

## Phytosociologie et analyse des données

**Guy BOUXIN**  
B-5101 ERPENT  
guy.bouxin@proximus.be

**Résumé.** Un tableau de relevés de parois calcaires, composé de 27 relevés et 42 taxons, est soumis à diverses analyses multivariées. L'analyse non symétrique des correspondances d'un tableau disjonctif simplifié assure la meilleure adéquation entre techniques statistiques et données de végétation. La classification est élaborée à partir de variables transformées issues de l'analyse multivariée. Les espèces caractéristiques ou différentielles sont définies au moyen d'un test par permutation de données de présence. Les groupements végétaux ainsi obtenus sont comparés avec les associations végétales préalablement décrites. Les différences entre les deux approches sont discutées, ainsi que des perspectives futures pour la phytosociologie.

**Mots clés :** analyse non symétrique des correspondances, classification, phytosociologie, espèces caractéristiques et différentielles.

**Abstract.** A relevé table of limestone wall, compound of 27 relevés and 42 taxa, is submitted to various multivariate analyses. The non-symmetrical correspondence analysis of a simplified disjunctive table ensures the best adequacy between statistical techniques and vegetation data. The classification is elaborated from transformed variables resulting from the multivariate analysis. The character or differential species are defined with a test by permutation of presence data. The plant groupings thus obtained are compared with the plant associations previously described. The differences between the two approaches are discussed, as well as future perspectives for phytosociology.

**Keywords :** non-symmetrical correspondence analysis, classification, phytosociology, character- and differential species.

## Introduction

Cette étude consiste à comparer les résultats de l'analyse statistique d'un tableau de relevés de végétation avec ceux provenant de l'analyse fournie par un phytosociologue (Royer, 2009) et de voir si cette analyse statistique apporte un réel avantage. Dans un relevé, chaque espèce reçoit un coefficient d'abondance-dominance (i, r, +, 1, 2, 3, 4, 5), un coefficient de sociabilité, parfois un coefficient de vitalité et quelques autres indications comme la localisation spatio-temporelle du relevé, le recouvrement de la végétation, la surface totale du relevé, ainsi que quelques indications mésologiques (pédologie, géologie, conditions hydriques) souvent très peu nombreuses. Le tableau de relevés présente traditionnellement les espèces en lignes et les relevés en colonnes. La manipulation des tableaux s'est d'abord faite en déplaçant simplement les lignes et colonnes, ce qui devient vite fastidieux avec de grands tableaux. Dès que les outils informatiques l'ont permis, l'analyse statistique s'est progressivement développée avec des techniques reprises sous le nom générique d'analyses factorielles, qui elles-mêmes font partie des analyses multivariées. Mais une confusion est apparue dès le début de l'utilisation de l'outil statistique avec le terme « ordination » qui a pris la place, chez beaucoup d'utilisateurs, de l'analyse statistique rigoureuse. 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. Ce processus synthétique est considéré comme un outil pour mettre les données floristiques en relation avec des 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 a écrit (traduction) : « L'analyse de gradient étudie la végétation en termes de gradient (ou variation plus ou moins continue) 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) ».

Selon Bourouche et Saporta (1980), l'analyse des données regroupe des méthodes statistiques aussi nombreuses que différentes. On distingue deux approches principales : la classification automatique, qui classe des unités statistiques et des variables sur base d'algorithmes préalablement définis, et l'analyse factorielle, qui utilise les propriétés de l'espace vectoriel euclidien pour décrire des individus ou des variables (Foucard, 1982). Selon Palm (1993), les méthodes d'analyse factorielle sont des outils fondamentaux pour analyser des tableaux de données qui n'ont pas de structure particulière. Ils ont essentiellement un but descriptif en condensant l'information contenue dans une table, composée souvent d'un grand nombre de rangées, en un petit nombre de représentations graphiques, accompagnées de tables contenant des valeurs numériques correspondant à des paramètres qui sont des aides à l'interprétation.

L'analyse factorielle, dans son acception 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. Le terme « ordination » a été et est actuellement toujours utilisé en lieu et place d'analyse factorielle (Wildi, 2013).

Parmi de nombreux exemples d'utilisation des analyses multivariées dans des études phytosociologiques, citons les travaux assez récents de Guyonneau (2005) avec l'analyse des correspondances, Bailly (2012), avec la même analyse des correspondances et la classification hiérarchique ascendante de données de présence (groupement ascendant du lien complet avec la distance de Bray-Curtis) ou des abondances-dominances (méthode de Ward avec la distance corde). Collaud et collaborateurs (2010) utilisent aussi l'analyse des correspondances et la classification hiérarchique ascendante en abondance-dominance (méthode de

Ward et distance corde) ainsi que Ferrey (2007, 2009 et 2011) avec la même technique de classification.

Les résultats sont généralement présentés sous forme de graphes construits à partir des coordonnées des espèces ou des relevés (parfois les deux en même temps) dans le plan des axes 1 et 2, 1 et 3 ou encore 2 et 3. Toutefois, les résultats des analyses multivariées ne sont pas toujours à la hauteur de ce que ces outils mathématiquement élaborés laissent espérer. Les raisons de ces échecs sont multiples :

- les algorithmes utilisés sont souvent mal connus ;
- la structure particulière du tableau phytosociologique, avec quelques espèces abondantes et fréquentes, quelques espèces moins fréquentes et beaucoup d'autres peu représentées, a souvent de lourdes conséquences sur les résultats, à cause du trop grand nombre de cellules vides. Les analyses multivariées n'ont pas été élaborées pour de telles données mais pour des tableaux presque totalement ou complètement remplis ;
- un modèle sous-jacent associé systématiquement à l'analyse, comme l'analyse de gradient, n'est pas nécessairement compatible avec la structure des données.

L'objectif de ce texte est de montrer comment on peut arriver à une bonne utilisation des techniques statistiques avec des tableaux de relevés. Pour illustrer la démarche, nous utilisons un petit tableau publié par Royer (2009).

## Le tableau de relevés et l'analyse du phytosociologue

Les données sont présentées dans le tableau ci-dessous. Seules les abondances sont reprises dans cette étude. La sociabilité n'est de toute manière jamais prise en compte dans les analyses numériques. La nomenclature a été adaptée à celle de *Flora Gallica* (Tison et de Foucault, 2014).

Espèces	Relevés	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		+	.	1	.	2	+	1	+	1	+	3	.	1	+	.	.	+	.	1	1	+	1	.	.	1	1	.
<i>Potentilla verna</i>		1	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	+	.	.	.	.	+
<i>Asplenium ceterach</i>		2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
<i>Helleborus foetidus</i>		+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Poterium sanguisorba</i>		+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Bromopsis erecta</i>		1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		.	2	.	.	.	2	+	.	.	.	.	1	.	.	.	1	.	1	.	.	.	.	1	.	.	.	.
<i>Poa nemoralis</i>		.	+	.	.	+	+	1	.	+	.	2	.	1	.	.	.	.	.	.	.	.	.	.	+	.	.	.
<i>Geranium robertianum</i>		.	+	+	.	.	.	.	+	.	1	1	.	1	+	.	+	.	.	.	1	1	.	.	.	.	.	.
<i>Asplenium scolopendrium</i>		.	.	2	.	.	.	.	1	.	2	.	.	1	1	.	.	.	.	.	.	1	.	.	+	1	.	.
<i>Cystopteris fragilis</i>		.	.	1	.	.	+	.	+	.	.	.	.	1	2	.	+	.	+	.	+	+	.	.	1	+	.	.
<i>Cardamine impatiens</i>		.	.	+	.	.	.	.	.	+	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	+	.	.
<i>Hedera helix</i>		.	.	+	.	.	.	1	+	.	+	.	1	.	.	.	.	.	+	.	.	.	+	1	.	.	+	.
<i>Moehringia trinervia</i>		.	.	+	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	+	.	.	+	.	.
<i>Vinca minor</i>		.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Asplenium trichomanes</i> subsp. <i>pachyrachis</i>		.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1
<i>Asplenium ruta-muraria</i>		.	.	.	+	1	.	+	.	.	.	+	+	.	.	+	.	+	.	+	1	.	1	+	1	.	1	+
<i>Lactuca muralis</i>		.	.	.	.	+	.	.	.	1	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i>		.	.	.	1	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	+	+	1	.	.	1	+	.
<i>Campanula rotundifolia</i>		.	.	.	.	+	.	+	.	.	.	.	.	.	.	+	.	+	.	+	.	1	.	.	.	.	+	.
<i>Sesleria caerulea</i>		.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.
<i>Lamium galeobdolon</i>		.	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.
<i>Mercurialis perennis</i>		.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Oxalis acetosella</i>		.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	+	+	.	.
<i>Campanula trachelium</i>		.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Galium aparine</i>		.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Chrysosplenium alternifolium</i>		.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Arabis hirsuta</i>		.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Origanum vulgare</i>		.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Thymus praecox</i> s.l.		.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>Ribes uva-crispa</i>		.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+
<i>Polypodium interjectum</i>		.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	+	.	.	.	.	.
<i>Fragaria vesca</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.	.	.	.	.
<i>Viola alba</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
<i>Silene nutans</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.
<i>Potentilla micrantha</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.	.
<i>Poa compressa</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.	.	.
<i>Melica uniflora</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.
<i>Clematis vitalba</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.
<i>Phedimus spurius</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.	.	.
<i>Ribes alpinum</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	.
<i>Sedum album</i>		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+

**Tableau 1.** Les 27 relevés de parois calcaires

L'approche méthodologique est basée sur le concept de syntaxon élémentaire développé par de Foucault (1984). Un syntaxon élémentaire regroupe un ensemble de relevés dans une table homotone. Les syntaxons élémentaires constituent les éléments élémentaires pour l'élaboration de la typologie phytosociologique et ne doivent pas être confondus avec les associations.

