

Chapitre 5. L'analyse des données ou l'ordination, réflexion illustrée avec un tableau floristique

par Guy BOUXIN[°]

Sommaire

Introduction	1
Méthodes	4
Le tableau de relevés et la nature des données	4
Analyses factorielles et techniques d'ordination	5
Données	8
Interprétation des résultats	9
Analyse du tableau floristique originel	10
Première transformation des données	14
Analyse en composantes principales des données ordinales.....	14
Analyse des correspondances des données ordinales.....	17
Analyse des principales coordonnées des données ordinales	20
Analyse non-métrique multidimensionnelle des données ordinales	23
Seconde transformation des données	25
Analyse des correspondances multiples.....	26
Analyse non symétrique des correspondances multiples	26
Analyse des principales coordonnées du tableau disjonctif simple	29
Une troisième voie pratique avec la transformation en tableau disjonctif 113 et en logarithme d'abondances	32
Transformation 113	32
Transformation logarithmique	35
Analyses factorielles multiples	38
Fichier disjonctif simple	39
Fichier 113	41
Fichier logarithmique.....	44
Fichiers floristique et environnemental réunis.....	47
Discussion et conclusions	50
Références	53

Abréviations: CA = analyse des correspondances; MFA = analyse factorielle multiple (suivie par CA, NSCA); NMDS = « non-metric multidimensional scaling »; NSCA = analyse non symétrique des correspondances; PCA = analyse en composantes principales. PCoA = analyse des principales coordonnées.

Introduction

Ce chapitre peut être lu seul. Il explique, pas à pas, notre choix final des techniques les mieux adaptées à une analyse approfondie d'un tableau de relevés de végétation.

Analyse statistique des données de végétation

Dans les tableaux de relevés de végétation, les lignes correspondent aux espèces et les colonnes aux relevés (individus). Chaque cellule correspond à la présence ou à l'abondance d'une espèce. En premier lieu, il est nécessaire de mettre en avant un ensemble de caractéristique des données de végétation : une grande diversité de types de données (présences, abondances, couvert, nombre d'individus, ...), une grande variabilité entre les sites, un mélange d'espèces communes, moins communes et rares, un mélange de motifs de dispersion et tout spécialement une forte proportion de cases vides (habituellement beaucoup plus de 50 %), chacune correspondant à l'absence d'une espèce et représentée par zéro. Toutes ces caractéristiques rendent l'analyse des données très difficile.

C'est la raison pour laquelle le concept d'analyse de données (sensu BOUROCHE & SAPORTA, 1980) est préféré. Il inclut un ensemble de méthodes très différentes d'analyse statistique qui sont adaptées aux tableaux. On distingue deux processus ; la classification automatique d'individus ou de variables et l'analyse factorielle, utilisant les propriétés de l'espace vectoriel euclidien pour décrire les individus ou les variables (FOUCART, 1982). Cet article ne porte que sur l'analyse factorielle. Selon PALM (1993), les méthodes d'analyse factorielle sont des outils de base de l'analyse des tableaux de données qui ne présentent aucune structure particulière. L'analyse des correspondances, créée pour traiter d'importantes données linguistiques, fut la première étape dans un large ensemble de techniques statistiques qui proposent de traiter les tableaux statistiques indépendamment de toute hypothèse mathématique ou probabilistique sur la structure des données (HUSSON *et al.*, 2009). Elles ont essentiellement un but descriptif, en condensant l'information d'un tableau, souvent composé d'un grand nombre de lignes et colonnes, en quelques représentations graphiques à deux dimensions, accompagnées de tableaux résumant les valeurs numériques dans le but d'aider l'utilisateur dans son interprétation. L'analyse des données est donc multidimensionnelle et descriptive. Elle produit des variables synthétiques qui représentent plusieurs autres, elles-mêmes liées entre elles.

Ces techniques semblent *a priori* bien adaptées à l'analyse des tableaux de relevés de végétation. Elles ont été utilisées depuis plusieurs dizaines d'années dans l'étude de la végétation sous le vocable « ordination ». L'ordination *sensu stricto* s'efforce de mettre la végétation en relation avec un ou plusieurs gradients environnementaux ou axes (GOODALL, 1954). En partant du concept que la végétation est un continuum, elle consiste à placer chaque relevé en relation avec un ou plusieurs axes de manière que leur position relative apporte un maximum d'information sur leur composition. Le processus synthétique est considéré comme un outil pour mettre les données floristiques en relation avec les facteurs environnementaux (ou autres) qui sont susceptibles d'influencer 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 écrit (traduction) : « L'analyse de gradient étudie la végétation en termes de gradient (ou variations plus ou moins continues) de l'environnement, de populations spécifiques et de communautés. L'arrangement des échantillons en fonction de gradients environnementaux comme base d'analyse de gradient est appelé ordination. L'ordination est soit directe (localisation des échantillons le long d'un ou plusieurs gradients environnementaux), soit indirecte (les échantillons sont rangés, par des calculs élaborés, le long d'axes qui indiquent les changements dans les communautés ». 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) appelle « diversité bêta ». Un axe ou un gradient avec une forte diversité bêta présente des compositions 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.

Analyse statistique des données de végétation

L'analyse factorielle, dans sa conception statistique, et l'ordination ont donc des bases conceptuelles très différentes en ce sens que l'analyse factorielle n'émet aucune idée préconçue sur la signification des axes. La manière d'exploiter les résultats devient complètement différente. Il y a une certaine confusion dans l'utilisation du terme « ordination ». Il est parfois utilisé en lieu et place d'analyse factorielle (WILDI, 2010 & 2013). Par exemple, selon GAUCH (1982) and beaucoup d'auteurs à sa suite, l'analyse des correspondances (CA) se révèle supérieure à l'analyse en composantes principales (PCA) comme technique d'ordination. CA et PCA produisent l'effet « fer à cheval » et spécialement avec PCA, les extrémités de l'arc se rejoignent parfois.

Dans son analyse de la littérature au niveau mondial, VON WEHRDEN *et al.* (2009) ont observés que CCA, DCA, CA, PCA ou NMDS sont les techniques communément utilisées. Si l'effet « fer à cheval » est généralement considéré comme un important défaut et conduit à rejeter PCA pour beaucoup de données de végétation, de nombreuses discussions méthodologiques et la diversité des méthodes disponibles ont conduit à une importante incertitude parmi les chercheurs et étudiants.

Les méthodes d'ordination courantes ont été revues et évaluées par AUSTIN (2012). Trois principaux paradigmes sont discutés :

- La phytosociologie traditionnelle avec la reconnaissance et la définition d'associations, la végétation est considérée comme discontinue ;
- Les analyses multivariées mettent l'accent sur l'utilisation de techniques d'ordination et de classification, avec trois paradigmes subordonnés associés à plusieurs méthodes – CA, PCoA et NMDS ;
- Le paradigme émergent est l'utilisation de méthodes de régression statistique.

CA est une méthode qui utilise la distance χ^2 comme mesure de similitude, PCoA (ou NMS) produit une ordination à partir d'une représentation euclidienne de relevés basées sur toute mesure de similitude choisie par l'utilisateur et NMDS construit une ordination dans laquelle la distance entre relevés trouve un accord maximum entre le rangement dans l'ordre et les mesures de similitude entre relevés (AUSTIN, 2012). Les différences entre ces paradigmes sont moins importantes que beaucoup d'écologues de la végétation le croient. Dans un autre article, l'utilisation de données artificielles sont mises en valeur dans le but d'évaluer les techniques ; la standardisation des données est aussi une étape de l'analyse importante et mal comprise.

On en déduit que le concept de gradient domine de manière permanente la recherche en écologie végétale. Quand le gradient est temporel, l'étude s'appelle une succession (AUSTIN, 1977 ; DEBUSSCHE *et al.*, 1996).

On constate aisément que, dans la plupart des études, seules les valeurs propres des axes et les scores des relevés et espèces sur les axes de l'ordination sont pris en compte dans l'interprétation des résultats. Les aides à l'interprétation, telles que les contributions relatives des espèces et relevés, sont rarement utilisées, probablement parce que non fournies par la plupart des programmes disponibles. Elles ont été utilisées par BONIN & ROUX (1978), BONIN *et al.* (1983) et BOUXIN (1986, 1987a & b, 1991, 1995 & 1999). Les pourcentages d'inertie associés aux axes sont présentés. Ils permettent d'évaluer le nombre d'axes à prendre en considération. Quand la somme des

Analyse statistique des données de végétation

pourcentages pour les deux ou trois premiers axes est importante (sans qu'il soit possible de fixer une norme), l'analyse multivariée est considérée comme réussie (voir ESPINAR *et al.*, 2002).

De nouvelles techniques sont régulièrement présentées, par exemple l'analyse des correspondances des données de lignes ou « raw data » (GREENACRE, 2010) ou le codage flou (fuzzy coding) dans les ordinations sous contrainte (GREENACRE, 2013), toujours en relation avec le paradigme de gradient. Toutefois, certaines analyses multivariées sont presque complètement ignorées, comme l'analyse non symétrique des correspondances (GIMARET-CARPENTIER, 1998) ou les analyses factorielles multiples qui analysent simultanément plusieurs tableaux correspondant au même ensemble de relevés (ESCOFIER & PAGES, 1984, 1986, 1989).

L'objectif de cette étude est de donner un exemple d'une approche scientifique conduisant à une analyse en profondeur de la structure d'un tableau de relevés, en utilisant des analyses multivariées, en dehors des paradigmes communs. Dans ce but, une végétation déjà analysée par l'auteur (BOUXIN 1975a, b and 1983) est prise comme exemple.

La question posée est la suivante : en utilisant les analyses multivariées, une approche dans l'esprit de l'analyse des données est-elle un meilleur outil dans l'étude de la végétation qu'une approche basée sur le concept d'ordination ?

Méthodes

Le tableau de relevés et la nature des données

Parmi les trois grandes catégories de données collectées (nominales, ordinales et métriques, voir WILDI, 2010 & 2013), il y a beaucoup de variantes de données floristiques : simples données de présence-absence (0-1), la fréquence dans des quadrats ou les contacts avec des aiguilles (comptages), la densité, l'abondance-dominance selon BRAUN-BLANQUET (échelle ordinale, 0, +, 1 à 5) souvent associée à un coefficient de sociabilité (1 à 5) et parfois convertie en pourcentage moyen pour chaque classe (1-2.5; 2-15.0; 3-37.5; 4-62.5; 5-87.5), ignorant les classes inférieures r et + (TALBOT *et al.* 2010), l'échelle de van der MAAREL, la couverture (voir FEHMI, 2010) pour le couvert des arbres ou la strate herbacée dense (%), l'aire basale (données métriques) et rarement la biomasse.

Certaines données sont mesurées, d'autres estimées et WILSON (2011) considère qu'il y a de sérieux problèmes avec les échelles de couverture-abondance, comme celle de BRAUN-BLANQUET. L'estimation de l'abondance « à l'œil » ou par un « invité bien informé », etc. donne des résultats trop souvent compromis dans la science de la végétation. Plusieurs opérations mathématiques sont possibles avec des présences, des densités, la couverture, l'aire de base mais pour d'autres données comme les données ordinales, certaines opérations mathématiques sont illicites. Par exemple, dans un relevé de forêt avec des arbres, des arbustes, des plantes herbacées, des mousses, que signifie réellement une somme d'abondance-dominances avec un tel ensemble d'espèces ? dans CA (ou DCA), les sommes des scores spécifiques des lignes ou des colonnes d'un tableau entre dans le calcul de la distance χ^2 et le choix du score est primordial (STEVENS *et al.*, 2011).

Analyse statistique des données de végétation

Les variables originelles sont aussi parfois transformées en variables secondaires (ORLÓCI & KENKEL, 1985): racine carrée, logarithme, sinus inverse, parmi beaucoup d'autres, ce qui change la forme de la relation entre les variables. Plusieurs autres transformations ou standardisations sont aussi possibles (WILDI, 2010 & 2013).

Un moyen commun de changer les scores originels des espèces en analyse des données est de transformer le tableau originel en un tableau ordinal, en remplaçant les variables primaires par une variable ordinale en utilisant plusieurs scores entiers ou modalités (0, 1, 2, ...) dont le nombre dépend de la variabilité des données. Une seconde étape conduit à un tableau disjonctif simple, dans lequel chaque variable est remplacée par une ou plusieurs nouvelles variables, chaque correspondant à une des modalités. Dans chaque ligne, la présence de la modalité est représentée par l'unité, et l'absence par zéro. Un tableau de relevés avec seulement des données de présence-absence est la forme la plus simple du tableau disjonctif simple. Une autre transformation est le tableau disjonctif complet, dans lequel le zéro des variables ordinales est aussi considéré comme une modalité (FOUCART, 1982). Un exemple simple est donné avec les abondances d'une espèce d'arbre le long d'un ruisseau, dans un ensemble de vingt relevés.

Alnus glutinosa 0+2142121220+5101022

Le tableau disjonctif simple avec les modalités +, 1, 2, 4 or 5 est

Alnus glutinosa+ 01000000000010000000

*Alnus glutinosa*1 00010010100000101000

*Alnus glutinosa*2 00100101011000000011

*Alnus glutinosa*4 00001000000000000000

*Alnus glutinosa*5 00000000000010000000

Pour obtenir un tableau disjonctif complet, une ligne est ajoutée :

Alnus glutinosa absences 10000000000100010100

Dans ce dernier tableau, les sommes sur les colonnes sont identiques.

Comme beaucoup d'analyses multivariées sont fortement influencées par les espèces rares dans un petit nombre de relevés (voir GAUTHIER, 2010 parmi beaucoup d'autres), une solution commune est d'éliminer les espèces rares à partir d'un seuil fixé arbitrairement. Cette solution est difficile à accepter et discutée dans les résultats.

Les relations entre les variables floristiques ne sont pas simples. Elles s'expriment rarement sous forme d'une élégante courbe linéaire ou curvilinéaire et entrent difficilement dans un modèle théorique. La variabilité des données biologiques est souvent très grande et les relations entre ces variables sont souvent très complexes, multifactorielles.

Toutes ces données doivent être intégrées dans un tableau et une adéquation entre les données et le modèle mathématique choisi est nécessaire.

Analyses factorielles et techniques d'ordination

Les étapes de calcul de l'analyse en composantes principales, de l'analyse des correspondances, l'analyse non symétrique des correspondances et l'analyse factorielle multiple sont rappelées :

1. Calcul d'un tableau transformé, avec le même nombre de lignes et de colonnes que le tableau originel
2. Calcul d'une matrice carrée de covariance (*sensu lato*), dont la dimension est égale au nombre d'espèces ou de relevés.
3. Calcul des valeurs propres et des vecteurs propres.

Analyse statistique des données de végétation

4. Calcul des coordonnées des espèces et des relevés sur les axes principaux et des contributions et probabilités associées.

Le tableau transformé contient :

- Pour l'analyse en composantes principales, les données de chaque ligne sont sommées et centrées avec une variance unitaire. La somme dans chaque ligne est donc égale à zéro. Les lignes sont sommées mais pas les colonnes. Les espèces rares ont des scores relativement élevés et les cases vides reçoivent un score qui n'est pas le même dans les différentes lignes. Dans le calcul du coefficient de corrélation, une grande variété de variables peut être traitée. Théoriquement, les meilleures conditions se présentent avec des variables monotones qui donnent des coefficients positifs ou négatifs élevés mais des coefficients élevés sont aussi observés avec d'autres distributions de variables. Par exemple, la cooccurrence d'espèces rares dans un relevé génère des coefficients de corrélation élevés.

- Dans l'analyse des correspondances, les lignes et colonnes sont sommées et sont pondérées par les totaux marginaux. On obtient les variables transformées en multipliant les données originelles par le total de toutes les données et en le divisant par le produit des totaux marginaux sur les lignes et colonnes correspondantes ; à ce quotient, on soustrait la valeur 1. Les scores des cases du tableau transformé varient entre -1 (espèce absente) et une valeur positive, qui est d'autant plus élevée que les totaux marginaux des lignes et colonnes correspondantes ne sont pas très élevés (espèce rare dans un relevé peu diversifié, par exemple). La matrice de covariance est alors calculée. Les espèces rares sont donc souvent surestimées. Cette technique est adaptée aux données de présence ou de fréquence (tableaux de contingence). Elle produit les mêmes valeurs propres avec le tableau de contingence transposé. L'utilisation de variables continues est proscrite. Considérant l'algorithme, il est clair que la sommation de plusieurs types de variables sur les colonnes n'est pas recommandée ; faire une somme concernant des présences and plusieurs mesures d'abondance (couvert, abondance-dominance) est un non-sens. Toutefois, CA et les techniques dérivées (comme l'analyse des correspondances détendancée) sont communément utilisées pour des tableaux d'abondance-dominance.

- Dans l'analyse non symétrique des correspondances, les variables transformées sont obtenues en divisant les données originelles par le total de la colonne correspondante et en soustrayant de ce résultat le quotient entre la somme de la ligne correspondante et le grand total du tableau. Dans le tableau transformé, les sommes sur les colonnes valent zéro et les totaux sur les lignes sont d'autant plus élevés que les espèces sont plus fréquentes ou abondantes. Comme pour PCA, les cases vides reçoivent un score qui n'est pas le même dans les différentes lignes. La matrice de variance est alors calculée. La restriction au sujet du type de donnée est la même que pour CA. Avec un tel algorithme, le poids des espèces est d'autant plus important que la somme sur les lignes est élevée. Le poids des espèces rares n'est pas surestimé comme dans CA. Le calcul des valeurs propres et des vecteurs propres est le même pour les deux analyses.

Les analyses multivariées sont le plus souvent utilisées comme techniques descriptives. L'inférence statistique est rarement prise en compte. Plusieurs techniques facilitent le passage à l'inférence statistique : ce sont les méthodes de randomisation, de bootstrap et de Monte Carlo (MANLY, 1997). Plusieurs techniques ont été développées pour évaluer le nombre d'axes significatifs (voir KNOX & PEET, 1989 ; REED *et al*, 1993 ; PILLAR, 1999).

J'ai développé un algorithme automatique avec randomisation (BOUXIN, 1999) appliqué à l'analyse en composantes principales et à l'analyse des correspondances ; il s'applique aux contributions relatives des axes,

Analyse statistique des données de végétation

colonnes et lignes. Le test est facilement construit sur les contributions relatives comme ces dernières sont toujours positives ; ce sont des éléments importants dans l'interprétation des résultats comme elles entrent directement dans le calcul des scores des espèces et relevés sur les axes. A chaque étape de permutation, toutes les lignes de la table (ici les espèces) sont permutées séparément. C'est une longue procédure, comme une PCA, une CA ou une NSCA est calculée pour chaque table permutée ; les contributions relatives des axes, des espèces et des relevés sont comparées avec les mêmes paramètres de la table originale. Si la proportion de contributions relatives des tables permutées supérieures à la contribution relative de la table originelle est inférieure à 0,05, alors la contribution relative est significative. Le nombre de permutations est fixé en fonction de nombre d'espèces (pour PCA) ou le nombre de relevés (pour CA ou NSCA) ; il varie entre 1000 pour de grands tableaux et 100.000 (le nombre idéal) pour les petits tableaux.

L'analyse factorielle multiple est une méthode d'analyse simultanée de plusieurs tables. La MFA est ainsi une généralisation des techniques de couplage de tableaux. C'est une des méthodes les plus pratiques parmi celles d'analyse simultanée de K tableaux (DRAY *et al.*, 2007). Avec cette technique, les K tableaux doivent avoir le même ensemble d'individus (relevés) en commun, chacun correspondant à un groupe différent de variables (espèces ou paramètres environnementaux). La MFA rend uniforme l'incidence de chaque tableau dans l'analyse simultanée ; elle consiste à mettre les premières valeurs propres (de PCA, par exemple) de chaque tableau sur une même échelle. Trois étapes principales sont nécessaires pour réaliser une analyse factorielle multiple. Premièrement, une PCA, une CA ou une NSCA du tableau de relevés complet (obtenu en réunissant tous les tableaux correspondant au même ensemble de relevés) est calculée. Deuxièmement, la même analyse est appliquée à chaque tableau élémentaire. Troisièmement, une nouvelle analyse est appliquée au tableau complet, mais en introduisant les premières valeurs propres des tableaux séparés comme paramètre de pondération (exactement la racine carrée de cette première valeur propre). Cela implique qu'aucune colonne ne soit vide dans les tableaux séparés. Cette condition rend la MFA impossible avec certains tableaux.

Cette technique a rarement été utilisée (voir BOUXIN, 2011 & 2013 ; DEVINEAU, 2001 ; HAURY & BAGLINIÈRE, 1990). Ce type d'analyse est adapté à de nombreuses études de végétation, par exemple aux végétations pluristratales. Les logiciels répandus produisent les coordonnées des espèces et des relevés le long des premiers axes de l'analyse, mais sans aucune aide à l'interprétation. Mais le calcul des contributions relatives des relevés et espèces pour chaque axe et les probabilités associées sont utiles. Ainsi, les relevés et les espèces peuvent être rangés en fonction de leurs contributions relatives qui ont priorité dans l'interprétation. Aucun test n'est calculé pour les analyses multiples comme les tests par permutation sont appliqués aux tableaux séparés. Mes logiciels produisent aussi le tableau transformé, la matrice de covariance, la distribution des coefficients de covariance en 20 classes et les vecteurs propres.

L'analyse des principales coordonnées (PCoA) est une généralisation de l'analyse en composantes principales. Elle a été proposée par GOWER (1966). Par rapport à une PCA, elle calcule la distance entre points de l'espace défini par les espèces ou les relevés en utilisant la distance euclidienne. Toutefois, PCoA permet l'utilisation d'autres mesures de distances afin de mieux refléter la distance entre relevés, par exemple. La distance peut être le nombre d'espèces présentes dans un relevé et un autre mais pas dans les deux (PIELOU, 1977), le coefficient de BRAY & CURTIS ou encore une distance corde.

Analyse statistique des données de végétation

L'analyse non métrique multidimensionnelle est une technique originale, proposée par KRUSKAL (1964a, b). Elle est expliquée par ORLÓCI & KENKEL (1985) et MCCUNE & MEFFORD (1999) mais elle appelle quelques remarques :

- elle commence aussi avec le calcul d'une matrice de distance ; comme pour PCoA, le choix du coefficient de distance est une étape importante (ROOT & NELSON, 2011),
- la première dimension dans un espace à deux dimensions n'est pas la même que la première dimension dans un espace à trois dimensions. Pour un nombre donné de dimensions, la solution pour un axe particulier est unique,
- le nombre approprié d'axes devrait être déterminé par l'évolution du stress final en fonction du nombre d'axes. Il faut choisir un nombre d'axes au-delà duquel la réduction du stress est faible.

La solution est influencée par l'ordre initial de relevés du tableau originel. Si on choisit l'option aléatoire, plusieurs analyses produisent des résultats différents. Il y a beaucoup de variantes dans cette technique (PRENTICE, 1977 et 1980 ; MINCHIN, 1987). C'est donc une technique complètement différente de CA et PCA comme il n'y a pas de calcul de vecteur propre.

Données

La démarche scientifique est illustrée avec un tableau floristique, avec 238 variables floristiques, provenant d'un ensemble de 80 relevés. Les données de savane et les principaux résultats ont été présentés dans BOUXIN (1975a, b and 1983). Les 80 relevés ont été placés dans un transect de 2400 x 200 mètres entre le lac Ihéma et la colline Kionja. Les relevés de 25 x 10 mètres ont été localisés le long de 40 traverses de 200 mètres de long. Le long de chaque traverse, deux relevés ont été sélectionnés au hasard parmi huit possibilités. La mesure de l'abondance des espèces change avec le type biologique. Pour les arbres, et chaque fois que c'était possible, la circonférence du tronc à hauteur de poitrine (d.b.h.) a été mesurée et les données transformées en aire basale (ou surface terrière). L'abondance des arbres adultes de chaque espèce dans un relevé est donnée par la somme des aires de base de tous les individus mesurables.

Pour tous les jeunes arbres et les arbustes de moins de 1,5 mètre de haut, l'abondance est exprimée par le nombre d'individus. L'abondance de chaque espèce d'arbre et d'arbuste est donc donnée par deux variables : la première pour les individus adultes et la seconde pour les petits. Comme beaucoup d'espèces herbacées forment des touffes, l'abondance a été estimée dans chaque relevés par un ensemble de 125 points régulièrement espacés le long de cinq lignes parallèles au petit côté du rectangle, à égale distance ; chaque point d'échantillonnage est matérialisé par une fine tige ; pour les espèces moins abondantes trouvées entre les lignes, la présence est notée.

Dans les tableaux de relevés, les variables floristiques des arbres avec une surface terrière mesurables sont suivis de A et les plus jeunes de J.

Pour rendre l'interprétation des divers résultats plus explicite, nous avons procédé à une classification des relevés avec recherche des espèces caractéristiques, en se basant sur l'analyse factorielle multiple d'un fichier 113, suivant les techniques décrites dans le chapitre sur la classification des relevés.

