie physique, 10 pour la géologie, 3 pour la physico-chimie de l'eau, 10 pour les caractéristiques de la végétation et 92 pour les macrophytes. Les résultats des cinq premiers axes ont été exploités et mettent clairement en évidence les relations entre divers facteurs mésologiques et des groupes floristiques. DEVINEAU (2001) présente une analyse conjointe de quatre tableaux, respectivement un tableau floristique avec des indices d'importance d'arbres et d'arbustes, un tableau floristique des fréquences des « régénérations et sous-arbustes », un tableau relatif aux profils pédologiques et un dernier tableau avec les caractéristiques chimiques et texturales de l'horizon supérieur. CHUA *et al.* (2016), dans l'étude sur le long terme de la reconstitution de forêts tropicales secondaires, examinent la correspondance entre quatre tableaux correspondant respectivement aux plantules, aux plantes adultes, aux conditions de germination et à la distance des sources potentielles de semences.

Cette technique, par ses nombreuses possibilités, est une des plus prometteuses de ces dernières années.

Les étapes de l'analyse sont :

- Le calcul de la matrice de corrélation entre les variables du tableau originel, ainsi que de ses valeurs propres ;
- Le calcul des matrices de corrélation et des premières valeurs propres des sous-tableaux ;
- Dans le tableau originel, chaque sous-tableau est pondéré par la première valeur propre correspondante ; pour chaque ligne, les variables originelles sont divisées par la racine carrée de la première valeur propre correspondante ;
- La matrice de corrélation du tableau pondéré est calculée, ainsi que ses vecteurs et valeurs propres ; les coordonnées et contributions relatives des individus sont calculées comme dans une analyse classique,

les coordonnées des espèces sont obtenues en multipliant les lignes des vecteurs propres par la racine carrée des valeurs propres du tableau pondéré et la racine carrée du paramètre de pondération.

De cette manière, nos résultats correspondent à ceux produits par le logiciel Rcmdr (chapitre 13).

Nous avons essayé d'adapter cette technique à l'analyse des correspondances classique ainsi qu'à l'analyse non-symétrique des correspondances.

Se pose toutefois le problème de la pondération des sous-tableaux. Prenons l'exemple de l'analyse en composantes principales : la première valeur propre est, dans une certaine mesure, fonction de la trace de la matrice, qui elle-même correspond au nombre de variables (dans le cas précis du calcul d'une matrice de corrélation). Un petit tableau produit une plus petite valeur propre qu'un grand tableau et le mélange de sous-tableaux avec un nombre bien différent de variables favorise grandement les sous-tableaux avec un petit nombre de variables. Il est donc déconseillé d'utiliser cette technique en mélangeant des données floristiques (souvent très nombreuses) et des données mésologiques, pédologiques, chimiques, ... souvent moins nombreuses. Une solution existe toutefois, en pondérant, non plus par la première valeur propre mais par le quotient entre la première valeur propre et la trace de la matrice de corrélation ; cela évite l'écueil d'une pondération due uniquement à la taille de la matrice.

Dans l'analyse des correspondances, la trace de la matrice et par conséquent les valeurs propres, dépendent en partie du grand total du tableau de données. Prenons l'exemple d'un tableau de présences : si la plupart des cellules du tableau correspondent à 1 (tableau dense) et donc que peu de cellules sont vides, la trace de la matrice sera plus grande que si beaucoup de cellules sont vides. Les premières valeurs propres sont donc plus grandes quand un tableau est riche de données. Ainsi, si on mélange des sous-tableaux avec des densités de données bien différentes, les sous-tableaux peu denses sont écrasés par les sous-tableaux denses. Nous avons ainsi eu l'idée de pondérer les sous-tableaux floristiques ne contenant que des « 0 » et des « 1 » par la proportion des cellules occupées par des « 1 » (que nous appelons **densité** par la suite) et de rééquilibrer ainsi l'influence de chaque sous-tableau.

Pour l'AC et l'ANSC, les sommes sont calculées sur les lignes et colonnes et la pondération par la racine carrée du paramètre adopté porte directement sur les variables originelles. Les sommes et le grand total sont soumis à cette pondération. Les diverses étapes de calcul se font ensuite comme dans les analyses classiques.

Analyse de co-inertie (MCoA)

L'analyse de co-inertie multiple a été expliquée mathématiquement par CHESSEL & HANAFI (1996). Suivant APONTE *et al.* (2011), MCoA est une technique d'analyse multivariée multi-tables, qui autorise l'analyse simultanée de plusieurs tableaux de données. Elle maximise la variance au sein de chaque tableau individuel et la corrélation entre les scores de chaque tableau individuel (analyse individuelle) ainsi que les scores de synthèse (analyse des tableaux juxtaposés) de manière à fournir une structure de référence. MCoA optimise donc la covariance entre plusieurs analyses individuelles et une analyse de référence qui peut être vue comme une structure commune partagée par les analyses individuelles.