Cinq syntaxons élémentaires ont été identifiés : le syntaxon A avec les relevés 4, 15, 27, le B avec les relevés 1 et 22, le syntaxon C avec les relevés 5, 7, 9, 11, 17, 19 et 26, le syntaxon D avec les relevés 3, 8, 10, 13, 14, 20, 21, 24, 25 et le syntaxon E avec les relevés 2, 6, 12, 16, 18, 23. Les espèces constantes du syntaxon A sont *Asplenium trichomanes* subsp. *pachyrachis* et *Asplenium ruta-muraria*, les espèces différentielles positives sont *Asplenium trichomanes* subsp. *pachyrachis*, *Potentilla verna* (cette dernière par rapport aux syntaxons C, D et E) et les espèces différentielles négatives *Asplenium trichomanes* subsp. *quadrivalens* et *Hedera helix*. Le syntaxon D possède davantage d'espèces constantes : *Asplenium scolopendrium*, *Cystopteris fragilis*, *Geranium robertianum*, *Asplenium trichomanes* subsp. *quadrivalens*. Ses espèces différentielles positives sont *Cardamine impatiens*, *Polypodium interjectum*, *Moehringia trinervia*, *Lamium galeobdolon*, *Oxalis acetosella*, *Cystopteris fragilis* (ce dernier par rapport à A-B-C mais pas par rapport à E), ses différentielles négatives *Asplenium trichomanes* subsp. *pachyrachis*

(par rapport à A), *Asplenium trichomanes* subsp. *hastatum* (par rapport à E). Grâce aux données de la littérature, le syntaxon A est rattaché à l'*Asplenietum pachyrachidis*, le syntaxon B à l'*Asplenietum trichomano-rutae-murariae ceterachetosum officinarum*, le syntaxon C à l'*Asplenietum trichomano-rutae-murariae cardaminopsietosum borbassii*, le syntaxon D au *Cystopterido fragilis-Phyllitidetum scolopendrii* et le syntaxon E à l'*Asplenietum hastati*.

## L'analyse des données

Nous n'utilisons ici que les techniques issues de l'analyse des données puisque que nous avons suffisamment montré que les autres étaient mal adaptées aux données de végétation (Bouxin, 2016 et 2024). Deux techniques principales sont considérées : l'analyse en composantes principales (ACP) et l'analyse des correspondances (AC). Elles opèrent globalement de la même manière en passant par plusieurs étapes :

- le calcul d'une matrice transformée,
- le calcul d'une matrice de corrélation ou de distance (*sensu lato*),
- le calcul de valeurs et vecteurs propres,
- le calcul d'un nombre limité de variables pour les lignes et colonnes du tableau, ainsi que des paramètres associés utilisés comme aides à l'interprétation. Les variables sont utilisées comme coordonnées dans les représentations graphiques.

Dans l'analyse en composantes principales, les variables originelles de chaque ligne (correspondant aux espèces) sont transformées en variables centrées réduites et les corrélations sont calculées à partir de ces variables réduites. Ce calcul des variables centrées réduites se fait ligne par ligne, indépendamment les unes des autres. Cette analyse est ainsi bien adaptée à des données d'abondance et il semblerait logique de la privilégier. Un premier problème vient toutefois de la nature des données représentées par des variables ordinales, avec lesquelles les calculs de moyennes ou de variances ont une signification limitée. L'analyse des correspondances n'est pas adaptée à des variables continues ou ordinales mais bien à des présences ou des fréquences. Elle est néanmoins utilisée très fréquemment avec de telles données, ce qui génère très souvent des déceptions. En plus, le calcul des variables transformées prend en compte des sommes sur les lignes et les colonnes, ce qui engendre des cotes particulièrement élevées pour les espèces peu fréquentes. Comme un tableau phytosociologique comprend en général quelques espèces abondantes et fréquentes, quelques espèces moins fréquentes et une proportion souvent assez importante d'espèces peu communes ou une seule fois présentes, l'analyse des correspondances génère des résultats souvent pauvres en informations.

Un double choix s'impose donc, celui de la transformation des données pour les mettre en adéquation avec une technique statistique et le choix même de la technique. En effet, tous les coefficients d'abondance-dominance ne peuvent être utilisés tels quels dans des opérations mathématiques. Il faut donc remplacer les coefficients comme +, r ou i par une valeur numérique. Ces derniers sont arbitrairement remplacés par 0,1 ou 0,5 (0,1 dans ce texte).

Une fois faite la transformation vers le "tout numérique", les analyses en composantes principales ou des correspondances sont directement utilisables. Mais d'autres transformations sont possibles :

- un tableau de présence, avec uniquement des 0 et 1 ; il y a autant de lignes que d'espèces ;
- un tableau d'abondance avec des coefficients 0,1 - 1 - 2 - 3 - 4 et 5 ; il y a aussi autant de lignes que d'espèces ;
- un tableau disjonctif complet, avec pour chaque espèce une ligne pour les absences (les absences sont notées 1 et les autres coefficients sont notés 0) et autant de lignes qu'il y a de coefficients d'abondance différents; les coefficients correspondant à la ligne des "1" sont notés 1 et les autres sont notés 0 et ainsi de suite pour les autres coefficients ; les sommes sur les colonnes sont toutes identiques; le nombre de lignes est fortement augmenté ;
- un tableau disjonctif simple, avec la seule différence qu'il n'y a pas de ligne pour les absences ;
- un tableau disjonctif simplifié (dit  $1 > 1 > 3$ ), obtenu par exemple en groupant les lignes +, r, i et 1, en groupant les lignes 2 et 3 et en groupant les lignes 4 et 5 ; pour chaque espèce, il y a trois lignes maximum ; ceci limite le nombre de lignes ; ce choix est purement empirique mais a produit des résultats utiles et est régulièrement utilisé dans nos analyses. Dans le tableau utilisé, il n'y a aucune abondance supérieure à 3. Dans les lignes correspondant aux abondances 2 et 3, les noms spécifiques sont suivis du signe ">1".

L'avantage des tableaux de présence et des tableaux disjonctifs est certain puisqu'il n'utilise que des données 0 et 1 et les analyses ne sont pas dominées par quelques cotes beaucoup plus élevées que les autres.

Une variante de CA, appelée analyse non symétrique des correspondances (ANSC), a été proposée par Gimaret-Carpentier *et al.* (1998) et utilisée notamment par Kroonenberg et Lombardo (1999) et résout le problème des espèces rares. Pour comprendre son intérêt, il faut considérer la transformation des données :

- pour CA, on calcule :  $\text{don1}(I, J) = \text{don}(I, J) \times \text{somme} / (\text{hc}(J) \times \text{hr}(I)) - 1$ ,
- pour NSCA, on calcule :  $\text{don1}(I, J) = \text{don}(I, J) / \text{hc}(J) - \text{hr}(I) / \text{somme}$ .

On voit tout de suite qu'avec CA les données transformées prennent une cote élevée si les variables hc et hr sont peu élevées, tout particulièrement si hr est petit. On calcule ensuite la matrice de distance comme suit :

- pour CA,  $y1(J1, J2) = y1(J1, J2) + \text{don1}(I, J1) \times \text{don1}(I, J2) \times \text{hr}(I) \times (\text{hc}(J1) \times \text{hc}(J2))^2$
- pour NSCA,  $y1(J1, J2) = y1(J1, J2) + \text{don1}(I, J1) \times \text{don1}(I, J2) \times (\text{hc}(J1) \times \text{hc}(J2))^2$ .

Avec don() : données de départ (par exemple, présence d'une espèce dans un relevé), don1() : données transformées avant calcul de la matrice de distance, somme : somme totale sur toutes les lignes et colonnes, hc() : somme sur une colonne, hr() : somme sur un ligne, y1() : distance entre colonnes.

Les valeurs propres sont ensuite calculées à partir de cette matrice de distance. L'incidence des quelques cotes très élevées obtenues par les espèces rares dans l'analyse des correspondances est très forte et ne met vraiment en évidence que les espèces rares. Avec l'analyse non symétrique, ce sont les espèces les plus fréquentes qui ressortent en premier lieu.

L'inférence statistique est rarement utilisée dans les analyses multivariées. Des techniques récentes facilitent le passage à cette technique : ce sont les tests de randomization, de bootstrap et de Monte Carlo (Manly, 1997). Les tests de randomization ont été appliqués par Bouxin (1999) aux analyses en composantes principales et aux analyses des correspondances. La technique de permutation porte sur les lignes (correspondant aux espèces) et est appliquée à chaque ligne indépendamment. Dans chaque ligne, les contenus sont permutés de manière aléatoire et une nouvelle analyse multivariée est calculée. Le programme note si les valeurs propres et les contributions relatives de chaque variable de cette seconde analyse sont supérieures ou égales à celle

de la première analyse. La procédure est répétée un grand nombre de fois (en fonction de la taille du fichier, soit idéalement 10 000 mais souvent moins). Sont significatifs seulement les axes dont la proportion de valeurs propres, obtenues après permutations et supérieures ou égales aux valeurs propres des axes correspondant du tableau original est inférieure à 0,05 (risque de première espèce). On fixe ainsi un nombre de valeurs propres significatives et pour chaque axe avec les espèces ou autres variables avec une contribution relative significative. Ce sont ces espèces ou autres variables qui doivent retenir l'attention en premier lieu.

Pour chaque analyse, les résultats sont réunis dans un même tableau comprenant :

- une première ligne avec les valeurs propres, les pourcentages par rapport à la trace de la matrice de corrélation ou de distance et la probabilité associée à cette valeur propre et cela pour tous les axes retenus,
- les lignes suivantes avec pour chaque espèce ou relevé, les coordonnées sur les axes, les contributions relatives et les probabilités associées.

La classification des relevés se fait généralement en deux temps. Il y a d'abord le calcul d'une matrice de distance entre les relevés puis la mise en œuvre de l'algorithme de classification. Il y a donc deux choix importants à opérer. Quelle distance entre les relevés choisir ? Les possibilités sont nombreuses. La distance euclidienne est souvent privilégiée parmi un très large éventail de coefficients. Mais quelles données faut-il utiliser pour calculer ces distances ? Données brutes ou transformées ? Le plus souvent, les données brutes sont utilisées. Comme dans les analyses factorielles, ces choix sont susceptibles d'apporter des résultats bien différents !

Dès que l'on dispose de variables transformées par une analyse factorielle, il est plus rationnel de les utiliser en lieu et place des coefficients d'abondance brut. Ainsi, on ne retient plus que la variabilité utile du tableau avec le nombre de variables transformées (ici les coordonnées des relevés sur les axes) fixé par le test par permutation. Cette opération est d'autant plus utile que la variabilité dans le tableau de données est grande, avec de nombreuses espèces peu fréquentes.

Afin de ne pas trop allonger notre propos, nous avons choisi un algorithme fréquemment utilisé, à savoir la classification hiérarchique ascendante. Nous comparons les résultats obtenus en partant d'abord directement des données d'abondance et en partant ensuite avec des variables transformées par l'analyse non symétrique des correspondances du tableau  $1 > 1 > 3$ .