En quittant le lac Ihéma en direction de la colline Kionja, les communautés suivantes ont été rencontrées :

Analyse statistique des données de végétation

- Savane boisée et dense, étroite, avec des grands *Acacia polyacantha* subsp. *campylacantha* sur sol humifère, le long du lac, relevés A, B et C.
- Savane boisée à *Acacia hockii* et *Themeda triandra* >1, avec *Solenostemon thyrsoflorus*, *Plectranthus caninus*, relevés 1, 2, 3, 4, 5, 6, 7, 8, 28, 30, 32, 37, 42, 46, 48, 50, 51, 52, 54.
- Bosquets xérophiles à *Acacia brevispica*, *Teclea nobilis*, *Haplocoelum gallense*, *Grewia trichocarpa*, ..., relevés 9, 10, 11, 20, 29, 31, 38, 44.
- Savane arbustive à *Acacia brevispica*, *Boscia angustifolia*, *Markhamia obtusifolia*, *Commelina erecta* subsp. *livingstonii*, *Blepharis maderaspatensis*, *Plectranthus pauciflorus*, relevés 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 24, 39.
- Savane boisée à *Acacia gerrardii*, *Acacia siberana*, *Eustachys paspaloides* >1, *Indigofera colutea*, *Panicum atosanguineum*, relevés 22, 25, 26, 27, 58, 60, 61, 63.
- Savane herbacée à *Heteropogon contortus*, *Microchloa kundtii*, *Sporobolus stapfianus*, *Vernonia miombicola*, relevés 33, 35, 36, 40, 41, 43, 45, 47.
- Savane boisée assez dense à *Acacia senegal* (principalement au pied de la colline), *Pennisetum polystachion*, *Achyranthes aspera*, *Panicum maximum* >1, *Albuca kirkii*, relevés 34, 59, 62, 64, 65, 66, 67, 68, 69, 70.
- Savane herbacée à *Hyparrhenia dissoluta*, *Hibiscus cannabinus*, *Commelina purpurea*, relevés 49, 53, 55, 56, 57.
- Savane herbeuse ouverte ou boisée sur sol rocheux et peu profond, dominée par la graminée *Loudetia simplex*, relevés D, E, F, G, H, I, K, L.

Les motifs de dispersion ont aussi été décrits pour chaque espèce et pour chaque paramètre environnemental ou pédologique (BOUXIN, 1983).

Interprétation des résultats

L'interprétation des résultats se base sur :

- La distribution des coefficients de corrélations entre variables floristiques, des covariances ou des distances entre relevés. Une très forte asymétrie de ces paramètres a une forte incidence sur le calcul des valeurs propres et signifie que la technique statistique n'est pas en adéquation avec les données du tableau de relevés.
- L'importance des valeurs propres rangées par ordre décroissant.
- Les coordonnées des espèces et relevés ou des relevés seuls sur les axes de l'analyse multivariée, ce qui conduit à la représentation des variables floristiques et mésologiques dans des plans. Seuls les deux premiers axes sont illustrés, les résultats complets sont donnés en annexe.
- Les contributions relatives des espèces ou relevés sur les axes.
- Les probabilités associées aux valeurs propres ou aux contributions relatives des variables floristiques ou des relevés, ce qui révèle les espèces caractéristiques ou non, les simples espèces compagnes, au sens phytosociologique du terme.

Analyse statistique des données de végétation

Dans les analyses des principales coordonnées, les programmes fournissent seulement les coordonnées des relevés sur les axes et les valeurs propres. Dans l'analyse non-métrique multidimensionnelle, seules les coordonnées des relevés sont données.

Analyse du tableau floristique originel

Le tableau original comprend 238 variables floristiques (fichier Akagera80.txt dans les Annexes). Il y a une proportion très élevée de cases vides (81,01 %). Un petit nombre de relevés est caractérisé par la co-occurrence d'espèces rares. Avec les données originelles, PCA est la seule technique possible. CA et les techniques dérivées ne sont absolument pas adaptées à ces données ainsi que toute technique basée sur un calcul de distance ou de similitude.

Dans la table transformée de PCA, les variables centrées réduites correspondant aux absences des espèces varient entre -2 et des scores proches de zéro. La matrice de corrélation montre une grande asymétrie (figure 1) dans la distribution des coefficients, orientée vers le côté négatif de la distribution. Il y a une très forte proportion (82,25 %) de coefficients entre -0,168 et 0,123. Les cinq premières valeurs propres décroissent rapidement et leur somme correspond à 26,419 % de la trace de la matrice et ensuite, leur décroissance est lente et régulière. Les douze premières sont significatives (Figure 1).

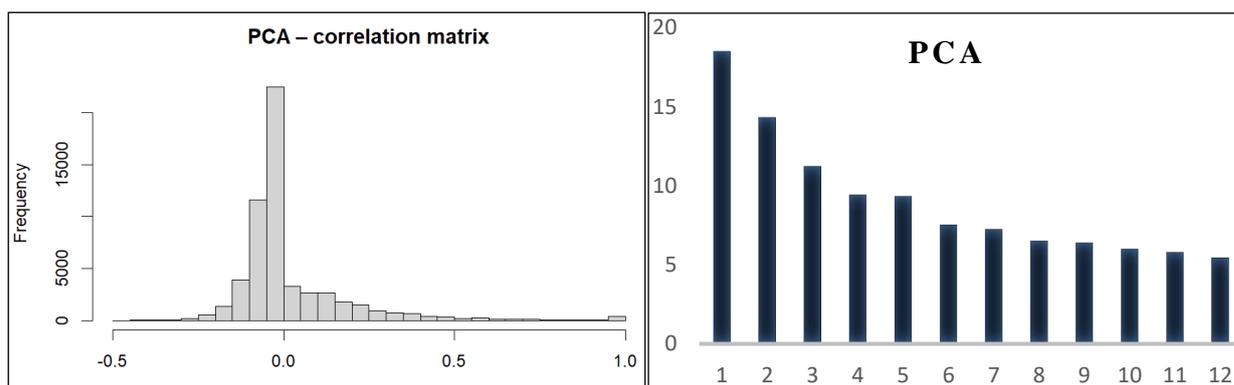


Figure 1. Distribution des coefficients de corrélation en 40 classes et les douze premières valeurs propres.

Les résultats partiels de la PCA sont illustrés dans le tableau 1 avec les deux premiers axes, avec les huit premières variables floristiques et les huit premiers relevés. Apparaissent : les coordonnées des variables et des relevés sur les axes, les valeurs propres ou les contributions relatives et les probabilités associées aux valeurs propres ou aux contributions relatives.

	coord1	Cr%1	P1	coord2	Cr%2	P2
Vp,Vp% et P	18,51314944	7,811455461	0	14,32723806	6,045248126	0
<i>Acacia senegal</i> A	-0,19571259	0,038303417	19,2	0,10248238	0,010502638	51,5
<i>Acacia senegal</i> J	-0,18995608	0,036083311	15,8	0,021164268	0,000447926	86,7
<i>Acacia hockii</i> A	-0,20201625	0,040810566	19,3	0,122969538	0,015121507	45,2
<i>Acacia hockii</i> J	-0,23943011	0,057326777	12,7	0,0874593	0,007649129	55
<i>Acacia gerrardii</i> A	-0,03428807	0,001175672	74,2	0,048950513	0,002396153	68,5
<i>Acacia gerrardii</i> J	0,048073944	0,002311104	62,9	0,075515768	0,005702631	58,9
<i>Acacia polyacantha</i> A	-0,04223656	0,001783927	67	-0,04340878	0,001884322	72,7
<i>Acacia sieberana</i> A	-0,00872359	7,61E-05	93,1	0,032843132	0,001078671	79,7
oooooooo	coord1	Cr%1	P1	coord2	Cr%2	P2
A	-1,29898443	0,003645468	54,3	-1,04010962	0,002337242	66,1

Analyse statistique des données de végétation

B	-0,74014334	0,002097109	67,6	-0,70639926	0,001910248	68,8
X1	-3,0228084	0,023309792	17,7	1,177807502	0,00353888	60,5
X2	-3,48515046	0,026031501	13,8	0,956005392	0,00195874	68,9
X3	-2,43930017	0,038617455	10,7	1,469432141	0,014013681	30,6
C	-0,97157414	0,003333479	57,9	-1,17653786	0,004888298	51,5
X4	-1,39749233	0,007006308	41,6	2,332630088	0,019520103	22,9
X5	-2,41115364	0,015005441	25,8	1,836433007	0,008704607	41,4

Tableau 1. Résultats partiels de l'analyse en composantes principales du fichier Akagera brut, pour les huit premières espèces et les huit premiers relevés. Coord : coordonnée sur les axes ou valeurs propres, Cr% : % d'inertie des valeurs propres ou contribution relative des espèces ou relevés sur les axes. P : probabilités associées aux valeurs propres ou aux contributions relatives.

Voici les résultats pour les quatre premiers axes avec les dix premières espèces et les dix premiers relevés rangés par ordre décroissant de contribution relative. Les résultats complets se trouvent dans le fichier Akagera80pca.xls (Annexes).

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Acacia brevispica</i> J	0,762585664	0,581536895	0	X9	24,53609198	0,577846353	0,7
<i>Jasminum schimperi</i>	0,751526263	0,564791724	0	X11	13,97306655	0,402345314	0,7
<i>Panicum brevifolium</i>	0,729652577	0,532392883	0	X10	10,60380036	0,25775096	1,3
<i>Tricalysia ruandensis</i> J	0,721874407	0,52110266	0,4	X18	5,466898187	0,154947798	1,3
<i>Pupalia lappacea</i>	0,699407098	0,489170288	0	X14	6,437972862	0,141537766	2,1
<i>Cyanotis lanata</i>	0,694475938	0,482296829	0	X13	7,266526484	0,14098438	2,2
<i>Psilotrichum elliotii</i>	0,676476692	0,457620715	1,4	X20	4,778346309	0,105484267	3,7
<i>Vernonia brachycalyx</i>	0,674331474	0,454722936	0	X66	-3,3395508	0,095853114	2,8
<i>Canthium lactescens</i> A	0,670569496	0,449663449	2,4	X15	5,134898167	0,088941845	3,2
<i>Allophylus rubifolius</i> J	0,66609473	0,443682189	1,7	X50	-3,50971725	0,0884272	3,3
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Pappea capensis</i> J	-0,82045405	0,673144852	0,2	F	-26,8941049	0,699190758	0
<i>Canthium vulgare</i> A	-0,79990289	0,639844633	0,3	K	-8,78060644	0,132258256	3
<i>Olea europaea</i> var. <i>africana</i> A	-0,79939663	0,639034965	0,4	H	-7,68915933	0,129159165	3
<i>Pappea capensis</i> A	-0,79939663	0,639034965	0,2	E	-6,27699606	0,087575895	3
<i>Allophylus africanus</i> J	-0,79939663	0,639034965	0,1	G	-6,38895057	0,068398766	8
<i>Indigofera dendroides</i>	-0,79939663	0,639034965	0,2	I	-4,09766745	0,055891745	#
<i>Ziziphus mucronata</i> A	-0,79327214	0,629280688	0,3	X9	6,940871032	0,046241248	8
Rubiaceae	-0,73151367	0,535112247	0,1	D	-2,74345102	0,032652776	#
<i>Haplocoelum gallaense</i> A	-0,73105462	0,534440854	0,4	X16	2,401468383	0,02612025	#
<i>Loudetia simplex</i>	-0,70341041	0,494786206	0	X11	3,409192179	0,023950714	#
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Kalanchoe integra</i>	-0,62626438	0,392207079	0,6	K	-13,5072317	0,312972711	1
<i>Rhoicissus tridentata</i>	-0,60106497	0,361279101	0,4	G	-11,6531707	0,227550308	1,2
<i>Tricalysia ruandensis</i> A	-0,55830165	0,311700729	0,2	F	12,14007294	0,142470215	2,4
<i>Parinari curatellifolia</i> A	-0,55220311	0,304928273	0,7	I	-6,46402548	0,139085024	2,9
<i>Dalbergia nitidula</i> J	-0,55025083	0,30277598	0,9	X9	-10,7225196	0,110355833	2,9
<i>Allophylus rubifolius</i> J	-0,49667153	0,246682609	1,3	H	-6,68669056	0,097676471	5,2
<i>Commelina bracteosa</i>	-0,48241912	0,232728212	2,2	X17	5,230411148	0,080876426	5,5
<i>Parinari curatellifolia</i> J	-0,47558396	0,226180102	1,5	X20	3,450101567	0,054991581	9,4
<i>Lannea schimperi</i> A	-0,45343907	0,205606993	2,7	X12	4,192425155	0,054787445	8,2
<i>Ficus glumosa</i> A	-0,45343907	0,205606993	2,2	B	-3,57801123	0,049008647	9,5
Species	coord4	Cr%4	P4	Relevés	Coord4	Cr%4	P4
<i>Bidens pilosa</i>	-0,56355414	0,317593273	0,1	X7	-10,7347085	0,234722375	0,9
<i>Achyranthes aspera</i>	-0,47454213	0,225190234	1,3	A	-9,62302007	0,200063655	1,4
<i>Panicum atrosanguineum</i>	0,452427304	0,204690465	0,3	X33	7,815465399	0,189212812	1,4
<i>Heliotropium strigosum</i>	0,438973219	0,192697487	0,5	X36	5,62058449	0,169181598	1,6
<i>Indigofera colutea</i>	0,431514461	0,18620473	0,8	B	-5,95482477	0,135746078	2,2
<i>Hibiscus cannabinus</i>	0,422760864	0,178726748	0,6	X32	3,887219281	0,092894753	4,4
<i>Justicia matammensis</i>	-0,41650734	0,173478361	2,5	X44	2,615658056	0,076035514	5,7

Analyse statistique des données de végétation

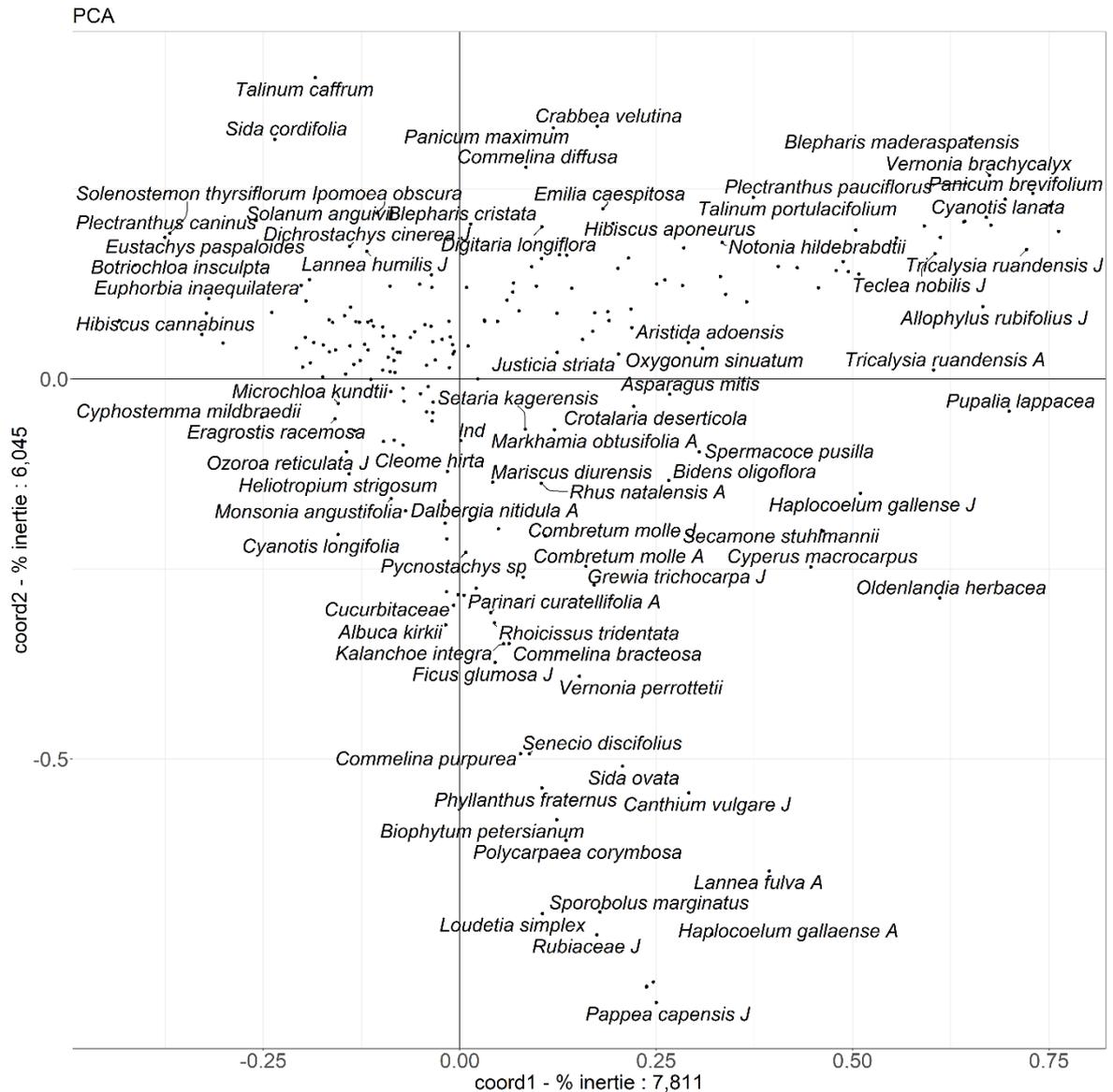


Figure 3. Analyse en composantes principales du fichier Akagera original. Coordonnées des espèces sur les axes 1 et 2.

Cette analyse met principalement en évidence des espèces localisées et se révèle peu intéressante sur le plan synthétique. Les principaux groupements végétaux et les espèces fréquentes sont assez bien écrasées par l'analyse.

Y a-t-il un moyen d'améliorer les résultats ? On pourrait être tenté d'éliminer les espèces rares, comme c'est souvent le cas, mais cette approche est difficilement acceptable en termes de phytosociologie. Premièrement, va-t-on éliminer les espèces avec 1, 2, 3 présences ou plus ? On se trouve en face d'un choix arbitraire. En plus, il apparaît souvent que certains relevés sont caractérisés par un ensemble d'espèces localisées, donc peu fréquentes dans le tableau de relevés. Enlever des espèces rares revient à amputer certains relevés d'un nombre important d'espèces. En plus, en forêt dense ou en savane, des espèces rares montrent parfois d'importantes biomasses et ont donc une grande importance dans le fonctionnement de l'écosystème. Une seconde étape serait l'élimination d'espèces qui n'ont pas de motif de dispersion structuré dans le tableau (BOUXIN, 1983) mais alors, certaines espèces communes sont éliminées du tableau et ce moyen n'est plus recommandé.

Analyse statistique des données de végétation

Nous explorons d'autres adaptations : la transformation des données et le choix d'analyses multivariées qui sont mieux adaptées à de telles données.

Première transformation des données

Avant de démarrer une nouvelle analyse, l'étape suivante est de transformer le tableau originel et de remplacer les abondances par une variable ordinaire variant de 1 à 5 (tableau 3), comme c'est recommandé par PODANI (2000), voir le fichier Akagera80phyto.txt en Annexes.

Variables	Les cinq coefficients et les limites de classe				
	1	2	3	4	5
Aire basale in dm ²	< 5	5-9.9	10-14.9	15-19.9	≥ 20
Nombre de contacts avec touffes	1-25	26-50	51-75	76-100	> 100
nombre d'individus	1-20	21-40	41-60	61-80	> 100

Tableau 3. Transformation des abondances originelles des espèces en valeurs ordinales.

Dans le tableau de relevés, il y a plusieurs lignes possibles pour chaque espèce ; le nom seul signifie que le coefficient est de 1, pour les abondances supérieures, les noms sont suivis de 2, 3, 4 ou 5 (en plus de A ou J pour les espèces ligneuses). Avec une échelle ordinaire, plusieurs analyses sont possibles : PCA, PCoA et NMDS. CA n'est pas vraiment recommandée car cette technique est adaptée à l'origine à des tableaux de présences ou de contingence. Toutefois, CA (ou DCA) est communément utilisée avec des échelles ordinales. NSCA n'est adaptée à aucune donnée d'abondance, puisque seulement les espèces abondantes sont mises en valeur.

Analyse en composantes principales des données ordinales

Les coefficients de corrélation sont un peu plus étalés que dans l'analyse précédente et les douze premières valeurs propres sont significatives (figure 4). Les principaux résultats sont présentés dans le tableau 4 et les figures 5 et 6. Les résultats complets se trouvent dans le fichier Akagera80phytopca.xls

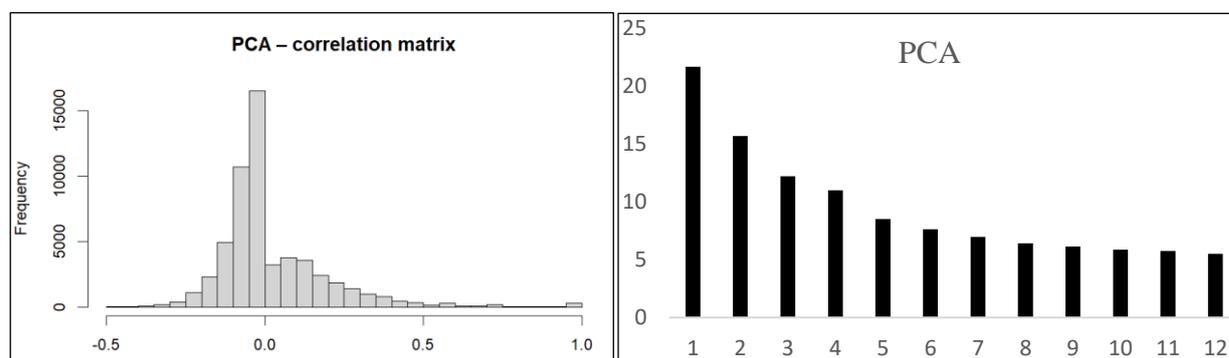


Figure 4. Distribution des coefficients de corrélation en 40 classes et les douze premières valeurs propres.

En rangeant sur l'axe 1, les espèces et relevés par ordre décroissant de contribution relative (tableau 4), on retrouve, comme dans l'analyse des données brutes, des espèces et relevés de bosquets xérophiiles.

L'axe 2 correspond aussi aux espèces et sites des savanes herbeuses sur sol superficiel avec des gros blocs ou dalles rocheuses. L'axe 3 est redondant.

Analyse statistique des données de végétation

L'axe 4 mélange des relevés de galerie forestière et de savane sur sol superficiel.

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Acacia brevispica</i> J	0,851106579	0,72438241	0	X9	20,1124359	0,4694869	0,48
<i>Sansevieria cylindrica</i>	0,778397524	0,60590271	0	X11	13,7051696	0,41105965	0,54
<i>Cyanotis lanata</i>	0,737597943	0,54405072	0	X10	12,279254	0,36155667	0,76
<i>Pupalia lappacea</i>	0,722954864	0,52266373	0	X18	9,1552484	0,30348263	0,89
<i>Panicum brevifolium</i>	0,719676534	0,51793431	0	X15	6,9404216	0,25915897	1,26
<i>Plectranthus pauciflorus</i>	0,718605418	0,51639375	0	X14	8,38343057	0,25271443	1,21
<i>Vernonia brachycalyx</i>	0,710244468	0,5044472	0	X20	8,17870591	0,22304754	1,55
<i>Blepharis maderaspatensis</i>	0,701139934	0,49159721	0	X13	9,18001198	0,20831384	1,3
<i>Jasminum schimperi</i> = <i>eminii</i>	0,681206363	0,46404211	0,04	X51	-5,3110084	0,14171469	2,48
<i>Teclea nobilis</i> A	0,668866734	0,44738271	0,04	X50	-3,9845708	0,13806026	2,4
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Loudetia simplex</i>	-0,81872329	0,67030783	0	K	-16,323878	0,39597085	0,18
<i>Parinari curatellifolia</i> J	-0,63174381	0,39910024	0,19	F	-17,390699	0,30019082	0,74
<i>Dalbergia nitidula</i> J	-0,62866935	0,39522515	0,16	G	-12,813812	0,2596882	0,87
<i>Parinari curatellifolia</i> A	-0,62866935	0,39522515	0,12	H	-10,307321	0,23760869	1,09
<i>Kalanchoe integra</i>	-0,61784043	0,3817268	0,03	I	-8,6931168	0,23121891	0,94
<i>Lankea fulva</i> A	-0,59487191	0,35387259	0,16	E	-7,2096235	0,1145078	3,4
<i>Rhoicissus tridentata</i>	-0,5892173	0,34717703	0,55	D	-4,2463931	0,07066902	6,83
<i>Canthium vulgare</i> A	-0,58619733	0,34362732	0,98	X19	3,91441116	0,06376302	7,8
<i>Ficus glumosa</i> J	-0,56236411	0,3162534	0,48	X16	3,60828686	0,05775999	9,89
Rubiaceae	-0,56010456	0,31371711	0,69	X21	3,66426113	0,05341233	10,4
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Olea europaea</i> var. <i>africana</i> A	0,582813921	0,33967207	0,44	F	18,0923783	0,3249037	0,35
<i>Pappea capensis</i> A	0,582813921	0,33967207	0,6	C	-5,4961781	0,17834744	1,61
<i>Allophyllus africanus</i> J	0,582813921	0,33967207	0,6	A	-9,5192302	0,17292535	1,84
<i>Indigofera dendroides</i>	0,582813921	0,33967207	0,48	B	-6,3635959	0,16079222	1,7
<i>Hyparrhenia dissoluta</i>	0,58065047	0,33715497	0,01	X36	4,40467955	0,11517091	3,4
<i>Macrotyloma maraguense</i>	0,578835396	0,33505042	0,01	K	-8,4370581	0,1057788	3,67
<i>Ziziphus mucronata</i> A	0,578215356	0,334333	0,23	X70	-3,0494483	0,10467571	3,96
<i>Haplocoelum gallense</i> J	0,548415449	0,3007595	0,43	I	-5,6393765	0,09730485	4,65
<i>Canthium vulgare</i> A	0,509390495	0,25947868	0,94	X68	-3,4573595	0,08909659	5,09
<i>Biophytum petersianum</i>	0,478255931	0,22872874	1,19	X29	3,81848752	0,08884279	4,85
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Achyranthes aspera</i> var. <i>sicula</i>	0,537389617	0,2887876	0,05	F	15,2722335	0,2315093	0,66
<i>Pappea capensis</i> A	0,518576842	0,26892194	0,84	A	10,3689279	0,20517418	0,8
<i>Indigofera dendroides</i>	0,518576842	0,26892194	0,65	B	6,25909813	0,15555478	1,78
<i>Olea europaea</i> var. <i>africana</i> A	0,518576842	0,26892194	0,54	X33	-5,3492606	0,12875056	2,76
<i>Allophyllus africanus</i> J	0,518576842	0,26892194	0,61	G	-8,9258375	0,12600665	2,66
<i>Setaria barbata</i>	0,48654066	0,23672181	0,54	X1	6,59384762	0,10740686	3,53
<i>Ziziphus mucronata</i> A	0,481913345	0,23224047	0,88	X2	6,02991072	0,0753779	6,39
<i>Acacia polyacantha</i> A	0,458508569	0,21023011	0,75	C	3,55923754	0,0747927	6,53
<i>Phyllanthus fraternus</i>	0,453356598	0,2055322	0,65	X7	5,33248403	0,06782763	7,79
<i>Sporobolus stapfianus</i>	-0,45198988	0,20429485	0,37	X36	-3,0499031	0,05521871	10,5

Tableau 4. Dix premières espèces et dix premiers relevés rangés par ordre décroissant de contributions relative. Axes 1 à 4.