Il est aussi possible de traiter deux ensembles de données qui n'ont pas été récoltées dans les mêmes sites, en ajoutant une matrice de voisinage entre les sites des deux plans d'échantillonnage (analyse RLQ spatiale, DOLÉDEC *et al.* 1996, DRAY *et al.*, 2002). La technique a aussi été utilisée par DRAY *et al.*, 2003 qui lie des tableaux de données écologiques. Deux exemples récents sont donnés par JAMIL *et al.* (2013) et WESULS *et al.* (2012). Ils expliquent la technique comme suit : l'analyse RLQ relie une matrice de variables environnementales dans des relevés (tableau **R**) à une matrice d'espèces par caractères fonctionnels (tableau **Q**) en utilisant des couvertures d'espèces dans les relevés (tableau **L**) comme liens. C'est une extension de l'analyse de co-inertie. Avant l'analyse RLQ, les tableaux sont soumis à des analyses multivariées séparées. L'analyse des correspondances a été utilisée pour la matrice des espèces dans les relevés (tableau **L**). La matrice des variables environnementales (tableau **R**) par relevé a été analysée par une analyse en composantes principales. Pour le tableau espèces par caractère fonctionnel (tableau **Q**), une analyse en composantes principales, adaptée à un mélange de données quantitatives et catégoriales a été calculée. Les scores des relevés de la CA du tableau **L** ont été utilisés comme poids des lignes dans la PCA du tableau **R** et les scores des espèces de la CA du tableau **L** ont été utilisées comme poids des lignes dans l'analyse du tableau **Q**. L'analyse **RLQ** est une extension de l'analyse de co-inertie qui tient compte simultanément des informations contenues dans les tableaux **R**, **L** et **Q**. Il existe aussi une autre forme d'analyse appelée RLQ partielle.

Autres techniques

Les techniques qui suivent sont fondamentalement différentes de celles que nous venons de présenter, en ce sens qu'elles ne font pas appel au calcul de valeurs propres et vecteurs propres issus d'une matrice de covariance ou de distances.

La technique de Bray & Curtis (Polar ordination)

Historiquement, une première méthode importante est l'ordination polaire de BRAY & CURTIS (1957, in COTTAM *et al.*, 1973). Soit un ensemble de relevés à ordonner. Des mesures de similitude sont calculées pour chaque relevé avec tous les autres. Deux relevés extrêmes sont définis, soit sur base de la plus

grande dissimilitude avec le reste de l'ensemble des relevés et entre eux, soit parce qu'ils représentent les extrêmes d'un gradient environnemental. Les mesures de similitude sont utilisées pour arranger tous les autres relevés dans une séquence le long d'un axe, entre la paire de points extrêmes. Une seconde paire de points extrêmes, différente de la première, est choisie et tous les relevés sont arrangés le long de l'axe entre la seconde paire. Les deux axes déterminent les coordonnées des relevés dans un plan, chaque relevé correspondant à un point sur lequel on peut surimposer diverses caractéristiques comme des variables de l'environnement. Les relevés peuvent encore être ordonnés par rapport à un troisième axe et ainsi de suite. De nombreuses variantes de cette technique ont été proposées. Une importante description de la végétation de l'État du Wisconsin a été réalisée avec l'aide de cette technique (CURTIS, 1959).

Gradation non métrique multidimensionnelle (Non-metric multidimensional scaling, NMDS)

Il s'agit d'une technique originale, imaginée par KRUSKAL (1964a et b). Nous reprenons ici les explications fournies par ORLÓCI & KENKEL (1985) et McCUNE & MEFFORD (1999).

Soit au départ, un tableau Y de relevés avec p lignes (espèces) et n relevés (colonnes). L'objectif est de fournir un tableau X à n colonnes et $t \leq p$ lignes, dans lequel les relevés sont placés dans un autre ordre, en fonction de leurs ressemblances (mesurée par une distance d), cet ordre étant le reflet d'un gradient de végétation ou traduisant l'influence prépondérante de l'un ou l'autre facteur. Les t axes sont indépendants. Les calculs commencent avec le calcul d'une matrice de distances $n \times n$ (d) calculée à partir du tableau Y . Dans ce premier tableau, les relevés (colonnes) sont placés dans l'ordre défini par l'observateur. Une autre matrice de distance $n \times n$ (D) est alors calculée, basée sur le tableau X , contenant n colonnes avec des coordonnées attribuées aux relevés et définies par l'observateur (nombres aléatoires ou coordonnées résultant d'une autre ordination) ainsi qu'un petit nombre de lignes t également fixé par l'observateur (1, 2 ou plus, $\leq p$). X est modifié au cours de cycles itératifs. L'objectif est de rendre les éléments de X de moins en moins aléatoires de manière que l'ordre des éléments de D se rapproche de plus en plus étroitement de l'ordre des éléments correspondants de d . Le degré avec lequel l'objectif est atteint est mesuré en termes d'un coefficient de "stress" $s(D;d)$. Si la valeur du stress est trop grande, les itérations continuent ; l'ordre des éléments de X sont changés et un nouveau tableau D est calculé puis comparé avec d . Quand le stress est suffisamment petit ou quand aucune amélioration ne peut plus être apportée, la configuration de X est définitive. Les éléments de X sont les coordonnées des n relevés sur les t axes. ORLÓCI & KENKEL (1985) recommandent de choisir une configuration de départ aléatoire et de ne pas choisir une configuration basée sur une autre ordination. Ils utilisent la distance euclidienne. Il est également possible d'utiliser le programme de McCUNE & MEFFORD (1999).