Une variante des classifications habituelles se calcule à partir d'un nombre prédéterminé de clusters. Elle suit idéalement une classification au sens strict, après que l'on a fixé un nombre idéal de clusters. Cette méthode s'appelle "kmeans". Elle permet de regrouper les objets en k clusters distincts (Della Vedova, 2022). Elle repose sur la minimisation de la somme des distances euclidiennes au carré entre chaque objet (ici les relevés) et le centroïde (ou point central) de son cluster. C'est une approche de clustering non hiérarchique (à l'intérieur d'un cluster les objets ne sont pas ordonnés en fonction de leur ressemblance). La distance euclidienne est à la base de la méthode des kmeans, puisqu'il s'agit d'attribuer un cluster à chaque point, de façon que la somme des distances euclidiennes au carré entre chaque point et le centroïde de son cluster, soit la plus faible possible. On parle de minimisation de la variation intra-cluster.

Deux techniques sont retenues pour fixer un nombre idéal de clusters : la méthode «elbow» (coude en français) et la méthode «average silhouette» (silhouette moyenne en français). L'idée de base est de définir des clusters de manière à minimiser la variation totale intra-cluster (somme totale des carrés au sein des clusters ou WSS), ce qui mesure leur compacité.

La méthode «elbow» donne la somme totale des carrés au sein des clusters en fonction du nombre  $k$  de clusters. Il faut choisir le nombre de clusters de manière telle que le fait d'ajouter un cluster n'améliore pas beaucoup la WSS. Idéalement, ce nombre est fixé quand on observe un coude marqué dans la courbe des sommes totales des carrés en fonction de  $k$ . Toutefois, la présence d'un coude n'est pas toujours marquée et la WSS diminue régulièrement en fonction de  $k$ . Cette méthode ne doit pas être utilisée seule.

La méthode «average silhouette» détermine dans quelle mesure un relevé est attaché à un cluster. Une silhouette moyenne élevée indique un bon clustering. Comme pour elbow, la silhouette moyenne est calculée en fonction du nombre de clusters et est à nouveau représentée par une courbe ; le maximum dans cette courbe indique le nombre approprié de clusters.

Après que l'on a obtenu des sous-ensembles de relevés, des tests par permutation sont aussi utilisés pour rechercher les espèces dont la fréquence est statistiquement plus élevée que dans l'ensemble du tableau. On dispose ainsi d'un outil statistique pour reconnaître des espèces caractéristiques, différentielles ou simplement compagnes.

Parmi les analyses multivariées, nous ne présentons les analyses avec des données d'abondance que pour l'analyse en composantes principales puisque l'expérience de très nombreuses analyses des correspondances montre que les résultats sont pauvres (Bouxin, 2016 et 2024), en plus du fait que ce type de données n'y est pas adapté. Il en va de même dans cette étude.

Tous les logiciels que nous utilisons sont en langage R. Ils ont été en partie transposés de la littérature, notamment de Foucard (1982) et Jember (2012) et sont en partie personnels. Ils sont spécialement adaptés aux analyses de tableaux de relevés. Les scripts sont disponibles sur simple demande.

## Analyse en composantes principales du tableau des abondances

La distribution des corrélations interspécifiques (Figure 1) montre une très grande majorité de coefficients entre 0 et -0,2 ; un très petit nombre est inférieur à -0,2 et un assez petit nombre est supérieur à 0. Le calcul des valeurs propres s'en trouve affecté.

En examinant les valeurs propres, exprimées en pourcentages de la trace de la matrice de corrélation (Figure 2), on constate une chute après la troisième valeur propre mais, avec le test par permutations, seules les valeurs propres 2 et 3 sont significatives. Nous avons choisi de considérer les trois premières par la suite, afin de mieux comparer tous les résultats.

Une seule espèce présente une contribution relative significative, et encore sur le troisième axe. Il s'agit de *Asplenium scolopendrium*. Sur le premier axe, les espèces qui présentent des contributions relatives les plus élevées sont respectivement et par ordre décroissant d'importance : *Poa compressa*, *Arabis hirsuta*, *Origanum vulgare* et *Thymus praecox* pour les coordonnées positives

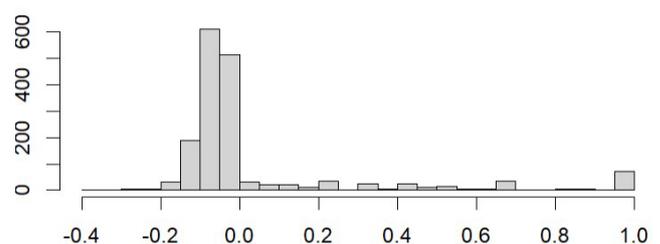


Figure 1. Distribution des coefficients de corrélation (ACP-matrice de corrélation).

et *Asplenium ceterach*, *Poterium sanguisorba*, *Helleborus foetidus*, *Bromopsis erecta* et *Poa nemoralis* pour les coordonnées négatives. Parmi ces espèces, certaines sont peu fréquentes et peu abondantes, d'autres sont plus fréquentes. Cela provient de la structure de la matrice de corrélation. Des corrélations élevées apparaissent entre des espèces rares si elles occupent les mêmes sites ou encore entre une espèce fréquente et une espèce rare si elles se retrouvent toutes deux ensemble. Le même scénario se reproduit avec les axes 2 et 3 (Figure 4).

En considérant les relevés, seuls R1 et R11 sont bien mis en évidence sur les axes 1 et 2 (Figure 3). Ils comprennent l'une ou l'autre espèce commune et plusieurs présentes une seule fois, celles déjà citées avec les contributions relatives élevées.

Sur l'axe 3 (Figure 4), les relevés avec des coordonnées élevées sont également occupés par des espèces peu fréquentes. Remarquons toutefois que dans la moitié supérieure, on trouve une bonne partie du syntaxon D.

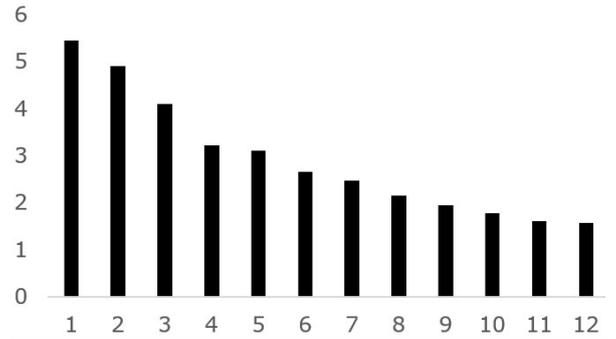


Figure 2. Les douze premières valeurs propres de la matrice de corrélation.

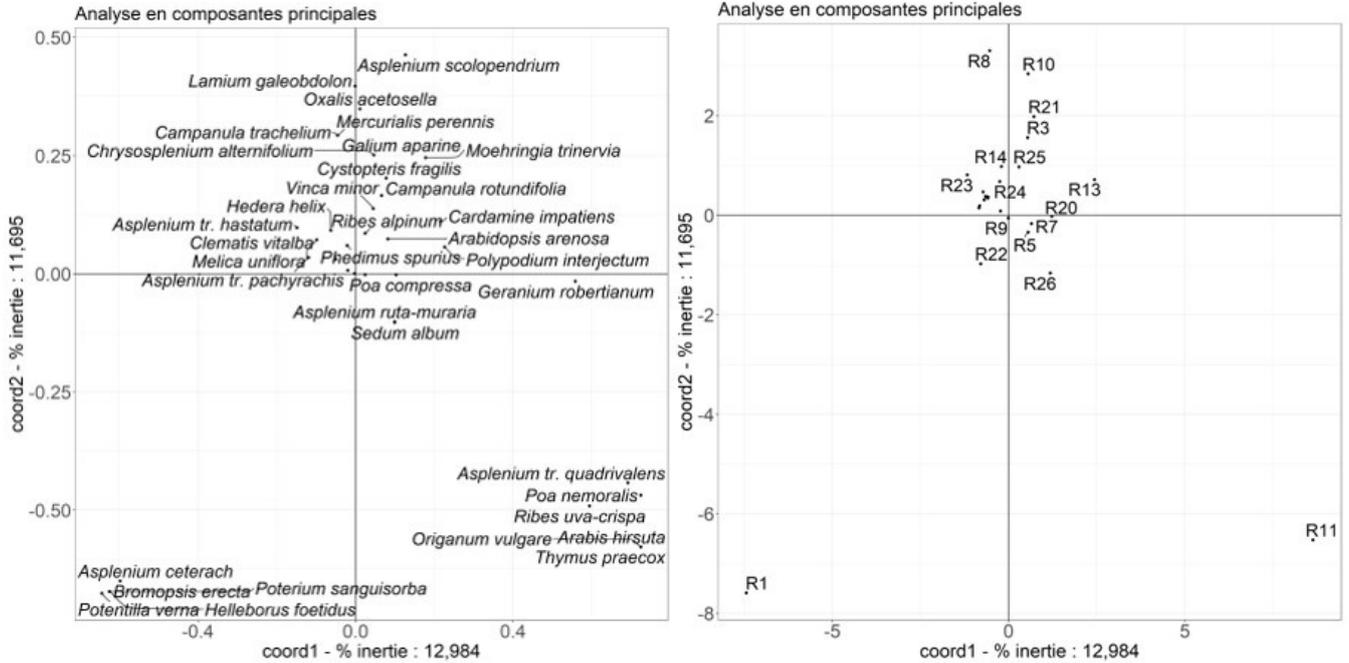


Figure 3. Analyse en composantes principales. Coordonnées des espèces et des relevés sur les axes 1 et 2