Analyse statistique des données de végétation

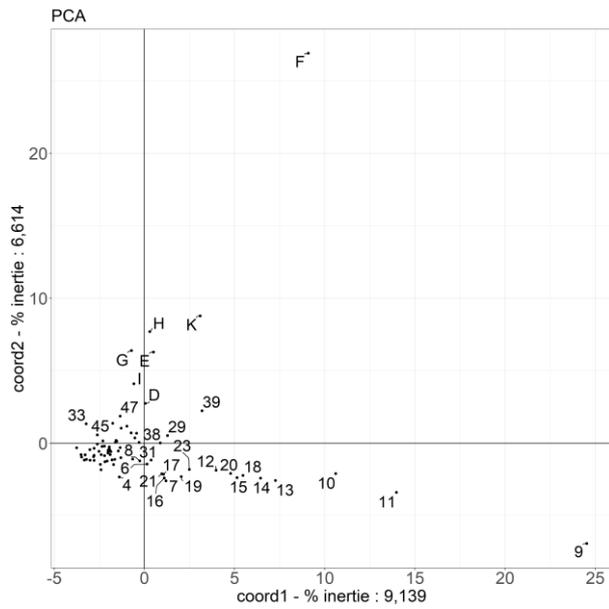


Figure 5. Analyse en composantes principales du fichier Akagera ordinal. Coordonnées des relevés sur les axes 1 et 2.

Analyse statistique des données de végétation

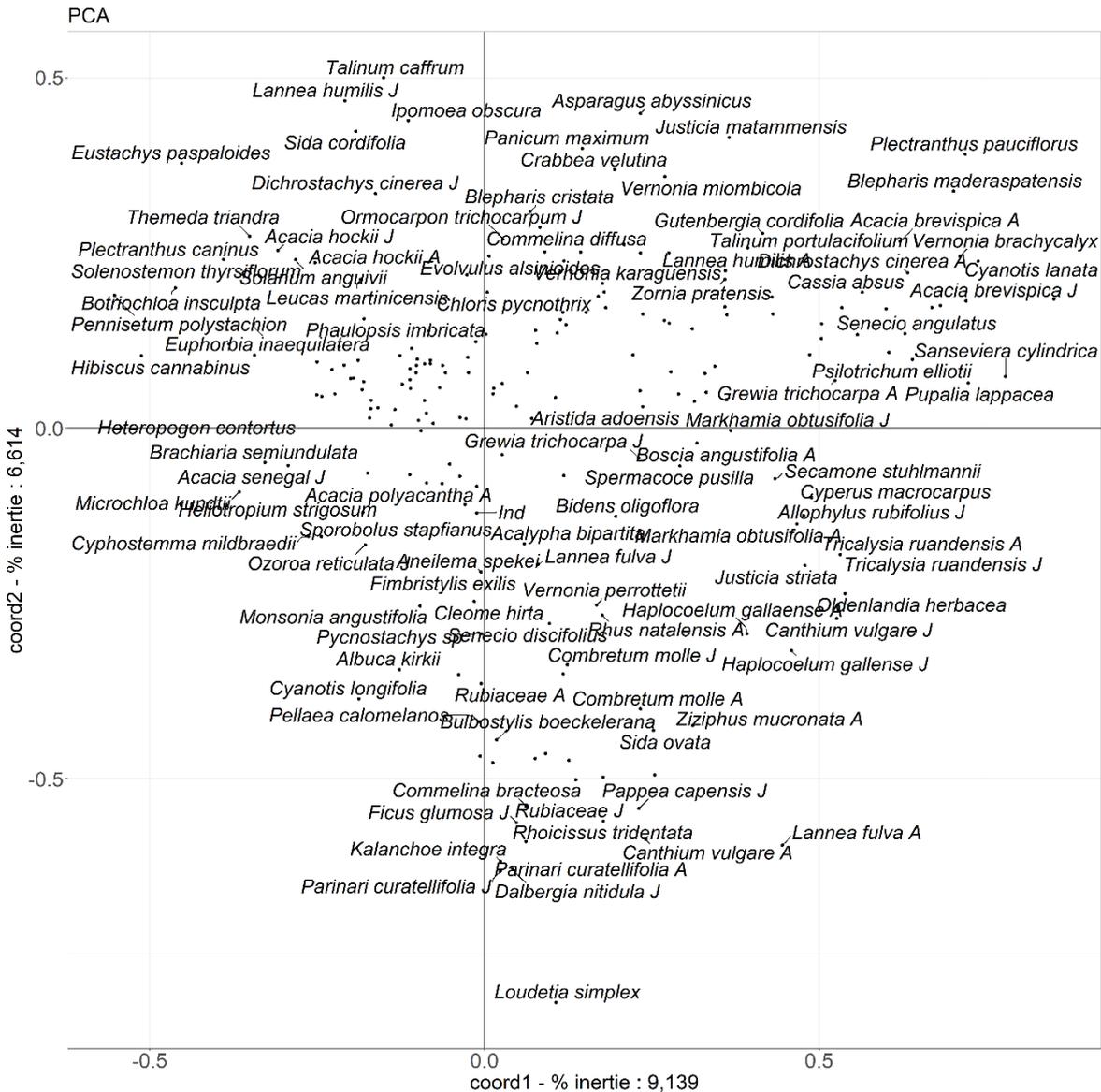


Figure 6. Analyse en composantes principales du fichier Akagera ordinal. Coordonnées des espèces sur les axes 1 et 2.

Cette analyse ne présente aucun avantage par rapport à celle réalisée avec des données brutes.

Analyse des correspondances des données ordinales

Les principaux résultats sont présentés dans le tableau 5 et les figures 7 à 9. Les résultats complets apparaissent dans le fichier Akagera80phyto.xls (Annexes).

Analyse statistique des données de végétation

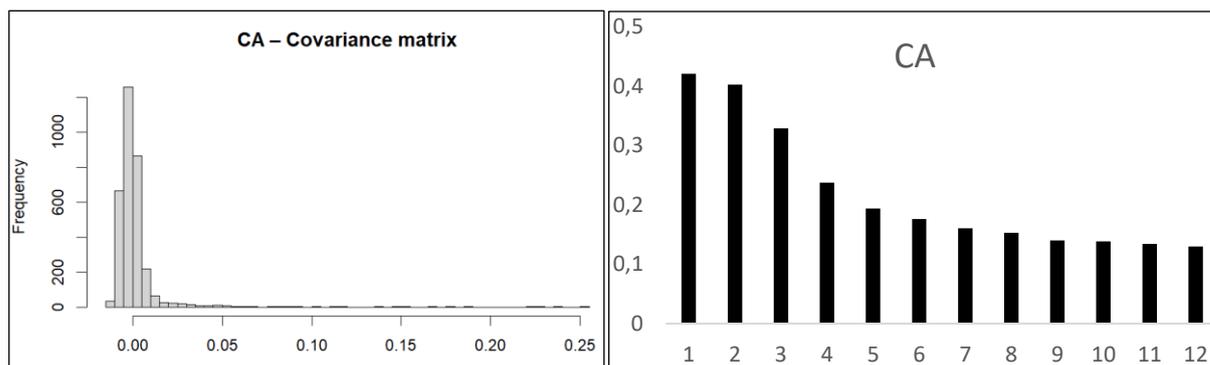


Figure 7. Distribution des coefficients de covariance en 40 classes et les douze premières valeurs propres.

La distribution des coefficients de covariance est fort asymétrique avec une très faible proportion de valeurs élevées et est caractéristique de l'analyse des correspondances de tableaux floristiques. Les douze premières valeurs propres sont significatives.

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Acacia polyacantha</i> A	-5,224924062	30,6225631	13,2	A	-3,33475792	24,94825616	1,2
<i>Acalypha bipartita</i>	-3,247728097	5,91576484	2,8	C	-4,74184334	19,61695093	1,5
<i>Setaria barbata</i>	-1,81964889	5,7775311	1,1	B	-3,7523013	19,30310424	1
<i>Loudetia simplex</i>	2,007963615	4,27139433	6,4	I	2,932765302	9,647975487	2,8
<i>Cissampelos mucronata</i>	-4,635691421	4,01753214	0,2	K	1,601864956	7,51551466	3,4
<i>Dalbergia nitidula</i> A	3,504596204	3,82696988	7,9	H	1,400051297	3,908831435	5,8
<i>Bidens pilosa</i>	-3,993636823	2,98172412	1,5	G	0,971946969	3,237851276	5,4
<i>Vigna macrorhyncha</i>	-3,664540235	2,51055252	2,9	X2	-0,84797273	2,150869574	6,8
<i>Cynodon dactylon</i>	-2,331718164	2,37169708	25,6	F	0,695372396	2,109317718	6,9
<i>Cissus petiolata</i>	-2,450311808	2,24493114	0,7	E	0,838636359	1,709310848	9,1
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P 2
<i>Acacia polyacantha</i> A	-3,213357061	12,1192487	30,3	I	-4,02455289	19,01047653	1,6
<i>Dalbergia nitidula</i> A	-5,064421216	8,36209048	3	K	-2,48131084	18,86884964	1,5
<i>Loudetia simplex</i>	-2,668488592	7,89341327	1	A	-1,99086045	9,303992665	4,4
<i>Acalypha bipartita</i>	-2,467076372	3,57184757	5,8	C	-3,18815363	9,278798333	3
<i>Dalbergia nitidula</i> J	-4,21176247	3,47003787	0,4	G	-1,51054823	8,183085739	3,1
<i>Parinari curatellifolia</i> A	-4,21176247	3,47003787	0,5	B	-2,17931661	6,813159498	4,3
<i>Parinari curatellifolia</i> J	-3,202087729	3,34288684	2,9	H	-1,70766989	6,084754174	4,5
<i>Ficus glumosa</i> A	-3,910986085	2,99212006	26,7	F	-1,10292785	5,552358024	5
<i>Kalanchoe integra</i>	-3,005429652	2,94488512	0,4	E	-0,63080415	1,011903051	16,9
<i>Cleome hirta</i>	-6,343409335	2,62379979	2,6	X9	-0,37680486	0,67583561	26,4
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Acacia brevispica</i> J	1,450030912	3,52279223	1,1	X9	1,549086652	13,97614819	2
<i>Grewia trichocarpa</i> A	1,205310131	3,36131199	39,8	X11	1,378952558	9,70938756	1,7
<i>Acacia brevispica</i> A	0,984211722	2,93679665	4,7	F	1,245907326	8,669274063	2,5
<i>Dalbergia nitidula</i> A	-2,696796784	2,90121331	10,4	I	-2,29733803	7,579412103	2,5
<i>Sansevieria cylindrica</i>	1,8108351	2,87782053	1,3	X10	1,130579999	7,342573555	2,4
<i>Cyanotis lanata</i>	0,906602806	2,29517273	1	X18	1,050440061	4,841929433	5,4
<i>Ziziphus mucronata</i> A	1,987386612	2,20584998	50,5	X20	0,87484964	4,274428804	5,1
<i>Pupalia lappacea</i>	1,603733862	2,05200215	1	X13	0,95149919	4,261692109	8,3
<i>Panicum brevifolium</i>	2,011106612	1,93613061	1,6	X14	0,973056503	4,003733867	6,6
<i>Vernonia brachycalyx</i>	1,731597589	1,91380262	0,7	X15	0,838858739	2,863261775	9,2
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Ziziphus mucronata</i> A	-4,201910497	13,644314	9,3	F	-2,85676322	63,06764315	0
<i>Canthium vulgare</i> A	-4,231451883	7,90676582	5	I	2,311204726	10,61474121	1,5
<i>Dalbergia nitidula</i> A	3,195287686	5,63572777	5,6	X9	0,756315756	4,609875827	7,5
<i>Olea europaea</i> var. <i>africana</i>	-5,858909216	3,78960207	1,3	E	-0,68977506	2,04851741	11,7
<i>Pappea capensis</i> A	-5,858909216	3,78960207	0,7	X11	0,537759824	2,04322762	11,5
<i>Allophyllus africanus</i> J	-5,858909216	3,78960207	1,2	H	-0,69148847	1,68919556	14,5
<i>Indigofera dendroides</i>	-5,858909216	3,78960207	1,7	K	0,539001519	1,507432615	17,1

Analyse statistique des données de végétation

<i>Haplocoelum gallense</i> J	-1,987249119	3,48782498	5,1	D	-0,78600516	1,500491531	17,4
<i>Rubiaceae</i>	-3,638538253	2,92310191	1,5	X18	0,492686642	1,473887502	16,7
<i>Sporobolus marginatus</i>	-3,636781253	2,92027954	1,6	X13	0,43231407	1,217337002	21,9

Tableau 5. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives. Axes 1 à 4.

L'analyse étant duale, les espèces et relevés sont souvent réunis dans les représentations graphiques. Nous les laissons séparés pour une lecture plus aisée.

Les scores du tableau transformé varient entre -1 et 210,722 et les plus hauts scores correspondent aux espèces avec une seule ou seulement quelques présences dans des sites très localisés, c'est-à-dire près du lac ou de la colline rocailleuse. Les quatre premières valeurs propres valent respectivement 8,394, 8,021, 6,556 et 4,738 % de la trace de la matrice.

L'axe 1 (tableau 5 et figures 8 et 9) sépare les espèces comme *Acacia polyacantha* A, *Acalypha bipartita*, ... et les relevés de la galerie forestière et les espèces comme *Loudetia simplex* et *Dalbergia nitidula* et les relevés sur sol peu profond. L'axe 2 les rapproche, ce qui donne une image en V habituelle de ce genre d'analyse. L'axe 3 met en évidence quelques relevés de bosquet xérophile. L'axe 4 sépare en deux petits groupes les relevés de sols superficiels.

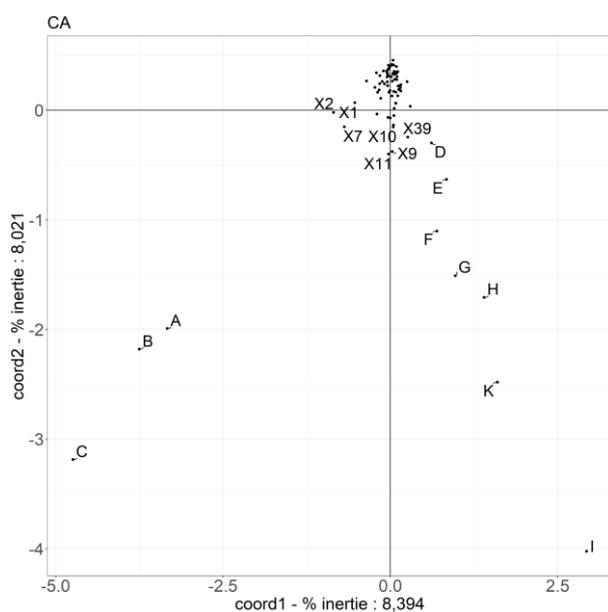


Figure 8. Analyse des correspondances du fichier Akagera ordinal. Coordonnées des relevés sur les axes 1 et 2.

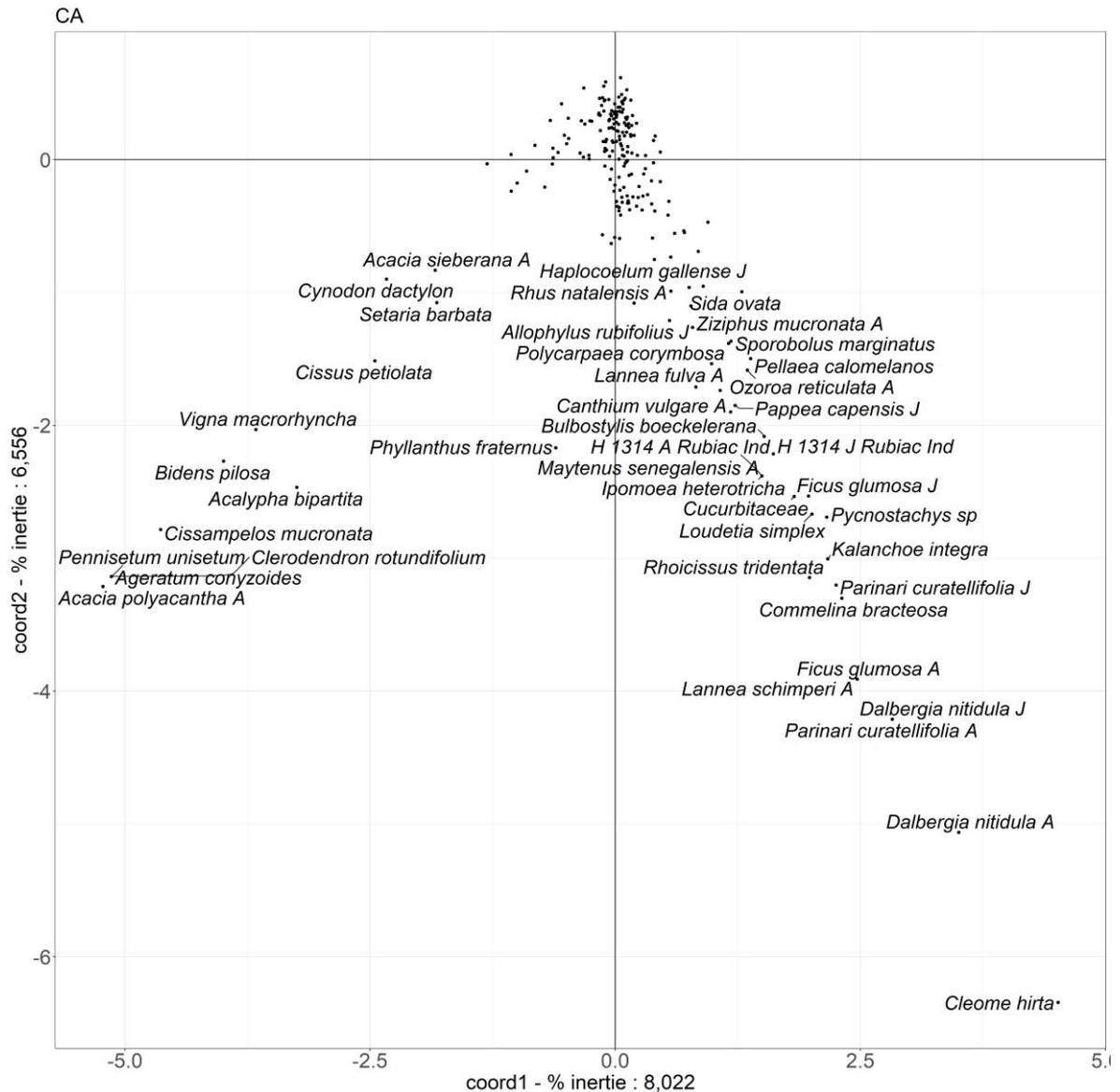


Figure 9. Analyse des correspondances du fichier Akagera ordinal. Coordonnées des espèces sur les axes 1 et 2.

Les résultats de cette analyse sont symptomatiques de l'analyse des correspondances qui privilégie les espèces peu fréquentes et les relevés localisés. Elle se prête mal à une analyse phytocologique.

Analyse des principales coordonnées des données ordinales

L'analyse des principales coordonnées part d'une matrice de distances entre relevés. Nous avons testé deux distances parmi les plus communément utilisées, à savoir la distance euclidienne et la distance corde. Cette analyse donne les valeurs propres de la matrice de distance et les coordonnées des relevés. Les logiciels existants ne produisent aucune aide à l'interprétation comme les contributions relatives.

Avec la distance euclidienne

Les principaux résultats figurent dans le tableau 6 et les figures 10 et 11. Les résultats complets se trouvent dans le fichier Akagera80phytoPCoAeucl.xls.

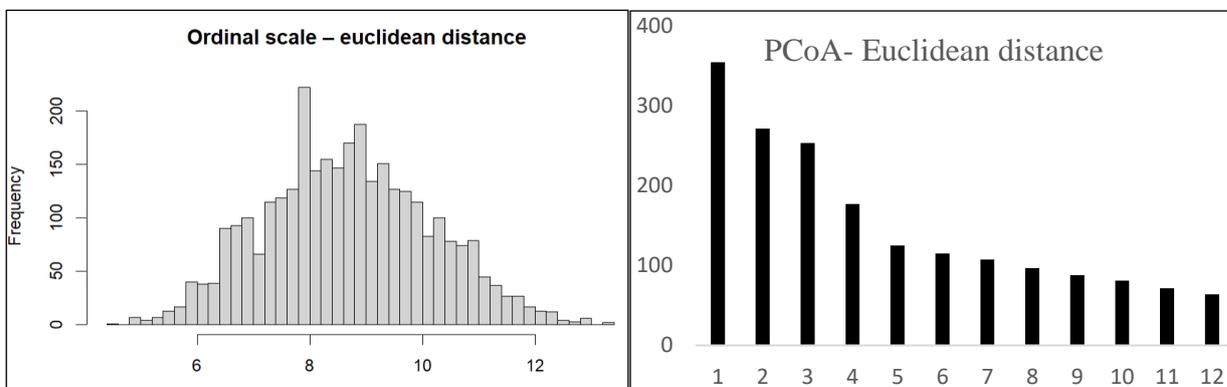


Figure 10. Distribution des distances euclidiennes en 40 classes et les douze premières valeurs propres.

	coord1	coord2	coord3	coord4	coord5	coord6
Cr%	11,57460606	8,866620745	8,2794365	5,784167651	4,089101845	3,761848914
A	-1,78063105	2,918819119	-3,811091664	-5,093437819	0,432814545	-0,933516571
B	-2,062567781	3,523156589	-3,610337866	-5,033908274	0,630946863	-1,452776756
X1	0,802006664	-1,180524779	-0,971970099	-1,87804676	0,648448127	0,08557696
X2	0,549851061	0,521122959	-1,15940105	-2,824141891	-0,031888058	-0,259018851
X3	1,532469015	-1,427122073	0,877773059	-1,009840159	0,605341079	0,563883549
C	-1,685759575	3,861853614	-2,744683887	-3,425247529	0,298243959	-2,210739172
X4	1,306918333	-2,535661797	-1,201394271	-0,095047924	2,096382465	-2,647171088
X5	1,434036756	-2,57543379	-1,80163057	-0,677534676	2,1995954	-1,715345853
X6	-0,089210838	-1,825705361	0,316728262	-0,674837153	1,931217997	0,775069856

Tableau 6. Résultats partiels de l'analyse des principales coordonnées du fichier Akagera avec données ordinales et la distance euclidienne, pour les neuf premiers relevés. Coord : coordonnée sur les axes.

Les scores de la matrice de distance, sont distribués suivant une courbe d'allure gaussienne (figure 10), Les quatre premières valeurs propres valent respectivement 11,574, 8,867, 8,279 et 5,784 % de la trace de la matrice de distance avec une diminution brusque pour la cinquième valeur. Le calcul des distances est évidemment nettement influencé par les fortes abondances des espèces et ce sont les relevés avec des espèces abondantes qui sont mis en évidence. Toutefois, la dispersion des relevés est beaucoup plus grande qu'avec une PCA ou une CA et les résultats peuvent conduire à une interprétation plus riche (figure 11).

Analyse statistique des données de végétation

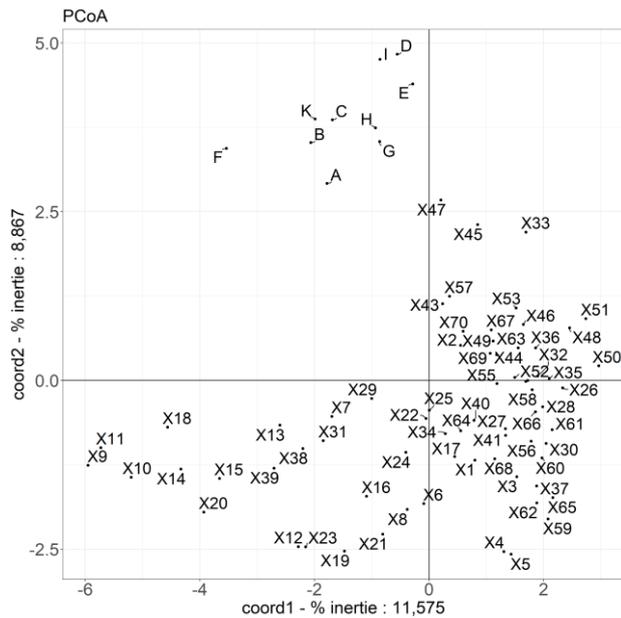


Figure 11. Analyse des principales coordonnées du fichier Akagera avec des données ordinales et la distance euclidienne, coordonnées 1 et 2.

L'axe 1 oppose les relevés de bosquets et environs à des relevés de savane boisée (à *Acacia hockii* ou *Acacia senegal*) ou herbacée.

L'axe 2 isole un groupe mélangé des trois relevés de galerie forestière et de relevés sur sol superficiel.

L'axe 3 sépare les relevés de savane boisée à *Acacia polyacantha* (galerie forestière) et à *Acacia senegal* à d'autres formations moins densément boisées ou herbeuses.

L'axe 4 isole un relevé de savane arbustive aux trois relevés de galerie forestière.