Cette technique particulière appelle quelques remarques :

- la première dimension dans un espace à deux dimensions n'est pas le même que la première dimension dans un espace à trois dimensions ou à une seule. Pour un nombre donné de dimensions, la solution pour un axe particulier est unique.
- les numéros des axes sont arbitraires, ce qui fait que le pourcentage de variance associé à un axe ne forme pas nécessairement une série descendante avec le nombre croissant d'axes.
- le nombre approprié d'axes devrait être déterminé par l'évolution de stress final en fonction du nombre d'axes choisi. Il faut choisir un nombre d'axes au-delà duquel la réduction du stress est petite.

La solution est influencée par l'ordre initial des relevés dans le tableau X . Si on choisit l'option "aléatoire", plusieurs analyses fournissent des résultats différents.

Il existe de nombreuses variantes de cette technique (PRENTICE, 1977 et 1980 ; MINCHIN, 1987). Les mesures de distance peuvent aussi être faites au moyen de la distance euclidienne, des coefficients de SØRENSEN, de JACCARD ou de corrélation. La distance "corde" que nous utiliserons est la suivante :

$\delta_{jk} = [2(1-q_{jk})]^{1/2}$ dans lequel

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

ZIMOWSKI *et al.* (2014) donnent un exemple très explicite de l'esprit dans lequel cette technique est utilisée.

Utilisation des techniques

Nous avons consulté 394 articles traitant de ces techniques, depuis le début de leur utilisation jusqu'en 2004. Les citer tous prendrait trop de place mais ils peuvent être trouvés presque tous en consultant des revues comme *Journal of Vegetation Science*, *Vegetatio*, *Journal of Ecology*, *Ecology*, *Ecological Monographs*, *Plant Ecology*, *Acta Oecologia*, *Phytocoenologia* parmi les plus représentées.

L'analyse factorielle *sensu stricto* était pratiquement la seule utilisée avant 1965, elle ne l'est plus. L'analyse des composantes principales, très utilisée entre 1965 et 1974, est peu utilisée seule actuellement pour des données de végétation mais sert toujours à synthétiser des tableaux de données monotones comme les variables environnementales. L'analyse de BRAY & CURTIS, peu utilisée, n'a plus été rencontrée après 1984. L'analyse des principales coordonnées, également peu utilisée, apparaît toujours sporadiquement. L'analyse factorielle multiple est toujours très rarement utilisée. Avec une fréquence plus élevée, l'analyse non-métrique multidimensionnelle, dans ses nombreuses variantes se maintient, ainsi que l'analyse des

redondances. Les plus utilisées actuellement sont CA (RA incluse), CCA et DCA (DCCA n'a pas fait de véritable percée). Les deux techniques les plus utilisées sont donc très différentes. CCA est une technique d'analyse directe de gradient alors que DCA et CA sont des techniques d'analyse indirecte de gradient. Il n'est pas certain que les chercheurs choisissent toujours l'une ou l'autre technique en parfaite connaissance de cause. Comme l'explique ØKLAND (1996), l'ordination *s.s.* et l'ordination sous contrainte (ou directe) suivent des objectifs différents. La persistance de CCA montre le besoin qu'il y a de coupler des tableaux floristiques et des tableaux mésologiques. Trop souvent, le choix des techniques est lié à la recherche d'un gradient dans la composition des espèces (VAN LOOY *et al.*, 2003 ; NATTA *et al.*, 2003) sans la moindre étude préalable sur la dispersion des espèces.

Nous avons ensuite repris tous les articles parus dans la revue « Journal of Vegetation Science » depuis son début en 1990 jusqu'au dernier volume de 2016 et construit la figure suivante. Seules les techniques les plus fréquemment utilisées ont été reprises.