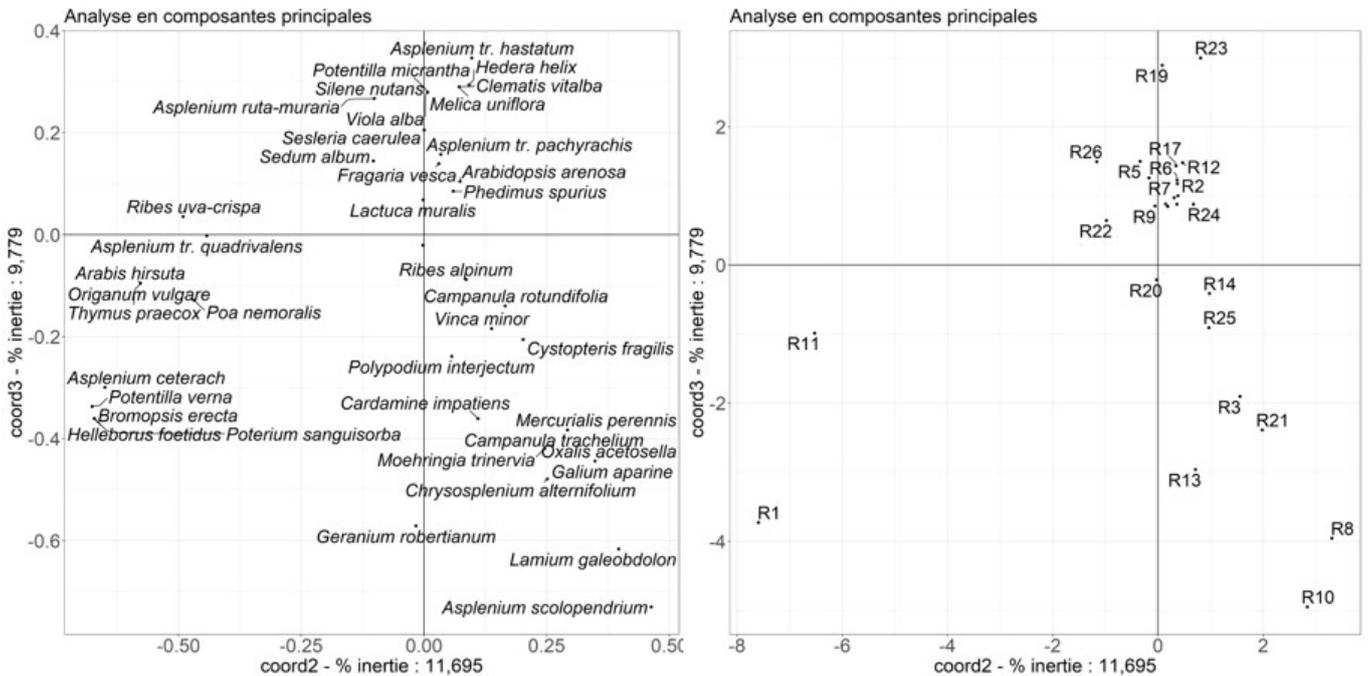


Figure 4. Analyse en composantes principales. Coordonnées des espèces et des relevés sur les axes 2 et 3.

On pourrait croire que le calcul d'une matrice de corrélation convient pour des données d'abondance, mais la structure même du tableau de relevés, avec des espèces présentant des fréquences fort différentes et surtout une proportion importante d'espèces peu fréquentes, parfois et même souvent ensemble dans un même site, rend l'analyse peu performante. Le calcul des valeurs



## Analyse non symétrique des correspondances du fichier 1-1-3

Ce n'est plus une analyse duale et on doit ainsi représenter séparément les espèces et relevés dans les plans définis par les coordonnées calculées. La distribution des coefficients de distance (Figure 9) est cette fois bien différente avec une asymétrie moins prononcée et une plus forte proportion de coefficients élevés, supérieurs à 0,1.

Les premières valeurs propres (Figure 10) sont relativement plus importantes que dans les précédentes analyses. Les premiers axes extraient ainsi une plus grande partie de la variabilité du tableau et apportent une meilleure aide à son interprétation. Le test par permutation donne les deux premiers axes comme significatifs. L'interprétation principale du tableau de relevés peut ainsi se faire en examinant la place des espèces et des relevés dans le plan des deux premiers axes.

Sur le premier axe (Figure 11), avec les contributions relatives les plus élevées et le plus souvent en même temps avec les coordonnées les plus élevées en valeur absolue, on trouve des espèces fréquentes : *Asplenium ruta-muraria*, *Cystopteris fragilis*, *Geranium robertianum*, *Asplenium trichomanes* subsp. *pachyrachis*, *Asplenium scolopendrium* et *Potentilla verna*. Aucune espèce avec une ou deux présences n'a de contribution importante. L'interprétation n'est donc plus dominée par les espèces rares. Remarquons également que les lignes avec les présences ont des contributions plus importantes que les lignes correspondantes avec des abondances > 1. Seuls *Asplenium trichomanes* subsp. *pachyrachis* et *Potentilla verna* ont des contributions relatives significatives. Sur le deuxième axe, avec toujours les contributions les plus élevées, on trouve des espèces moins fréquentes et parfois rares : *Asplenium trichomanes* subsp. *hastatum*, *Hedera helix*, *Asplenium trichomanes* subsp. *quadrivalens*, *Cystopteris fragilis* et *Campanula rotundifolia*, c'est-à-dire des espèces parmi les plus communes. Seuls *Asplenium trichomanes* subsp. *hastatum*, *Clematis vitalba* et *Melica uniflora* ont des contributions relatives significatives, ces deux dernières espèces une seule fois présentes n'ayant qu'un faible poids dans le calcul des coordonnées des espèces et relevés.

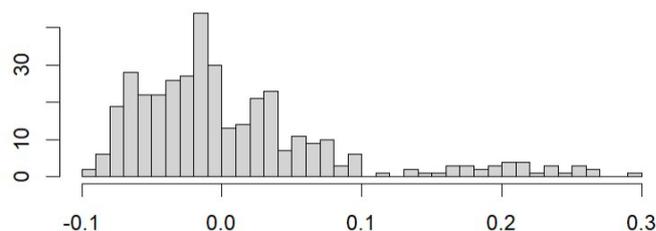


Figure 9. Distribution des coefficients de distance

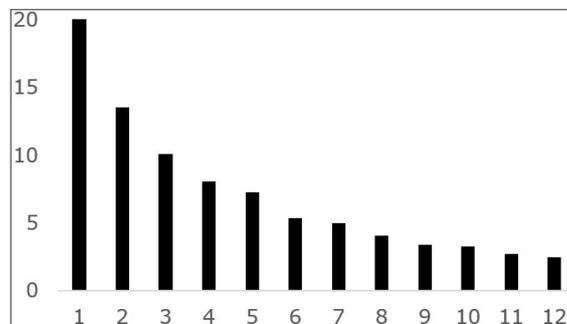


Figure 10. Les douze premières valeurs propres de la matrice de distance

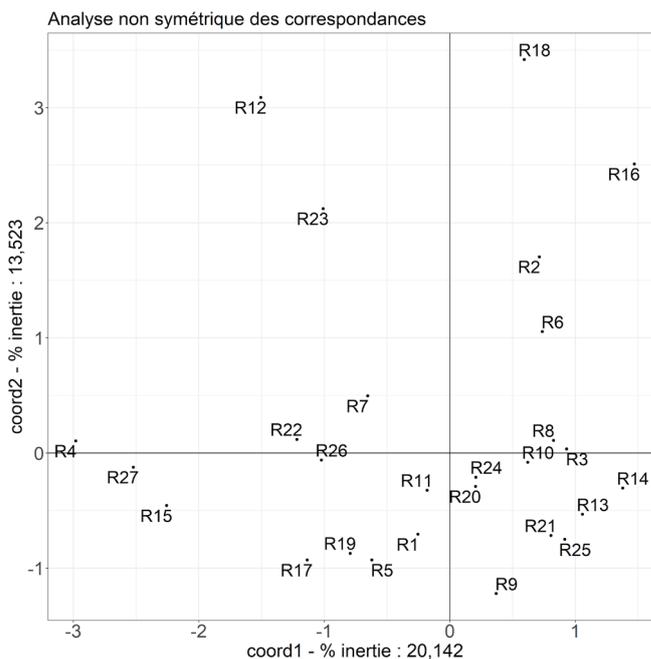
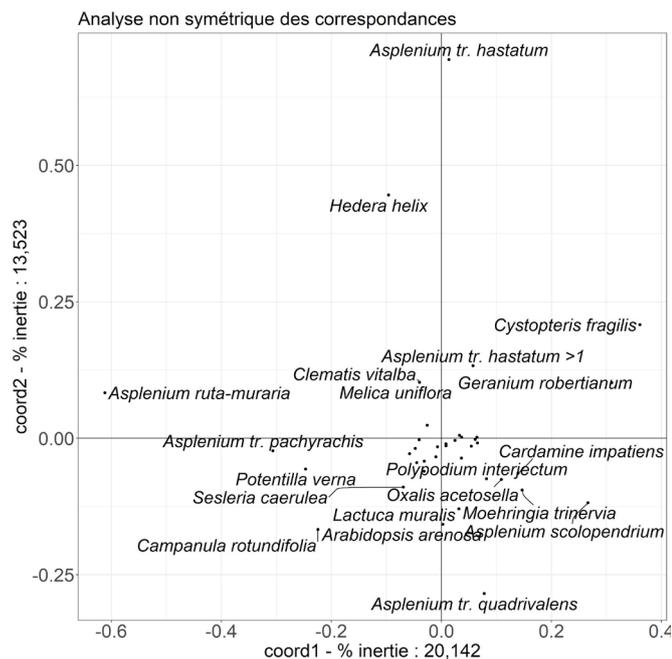


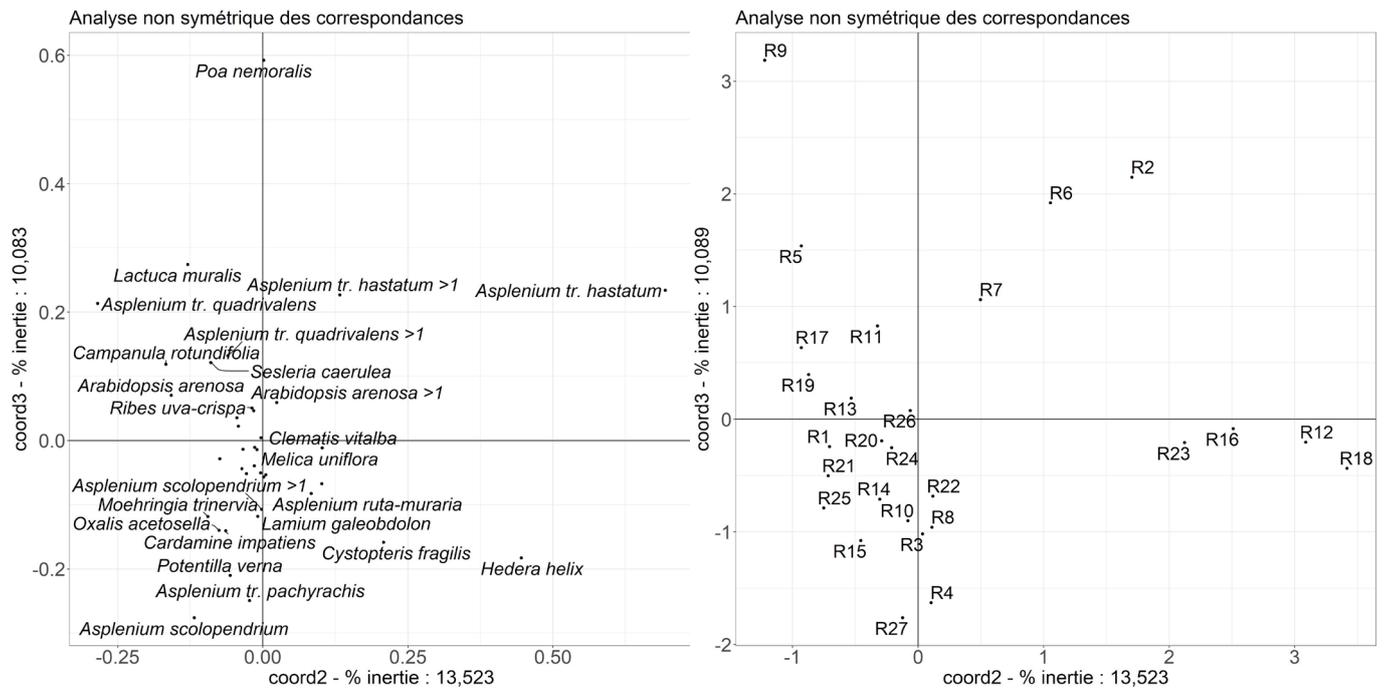
Figure 11. Analyse non symétrique des correspondances. Coordonnées des espèces et des relevés sur les axes 1 et 2 ; >1 : uniquement pour les abondances > 1

Remarquons aussi que sur le troisième axe (Figure 12) quatre espèces ont aussi des contributions relatives significatives : *Poa nemoralis*, *Asplenium trichomanes* subsp. *hastatum* > 1, *Lactuca muralis* et *Asplenium trichomanes* subsp. *pachyrachis*.