Avec la distance corde

La PCoA est répétée avec la distance corde. Les principaux résultats figurent dans le tableau 7 et les figures 12 et 13. Les résultats complets se trouvent dans les fichiers Akagera80phytoPCoAchord.xls (Annexes).

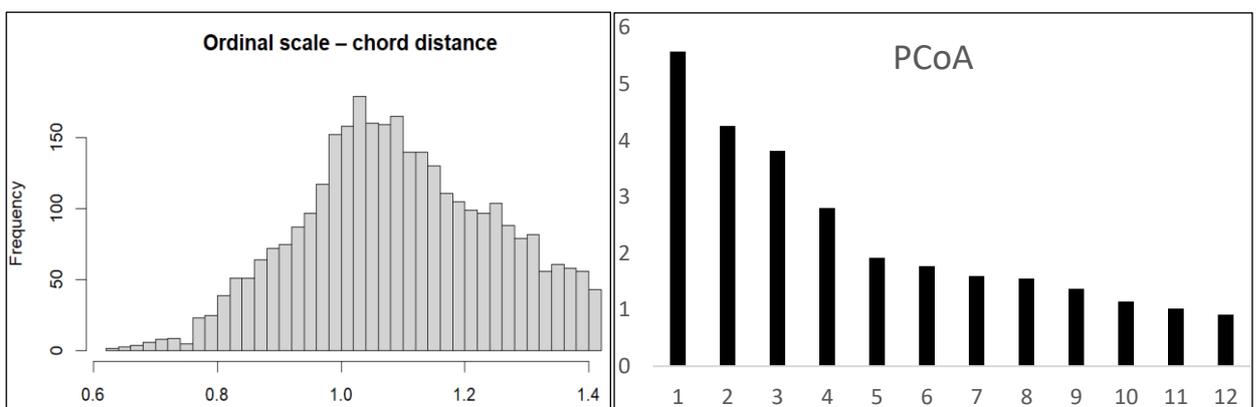


Figure 12. Distribution des distances corde en 40 classes et les douze premières valeurs propres.

Analyse statistique des données de végétation

	coord1	coord2	coord3	coord4	coord5	coord6
Cr%	11,5881316	8,84664043	7,934641	5,82939817	3,99252643	3,69045722
A	0,45686316	-0,1349234	0,35523253	-0,628872	0,00958839	0,13484783
B	0,51584364	-0,0854913	0,37731889	-0,6784195	0,06241839	0,19013745
X1	-0,0807726	-0,2130418	-0,0654457	-0,2372469	0,0019012	-0,1299279
X2	0,01675789	-0,097986	0,0803001	-0,3816422	-0,0463316	-0,0024833
X3	-0,2328349	-0,0455137	-0,1710282	-0,1241955	-0,0507581	-0,136172
C	0,54795815	0,00147261	0,44093107	-0,6745913	0,19965784	0,22304108
X4	-0,1230382	-0,2634471	-0,0706228	-0,0503504	0,41513178	-0,1403256
X5	-0,1052084	-0,3080524	-0,0144355	-0,1372276	0,27766753	-0,1713962

Tableau 7. Résultats partiels de l'analyse des principales coordonnées du fichier Akagera avec données ordinales et la distance corde, pour les huit premiers relevés. Coord : coordonnée sur les axes.

La distribution des distances est un peu asymétrique et les quatre premières valeurs propres sortent à nouveau du lot.

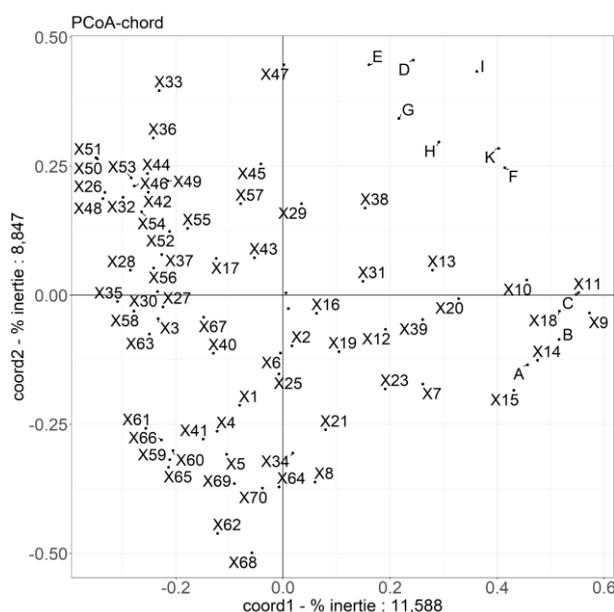


Figure 13. Analyse des principales coordonnées du fichier Akagera avec des données ordinales et la distance corde, coordonnées 1 et 2.

Sur l'axe 1, un groupe de relevés de galerie forestière, de bosquets xérophiles et de relevés sur sol superficiel est opposé à un groupe de relevés de savane boisée, principalement à *Acacia hockii*.

Sur l'axe 2, ce sont des relevés de savane à *Acacia senegal* qui sont opposés à un mélange de plusieurs formations.

Les autres axes sont assez redondants et n'apportent pas d'interprétation claire.

Analyse non-métrique multidimensionnelle des données ordinales

Le calcul de l'analyse non-métrique multidimensionnelle part aussi de la même matrice de distance euclidienne et est faite ici avec six axes. Le logiciel ne donne que les coordonnées de relevés sur les axes. Comme dans PCoA, il n'y a pas d'aide à l'interprétation.

La technique fournit des résultats divergents suivant que le logiciel démarre avec les coordonnées des relevés produites par une autre technique ou par un choix aléatoire des relevés. Les résultats numériques dépendent aussi du

Analyse statistique des données de végétation

nombre d'axes retenu ; les coordonnées des relevés ne sont pas les mêmes sur le premier ou le second axe, par exemple, si on choisit de faire les calculs avec quatre ou six axes. Plusieurs choix ont un caractère arbitraire.

Les résultats sont présentés avec six axes et un emplacement aléatoire des relevés.

Avec la distance euclidienne

Les principaux résultats apparaissent dans le tableau 8 et la figure 14. Les résultats complets apparaissent dans le fichier Akagera80phytonmdseucl.xls (Annexes).

	coord1	coord2	coord3	coord4	coord5	coord6
A	-1,0593515	3,12760937	-4,7860808	-5,4018757	-1,0867468	0,10703587
B	-1,9014348	3,79151737	-4,2093278	-5,3799414	-0,3059905	-1,1228015
X1	1,07924878	-1,090534	-0,9320908	-2,129708	0,93276033	0,24782789
X2	1,62514184	0,74209784	-1,008236	-3,9641807	-0,5175754	0,50017172
X3	1,26822522	-1,3246027	0,59853233	-0,8256296	0,57303183	0,50193654
C	-1,1492332	3,74730939	-2,7688435	-3,0081023	-0,66203	-1,3189415
X4	2,1618686	-2,8867863	-1,0103541	-0,1777476	1,845727	-2,2600538
X5	2,49535727	-2,7957172	-1,6668633	-0,794667	2,35721314	-0,4174664

Tableau 8. Résultats partiels de l'analyse non-métrique multidimensionnelle du fichier Akagera avec données ordinales et la distance euclidienne pour les huit premiers relevés. Coord : coordonnée sur les axes.

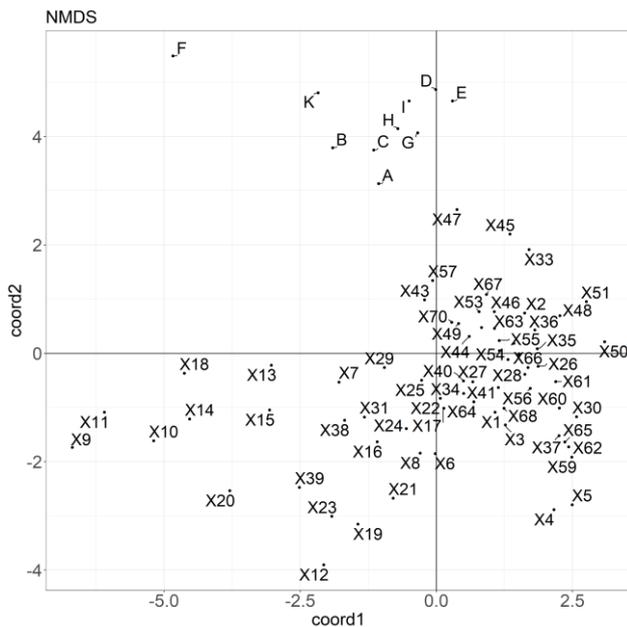


Figure 14. Analyse non-métrique multidimensionnelle du fichier Akagera avec des données ordinales et la distance euclidienne.

L'axe 1 oppose des relevés de bosquet xérophile à des relevés de savane boisée à *Acacia hockii* ou *Acacia senegal*.

L'axe 2 isole les relevés sur sol superficiel.

L'axe 3 isole un relevé de savane à *Acacia hockii* et l'oppose aux relevés de savane à *Acacia polyacantha* et *Acacia senegal*.

L'axe 4 isole un relevé de savane à *Acacia brevispica* proche des bosquets et l'oppose principalement à des relevés de savane à *Acacia polyacantha*.

Avec la distance corde

Les résultats apparaissent dans le tableau 9 et la figure 15. Les résultats complets se trouvent dans le fichier Akagera80phytonmdschord.xls (voir Annexes).

	coord1	coord2	coord3	coord4	coord5	coord6
A	0,50056657	-0,3228975	0,36586493	-0,6770017	-0,0541512	0,09112633
B	0,65689707	-0,2601893	0,38063064	-0,7125544	-0,0790779	0,24022396
X1	-0,1042833	-0,2181797	-0,0945688	-0,3113142	0,02639866	-0,0661905
X2	-0,0835822	-0,0958902	0,07457619	-0,5844833	-0,0216646	0,03071737
X3	-0,1837797	-0,0474034	-0,134469	-0,1039278	-0,0112858	-0,0933054
C	0,75462553	-0,1900815	0,51534941	-0,7669592	0,22949597	0,22437117
X5	-0,1592072	-0,3291422	-0,0794542	-0,1391825	0,16499064	-0,170016

Tableau 9. Résultats partiels de l'analyse non-métrique multidimensionnelle du fichier Akagera avec données ordinales et la distance corde, pour les huit premiers relevés. Coord : coordonnée sur les axes.

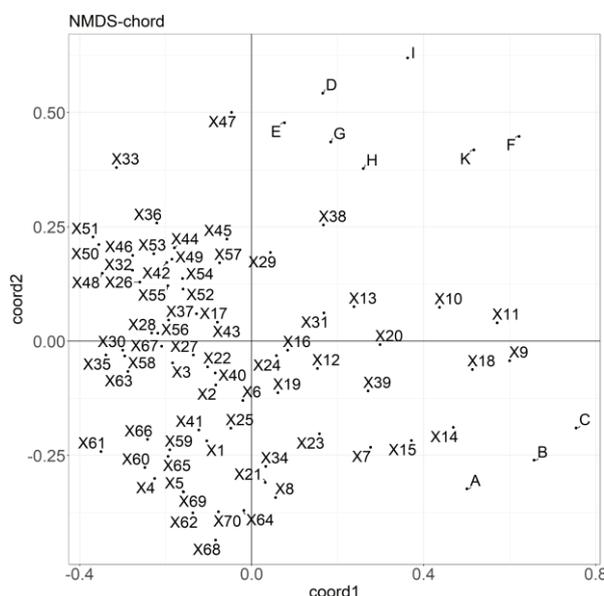


Figure 15. Analyse non-métrique multidimensionnelle du fichier Akagera avec des données ordinales et la distance corde, coordonnées 1 et 2.

Les résultats sont encore une fois assez comparables avec ceux de la distance euclidienne.

Seconde transformation des données

Il apparaît ainsi que les techniques choisies pour analyser un tableau de relevés de végétation présentant à la fois des abondances et de simples présences ne produisent pas de résultats satisfaisants. Une solution est de changer le tableau originel en un tableau disjonctif simple, ou en un tableau disjonctif complet (voir plus haut), ce qui permet alors l'utilisation d'analyses multivariées comme CA ou NSCA. Les analyses des correspondances avec des tableaux disjonctifs simples sont aussi appelées analyses des correspondances multiples. Les tableaux disjonctifs simples sont aussi utilisables avec PCoA. Dans le tableau disjonctif simple, 86,32 % des cases sont vides et dans le tableau disjonctif complet, la proportion est réduite à 56,81 %. Un tableau disjonctif complet de données de végétation fait la part belle aux absences, qui sont beaucoup plus nombreuses que les présences, ce n'est pas adapté à une étude de végétation. Seuls, les tableaux disjonctifs simples seront analysés.

Analyse des correspondances multiples

La CA du tableau disjonctif simple montre des résultats comparables avec ceux du tableau ordinal. Dans le tableau transformé, les scores varient entre -1 et 379,67. La distribution des covariances est toujours fortement asymétrique. Les quatre premières valeurs propres valent 6,125, 5,87, 4,513 et 4,064 % de la trace de la matrice et les suivantes décroissent régulièrement et lentement. Les premiers axes sont encore influencés par des espèces peu fréquentes et localisées (comme *Acacia polyacantha* et les espèces de bosquet) et aucun avantage n'est trouvé.

Analyse non symétrique des correspondances multiples

Avec NSCA du tableau disjonctif simple, on entre dans une autre catégorie de résultats. La distribution des covariances est faiblement asymétrique et régulière. Seules les cinq premières valeurs propres sont significatives, ce qui indique clairement le caractère synthétique de cette analyse (table 10 et figure 16).

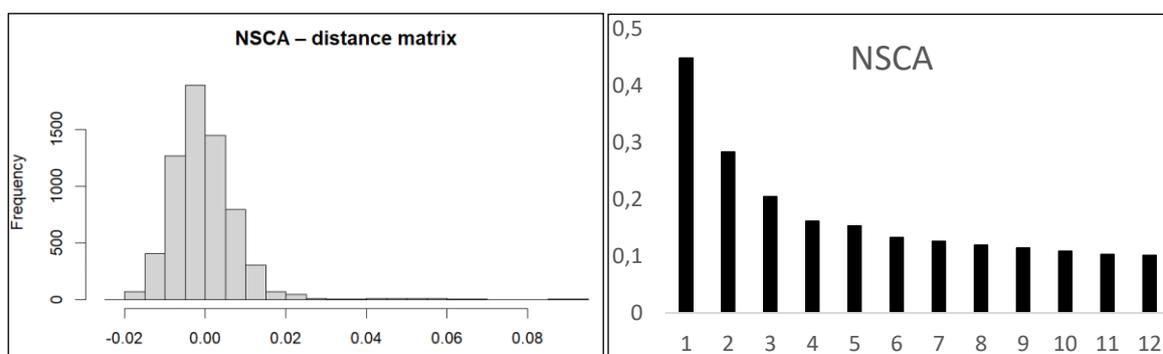


Figure 16. Distribution des covariances en classes et les douze premières valeurs propres.

Les principaux résultats apparaissent dans le tableau 10 et les figures 17 et 18. Les résultats complets se trouvent dans le fichier Akagera80phytodisj12345nsca.xls (Annexes).

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Hibiscus cannabinus</i>	0,21161587	4,47812784	7,2	X9	-1,7365976	5,81141295	2,8
<i>Heteropogon contortus</i>	0,21086076	4,44622592	5,4	X15	-2,0228431	5,73462547	3,2
<i>Pennisetum polystachion</i>	0,19760028	3,90458723	8,6	X14	-1,997668	5,35974194	3,8
<i>Albucca kirkii</i>	0,19352699	3,74526959	12,9	X11	-1,7500253	5,18628121	3,5
<i>Botriochloa insculpta</i>	0,18796352	3,53302859	9,1	X18	-1,7578419	4,60118023	4,5
<i>Panicum atrosanguineum</i>	0,16008675	2,5627766	14,5	X10	-1,3730275	3,57776049	8,5
<i>Indigofera colutea</i>	0,15630264	2,44305167	15,4	X13	-1,4058361	3,23144544	11,3
<i>Microchloa kundtii</i>	0,15456958	2,3891755	16,6	X51	1,46390263	2,94077408	14
<i>Plectranthus pauciflorus</i>	-0,1474352	2,17371294	9	X12	-1,2977361	2,70442469	14,4
<i>Solenostemon thyrsoiflorum</i>	0,14650623	2,14640769	17,6	X63	1,68436415	2,56787036	17,5
Species	coord2	Cr%2	P2	Relevés	Coord2	Cr%2	P2
<i>Achyranthes aspera</i> var <i>sicula</i>	-0,3050809	9,30743511	0,5	X62	-2,4919716	6,16459475	2,4
<i>Commelina benghalensis</i>	-0,223288	4,98575161	5,7	X70	-2,9897854	5,74172177	3,2
<i>Spermacoce pusilla</i>	0,2129926	4,53658489	7,3	X59	-2,0336598	5,07160409	4,2
<i>Hyparrhenia dissoluta</i>	0,18600733	3,45987262	13,2	E	2,05881846	4,08403517	7,3
<i>Justicia ruwenzoriensis</i>	-0,1773634	3,14577815	0,2	X1	-1,665733	3,72655941	8,5
<i>Panicum maximum</i> 2	-0,1727271	2,98346677	7,2	X47	1,62131944	3,6839849	8,7
<i>Panicum maximum</i>	0,17232925	2,9697369	14,6	X64	-1,7494567	3,66378247	8,4
<i>Pennisetum polystachion</i>	-0,1672342	2,79672908	13,2	X34	-1,6443993	3,63171554	9,5
<i>Solenostemon thyrsoiflorum</i>	-0,1599391	2,55805083	16,4	A	-2,1283562	3,30649647	10,9
<i>Plectranthus caninus</i>	-0,1599391	2,55805083	15,9	X5	-1,6356383	3,20256394	10,3

Analyse statistique des données de végétation

Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Lansea humilis</i> J	-0,1929802	3,72413474	8,1	K	3,40772896	14,2402891	0,1
<i>Cyanotis longifolia</i>	0,19081255	3,6409431	0,1	I	5,40454751	11,0866785	0,3
<i>Ipomoea obscura</i>	-0,1904212	3,6260222	6,4	H	3,16686959	8,19894199	1
<i>Bidens oligoflora</i>	0,18047158	3,25699928	10,1	G	2,34733243	8,04375114	0,9
<i>Blepharis Cr%istata</i>	-0,1772897	3,1431621	3,6	X32	-1,6561589	3,60376073	9,2
<i>Asparagus abyssinicus</i>	-0,1724299	2,97320601	8,2	X8	-1,5478956	2,86818149	12,6
<i>Vernonia miombicola</i>	-0,1697874	2,88277779	14,3	X4	-1,5307827	2,66827837	13,8
<i>Digitaria longiflora</i>	-0,1572147	2,47164757	18,5	C	3,09475618	2,51671949	19,1
<i>Acacia hockii</i> J	-0,1569337	2,4628185	11,9	A	1,72316586	2,16737268	18,9
<i>Kalanchoe integra</i>	0,15181364	2,30473822	0	F	1,11079922	2,05345604	19,5
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Justicia matammensis</i>	-0,2236196	5,00057325	4,6	B	-6,189709	13,423357	0,1
<i>Crabbea velutina</i>	0,21594148	4,66307233	5,3	D	-5,3578267	13,4102454	0,1
<i>Sporobolus stapfianus</i>	0,19801573	3,92102291	7,4	A	-3,6817695	9,89447212	0,2
<i>Digitaria longiflora</i>	-0,1975248	3,90160423	9,1	X45	-1,9619388	4,94496274	4
<i>Markhamia obtusifolia</i> J	0,1940661	3,76616507	8,2	C	-4,2068516	4,65046442	5,5
<i>Brachiaria semiundulata</i>	-0,1821637	3,31836235	0	X61	2,33401317	4,13541766	6,1
<i>Acacia polyacantha</i> A 5	-0,1804481	3,2561518	5,3	X2	-1,7710208	4,12096805	5,7
<i>Heliotropium strigosum</i>	-0,1788344	3,19817517	0	I	2,49679426	2,36618279	18,5
<i>Commiphora africana</i> J	0,16613102	2,75995148	15,6	X24	1,23145585	2,12529078	20,6
<i>Brachiaria brizantha</i>	-0,1579148	2,49370855	0,2	X55	1,32859764	2,06151439	21,1

Tableau 10. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives. Axes 1 à 4.

L'axe 1 oppose les relevés et espèces des sites fermés à ceux de savane ouverte. En considérant les contributions relatives, les résultats opposent les cinq relevés de bosquet (la plupart sont significatives) avec *Plectranthus pauciflorus* (P = 9) et *Blepharis maderaspatensis* (P = 1,2) à huit relevés avec des espèces de savane ouverte (*Albuka kirkii*, *Botriochloa insculpta*, *Heteropogon contortus*, *Hibiscus cannabinus*, *Pennisetum polystachion*), mais sans aucune contribution relative significative

Sur le second axe, deux groupes d'espèces et relevés sont clairement opposés : le premier avec plusieurs relevés et espèces (*Achyranthes aspera* var. *sicula*, *Commelina benghalensis*, *Justicia ruwenzoriensis* et *Commicarpus grandiflorus* avec les contributions relatives significatives) localisés au pied de la colline (savane boisée) et au bord ou à proximité du lac (savane boisée) et le second avec des espèces fréquentes dans les autres sites.

Le troisième axe est dominé par les quatre relevés du sommet du transect, sur la colline rocheuse, avec une contribution significative ; plusieurs espèces sont hautement caractéristiques : *Bulbostylis boeckelerana*, *Cleome hirta*, *Combretum molle* adulte, *Cyanotis longifolia*, *Dalbergia nitidula* adulte, *Kalanchoe integra*, *Loudetia simplex* (abondance 3) et *Parinari curatellifolia* jeune et adulte.

Avec l'axe 4, on note *Crabbea velutina*, une espèce très commune, et plusieurs autres espèces absentes des trois relevés du bord du lac ou d'un site rocheux.

Avec l'axe 5, les espèces *Solenostemon thyrsoflorum* et *Plectranthus caninus*, *Acacia hockii* adulte (les deux premières espèces avec une contribution relative significative) sont absentes des trois relevés avec de grands *Acacia polyacantha*, des relevés de la zone des bosquets et de la colline rocheuse; elles sont opposées à *Cassia falcinella*, *Commelina erecta*, *Indigofera colutea*, *Justicia matammensis*, *Talinum caffrum*, *Vernonia miombicola*, *Acacia senegal* jeune, certaines avec une contribution relative significative.

Analyse statistique des données de végétation

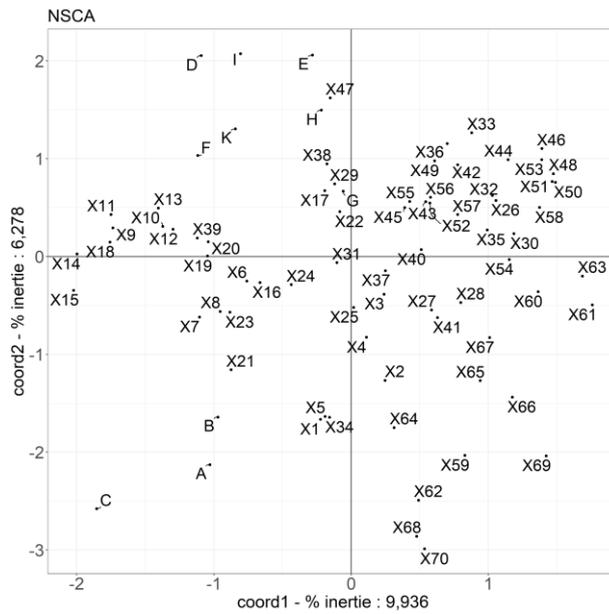


Figure 17. Analyse non symétrique des correspondances du fichier Akagera avec le tableau disjonctif simple. Coordonnées des relevés sur les axes 1 et 2.

Analyse statistique des données de végétation

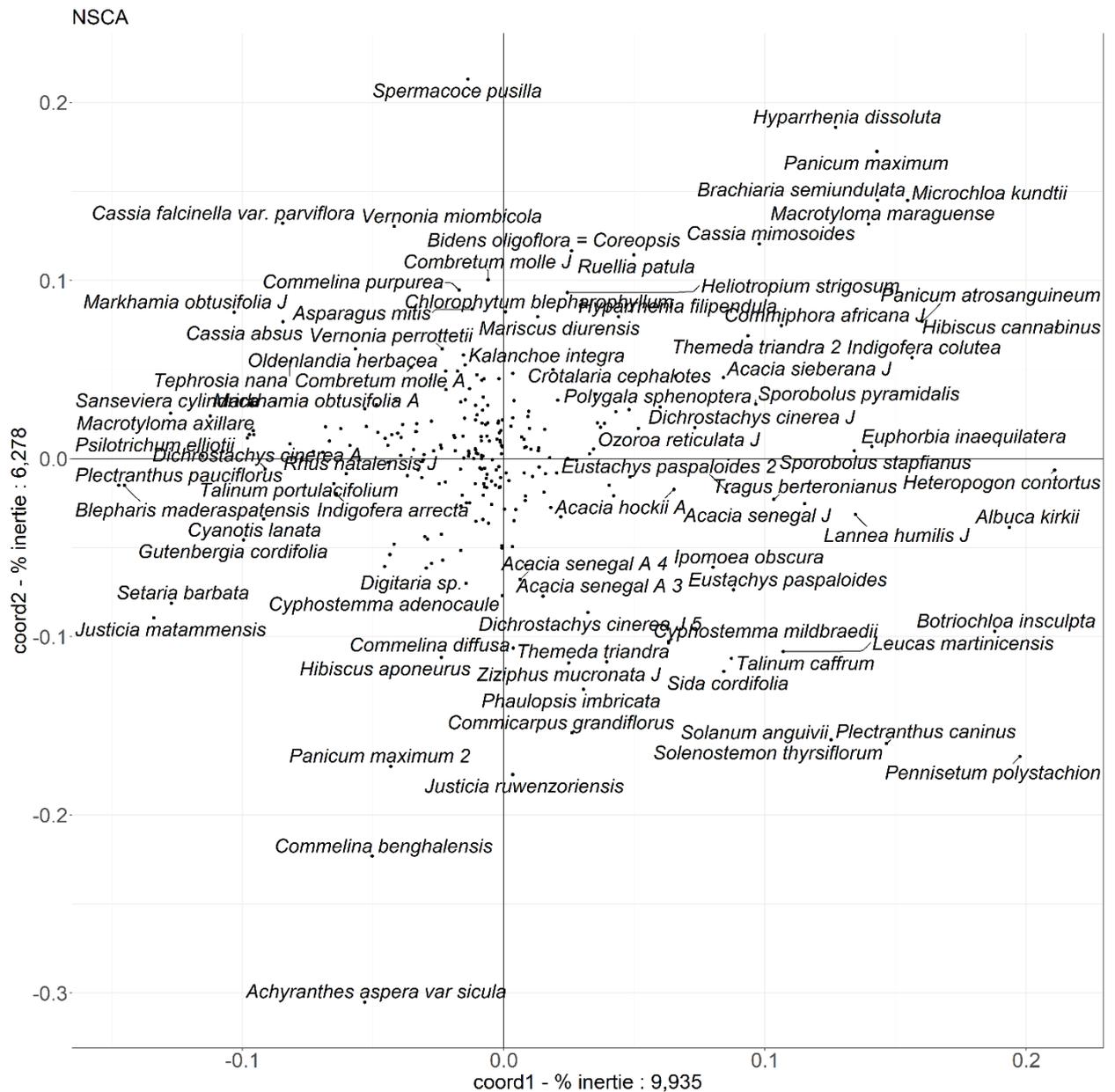


Figure 18. Analyse non symétrique des correspondances du fichier Akagera avec le tableau disjonctif simple. Coordonnées des espèces sur les axes 1 et 2, 2 et 3, 3 et 4 et 4 et 5.