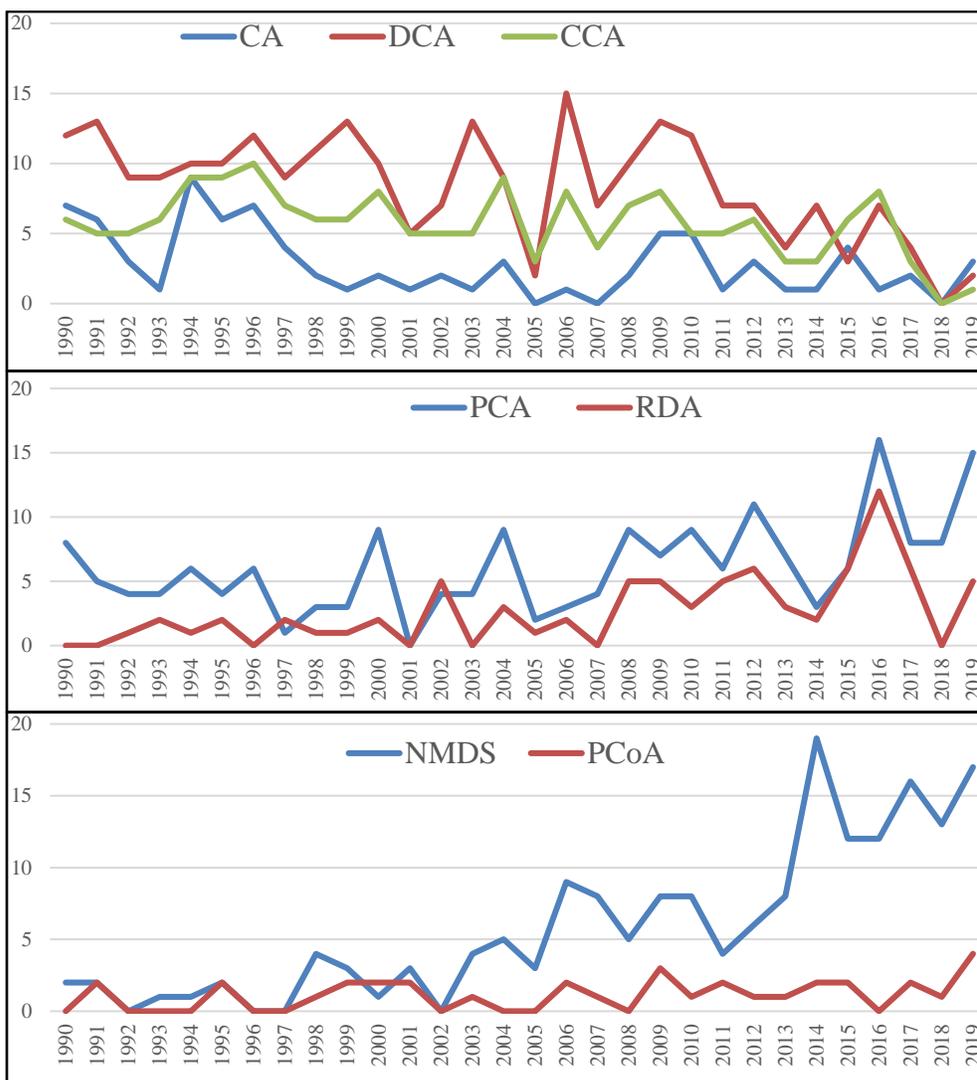


Figure 3. Fréquences annuelles d'utilisation des diverses techniques, de 1990 à 2019.

CA, DCA et CCA sont manifestement les techniques très fréquemment rencontrées jusqu'en 2010 et puis, leurs fréquences d'utilisation diminuent, spécialement à partir de 2017. PCA est fréquemment choisie pour des tableaux de paramètres mésologiques ou de caractères fonctionnels (*functional traits*), plutôt que pour des tableaux de végétation habituels ; l'utilisation de PCA pour les caractères fonctionnels est en grande partie la cause de son utilisation plus fréquente ces dernières années. La fréquence de RDA suit logiquement celle de PCA puisque RDA est une PCA sous contraintes. La fréquence de NMDS augmente régulièrement. Les autres techniques sont moins prisées. On déduit que le concept de gradient domine les recherches en écologie végétale. Quand le gradient est temporel, l'étude devient une étude de succession (AUSTIN, 1977 ; DEBUSSCHE *et al.*, 1996). On peut aussi se demander dans quelle mesure le choix d'une technique n'est

pas influencé par les ensembles de logiciels les plus connus. PC-ORD (Mc CUNE & GRACE, 2002) et CANOCO (ter BRAAK & ŠMILAUER, 2002) sont de plus en plus supplantés par les programmes en « R ». Ce qui est étonnant, c'est que l'analyse des correspondances non-symétrique est totalement ignorée et que l'analyse factorielle multiple n'apparaît que récemment (CHUA *et al.*, 2016).

