



HAL
open science

ANALYSES MULTIVARIEES DES RELATIONS ESPECES-MILIEU : STRUCTURE ET INTERPRETATION ECOLOGIQUE

R Prodon, J D Lebreton

► **To cite this version:**

R Prodon, J D Lebreton. ANALYSES MULTIVARIEES DES RELATIONS ESPECES-MILIEU : STRUCTURE ET INTERPRETATION ECOLOGIQUE. Vie et Milieu / Life & Environment, 1994, pp.69-91. hal-03047896

HAL Id: hal-03047896

<https://hal.sorbonne-universite.fr/hal-03047896v1>

Submitted on 9 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSES MULTIVARIEES DES RELATIONS ESPECES-MILIEU : STRUCTURE ET INTERPRETATION ECOLOGIQUE

*Multivariate analyses of species-environment relationships :
structure and ecological interpretation*

R. PRODON⁽¹⁾, J.D. LEBRETON⁽²⁾

⁽¹⁾ Laboratoire Arago, Université Paris-6, CNRS URA 117, 66650 Banyuls-sur-Mer, France

⁽²⁾ Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, BP 5051, 34033 Montpellier cedex, France

ÉCOLOGIE
COMMUNAUTÉ
RELATIONS ESPECES-MILIEU
ORDINATION
GRADIENT
ANALYSES MULTIVARIEES
AMPLITUDE D'HABITAT
DIVERSITÉ

RÉSUMÉ – L'étude des relations entre les communautés animales ou végétales et leur environnement implique souvent un ensemble d'échantillons associant un certain nombre d'espèces et un certain nombre de variables de milieu. Le présent article passe en revue une large part des méthodes multivariées, basées sur le modèle linéaire, qui permettent l'analyse des tableaux relevés × espèces, en offrant différentes possibilités de couplage avec les tableaux relevés × variables de milieu correspondants. L'accent est mis sur la structure des modèles statistiques, sur la robustesse et la flexibilité des méthodes, sur leur choix, sur leur adéquation aux problématiques biologiques, et sur l'interprétation des résultats en relation avec les concepts écologiques.

ECOLOGY
COMMUNITY
SPECIES-ENVIRONMENT
RELATIONSHIPS
ORDINATION
GRADIENT
MULTIVARIATE ANALYSES
HABITAT BREADTH
DIVERSITY

ABSTRACT – The study of the relationships between animal or plant communities and their environment often supposes a set of samples that associate several species and several environmental variables. This paper reviews the main multivariate methods, based on the linear model, allowing the analysis of samples × species matrices, and offering different possibilities of coupling these matrices with the corresponding sample × environmental variables matrices. The emphasis is placed on the structure of the statistical models, on the robustness and flexibility of the methods, on the choice among them, on their relevance to biological problems, and on the interpretation of the results in relation with ecological concepts.

1. INTRODUCTION

L'écologie des communautés a longtemps entretenu l'espoir que des modèles type analyse des systèmes, en éliminant les aspects descriptifs liés à la prise en compte des multiples cas d'espèces, pourraient faire de l'écologie une science « dure » (McIntosh, 1987). Mais la prise de conscience de la diversité des mécanismes impliqués et de l'importance des facteurs historiques ou aléatoires, notamment des perturbations (Connell 1978, Denslow 1985, Pickett et White 1985), ont engendré une période de doute et d'introspection bien résumée par McIntosh (*ibid.*). De fait, le corpus théorique en écologie des communautés est en plein renouvellement. Malgré l'intérêt de concepts comme ceux d'espèces « clés » (*key-stone species*) (Paine 1966), de groupes fonctionnels ou de

guildes (Simberloff et Dayan 1991), l'indispensable analyse de la dynamique des communautés ne saurait se réduire à une juxtaposition ou une amplification d'approches « fonctionnelles » de type dynamique des populations. Les approches empiriques gardent toute leur importance, et l'échantillonnage des communautés doit reprendre son rôle central, qui est de fournir les informations de base sur leur variabilité spatio-temporelle.

Dans ce contexte, l'objectif de cet article est de fournir au lecteur écologiste non statisticien une approche pédagogique, mais néanmoins critique, d'une large part des méthodes actuelles d'analyse statistique en écologie des communautés. Un premier problème pratique est de mettre en évidence, résumer et quantifier la diversité et la variabilité floro/faunistique contenue dans un ensemble de relevés. Un deuxième problème est de mettre cette variabilité et cette diversité en re-

lation avec des variables de milieu. Pour cela, l'utilisation des méthodes statistiques multivariées est naturelle et souvent préférable à celle des méthodes de classification, ne serait-ce que parce que les variations des facteurs environnementaux sont le plus souvent graduelles dans l'espace ou le temps.

Si les analyses factorielles, particulièrement l'Analyse des Correspondances, sont largement utilisées en écologie, c'est trop souvent sans prise de conscience de leurs propriétés en tant que modèles écologiques. Les exposés pédagogiques usuels reposent sur des notions abstraites de calcul matriciel ou de géométrie dans l'espace sans connotation écologique directe. Nous soulignerons plutôt les liens étroits avec des concepts de base de l'écologie des communautés, et par conséquent l'adéquation entre méthode statistique et modèle écologique, montrant par exemple que l'Analyse des Correspondances est à l'intersection des notions d'*optimum écologique* et d'*espèce indicatrice*. Nous envisagerons ensuite l'intervention dans l'analyse des variables mesurées dans le milieu : c'est le domaine des analyses à plusieurs tableaux, où sont analysées les relations entre *vie* et *milieu*, c'est-à-dire entre tableaux florofaunistiques et tableaux des variables externes. Nous évoquerons enfin divers développements méthodologiques récents, en proposant quelques éléments de réflexion sur le choix des méthodes, sur leur interprétation en termes écologiques, et sur les possibilités de passer de la simple description exploratoire à la modélisation de phénomènes définis et aux tests d'hypothèses.

2. CORRESPONDANCE ENTRE ESPÈCES ET RELEVÉS

Comment mettre en évidence l'information contenue, sous forme de variabilité florofaunistique, dans un tableau espèces \times relevés, autrement dit comment faire apparaître la structure d'un tableau souvent volumineux et illisible ? Certaines espèces sont-elles liées à certains groupes de relevés ? Existe-t-il donc un classement des espèces tel qu'on puisse lui faire correspondre un classement des milieux ? Bref, peut-on mettre en *correspondance* les espèces et les relevés ?

2.1. Optimum écologique et ordination des espèces

Considérons la répartition d'une espèce dans divers relevés différant entre eux par une variable écologique quelconque. Sur l'axe représentant cette variable, chaque relevé est à la valeur x_i prise par la variable en question (éventuellement

en classes). Il y a donc ordination *a priori* des relevés. La distribution des relevés contenant une espèce donnée sur cet axe figure ce qu'on peut appeler le « profil écologique » (Gounot 1969) de l'espèce (cf. Romane 1972, Daget et Godron 1982). L'optimum de l'espèce sera logiquement estimé par un paramètre de position de la distribution, le plus souvent la *moyenne* par souci de simplicité. On pourrait aussi utiliser le mode de l'approximation gaussienne du profil de l'espèce (Gauch *et al.* 1974, Ter Braak et Looman 1986), mais cette méthode est difficilement utilisable si l'on dispose de plus de deux ordinations concurrentes des relevés. Remarquons que la moyenne ne constitue un « résumé » satisfaisant de la position de l'espèce sur le gradient que si la distribution de cette dernière est unimodale.

Dès qu'on envisage plusieurs espèces se pose le problème de leur classement. Il est logique de les ordonner d'après les moyennes de leurs distributions, c'est-à-dire les moyennes des x_i des relevés où on les trouve. On obtient ainsi, à partir de l'ordination *a priori* des relevés, une ordination *a posteriori* y_j des espèces sur un axe-espèces correspondant (Fig. 1a). Sur cet axe, des espèces dont les optimums écologiques sont proches auront des y_j proches. Cette méthode de classement des espèces sur un gradient par leurs positions moyennes (moyennes pondérées, ou moyennes simples si les relevés sont en présence-absence), ou « *averaging* » simple (*weighted averages*), est classique (Whittaker 1948, in Whittaker 1960 ; « méthode des barycentres », Daget 1977). D'une façon générale, l'étude de la distribution des espèces sur un gradient écologique défini *a priori* est appelée « analyse de gradient directe ». Cette définition *a priori* s'effectue en général à partir d'une variable du milieu dont l'importance est reconnue (humidité, altitude, stade d'une succession écologique, etc). On a également imaginé utiliser une combinaison *a priori* de variables à l'aide d'indices synthétiques comme dans la méthode des « scalars » (cf. Togerson 1958, McIntosh 1967, Austin *et al.* 1984).

D'un point de vue statistique, le classement obtenu par les moyennes *maximise la corrélation* entre les indices x_i des classes et les indices y_j des espèces. Mais la méthode ne fournit aucune information sur les ressemblances et différences florofaunistiques entre relevés (c'est-à-dire entre les x_i puisque ceux-ci sont définis *a priori*). De plus, chaque espèce est considérée indépendamment des autres.

2.2. Espèces indicatrices et ordination des relevés

Selon un raisonnement symétrique du précédent, on peut désirer ordonner un ensemble de relevés d'après les espèces qu'ils contiennent, si les

préférences de ces espèces concernant une variable écologique donnée sont connues (notion d'espèce « caractéristique »). La présence d'une espèce est alors considérée comme *indicatrice* d'une certaine valeur probable de la variable. Dans ce cas, ce sont les y_j (optimums des espèces sur ce facteur, autrement dit leur « valeurs indicatrices ») qui sont fixés *a priori*. On en déduit l'ordination *a posteriori* des relevés, la coordonnée x_i d'un relevé étant égale à la moyenne des y_j des espèces qu'il contient.

Ainsi Ellenberg (1948), désirant prédire l'acidité d'un sol d'après la présence de certaines plantes, classe d'abord ces espèces de 1 (très acidophiles) à 9 (très alcalophiles); la moyenne des indices des espèces d'un échantillon lui donne une estimation semi-quantitative de l'acidité du sol (voir aussi Ter Braak et Gremmen 1987). Whittaker (1960), ou Whittaker et Niering (1965), étudiant un gradient d'aridité, attribuent aux espèces les indices $y_j = 1$ (« mésiques ») à 4 (« xériques »), puis ordonnent les stations d'après la moyenne de ces indices. De même, Demarq et Mourer-Chauviré (1976), cherchant à reconstituer les variations paléoclimatiques d'après des faunes d'oiseaux fossiles, indexent les espèces de $y_j = 1$ (« arctiques ») à 4 (« méditerranéennes »); ils utilisent la moyenne des indices des espèces comme « indice thermique » de l'échantillon (voir d'autres exemples de « *habitat adaptation numbers* » ou « *species position indices* » in McIntosh 1967 et Goff et Cottam 1967).

Mais pour une matrice relevés x espèces donnée, la corrélation globale entre les indices x_i des relevés et y_j des espèces ne peut être optimale dès lors que l'on fixe *a priori* soit les uns, soit les autres. La qualité des ordinations obtenues par le calcul des moyennes dépend en effet de la qualité de l'ordination *a priori*. Elle dépend donc étroitement de l'expertise de l'écologiste et de la connaissance qu'il possède, ou croit posséder, soit des espèces, soit des relevés.

2.3. De l'ordination réciproque des espèces et des relevés à l'Analyse des Correspondances

On peut donc n'accorder qu'une confiance relative à l'ordination *a priori* des relevés (ou des espèces), et préférer s'en remettre à une méthode objective pour obtenir la meilleure ordination réciproque des espèces et des relevés possible. Ainsi, Dix et Smeins (1967), étudiant l'influence du drainage sur la végétation, commencent par ranger « à l'estime » leurs relevés dans les 6 classes d'une échelle approximative d'humidité. Ils choisissent alors 48 espèces indicatrices, c'est-à-dire ayant un mode bien marqué sur cette échelle provisoire, et les ordonnent de 1 (celles dont le mode est dans des stations très sèches) à 6 (très humides). Ils

en déduisent une nouvelle ordination des relevés, placés à la moyenne des indices de leurs espèces, « *more accurate than that based on the original visual assessment of topography and drainage* » (Gauch 1982). Il s'agit-là d'un raisonnement qui conduit, par itérations, à l'*averaging* réciproque, autrement dit à l'AFC.

Une des techniques de l'AFC (Hill 1973a) consiste en effet, à partir d'une ordination *a priori* des espèces, à ordonner les relevés d'après la moyenne des espèces qu'ils contiennent. Cette nouvelle ordination des relevés est utilisée pour réordonner les espèces par la même technique, qui conduit à une nouvelle ordination des relevés. On réitère ces deux étapes jusqu'à ce que les deux ordinations se stabilisent (Fig. 1b). La solution, unique, ne dépend plus des ordinations initiales. Digby et Kempton (1987 p. 72), sur un exemple, visualisent la convergence vers une ordination unique, après une dizaine d'itérations, en partant de deux ordinations initiales distinctes.

Ici, l'écologiste renonce donc à intervenir *a priori* dans l'ordination. Il en résulte trois propriétés importantes :

a) La solution dépend exclusivement de la structure interne du jeu de données. Elle est donc objective. Elle est aussi optimale. En effet, à chaque itération, la corrélation entre les indices x_i des relevés et y_j des espèces augmente, jusqu'à stabilisation où cette corrélation r est maximale. On obtient alors une ordination réciproque des espèces et des relevés, qui correspond au *premier couple de facteurs de l'AFC*. Cette ordination réciproque peut être visualisée par le graphe de corrélation canonique relevés x espèces (Fig. 2a). L'ordination est optimale : tout déplacement d'une ligne ou d'une colonne aurait pour effet de faire baisser la corrélation globale relevés x espèces (qui n'est autre que la racine de la valeur propre du facteur considéré de l'analyse). Le maximum de points (de « présences ») est situé vers la diagonale. Les lignes (ou colonnes) sont d'autant plus proches que leurs moyennes sur le gradient sont plus proches : la proximité sur l'axe est donc une mesure de ressemblance à la fois pour les espèces et pour les relevés.

b) Au coefficient $1/r$ près, chaque espèce a une abscisse y_j égale à la moyenne des abscisses x_i des relevés où on la trouve. Réciproquement, chaque relevé a une abscisse x_i égale à la moyenne des y_j des espèces qu'il contient (Tabl. I). Si r est suffisamment élevé, on peut représenter espèces et relevés sur le même axe : les espèces sont alors au milieu de leurs relevés, et réciproquement. Mais on peut aussi choisir l'axe-espèces de variance unité et l'axe-relevés de variance égale à la valeur propre, auquel cas les points-relevés sont au barycentre des espèces qu'ils contiennent (mais non l'inverse). Si l'axe-relevés est de variance unité, et l'axe-espèces de variance égale à

la valeur propre, ce sont les espèces qui sont au barycentre des relevés où elles sont présentes.

c) Le «facteur» ainsi mis en évidence par l'analyse est un facteur purement abstrait qui ne traduit que la structure principale du jeu de données. Les coordonnées représentent de nouvelles variables synthétiques floro/faunistiques, de variance maximale. Le problème consistera à mettre en évidence *a posteriori* d'éventuelles corrélations de ce facteur abstrait (qu'on appelle quelquefois

«variable latente») avec les variations de variables externes mesurées sur le terrain (voir plus loin paragraphe 4).