Pour les relevés dans le plan des axes 1 et 2, on isole facilement trois groupes : le premier avec les relevés 4, 15 et 27 (= syntaxon A), le deuxième avec les relevés 2, 6, 12, 16, 18 et 23 (= syntaxon E) et le troisième avec les 18 autres.

Avec le troisième axe, le groupe formé des relevés 12, 16, 18 et 23 (partie du syntaxon E) est maintenant isolé et séparé des relevés 2 et 6 (seconde partie de E). Le relevé 9 est également séparé du reste.

L'analyse non symétrique, donnant le plus de poids aux espèces fréquentes, tout en ne négligeant pas les espèces peu communes, semble ainsi la mieux adaptée à l'analyse d'un tableau de relevés. La même analyse produit pratiquement les mêmes résultats avec des données uniquement de présence et il n'est pas nécessaire de les présenter. Avec ce fichier, les données d'abondance se révèlent non essentielles.



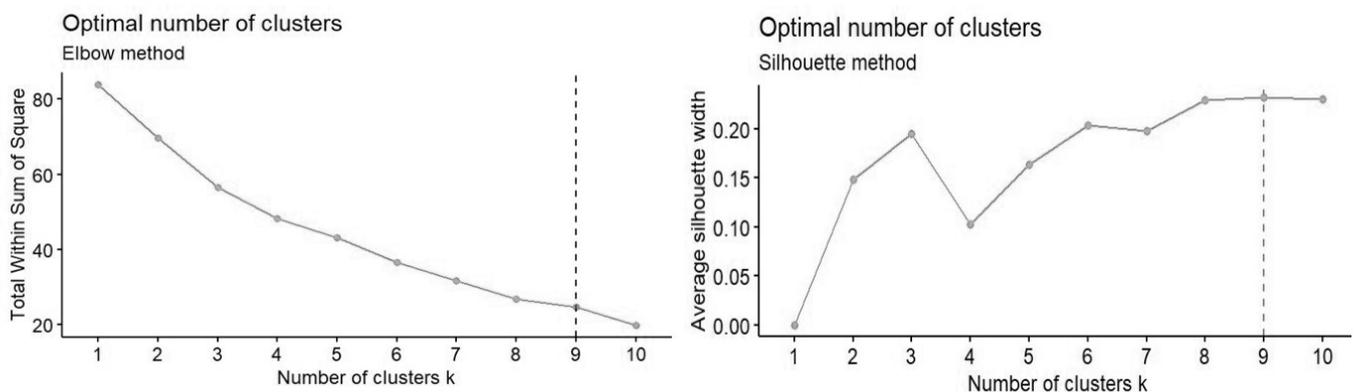
**Figure 12.** Analyse non symétrique des correspondances. Coordonnées des espèces et des relevés sur les axes 2 et 3 ; >1 : uniquement pour les abondances >1

### Classification des relevés

La classification se fait à partir de deux fichiers. D'abord avec les fichiers des abondances 0,1-1-2-3-4-5 puisque c'est ce type de fichier qui est généralement utilisé, ensuite avec les premières coordonnées issues de l'analyse non symétrique des correspondances. Cette seconde façon de procéder ne reprend ainsi que les coordonnées significatives et élimine une partie de la variabilité du fichier original, variabilité que l'on ne contrôle pas à cette échelle d'observation.

### Avec les données originelles

Dans la recherche du nombre optimal de clusters (**Figure 13**), la méthode «elbow» ne fournit pas de résultat clair. Avec «silhouette», on arrive à neuf clusters, avec un découpage produisant plusieurs relevés isolés. Cette analyse n'a donc guère de



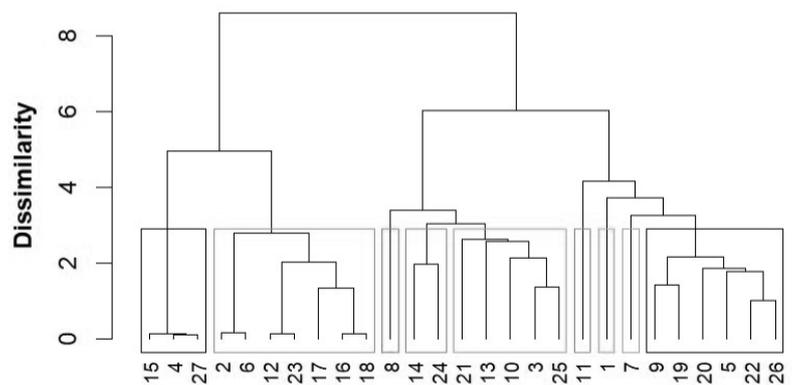
**Figure 13.** Détermination du nombre de clusters avec les techniques Elbow et Silhouette.

valeur synthétique. On s'éloigne assez bien de la classification sigmatiste. Remarquons toutefois que l'on retrouve les trois relevés (4, 15 et 27) du syntaxon A dans un des clusters (**Figure 14**).

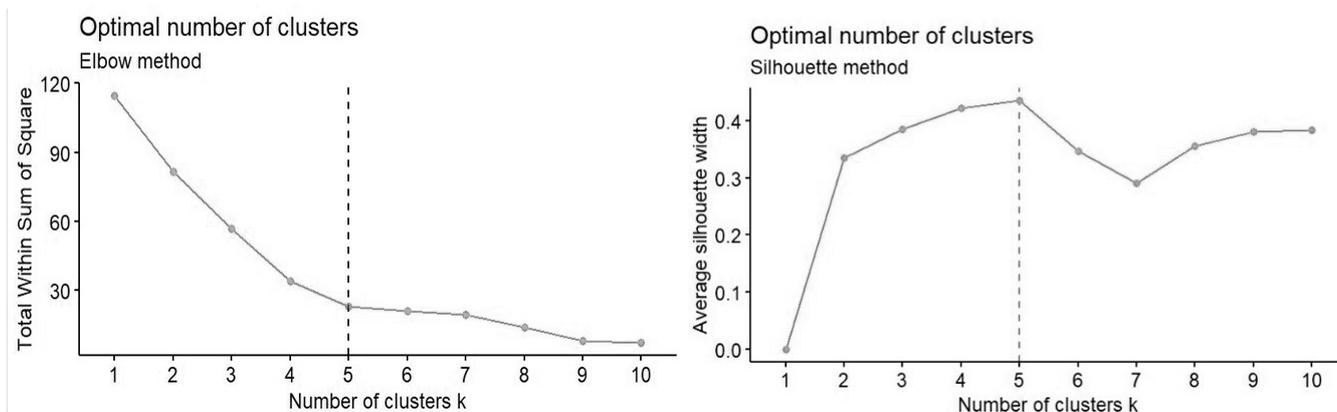
### Avec les trois premières coordonnées de nsca – tableau disjonctif 1-1-3.

La technique Elbow (Figure 15) montre bien un coude avec cinq clusters. Il en va de même avec Silhouette. Cette fois-ci le résultat semble exploitable.

Dans le dendrogramme (**Figure 16**), on trouve, de gauche à droite, un premier cluster avec sept des neuf relevés du syntaxon D, un second avec six des sept relevés du syntaxon C, un troisième avec deux des six relevés du syntaxon E et un du syntaxon



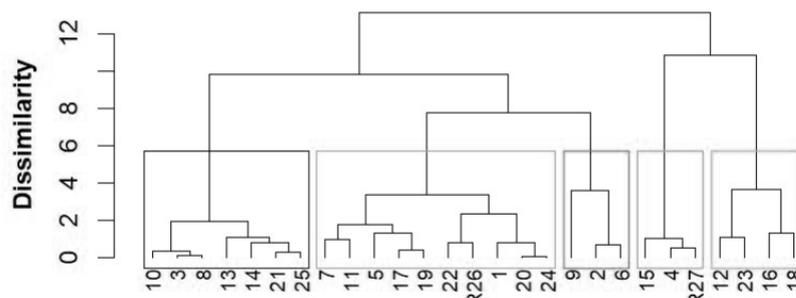
**Figure 14.** Dendrogramme de la classification hiérarchique ascendante avec les données d'abondance



**Figure 15.** Détermination du nombre de clusters avec les techniques Elbow et Silhouette

C, un quatrième cluster avec les trois relevés du syntaxon A et un cinquième avec quatre relevés sur les six du syntaxon E.

Il y a donc une constante dans les résultats, aussi bien dans l'analyse factorielle que dans la classification : le syntaxon A est bien mis en évidence et le syntaxon B n'est jamais reconnu. Remarquons qu'en partant avec un simple tableau de présence on arrive exactement à la même classification. À cette étape, on doit se demander comment opérer un choix parmi des résultats de classification parfois bien différents. Le choix se fait en faveur de la technique qui met le mieux en évidence des espèces caractéristiques.



**Figure 16.** Dendrogramme de la classification hiérarchique ascendante avec les données transformées

## Recherche des espèces caractéristiques

Les classifications sont évaluées en examinant dans quelles mesures elles produisent des combinaisons caractéristiques. Les syntaxons de Royer et les clusters retenus par l'analyse numérique sont soumis au test par permutation dans la recherche de ces caractéristiques.