Des conclusions très proches ont été tirées par VON WEHRDEN *et al.* (2009) dans une étude portant aussi sur les années 1990-2007, dans un nombre important de revues, notamment grâce aux informations fournies par l'« Institute of Scientific Information » ou ISI. L'utilisation des techniques d'analyse multivariée dans le domaine de l'écologie tend à augmenter depuis 1990. CCA est la technique la plus utilisée, suivie par CA/RA. DCA, NMDS, PCA et RDA ont récemment gagné en importance. Les auteurs ont également souligné l'importance des logiciels disponibles. Sur toute la période utilisée, CANOCO reste l'ensemble le plus utilisé. Depuis 1998, PC-ORD a pris de l'importance et plus récemment, les logiciels basés sur l'environnement R sont facilement accessibles. VON WEHRDEN *et al.* (2009) mettent en évidence la faiblesse de plusieurs techniques comme DCA notamment sur les paramètres influençant le « detrending » ou pour d'autres techniques l'importance de la standardisation des données du tableau de relevés ou du choix d'un indice dans le calcul de matrices de distance. Certaines informations mathématiques utiles à l'interprétation, comme les valeurs propres ne sont pas toujours utilisées. Les performances des techniques sont aussi toujours évaluées sous l'angle de l'ordination, c'est-à-dire dans leur capacité à révéler des gradients indirects.

Dans la plupart des études, les pourcentages d'inertie associés aux axes sont présentés. Ils permettent d'évaluer le nombre d'axes à prendre en considération. Quand les pourcentages cumulés des deux ou trois premiers axes sont importants (sans qu'il soit possible de fixer une norme), l'analyse multivariée est généralement considérée comme utile.

Force est de constater que, dans une très grande majorité des études, seuls les scores des relevés ou des espèces sur les axes de l'analyse multivariée sont pris en considération dans l'interprétation des résultats. Les aides à l'interprétation sont rarement utilisées, probablement parce qu'elles ne sont pas fournies par la plupart des programmes disponibles. Les contributions relatives des espèces ou des relevés sont pourtant des aides précieuses pour organiser l'interprétation. Elles sont prises en considération notamment par BONIN & ROUX (1978), BONIN *et al.* (1983), BOUXIN (1986, 1987a & b, 1991, 1995 & 1999).

Les analyses multivariées sont utilisées aussi comme technique pour réduire le "bruit de fond" d'un tableau de relevés (GAUCH, 1982a). En effet, les tableaux de relevés contiennent des variations intéressantes, interprétables en termes historiques ou environnementaux ainsi que des variations aléatoires, inutilisables. C'est donc ici une technique de réduction de données.

Elles sont aussi utilisées pour étudier les motifs plurispécifiques dans des grilles ou placettes de relevés ou tout simplement lorsque les relevés sont placés en lignes. Dans la plupart des études, les espèces, les relevés ou les deux ensembles sont présentés dans un espace à deux dimensions (axes 1 et 2, 1 et 3, 2 et 3, *etc...*). Dans l'étude d'un transect, ESTÈVE (1978) recommande que l'on représente chaque axe séparément, en prenant en compte la position des relevés dans l'espace. La fonction mettant en relation la variable de l'axe considéré (en ordonnée) et la position des relevés (en abscisse) met en évidence les zones homogènes, les zones de transition et les zones à forte variabilité floristique. Les axes sont placés à la suite l'un de l'autre, parallèlement. On peut de toute manière étendre la recommandation d'ESTÈVE à toute carte de relevés. Des exemples sont donnés par BACHACOU & CHESSEL (1979), BOUXIN (1983), BOUXIN & LE BOULENGÉ (1983), BOUXIN & GAUTIER (1979, 1982) et GAUTIER (1979). GALIANO (1983) utilise les coordonnées de relevés le long du premier axe d'une RA comme base d'étude de motifs plurispécifiques. Des exemples d'applications seront présentés dans ce chapitre ainsi que dans les chapitres sur l'analyse statistique des tableaux de végétation.

Les analyses multivariées ont également servi de base à des études phyto-écologiques ou phytosociologies importantes. Citons GUINOCHET (1973), BONIN & ROUX (1978), BONIN *et al.* (1983), MÄKIRINTA (1989).

De nombreux auteurs ne procèdent pas à l'analyse de tableaux complets mais éliminent les espèces rares comme BARUCH (1984), BELSKY (1983), BRZEZIECKI (1987), BURKE (2001), COOPER (1984), HALL & SWAINE (1976), HERMY & STIEPERAERE (1981) et NEWBERY (1991) parmi d'autres. Ces simplifications sont censées éliminer ou réduire les distorsions observées, comme l'effet fer à cheval. Certains auteurs n'hésitent pas à traiter de très grands tableaux (1117 relevés), qui sont ensuite simplifiés par l'élimination d'espèces rares et de relevés marginaux (BERGMEIER & DIMOPOULOS, 2001).