La littérature offre des exemples de tentatives *ad hoc* visant le même but que l'AFC. Ainsi Glowacinski (1975), après avoir fixé *a priori* l'ordre des colonnes (= stades d'une succession écologique) de son tableau, permute les lignes-espèces «with the principle of splitting up the columns of each succeeding community to the smallest pos-

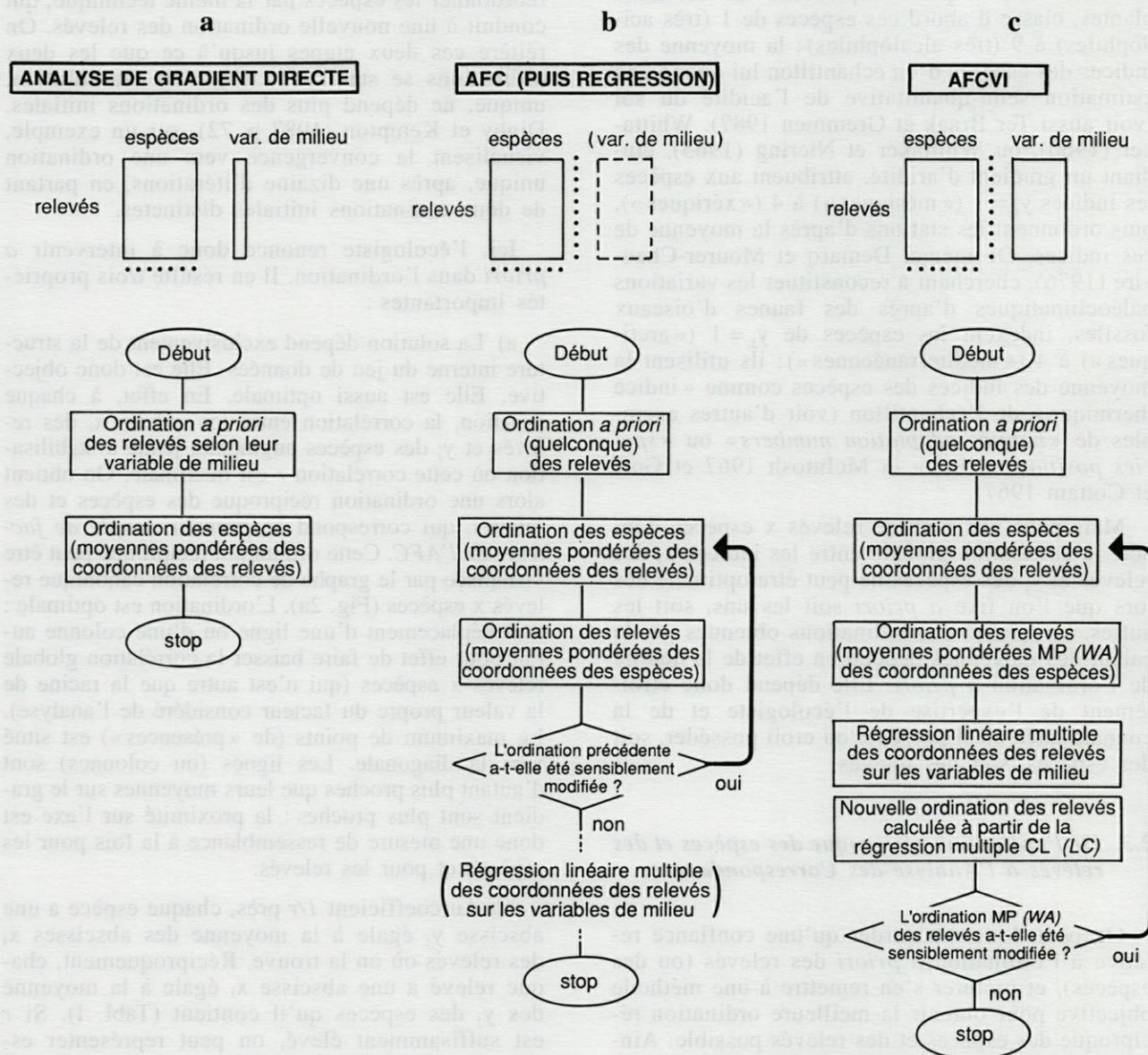


Fig. 1. — Algorithmes correspondant à l'analyse de gradient directe avec *averaging* simple (a), à l'Analyse Factorielle des Correspondances (AFC) éventuellement suivie d'une régression (b), et à Analyse Factorielle des Correspondances sur Variables Instrumentales (AFCVI) ou Analyse Canonique des Correspondances (ACC) (c) (modifié et complété d'après Palmer, 1993).

Algorithms for direct gradient analysis by Weighted Averaging (WA) (a), for Correspondence Analysis (CA) eventually followed by a regression (b), and for Correspondence Analysis with respect to Instrumental Variables (CAIV), or Canonical Correspondence Analysis (CCA) (c).

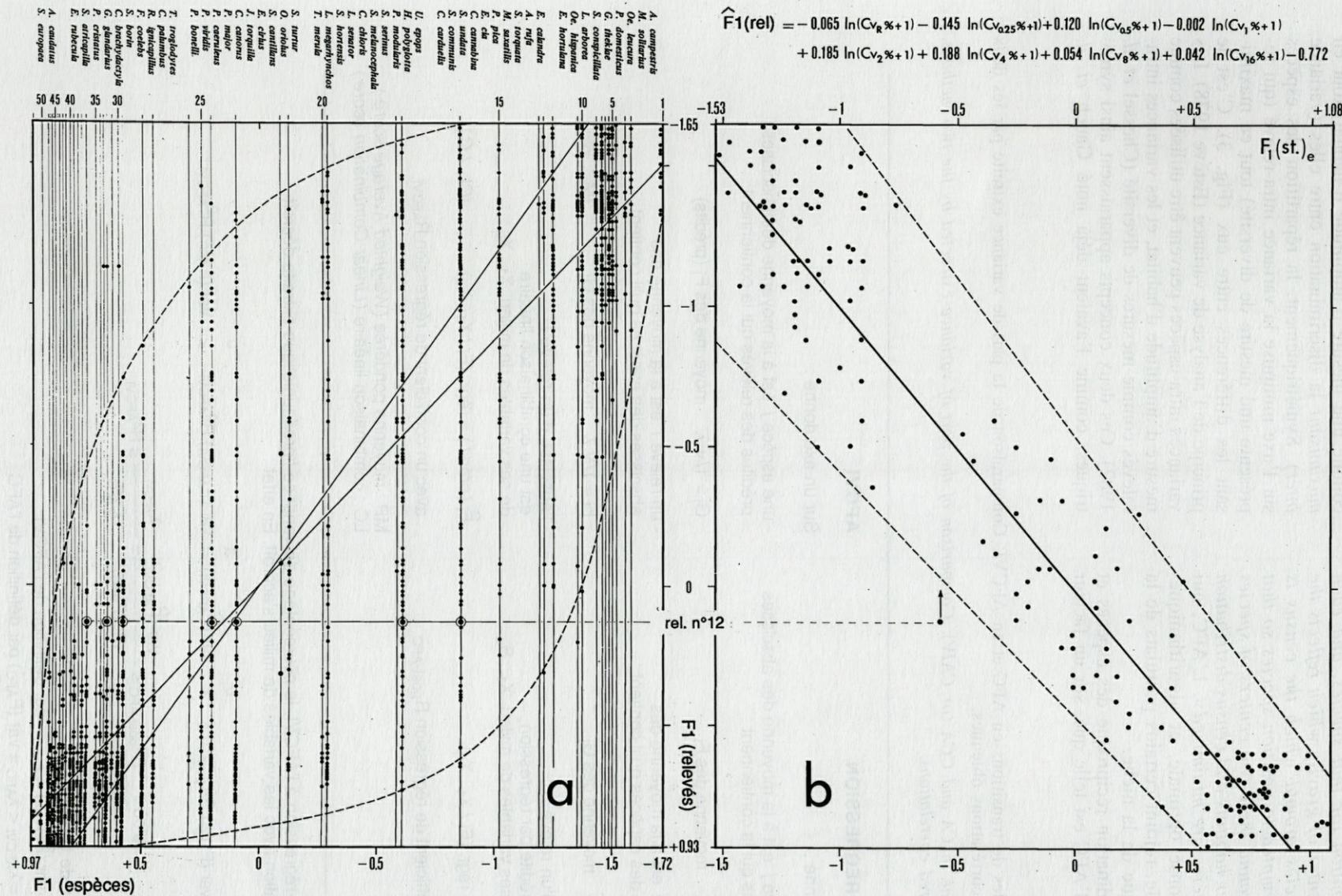


Fig. 2. - a, Exemple de graphe canonique montrant la corrélation entre l'ordination des espèces (axe-espèces horizontal) et l'ordination des relevés (axe-relevés vertical) selon le premier facteur de l'Analyse Factorielle des Correspondances d'un gradient avifaunistique pelouses-forêts. Une seule des lignes correspondant aux 182 relevés (celle du relevé n° 12) a été tracée. b, Graphe de régression multiple de l'ordination des relevés selon le premier facteur de l'AFC sur 8 variables de milieu (recouvrements de 8 strates végétales dans cet exemple) (d'après Prodon et Lebreton 1981).

a. Example of canonical graph showing the correlation between the species ordination (species-axis horizontal) and the sample ordination (sample-axis vertical) corresponding to the first factor of Correspondence Analysis of an avifauna gradient. Being too numerous, the lines corresponding to the 182 samples are omitted.
 b. Multiple regression graph of the relevé ordering, according the first axis of CA, on 8 environmental variables (vegetation cover values in this example).

sible (vertical) extent»; et Cody (1975): «each row... has a center of gravity which reflects the center of distribution for that species, and each column has a center of gravity which reflects the rank of the species around which the census is centered... I attempted to position species so that the variances around these two centers of species distributions over habitats and census distribution over ranked species are minimized». L'AFC, qui fournit une réponse optimale, et algorithmiquement bien définie, rejoint certains postulats de la théorie écologique de la niche.

En effet, l'ordination réciproque des espèces et des relevés par l'AFC est telle que, sur un facteur

donné, la répartition des relevés minimise les variances intra-espèces (qui peuvent alors être utilisées comme une mesure d'amplitude d'habitat) tout en maximisant la discrimination entre elles (variance inter). Symétriquement, la répartition des espèces sur l'axe minimise la variance intra-relevé (qui représente une mesure de diversité) tout en maximisant les différences entre eux (Fig. 3). C'est le principe de l'analyse de variance (Estève 1978). Les variances intra-espèces peuvent être utilisées comme mesure d'amplitude d'habitat, et les variances intra-relevés comme mesures de diversité (Chessel *et al.* 1982). Ces deux concepts apparaissent ainsi symétriques, comme l'avaient déjà noté Gauch *et al.*

Tabl. I. – Formules de transitions en AFC et en AFCVI. Comparaison de la part de variance extraite par les deux méthodes, et des corrélations obtenues.

Transition formulae in CA and CCA (or CAIV). Comparison of the part of variance extracted by the two methods, and of the measured correlations.

AFC, PUIS REGRESSION	AFCVI
Sur un axe donné :	Sur un axe donné :
1) - une espèce <i>j</i> est à la moyenne des abscisses des relevés qui la contiennent :	- une espèce <i>j</i> est à la moyenne des abscisses prédites des relevés qui la contiennent :
$G_j = 1/\sqrt{\lambda}$ moyenne des F_i	$G_j = 1/\sqrt{\lambda}$ moyenne des \hat{F}_i (prédits)
- un relevé <i>i</i> est à la moyenne des abscisses des espèces qu'il contient :	- un relevé <i>i</i> est à la moyenne des abscisses des espèces qu'il contient :
$F_i = 1/\sqrt{\lambda}$ moyenne des G_j	$F_i = 1/\sqrt{\lambda}$ moyenne des G_j [MP, WA]
2) l'abscisse d'un relevé <i>i</i> peut être prédite par régression à partir de ses variables de milieu $X_1...X_q$:	- l'abscisse d'un relevé <i>i</i> est une combinaison linéaire de ses variables de milieu $X_1...X_q$:
\hat{F}_i (prédit) = régr. ($F_i / X_1...X_q$)	\hat{F}_i (prédit) = régr. ($F_i / X_1...X_q$) [CL, LC]
avec un coefficient de régression $R_{\text{post-AFC}}$	avec un coefficient de régression R_{AFCVI}
	MP : moyenne pondérée (Weighted Average score) LC : combinaison linéaire (Linear Combination score)

De l'AFC-puis-régression à l'AFCVI, le pourcentage de variance flor/faunistique extraite décroît, mais la corrélation avec les variables de milieu s'accroît. En effet :

$$\text{var}(\hat{F}_{\text{AFC}}) = \text{var}(F_{\text{AFC}}) R_{\text{post-AFC}}^2 \leq \text{var}(\hat{F}_{\text{AFCVI}}) = \text{var}(F_{\text{AFCVI}}) R_{\text{AFCVI}}^2 \leq \text{var}(F_{\text{AFC}}) R_{\text{AFCVI}}^2$$

$$\text{d'où : } R_{\text{post-AFC}}^2 \leq \frac{\text{var}(\hat{F}_{\text{AFCVI}})}{\text{var}(F_{\text{AFC}})} = \frac{\lambda_{\text{AFCVI}}}{\lambda_{\text{AFC}}} \leq R_{\text{AFCVI}}^2$$

a) par définition de la régression

b) $\text{var}(\hat{F}_{\text{AFC}}) \leq \text{var}(\hat{F}_{\text{AFCVI}}) = \lambda_i$ par définition de l'AFCVI

c) $\text{var}(F_{\text{AFCVI}}) = \lambda_{\text{AFCVI}} < \lambda_{\text{AFC}} = \text{var}(F_{\text{AFC}})$ par définition de l'AFC

(1977). Il est possible d'en donner une représentation graphique simultanée dans un plan factoriel sous forme d'ellipses de dispersion (Thioulouse et Chessel 1992).