### Avec les syntaxons préalablement définis

Dans le syntaxon A (**Tableau 2**), une seule espèce, soit *Asplenium trichomanes* subsp. *pachyrachis*, voit sa fréquence statistiquement supérieure à celle du reste des relevés. *Potentilla verna* montre une probabilité à peine supérieure à la norme 0,05 et les différentielles négatives sont *Asplenium trichomanes* subsp. *quadrivalens* et *Hedera helix*. Ce syntaxon est peu diversifié avec seulement quatre espèces et donc beaucoup d'espèces absentes par rapport aux 24 autres relevés.

Espèces	Relevés	4	15	27	S	Fr	Pr
<i>Asplenium trichomanes</i> subsp. <i>pachyrachis</i>		1	1	1	3	1	4,00E-04
<i>Potentilla verna</i>		0	1	1	2	0,667	0,0526
<i>Asplenium ruta-muraria</i>		1	1	1	3	1	0,1331
<i>Campanula rotundifolia</i>		0	1	0	1	0,333	0,6056

**Tableau 2.** Syntaxon A avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans le syntaxon B (**Tableau 3**) un peu plus diversifié, il y a deux différentielles positives avec une probabilité inférieure à 0,05, soit *Asplenium ceterach* et à nouveau *Potentilla verna*. La fougère n'est présente que dans les deux relevés du syntaxon tandis que la potentille est aussi présente dans deux autres. Dans l'ensemble des 25 autres relevés, aucune espèce ne montre de probabilité inférieure à 0,3407. Il n'y a donc pas de différentielle négative.

Espèces	Relevés	1	22	S	Fr	Pr
<i>Asplenium ceterach</i>		1	1	2	1	0,0028
<i>Potentilla verna</i>		1	1	2	1	0,0160
<i>Poterium sanguisorba</i>		1	0	1	0,5	0,0735
<i>Helleborus foetidus</i>		1	0	1	0,5	0,0740
<i>Bromopsis erecta</i>		1	0	1	0,5	0,0777
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		1	1	2	1	0,4426
<i>Hedera helix</i>		0	1	1	0,5	0,5697
<i>Asplenium ruta-muraria</i>		0	1	1	0,5	0,7818

**Tableau 3.** Syntaxon B avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans le syntaxon C (**Tableau 4**) plus diversifié, *Campanula rotundifolia*, *Asplenium trichomanes* subsp. *quadrivalens* et *Asplenium ruta-muraria* sont des différentielles positives ; la première espèce est la seule à bien discriminer ce syntaxon, puisqu'elle n'a qu'une présence en dehors de celui-ci. Deux autres espèces sont assez proches de la limite 0,05. *Arabidopsis arenosa* est loin du compte. *Cystopteris fragilis* est une différentielle négative, *Asplenium scolopendrium* est proche de la limite.

Espèces	Relevés	5	7	9	11	17	19	26	S	Fr	Pr
<i>Campanula rotundifolia</i>		1	1	0	0	1	1	1	5	0,7143	0,0045
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		1	1	1	1	1	1	1	7	1	0,0328
<i>Asplenium ruta-muraria</i>		1	1	0	1	1	1	1	6	0,8571	0,0464
<i>Ribes uva-crispa</i>		0	0	0	1	0	0	1	2	0,2857	0,0586
<i>Sesleria caerulea</i>		1	0	0	0	1	0	0	2	0,2857	0,0589
<i>Poa nemoralis</i>		1	1	1	1	0	0	0	4	0,5714	0,0886
<i>Lactuca muralis</i>		1	0	1	0	0	0	0	2	0,2857	0,1616
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i>		1	1	0	0	0	1	0	3	0,4286	0,2408
<i>Fragaria vesca</i>		0	0	0	0	1	0	0	1	0,1429	0,2578
<i>Sedum album</i>		0	0	0	0	0	0	1	1	0,1429	0,2578
<i>Thymus praecox</i>		0	0	0	1	0	0	0	1	0,1429	0,2598
<i>Viola alba</i>		0	0	0	0	0	1	0	1	0,1429	0,2611
<i>Potentilla micrantha</i>		0	0	0	0	0	1	0	1	0,1429	0,2616
<i>Origanum vulgare</i>		0	0	0	1	0	0	0	1	0,1429	0,262
<i>Arabis hirsuta</i>		0	0	0	1	0	0	0	1	0,1429	0,266
<i>Silene nutans</i>		0	0	0	0	0	1	0	1	0,1429	0,268
<i>Hedera helix</i>		0	1	0	0	0	0	1	2	0,2857	0,7742
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		0	1	0	0	0	0	0	1	0,1429	0,9168
<i>Geranium robertianum</i>		0	0	0	1	0	0	0	1	0,1429	0,9777

**Tableau 4.** Syntaxon C avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Ce syntaxon a été rattaché à l'*Asplenietum trichomano-rutae-murariae cardaminopsietosum borbasii*. Dans le syntaxon D (**Tableau 5**), huit espèces ont une fréquence significative et trois sont assez constantes. Il y a deux différentielles négatives : *Asplenium trichomanes* subsp. *hastatum* et *Asplenium ruta-muraria*. Ce syntaxon est donc assez bien discriminé par la combinaison des trois espèces les plus communes.

Espèces	Relevés	3	8	10	13	14	20	21	24	25	S	Fr	Pr
<i>Asplenium scolopendrium</i>		1	1	1	1	1	0	1	1	1	8	0,889	0
<i>Cystopteris fragilis</i>		1	1	0	1	1	1	1	1	1	8	0,889	0,0006
<i>Geranium robertianum</i>		1	1	1	1	1	1	1	0	0	7	0,778	0,0035
<i>Cardamine impatiens</i>		1	0	1	1	0	0	0	0	1	4	0,444	0,0059
<i>Oxalis acetosella</i>		0	1	0	0	0	0	1	1	1	4	0,444	0,0064
<i>Moehringia trinervia</i>		1	0	0	1	0	0	1	0	1	4	0,444	0,0071
<i>Lamium galeobdolon</i>		0	1	1	0	0	0	0	1	0	3	0,333	0,0287
<i>Polypodium interjectum</i>		0	0	0	1	0	1	1	0	0	3	0,333	0,0297
<i>Asplenium trichomanes</i> s. <i>quadrivalens</i>		1	1	1	1	1	1	1	0	1	8	0,8889	0,0952
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i>		0	0	0	0	0	1	1	1	1	4	0,4444	0,1398
<i>Chrysosplenium alternifolium</i>		0	0	1	0	0	0	0	0	0	1	0,1111	0,3292
<i>Vinca minor</i>		1	0	0	0	0	0	0	0	0	1	0,1111	0,3325
<i>Ribes alpinum</i>		0	0	0	0	0	0	0	0	1	1	0,1111	0,3326
<i>Campanula trachelium</i>		0	1	0	0	0	0	0	0	0	1	0,1111	0,3335
<i>Phedimus spurius</i>		0	0	0	0	0	0	0	1	0	1	0,1111	0,3335
<i>Poa compressa</i>		0	0	0	0	0	1	0	0	0	1	0,1111	0,3341
<i>Galium aparine</i>		0	0	1	0	0	0	0	0	0	1	0,1111	0,3360
<i>Hedera helix</i>		1	1	1	0	0	0	0	0	0	3	0,3333	0,6616
<i>Lactuca muralis</i>		0	0	0	1	0	0	0	0	0	1	0,1111	0,7259
<i>Poa nemoralis</i>		0	0	0	1	0	0	0	1	0	2	0,2222	0,8483
<i>Campanula rotundifolia</i>		0	0	0	0	0	0	1	0	0	1	0,1111	0,9639
<i>Asplenium ruta-muraria</i>		0	0	0	0	0	1	0	1	0	2	0,2222	0,9955

**Tableau 5.** Syntaxon D avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans ce syntaxon E (**Tableau 6**), une seule espèce est nettement plus fréquente que dans les 21 autres relevés ; c'est *Asplenium trichomanes* subsp. *hastatum*. Une seule différentielle négative apparaît, c'est *Asplenium trichomanes* subsp. *quadrivalens*.

Espèces	Relevés	2	6	12	16	18	23	S	Fr	Pr
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		1	1	1	1	1	1	6	1	0
<i>Clematis vitalba</i>		0	0	0	0	0	1	1	0,1667	0,2256
<i>Melica uniflora</i>		0	0	0	0	0	1	1	0,1667	0,2270
<i>Hedera helix</i>		0	0	1	0	1	1	3	0,5000	0,3070
<i>Cystopteris fragilis</i>		0	1	0	1	1	0	3	0,5000	0,4737
<i>Poa nemoralis</i>		1	1	0	0	0	0	2	0,3333	0,5926
<i>Geranium robertianum</i>		1	0	0	1	0	0	2	0,3333	0,7535
<i>Asplenium ruta-muraria</i>		0	0	1	0	0	1	2	0,3333	0,9283
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		0	1	0	0	0	0	1	0,1667	0,9997

**Tableau 6.** Syntaxon E avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Le syntaxon E correspond à l'*Asplenietum hastati*. Toutes ces données servent à établir une comparaison avec les classifications numériques.

## Avec les syntaxons définis par l'analyse non symétrique des correspondances

Dans le premier cluster (**Tableau 7**), il y a sept espèces avec des fréquences significatives et surtout une espèce différentielle négative très nette avec *Asplenium ruta-muraria*. Cette dernière est celle qui a la contribution relative la plus élevée sur l'axe 1 de l'analyse non symétrique des correspondances. C'est donc un élément solide pour objectiver un syntaxon. On retrouve sept des neuf relevés du syntaxon D. La différence se marque avec la constance parfaite de *Asplenium scolopendrium* et d'*Asplenium trichomanes* subsp. *quadrivalens* mais aussi avec l'absence d'*Asplenium ruta-muraria*. Ce syntaxon est à ranger dans l'association *Cystopterido fragilis* - *Phyllidetum scolopendrii* (Royer et al. 2006).