Les méthodes d'analyse multivariée sont aussi des techniques de réduction de données : elles permettent tout simplement de transformer un tableau à n lignes (espèces) en un tableau à deux, trois ou un petit nombre de lignes. Cette réduction de dimensionalité est considérée comme un des objectifs de l'analyse par DALE (1975) et AUSTIN (1976), pour autant qu'elle soit associée à une recherche de gradient. Les tableaux simplifiés, dont on a éliminé les redondances, peuvent aussi être utilisés dans des algorithmes de classification, par exemple (BERTHET *et al.*, 1976 ; BOUXIN, 1978, 1983, 1986, 1987a,b, 1995 et 1999; BOUXIN & DEFLANDRE, 1988) ou comme on l'a vu, comme base d'étude des motifs plurispécifiques.

Le problème des distorsions qui apparaissent lors de la représentation des variables transformées dans un espace à 2 ou 3 dimensions a fait couler beaucoup d'encre. ESTÈVE (1978) met clairement en cause la

définition de la distance calculée entre les relevés et la méthode d'application de la configuration sur un plan plutôt que la linéarité du modèle. Il est aberrant de mettre en cause l'absence de linéarité entre la performance des espèces et les facteurs de l'environnement, puisque ces facteurs ne sont pas pris en compte, dans l'ordination indirecte du moins. Dans l'interprétation, il faut donc porter son attention sur la distance mutuelle entre les relevés pris en compte. Les distorsions sont créées par la différence entre la richesse spécifique des relevés (diversité β), mais c'est loin d'être la seule cause. La diversité β est une autre cause (DARGIE, 1986). DEL MORAL (1980) constate que PCA, RA ou NMDS donnent de très bons résultats avec une diversité β faible alors que la technique de BRAY & CURTIS (1957) en donne de mauvais. Les avis sont toutefois fortement partagés; ainsi GAUCH & SCRUGGS (1979) considèrent que RA est assez tolérant à une forte diversité β , du moins si on compare avec PCA; RA est d'ailleurs utilisé pour calculer cette diversité β [selon WHITTAKER & WOODWELL (1978), c'est la mesure du changement dans la composition en espèces le long d'un gradient environnemental; elle est calculée par le degré de diminution du logarithme de la similitude entre relevés au fur et à mesure qu'augmente la séparation le long du gradient déterminé par le premier axe de RA; elle est exprimée en "demi-changements" dans la composition des espèces]. OLSVIG (1979) présente différentes mesures de cette diversité. D'autres techniques de calcul ont été proposées par WILSON & MOHLER (1983) qui insistent sur le fait que le "bruit de fond" présent dans les tableaux de relevés peut obscurcir les motifs écologiques. Des exemples et calculs sont également donnés par ITOW (1991) et OKSANEN & TONTERI (1995). DEL MORAL & WATSON (1978) considèrent que les espèces rares contribuent fortement aux distorsions. LLOYD *et al.* (1994) expriment des idées proches de celles d'ESTÈVE (1978) quand affirment préférer une ordination indirecte (DCA) à une ordination sous contrainte (CCA et DCCA) car ils souhaitent estimer la végétation indépendamment des variables environnementales.

Un autre problème lié à l'analyse des correspondances est l'effet GUTTMAN (ESTÈVE, 1978) qui est l'apparition de facteurs artificiels fonctions du premier; lorsque cet effet est marqué, il est parfois utile de ne considérer que les facteurs 1 et 3.

Les méthodes d'analyse multivariée sont des méthodes le plus souvent utilisées d'un point de vue descriptif. L'inférence est rarement utilisée. Des techniques récentes facilitent le passage à l'inférence statistique: ce sont les tests de randomization, de bootstrap et de Monte Carlo (MANLY, 1997). Les tests de randomization ont été appliqués par BOUXIN (1999) aux analyses en composantes principales et aux analyses des correspondances; ils portent sur les contributions relatives des axes, des relevés et des espèces; les permutations concernent chaque ligne indépendamment (ici les espèces); on peut ainsi déterminer, avec un risque $\alpha = 0,05$ le nombre d'axes significatifs et les espèces ou les relevés présentant des contributions relatives significatives. La technique du bootstrap est utilisée par KNOX & PEET (1989) pour retenir le nombre d'axes utiles dans une DCA et par PILLAR (1999), pour déterminer le nombre de dimensions dans l'analyse d'un échantillon de relevés. Dans une "fuzzy set ordination", FULTON (1996) utilise un test de

Monte Carlo pour tester la robustesse de l'ensemble des données par rapport aux préférences environnementales calculées pour chaque espèce. ter BRAAK(1987 et 1990) et ter BRAAK & ŠMILAUER (2002), dans l'analyse canonique des correspondances, utilisent la distribution obtenue en permutant de manière aléatoire les variables environnementales entre les sites, de manière à évaluer la signification statistique des effets des variables environnementales. ALLEN *et al.* (1991) utilisent la technique du bootstrap pour évaluer la relation entre l'altitude et le premier axe de CCA. REED *et al.* (1993) utilisent un test de Monte Carlo pour évaluer la signification des axes produits par DCA et DCCA.