2.4. L'aspect multidimensionnel de l'Analyse des Correspondances

S'il existe bien une et une seule meilleure ordination réciproque des espèces et des relevés, il

en existe aussi d'autres possibles, non corrélées à la première, et de coefficients de corrélation moins élevés. L'AFC fournit ces ordinations sous forme de couples de facteurs successifs.

Certains de ces facteurs peuvent être artefactuels. Ainsi, lorsque la première ordination extrait presque toute la variabilité de la matrice, c'est-à-dire lorsque la structure du tableau est essentiellement unidimensionnelle. Le second axe, contraint par le calcul d'être non corrélé au premier, en est alors une fonction quadratique : les points espèces et relevés

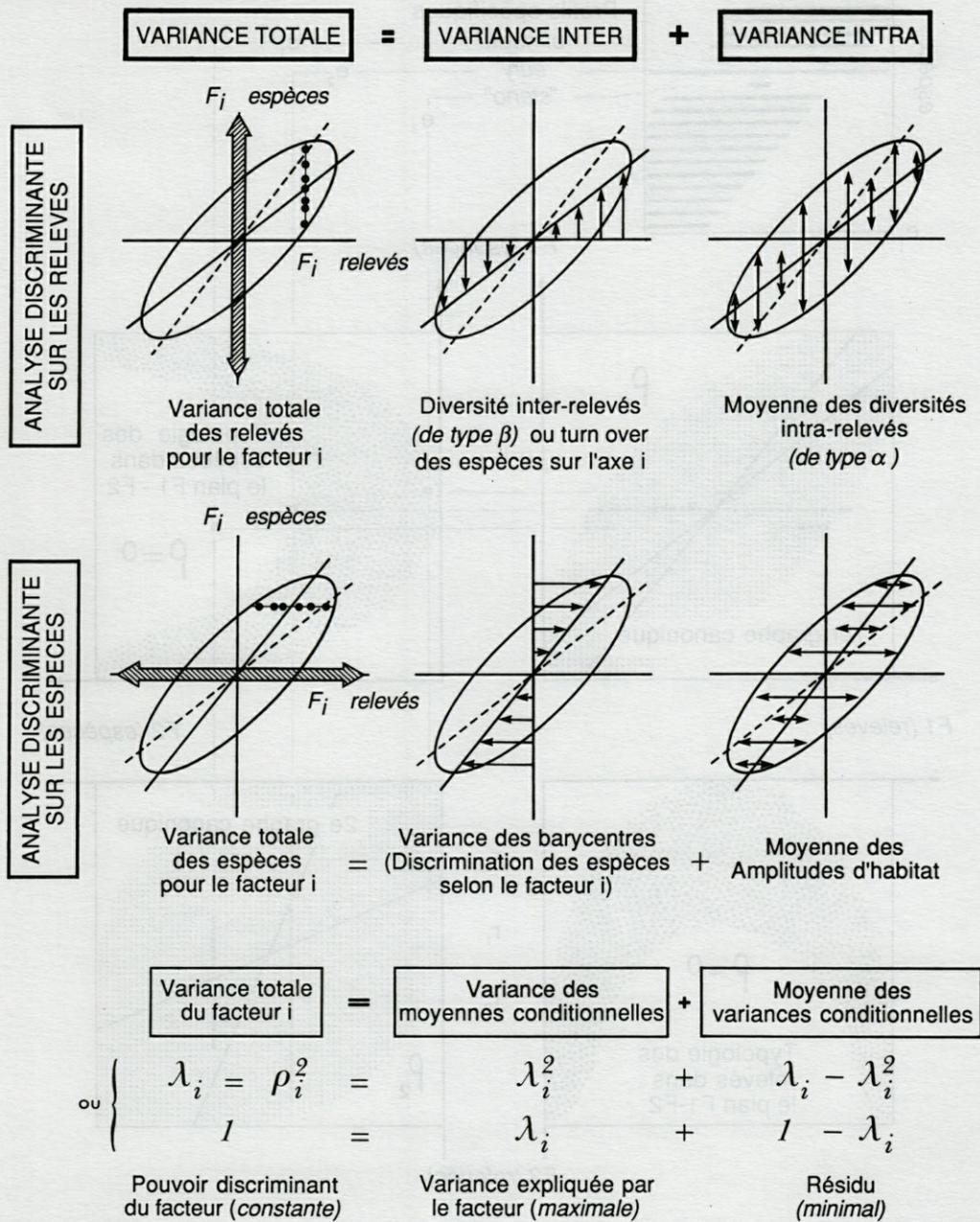


Fig. 3. – L'Analyse Factorielle des Correspondances comme double analyse discriminante (sur les espèces et sur les relevés).

Correspondence Analysis as a double discriminant analysis (on the species and on the samples).

sont disposés en parabole dans le plan des 2 premiers facteurs. C'est l'« effet Guttman » ou « arch effect ». Cet effet peut s'observer même lorsqu'il existe bien un deuxième gradient écologique; il suffit pour cela que la séparation des espèces soit meilleure sur le premier gradient « replié » sous

forme quadratique que sur ce deuxième gradient, lequel se voit alors relégué à un rang ultérieur. La figure 4 illustre les différents modes de représentation selon les premier et deuxième axes-espèces et axes-relevés dans le cas d'un gradient simple avec effet Guttman. Dans ce cas, le second

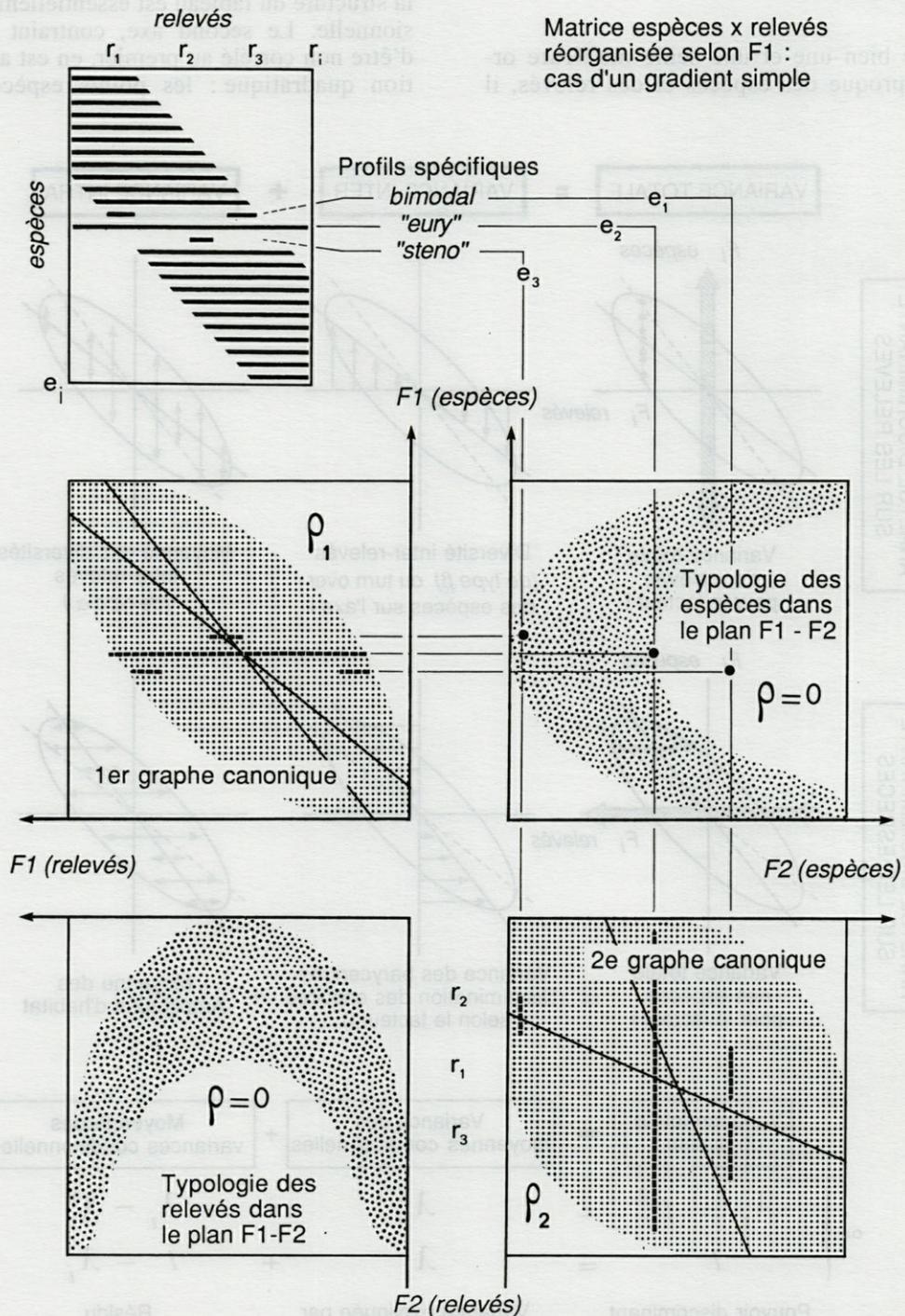


Fig. 4. – Graphes de corrélations canoniques et plans factoriels correspondant aux deux premiers facteurs d'une Analyse Factorielle des Correspondances dans le cas d'un gradient unidimensionnel avec « effet Guttman » (ρ = corrélation correspondant à un graphe donné).

Canonical correlation graphs and factorial planes corresponding to the two first factors of Correspondence Analysis in the case of an unidimensional gradient with an « arch effect » (ρ = correlation corresponding to a given plot).

axe sépare néanmoins au centre deux groupes d'espèces/de relevés situés au milieu du premier axe :

— les points correspondants aux espèces ou aux relevés à faibles variances conditionnelles (respectivement à faibles amplitudes d'habitat et faibles diversités) caractéristiques du milieu du gradient, et de lui seul, sont à l'extérieur de la parabole,

— les espèces/les relevés à forte variance conditionnelle (forte amplitude d'habitat/forte diversité) sont à l'intérieur de la parabole.

Le classement similaire de ces deux groupes écologiquement différents au milieu du premier axe est une conséquence logique de l'ordination d'après la moyenne des profils.

Mais la relation parabolique entre les deux premiers axes n'est pas toujours un artefact de calcul; elle peut simplement résulter de relations non linéaires entre variables du milieu (cf. par exemple Prodon et Lebreton 1981, Balent 1986, Lebreton *et al.* 1988a). C'est pourquoi on ne saurait conseiller l'utilisation de l'AFC « détendancée » (*Detrended Correspondence Analysis*) de Hill et Gauch (1980). Dans cette modification *ad hoc* de l'AFC, une étape supplémentaire est introduite dans le calcul itératif pour supprimer toute relation parabolique entre les deux premiers axes. Lebreton et Yoccoz (1987), Wartenberg *et al.* (1987) et Palmer (1993) dénoncent l'arbitraire d'une telle manipulation, d'ailleurs rendue inutile par l'avènement des analyses factorielles sous contrainte linéaire (cf. paragraphe 4.2.1.).

3. GRADIENTS D'ABONDANCE : L'ANALYSE EN COMPOSANTES PRINCIPALES

Au lieu que les distributions des espèces soient unimodales, leurs optimums se succédant le long du gradient, les variations d'abondance de ces espèces peuvent être continues le long du gradient (c'est-à-dire monotones croissantes ou décroissantes, voire même linéaires après transformation éventuelle). Ce sera le cas si l'on étudie un tronçon de gradient suffisamment court pour que les optimums de la plupart des espèces soient en dehors des bornes de l'échantillonnage. Dans ce cas, on peut raisonner directement en termes de *corrélation* entre les abondances des espèces et le facteur responsable du gradient. Si le facteur écologique responsable est bien identifié, sa corrélation directe avec une variable-espèce donnée est aisée à calculer. S'il ne l'est pas, on recherchera une « variable latente » (autrement dit un « gradient sous-jacent ») qui soit en meilleure corrélation linéaire possible avec l'ensemble des variables-espèces. La méthode naturelle pour cette

analyse indirecte de gradient est l'Analyse en Composantes Principales (ACP).

Ter Braak (1987b, 1988) montre que la recherche de ce facteur latent peut s'effectuer, comme dans l'AFC, par une suite d'itérations. Soit y_{ij} l'abondance d'une espèce j dans un relevé i :

— on commence par une indexation x_i arbitraire, *a priori* et de variance unité, des relevés,

— on calcule, par *régression* linéaire classique, la pente b_j exprimant les variations des abondances des espèces. Les x_i étant centrés, cette équation s'écrit $y_{ij} = b_j \cdot x_i$,

— on calcule les nouveaux indices x_i des relevés par régression inverse de x en y (c'est-à-dire par *calibration*),

— et ainsi de suite jusqu'à stabilisation de b_j et des x_i , qui alors ne dépendent plus des valeurs initiales arbitraires.

Comme dans l'AFC, l'ACP met en évidence des facteurs successifs non corrélés, facteurs qui résument la réponse commune de l'ensemble des espèces à des facteurs communs.