Les deux premiers axes apportent un maximum d'informations sur les principaux groupements végétaux et les axes suivants mettent en évidence les relevés de groupements moins importants en nombre de relevés.

Analyse des principales coordonnées du tableau disjonctif simple

L'analyse des principales coordonnées est aussi testée avec un tableau disjonctif simple.

Avec la distance euclidienne

Les principaux résultats apparaissent dans les figures 19 et 20. Les résultats complets se trouvent dans le fichier Akagera80disj12345PCoAeucl.xls (Annexes)

Analyse statistique des données de végétation

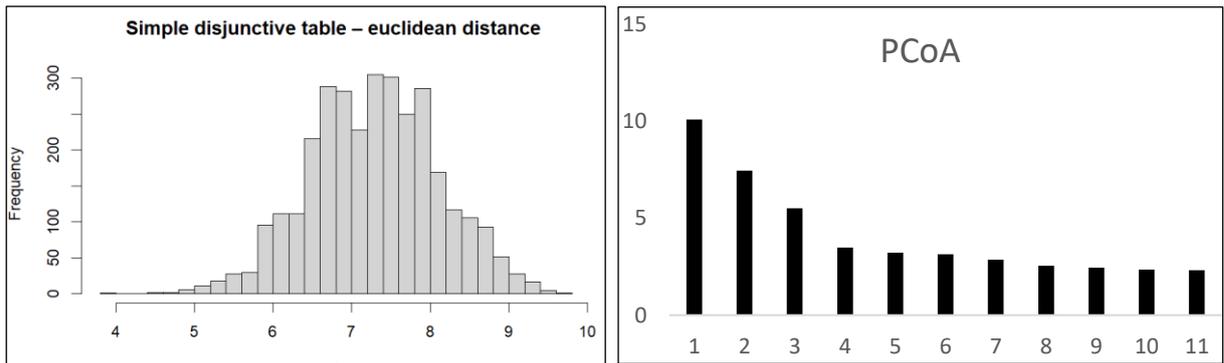


Figure 19. Distribution des distances euclidiennes en 40 classes et les douze premières valeurs propres.

PCoA est calculée avec une matrice de distances euclidiennes et la distribution des distances a une forme proche d'une distribution gaussienne mais avec une faible asymétrie (figure 19). Les quatre premières valeurs propres valent respectivement 10,080, 7,444, 5,500 et 3,487 % de la trace de la matrice et la décroissance continue de manière lente et régulière. Les principaux résultats apparaissent dans la figure 20.

L'axe 1 oppose les relevés de bosquets et environs à ceux de savane à *Acacia hockii* et *Themeda triandra*.

L'axe 2 oppose des relevés de savane plus ou moins boisée aux relevés de galerie forestière et ceux de végétation herbacée ou ligneuse sur sol superficiel.

L'axe 3 met principalement en évidence les relevés sur sol superficiel.

L'axe 4 n'engendre aucune interprétation claire.

Seuls les trois premiers axes fournissent une interprétation claire de la végétation.

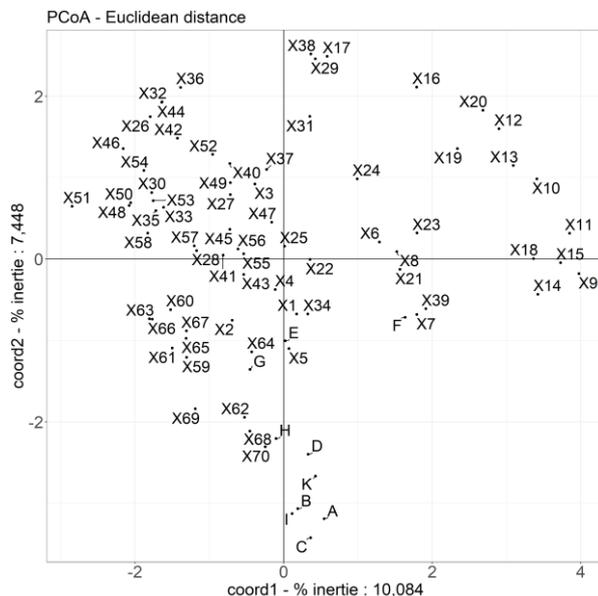


Figure 20. Analyse des principales coordonnées du fichier Akagera avec le tableau disjonctif simple et la distance euclidienne. Coordonnées des relevés sur les axes 1 et 2.

Avec la distance corde

L'analyse est répétée avec la distance corde (Figures 21 et 22). Les résultats complets se trouvent dans le fichier Akagera80disj12345PCoAchord.xls (Annexes).

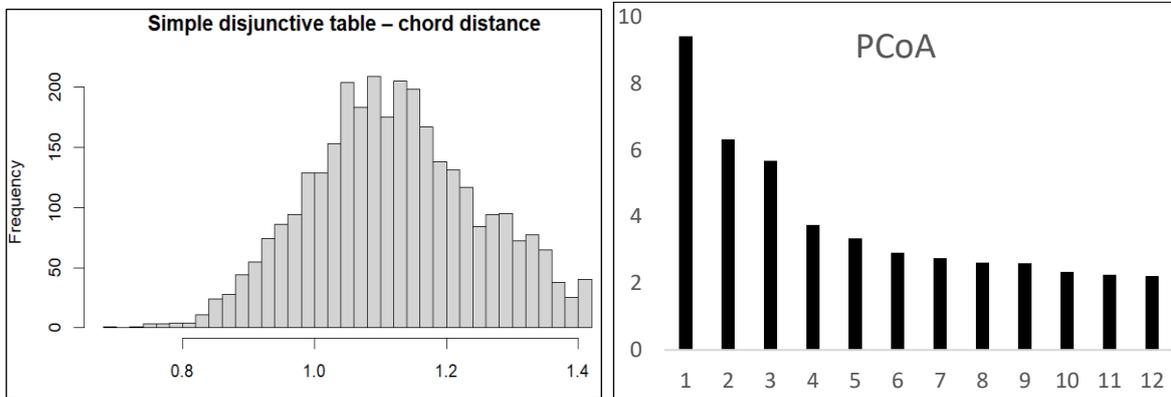


Figure 21. Distribution des distances cordes en 40 classes et les douze premières valeurs propres.

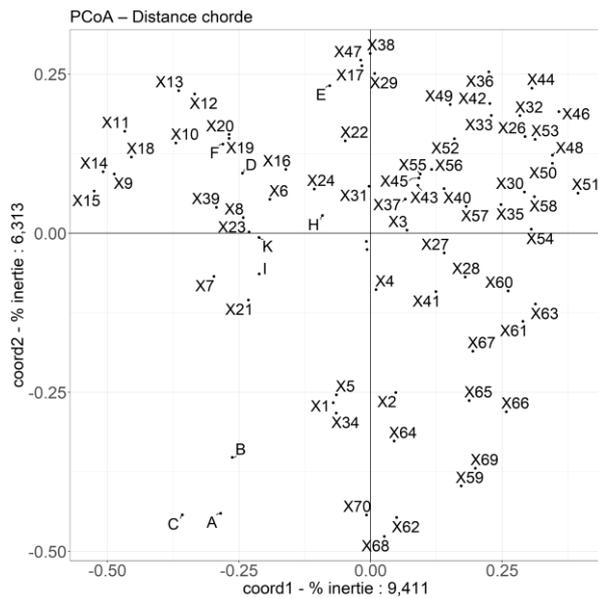


Figure 22. Analyse des principales coordonnées du fichier Akagera avec le tableau disjonctif simple et la distance corde. Coordonnées des relevés sur les axes 1 et 2.

L'axe 1 oppose des relevés de savane plus ou moins boisée à un ensemble de relevés sur sol superficiel ou de galerie forestière.

L'axe 2 isole les relevés de savane boisée à *Acacia senegal* et ceux de la galerie forestière à *Acacia polyacantha*.

Avec l'axe 3, ce sont les relevés sur sol superficiel qui sont isolés.

L'axe 4 isole les relevés sur sol superficiel avec aussi ceux de galerie forestière.

Mais comme déjà vu précédemment, il n'y a aucune aide à l'interprétation dans les logiciels disponibles, comme les \cos^2 sur les axes. En plus, PCoA est calculée seulement pour les relevés.

Une troisième voie pratique avec la transformation en tableau disjonctif 113 et en logarithme d'abondances

La transformation en un tableau disjonctif avec une ligne possible pour chaque degré d'abondance semble la mieux adaptée et la plus rationnelle aux données des tableaux phytosociologiques. Toutefois, elle rend l'interprétation plus lourde et est susceptible de décourager plus d'un utilisateur peu habitué à ce type d'analyse.

Les phytosociologiques privilégient les données de présence mais l'abondance, expression de la biomasse, permet parfois de discriminer certains sites. Nous avons donc testé la solution suivante : pour chaque ligne d'un tableau, il y a trois lignes possibles :

- Une première dans laquelle toutes les abondances (+, i, 1, 2, 3, 4, 5) deviennent égales à 1, le nom spécifique apparaît seul :
- Une deuxième dans laquelle les abondances supérieures à 1 deviennent égales en 1, le nom spécifique est suivi de >1 ;
- Une troisième dans laquelle les abondances supérieures à 3 deviennent égales en 1, le nom spécifique est suivi de >3.

Ces fichiers se terminent par "113".

Une autre solution encore plus simple est proposée en tenant compte des particularités de NSCA. Dans un tableau phytosociologique avec des abondances transformées en scores 0,1 ou 0,5 pour les abondances r ou +, et en scores 1, 2, 3, 4, 5 pour les autres abondances, NSCA est dominée par les espèces abondantes. De manière à diminuer cette prépondérance, il suffit de transformer les abondances en logarithmes. Il n'y a plus alors qu'une ligne par espèce.

Transformation 113

La transformation 113 est maintenant appliquée avec l'analyse non symétrique des correspondances. Les principaux résultats apparaissent dans le tableau 11 et les figures 23 à 25. Les résultats complets se trouvent dans le fichier Akagera80disj113nsca.xlsx

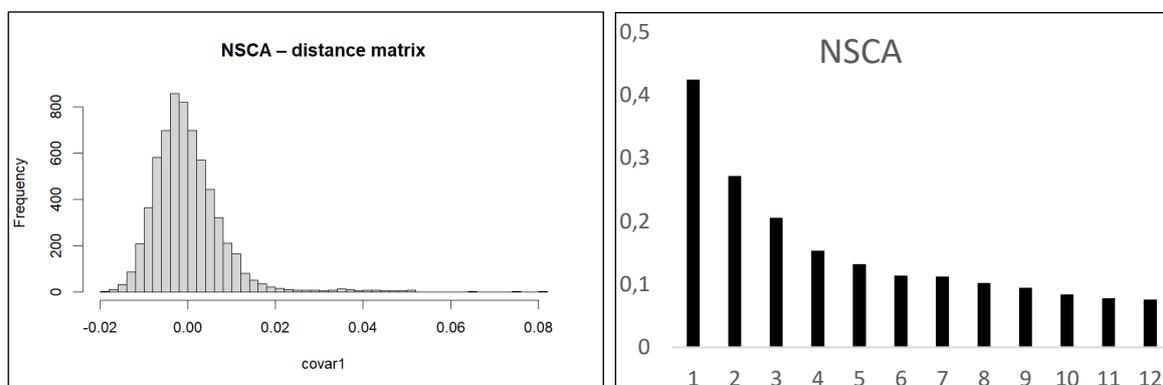


Figure 23. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Les huit premières valeurs propres sont significatives.

Analyse statistique des données de végétation

Species	coord1	Cr%1	P 1	Relevés	coord1	Cr%1	P1
<i>Pennisetum polystachion</i>	-0,205654	4,229356757	7,4	X9	1,931746786	7,279287087	0,6
<i>Hibiscus cannabinus</i>	-0,19457641	3,785997997	10	X11	1,896731568	6,043094547	2,7
<i>Heteropogon contortus</i>	-0,19391402	3,760264825	8,9	X18	1,90386914	5,303027796	3,3
<i>Botriochloa insculpta</i>	-0,17567716	3,086246343	13	X14	1,948410712	5,245503101	3,1
<i>Eustachys paspaloides</i>	-0,17247503	2,974763604	0,8	X15	1,917336658	5,07952243	4,5
<i>Cyanotis lanata</i>	0,151068063	2,282155958	15	X10	1,500473981	4,330831045	6,4
<i>Panicum atrosanguineum</i>	-0,14585373	2,127330928	19	F	1,30081292	2,934037174	14
<i>Themeda triandra</i>	-0,14417659	2,078688907	24	X13	1,314346486	2,714586501	14
<i>Microchloa kundtii</i>	-0,14338215	2,055844003	20	X61	-1,76454138	2,446344688	17
<i>Setaria barbata</i>	0,142631628	2,034378125	12	X20	1,136825992	2,380964153	19
Species	coord2	Cr%2	P 2	Relevés	coord2	Cr%2	P2
<i>Achyranthes aspera</i> var <i>sicula</i>	-0,29649337	8,790831752	0,7	X68	-2,45938489	5,899449087	2,5
<i>Panicum maximum</i> >1	-0,23178046	5,372217916	4,9	X47	1,900134963	4,890968408	4,9
<i>Commelina benghalensis</i>	-0,2085065	4,347496109	9,6	E	2,140714012	4,593830663	3,9
<i>Spermacoce pusilla</i>	0,200329963	4,013209412	8	X62	-1,95519421	4,142817031	6
<i>Hyparrhenia dissoluta</i>	0,171534636	2,94241312	16	I	3,020812315	3,955700694	8,4
<i>Microchloa kundtii</i>	0,16503794	2,723752157	14	X70	-2,42736541	3,831223722	6,7
<i>Acacia senegal</i> A	-0,15652509	2,450010399	4,9	A	-2,05688627	3,553351805	10
<i>Justicia ruwenzoriensis</i>	-0,14658204	2,148629557	1,6	D	2,62206297	3,53912409	9,9
<i>Acacia senegal</i> A >1	-0,14350156	2,059269714	1,1	X33	1,664920475	3,379523393	10
<i>Brachiaria semiundulata</i>	0,139663016	1,950575798	15	X64	-1,57330265	2,950755184	14
Species	coord3	Cr% 3	P3	Relevés	coord3	Cr%3	P3
<i>Albucca kirkii</i>	-0,20082014	4,032872862	6,3	I	-4,91201218	10,45911184	0,1
<i>Acacia hockii</i> J	0,192182098	3,693395863	10	K	-2,7815723	9,223363737	0,4
<i>Ormocarpon trichocarpum</i> J	0,18186171	3,307368163	13	G	-2,13411921	6,663264697	1,2
<i>Asparagus abyssinicus</i>	0,176820801	3,126559559	8	H	-2,85677371	6,633288565	1,8
<i>Lannea humilis</i> J	0,171565833	2,94348351	12	C	-3,48272441	3,614821521	9,1
<i>Loudetia simplex</i>	-0,17082224	2,918023813	0	A	-2,03928024	3,492781921	8,6
<i>Blepharis cristata</i>	0,15622873	2,440741594	4,9	X69	-2,00399413	3,046539937	14
<i>Loudetia simplex</i> >1	-0,15509669	2,405498437	0	X32	1,468564246	2,687816945	15
<i>Ipomoea obscura</i>	0,153544795	2,357600413	13	B	-2,46888443	2,642271618	14
<i>Vernonia miombicola</i>	0,151890874	2,307083759	20	F	-1,18295331	2,42644883	16
Species	coord4	Cr%4	4	Relevés	coord4	Cr%4	P4
<i>Acacia polyacantha</i> A	0,239734653	5,747270377	0	B	6,092445228	16,09011709	0
<i>Setaria barbata</i>	0,222355623	4,944202309	0,8	C	6,542512286	12,75668212	0,1
<i>Acacia polyacantha</i> A >3	0,191574746	3,670088313	0	A	3,518284528	10,39631825	0
<i>Brachiaria brizantha</i>	0,18452798	3,405057557	0	X2	2,311738273	6,805047157	1,9
<i>Crabbea velutina</i>	-0,18290449	3,34540529	10	X70	-2,15794219	3,027936819	11
<i>Cissampelos mucronata</i>	0,166762532	2,780974212	0	X61	-1,87372444	2,758451765	15
<i>Cissus petiolata</i>	0,164189639	2,695823747	0	X1	1,426162794	2,755270001	15
<i>Sporobolus stapfianus</i>	-0,15317559	2,346276275	17	X69	-1,69170935	2,171028299	19
<i>Albucca kirkii</i>	-0,15247468	2,324852832	20	X68	-1,48125323	2,140016273	19
<i>Acacia hockii</i> J	0,151291024	2,288897403	19	X51	1,141877958	1,730977466	27

Tableau 11. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives Axes 1 à 4.

Le premier axe (table 11 et figures 24 et 25) oppose clairement des espèces et relevés de savane herbeuse (coordonnées positives) aux espèces et relevés de bosquets xérophiles ; on a donc, d'un côté, les formations herbacées les plus ouvertes opposées aux formations de bosquet les plus densément ligneuses.

L'axe 2 oppose des espèces et relevés de sols superficiels aux espèces et relevés de savane à *Acacia senegal* et *Acacia polyacantha*, c'est-à-dire les savanes boisées les plus denses.

Les axes suivants se rapportent à des groupements divers représentés par un plus petit nombre de relevés.

Analyse statistique des données de végétation

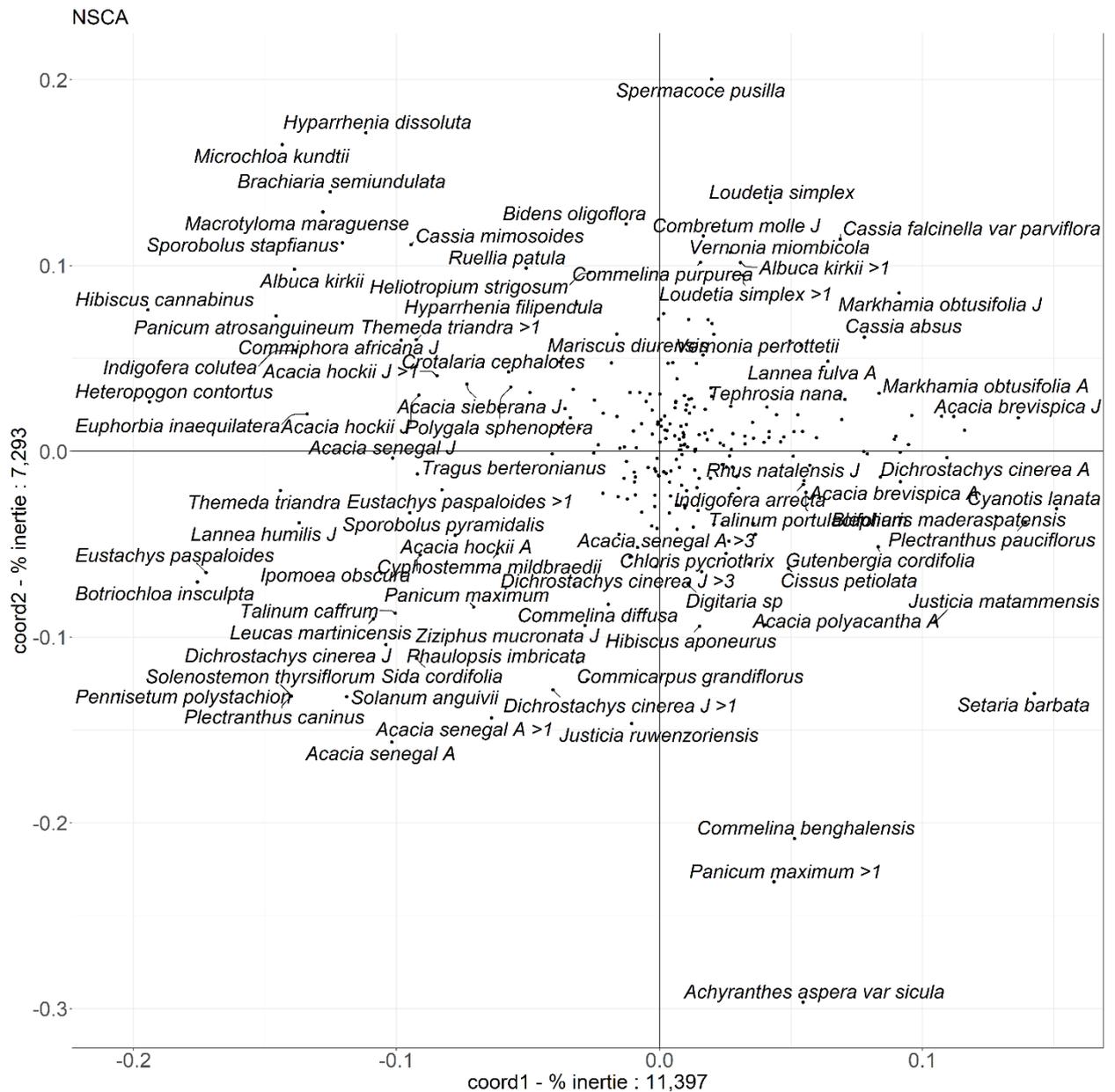


Figure 25. Analyse non symétrique des correspondances du fichier Akagera disjonctif 113. Coordonnées des espèces sur les axes 1 et 2.

Cette technique offre une bonne base conduisant à une interprétation approfondie de la végétation.

Transformation logarithmique

L'analyse non symétrique avec des données d'abondance de type 12345 n'est pas applicable car trop fortement influencée par les espèces abondantes qui écrasent toutes les autres données. La transformation logarithmique est testée comme une forme intermédiaire entre simples données de présence et données d'abondance (Figures 26 et 27, tableau 12). Les résultats complets se trouvent dans le fichier Akagera80lognsca.xlsx.

Analyse statistique des données de végétation

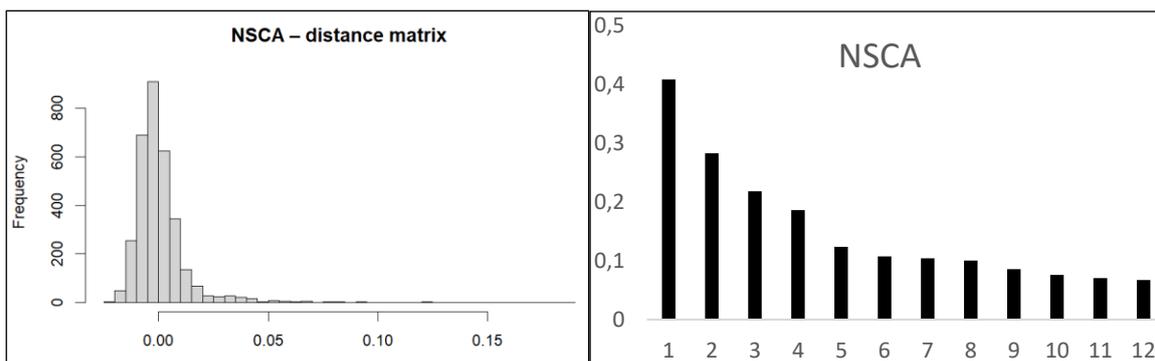


Figure 26. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Les huit premières valeurs propres sont significatives.

