

---

# Chapitre 1. Introduction générale

---

par Guy BOUXIN<sup>°</sup>

Ce texte sur l'analyse des données de végétation porte principalement sur **le traitement statistique des tableaux de relevés floristiques et mésologiques**. Il n'est pas exhaustif mais il insiste surtout sur la compréhension des étapes successives de la démarche analytique, seule manière pour nous d'arriver à une bonne utilisation des techniques statistiques.

L'étude de la végétation a connu de nombreux développements depuis quelques décennies, notamment grâce à l'accès généralisé aux techniques informatiques et aux logiciels d'analyses multivariées. Cela a favorisé l'emploi de techniques statistiques autrefois décriées ou tout au moins regardées avec une grande méfiance. Mais les résultats n'ont pas toujours suivi car l'utilisation automatique des outils disponibles a trop souvent fait disparaître la réflexion préalable et indispensable pour mener à bien toutes les étapes d'une étude de végétation.

Un des objectifs de ce texte est de montrer qu'une étude de végétation est un tout, que les étapes successives de la démarche scientifique sont intimement liées, depuis la collecte sur le terrain jusqu'à la synthèse finale. Les botanistes font généralement leur étude de terrain en récoltant des données dans des relevés d'assez petite taille, peu nombreux par rapport à l'aire considérée, séparés les uns des autres par des distances importantes. Souvent, ce n'est qu'après, c'est-à-dire trop tard, qu'ils se soucient de l'analyse des données. Cette façon de faire n'apporte que des résultats limités avec la fausse sécurité d'un traitement mathématique. Souvent, l'utilisateur n'est d'ailleurs satisfait que si les résultats de l'analyse confirment les impressions de terrain.

Reprenons les étapes habituelles de l'analyse des données de végétation :

- la récolte de données sur le terrain; il s'agit de données floristiques ou mésologiques;
- la présentation des données dans un tableau à double entrée, les lignes correspondant aux espèces ou aux facteurs de l'environnement et les colonnes aux sites décrits;
- l'analyse et la synthèse des données;

---

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

- la mise en relation des données de végétation avec les données de l'environnement.

Toutefois, pour réussir ce parcours, l'observateur doit être parfaitement conscient du type de travail qu'il entreprend et surtout d'avoir bien défini ses objectifs (JONGMAN *et al.*, 1987). Dans quelle démarche son travail s'inscrit-il ? Toutes les étapes devraient logiquement être adaptées en fonction des objectifs.

Pourtant, comme le souligne GOUNOT (1969), l'utilisation de l'outil statistique devrait aboutir à dégager des principes utilisables en vue d'une synthèse dépassant et intégrant dans toute la mesure du possible les théories et les méthodes existantes sur l'étude de la végétation. L'approche quantitative de cette étude a été rendue possible grâce aux efforts des chercheurs anglo-saxons des précédentes décades. Mais, on doit tâcher d'y incorporer tout l'apport des écoles européennes continentales, empiriques, certes mais nourries d'une expérience concrète inégalée. GOUNOT (1969) ajoute encore : "Il nous semble en effet essentiel de bien dégager dès le départ la spécificité propre des problèmes posés par la végétation, car les méthodes ne sont jamais neutres en dépit des apparences. Il importe donc de s'assurer que le modèle théorique que suppose (implicitement ou explicitement) toute technique statistique est bien compatible avec la réalité que nous cherchons à appréhender". Il n'est pas inutile de rappeler brièvement le problème des relations de la phytosociologie et de l'écologie (GUINOCHET, 1973). Pour ce dernier, la phytosociologie est l'étude des associations végétales dont l'existence même, ou plutôt, la possibilité de les mettre en évidence en se livrant à un certain nombre d'opérations bien définies, constitue un phénomène en soit dont il est légitime de s'occuper pour lui-même. La phytosociologie apparaît bien comme une discipline à part entière, dont l'objectif n'est pas uniquement la diagnose floristique et la classification des associations végétales : il comporte également l'étude de leur dynamique, de leurs relations avec les variables de l'environnement, de leur histoire, c'est-à-dire de leur évolution et de leur genèse. Pour GUINOCHET, la phytosociologie ne doit pas être inféodée à l'écologie. Ceci montre l'opposition avec GREIG-SMITH (1964) qui fut un des pionniers de l'étude quantitative de la végétation et qui écrit notamment au début du chapitre 1 de son livre "Many ecological data take the form of description of vegetation with or without concurrent recording of factors of the environment. Such data have formed the main basis of most ecological theory and concepts and are likely to continue to do so". Dans le même ordre d'idées, GAUCH (1982) parle de "community ecology" et ORLÓCI (1978) de "vegetation ecology". Cette conception différente de la place de l'étude de la végétation a eu une influence majeure sur la manière d'organiser le travail scientifique. La phytosociologie, *sensu stricto*, a développé un formalisme très élaboré, la rendant difficile d'accès aux "non-initiés" et l'isolant, tout en limitant son développement.