Dans les cas, fréquents en écologie, où se succèdent sur un gradient des espèces à distribution essentiellement unimodales (c'est-à-dire où le tableau des espèces est susceptible d'être réorganisé de façon à ce que les présences soient plus ou moins concentrées sur la diagonale, et les absences dans deux angles) l'utilisation de l'ACP devient inappropriée et l'AFC s'impose. En effet, la distance euclidienne sous-jacente à l'ACP rend cette analyse très sensible aux forts effectifs et aux doubles zéros; il apparaît de fausses similarités entre échantillons par absences d'espèces. L'utilisation de cette distance pour mesurer les similarités entre espèces ou entre relevés d'un gradient induit des distortions curvilinéaires maintes fois dénoncées, en particulier une involution des extrémités d'un gradient simple sur le plan des deux premiers facteurs (effet « fer à cheval ») (Beals 1973, Whittaker et Gauch 1973, Gauch *et al.* 1977, Wartenberg *et al.* 1987) (voir aussi Fig. 5). La distance du khi-2, sous-jacente à l'AFC, est moins sensible à ce problème; son comportement l'apparente d'ailleurs plutôt à une distance simple déduite du coefficient de Jaccard (1901), de la forme $1 - w/u$ (où w est le nombre d'espèces présentes dans les deux relevés comparés, u le nombre d'espèces présentes dans l'un ou l'autre). Cette distance, qui exclut les doubles absences, ou celle assez proche dérivée de l'indice de Sørensen (Bray et Curtis 1957, Beals 1973), a démontré depuis longtemps son efficacité en tant que mesure de similarité dans des tableaux de comptage (Shi 1993). En pratique, on réservera donc l'ACP (en général normée) à l'analyse des tableaux de variables de milieu, dont les variations sont souvent monotones.

4. MISE EN RELATION DES VARIATIONS FLORO/FAUNISTIQUES AVEC LES VARIATIONS DES VARIABLES DE MILIEU

L'échelle des axes factoriels issus de l'analyse factorielle d'un tableau espèces x relevés est une « échelle d'intervalle » où unité et position de l'origine sont arbitraires. Certains auteurs, à la suite de Hill (1973), graduent par commodité leurs axes de 0 (choisi à un extrême de l'axe) à 100 (à l'autre extrême). C'est une graduation abstraite, et il n'y a pas de raison de supposer « *that (the habitat) characteristics change twice as much from 0 et 2 as they do from 0 to 1* » (Routledge 1984). L'ambiguïté est augmentée par l'usage du mot facteur, qui dans l'analyse ne se réfère pas à un « facteur du milieu » (c'est-à-dire une variable externe) explicite. Quelle est donc la signification écologique d'un axe factoriel ? Plus précisément, est-il possible de mettre en correspondance l'ordination floro/faunistique avec des variables de milieu, en ne se contentant pas, comme trop souvent, d'interpréter qualitativement les axes grâce aux connaissances acquises par ailleurs sur l'écologie des espèces ?

Une solution *ad hoc* efficace consiste, sur le plan factoriel issu de l'AFC du tableau floro/faunistique, à associer chaque point-relevé à la valeur (ou la modalité) prise par la variable étudiée, ou à calculer la moyenne des coordonnées des relevés porteurs de la même modalité (Ayyad et Dix 1964). Selon cette représentation, on représentera les variables de milieu (codées en classes) de la même façon que les espèces participant à l'analyse, mais sans influencer celle-ci. Les proximités des espèces et des relevés avec les points représentatifs des classes de variables peuvent alors donner d'utiles indications.

Mais, pour rester dans la logique du modèle linéaire, il est préférable de chercher à relier par régression les ordinations du tableau des espèces aux variables contenues dans le tableau des variables de milieu. Ce problème d'étalonnage des axes de l'analyse floro/faunistiques (AFC ou autre) par rapport à une ou plusieurs variables quantitatives décrivant le gradient (altitude dans le cas d'un gradient en montagne, durée d'émersion moyenne dans une étude de la zone intertidale, volume de la végétation pour un gradient successional, etc.) se pose dans les mêmes termes lorsque les variables sont qualitatives. Dans ce cadre simple de régression d'axes factoriels par rapport à des descripteurs de milieu, plusieurs approches sont possibles ; elles correspondent à une imbrication plus ou moins poussée de l'étape analyse factorielle et de l'étape régression.

4.1. Analyse factorielle, puis régression

Dans cette approche, la régression *suit* l'analyse factorielle du tableau relevés x espèces, qui peut être une Analyse des Correspondances (Fig. 1b) ou toute autre analyse factorielle. On peut ainsi faire la régression d'un axe de l'AFC floro/faunistique sur un facteur du milieu (régression simple après une ordination *ad hoc* : Ayyad et Dix 1964, ou après ACP : Ibanez 1968), ou sur plusieurs variables (régression multiple), ou chercher une combinaison linéaire de plusieurs axes avec plusieurs variables (corrélation canonique). Ainsi Prodon et Lebreton (1981) relient-ils des variations de composition de l'avifaune (AFC) avec des variations de densités de plusieurs strates de végétation. De même Chang et Gauch (1986), après une AFC « détendancée », relient les axes de variations floristiques à des facteurs thermiques, hydriques et édaphiques. Lorsque certains facteurs ont des variations non linéaires, il en résulte une certaine perte d'information, que l'on cherchera à limiter par des transformations appropriées.

Cette approche se résume en trois formules : les deux formules de transition de l'AFC reliant coordonnées des espèces et des relevés (Tabl. I), et la formule de régression simple ou multiple (voir l'exemple de la Fig. 2b). Lorsque le coefficient de corrélation multiple, noté $R_{\text{post-AFC}}$, est élevé (ce qui sera souvent le cas dans une étude de gradient si les variables ont été bien choisies), les coordonnées des relevés obtenus par l'AFC (F_i observés, c'est-à-dire les moyennes pondérées des scores des espèces) et celles prédites par régression (F_i prédits) seront proches. Les axes factoriels pourront alors être considérés comme des variables de milieu synthétiques.

Cette approche est efficace lorsque l'adéquation entre le, ou les, premier(s) axe(s) de l'analyse et le gradient étudié est bonne, ce que seule une planification appropriée des observations est à même de garantir. Sinon, il est possible que des variations imprévues de composition taxonomique viennent dans l'analyse masquer le phénomène d'intérêt principal : des présences d'espèces rares peuvent fragmenter en blocs relevés x espèces un gradient *a priori* régulier ; des variations de paysage peuvent venir perturber l'effet d'un gradient d'altitude, etc. L'éventuelle incapacité à expliquer le premier (ou un des premiers) facteur(s) de l'analyse par les variables de milieu mesurées n'implique donc pas forcément l'absence d'effet floro/faunistique de ces variables.

Il en est un peu de même pour une autre approche consistant à réaliser indépendamment une analyse factorielle pour chacun des deux tableaux (par exemple une AFC pour le tableau floro/faunistique, et une ACP pour celui des variables de milieu), puis à calculer la régression de la pre-

mière ordination sur la deuxième. Cette méthode, souvent efficace, assure une bonne relation entre les deux tableaux puisque la réduction extrait de chacun d'eux les sources non aléatoires de variation. Elle assure une bonne robustesse puisque les processus d'optimisation inhérents aux deux méthodes n'interviennent pas dans la liaison entre les deux tableaux. Mais il n'y a toutefois pas de raisons qu'une combinaison linéaire des variables du deuxième tableau soit précisément celle qui explique au mieux les variations du tableau flor/faunistique (Lebreton *et al.* 1991).

4.2. Régression au sein d'une analyse factorielle : les analyses sous contrainte linéaire

4.2.1. Principe général

Pour être plus sûr de bien détecter l'action des variables de milieu, une imbrication plus poussée de l'analyse factorielle et de la régression est possible. Ainsi, dans le cas de l'AFC, on peut contraindre les coordonnées factorielles des relevés à être une combinaison linéaire de ces variables : c'est « l'Analyse Canonique des Correspondances » (Ter Braak 1986, 1987), ou « Analyse Factorielle des Correspondances par rapport à des Variables Instrumentales » (AFCVI; Lebreton *et al.*, 1988a), ou encore « AFC sous Contrainte Linéaire » (Sabatier 1987). Dans cette analyse, les formules de transition incorporent désormais une régression (Tabl. I). Ces trois formules se prêtent à un calcul itératif, qui donne le premier couple de facteur de l'AFCVI (Fig. 1c), tout comme le double *averaging* conduisait au facteur de l'AFC (cf. ci-dessus 2.3.). La même imbrication des deux étapes, ordination et régression, peut d'ailleurs être réalisée à partir de toute ACP du tableau relevés x espèces, dans ce qu'on appelle du nom générique d'ACPVI (Sabatier 87, Sabatier *et al.*, 1989; Lebreton *et al.*, 1991).

Dans l'AFCVI, à la différence de l'AFC, on obtient explicitement deux coordonnées pour chaque relevé : celle obtenue par moyenne pondérée à partir des coordonnées de ses espèces, comme dans l'AFC ordinaire (F_i observé, c'est-à-dire le *Weighted Average score* (WA) de Palmer 1993), et celle prédite à partir de la combinaison linéaire de ses variables de milieu (F_i prédit, le *Linear Combination score* (LC) de Palmer). Ces deux ordinations sont reliées par un coefficient de corrélation multiple noté R_{AFCVI} . Le gain correspondant à l'imbrication de la régression dans l'analyse elle-même est que R_{AFCVI} est toujours supérieur ou égal à $R_{post-AFC}$ (Tabl. I).

En pratique, les coordonnées prédites par la régression multiple (« LC-scores ») exprimeront bien l'influence des variables de milieu mesurées sur la composition taxonomique. L'AFCVI sera capa-

ble d'« ignorer » les autres sources de variation éventuelles, connues ou non, de cette composition taxonomique. Le gain sera peut-être limité dans le cas de gradient simple ; il pourra être important en cas de données complexes ou fortement bruitées (Palmer 1993). Insistons aussi sur trois autres avantages très importants de l'AFCVI par rapport à l'AFC simple :

- a) une moindre sensibilité à l'« effet Guttman », avec la possibilité de séparer les cas artéfactuels des liaisons réellement curvilinéaires,
- b) une moindre sensibilité aux espèces rares,
- c) une moindre sensibilité à la troncature des gradients qui entraîne un relatif tassement des espèces et des relevés aux extrémités des axes dans l'AFC simple.

Tous ces avantages rendent sans objet le recours aux analyses « détendancées » (*detrended analyses*). Enfin, l'AFCVI facilite l'interprétation des résultats en permettant la représentation dans le même plan, et éventuellement sur la même figure, des points-espèces (typologie des espèces) et des points-relevés (typologie des relevés), comme dans une AFC, et des vecteurs-variables, dont les angles et les longueurs s'interprètent comme en ACP (représentation « *triplot* »).

4.2.2. Nombre, nature, et pertinence des variables

Lebreton *et al.* (1991) rappellent que le rapport des variances $I(AFCVI)/I(AFC)$, ou des inerties, ou encore des traces (= somme des valeurs propres de tous les axes de l'analyse), est comme dans toute ACPVI un rapport de corrélation multivarié ; c'est une moyenne pondérée des carrés des coefficients de corrélation entre les profils de chaque espèce et les variables de milieu. Ce rapport peut être interprété comme le carré d'un coefficient de corrélation. Il constitue donc une mesure *globale* (c'est-à-dire qui tient compte de tous les axes) de l'influence de l'ensemble des variables de milieu sur le peuplement, et donc de la pertinence de leur choix.

En tant que moyenne, ce rapport des variances est relativement insensible au nombre de termes utilisés, c'est-à-dire au nombre d'espèces. Cette robustesse pour un nombre d'espèces élevé est une des propriétés essentielles des ACPVI pour l'analyse des relations espèces-milieu. Les analyses sous contraintes perdent cependant de leur robustesse lorsque le nombre de variables approche le nombre de relevés ; les coefficients des variables dans les régressions multiples inhérentes à la méthode deviennent alors instables, et les rôles respectifs des variables ne peuvent être déterminés (Ter Braak 1986, 1987, Mercier *et al.* 1992). Autrement dit, l'intensité de la contrainte linéaire varie avec le nombre de variables de mi-

lieu. Si leur nombre dépasse celui des espèces, il n'y a plus de contrainte du tout, et l'AFCVI reproduit les résultats de l'AFC. S'il n'y a qu'une seule variable, l'AFCVI reproduit le résultat d'un simple *averaging* direct, et les espèces sont alors à la moyenne de leur variable de milieu.

Lorsque les relevés sont groupés en n classes (par exemple selon 4 types de sol), on a alors une variable qualitative à n classes ou modalités (soit 4 variables en 0/1 dans notre exemple). Une telle variable qualitative, qui constitue un « facteur » au sens de l'Analyse de Variance (on parlera souvent de « covariables »), pourra être utilisée par elle-même dans l'analyse, ou en conjonction avec d'autres variables qualitatives ou quantitatives.

4.2.3. Analyses « partielles »

Supposons que l'on souhaite étudier l'effet d'un groupe de variables B (par exemple les différentes modalités de la variable qualitative exposition), sachant qu'un autre groupe de variables A (par exemple l'altitude et le type de végétation) influence déjà le tableau relevés \times espèces dans une étude d'avifaune en montagne. Il s'agit donc de déterminer la part de variation dans le tableau relevés \times espèces expliquée par B, une fois l'influence de A prise en compte. Inversement, B peut être un ensemble de covariables dont on souhaite éliminer l'effet, et A le tableau des variables quantitatives d'intérêt principal. Comme avec les techniques classiques de régression partielle, on peut réaliser des analyses canoniques partielles, ou des ACPVI Partielles (Ter Braak, 1988; voir également Yoccoz et Chessel 1988), c'est-à-dire conditionnellement à une variable multivariée donnée. On parlera par exemple de l'ACPVI du tableau X par rapport à B/A (« B sachant A »), c'est-à-dire par rapport à B une fois tenu compte de l'influence de A, ou l'inverse. La décomposition en effets successifs lorsqu'on a affaire à des facteurs multiples est développée par Sabatier *et al.* (1989). C'est une des voies les plus prometteuses des ACPVI.

4.2.4. AFCVI et Analyse discriminante

Parmi les Analyses par rapport à des variables instrumentales, l'AFCVI a un statut très particulier pour l'écologiste du fait de ses relations étroites avec certaines analyses discriminantes classiquement utilisées en écologie des communautés. L'usage de l'analyse discriminante est licite pour comparer les habitats de diverses espèces dès lors que chaque relevé concerne un seul individu d'une espèce donnée, accompagné de ses variables de milieu (voir par ex. Nudds 1982). Chaque relevé constitue donc un « individu statistique » dans la classe espèce considérée. L'analyse discriminante entre les classes ainsi constituées fournit des

axes discriminants qui sont des variables de milieu synthétiques, c'est-à-dire des combinaisons linéaires des variables de milieu. Chaque espèce est placée à la moyenne des relevés où elle est présente, et le chevauchement entre les distributions des différentes espèces est minimisé.