Espèces	Relevés	3	8	10	13	14	21	25	S	Fr	Pr
<i>Asplenium scolopendrium</i>		1	1	1	1	1	1	1	7	1	0
<i>Cardamine impatiens</i>		1	0	1	1	0	0	1	4	0,5710	0,0018
<i>Moehringia trinervia</i>		1	0	0	1	0	1	1	4	0,5710	0,0019
<i>Geranium robertianum</i>		1	1	1	1	1	1	0	6	0,8570	0,0034
<i>Cystopteris fragilis</i>		1	1	0	1	1	1	1	6	0,8570	0,008
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		1	1	1	1	1	1	1	7	1	0,0344
<i>Oxalis acetosella</i>		0	1	0	0	0	1	1	3	0,429	0,0406
<i>Asplenium scolopendrium</i> >1		1	0	1	0	0	0	0	2	0,2857	0,0599
<i>Lamium galeobdolon</i>		0	1	1	0	0	0	0	2	0,2857	0,1534
<i>Polypodium interjectum</i>		0	0	0	1	0	1	0	2	0,2857	0,1565
<i>Mercurialis perennis</i> >1		0	1	0	0	0	0	0	1	0,1429	0,2548
<i>Cystopteris fragilis</i> >1		0	0	0	0	1	0	0	1	0,1429	0,2566
<i>Galium aparine</i>		0	0	1	0	0	0	0	1	0,1429	0,2574
<i>Vinca minor</i>		1	0	0	0	0	0	0	1	0,1429	0,2591
<i>Ribes alpinum</i>		0	0	0	0	0	0	1	1	0,1429	0,264
<i>Campanula trachelium</i>		0	1	0	0	0	0	0	1	0,1429	0,2646
<i>Chrysosplenium alternifolium</i>		0	0	1	0	0	0	0	1	0,1429	0,2654
<i>Hedera helix</i>		1	1	1	0	0	0	0	3	0,4286	0,4281
<i>Lactuca muralis</i>		0	0	0	1	0	0	0	1	0,1429	0,6084
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i>		0	0	0	0	0	1	1	2	0,2857	0,6095
<i>Campanula rotundifolia</i>		0	0	0	0	0	1	0	1	0,1429	0,9132
<i>Poa nemoralis</i>		0	0	0	1	0	0	0	1	0,1429	0,9433

**Tableau 7.** Premier des cinq clusters avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans le deuxième cluster (**Tableau 8**), il y a six relevés sur dix en commun avec le syntaxon C. Cette fois, c'est la présence d'*Asplenium ruta-muraria* qui s'impose. *Asplenium trichomanes* subsp. *quadrivalens* a aussi une contribution relative sur le deuxième axe ; elle est constante dans ce cluster. On rencontre ici des espèces caractéristiques des *Asplenietaea trichomanis* ; on doit rattacher ce syntaxon à l'association *Asplenetum trichomano - ruta-murariae*. Par rapport à l'analyse phytosociologique, aucune distinction n'apparaît plus entre les sous-associations *ceterachetosum* et *cardaminopsietosum borbasii*.

Espèces	Relevés	1	5	7	11	17	19	20	22	24	26	S	Fr	Pr
<i>Asplenium ruta-muraria</i>		0	1	1	1	1	1	1	1	1	1	9	0,9	0,0026
<i>Campanula rotundifolia</i>		0	1	1	0	1	1	0	0	0	1	5	0,5	0,0409
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i>		0	1	1	0	0	1	1	0	1	0	5	0,5	0,0413
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		1	1	1	1	1	1	1	1	0	1	9	0,9	0,0578
<i>Sesleria caerulea</i>		0	1	0	0	1	0	0	0	0	0	2	0,2	0,1267
<i>Ribes uva-crispa</i>		0	0	0	1	0	0	0	0	0	1	2	0,2	0,1295
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i> >1		0	1	0	1	0	0	0	0	0	0	2	0,2	0,1312
<i>Asplenium ceterach</i>		1	0	0	0	0	0	0	1	0	0	2	0,2	0,1358
<i>Poa nemoralis</i>		0	1	1	1	0	0	0	0	1	0	4	0,4	0,3203
<i>Phedimus spurius</i>		0	0	0	0	0	0	0	0	1	0	1	0,1	0,3594
<i>Fragaria vesca</i>		0	0	0	0	1	0	0	0	0	0	1	0,1	0,3630
<i>Bromopsis erecta</i>		1	0	0	0	0	0	0	0	0	0	1	0,1	0,3653
<i>Poterium sanguisorba</i>		1	0	0	0	0	0	0	0	0	0	1	0,1	0,3658
<i>Origanum vulgare</i>		0	0	0	1	0	0	0	0	0	0	1	0,1	0,3668
<i>Asplenium ceterach</i> >1		1	0	0	0	0	0	0	0	0	0	1	0,1	0,3676
<i>Helleborus foetidus</i>		1	0	0	0	0	0	0	0	0	0	1	0,1	0,3693
<i>Viola alba</i>		0	0	0	0	0	1	0	0	0	0	1	0,1	0,3693
<i>Sedum album</i>		0	0	0	0	0	0	0	0	0	1	1	0,1	0,3694
<i>Arabis hirsuta</i>		0	0	0	1	0	0	0	0	0	0	1	0,1	0,3696
<i>Potentilla micrantha</i>		0	0	0	0	0	1	0	0	0	0	1	0,1	0,3731
<i>Poa nemoralis</i> >1		0	0	0	1	0	0	0	0	0	0	1	0,1	0,3737
<i>Poa compressa</i>		0	0	0	0	0	0	1	0	0	0	1	0,1	0,3744
<i>Silene nutans</i>		0	0	0	0	0	1	0	0	0	0	1	0,1	0,3747
<i>Arabidopsis arenosa</i> subsp. <i>borbasii</i> >1		0	0	1	0	0	0	0	0	0	0	1	0,1	0,3751
<i>Thymus praecox</i>		0	0	0	1	0	0	0	0	0	0	1	0,1	0,377
<i>Potentilla verna</i>		1	0	0	0	0	0	0	1	0	0	2	0,2	0,4656
<i>Hedera helix</i>		0	0	1	0	0	0	0	1	0	1	3	0,3	0,7599
<i>Lactuca muralis</i>		0	1	0	0	0	0	0	0	0	0	1	0,1	0,7638
<i>Polypodium interjectum</i>		0	0	0	0	0	0	1	0	0	0	1	0,1	0,7669
<i>Lamium galeobdolon</i>		0	0	0	0	0	0	0	0	1	0	1	0,1	0,7682

<i>Oxalis acetosella</i>	0	0	0	0	0	0	0	0	1	0	1	0,1	0,8617
<i>Geranium robertianum</i>	0	0	0	1	0	0	1	0	0	0	2	0,2	0,9694
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>	0	0	1	0	0	0	0	0	0	0	1	0,1	0,9797
<i>Cystopteris fragilis</i>	0	0	0	0	0	0	1	0	1	0	2	0,2	0,9832
<i>Asplenium scolopendrium</i>	0	0	0	0	0	0	0	0	1	0	1	0,1	0,9882

**Tableau 8.** Deuxième des cinq clusters avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans le troisième syntaxon (**Tableau 9**), il y a une combinaison d'une espèce abondante et d'une autre simplement présente, mais dans seulement deux des trois relevés. *Poa nemoralis* est constante et deux autres espèces sont présentes dans deux relevés. C'est une combinaison originale qui tient compte de l'importance de *Poa nemoralis* dans la classification. On se rapproche de l'*Asplenietum hastati*. Précisons toutefois que la même combinaison de relevés apparaît avec simplement des données de présence.

Espèces	Relevés	R2	R6	R9	S	Fr	Pr
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i> >1		1	1	0	2	0,6667	0,0077
<i>Poa nemoralis</i>		1	1	1	3	1	0,0193
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		1	1	0	2	0,66667	0,1626
<i>Lactuca muralis</i>		0	0	1	1	0,33333	0,3069
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		0	1	1	2	0,66667	0,7470
<i>Geranium robertianum</i>		1	0	0	1	0,33333	0,7678
<i>Cystopteris fragilis</i>		0	1	0	1	0,33333	0,8085

**Tableau 9.** Troisième des cinq clusters avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Dans le quatrième cluster (**Tableau 10**), on retrouve le même syntaxon de trois relevés qui apparaît, quelle que soit la technique utilisée.

Espèces	Relevés	R4	R15	R27	S	Fr	Pr
<i>Asplenium trichomanes</i> subsp. <i>pachyrachis</i>		1	1	1	3	1	0,0004
<i>Potentilla verna</i>		0	1	1	2	0,667	0,0526
<i>Asplenium ruta-muraria</i>		1	1	1	3	1	0,1331
<i>Campanula rotundifolia</i>		0	1	0	1	0,333	0,6056

**Tableau 10.** Quatrième des cinq clusters avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Le cinquième syntaxon (**Tableau 11**), un peu plus diversifié, est caractérisé par *Asplenium trichomanes* subsp. *hastatum* qui présente une forte contribution relative sur le second axe de l'analyse non symétrique. Dans cet ordre, *Hedera helix* vient en seconde position. Il y a ici quatre des six relevés du syntaxon E ; les deux autres relevés de ce syntaxon sont marqués par l'abondance d'*Asplenium trichomanes* subsp. *hastatum*. En suivant Royer *et al.* (2006), ce syntaxon est facilement rattaché à l'*Asplenietum hastati*. Ce cluster se marque aussi par l'absence d'*Asplenium trichomanes* subsp. *quadrivalens* (la seule significative statistiquement), de *Poa nemoralis*, *Asplenium scolopendrium*, *Arabidopsis arenosa* subsp. *borbasii* et *Campanula rotundifolia*, parmi les plus communes.

Espèces	Relevés	12	16	18	23	S	Fr	Pr
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		1	1	1	1	4	1	0,0022
<i>Hedera helix</i>		1	0	1	1	3	0,75	0,0928
<i>Melica uniflora</i>		0	0	0	1	1	0,25	0,1458
<i>Clematis vitalba</i>		0	0	0	1	1	0,25	0,1507
<i>Cystopteris fragilis</i>		0	1	1	0	2	0,5	0,5469
<i>Asplenium ruta-muraria</i>		1	0	0	1	2	0,5	0,7299
<i>Geranium robertianum</i>		0	1	0	0	1	0,25	0,8593

**Tableau 11.** Cinquième des cinq clusters avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

## Discussion et nouvelle analyse complémentaire

Le premier cluster de notre dernière analyse (relevés 3, 8, 10, 13, 14, 21, 25) est pratiquement identique au syntaxon de Royer. Il s'agit d'une végétation relativement variée avec sept espèces caractéristiques. Un autre cluster (relevés 4, 15, 27) correspond exactement au syntaxon de Royer ; il s'agit d'un groupement peu diversifié avec une caractéristique exclusive et une seconde peu commune. Il n'en va plus de même pour les autres clusters.