Quelques conclusions

Quelle technique choisir ? GREIG-SMITH (1964) constatait que le choix d'une technique à utiliser dans un cas particulier n'était pas facile à faire. Cela dépendait du type de données, de la diversité des données récoltées ou parfois des logiciels disponibles. Le choix dépendait surtout de l'objectif visé, du type d'information que l'on cherchait à extraire du tableau de relevés. Il était donc difficile de donner des recommandations générales. Mais le même auteur notait que, dans les cas où plusieurs techniques avaient été appliquées à un même ensemble de données, elles apportaient des résultats cohérents, ce qui était encourageant. Cette constatation a également été faite récemment par plusieurs auteurs, dont HUEBNER & VANKAT (2003) en ce qui concerne DCA, CCA et NMDS et SCHADE *et al.* (2003) en ce qui concerne DCA, PCA & NMDS. La robustesse des diverses techniques a été constatée mais les limites à cette robustesse sont très mal connues (AUSTIN *et al.*, 1994) ; ces auteurs soulignent l'importance qu'il y a de bien connaître la forme de la courbe de réponse des espèces le long de gradients environnementaux. KARADŽIĆ & POPOVIĆ (1994) augmentent les performances de PCA, dans la recherche de gradients, par une double standardisation des tableaux de relevés, portant à la fois sur les lignes et colonnes ; dans ce cas PCA est même supérieure à CA. Une bonne façon d'évaluer la qualité de la représentation des résultats d'une technique est d'en utiliser plusieurs et de comparer leurs apports respectifs (ØKLAND, 1996).

Il faut rappeler le manque d'aides à l'interprétation de plusieurs techniques. Seuls les logiciel ADE-4 et Rcmdr apportent ces aides, notamment les contributions relatives ou Cos^2 des variables ou individus.

Nous allons tenter de répondre à ces questions dans les chapitres qui suivent.

janvier 2024

Références

- 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 phyto-écologique 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 éodévoniennes, 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 Molignée (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.

- CHESSSEL, 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., CHESSSEL, 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. & CHESSSEL, 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, M.O. & 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.
- PERMUTATION
- 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). *Regression 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 : calcul des valeurs et vecteurs propres

Données d'origine

Nous présentons un exemple simple avec trois espèces, extrait du tableau Tailfer5 utilisé dans ce chapitre. Par rapport à ce tableau, plusieurs relevés vides ont été enlevés.

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

Exemples d'analyse en composantes principales

Matrice des données centrées réduites

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

Remarque : les totaux nuls le sont aux erreurs d'arrondis près.

Matrice de corrélation

	<i>Care pen</i>	<i>Care</i>	<i>Desc ces</i>
<i>Care</i>	1		
<i>Care</i>	-0,60247	1	
<i>Desc</i>	-0,54660	0,38490	1

Valeurs propres

$$\lambda_1 = 2,027863$$

$$\lambda_2 = 0,6118351$$

$$\lambda_3 = 0,353785$$

Vecteurs propres

	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

Exemple de calcul

Les trois déterminants

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

1 - 0,353785	-0,60247	-0,5466
-0,60247	1 - 0,353785	0,3849
-0,5466	-0,3849	1 - 0,353785

sont égaux à zéro.

Le produit de la matrice **A** et du vecteur **v**₁ est égal au produit du vecteur **v**₁ par la valeur propre λ_1 (aux erreurs d'arrondi près).

$$\begin{bmatrix} 1 & -0,60247 & -0,54660 \\ -0,60247 & 1 & 0,38490 \\ -0,54660 & 0,38490 & 1 \end{bmatrix} * \begin{bmatrix} -0,62046 \\ 0,566674 \\ 0,542136 \end{bmatrix} = \begin{bmatrix} -0,62046 \\ 0,566674 \\ 0,542136 \end{bmatrix} * 2,027863$$

De même, les deux autres produits impliquant respectivement les deuxième et troisième vecteurs propres et les deuxième et troisième valeurs propres conduisent aux mêmes égalités (aux erreurs d'arrondi près).

$$\begin{bmatrix} 1 & -0,60247 & -0,54660 \\ -0,60247 & 1 & 0,38490 \\ -0,54660 & 0,38490 & 1 \end{bmatrix} * \begin{bmatrix} -0,07933 \\ 0,642395 \\ -0,76226 \end{bmatrix} = \begin{bmatrix} -0,07933 \\ 0,642395 \\ -0,76226 \end{bmatrix} * 0,6118351$$

$$\begin{bmatrix} 1 & -0,60247 & -0,54660 \\ -0,60247 & 1 & 0,38490 \\ -0,54660 & 0,38490 & 1 \end{bmatrix} * \begin{bmatrix} 0,780219 \\ 0,515952 \\ 0,353624 \end{bmatrix} = \begin{bmatrix} 0,780219 \\ 0,515952 \\ 0,353624 \end{bmatrix} * 0,353785$$

Les coordonnées des variables floristiques s'obtiennent en multipliant chaque vecteur propre par la racine carrée de la valeur propre correspondante.