Ce sont ces propriétés que l'AFCVI généralise au cas de relevés comportant plusieurs espèces. L'AFCVI fournit la combinaison linéaire de variables du milieu qui maximise la dispersion des espèces, c'est-à-dire la discrimination entre elles, puis la suivante non corrélée à la première, etc. (Ter Braak 1987a,b). Dans l'analyse, chaque présence d'une espèce donnée constitue un individu statistique qui se voit attribuer les variables de milieu du relevé concerné (voir Tabl. I in Lebreton *et al.*, 1988a). A noter, *contra* Nudds (1982), que les tests associés à l'analyse discriminante ne sont plus valides du fait de l'interdépendance induite par la duplication des variables de milieu inhérente à cette analyse.

4.3. Autres méthodes pour relier deux tableaux : les analyses symétriques

En insérant la régression au sein même de l'analyse factorielle, l'approche précédente relie directement le tableau relevés \times espèces et le tableau relevés \times milieu par une analyse factorielle d'un type particulier. A ce titre elle ne constitue qu'une des analyses qui permettent de relier deux tableaux de données portant sur les mêmes unités statistiques, c'est-à-dire les mêmes relevés. Parmi ces analyses, dont on pourra trouver une revue synthétique dans Mercier (1991), certaines font jouer un rôle symétrique aux deux tableaux.

4.3.1. Analyse des corrélations canoniques

Une première méthode consiste à chercher les couples de facteurs (il y en a autant que de relevés), de corrélation maximale, formés d'une combinaison linéaire des variables du premier tableau, et d'une combinaison linéaire des variables du second tableau : c'est l'Analyse des Corrélations Canoniques. Cette méthode représente de façon théoriquement optimale les dépendances entre les deux tableaux. Mais, en pratique, ceci n'est vrai que quand le nombre de relevés est très élevé par rapport au nombre de taxons et au nombre de variables de milieu (Hitier 1993). Sinon, apparaît un problème de robustesse qui se traduit par une nette tendance à l'instabilité des résultats (par exemple Gauch et Wentworth 1976, Gittins 1985); on pourra toujours s'en convaincre en comparant les résultats obtenus par les analyses séparées des deux moitiés d'un même jeu de données. De plus, les relations entre les abondances des espèces et les variables de milieu sont en général non li-

néaires, rendant la recherche directe de corrélations problématique (Austin 1968).

4.3.2. Analyses de Costructure (ou de Co-inertie)

Les Analyses de Costructures, introduites dans un cas simple par Tucker (1958), ont récemment été systématisées et généralisées par Chessel et Mercier (1993) dans le contexte de l'analyse de données écologiques. On peut en comprendre le principe en simplifiant à l'extrême l'approche par corrélation. Réduisons tout d'abord le tableau relevés x taxons Y et, indépendamment, le tableau relevés x variables de milieu X chacun à leur plus important axe de variation; celui-ci sera le premier axe factoriel d'une analyse appropriée, souvent une AFC pour Y , et une ACP normée pour X . Soient y et x ces deux axes; ils sont bien représentatifs de la variabilité de chaque tableau car, parmi les combinaisons linéaires possibles des variables des tableaux respectifs Y et X , ce sont par définition celles de variance maximale. On peut alors étudier la liaison entre y et x sur un simple graphique et, si elle est raisonnablement linéaire, la quantifier par un coefficient de corrélation r_{xy} . Mais on peut chercher, dans un second temps, à modifier les coefficients des combinaisons linéaires de x et y de façon à augmenter r_{xy} . Ici, x et y étant chacun de variance maximum (σ_x^2 et σ_y^2), toute modification de l'un ou de l'autre diminuera sa variance. En améliorant la liaison entre x et y , on amoindrit donc dans une certaine mesure leur représentativité en tant que résumés respectifs de X et Y . Un compromis consiste alors à maximiser le produit $\sigma_x \sigma_y r_{xy}$. C'est l'objet de l'Analyse de Costructure, qui cherche à rapprocher les F1 des deux ACP sans détruire trop leur représentativité en tant que résumés linéaires de X et Y . Le même raisonnement se poursuit pour les axes factoriels successifs. Les facteurs obtenus, tout comme ceux des ACP originelles, sont par construction bien corrélés avec les variables de Y et X respectivement. Pour cette raison, les analyses de costructure sont relativement insensibles au nombre de taxons et de variables de milieu. En comparaison, les ACPVI sont peu robustes pour un nombre élevé de variables dans X (le tableau relevé x variables de milieu dans notre cas). Elles supporteront par contre éventuellement plus de taxons que de relevés dans Y (voir un exemple dans Lebreton *et al.*, 1991).

Les deux approches symétriques ci-dessus peuvent être critiquées précisément pour leur symétrie. Un des avantages des ACPVI par rapport à elles est en effet de prendre en compte explicitement l'influence supposée des variables de milieu sur la composition spécifique, et non l'inverse. Le tableau floro/faunistique constitue la structure à expliquer, l'autre tableau contient les variables explicatives. Il est vrai que dans certains cas, on

peut être aussi intéressé par les possibilités de prédiction des variables de milieu à partir du peuplement que l'inverse. De plus, si les méthodes symétriques s'adressent théoriquement à des tableaux X et Y obtenus par échantillonnage entièrement aléatoire des relevés, la recherche d'une liaison milieu - espèce conditionnellement aux variables de milieu (par exemple lorsque les relevés sont organisés selon un plan d'expérience) reste acceptable par les méthodes symétriques, avec l'avantage d'une robustesse très appréciable en cas de nombres de variables de milieu ou de taxons élevés (Tabl. II) (Chessel et Mercier 1993, p. 34).

Advantages and drawbacks of the three categories of multivariate analyses coupling two matrices (Y : species matrix, X : environmental variable matrix), in particular with respect to robustness.

Méthodes de couplage de Y et X	ANALYSE CANONIQUE	ACPVI (s.l.) de Y/X	ANALYSES DE COSTRUCTURE
Robustesse pour un nombre d'espèces élevées	NON (1,2)	OUI (1, 2)	OUI (1)
Robustesse pour un nombre de variables de milieu élevé (a)	NON	NON	OUI (1)
Flexibilité concernant la nature des mesures dans Y et X	potentiellement OUI	OUI (3)	potentiellement OUI
Possibilités d'analyses conditionnelles (ou "partielles")	OUI	OUI (3, 4)	potentiellement OUI
Possibilités d'inférences	OUI (5)	OUI (2)	OUI (2)

Note :

(a) une variable de milieu qualitative à m modalités compte pour $m-1$ variables

Références : 1 - Chessel et Mercier (1993); 2 - Hitier (1993); 3 - Sabatier *et al.* (1989); 4 - Ter Braak (1988); 5 - Seber (1984)

peut être aussi intéressé par les possibilités de prédiction des variables de milieu à partir du peuplement que l'inverse. De plus, si les méthodes symétriques s'adressent théoriquement à des tableaux X et Y obtenus par échantillonnage entièrement aléatoire des relevés, la recherche d'une liaison milieu - espèce conditionnellement aux variables de milieu (par exemple lorsque les relevés sont organisés selon un plan d'expérience) reste acceptable par les méthodes symétriques, avec l'avantage d'une robustesse très appréciable en cas de nombres de variables de milieu ou de taxons élevés (Tabl. II) (Chessel et Mercier 1993, p. 34).

En terme de flexibilité (possibilité d'adaptation à différents types de données), les trois approches sont équivalentes. Il est ainsi parfaitement possible de réaliser une analyse canonique entre un tableau relevés x espèces en présences-absences, justifiable par exemple d'une AFC, et des variables de milieu qualitatives, et ceci en jouant sur les transformations préalables et les pondérations (*cf.* 5.2.). En ce qui concerne les possibilités de tests, l'analyse canonique l'emporte, au prix d'hypothèses restrictives sur la nature des distributions sous-jacentes (normalité, linéarité...). Pour les autres analyses, des tests de permutation approchés sont possibles, soit par tirages au hasard (Ter Braak 1988), soit en utilisant une approximation de la distribution de permutation (Hitier 1993). Pour l'instant, les meilleures possibilités sont celles du

logiciel CANOCO (Ter Braak 1987c) dans le cadre de l'AFCVI.

Les avantages en matière de robustesse et pertinence des trois approches analyse canonique, ACPVI, analyses de costructure sont résumées dans le Tableau II.

5. CHOIX D'UNE STRATÉGIE POUR UNE ÉTUDE ESPÈCES X MILIEU

5.1. Planification, choix des variables

Les trois principales décisions à prendre lors de la mise en place d'une étude de relations taxons - milieu sont les suivantes (d'après Lebreton et Yoccoz, 1987 complété; voir aussi Wiens 1989) :

1) Quel ensemble de taxons échantillonner, et avec quelle méthode de mesure d'abondance ou de simple présence ?

2) Quelles variables de milieu mesurer ?

3) Selon quelle planification des relevés ?

Les réponses à ces questions dépendent en partie des objectifs et des questions biologiques, et en partie de considérations statistiques dont Frontier (1983) offre une bonne vue d'ensemble. L'ensemble de taxons échantillonné est généralement un groupe systématique bien défini, ou un ensemble d'espèces sensibles aux mêmes variations de milieu, susceptibles d'interactions, et échantillonnables par une même méthode (Coléoptères capturés par des pots de Barber, passereaux forestiers chanteurs, poissons pêchés au chalut, etc). Dans des études de groupes taxonomiques homogènes en relation avec la théorie de la niche, un plan d'échantillonnage selon un ou plusieurs gradients sera souvent retenu, alors qu'en biologie de la conservation, un plan d'échantillonnage systématique sera souvent approprié pour couvrir la région d'étude d'une façon homogène.

Choix des variables de milieu et planification des relevés sont d'ailleurs intimement liés. En effet par variables de milieu, il faut entendre toute variable caractérisant les relevés, y compris des variables qualitatives qui sont en fait des facteurs au sens analyse de variance du terme. Ainsi par exemple on pourra choisir 30 relevés dans un type de forêt, et 30 autres relevés dans un autre type, ou bien associer une variable qualitative (le versant) à une variable quantitative (l'altitude). Pour éliminer l'effet d'une variable, il est bien sûr toujours possible de limiter en conséquence l'échantillonnage, mais il sera souvent intéressant de pouvoir « éliminer » ces effets *a posteriori* (cf. par ex. Yoccoz et Chessel 1988, Dolédec et Chessel 1989). Ainsi, dans la planification d'une typologie

de stations forestières, pour étudier l'effet du type de sol sur la composition floristique, on pourra :

— soit minimiser l'action du facteur altitude en répartissant strictement tous les relevés à l'intérieur d'un même étage altitudinal,

— soit procéder à une analyse, par exemple sur des variables instrumentales définissant les types de sol, conditionnellement à l'altitude (§ 4.2.3.).

C'est ainsi que dans une étude d'invertébrés fluviatiles, Richardot-Coulet *et al.* (1986; voir également Lebreton *et al.*, 1988b) caractérisent des relevés par une série de facteurs qualitatifs croisés, ce qui permet par exemple de tester l'effet de différentes méthodes d'échantillonnage compte-tenu d'autres sources de variation de composition spécifique comme la saison, le courant ou la station (Sabatier *et al.* 1989).

On veillera aussi à limiter au maximum, par une planification soignée, les confusions de facteurs. Par exemple des comparaisons d'avifaunes entre régions par des méthodes de points d'écoute ne pourront faire l'économie, si les différences à mettre en évidence sont subtiles, d'un contrôle de l'effet observateur : croiser observateur et régions, chaque observateur effectuant une part des relevés de chaque région est idéal. Un seul observateur partout semblerait encore préférable, mais le risque est alors d'introduire une confusion région-saison en étalant les relevés dans le temps !

Il est bien sûr recommandé de réduire le nombre de variables de milieu, à la fois pour les raisons de robustesse déjà discutées, mais aussi pour ménager les possibilités de mise en évidence d'effets car au delà de 4 variables des interactions complexes empêchent souvent la construction d'un modèle raisonnable.

D'après Chessel (1992), études théoriques et analyses concrètes permettent d'énoncer d'autres principes importants à prendre en considération dans la construction des plans d'échantillonnage : miniaturisation du sondage élémentaire (pour tenir compte de l'hétérogénéité du milieu et pouvoir augmenter le nombre des échantillons), emboîtement des mesures (plans de type hiérarchique), « abandon impératif de l'échantillonnage au hasard (constat d'ignorance et entretien de celle-ci) », contrôle des structures à plusieurs échelles... Les problèmes d'échelle sont au cœur de l'étude des relations espèces-milieu, depuis la définition des questions posées jusqu'à l'interprétation des résultats. En dépendent en effet tous les patrons de variabilité mesurés (Wiens 1989).