Il est impossible de s'intéresser à l'étude de la végétation sans faire allusion à une autre opposition passionnée entre ceux qui voient la végétation comme un *discontinuum* (école sigmatiste, voir GUINOCHET, 1973, pages 159 et suivantes) et ceux qui ont une interprétation individualiste du comportement des espèces

et qui basent leurs études sur le concept de *continuum* (WHITTAKER 1973). L'analyse approfondie du tapis végétal montre à quel point cette opposition est vaine. Certains auteurs confondent ainsi modèle et réalité, c'est-à-dire qu'ils regardent la végétation en fonction du modèle de description qu'ils utilisent. Leur démarche de description est forcément influencée par ce modèle qui devient paradigme, voire dogme (voir à ce sujet LEGAY, 1996).

Le problème de la récolte de données de végétation dans des placettes ou relevés, ainsi que celui de l'élaboration d'un plan d'échantillonnage ne seront pas développés dans un chapitre particulier puisque des informations détaillées ont été présentées par GOUNOT (1969), GREIG-SMITH (1964), GUINOCHET (1973), KERSHAW (1964), KNAPP (1984), ORLÓCI & KENKEL (1985), PODANI (2000) et VANDEN BERGHEN (1973). L'échantillonnage sera forcément évoqué dans les exemples choisis.

Il est ainsi possible de construire beaucoup de tableaux différents, en fonction du mode de collecte des données (par exemple dans une série de relevés situés soit dans un seul site, soit dans des sites d'un ou de plusieurs types de communautés). Chaque colonne du tableau correspond à une placette, à un site ou à un relevé et chaque ligne à des espèces ou des paramètres de l'environnement. Le tableau phytosociologique forme ainsi une matrice primaire à partir de laquelle l'écologue espère arriver à une synthèse de la description du motif de variation dans la nature et par conséquent de l'interpréter.

Avec le tableau construit, l'analyse des données peut commencer mais tout d'abord qu'est-ce qu'une analyse des données ? Reprenons les termes de BOUROCHE & SAPORTA (1980) : "La statistique classique s'est axée sur l'étude d'un seul caractère (ou variable) mesuré sur un petit ensemble d'individus. Elle a développé les notions d'estimation et de tests fondées sur des hypothèses très restrictives. Cependant, dans la pratique, les individus observés sont fréquemment caractérisés par un grand nombre de caractères. Les méthodes d'analyse des données permettent une étude globale de ces variables, mettant en évidence des liaisons, des ressemblances ou des différences. Pour cela, on plonge individus et variables dans des espaces géométriques tout en faisant la plus grande économie d'hypothèses et on transforme les données pour les visualiser dans un plan ou les classer en groupes homogènes et ceci tout en perdant le minimum d'information". Il s'agit donc d'une "approche descriptive multidimensionnelle". "On distingue généralement deux ensembles : les individus et les caractères relatifs à ces individus". En citant toujours les auteurs : « L'ensemble des individus observés provient soit d'un échantillonnage dans une population (dans le cas d'un sondage) ou alors de la population entière. L'objectif n'est pas de rechercher des caractéristiques d'une population mais bien de décrire la structure de l'ensemble d'individus observés, abandonnant toute volonté pseudo-inductive".