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Eustachys paspaloides</i>	-0,21339249	4,553635498	0,5	X9	1,952533953	7,367856611	1,8
<i>Pennisetum polystachion</i>	-0,20511014	4,207016862	1	X11	1,909471698	6,122380793	2,2
<i>Dichrostachys cinerea J</i>	-0,19309178	3,728443555	64,6	X18	1,8603817	5,086075828	4,6
<i>Themeda triandra</i>	-0,18710563	3,500851597	25,1	X14	1,886364105	4,893273176	4,4
<i>Cyanotis lanata</i>	0,177081351	3,1357805	13	X10	1,531669499	4,452027512	6,3
<i>Setaria barbata</i>	0,176206754	3,104882028	6,3	X15	1,758069429	4,250308548	6,1
<i>Heteropogon contortus</i>	-0,17399798	3,027529808	2,9	F	1,434897127	3,604385453	9,7
<i>Hibiscus cannabinus</i>	-0,17141107	2,938175565	5,2	X61	-1,96419565	3,003778085	12,6
<i>Acacia hockii J</i>	-0,16620387	2,762372679	85,2	C	2,948027705	2,683033182	12,8
<i>Acacia senegal A</i>	-0,16547459	2,738183838	70,9	X20	1,182870945	2,593939431	17,2
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Achyranthes aspera var sicula</i>	0,315712561	9,967442107	0,1	B	4,291176443	8,22949935	1,1
<i>Acacia senegal A</i>	0,251774941	6,339062093	43,8	A	2,661511406	5,939949577	2,9
<i>Acacia polyacantha A</i>	0,25118667	6,309474308	0,2	I	-3,49130735	5,2504266	3,7
<i>Panicum maximum</i>	0,220742022	4,872704045	7,2	D	-3,09781429	4,929298421	3,9
<i>Setaria barbata</i>	0,217179173	4,716679337	3,6	X68	2,239873514	4,746260437	4,6
<i>Dichrostachys cinerea J</i>	0,201490279	4,059833273	51,9	E	-2,15561563	4,591882107	5,1
<i>Commelina benghalensis</i>	0,19509359	3,806150877	2,9	C	3,777524666	4,405320806	6
<i>Loudetia simplex</i>	-0,19334108	3,738077281	1,3	X47	-1,76151289	4,216808347	6,8
<i>Spermacoce pusilla</i>	-0,17369181	3,016884354	4	X62	1,828634006	3,519061607	8,3
<i>Sporobolus stapfianus</i>	-0,16584379	2,750416248	11,2	X33	-1,57779386	3,039064015	11,3
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Albica kirkii</i>	0,320772136	10,28947635	7,6	I	4,918342153	10,41971651	0,5
<i>Acacia polyacantha A</i>	0,268158828	7,190915702	0,1	C	4,988695638	7,683108024	1,1
<i>Loudetia simplex</i>	0,258640933	6,689513211	0	B	4,016534795	7,209808647	1,8
<i>Acacia hockii J</i>	-0,19817944	3,927509129	41,2	D	3,336491382	5,718134541	2,6
<i>Ormocarpon trichocarpum J</i>	-0,19423327	3,772656268	21,6	A	2,523450027	5,339683458	3,5
<i>Sporobolus stapfianus</i>	0,176529656	3,116271936	11,1	K	1,964411181	4,755168009	4,8
<i>Acacia senegal A</i>	0,169496043	2,87289087	52,6	H	2,355974997	4,602303709	6,7
<i>Asparagus abyssinicus</i>	-0,14677704	2,154349816	9,7	G	1,75548669	4,470948348	5,8
<i>Plectranthus pauciflorus</i>	-0,13858614	1,920611896	8,8	E	1,854664298	3,399215554	9,6
<i>Acalypha bipartita</i>	0,138441191	1,916596336	0,6	X12	-1,22065085	2,395488469	14,9
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Acacia polyacantha A</i>	-0,41020576	16,82687666	0	B	-5,77213975	14,88998795	0
<i>Acacia senegal A</i>	0,317230669	10,06352975	13,2	C	-6,48694008	12,99100336	0,1
<i>Acacia hockii J</i>	-0,28321225	8,020917793	15	A	-2,84251455	6,775346131	1,7
<i>Setaria barbata</i>	-0,25613971	6,560755264	1,3	X68	2,307287883	5,036260053	5,6
<i>Dichrostachys cinerea J</i>	0,220299849	4,853202356	29,1	X70	2,60541073	4,424867295	6,4
<i>Panicum maximum</i>	0,190616031	3,633447115	16,5	X69	2,265682377	3,972544313	8,9
<i>Albica kirkii</i>	0,180305763	3,251016817	36	X50	-1,83469447	3,791576906	7,5
<i>Brachiaria brizantha</i>	-0,14319256	2,050410909	0,3	X62	1,714753917	3,094403142	11,4
<i>Acalypha bipartita</i>	-0,14120115	1,993776489	0,5	X64	1,58127545	2,993887655	11,2
<i>Cissampelos mucronata</i>	-0,12907242	1,665968992	0	X2	-1,43397922	2,632600802	15,7

Tableau 12. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives. Axes 1 à 4.

Analyse statistique des données de végétation

Sur l'axe 1, on trouve d'un côté des espèces de *Poaceae* et autres de savane herbeuse et de l'autre, des espèces et relevés de bosquets xérophiiles denses.

L'axe 2 oppose des espèces et relevés de savane densément boisée à *Acacia polyacantha* et *Acacia senegal* et de l'autre, des espèces de sols superficiels.

L'axe 3 oppose ces éléments de l'axe 2 aux espèces et relevés de savane arbustive.

L'axe 4 oppose les relevés et espèces de galerie forestière et ceux de savane à *Acacia senegal*.

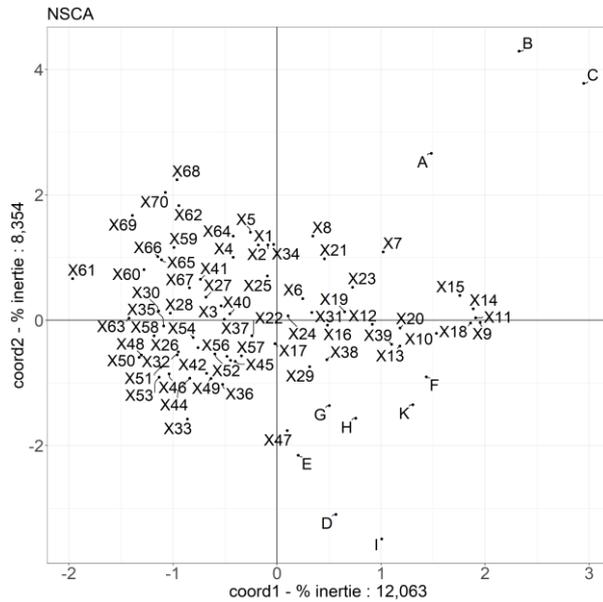


Figure 27. Analyse non symétrique des correspondances du fichier Akagera avec transformation logarithmique. Coordonnées des relevés sur les axes 1 et 2.

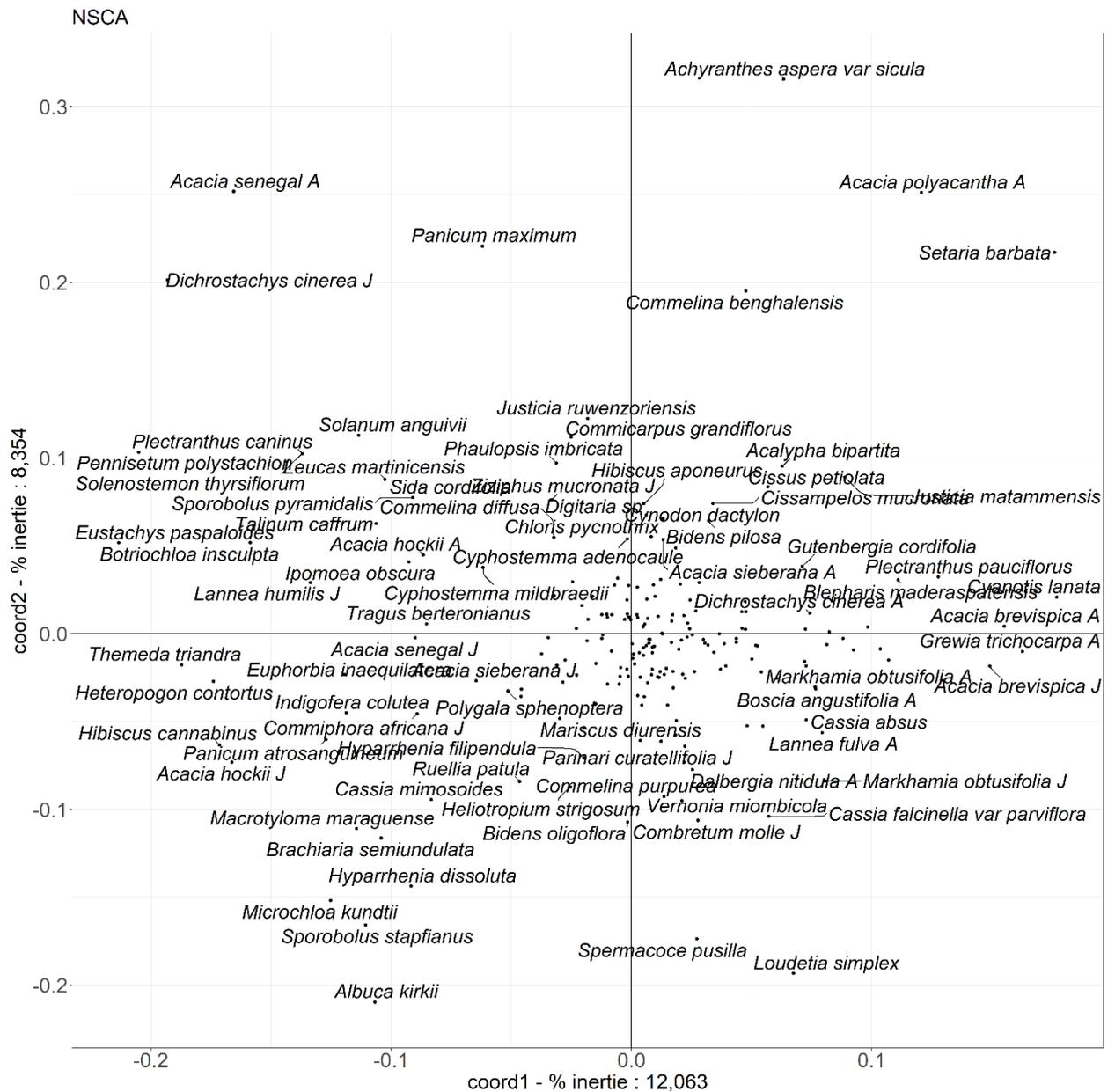


Figure 28. Analyse non symétrique des correspondances du fichier Akagera avec transformation logarithmique. Coordonnées des espèces sur les axes 1 et 2.

Cette transformation logarithmique des abondances apporte des résultats fort proches de celles de la transformation 113.

Analyses factorielles multiples

Un autre problème est la présence, dans la même table, de plantes herbacées et d'espèces ligneuses de grande taille avec des motifs de dispersion très différents. Dans une végétation de savane, par exemple, les plantes ligneuses, des *Poaceae* et *Cyperaceae*, formant des touffes denses et d'autres espèces montrent une grande diversité de motifs, depuis les petits agrégats, les plaques, les variations de densité, les gradients et les motifs complexes. Si chaque catégorie de plantes est soumise séparément à une analyse multivariée, il apparaît que chaque sous-ensemble a sa propre structure. Ainsi, les espèces herbacées, par exemple, La différence de structure peut s'exprimer, par exemple,

Analyse statistique des données de végétation

par la première valeur propre de NSCA. Si ces valeurs sont très différentes, le choix d'une analyse multiple est nécessaire pour rééquilibrer les sous-ensembles.

Fichier disjonctif simple

L'analyse est conduite avec trois sous-tableaux : un de 116 lignes pour les plantes ligneuses, un de 57 lignes pour les *Poaceae* et *Cyperaceae* et un de 139 lignes pour les autres plantes herbacées. Les douze premières valeurs propres sont significatives. Les densités des sous-tableaux sont respectivement de 0,0742, 0,17982456 et 0,17230216 ; cela signifie que les espèces ligneuses sont privilégiées par rapport aux herbacées. Les principaux résultats sont présentés dans les figures 29 à 31 et dans le tableau 13. Les résultats complets se trouvent dans le fichier Akagera80disj12345mfansca.xlsx

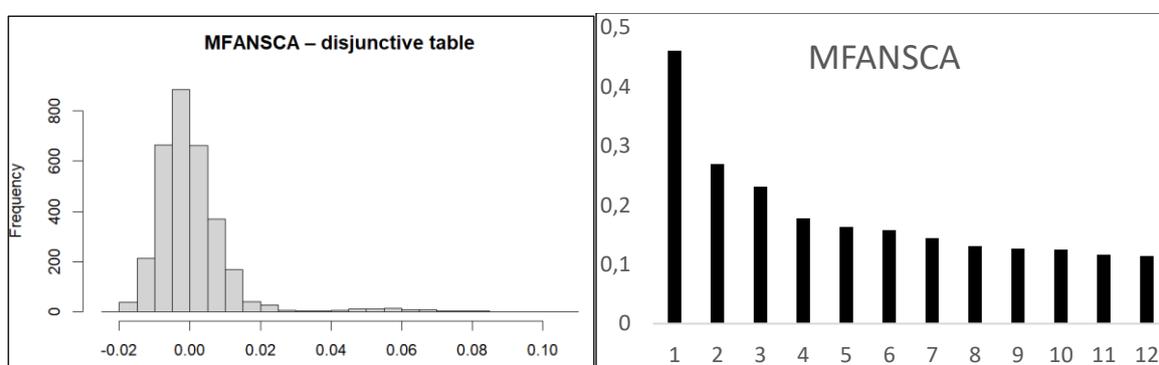


Figure 29. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Lansea humilis</i> J	0,228837702	5,236669379	0	X9	-1,69735155	5,911751981	2,4
<i>Hibiscus cannabinus</i>	0,189627734	3,595867733	0,2	X15	-1,94795826	5,349843569	3,5
<i>Pennisetum polystachion</i>	0,181837929	3,306503247	0,7	X18	-1,82717711	5,056154637	4,3
<i>Heteropogon contortus</i>	0,180926039	3,27342316	0,6	X14	-1,93909954	5,0495039	4
<i>Albica kirkii</i>	0,172709029	2,982840887	1,3	X11	-1,59936695	4,531080513	5,3
<i>Botriochloa insculpta</i>	0,167252781	2,797349259	0,9	X10	-1,25767857	3,137842808	12
<i>Commiphora africana</i> J	0,15996519	2,558886205	1,8	F	-1,30470719	2,991867615	13
<i>Acacia senegal</i> J	0,156040086	2,434850835	0,1	X13	-1,25829634	2,544641739	13
<i>Acacia brevispica</i> J	-0,15347631	2,355497799	0	X61	1,80302273	2,495695742	17
<i>Markhamia obtusifolia</i> J	-0,15158521	2,297807691	1,9	X39	-1,28710761	2,488907276	16
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Achyranthes aspera</i>	0,26539892	7,043658699	0	K	-2,20166576	6,934855355	1,5
<i>Commelina benghalensis</i>	0,202324818	4,093533212	0,1	X68	2,31181371	4,815076888	5
<i>Spermacoce pusilla</i>	-0,20106872	4,042862821	0,2	I	-3,31819606	4,550712087	4,7
<i>Combretum molle</i> J	-0,18251188	3,33105851	0	H	-2,3482996	4,538522492	5,9
<i>Panicum maximum</i>	-0,16498708	2,722073685	1,2	X1	1,82599952	4,277440785	6,9
<i>Hyparrhenia dissoluta</i>	-0,15980916	2,55389661	2,1	X70	2,56061586	4,261242898	4,3
<i>Ziziphus mucronata</i> J	0,150224349	2,256735512	0	X62	2,05288651	4,020057363	7
<i>Justicia ruwenzoriensis</i>	0,150012464	2,250373927	0	X5	1,80130522	3,926828505	8,7
<i>Panicum maximum</i> 2	0,149699483	2,240993511	1	X59	1,73550218	3,632982171	8,1
<i>Markhamia obtusifolia</i> J	-0,14775075	2,183028419	2,7	E	-1,96365876	3,565049738	10
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Lansea humilis</i> J	-0,24837846	6,169185773	0	K	2,42066137	8,383062603	0,5
<i>Ormocarpon</i>	-0,22122611	4,894099095	0	G	1,9903612	6,112347677	1,8
<i>Acacia hockii</i> J	-0,20206673	4,08309643	0	I	3,40633722	4,795683921	4,5
<i>Acacia senegal</i> J	0,187756997	3,525268984	0	X32	-1,77281425	4,035795282	6,3
<i>Vernonia miombicola</i>	-0,16109209	2,595066165	1,2	H	2,19377105	3,96086539	6,6

Analyse statistique des données de végétation

<i>Blepharis cristata</i>	-0,15563974	2,422372935	0	X68	2,09050257	3,9373042	7,9
<i>Boscia angustifolia</i> J	-0,15402882	2,372487741	0	C	3,7106775	3,854565547	9,3
<i>Commiphora africana</i> J	-0,14934833	2,230492445	3,3	X70	2,35094842	3,59197769	9,1
<i>Ruellia patula</i>	-0,14373713	2,066036268	3,5	A	2,25915251	3,507742517	9,3
<i>Asparagus abyssinicus</i>	-0,13917043	1,936840831	2,2	X62	1,91269199	3,489734982	10
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Markhamia obtusifolia</i> J	-0,35678743	12,72972681	0	K	-2,19215992	6,875101195	1,9
<i>Commiphora africana</i> J	-0,27523249	7,575292375	0	I	-4,01389826	6,658984326	2,4
<i>Dichrostachys cinerea</i> J	0,267925698	7,178417991	0	X45	2,21807026	5,976539014	2
<i>Acacia hockii</i> A	-0,25245989	6,373599492	0	D	3,54650167	5,291053076	4,2
<i>Dichrostachys cinerea</i> J 2	-0,1970746	3,883839617	0	X55	-2,06475049	5,214376457	4
<i>Justicia matamensis</i>	0,164505674	2,706211675	1	X36	1,72288125	4,398930065	5,9
<i>Microchloa kundtii</i>	0,163733475	2,680865097	1,1	G	-1,55685362	3,739726277	8,8
<i>Markhamia obtusifolia</i> A	-0,14796055	2,189232467	0,1	X31	1,36029284	3,043730388	12
<i>Indigofera colutea</i>	0,145747741	2,124240413	1,4	X11	1,2918838	2,956326859	13
<i>Dalbergia nitidula</i> J	-0,13144247	1,727712249	0	B	2,95900816	2,895118005	13

Tableau 13. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions. Axes 1 à 4.

Avec l'axe 1, on se retrouve avec l'opposition entre espèces de savane herbeuse et bosquets xérophiles mais cette fois plusieurs espèces ligneuse sont mises en évidence comme *Lannea humilis* J, *Commiphora africana* J pour les milieux ouverts et *Acacia brevispica* J, *Markhamia obtusifolia* et *Acacia senegal* pour les milieux fermés.

Dans l'axe 2 qui oppose les végétations de sols superficiels et de savane boisée à *Acacia senegal*, deux espèces ligneuses sont aussi mises en évidence à savoir *Combretum molle* J et *Markhamia obtusifolia*. On peut aussi remarquer *Panicum maximum* avec l'abondance 2 qui se développe bien sous les *Acacia senegal*.

L'axe 3 est fortement influencé par les espèces ligneuses jeunes (J) qui accompagnent plusieurs espèces herbacées caractéristiques dans les milieux ouverts alors que les jeunes *Acacia senegal* préfèrent les savanes boisées.

L'axe 4 est fortement influencé par les espèces de savane boisée à *Acacia hockii*, nettement moins densément couverte que des savanes à *Acacia senegal* ou à *Acacia polyacantha*.

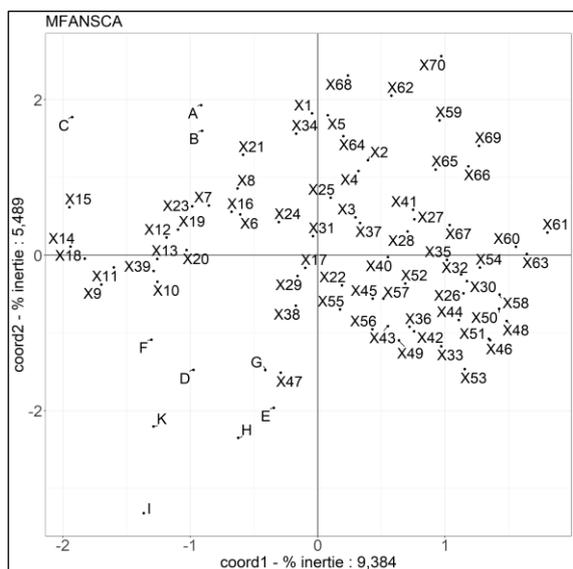


Figure 30. Analyse factorielle multiple basée sur l'analyse non symétrique des correspondances du fichier Akagera disjonctif simple. Coordonnées des relevés sur les axes 1 et 2.

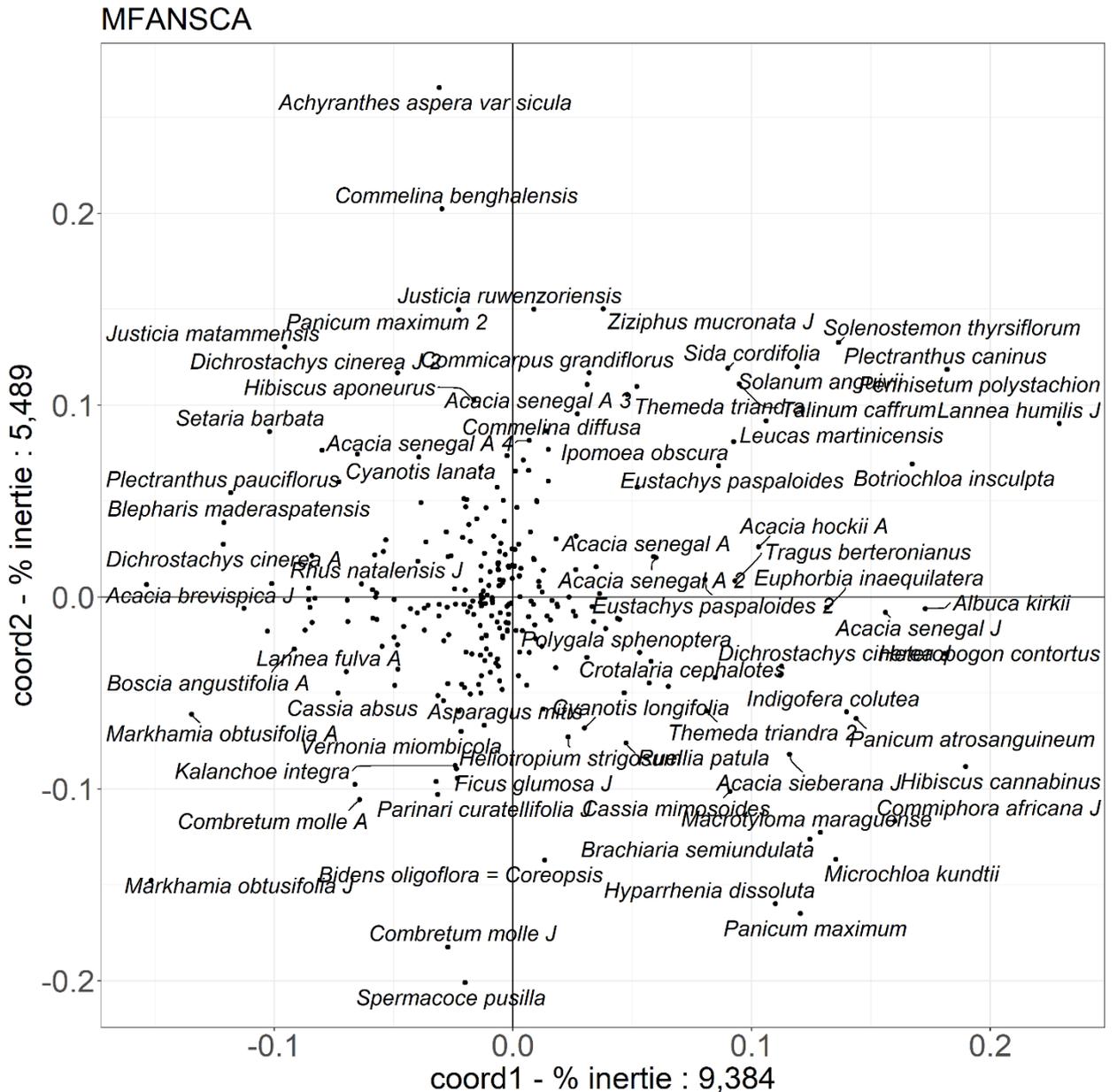


Figure 31. Analyse factorielle multiple basée sur l'analyse non symétrique des correspondances du fichier Akagera disjonctif simple. Coordonnées des espèces sur les axes 1 et 2.

Fichier 113

L'analyse est conduite avec trois sous-tableaux : un de 105 lignes pour les plantes ligneuses, un de 52 lignes pour les *Poaceae* et *Cyperaceae* et un de 136 lignes pour les autres plantes herbacées. Les densités des sous-tableaux sont respectivement de 0,0995, 0,2199 et 0,17830882 ; les plantes ligneuses sont à nouveau favorisées par rapport aux espèces herbacées, et parmi celles-ci les plantes herbacées autres que les *Poaceae* et *Cyperaceae* sont un peu favorisées aussi. Les principaux résultats sont présentés dans les figures 32 à 34 et le tableau 14. Les résultats complets se trouvent dans le fichier Akagera80disj113mfansca.xlsx.

Les huit premières valeurs propres sont significatives.