5.2. Choix des transformations, des métriques, et de la méthode de couplage

Une fois les tableaux de données obtenus, trois questions concernant l'analyse statistique doivent être examinées :

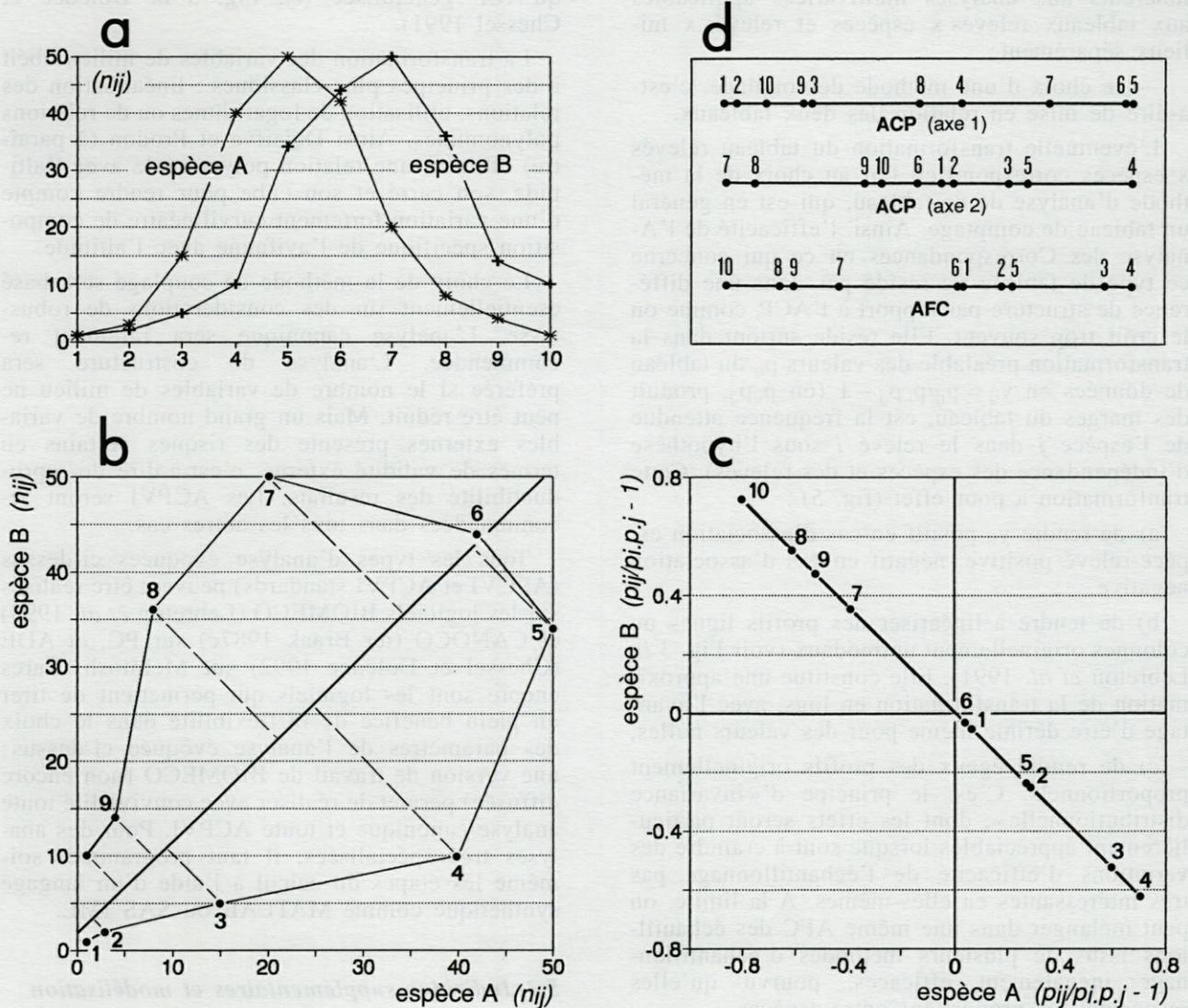


Fig. 5. – a, Exemple théorique d'un transect de 10 relevés ($i = 1$ à 10) contenant 2 espèces j_1 (« espèce A ») et j_2 (« espèce B ») en proportion variable (abondances de A : 1 ; 3 ; 15 ; 40 ; 50 ; 42 ; 20 ; 8 ; 4 ; 1, soit $n_1 = 184$, abondances de B : 1 ; 2 ; 5 ; 10 ; 34 ; 44 ; 50 ; 36 ; 14 ; 10, soit $n_2 = 206$).

b, Les 10 points-relevés dans l'espace défini par les deux axes-espèces, sans transformation des abondances brutes n_{ij} . L'ordination selon l'axe principal d'allongement du nuage est une approximation du premier facteur de l'ACP ; dans cet exemple, c'est un axe d'abondance trivial. L'ordination selon l'axe perpendiculaire au précédent est une approximation du deuxième facteur de l'ACP.

c, Les 10 points-relevés dans l'espace défini par les deux axes-espèces, après transformation des abondances en $y_{ij} = p_{ij}/p_i.p_j - 1$ (avec $p_{ij} = n_{ij}/n_{..}$) : même sans les pondérations par les p_i et les p_j propres à l'AFC, l'ordination obtenue est une bonne approximation de l'AFC.

d, ordination des 10 points-relevés après ACP et après AFC. On remarque une certaine analogie entre l'ordination obtenue par AFC et celle obtenue sur le deuxième axe de l'ACP (inspiré et modifié de Wartenberg *et al.* 1987).

a, Theoretical example of 10 samples ($i = 1$ à 10) containing 2 species j_1 (species A) and j_2 (species B) in different proportions.

b, The 10 relevés in the space defined by the 2 species-axes, with the raw species abundance values. The ordering on the principal axis of the scatter of points is an approximation of the first axis of PCA. The ordering on the second principal axis is an approximation of the second axis of PCA.

c, The 10 relevés in the space defined by the 2 species-axes, after transformation of the abundances in $y_{ij} = p_{ij}/p_i.p_j - 1$ (with $p_{ij} = n_{ij}/n_{..}$). Even without the weighting by p_i and p_j , the resulting ordering is a good approximation of the CA ordering.

d, One-dimensional ordering of the relevés after PCA (first and second axes) and after CA.

— le choix d'une transformation des données, et le choix des pondérations. Ces deux choix sont inhérents aux analyses multivariées applicables aux tableaux relevés x espèces et relevés x milieux séparément;

— le choix d'une méthode de couplage, c'est-à-dire de mise en relation des deux tableaux.

L'éventuelle transformation du tableau relevés x espèces correspond en fait au choix de la méthode d'analyse de ce tableau, qui est en général un tableau de comptage. Ainsi, l'efficacité de l'Analyse des Correspondances en ce qui concerne ce type de tableau ne réside pas dans une différence de structure par rapport à l'ACP, comme on le croit trop souvent. Elle réside surtout dans la transformation préalable des valeurs p_{ij} du tableau de données en $y_{ij} = p_{ij}/p_{i.}p_{.j} - 1$ (où $p_{i.}p_{.j}$, produit des marges du tableau, est la fréquence attendue de l'espèce j dans le relevé i sous l'hypothèse d'indépendance des espèces et des relevés). Cette transformation a pour effet (fig. 5) :

a) de rendre y_{ij} positif en cas d'association espèce-relevé positive, négatif en cas d'association négative,

b) de tendre à linéariser des profils lignes ou colonnes originellement unimodaux (voir Fig. 3 in Lebreton *et al.* 1991). Elle constitue une approximation de la transformation en logs, avec l'avantage d'être définie même pour des valeurs nulles,

c) de rendre égaux des profils originellement proportionnels. C'est le principe d'« invariance distributionnelle », dont les effets seront particulièrement appréciables lorsque sont à craindre des variations d'efficacité de l'échantillonnage pas très intéressantes en elles-mêmes. A la limite, on peut mélanger dans une même AFC des échantillons issus de plusieurs méthodes d'échantillonnage inégalement efficaces, pourvu qu'elles respectent les proportions entre espèces.

En revanche, cette transformation peut être inopportune lorsque les variations d'abondance absolue présentent un intérêt en elles-mêmes, ou lorsque l'accent doit être mis sur les espèces les plus abondantes. Diverses autres variantes sont possibles pour des tableaux de comptage. Elles peuvent être vues comme des solutions intermédiaires entre ACP et AFC : par exemple les transformations $y_{ij} = p_{ij} - p_{i.} - p_{.j} + 1$ (ACP bicentree), ou $y_{ij} = p_{ij}/p_{.j} - p_{i.}$ (Analyse non symétrique des Correspondances). L'exploration des performances pratiques de ces variantes est en cours.

Le choix d'une transformation préalable du tableau de comptage est donc essentiel. Celui des pondérations est moins critique (Blanc *et al.* 1976) et résultera souvent du choix de la méthode sélectionnée pour la transformation qu'elle implique. Ainsi, dans le cas de l'AFC, une pondération uniforme des relevés pourra être souhaitable pour donner un sens à des tests de permutations. Quoi

qu'il en soit, les étapes qui suivent transformations et pondérations peuvent être décrites en tant qu'ACP généralisée (cf. Fig. 3 in Dolédec et Chessel 1991).

La transformation des variables de milieu obéit à des principes plus classiques : linéarisation des relations, utilisation de logarithmes ou de relations polynomiales. Ainsi Dejaifve et Prodon (à paraître) utilisent une relation polynomiale avec l'altitude, son carré et son cube pour rendre compte d'une variation fortement curvilinéaire de composition spécifique de l'avifaune avec l'altitude.

Le choix de la méthode de couplage sera basé essentiellement sur des considérations de robustesse. L'analyse canonique sera rarement recommandée. L'analyse de costructure sera préférée si le nombre de variables de milieu ne peut être réduit. Mais un grand nombre de variables externes présente des risques certains en termes de validité externe, c'est-à-dire de reproductibilité des résultats. Les ACPVI seront recommandées dans tous les autres cas.

Tous les types d'analyse évoqués ci-dessus (AFCVI et ACPVI standards) peuvent être réalisés sur les logiciels BIOMEKO (Lebreton *et al.* 1990) et CANOCO (ter Braak 1987c) sur PC, et ADE (Chessel et Dolédec 1992) sur McIntosh. Rares encore sont les logiciels qui permettent de tirer un plein bénéfice de la flexibilité dans le choix des paramètres de l'analyse évoquée ci-dessus ; une version de travail de BIOMEKO (non encore diffusée) permet de réaliser avec convivialité toute analyse canonique et toute ACPVI. Pour des analyses très spécialisées, il faut programmer soi-même les étapes du calcul à l'aide d'un langage synthétique comme MATLAB ou SAS IML.

5.3 Individus supplémentaires et modélisation

En analyse multivariée, où le modèle linéaire est partout sous-jacent, on oublie trop que le principal intérêt d'une équation de régression $y = a + bx$ est de pouvoir prédire y lorsque seul x est connu. L'analyse des relations espèces - milieux ne fait pas exception. Ainsi peut-on considérer la représentation obtenue dans une AFC suivie d'une régression, ou dans une AFCVI, comme un modèle des relations espèces milieu. La composition taxonomique permet de prédire les coordonnées de relevés supplémentaires par simple *averaging* des coordonnées des espèces. La mesure des variables de milieu permet une autre prédiction des coordonnées de ces mêmes relevés, qui peut être en forte liaison linéaire avec la première (cf. par exemple Prodon et Lebreton 1981). Comme toujours, ce sont les cas où le modèle est pris en défaut qui sont les plus instructifs. Ainsi, après perturbations, peut-on observer des distorsions entre la composition faunistique et les va-

riables de milieu (dans cet exemple la couverture végétale). Cette distortion est due ici à des effets de rémanence des avifaunes forestières après incendie (Prodon et Lebreton 1983).

La même approche pourrait être utilisée pour effectuer une validation croisée, c'est-à-dire pour obtenir, sur un lot de relevés constituant une réplique du lot de départ, une estimation de la corrélation entre l'ordination floro/faunistique et les variables de milieu indépendante, et donc de validité externe maximale. C'est une manière de rétablir la robustesse des ACPVI en cas de doute dû à un grand nombre de variables de milieu.

6. DISCUSSION

« Par le choix d'une méthode, par la direction donnée à l'interprétation, l'analyse des données est modélisatrice »

Estève 1978

Parmi les méthodes d'ordination utilisées en écologie, on distingue classiquement deux catégories :

– les analyses directes (par ex. Whittaker 1967) ou « *environmental ordinations* » (Austin 1968) relient la distribution des espèces à des variables externes mesurables et d'importance reconnue. C'est l'« *extrinsic analysis* » de William et Lance (1968), ou « *ecological analysis* » (Whittaker et Gauch 1978). C'est le domaine des profils écologiques, de l'*averaging* simple ou de la régression logistique, de la régression simple ou multiple sur variables de milieu, espèce par espèce. Une telle approche peut être pertinente à condition que ces variables de milieu rendent bien compte de la variabilité floro/faunistique observée (Mercier *et al.* 1992). Mais l'analyse ignore les éventuelles covariations entre espèces, c'est-à-dire l'aspect intégré du peuplement.

— les analyses indirectes sont basées, comme dans la phytosociologie classique, sur une analyse du seul tableau relevés x espèces (« *intrinsic analysis* » de William et Lance 1968, « *taxonometric analysis* » de Whittaker et Gauch 1978 ou encore « *continuum analysis* » de McIntosh 1958). Pour cela, l'AFC et ses variantes éventuelles sont un des outils les mieux adaptés. L'influence éventuelle de variables de milieu, par exemple sous forme de gradients écologiques, peut alors être inférée *a posteriori*.

Mais la distinction entre ces deux types d'analyse tend à s'estomper. Si une démarche en deux temps (ordination du tableau floro/faunistique, puis corrélation d'un ou plusieurs axes avec des descripteurs externes) distingue encore clairement

analyse interne et externe, les analyses sous contraintes comme l'AFCVI (pourtant qualifiées de directes par Ter Braak 1986), et plus encore les analyses symétriques, mêlent les deux approches. Elles rapportent en effet les principaux axes de la variabilité floro/faunistique à la variabilité du milieu, tout en fournissant des mesures de l'intensité de la relation entre les deux. Ainsi les multiples possibilités de couplage des deux tableaux réconcilient-elles méthodes d'ordinations directes et indirectes.