C'est dans cet esprit qu'est née l'analyse factorielle des correspondances. HUSSON *et al.* (2009) expliquent que l'analyse factorielle des correspondances fut inventée pour traiter de vastes données

linguistiques et qu'elle a été le premier maillon d'un grand ensemble de techniques statistiques qui se proposaient d'aborder les tableaux statistiques indépendamment de toute hypothèse mathématique ou probabilistique sur la structure des données. Toujours en citant les mêmes auteurs, « Une approche multidimensionnelle implique la prise en compte simultanée de l'ensemble des liaisons entre les variables prises deux à deux. C'est bien ce qui est fait, par exemple dans la mise en évidence de variables synthétiques : une telle variable en représente plusieurs autres, ce qui implique qu'elle soit liée à chacune d'entre elles, ce qui n'est possible que si ces dernières sont elles-mêmes liées entre elles deux à deux. La notion de variable synthétique est donc bien intrinsèquement multidimensionnelle ; elle est un outil puissant de description d'un tableau individus x variables ».

Pourquoi donner une telle importance à l'outil statistique ? La définition de SCHWARTZ (1994) est déjà un élément de réponse : « La statistique est un mode de pensée permettant de recueillir, de traiter et d'interpréter les données qu'on rencontre dans divers domaines, et tout particulièrement dans les sciences de la vie, du fait que ces données présentent une caractéristique essentielle : la variabilité ». Ou encore : « L'individu diffère des autres individus, il diffère de lui-même d'un moment à l'autre. Ainsi le domaine du vivant est fait de cas particuliers. Mais il n'y a de science que du général. Comment alors peut-il y avoir une science du vivant ? Voilà posé le problème : il faut adapter la science au domaine du particulier. Du fait de la variabilité, on est dans le domaine de l'incertain. Adapter une science au particulier, c'est inventer une science de l'incertain ».

Notre intention n'est pas de rédiger des notes théoriques sur tous les aspects de l'analyse des données de végétation mais bien de faire comprendre la logique d'une approche statistique complète dans une étude de végétation, en sachant parfaitement qu'il sera difficile de convaincre ceux qui continuent à utiliser la simple technique de permutation de lignes et tableaux ou qui utilisent des techniques que je qualifierai de mixtes, c'est-à-dire une collecte de données suivant des techniques traditionnelles, suivie d'une analyse statistique sommaire.

Les fondements conduisant à une analyse complète des données ont été posés par GOUNOT (1969) et GREIG-SMITH (1979). GOUNOT s'est beaucoup intéressé à la structure du tapis végétal et il déclarait notamment : "Tout d'abord, le tapis végétal a souvent, sinon toujours, un aspect en mosaïque. Elle est constituée par la répétition de motifs structuraux, généralement en petit nombre, liés à des variations concomitantes de composition floristique quantitative et qualitative. L'homogénéité du tapis végétal ne peut exister que si la mosaïque est répétitive, c'est-à-dire résulte d'un agencement plus ou moins régulier de ses différentes parties. L'homogénéité est généralement une question d'échelle. Les parties de la mosaïque qui semblent homogènes à une certaine échelle se révèlent en général avoir elles aussi une structure complexe si l'on passe à une échelle plus fine. L'existence de ces structures emboîtées implique qu'il y a des discontinuités