Analyse statistique des données de végétation

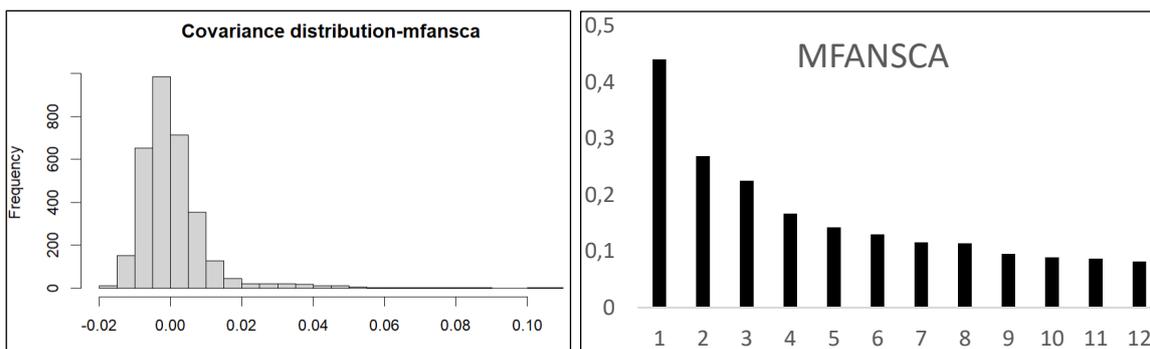


Figure 32. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Species	coord 1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Lamnea humilis</i> J	-0,19580723	3,83404713	0,3	X9	1,922770785	7,566415081	1,2
<i>Pennisetum polystachion</i>	-0,18126604	3,85737582	0,1	X11	1,821345519	5,815800319	1,6
<i>Hibiscus cannabinus</i>	-0,17822665	3,176473905	0,5	X18	1,914920945	5,421226903	1,9
<i>Acacia brevispica</i> J	0,17037193	2,902659457	0	X10	1,507844311	4,549129674	5,3
<i>Heteropogon contortus</i>	-0,15735716	2,476127559	1,9	X14	1,796394474	4,526544571	4,6
<i>Dichrostachys cinerea</i> J	-0,15492261	2,400101361	0	X15	1,732678895	4,184965157	6,6
<i>Eustachys paspaloides</i>	-0,15437563	2,383183386	0	F	1,454645365	3,857749558	8,7
<i>Botriochloa insculpta</i>	-0,14976169	2,24285623	2,2	K	1,498154028	2,972721932	13
<i>Grewia trichocarpa</i> A	0,145613794	2,120337708	0	X39	1,280153875	2,733856701	17
<i>Acacia senegal</i> A	-0,14246327	2,029578362	0,3	X20	1,157426961	2,587225225	16
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Achyranthes aspera</i> var. <i>sicula</i>	0,286930682	8,2329216	0	X68	2,410010117	5,676863802	2,5
<i>Commelina benghalensis</i>	0,208527626	4,348377064	0,2	X70	2,619458378	4,476846065	5,8
<i>Acacia senegal</i> A	0,205037708	4,204046187	0	X62	2,034383213	4,4566812	5,6
<i>Panicum maximum</i> >1	0,198629824	3,945380715	0	X47	-1,74845967	4,013447238	7,3
<i>Acacia senegal</i> A >1	0,191705225	3,675089327	0	I	-2,84318778	3,691682954	7,9
<i>Spermacoce pusilla</i>	-0,18626242	3,469368907	0,6	E	-1,94021047	3,583206121	9,2
<i>Dichrostachys cinerea</i> J >1	0,164647188	2,710869641	1,1	A	2,099624985	3,560241665	9,9
<i>Hyparrhenia dissoluta</i>	-0,15825073	2,504329402	2,7	K	-1,56482723	3,243203086	11
<i>Combretum molle</i> J	-0,15157021	2,297352711	0	X33	-1,66258394	3,217648424	12
<i>Microchloa kundtii</i>	-0,1452494	2,109738753	3,6	X64	1,63837519	3,200054841	13
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Acacia hockii</i> J	-0,23925453	5,724273189	0	K	2,527092123	8,458308139	0,5
<i>Ormocarpon trichocarpum</i> J	-0,23801289	5,665013683	0	I	4,269082524	8,323048164	0,5
<i>Lamnea humilis</i> J	-0,20047461	4,019007086	0	G	2,128827922	6,913838487	1,5
<i>Albucca kirkii</i>	0,199639776	3,985604011	0,1	H	2,62448696	5,618466675	3,8
<i>Acacia senegal</i> A >1	0,160798042	2,585601017	0	X69	2,208788398	3,575643925	11
<i>Acacia senegal</i> A	0,159359458	2,53954368	0	X68	1,771017296	3,065606078	13
<i>Asparagus abyssinicus</i>	-0,15225308	2,318100076	1,5	X62	1,594906788	2,739155292	14
<i>Acacia hockii</i> A	-0,14717191	2,165957161	0,8	X70	2,011577156	2,640113194	14
<i>Blepharis cristata</i>	-0,14156728	2,004129545	0,3	X32	-1,4487688	2,539299422	14
<i>Plectranthus pauciflorus</i>	-0,1344135	1,806698981	1,2	C	2,818127445	2,506484055	17
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Acacia polyacantha</i> A	-0,31237972	9,75810925	0	B	-6,03624455	15,18214296	0
<i>Acacia polyacantha</i> A >3	-0,24032899	5,775802392	0	C	-6,86901302	14,89129155	0
<i>Acacia hockii</i> A	-0,1998283	3,993135007	0	A	-3,74618424	11,33376696	0
<i>Acacia hockii</i> J	-0,19585114	3,835767094	0,6	X2	-2,20770499	5,857079683	3,1
<i>Setaria barbata</i>	-0,17852863	3,187247085	0	X50	-1,61767367	2,94866308	12
<i>Acacia senegal</i> A	0,170728809	2,914832634	0	X1	-1,4957051	2,931221497	14
<i>Cissampelos mucronata</i>	-0,15528265	2,411270007	0	X7	-1,42820648	2,827697034	15
<i>Crabbea velutina</i>	0,152539492	2,32682965	2,5	X61	1,869004586	2,754013813	14
<i>Albucca kirkii</i>	0,151659897	2,300072442	2,5	X70	1,929537251	2,42915648	17
<i>Cissus petiolata</i>	-0,15133498	2,290227474	0	X5	-1,35619787	2,338373745	19

Tableau 14. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives. Axes 1 à 4.

Analyse statistique des données de végétation

Sur l'axe 1, on retrouve l'opposition entre espèces et relevés de bosquets xérophiles (*Grewia trichcarpa* A et *Acacia brevispica* J) et des espèces de savane avec de jeunes individus de *Lannea humilis*, *Dichrostachis cinerea* ou des *Acacia senegal* adultes.

Sur l'axe 2, il y a d'un côté les relevés de savane boisée dense avec *Acacia senegal* présent et abondant et *Acacia polyacantha* présent et abondant avec *Achyranthes aspera* var. *sicula* présent dans les deux formations et de l'autre côté des relevés à sol superficiel. Plusieurs espèces ligneuses sont encore mises en évidence.

Sur l'axe 3, les jeunes individus d'*Acacia hockii*, *Ormocarpon trochocarpon* et *Lannea humilis* en association avec les *Acacia hockii* adultes s'opposent aux *Acacia senegal* adultes.

Sur l'axe 4, c'est surtout la savane à *Acacia senegal* qui est mise en évidence, en opposition à la galerie forestière à *Acacia polyacantha*.

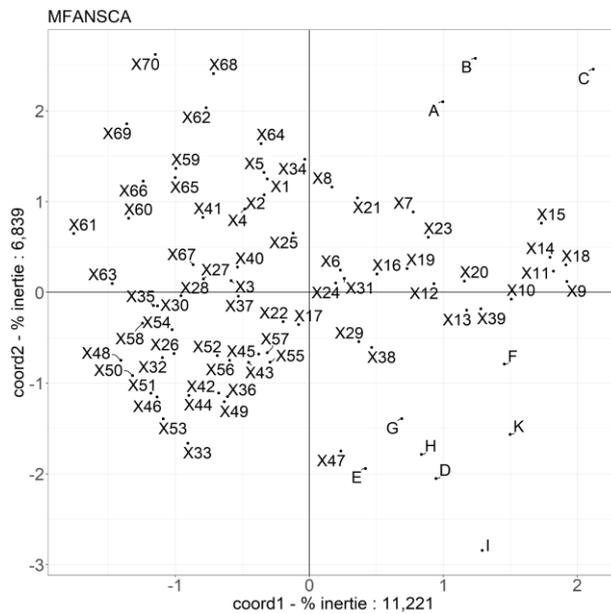


Figure 33. Analyse factorielle multiple du fichier Akagera disjonctif 113. Coordonnées des relevés sur les axes 1 et 2.

Analyse statistique des données de végétation

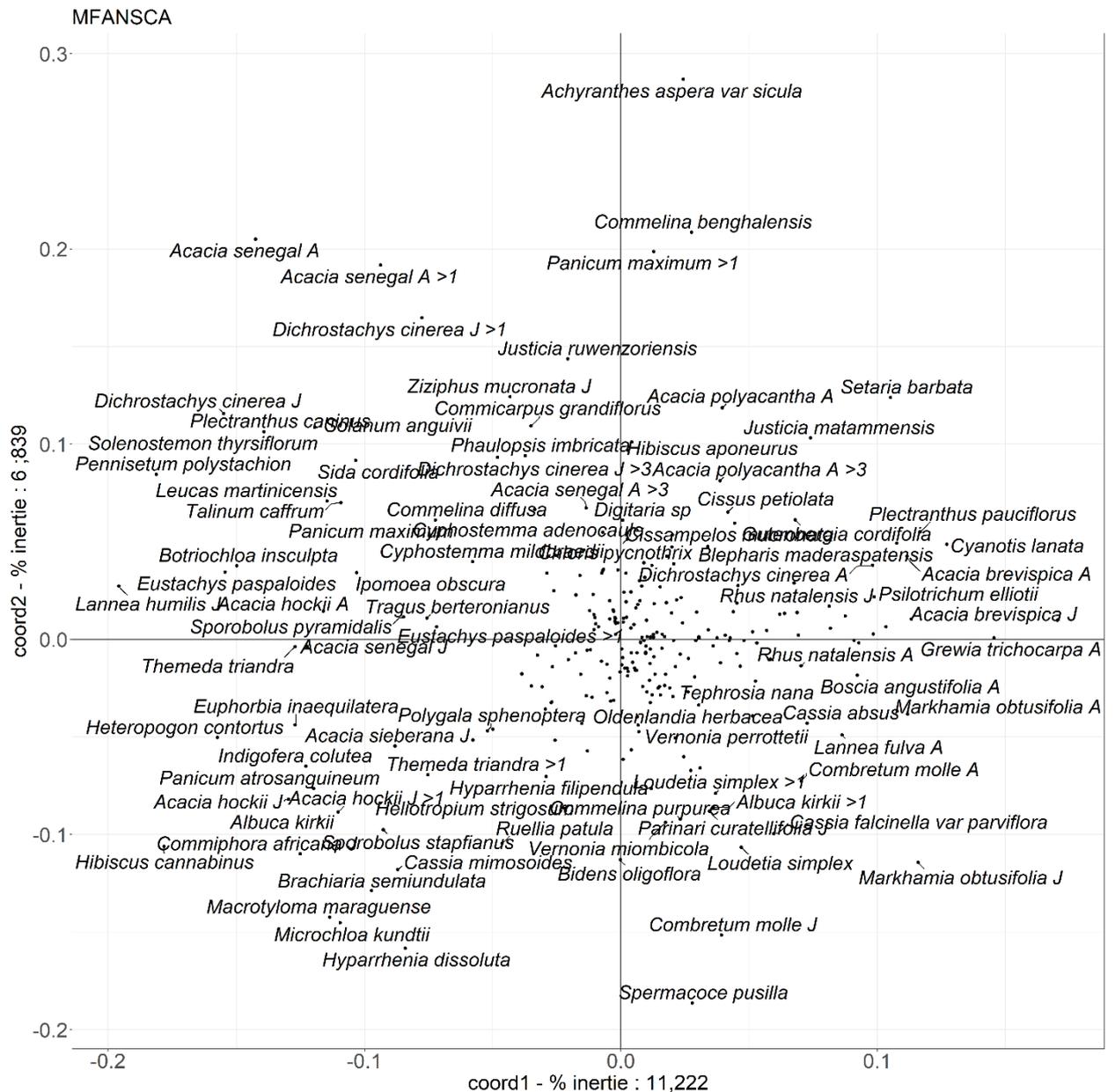


Figure 34. Analyse factorielle multiple du fichier Akagera disjonctif 113. Coordonnées des espèces sur les axes 1 et 2.

Fichier logarithmique

Les densités des sous-tableaux sont respectivement de 0,0981 pour les plantes ligneuses, 0,2011 pour les *Poaceae* et *Cyperaceae* et 0,1308 pour les autres plantes herbacées.

Les huit premières valeurs propres sont significatives. Les principaux résultats apparaissent dans figures 35 à 37 et le tableau 15. Les résultats complets apparaissent dans le fichier Akagera80phytologmfansca.xlsx.

Analyse statistique des données de végétation

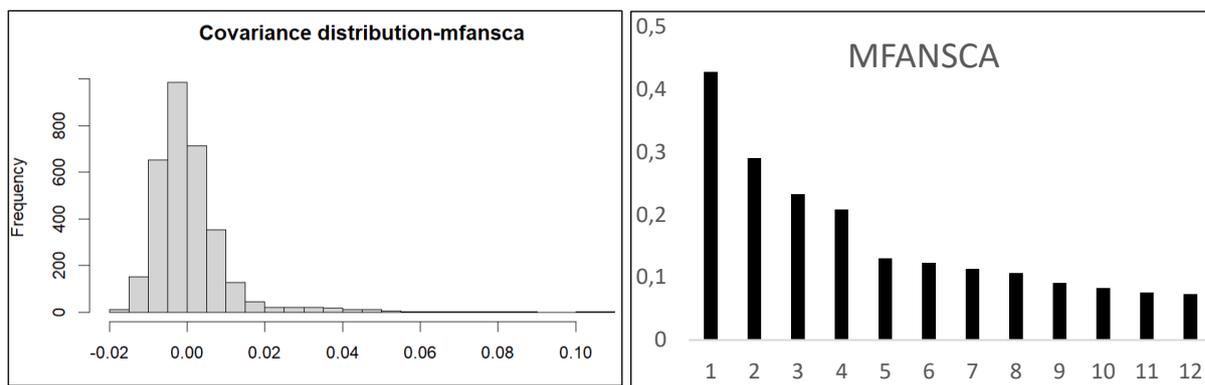


Figure 35. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Species	coord1	Cr%1	P1	Relevés	coord1	Cr%1	P1
<i>Dichrostachys cinerea</i> J	-0,259425669	6,730167794	9,8	X9	1,907155037	7,312626847	1,6
<i>Acacia senegal</i> A	-0,217828126	4,744909248	22,8	X11	1,845086894	5,944282292	3
<i>Acacia hockii</i> J	-0,197421614	3,897529355	65,3	X18	1,845976745	5,079864247	3,7
<i>Grewia trichocarpa</i> A	0,190712865	3,63713967	1,5	X10	1,532706249	4,605218095	5,1
<i>Eustachys paspaloides</i>	-0,176273823	3,107246078	0	X14	1,722529192	4,186893032	6,9
<i>Acacia brevispica</i> A	0,175048854	3,064210133	6,9	F	1,497387597	4,0847892	7,3
<i>Pennisetum polystachion</i>	-0,170839478	2,918612731	0	X15	1,575643301	3,496978094	8
<i>Acacia brevispica</i> J	0,169707648	2,880068563	0	X61	-2,02020901	3,118859787	11
<i>Hibiscus cannabinus</i>	-0,165536144	2,740221499	0	K	1,508590277	3,006253768	12
<i>Lankea humilis</i> J	-0,164808645	2,716188931	0,1	X20	1,195296654	2,757832704	11
Species	coord2	Cr%2	P2	Relevés	coord2	Cr%2	P2
<i>Acacia polyacantha</i> A	-0,343403329	11,79258462	0	B	-5,09063644	11,37806905	0,2
<i>Achyranthes aspera</i> var <i>sicula</i>	-0,333722427	11,1370658	0	A	-3,12140265	8,007325565	1
<i>Acacia senegal</i> A	-0,313871549	9,851534909	1,7	C	-4,90324743	7,865918784	1,2
<i>Commelina benghalensis</i>	-0,204469025	4,180758229	0	X68	-2,31742992	5,030181317	4,3
<i>Setaria barbata</i>	-0,19715431	3,886982178	0	X62	-2,05580598	4,444753951	5,7
<i>Dichrostachys cinerea</i> J	-0,196007732	3,841903118	25,6	X70	-2,45950794	3,93575739	7,8
<i>Acacia hockii</i> J	0,170017445	2,890593154	39,5	X47	1,467290986	2,870537676	16
<i>Spermacoce pusilla</i>	0,167671184	2,811362604	0,1	E	1,685163394	2,689228215	15
<i>Panicum maximum</i>	-0,157556095	2,482392297	2,6	I	2,448476484	2,674275654	13
<i>Albica kirkii</i>	0,152955054	2,339524843	17,4	X69	-1,85657778	2,593679573	15
Species	coord3	Cr%3	P3	Relevés	coord3	Cr%3	P3
<i>Albica kirkii</i>	0,356480508	12,70783526	0	I	5,03395195	11,30400726	0,1
<i>Ormocarpon trichocarpum</i> J	-0,230098095	5,29451334	0,1	G	2,034947753	6,147117434	3
<i>Loudetia simplex</i>	0,22278379	4,963261716	0	K	2,133156416	6,010746783	2,5
<i>Acacia polyacantha</i> A	0,214859697	4,616468954	0	D	3,488016295	5,802830858	2,2
<i>Acacia hockii</i> J	-0,204818602	4,195065973	15,3	H	2,53774337	5,360175937	4
<i>Sporobolus stapfianus</i>	0,164418129	2,70332109	0,5	E	2,275332401	4,902677529	5,8
<i>Acacia senegal</i> A	0,161418753	2,605601391	31,9	C	3,839022577	4,821955208	5,2
<i>Lankea humilis</i> J	-0,155250311	2,410265904	0,4	B	2,800794373	3,444188019	9,1
<i>Acacia brevispica</i> A	-0,148847552	2,215559386	17,8	A	1,884626482	2,91902945	14
<i>Dalbergia nitidula</i> A	0,146879675	2,157363896	0	X21	-1,40803445	2,817180358	15
Species	coord4	Cr%4	P4	Relevés	coord4	Cr%4	P4
<i>Acacia polyacantha</i> A	-0,478723905	22,91765776	0	C	-6,7479814	14,8980557	0
<i>Acacia senegal</i> A	0,3478979	12,10329488	0	B	-5,81119363	14,82706169	0,1
<i>Acacia hockii</i> J	-0,332564836	11,05993703	0,4	A	-3,23460269	8,598640353	1,2
<i>Dichrostachys cinerea</i> J	0,237440143	5,637782157	5,8	X68	2,202905527	4,545295941	5,6
<i>Setaria barbata</i>	-0,200190687	4,007631099	0,1	X70	2,435470277	3,85920216	7,1
<i>Albica kirkii</i>	0,166775857	2,781418651	13,1	X50	-1,79434528	3,56810159	8,6
<i>Acacia hockii</i> A	-0,157089785	2,467720052	1,2	X69	2,050446365	3,163638515	9,5
<i>Acalypha bipartita</i>	-0,143101459	2,047802765	0	X62	1,705678588	3,05969312	12
<i>Acacia brevispica</i> A	0,131518453	1,729710349	25,6	X64	1,47589798	2,614536218	16
<i>Cissampelos mucronata</i>	-0,126383686	1,597283615	0	X2	-1,38304666	2,344620374	16

Tableau 15. Dix premières espèces et dix premiers relevés rangés par ordre décroissant des contributions relatives sur les axes. Axes 1 à 4.

Analyse statistique des données de végétation

Sur l'axe 1, la place prise par les espèces ligneuses est encore plus importante que dans les analyses multivariées précédentes et l'on retrouve l'opposition entre bosquets et savane, principalement boisée puisque cette analyse privilégie les espèces ligneuses.

Sur l'axe 2, ce sont les savanes les plus densément boisées, à *Acacia polyacantha* ou *Acacia senegal* qui sont opposées aux formations les plus ouvertes sur sol superficiel.

Sur l'axe 3, les sites ouverts avec un grand nombre de petits individus ligneux d'*Acacia hockii*, *Lannea humilis* ou encore *Ormocarpum trichocarpum* qui s'opposent aux savanes boisées avec de grands individus.

L'axe 4 sépare la savane boisée à *Acacia polyacantha* à la savane à *Acacia senegal*.

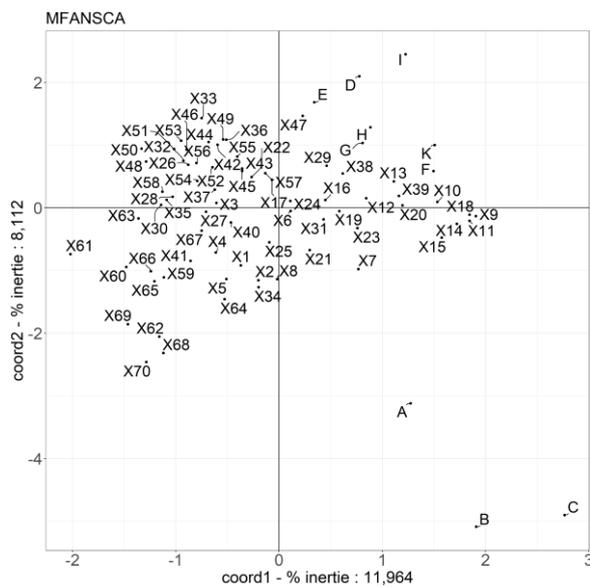


Figure 36. Analyse factorielle multiple du fichier Akagera logarithmique. Coordonnées des relevés sur les axes 1 et 2.

Analyse statistique des données de végétation

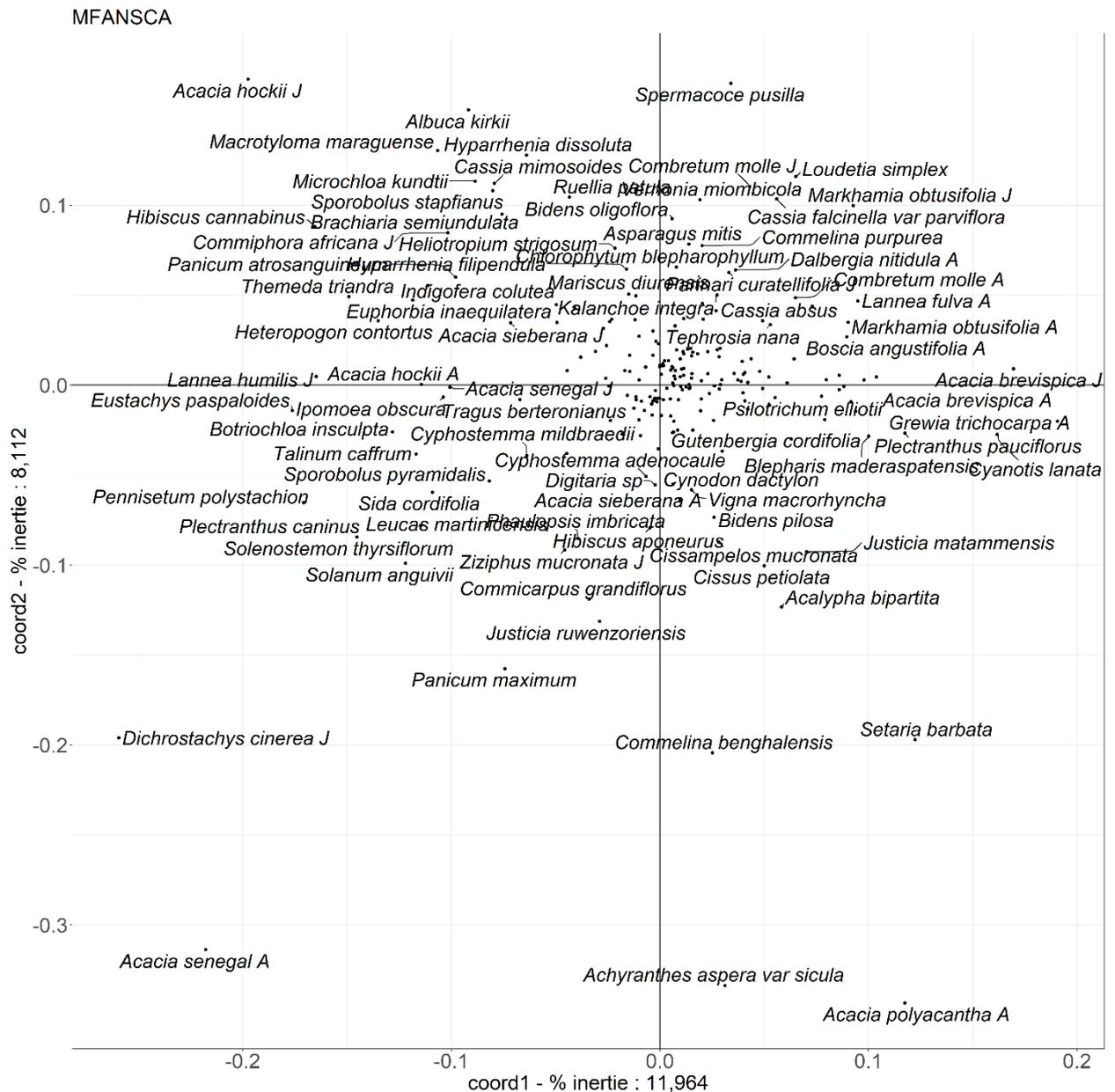


Figure 37. Analyse factorielle multiple du fichier Akagera logarithmique. Coordonnées des espèces sur les axes 1 et 2.

Comme on pouvait s'y attendre, ce sont les espèces ligneuses qui dominent l'analyse, par rapport aux espèces herbacées, et notamment les graminées de savane ouverte.

Fichiers floristique et environnemental réunis

A titre d'exemple, nous groupons le fichier floristique disjonctif de type 113 et le fichier environnemental de type disjonctif simple. Ce choix est le fruit des analyses précédentes. C'est le fichier disjonctif 113 soumis à une analyse factorielle multiple qui nous apporte la meilleure aide à la compréhension de la végétation. Pour les données mésologiques, le fichier disjonctif simple est aussi le mieux adapté à ce type de données et aux relations entre facteurs

Analyse statistique des données de végétation

mésologiques et variables floristiques qui ne peuvent être modélisées (voir conclusions). Il a été quelque peu simplifié par rapport au fichier complet originel.