Par exemple, la coordonnée de *Carex pendula* sur le premier axe est égale à :
 $-0,62046 * \sqrt{2,027863} = -0,883554$.

Les coordonnées des relevés sur un axe s'obtiennent en multipliant les éléments des colonnes du tableau des données centrées réduites par les valeurs correspondantes des vecteurs propres de l'axe considéré. Exemple, la coordonnée du relevé 1 sur le premier axe est égal à :

$$(-0,7746 * 0,62046) + (-1 * 0,566674) + (0,57735 * 0,542136) = 0,226936$$

Exemple d'analyse des correspondances

Comme dans cet exemple, il n'y a que trois espèces pour 24 relevés.

Matrice pondérée

Rel	<i>Carex pendula</i>	<i>Carex remota</i>	<i>Deschampsia cespitosa</i>	Somme
-----	----------------------	---------------------	------------------------------	-------

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

Matrice de covariance

	<i>Care pen</i>	<i>Care rem</i>	<i>Desc ces</i>
<i>Care pen</i>	0,52849		
<i>Care rem</i>	-0,2344	0,220085	
<i>Desc ces</i>	-0,18232	-0,01396	0,140313

Valeurs propres

$$\lambda_1 = 0,698526$$

$$\lambda_2 = 0,190362$$

Vecteurs propres

<i>Carex pendula</i>	1,804984	-0,27453
<i>Carex remota</i>	-0,75147	-1,29819
<i>Deschampsia</i>	-0,40151	1,002723

Exemple de calcul

Les deux déterminants

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

sont égaux à zéro.

Les coordonnées des variables floristiques s'obtiennent en multipliant chaque vecteur propre par la racine carré de la valeur propre correspondante.

Par exemple, la coordonnée de *Carex pendula* sur le premier axe est égale à :

$$1,804984 * \sqrt{0,698526} = 1,508567.$$

Les coordonnées pour chaque relevé s'obtiennent en additionnant les produits des scores des espèces du tableau original (ici valant toujours 1) par les valeurs correspondantes des vecteurs propres, le tout étant pondéré par le totaux marginaux de chaque relevé.

Par exemple, la coordonnée du relevé 1 est égale à $-0,40151$, comme c'est la seule espèce présente. Pour le relevé 2, on a $(1 * (-0,75147) + 1 * (-0,40151))/2$ soit $-0,57649$.

Exemple d'analyse non-symétrique des correspondances

Les calculs se font à partir des colonnes (ici les relevés).

Matrice pondérée

Par souci de présentation, la matrice est permutée

	<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

Remarque : les totaux nuls le sont aux erreurs d'arrondis près.

Valeurs propres

$$\lambda_1 = 0,548992$$

$$\lambda_2 = 0,214318$$

Vecteurs propres

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

Le calcul des déterminants n'est pas présenté ici car trop long.

Les coordonnées des relevés s'obtiennent en multipliant la valeur du vecteur propre sur l'axe correspondant divisé par la somme sur la colonne du relevé, le tout multiplié par la somme totale du tableau.

Par exemple, la coordonnée du relevé 1 sur le premier axe est égale à :
 $-0,1006226/\sqrt{1}*\sqrt{39} = -0,6283878$.

Les coordonnées pour chaque espèce s'obtiennent comme indiqué sur la ligne suivante :
 $\text{cosp} = \text{cosp} + \text{don}[i,j]*V[j, b]/\sqrt{\text{hc}[j]*\sqrt{\text{nbrsp}/(\text{Lamb}[b]*\text{somme})}}$.

On calcule d'abord la constante $\sqrt{\text{nbrsp}/(\text{Lamb}[b]*\text{somme})}$:
 Pour le premier axe, on a : 0,3743219684

Par exemple, la coordonnée de *Carex pendula* (espèce présente dans les relevés 16, 17 et de 23 à 29) sur l'axe 1 est égal à

Pour R16 : $0,351624*0,3743219684$ (pour hc=1)
 Pour R17 : $0,1774848/\sqrt{2}*0,3743219684$ (pour hc=2)
 Pour R23 : $0,1774848/\sqrt{2}*0,3743219684$ (pour hc=2)
 Pour R24 : $0,351624*0,3743219684$ (pour hc=1)
 Pour R25 : $0,351624*0,3743219684$ (pour hc=1)
 Pour R26 : $0,351624*0,3743219684$ (pour hc=1)
 Pour R27 : $0,351624*0,3743219684$ (pour hc=1)
 Pour R28 : $0,0797998/\sqrt{3}*0,3743219684$ (pour hc =3)
 Pour R29 : $0,1774848/\sqrt{2}*0,3743219684$ (pour hc=2)

ce qui fait 0,816253037.