plus ou moins nettes dans le tapis végétal, donc des niveaux privilégiés pour l'analyse de la végétation". GOUNOT avait observé combien le tapis végétal peut être complexe dans certains cas, et combien en conséquence il est difficile de donner des critères permettant de définir l'unité d'échantillonnage dans le cas général. Un point sur lequel nous insistons est l'opinion de l'auteur sur la courbe aire-espèces : en définitive, il semble bien que l'utilisation de la courbe aire-espèces ne soit pas une méthode valable de définition de l'aire minimale. Les méthodes de détermination pratique de la courbe sont, soit contestables, soit difficilement réalisables. D'autre part, il n'est même pas certain que la courbe aire-espèces présente une asymptote horizontale. Il semble plutôt que le nombre d'espèces croît d'une manière plus ou moins linéaire en fonction du logarithme de la surface inventoriée, dans la gamme de surfaces où l'on situe généralement l'aire minimale. Enfin l'allure de la courbe est certainement très influencée par la densité et le mode de distribution des espèces. Le résultat dépend donc de causes si diverses que son interprétation théorique est difficile. Tout lecteur intéressé par la relation aire-nombre d'espèces consultera avantagement les travaux de WILLIAMSON (2003) et de WILLIAMSON *et al.* (2001 et 2002). Bien qu'ayant abandonné la technique de la courbe aire-espèces, GOUNOT (1969) persiste néanmoins à rechercher de vraies communautés homogènes en utilisant des critères écologiques, structuraux et dynamiques afin de discriminer des mosaïques. Et il ajoute : "De toute façon, il y a une limite inférieure au pouvoir résolutif de la végétation qui est du même ordre de grandeur que l'aire minimale (dans la mesure où une surface inférieure à l'aire minimale n'est pas représentative de la communauté)". Il nous est difficile de souscrire à cette dernière affirmation. En 1979, GREIG-SMITH déclarait: "Though it is convenient to distinguish variation between communities and variation within communities, because techniques of investigation are generally different, there is clearly a continuum both of scale and of intensity of pattern. Ecologists are increasingly aware of the need to "build a bridge" between community ecology and population ecology if we are to further our understanding of the functioning of ecosystems and ultimately to predict or control changes in them". Nous verrons comment il est dorénavant possible de "jeter ce pont" et l'impact de cette constatation dans la définition des communautés végétales.

La démarche scientifique que nous suivons dans cet ouvrage se place dans la logique de création et de rejet de modèles (FOUREZ, 1992). Un modèle n'apporte pas la vérité mais c'est son efficacité qui nous intéresse. Le modèle de description de la végétation auquel nous arrivons a déjà été présenté dans BOUXIN (1995, 1999) mais il évolue constamment (BOUXIN 2011 et 2013). Nous avons tenu à le développer complètement et surtout à expliciter les étapes qui ont conduit à son élaboration.

Le chapitre « Base de données et programmes » présente d'abord le système R et des « packages » ou programmes facilement accessibles. Par rapport à d'anciennes versions, nous publions plusieurs scripts personnels d'analyse multivariée en langage R, permettant la transformation et l'analyse des tableaux dans l'esprit que nous développons. Y figurent les programmes de transformations de tableaux d'abondance-

dominance en tableaux de présence et en tableaux disjonctifs, l'analyse en composantes principales, l'analyse des correspondances duale, l'analyse non-symétrique des correspondances, associées à un test par permutations, ainsi que des analyses factorielles multiples basées sur les deux analyses des correspondances, avec également les tests par permutations. D'autres logiciels importants et programmes sont encore expliqués. Les fichiers servant d'exemples sont repris en annexe.

Cet ouvrage est aussi l'occasion de rendre hommage à quelques pionniers de l'analyse statistique de la végétation et qui m'ont directement aidé ou influencé d'une manière ou d'un autre : P. BERTHET, D. CHESSEL, P. DAGNELIE, M. GOUNOT, L. ORLÓCI et E. van der MAAREL. Sans eux, ce travail n'aurait jamais vu le jour.