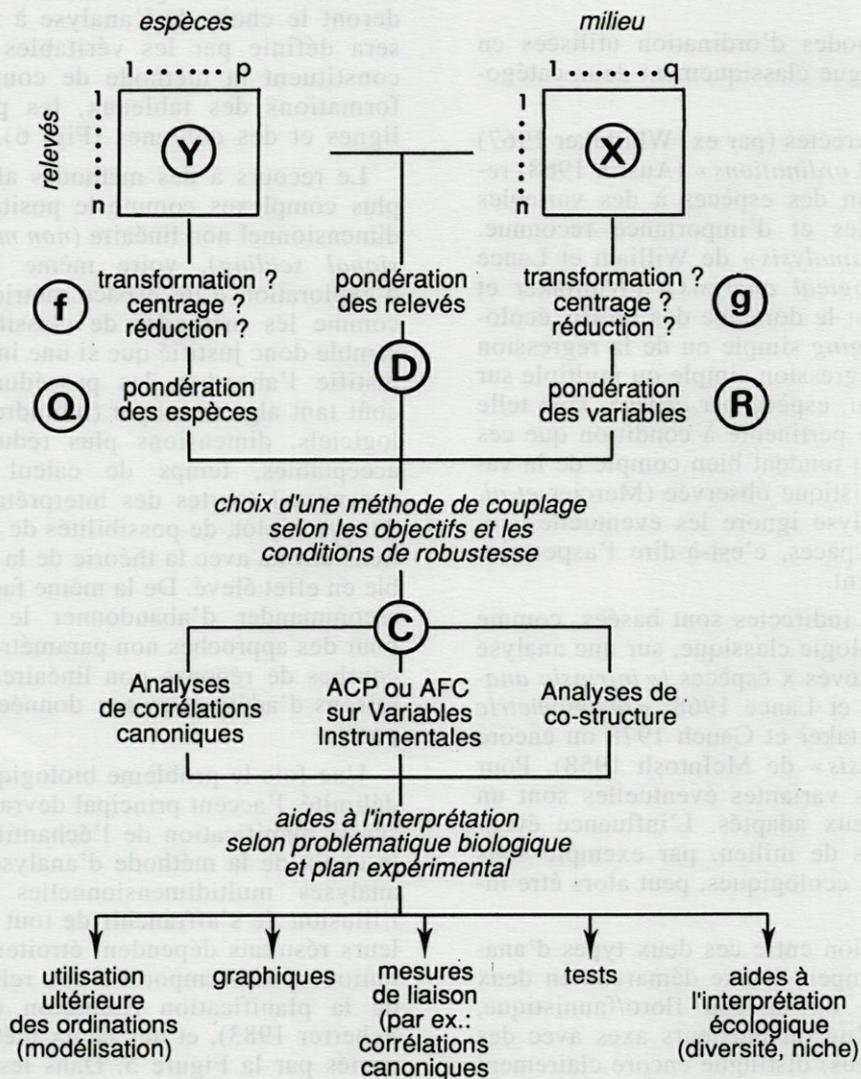
Comme le pressentaient Blanc *et al.* (1976) la flexibilité des analyses multivariées linéaires est loin d'avoir été épuisée. Permettant de se libérer des limites de telle ou telle méthode rigide, cette flexibilité rend possible en effet une meilleure mise en adéquation entre problématique et données d'une part, et analyse statistique d'autre part. Ce sont la nature des données (par exemple espèces notées en présence-absence), les considérations de robustesse (par exemple un nombre élevé de taxons), et surtout les questions biologiques (par exemple étude de l'effet de perturbations sur un gradient de milieux) qui guideront le choix de l'analyse à réaliser. Celle-ci sera définie par les véritables paramètres que constituent la méthode de couplage, les transformations des tableaux, les pondérations des lignes et des colonnes (Fig. 6).

Le recours à des méthodes algorithmiquement plus complexes comme le positionnement multidimensionnel non linéaire (*non metric multidimensional scaling*), voire même à des méthodes d'exploration d'un espace métrique non linéaires comme les méthodes de classification, ne nous semble donc justifié que si une inadéquation claire justifie l'abandon des procédures linéaires. Le coût tant algorithmique (moindre disponibilité en logiciels, dimensions plus réduite des tableaux acceptables, temps de calcul important) que conceptuel (pertes des interprétations graphiques de type biplot, de possibilités de modélisation, des liens étroits avec la théorie de la niche) nous semble en effet élevé. De la même façon, on ne saurait recommander d'abandonner le modèle linéaire pour des approches non paramétriques ou pour des courbes de réponse non linéaires sans de solides raisons d'adéquation aux données et au problème posé.

Une fois le problème biologique à étudier bien délimité, l'accent principal devra porter davantage sur la planification de l'échantillonnage, que sur le choix de la méthode d'analyse elle-même. Les analyses multidimensionnelles peuvent donner l'illusion de s'affranchir de tout *a priori*. En fait, leurs résultats dépendent étroitement de la distribution spatio-temporelle des relevés, c'est-à-dire de la planification (Lebreton et Yoccoz 1987, Scherrer 1983), et des choix méthodologiques résumés par la Figure 5. Dans les analyses à deux

tableaux, le choix des variables externes explicatives est, nous l'avons vu, particulièrement délicat. Or ces variables sont en nombre infini. Pour leur choix, il sera indispensable de faire appel aux connaissances *a priori* de l'écologiste et à son intuition. Il est aussi recommandé de se soucier de l'interprétation des résultats dès la planification : «*I suspect that a lot of unintelligible linear combinations have been lost somewhere between analysis and publication*» (Johnsson 1981). Il est donc clair que toutes les connaissances *a priori*, même intuitives, des espèces, de la région d'étude, ou de la littérature, doivent être mobilisées dans la phase de planification. Elles jouent un rôle essentiel et décisif que l'objectivité des calculs ultérieurs ne saurait masquer. On ne peut donc suivre Benzecri (1986) lorsqu'il affirme que «l'analyse statistique affranchit des idées *a priori*» en permettant de traiter le complexe «sans y projeter de structures *a priori*».

L'accent mis sur la planification résulte d'une évolution de l'outil statistique : on peut dire qu'avec le développement des méthodes de couplage entre tableaux, les méthodes d'ordination pour l'analyse des relations espèces milieux ont rejoint la statistique classique. En effet, utilisées au départ comme méthodes exploratoires pour révéler des gradients ou des groupements dans des ensembles de données complexes, les analyses factorielles tendent désormais à être utilisées en écologie pour tester des hypothèses sur les relations entre faune/flore et environnement, par exemple dans la modélisation de gradients définis *a priori* comme objets d'étude (Prodon et Lebreton 1981, Balent 1986a et b, Dakki 1986, Mercier 1988, etc). Le rapide développement de procédures inférentielles ne fera qu'accentuer cette tendance. Parmi les directions les plus prometteuses, l'utilisation dans des situations de type expérimental est à peine amorcée. Lavorel *et al.* (1991)



et Lavorel et Lebreton (1992) testent ainsi des hypothèses sur les mécanismes de succession végétale entre des relevés traités et des témoins. Les « variables de milieu » sont alors des variables qualifiant les relevés, c'est-à-dire des facteurs au sens exact de l'analyse de variance. Les mêmes auteurs associent également des facteurs hiérarchiques à différentes échelles spatiales et délimitent ainsi les échelles de plus forte variation de composition spécifique. A une échelle appropriée, est ensuite étudié l'effet de « variables de milieu » biotique, en l'occurrence le contenu de la banque de graines, sur la composition de la végétation épigée, pour tester des hypothèses concernant la niche de régénération.

Beaucoup de critiques sur l'utilisation de l'analyse multivariée dans l'étude des communautés résultent de l'abus qui en a été fait dans des situations uniquement descriptives. On rejoint ici les raisons de la crise de la phytosociologie classique. D'autres se rapportent aux difficultés infé-

rentielles, c'est-à-dire à la difficulté de cerner la validité des résultats. La flexibilité se développant actuellement vaut-elle la peine de l'effort de compréhension que cela suppose de la part de l'écologiste? La réponse peut être non en termes de compréhension des patrons principaux de variation, ou d'obtention de graphiques démonstratifs. L'échantillonnage dans l'étude des communautés est souvent si contrasté que les faits majeurs peuvent être appréhendés par bien des techniques statistiques différentes. Mais, même si l'évolution encore en cours des méthodes que nous avons passées en revue, et les difficultés qui en résultent pour l'écologiste, sont encore un frein, la réponse est certainement oui dans trois directions qui ne peuvent laisser indifférent :

— la possibilité d'un contrôle de la robustesse, c'est-à-dire de la validité externe des résultats, notamment lorsque les effets étudiés sont subtils,

— les liens étroits avec les concepts écologiques, par exemple avec les différents aspects de

Fig. 6. — Etapes de la réalisation et choix des options dans une analyse de relations espèces-milieu, basée sur un tableau relevés x espèces (Y) associé à un tableau relevés x variables (X).

Une telle analyse peut se résumer par la formule

$C((f(Y), Q, D), (g(X), R, D))$

dont les huit paramètres sont :

C méthode de couplage (canonique, variables instrumentales, co-inertie, régression après ACP de Y,...) (voir revue in Mercier 1991, Chessel et Mercier 1993)

f transformation (y compris centrage et réduction éventuels) du tableau relevés x espèces (par exemple $(p_{ij}/p_{i.p_j} - 1)$ dans le cas de l'AFC)

Y tableau relevés x espèces

Q pondération (ou « métrique ») des espèces (inutilisée en analyse canonique) (par exemple (p_i) en AFC)

D pondération (ou « métrique ») des relevés (c'est-à-dire des lignes de Y et de X)

g transformation des variables de milieu (y compris centrage et réduction éventuels) (par exemple en Logs)

X tableau relevés x variables de milieu

R pondération (« métrique ») des variables de milieu (inutilisée en analyse canonique et sur variables instrumentales) (par exemple Identité, c'est-à-dire poids uniformes, pour l'ACP).

Ainsi, par exemple :

ACPVI $((p_{ij}/p_{i.p_j} - 1), (p_j), (p_i))$, (X, Identité, (p_i)) est l'Analyse Canonique des Correspondances (ACC; ter Braak 1986, 1987) ou Analyse Factorielle des Correspondances sur Variables Instrumentales (AFCVI; Lebreton et al. 1988 a et b).

Steps and methodological options in the realization of a species-environment analysis, from a relevés x species matrix (Y) associated with a relevés x environmental variables (X). This analysis can be summarized by the formula

$C((f(Y), Q, D), (g(X), R, D))$

the eight parameters of which are :

C the method of coupling (canonical, instrumental variables, coinertia, regression after PCA of Y...) (see review in Mercier 1991, Chessel et Mercier 1993)

f transformation (including possible centering and reduction) of the sample-species matrix (e.g. $(p_{ij}/p_{i.p_j} - 1)$ for Correspondence Analysis)

Y sample x species matrix

Q weighting of the species (unused in canonical analysis) (e.g. (p_i) for CA)

D weighting of the samples (i.e. lines of Y and X matrices)

g transformation of environmental variables (including possible centering and reducing)

X sample x environmental variables matrix

R weighting of the environmental variables (unused in canonical analysis and instrumental variable analysis) (e.g. identity, i.e. uniform weighting, in PCA)

As an example :

CAIV $((p_{ij}/p_{i.p_j} - 1), (p_j), (p_i))$, (X, Identity, (p_i)) is the Canonical Correspondence Analysis (CCA; ter Braak 1986, 1987) or Correspondence Analysis with respect to Instrumental Variables (CAIV; Lebreton et al. 1988 a et b).

la théorie de la niche et de la diversité, liens sur lesquels nous nous proposons de revenir,

— les possibilités d'utilisation comme modèle (étude des perturbations, des échelles spatiales et temporelles...).

On ne peut donc qu'encourager les écologistes à faire preuve d'imagination et à concevoir des plans d'observation ou d'expérience astucieux ciblés sur des problèmes précis.

REMERCIEMENTS — Nous remercions MM. J. Lepart, P. Chardy et S. Hitier pour les utiles remarques dont ils ont bien voulu nous faire part dans des délais très courts.

BIBLIOGRAPHIE

- AUSTIN M.P., 1968. An ordination study of a chalk grassland community. *J. Ecol.* 56 : 739-757.
- AUSTIN M.P., R.B. CUNNINGHAM & P.M. FLEMING, 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55 : 11-27.
- AYYAD M.A.G. & R.L. DIX, 1964. An analysis of a vegetation microenvironmental complex on prairie slopes in Saskatchewan. *Ecol. Monogr.* 34 : 421-442.
- BALENT G., 1986a. Modélisation de l'évolution des surfaces pastorales dans les Pyrénées centrales. Mise au point d'un référentiel micro-régional de diagnostic au niveau de la parcelle. *Cah. Recherche-Développement* 9-10 : 92-98.
- BALENT G., 1986b. The influence of grazing on the evolution of botanical composition of previously cultivated fields. The example of the Pyrenees. In : Joss P.J., Lynch P.W., Williams O.B. (eds), *Rangeland : a resource under siege*. Australian Academy of Sc., Canberra.
- BEALS E.W., 1973. Ordination : mathematical elegance and ecological naïveté. *J. Ecol.* 61 : 23-35.
- BENZECRI J.P., 1972. La place de l'a priori. *Encyclopedia Universalis Organum* 17 : 11-24.
- BLANC F., P. CHARDY, A. LAUREC & J.P. REYS, 1976. Choix des métriques qualitatives en analyse d'inertie. Implications en écologie marine benthique. *Marine Biology* 35 : 49-67.
- BRAY J.R. & J.T. CURTIS, 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27 : 325-349.
- CHANG D.H.S. & H.G. GAUCH, 1986. Multivariate analysis of plant communities and environmental factors in Ngari, Tibet. *Ecology* 67 : 1568-1575.
- CHEssel D., 1992. Echanges interdisciplinaires en analyse des données écologiques. Mém. Univ. Claude Bernard — Lyon I, 108 p. (non publié).
- CHEssel D. & S. DOLEDEC, 1992. ADE version 3.3. URA CNRS 1451, Université Lyon I, 69633 Villeurbanne cedex.
- CHEssel D. & P. MERCIER, 1993. Couplage de triplet statistiques et liaisons espèces-environnement. In : Lebreton J.D. et Asselain B. (eds), *Biométrie et environnement*, p. 15-43. Masson, Paris.
- CHEssel D., J.D. LEBRETON & R. PRODON, 1982. Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés : cas d'un gradient simple. *C. R. Séances Acad. Sci. Sér. III Sci. Vie* 295 : 83-88.
- CONNELL J.H., 1978. Diversity in tropical rain forests and coral reefs. *Science* 199 : 1302-1310.
- DAGET P., 1977. Ordination des profils écologiques. *Naturalia Monspeliensa*, Sér. Bot. 26 : 109-128.
- DAGET P., 1979. Préférence écologique et valeur caractérisante. *Naturalia Monspeliensa* 27 : 151-166.
- DAGET P. & M. GODRON, 1982. Analyse fréquentielle de l'écologie des espèces dans les communautés. Masson, Paris.
- DAKKI M., 1986. Biotypologie et gradient thermique spatio-temporel. Etude sur un cours d'eau du Moyen Atlas (Maroc). *Bull. Ecol.* 17 : 79-85.
- DEMARCO G. & C. MOURER-CHAUVIRE, 1976. L'indice thermique, un essai de méthode pour comparer et utiliser climatologiquement les gisements quaternaires. Application aux oiseaux du Pléistocène français. *Géobios* 9 : 125-141.
- DENSLOW J.S., 1985. Disturbance-mediated coexistence of species. In : Pickett S.T.A. et White P.S. (eds), *The ecology of natural disturbance and patch dynamics*. Academic Press, Orlando, Florida.
- DIGBY P.G.N. & R.A. KEMPTON, 1987. Multivariate analysis of ecological communities. Chapman and Hall, London, New York 206 p.
- DIX R.L. & F.E. SMEINS, 1967. The prairie, meadow, and marsh vegetation of Nelson County, North Dakota. *Can. J. Bot.* 45 : 21-58.
- DOLEDEC S. & D. CHEssel, 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables. *Acta Oecologica, Oecol. Gener.* 8 : 207-232.
- DOLEDEC S. & D. CHEssel, 1989. Rythmes saisonniers et composantes stationnelles en milieu aquatique. II. Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica, Oecol. Gener.* 10 : 341-352.
- DOLEDEC S. & D. CHEssel, 1991. Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology* 1 : 133-155.
- ELLENBERG H., 1948. Unkrautgesellschaften als Mass für den Säuregrad, die Verdichtung und andere Eigenschaften des Ackerbodens. *Berichte über Landtechnik, Kuratorium für Technik und Bauwesen in der Landwirtschaft.* 4 : 130-146.
- ELLENBERG H., 1979. Zeigerwerte der Gefäßpflanzen Mitteleuropas. *Scripta Geobotanica*, Göttingen, 121 p.
- ESTEVE J., 1878. Les méthodes d'ordination : éléments pour une discussion. In : Legay J.M., Tomassone R. (eds.), *Biométrie et écologie*. Soc. Française de Biométrie, Paris 1 : 223-250.
- FRONTIER S., 1983. Stratégies d'échantillonnage en écologie. Masson, Paris.