Il y a un syntaxon qui ne résiste pas aux analyses statistiques de ce tableau, c'est celui composé des relevés 1 et 22 qui se retrouvent dans un cluster avec des relevés comprenant aussi *Asplenium trichomanes* subsp. *quadrivalens*, dont la contribution relative est nettement supérieure à celle d'*Asplenium ceterach*. De toute manière, il paraît difficile d'apporter une signification statistique à un groupement constitué d'un aussi petit nombre de relevés. Cela peut aussi signifier qu'il serait utile soit de retourner sur le terrain afin de disposer d'un plus grand nombre de relevés, soit de comparer avec d'autres tableaux publiés. Nous avons repris l'étude de Ferrez (2009) sur les groupements végétaux des parois calcaires du massif jurassien et de la Franche-Comté. Les deux relevés 1 et 22 y figurent (avec les étiquettes AT4 et AT5). Nous avons reconstitué un tableau de 68 relevés comprenant les 19 relevés de l'*Asplenio viridis* - *Cystopteridetum fragilis*, les 22 relevés du *Cystopterido fragilis* - *Phyllitidetum scolopendrii*, les 13 relevés du *Moehringio trinerviae* - *Geranietum robertiani*, les 6 relevés de l'*Asplenietum trichomano* - *rutae-murariae* et les 8 relevés de l'*Asplenio fontani* - *Ceterachetum officinarum*. L'analyse statistique réalisée comme précédemment fournit six clusters dont un cluster de sept relevés comprenant les deux de Royer ; ce cluster comprend cinq des six relevés de l'*Asplenietum trichomano* - *rutae-murariae* présenté mais aucun de l'*Asplenio fontani* - *Ceterachetum officinarum* présenté par le même auteur, auquel les relevés 1 et 22 de Royer sont pourtant associés. Nous tentons d'expliquer cette différence grâce à

la recherche des espèces caractéristiques ou différentielles du cluster de 7 relevés comprenant les deux relevés litigieux parmi les 68 relevés, ce qui fournit le tableau 12.

Espèces	Relevés	AH4	AT1	AT2	AT3	AT4	AT5	AT6	S	Pr
<i>Potentilla verna</i>		0	1	0	0	1	1	0	3	0,001
<i>Asplenium ruta-muraria</i>		0	1	1	1	1	0	1	5	0,003
<i>Ribes uva-crispa</i>		0	1	1	0	0	0	0	2	0,050
<i>Hieracium murorum</i>		0	1	0	1	0	0	1	3	0,058
<i>Bromopsis erecta</i>		0	0	0	0	0	1	0	1	0,097
<i>Festuca lemanii</i>		0	1	0	0	0	0	0	1	0,102
<i>Festuca burgundiana</i>		0	1	0	0	0	0	0	1	0,103
<i>Teucrium montanum</i>		0	1	0	0	0	0	0	1	0,103
<i>Reseda lutea</i>		0	0	0	1	0	0	0	1	0,103
<i>Brachypodium pinnatum</i>		0	0	0	0	0	1	0	1	0,103
<i>Poterium sanguisorba</i>		0	0	0	0	0	1	0	1	0,104
<i>Asplenium ceterach &gt;1</i>		0	0	0	0	0	1	0	1	0,106
<i>Asplenium trichomanes</i> subsp. <i>hastatum</i>		1	0	0	0	0	0	0	1	0,106
<i>Asplenium ceterach</i>		0	0	0	0	1	1	0	2	0,119
<i>Chelidonium majus</i>		0	1	0	0	0	0	0	1	0,189
<i>Helleborus foetidus</i>		0	0	0	0	0	1	0	1	0,192
<i>Euphorbia cyparissias</i>		0	1	0	0	0	0	0	1	0,194
<i>Sedum album</i>		0	0	1	0	0	0	0	1	0,199
<i>Campanula rotundifolia</i>		0	1	1	0	0	0	0	2	0,444
<i>Cardamine impatiens</i>		0	0	0	0	0	0	1	1	0,549
<i>Hedera helix</i>		1	0	1	0	1	0	0	3	0,668
<i>Asplenium trichomanes</i> subsp. <i>quadrivalens</i>		1	1	1	1	1	1	1	7	0,719
<i>Lactuca muralis</i>		1	0	0	0	0	0	0	1	0,977
<i>Geranium robertianum</i>		0	1	0	1	0	0	0	2	1,000

**Tableau 12.** Cluster de sept relevés avec la présence des espèces, la somme (S), la fréquence relative (Fr) et la probabilité (Pr) associées à chacune.

Trois espèces sont des différentielles négatives : *Asplenium scolopendrium*, *Cystopteris fragilis* et *Geranium robertianum*. C'est donc en grande partie l'absence combinée des deux premières qui caractérise ce groupement et les deux relevés litigieux, qui devraient être rangés dans l'*Asplenietum trichomano - rutaemurariae*.

La même démarche est appliquée avec un plus grand tableau de relevés de savane (Bouxin, 2024) et le sera ensuite avec de très grands tableaux, comprenant jusqu'à plusieurs centaines de relevés.

## Conclusions

L'avantage de commencer l'analyse d'un tableau de relevés par une analyse multivariée apparaît clairement. Celle-ci nous indique quelles sont les espèces avec les contributions relatives les plus élevées, c'est-à-dire celles qui participent le plus à la structure du tableau. Ce n'est pas un simple complément à l'analyse phytosociologique, mais bien une étape de la démarche de classification puisque ce sont ensuite les coordonnées sur les axes qui deviennent les données de base de la classification en lieu et place des abondances. Cela permet d'éliminer du tableau une certaine variabilité difficile à expliquer à cette échelle d'étude, ce qui se manifeste dans presque tous les tableaux de relevés. Les contributions relatives des axes restent de toute manière assez faibles, ce qui apparaît dans presque toutes les analyses de tableaux de relevés. Il ne faut pas oublier non plus que les choix de l'emplacement du relevé et de son aire conservent toujours une part d'arbitraire.

D'autres outils très performants sont peu exploités, comme l'analyse factorielle multiple qui conduit à pondérer l'influence respective des strates dans une végétation forestière, par exemple ou encore de coupler des fichiers floristiques et environnementaux, sans préjuger de leur structure particulière ou du type de lien qui existe entre eux.

L'analyse statistique des tableaux de relevés ne devrait pas être un simple outil élégant et complémentaire à l'analyse phytosociologique, mais peut parfaitement devenir une des étapes importantes dans la démarche conduisant à une classification de la végétation. Nous sommes ici dans le domaine de la variabilité, de l'incertain et il est impératif d'adapter nos outils à ce domaine particulier.

## Bibliographie

- Bailly G., 2012 - Contribution à l'étude des aulnaies marécageuses comtoises. *Les Nouvelles Archives de la Flore jurassienne et du nord-est de la France* **10** : 57-102
- Bourouche J.-M. & Saporta G., 1980 - *L'analyse des données*. P.U.F. Collection Que sais-je ?, 125 p.
- Bouxin G., 1999 - Description de la végétation aquatique et du bord de l'eau dans le bassin hydrographique de la Moline (Condruz, Belgique) par l'analyse multiscalaire des motifs. *Ecologie* **30** : 139-163
- Bouxin G., 2016 - *Analyse statistique des tableaux de relevés de végétation. Recherche d'adéquation entre les données de végétation et les techniques statistiques, au moyen d'exemples*. Éditions universitaires européennes, 440 p.
- Bouxin G., 2024 - *Analyse statistique des tableaux de relevés de végétation*. <http://guy-bouxin.e-monsite.com> (dernière mise à jour 23/01/2024)
- Collaud R. avec la collaboration de Ferrey Y., Simler N. et Bailly G., 2010 - Contribution à l'étude des forêts hygroscoaphiles de ravins à affinités atlantiques en Franche-Comté. *Les Nouvelles Archives de la Flore jurassienne et du nord-est de la France* **8** : 87-122

Della Vedova C., 2022 - *La méthode des Kmeans*. <https://delladata.fr> (dernière consultation en août 2022).

Ferrez Y., 2007 - Contribution à l'étude phytosociologique des prairies mésophiles de Franche-Comté. *Les Nouvelles Archives de la Flore jurassienne* **5** : 67-159

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

Ferrez Y., 2011 - Contribution à la connaissance des prairies humides oligotrophes (*Molinion caeruleae* Koch 1926) de Franche-Comté. *Les Nouvelles Archives de la Flore jurassienne* **9** : 9-24

Foucard T., 1982 - *Analyse factorielle. Programmation sur micro-ordinateurs*. Masson, Paris, 286 p.

Foucault de B., 1984 - *Systémique, structuralisme et synsystématique des prairies hygrophiles des plaines atlantiques françaises*. t. 1. Thèse de doctorat, Sciences Naturelles, Université de Rouen – Laboratoire d'Écologie Végétale. Université de Lille II – Laboratoire de Botanique. 409 p.

Gimaret-Carpentier C., Chessel D. & Pascal J.-P., 1998 - Non-symmetric correspondence analysis: an alternative for species occurrences data. *Pl. Ecol.* **138** : 97-112

Goodall D. W., 1954 - Objective methods in the classification of vegetation, III - An essay on the use of factor analysis. *Austr. J. Bot.* **2** : 304-324.

Greig-Smith P., 1964 - *Quantitative plant ecology (second edition)*. Butterworths, London, 256 p.

Guyonneau J., 2005 - Étude de la végétation et de l'hydrologie du marais de Levresses, réserve naturelle régionale des tourbières de Frasné. *Les Nouvelles Archives de la Flore jurassienne et du nord-est de la France* **3** : 69-126

Jember T.G., 2012 - *Multivariate data analysis using R software*. Lambert Academic publishing, 104 p.

Kroonenberg P. M. & Lombardo R., 1999 - Nonsymmetric correspondence analysis: a tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research* **34** : 367-396.

Manly B.F.J., 1997 - *Randomization, bootstrap and Monte Carlo methods in biology*. Second edition. Chapman & Hall, 399 p.

Palm R., 1993 - Les méthodes d'analyse factorielle : principes et applications. *Notes de Statistique et d'Informatique*. (Gembloux) 93/1, 38 p.

Royer J.-M., 2009 - Petit précis de phytosociologie sigmatiste. *Bull. Soc. Bot. Centre-Ouest*, n. s., n° sp. **33** : 1-86.

Royer J.-M., Felzines J.-C., Misset C. & Thévenin S., 2006 - Synopsis commenté des groupements végétaux de la Bourgogne et de la Champagne-Ardenne. *Bull. Soc. Bot. Centre-Ouest*, n. s., n° sp. **25** : 1-394.

Tison J.-M. & de Foucault B. (coords), 2014 - *Flora Gallica, Flore de France*. Biotopie, Mèze, xx + 1 196 p.

Wildi O., 2013 - *Data analysis in vegetation ecology (second edition)*. Wiley-Blackwell, 301 p.

Whittaker R.H., 1973 - *Ordination and classification of communities*. Handbook of vegetation science, V. W. Junk b.v. Publishers, The Hague, 737 p.