Janvier 2024

## Références

- BOUROCHE, J.-M. & SAPORTA, G. (1980). *L'analyse des données*. Presses universitaires de France. Collection "Que sais-je ?". 127 pp.
- 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* **30** : 139-163.
- BOUXIN, G. (2011). Évolution de la végétation macrophytique et trophie dans deux ruisseaux du bassin hydrographique de la Molinee (Condroz, Belgique). *Rev. Sci. Eau*, **24**, 253-266.
- BOUXIN, G. (2013). Végétation macrophytique, environnement et qualité d'eau dans le bassin versant du Bocq (Belgique, Wallonie). *Rev. Sci. Eau*, **26**, 1-19.
- FOUREZ, G. (1992). *La construction des sciences*. Deuxième édition revue. De Boeck Université. 288 pp.
- GAUCH, H.G., Jr. (1982). *Multivariate analysis in community ecology*. Cambridge University Press. 298 pp.
- GOUNOT, M. (1969). *Méthodes d'étude quantitative de la végétation*. Masson, Paris. 314 pp.
- GREIG-SMITH, P. (1964). *Quantitative plant ecology (2<sup>nd</sup> ed.)*. Butterworths, London. 256 pp.
- GREIG-SMITH, P. (1979). Pattern in vegetation. *Journal of Ecology* **67**: 755-779.
- GUINOCHET, M. (1973). *Phytosociologie*. Masson et Cie. 227 pp.
- HUSSON, R., LÊ, S. & PAGÈS, J. (2009). *Analyse des données avec R*. Presses Universitaires de France. 224 pp.
- JEANMONOD, D. & GAMISANS, J. (2007). *Flora Corsica*. Édisud. 921 pp. et 136 figures.

- JONGMAN, R.H.G., ter BRAAK, C.J.F. & van TONGEREN, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc Wageningen. 299 pp.
- KERSHAW, K.A. (1964). *Quantitative and dynamic ecology*. Edward Arnold, London. 183 pp.
- KNAPP, R. (1984). *Sample (relevé) areas (distribution, homogeneity, size, shape) and plot-less sampling*. In Knaap, R. (ed.), *Sampling methods and taxon analysis in vegetation science*. Dr W. Junk Publishers, The Hague. 105-119.
- LEGAY, J.-M. (1996). *L'expérience et le modèle. Un discours sur la méthode*. INRA editions. 112 pp.
- ORLÓCI, L. (1978). *Multivariate analysis in vegetation research* (2<sup>nd</sup> ed.). Dr. W. Junk, Publishers, The Hague. 451 pp.
- ORLÓCI, L. & KENKEL, N.C. (1985). *Statistical Ecology Monographs. Vol. I. Introduction to data analysis*. International Co-operative Publishing House. 339 pp.
- PODANI, J. (2000). *Introduction to the exploration of multivariate biological data*. Backhuys Publishers, Leiden. 407 pp.
- SCHWARTZ, D. (1994). *Le jeu de la science et du hasard. La statistique et le vivant*. Flammarion. 111 pp.
- VANDEN BERGHEN, C. (1973). *Initiation à l'étude de la végétation*. Les Naturalistes Belges. 236 pp.
- WHITTAKER, R.H. (1973). *Ordination and classification of communities*. Handbook of vegetation science. Part V. Dr. W. Junk b.v. - Publishers - The Hague. 737 pp.
- WILLIAMSON, M. (2003). Species-area relationships at small scales in continuum vegetation. *Journal of Ecology*, **91**: 904-907.
- WILLIAMSON, M., GASTON, K.J. & LONSDALE, W.M. (2001). The species area relationship does not have an asymptote! *Journal of Biogeography*, **28**: 827-830.
- WILLIAMSON, M., GASTON, K.J. & LONSDALE, W.M. (2002). An asymptote is an asymptote and not found in species-area relationships. *Journal of Biogeography*, **29**: 1713.