- GAUCH H.G., 1982. Multivariate analysis in community ecology. Cambridge Univ. Press., Cambridge 298 p.
- GAUCH H.G. & E.L. STONE, 1979. Vegetation and soil pattern in a mesophytic forest at Ithaca, New York. *Am. Midl. Nat.* 102 : 332-345.
- GAUCH H.G. & T.R. WENTWORTH, 1976. Canonical Correlation Analysis as an ordination technique. *Vegetatio* 33 : 17-22.
- GAUCH H.G., G.B. CHASE & R.H. WHITTAKER, 1974. Ordination of vegetation samples by Gaussian species distributions. *Ecology* 55 : 1382-1390.
- GAUCH H.G., R.H. WHITTAKER & T.R. WENTWORTH, 1977. A comparative study of reciprocal averaging and other ordination techniques. *J. Ecol.* 65 : 157-174.
- GAUSE G.F., 1934. The struggle for existence (reprinted 1969). Hafner Publ. Comp., New York.
- GITTINS R., 1985. Canonical analysis. A review with applications in ecology. Springer Verlag, Berlin.
- GLOWACINSKI Z., 1975. Succession of bird communities in the Niepolomice forest (Southern Poland). *Ecologia Polska* 23 : 231-263.
- GOFF F.G. & G. COTTAM, 1967. Gradient analysis : the use of species and synthetic indices. *Ecology* 48 : 793-805.
- GOUNOT M., 1969. Méthodes d'étude quantitative de la végétation. Masson, Paris 314 p.
- GREIG-SMITH, 1979. The development of numerical classification and ordination. *Vegetatio* 42 : 1-9.
- GUINOCHET M., 1973. Phytosociologie. Masson, Paris.
- HILL M.O., 1973. Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61 : 237-249.
- HILL M.O. & H.G. GAUCH, 1979. Detrended correspondence analysis : an improved ordination technique. *Vegetatio* 42 : 47-58.
- HITIER S., 1993. Quelques éléments pour tester l'indépendance de deux tableaux. Rapport de D.E.A., Univ. Montpellier I, II, ENSAM, 38 p. (non publié).
- HUTCHINSON G.E., 1957. Concluding remarks : The formalization of the niche and the Volterra-Gause principle. *Cold Spring Harbor Symposium in Quantitative biology* 22 : 415-427.
- HUTCHINSON G.E., 1978. An introduction to population ecology. Yale Univ. Press, New Haven, Connecticut.
- IBANEZ F., 1968. Application de la méthode d'analyse en composantes principales à l'étude des populations planctoniques à l'ouest de la Sardaigne. *C.R. Séances Acad. Sci.* D267 : 1215-1258.
- JACCARD P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull. Soc. Vaudoise Sci. nat.* 37 : 547-579.
- JAMES F.C. & C.E. MCCULLOCH, 1990. Multivariate analysis in ecology and systematics : panacea or Pandora's box ? *Annu. Rev. Ecol. Syst.* 21 : 129-166.
- JOHNSON D.H., 1981. The use and misuse of statistics in wildlife habitat studies. In : Capen D.E. (ed.), The use of multivariate statistics in studies of wildlife habitat, 11-19. U.S. Dep. Agric. For. Serv., Gen. Techn. Report RM-87.
- KOSTITZIN V.A., 1935. Evolution de l'atmosphère. Exposés de biométrie et de statistique biologique. VIII. Hermann, Paris.
- LAVOREL S. & J.D. LEBRETON, 1992. Evidence for lottery recruitment in Mediterranean old field. *J. Vegetation Sc.* 3 : 91-100.
- LAVOREL S., J.D. LEBRETON, M. DEBUSSCHE & J. LEPART, 1991. Nested spatial patterns in seed bank and vegetation of Mediterranean old-fields. *J. Vegetation Sc.* 2 : 367-376.
- LEBRETON J.D., 1980. Quelques éléments sur l'analyse des correspondances à l'intention des ornithologues. *Bièvre* 2 : 47-62.
- LEBRETON J.D., 1991. Principal component and Correspondence analysis with respect to instrumental variables : on overview of their role in studies of structure-activity and species-environment relationships. In *Applied multivariate analysis*, ed Devillers J., Karcher W., Kluwer Pub., Dordrecht.
- LEBRETON J.D. & N. YOCCOZ, 1987. Multivariate analysis of bird count data. In : Blondel J., Frochot B. (eds.), Bird census and atlas studies. Proc. 9th Int. Conf. Bird Census and Atlas Work. *Acta Oecologica, Oecol. Gener.* 8 : 125-144.
- LEBRETON J.D., D. CHESSEL, R. PRODON & N. YOCCOZ, 1988a. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. *Acta oecologica, Oecol. Gener.* 9 : 53-67.
- LEBRETON J.D., D. CHESSEL, M. RICHARDOT-COULET & N. YOCCOZ, 1988b. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II. Variables de milieu qualitatives. *Acta Oecologica, Oecol. Gener.* 9 : 137-151.
- LEBRETON J.D., M. ROUX, A.M. BACOU, & G. BANCO, 1990. Bioméco (Biométrie-Ecologie), version 3.9, CNRS-CEFE Montpellier.
- LEBRETON J.D., R. SABATIER, G. BANCO & A.M. BACOU, 1991. Principal Component and Correspondence analyses with respect to Instrumental Variables : an overview of their role in studies of structure - activity and species - environment relationships. In : Devillers J. and Karcher. *Applied Multivariate Analysis in SAR and Environmental studies* : 85-114.
- MACARTHUR R.H., 1958. Population ecology of some warblers of northeastern coniferous forests. *Ecology* 39 : 599-619.
- MCINTOSH R.P., 1958. Plant communities. *Science* 128 : 115-120.
- MCINTOSH R.P., 1967. The continuum concept of vegetation. *Bot. Rev.* 33 : 130-187.
- MCINTOSH R.P., 1987. Pluralism in ecology. *Ann. Rev. Ecol. Syst.* 18 : 321-341.
- MERCIER P., 1988. Approche méthodologique de l'étude des relations entre la structure spatiale du peuplement ligneux et la régénération du sous-bois. *Ann. Sci. For.* 45 : 275-290.
- MERCIER P., 1991. Etude des relations espèces-environnement et analyse de la co-structure d'un couple

- de tableaux. Thèse Univ. Claude Bernard, Lyon I, 178 p.
- MERCIER P., D. CHESSEL & S. DOLEDEC, 1992. Complete Correspondence Analysis of an ecological profile data table : a central ordination method. *Acta Oecologica* 13 : 25-44.
- NOY-MEIR I. & M.P. AUSTIN, 1970. Principal component ordination and simulated vegetational data. *Ecology* 51 : 551-552.
- NUDDS T.D., 1982. Ecological separation of grebes and coots. Interference competition or microhabitat selection?. *Wilson Bull.* 94 : 505-514.
- PAINE R.T., 1966. Food web complexity and species diversity. *Am. Nat.* 100 : 65-75.
- PALMER M.W., 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74 : 2215-2230.
- PICKETT S.T.A. & P.S. WHITE, 1985. Patch dynamics : a synthesis. In : The ecology of natural disturbance and patch dynamics, ed. Pickett S.T.A., White P.S. Academic Press.
- PRODON R., 1988. Dynamique des systèmes avifaune-végétation après déprise rurale et incendies dans les Pyrénées méditerranéennes siliceuses. Thèse Doct. Sci., Paris, Univ. Pierre et Marie Curie 333 p.
- PRODON R. & J.D. LEBRETON, 1981. Breeding avifauna of a Mediterranean succession : the holm oak and cork oak series in the eastern Pyrenees, 1. Analysis and modelling of the structure gradient. *Oikos* 37 : 21-38.
- PRODON R. & J.D. LEBRETON, 1983. Prediction of bird census from vegetation structure. Application to the study of a postfire succession. In : Purroy F.J. (ed.), Bird census and Mediterranean landscape. Proc. 7th Int. Conf. Bird Census. Univ. Leon, Spain 190-194.
- RICHARDOT-COULET M., D. CHESSEL & M. BOURNAUD, 1986. Typological value of the benches of old beds of a large river. Methodological approach. *Archiv. Hydrobiol.* 107 : 363-383.
- ROMANE F., 1972. Application à la phytoécologie de quelques méthodes d'analyse multivariée. Thèse Doct.-Ing., Univ. des Sc. et Techn. du Languedoc, Montpellier.
- ROUTLEDGE R.D., 1984. Estimating ecological components of diversity. *Oikos* 42 : 23-29.
- SABATIER R., 1987. Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes. Thèse Doct. ès-Sc., Univ. Montpellier, 184 p.
- SABATIER R., J.D. LEBRETON & D. CHESSEL, 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. In : Coppi R. and Bolasco S. (eds), Multiway data analysis. Elsevier, Amsterdam.
- SCHERRER B., 1983. Techniques de sondage en écologie. In : Frontier S. (ed.), Stratégies d'échantillonnage en écologie. Masson, Paris, et Les Presses de l'Univ. du Québec 63-162.
- SCHOENER T.W., 1989. The ecological niche. In : Cherratt, J.M. (ed.), Ecological concepts. The contribution of ecology to an understanding of the natural world. Blackwell Sc. Publ., Oxford.
- SEBER G.A.S., 1984. Multivariate observations. Wiley, New York.
- SHI G.R., 1993. Multivariate data analysis in palaeoecology and palaeobiogeography – A review. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* 105 : 199-234.
- SIMBERLOFF D. & T. DAYAN, 1991. The guild concept and the structure of ecological communities. *Annu. Rev. Ecol. Syst.* 22 : 115-1143.
- TER BRAAK C.J.F., 1986. Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67 : 1167-1179.
- TER BRAAK C.J.F., 1987a. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69 : 69-77.
- TER BRAAK C.J.F., 1987b. Ordination. In : Jongmann R.H.G., ter Braak C.J.F., van Tongeren O.F.R. (eds), Data analysis in community and landscape ecology, 91-173. Centre Agric. Publ. Docum. (Pudoc), Wageningen.
- TER BRAAK C.J.F., 1987c. CANOCO. Version 2.1. ITI-TNO, Wageningen, The Netherlands.
- TER BRAAK C.J.F., 1988. Partial Canonical Correspondence Analysis. In : Bock H.H. (ed.), Classification and related methods of data analysis, 551-558. North-Holland, Amsterdam.
- TER BRAAK C.J.F. & N.J.M. GREMMEN, 1987. Ecological amplitude of plant species and the internal consistency of Ellenberg's indicator values for moisture. *Vegetatio* 69 : 79-87.
- TER BRAAK C.J.F. & C.W.N. LOOMAN, 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65 : 3-11.
- TER BRAAK C.J.F. & PRENTICE I.C., 1988. A theory of gradient analysis. In : Begon M., Fitter A.H., Ford E.D. and MacFadyen A. (eds), Advances in Ecological Research, Vol. 18 : 271-317.
- THIOULOUSE J. & D. CHESSEL, 1992. A method for reciprocal scaling of species tolerance and sample diversity. *Ecology* 73 : 670-680.
- TOGERSON W.S., 1958. Theory and methods of scaling. Wiley, New York.
- TUCKER L.R., 1958. An inter-battery method of factor analysis. *Psychometrika* 23 : 111-136.
- WARTENBERG D., S. FERSON & F.J. ROHLF, 1987. Putting things in order : a critique of detrended correspondence analysis. *Am. Nat.* 129 : 434-448.
- WHITTAKER R.H., 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30 : 279-338.
- WHITTAKER R.H., 1967. Gradient analysis of vegetation. *Biol. Rev.* 42 : 207-264.
- WHITTAKER R.H. & H.G. GAUCH, 1978. Evaluation of ordination techniques. In : Whittaker R.H. (ed.), Ordination of plant communities, 277-336. Junk, The Hague.
- WHITTAKER R.H. & W.A. NIERING, 1964. Vegetation of the Santa Catalina Mountains, Arizona. I. Ecological classification and distribution of species. *J. Arizona Acad. Sc.* 3 : 9-34.

WHITTAKER R.H., S.A. LEVIN & R.B. ROOT, 1973. Niche, habitat and ecotope. *Am. Nat.* 107 : 321-338.

WIENS J.A., 1989. The ecology of bird communities. 2 Vol. Cambridge Univ. Press., Cambridge.

WILLIAM W.T. & G.N. LANCE, 1968. Choice of strategy in the analysis of complex data. *Statistician* 18 : 31-43.

YOCCOZ N. & D. CHESSEL, 1988. Ordination sous contrainte de relevés d'avifaune : élimination d'effets dans un plan d'observation à deux facteurs. *C. R. Séances Acad. Sci. Sér. III Sci. Vie* 307 : 189-194.

Reçu le 16 mai 1994 ; received May 16, 1994
Accepté le 14 juin 1994 ; accepted June 14, 1994