La transformation du fichier mésologique brut en fichier disjonctif se fait de la manière suivante ; les noms des paramètres sont représentés en caractères gras dans la figure 37 :

- Profondeur du sol : trois classes (**sol profond, sol moy profond, lithosol**),
- Pente : trois classes de 2 à 50 % (**pente-1 à 3**),
- Milieu ouvert ou fermé : trois classes (**milieu ouvert, milieu moy ferme, milieu fermé**),
- Type de sol : tropical récent (**sol recent**), **ferralsol**, ferrisol humifère, (**ferrisol**),
- Couvert léger : cinq classes sur base de la somme des surfaces terrières, de 0 à 35 dm² (**couvert1-1 à 4**),
- Couvert dense : trois classes sur base de la somme des surfaces terrières, de 0 à 35 dm² (**couvert2-1 à 2**),
- Sol nu : cinq classes sur base de la proportion de contact avec le sol nu dans un ensemble de 125 points (**vide-1 à 4**),
- Horizon humifère : cinq classes sur base de l'épaisseur en cm (**humus-1 à 4**),
- La capacité de rétention en eau (ou capacité au champ), cinq classes de 17 à 46 % (**ret eau-1 à 4**),
- Le pH en surface (horizon A1) et profondeur : cinq classes de 4,9 à 7,5 (**pH S-1 à 4 et pH P-1 à 4**),
- La somme des cations totaux échangeables en surface (horizon A1) et profondeur, quatre classes de 3,9 à 21,5 méq/100g (**S S-1 à 4 et S P-1 à 4**),
- Le pourcentage de saturation en bases en surface (horizon A1) et profondeur, quatre classes de 29,7 à 87 % (**V S-1 à 4 et V P-1 à 4**),
- La teneur en matières organiques de l'horizon A1 : quatre classes de 3,9 à 21,5 % (**MO-1 à 4**).

L'analyse factorielle multiple est faite sur base de trois sous-tableaux, un de 105 lignes pour les plantes ligneuses (densité = 0,0995), un de 188 lignes pour les plantes herbacées (densité = 0,1898) et un de 62 lignes pour les paramètres mésologiques (densité = 0,2419). C'est donc le fichier « plantes ligneuses » qui est ici favorisé. Les principaux résultats sont présentés dans les figures 38 et 39 et dans la table 16. Les résultats complets se trouvent dans le fichier Akagera80disj113envmfansca.xlsx.

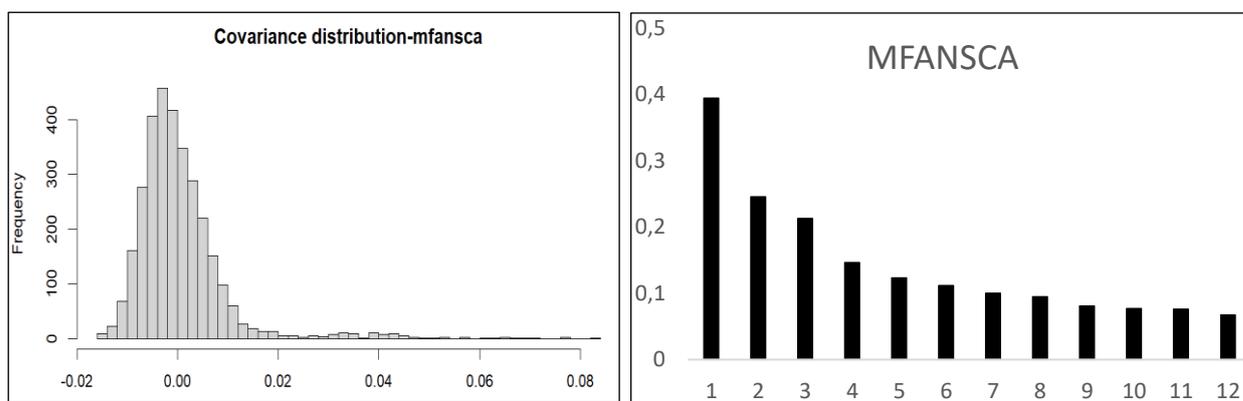


Figure 38. Distribution des covariances en 40 classes et les douze premières valeurs propres.

Les huit premières valeurs propres sont significatives. Seules les données floristiques et mésologiques sont ici présentées.

Analyse statistique des données de végétation

Sp et par. mésologiques	coord1	CR%1	P1	Sp et par. mésologiques	coord2	CR%2	P2
<i>Pennisetum polystachion</i>	0,173654	3,0155712	2,7	<i>Achyranthes aspera var sicula</i>	-0,230961	5,33431645	0,2
<i>Acacia brevispica J</i>	-0,1675273	2,8065384	12,4	<i>Panicum maximum >1</i>	-0,1809	3,27246696	2,3
<i>Lannea humilis J</i>	0,15550274	2,4181103	49,7	<i>Spermacoce pusilla</i>	0,172541	2,97705291	3,4
<i>Hibiscus cannabinus</i>	0,15006793	2,2520384	6,3	<i>Commelina benghalensis</i>	-0,169021	2,85679495	2,8
<i>Heteropogon contortus</i>	0,14653509	2,1472533	5,4	<i>Microchloa kundtii</i>	0,159389	2,54048048	3,4
S P-2	0,14589001	2,1283895	1,9	<i>Hyparrhenia dissoluta</i>	0,154821	2,39695008	6,5
<i>Botriochloa insculpta</i>	0,14259358	2,033293	4,9	<i>Acacia senegal A</i>	-0,151417	2,29272269	22
<i>Eustachys paspaloides</i>	0,13698305	1,8764357	1,4	S S-3	-0,150782	2,27351545	0,2
pH P-3	0,13675553	1,8702075	2,1	V P-3	-0,149531	2,23595068	2,2
<i>Markhamia obtusifolia J</i>	-0,1319613	1,7413775	64,4	<i>Acacia senegal A >1</i>	-0,148442	2,20349482	6,5
Milieu ouvert	0,12596826	1,5868002	1,7	<i>Combretum molle J</i>	0,14687	2,15708062	12,6
<i>Grewia trichocarpa A</i>	-0,1251961	1,5674072	13,2	<i>Macrotyloma maraguense</i>	0,138324	1,91334565	4
<i>Acacia senegal A</i>	0,12447561	1,5494178	28,1	<i>Brachiaria semiundulata</i>	0,13707	1,87881728	5,8
pH S-4	0,12356549	1,5268431	3,6	rét. eau S-3	-0,132034	1,74330344	2,9
<i>Cyanotis lanata</i>	-0,1214868	1,4759042	7,8	<i>Dichrostachys cinerea J >1</i>	-0,128697	1,65628157	55,5
Sp et par. mésologiques	coord3	CR%3	P3	Sp et par. mésologiques	coord4	CR%4	P4
<i>Acacia hockii J</i>	0,22621717	5,1174209	22,3	<i>Acacia polyacantha A</i>	0,245719	6,03777565	0
<i>Ormocarpon trichocarpum J</i>	0,20461047	4,1865443	23,2	<i>Acacia hockii A</i>	0,211422	4,46992869	14,7
<i>Lannea humilis J</i>	0,19413655	3,7689001	28,5	pH S-3	0,187841	3,52843639	0,1
Milieu ouvert	-0,1893768	3,5863567	0	<i>Acacia polyacantha A >3</i>	0,169247	2,86444536	0
<i>Albucca kirkii</i>	-0,1558881	2,4301108	5,4	humus-1	-0,167775	2,81486162	0,5
<i>Acacia hockii A</i>	0,14487661	2,0989233	35,9	<i>Acacia senegal A</i>	-0,164452	2,70445098	15,7
<i>Asparagus abyssinicus</i>	0,13773398	1,897065	4,6	MO-1	-0,157241	2,47247237	0,7
<i>Loudetia simplex</i>	-0,1368173	1,8718973	0,1	<i>Albucca kirkii</i>	-0,155844	2,42872145	6,2
S S-2	0,13230747	1,7505267	4	<i>Acacia hockii J</i>	0,153471	2,35533991	33,4
<i>Blepharis cristata</i>	0,12692152	1,6109072	3,4	<i>Sporobolus stapfianus</i>	-0,151996	2,31028044	4,6
V S-4	-0,1232273	1,5184971	0,7	<i>Setaria barbata</i>	0,14058	1,97626066	5
<i>Loudetia simplex >1</i>	-0,1224702	1,4998949	0	<i>Acacia senegal A >1</i>	-0,137861	1,90056432	12,2
V P-2	0,12073345	1,4576565	3,3	<i>Dichrostachys cinerea J</i>	-0,135306	1,83077249	14,2
V S-3	0,1199625	1,4391002	7,4	<i>Acacia brevispica A</i>	-0,13352	1,78276564	34,8
<i>Parinari curatellifolia J</i>	-0,1190851	1,418125	0	V S-3	0,132331	1,75115373	4,6

Tableau 16. Quinze premières espèces et paramètres mésologiques (caractères gras) rangés par ordre décroissant des contributions relatives sur les axes 1 à 4.

Sur l'axe 1, on trouve ensemble plusieurs espèces de *Poaceae* comme *Pennisetum polystachion*, *Botriochloa insculpta*, *Heteropogon contortus*, *Eustachys paspaloides*, avec *Hibiscus cannabinus*, les jeunes individus de *Lannea humilis* et *Acacia senegal* adulte, le tout associé à un milieu ouvert, une teneur en cations échangeables de niveau 2 en profondeur, un pH de niveau 3 en profondeur et de niveau 4 en profondeur. Le tout est opposé à *Acacia brevispica* jeune, *Markhamia obtusifolia* jeune, *Grewia trichocarpa* adulte (espèces de bosquets) et *Cyanotis lanata*.

Sur l'axe 2, sont réunis *Achyranthes aspera var. sicula*, *Panicum maximum >1*, *Commelina benghalensis*, *Acacia senegal* adulte présent et >1 , *Dichrostachys cinerea* jeune, le tout associé à une teneur en cations échangeables de niveau 3 en surface, un pourcentage de saturation en bases de niveau 3 en profondeur, une capacité au champ de niveau 3 en surface. A l'opposé, on trouve *Combretum molle* jeune, *Spermacoce pusilla*, *Hyparrhenia dissoluta*, *Brachiaria semiundulata* et *Macrotyloma maraguense*.

Sur l'axe 3, on oppose les milieux les plus ouverts avec *Albucca kirkii*, *Loudetia simplex* et *Parinari curatellifolia* à la savane à *Acacia hockii* avec des concentrations en cations échangeables de niveau 2 en surface, un taux de saturations en bases de niveau 3 en surface et de niveau 2 en profondeur.

Sur l'axe 4, on voit clairement que la galerie forestière à *Acacia polyacantha* et quelques relevés proches se trouvent sur sols riches en matières organiques et minérales, bien structurés au bord et à proximité du lac. Le tout est opposé à divers milieux plus ouverts.

Analyse statistique des données de végétation



Figure 37. Analyse factorielle multiple des fichiers Akagera floristique et environnemental disjonctifs réunis. Coordonnées des espèces et paramètres méso-climatiques réunis sur les axes 1 et 2.

Cet exemple d'analyse combinant des variables floristiques et environnementales montre l'intérêt de transformer d'abord les tableaux bruts en tableaux disjonctifs.

Discussion et conclusions

Cet exemple développe principalement une démarche expliquant comment on obtient progressivement des résultats interprétables en adaptant les données et en s'efforçant de trouver le meilleur accord entre les données et les analyses multivariées. Il apparaît clairement qu'il n'est jamais superflu d'examiner toutes les étapes de calcul des analyses multivariées de manière à optimiser leur utilisation : tableau transformé pour PCA, CA ou NSCA, matrice

Analyse statistique des données de végétation

de corrélation, de covariance ou de distance pour toutes les techniques, valeurs propres et leur caractère significatif, coordonnées des lignes et colonnes sur les axes. En partant des données présentées et de beaucoup d'autres, je peux conclure que la forme de la distribution des coefficients de corrélation de distance est une étape essentielle pour décider si l'analyse multivariée est bien adaptée aux données. Une dissymétrie importante, avec une très faible proportion de coefficients extrêmes est une cause de non-utilisation puisqu'une très petite partie du tableau de données est mise en évidence. Une symétrie parfaite avec une distribution gaussienne des coefficients est une situation idéale, spécialement quand la proportion de valeurs centrales de la distribution n'est pas trop élevée (moins de 20 %). Avec des données de végétation, PCA et CA souffrent de cette dissymétrie, en partie parce que la proportion de cases vides du tableau de données est importante, ce qui est une situation commune. Avec NSCA d'un tableau disjonctif simple l'asymétrie est moins prononcée qu'avec CA. Dans les tableaux de distance ou de similitude, la distribution des coefficients est souvent symétrique, mais parfois avec une faible proportion de valeurs extrêmes. En plus, PCA et plus spécialement CA sont très sensibles aux motifs spécifiques locaux ; la présence de plusieurs espèces rares dans un ou dans un petit nombre de relevés rend ces techniques pratiquement inutilisables. Les performances de PCA ou CA sont aussi souvent associées à des relations linéaires ou unimodales entre les variables (ESPINAR *et al.*, 2002), ce qui n'est pas très commun avec des données de végétation, particulièrement à cause des techniques d'échantillonnage qui consistent en relevés séparés par des espaces relativement grands.

L'utilisation de tests par permutation est un complément très appréciable. Je recommande d'abord de ranger les espèces en ordre descendant de contributions relatives. Ensuite, quand un test par permutation est associé aux contributions relatives, on ajoute une dimension inférentielle à l'interprétation.

Voici maintenant les conclusions plus précises sur chacune des analyses présentées :

Une analyse en composantes principales de données brutes n'apporte que des résultats assez pauvres, dominés par des espèces localisées. Une transformation en données ordinales n'améliore pas les résultats.

Une analyse des correspondances de données ordinales donne des résultats encore moins bons que l'analyse en composantes principales en exacerbant les variations locales par rapport aux formations végétales plus importantes qui sont nettement écrasées dans l'analyse.

L'analyse des principales coordonnées de données ordinales apporte une meilleure dispersion des relevés et mais ce sont surtout les données extrêmes qui sont mises en évidence et l'interprétation n'est pas aisée. Seules les coordonnées des relevés sont fournies. En plus, il n'y a pas d'aide à l'interprétation en dehors des valeurs propres. Le type de distance a peu d'incidence sur l'interprétation.

Les analyses non-métriques multidimensionnelles de données ordinales produisent aussi une bonne dispersion des relevés mais ne produisent encore une fois que les coordonnées des relevés et aucune aide à l'interprétation. En plus, le choix, au départ du nombre d'axes à extraire, a une incidence sur les résultats, ce qui rend l'interprétation des résultats assez délicate.

Avec l'analyse d'un tableau disjonctif simple, l'analyse des correspondances s'avère totalement inadaptée et elle ne devrait plus être utilisée. Il en va de même pour les variantes créées comme CCA ou DCA, ... qui souffrent aussi du même défaut de base. Elles sont vivement déconseillées.

L'analyse non symétrique des correspondances du tableau disjonctif simple produit une belle vue de l'ensemble du transect et ensuite des variations plus locales. Les aides à l'interprétation sont utiles. Le caractère synthétique de l'analyse est certain avec seulement les cinq premières valeurs propres significatives.

Analyse statistique des données de végétation

L'analyse des principales coordonnées produit une bonne dispersion des relevés assez facile à interpréter sur les premiers axes mais n'apporte pas de coordonnées pour les espèces ni d'aide à l'interprétation.

Une analyse non symétrique des correspondances d'un tableau disjonctif de type 113 produit aussi de bons résultats assez faciles à interpréter comme ceux du tableau disjonctif simple.

L'analyse non symétrique des correspondances des données d'abondance en version logarithmique privilégie toujours les espèces abondantes, mais de façon atténuée par rapport à une version non transformée. Il y a donc un choix à opérer si l'on veut privilégier les abondances sans leur donner un poids trop lourd. Dans une telle analyse, il n'y a qu'une seule ligne par espèce.

L'analyse factorielle multiple d'un tableau disjonctif, avec trois sous-tableaux, dont un pour les espèces ligneuses, favorise nettement ces dernières. Il y a donc un choix à opérer si l'on souhaite mettre en évidence une strate ou une catégorie d'espèces. La même analyse avec le fichier 113 apporte aussi des résultats faciles à interpréter et c'est à nouveau un choix à faire entre les deux types de tableau disjonctif, le dernier donnant un peu plus d'importance aux simples présences par rapport aux abondances. Avec le fichier logarithmique, les espèces abondantes sont à nouveau privilégiées, mais de manière moins nette qu'avec les abondances 12345.

Si l'on souhaite réunir les tableaux floristiques et mésologiques, une analyse factorielle multiple basée sur l'analyse non symétrique des correspondances de tableaux disjonctifs apparaît comme une technique très puissante pour comprendre les parallélismes entre espèces et facteurs mésologiques, sans préjuger du type de relation entre ces paramètres.

Le problème des analyses multivariées est de trouver une pondération adéquate pour les sous-tableaux. Avec des tableaux disjonctifs, la proportion de présences de chaque sous-tableau a été choisie. Dans ce cas, le fait d'avoir une ou plusieurs colonnes vides dans un sous-tableau n'empêche pas l'analyse, comme c'est le cas avec le choix des premières valeurs propres comme paramètres de pondération.

Nous recommandons fortement l'analyse non symétrique des correspondances. Sa principale qualité est sa faible sensibilité aux espèces rares ; ainsi un tableau complet, avec toutes les espèces enregistrées, est analysable. Avec une échelle ordinale quelconque ou une échelle de Braun-Blanquet en particulier, je recommande toujours de transformer les données afin d'éviter que le poids du tableau ne soit concentré dans un petit nombre d'espèces abondantes. Le plaidoyer pour l'utilisation de tableaux disjonctifs n'est ni neuf ni original. BOUROCHE et SAPORTA (1980) le soulignaient en ces termes : " La mise sous forme disjonctive est bien plus qu'une commodité mathématique et cela pour plusieurs raisons. Puisqu'un caractère numérique peut être transformé en un caractère qualitatif par découpage en classes de ses valeurs, il est possible d'étudier des tableaux comportant un mélange de caractères numériques et qualitatifs : il suffit de tout rendre qualitatif et d'effectuer une analyse des correspondances multiples. L'intérêt est qu'en procédant ainsi on peut prendre en compte des liaisons non linéaires entre caractères".

Notre analyse du fichier Akagera ne prend en compte aucun paradigme *a priori*. Le paradigme du gradient aurait pu être favorisé comme le transect étudié part du bord d'un lac jusqu'à une colline sèche et rocheuse mais il ne semble pas possible de réduire cette végétation à un ou plusieurs gradients simples.

janvier 2004

Références

- AUSTIN, M.P. 1977. Use of ordination and other multivariate descriptive methods to study succession. *Vegetatio* 35: 165-175.
- AUSTIN, M.P. 2012. *Vegetation and environment: discontinuities and continuities*. In *Vegetation Ecology*, second edition. Edited by Eddy van der Maarel and Janet Franklin. Wiley-Blackwell.
- AUSTIN, M.P. 2013. Inconsistencies between theory and methodology: a recurrent problem in ordination studies. *Journal of Vegetation Science* 24: 251-268.
- 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.
- 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. *Oecologia Plantarum* 13: 121-138.
- BONIN, G., AUBERT, G., BARBERO, M., GAMISANS, J., GRUBER, M., LOISEL, R., QUEZEL, P., SANDOZ, H., THINON, M. & VEDRENNE, 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.
- BOUROUCHE, J.-M. & SAPORTA, G. 1980. *L'analyse des données*. P.U.F. Collection Que sais-je ?
- BOUXIN, G. 1975a. Ordination and classification in the savanna vegetation of the Akagera park (Rwanda, Central Africa). *Vegetatio* 29: 155-167.
- BOUXIN, G. 1975b. Ordination of quantitative and qualitative data in a savanna vegetation (Rwanda, Central Africa). *Vegetatio* 30:197-200.
- BOUXIN, G. 1983. Multi-scaled pattern analysis: an example with savannah 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. 1995. Démarche conduisant à la mise au point d'une technique de description de la végétation. *Acta Botanica Gallica* 142: 533-540.
- BOUXIN, G. 1999. Description de la végétation aquatique et du bord de l'eau dans le bassin hydrographique de la Molinee (Condroz, Belgique) par l'analyse multiscalaire des motifs. *Ecologie*, t. 30: 139-163.
- BOUXIN, G. 2011. Évolution de la végétation macrophytique et trophie dans deux ruisseaux du bassin hydrographique de la Molinee (Condroz, Belgique). *Revue des Sciences de l'Eau* 24: 253-266.
- BOUXIN, G. 2013. Végétation macrophytique, environnement et qualité d'eau dans le bassin versant du Bocq (Belgique, Wallonie). *Revue des Sciences de l'Eau* 26: 1-19.
- 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.

Analyse statistique des données de végétation

- 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.
- DRAY, S., DUFOUR, A.B. & CHESSEL, D. 2007. The ade4 package - II: Two-table and K-table methods. *R News* 7(2): 47-52.
- 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. 179-191. Diday, E. & Coll. (Eds.) Elsevier, North-Holland.
- ESCOFIER, B. & PAGES, J. 1989. *Multiple factor analysis: results of a three-year utilization*. In: *Multway data analysis*. 277-285. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland.
- ESPINAR, J.L., GARCÍA, L.V., GARCÍA MURILLO, P. & TOJA, J. 2002. Submerged macrophyte zonation in a Mediterranean salt marsh: a facilitation effect from established helophytes? *Journal of Vegetation Science* 13: 831-840.
- FEHMI, J. 2010. Confusion among three common plant cover definitions may result in data unsuited for comparison. *Journal of Vegetation Science* 21: 273-279.
- FOUCART, T. 1982. *Analyse factorielle. Programmation sur micro-ordinateurs*. Masson, Paris.
- GAUCH, H.G., Jr. 1982. *Multivariate analysis in community ecology*. Cambridge University Press.
- GAUTHIER, S., BOUCHER, D., MORISSETTE, J. & DE GRANDPRÉ, L. 2010. Fifty-seven years of composition change in the eastern boreal forest of Canada. *Journal of Vegetation Science* 21: 772-785.
- GIMARET-CARPENTIER, C., CHESSEL, D. & PASCAL, J.-P. 1998. Non-symmetrical 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.
- GREENACRE, M. 2013. Fuzzy coding in constrained ordinations. *Ecology* 94: 280-286.
- GREIG-SMITH, P. 1964. *Quantitative plant ecology (second edition)*. Butterworths, London.
- HAURY, J. & BAGLINIERE, J.-L. 1990. Relations entre la population de truites communes (*Salmo trutta*), les macrophytes et les paramètres du milieu sur un ruisseau. *Bull. Fr. Pêche Piscic.* 318: 118-131.
- HUSSON, F., LÊ, S. ET PAGÈS, J. 2009. *Analyse des données avec R*. Presses Universitaires de Rennes.
- KNOX, R.G. & PEET, R.K. 1989. Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. *Vegetatio* 80: 153-165.
- 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.
- MANLY, B.F.J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Second edition. Chapman & Hall.
- MCCUNE, B & MEFFORD, M.J. 1999. *PC-Ord. Multivariate analysis of ecological data, version 4*. MjM Software Design, Gleneden Beach, OR, US.

Analyse statistique des données de végétation

- MINCHIN, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89-107.
- ORLÓCI, L. & KENKEL, N.C. 1985. Introduction to data analysis, with examples from population and community ecology. International Co-operative Publishing House. *Statistical Ecology Monographs*, 1: 1-339.
- PALM, R. 1993. Les méthodes d'analyse factorielle : principes et applications. *Notes Stat. Inform.* (Gembloux) 93/1.
- PIELOU, E.C. 1977. *Mathematical Ecology*. Wiley-Interscience.
- PILLAR, V.D.P. 1999. The bootstrapped ordination re-examined. *Journal of Vegetation Science* 10: 895-902.
- PODANI, J. 2006. Braun-Blanquet's legacy and data analysis in vegetation science. *Journal of Vegetation Science* 17: 113-117.
- 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.
- ROOT, H.T. & NELSON, P.R. 2011. Does phylogenetic distance aid in detecting environmental gradients related to species composition? *Journal of Vegetation Science* 21: 1143-1148.
- STEVENS, C., DUPRÉ, C., GAUDNIK, C., DORLAND, E., DISE, N., GOWING, D., BLEEKER, A., ALARD, D., BOBBINK, R., FOWLER, D., VANDVIK, V., CORCKET, E., MOUNTFORD, J.O., AARRESTAD, P.A., MULLER, S. & DIEKMANN, M. 2011. Changes in species composition of European acid grasslands observed along a gradient of nitrogen deposition. *Journal of Vegetation Science* 22: 207-215.
- TALBOT, J., RICHARD, P.J.H., ROULET, N.T. & BOOTH, R.K. 2010. Assessing long-term hydrological and ecological responses to drainage in a raised bog using paleoecology and a hydrosequence. *Journal of Vegetation Science* 21: 143-156.
- TROUPIN, G. 1982. *Flore des plantes ligneuses du Rwanda*. Institut national de la Recherche scientifique, Butare, République Rwandaise. Publication n° 21.
- TROUPIN, G. 1983. *Flore du Rwanda. Spermatophytes, Volume II*. Musée royal de l'Afrique centrale – Tervuren, Belgique. Annales-Série IN-8°-Sciences économiques – n°13.
- TROUPIN, G. 1985. *Flore du Rwanda. Spermatophytes, volume III*. Institut national de la Recherche scientifique, Butare. Publication n° 30.
- 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.
- 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.
- WILDI, O. 2010. *Data analysis in vegetation ecology*. Wiley-Blackwell. 211 pp.
- WILDI, O. 2013. *Data analysis in vegetation ecology* (second edition). Wiley-Blackwell. 301 pp.
- WILSON, J.B. 2011. Cover plus: ways of measuring plant canopies and the terms used for them. *Journal of Vegetation Science* 22: 197-206.