

LILA ABDESSEMED

BRIGITTE ESCOFIER

**Analyse factorielle multiple de tableaux de
fréquences ; Comparaison avec l'analyse
canonique des correspondances**

Journal de la société statistique de Paris, tome 137, n° 2 (1996), p. 3-18

http://www.numdam.org/item?id=JSFS_1996__137_2_3_0

© Société de statistique de Paris, 1996, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

I

ARTICLES

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCES ; COMPARAISON AVEC L'ANALYSE CANONIQUE DES CORRESPONDANCES

Lila ABDESSEMED, Brigitte ESCOFIER¹
IRISA-INRIA Rennes

Résumé

L'analyse factorielle multiple est une technique qui permet seulement d'étudier de façon simultanée des groupes de variables numériques et/ou qualitatives. Nous proposons ici, une méthode qui permet également l'étude des tableaux de fréquences.

Nous présentons cette méthode dans le cas d'un seul tableau de fréquences et d'un tableau de variables numériques. Celle-ci consiste en une transformation du tableau de fréquences et l'introduction de pondérations adéquates sur les lignes et les colonnes.

Nous illustrons cette méthode par un exemple d'application sur des données écologiques. Nous comparons cette méthode avec l'analyse canonique des correspondances, et nous discutons le cas de plusieurs tableaux de fréquences.

Abstract

Multiple Factor Analysis is a method which only allows to analyze together groups of numerical variables and/or groups of qualitative variables. We propose here a method which allows to analyze simultaneously two-way contingency tables, quantitative and qualitative tables.

We present this method in the case of one contingency table only and one group of numerical variables. It consists of the transformation of the frequency table and the introduction of metrics on rows and columns of the two groups.

A comparison is made with the Canonical Correspondence Analysis, and the case of multiple two-way contingency tables is studied.

1. Mme Escofier était professeur à l'IUT de Vannes et est décédée au mois de Juillet 1994.

I. Introduction

L'étude de la liaison entre plusieurs groupes de variables a toujours été et est toujours un sujet d'actualité, qui a eu des éléments de réponse selon la problématique d'une part, et selon la nature des variables.

Nous nous intéressons dans ce cadre, à l'étude de la liaison entre les structures induites par un tableau de fréquences et par un tableau de variables descriptives, ces dernières pouvant être de type numérique, ou qualitatif.

Parmi les méthodes qui permettent d'étudier de façon simultanée des groupes de variables de nature différente, l'Analyse Factorielle Multiple (AFM), ESCOPIER (1990), peut apporter des réponses à notre problème, mais celle-ci ne s'applique qu'à des groupes de variables numériques ou qualitatives et ne prend pas en compte des informations de type profils qui caractérisent les tableaux de fréquence. Dans cette étude, nous nous proposons de généraliser l'AFM à l'étude des tableaux de fréquences.

Nous présentons notre méthode d'abord dans le cas d'un tableau descriptif numérique et d'un tableau de fréquences. Nous discutons ensuite, le cas de plusieurs tableaux de variables numériques et/ou qualitatives et d'un tableau de fréquences, et enfin, nous examinons le cas de plusieurs tableaux de fréquences.

Par ailleurs, l'Analyse Canonique des Correspondances (ACC) étant une technique permettant d'étudier la liaison entre les structures induites par un tableau de fréquences, et par un tableau de variables numériques, nous la comparons avec la méthode que nous proposons.

II. Les notations, les données et les objectifs

Nous considérons un tableau de fréquences F de terme général f_{ij} , ayant I lignes et J colonnes ; ses marges sur I et sur J sont notées respectivement $f_{i.}$ et $f_{.j}$.

Par ailleurs, nous considérons un tableau de variables descriptives X de terme général x_{ik} , décrivant les I individus à l'aide de K variables numériques.

Pour notre étude, nous avons repris les données écologiques illustratives de l'ACC présentées dans LEBRETON (1991). Ces données se présentent sous la forme d'un tableau $F=(f_{ij})$ indiquant la présence ou l'absence de 22 espèces de plantes² sur 12 sites, qui sont par ailleurs décrits par 4 variables d'environnement dans le tableau $X=(x_{ik})$. (*sal* : indice de salinité en g/m^2 , *dist* : distance à la mer (m), *soil* : profondeur du sol (cm), *slop* : pente (degré)). A travers cette étude, nous cherchons à mettre en évidence des relations entre la fréquence des espèces sur les sites et les valeurs des variables descriptives.

2. On trouvera en annexe l'intitulé et les noms des 22 espèces de plantes.

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCE

Aussi, notre premier objectif consiste à voir si les structures induites sur les sites par les deux tableaux F et X sont liées, ce qui se traduira par des directions de dispersion communes.

On étudie aussi les relations entre les espèces caractérisées par les fréquences sur les différents sites, ainsi que les corrélations entre les variables descriptives. On cherche surtout à mettre en relief, si elle existe, une relation spécifique espèce-variable.

L'objectif suivant est de trouver les sites qui se ressemblent de par leur répartition sur les espèces et par les valeurs des variables descriptives, et on veut détecter ceux qui sont proches seulement au sens de la répartition des espèces ou ceux qui le sont uniquement au sens des variables environnementales. On cherche alors, une représentation simultanée des sites au sens de la répartition "espèces et variables", et au sens des deux autres points de vue.

Dans l'étude de la liaison entre un tableau de fréquences et un groupe de variables, ces objectifs sont donc exactement les mêmes que ceux de l'AFM, pour l'analyse de la liaison entre plusieurs groupes de variables numériques.

		Espèces																Variables descriptives			
																		sali	dist	soil	slop
Sites		1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1483	23	3	30
		1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	777	29	2	20
		0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	298	35	11	25
		0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	875	57	1	25
		0	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	509	57	11	20
		0	0	1	1	1	0	0	1	1	0	0	0	1	0	0	0	542	60	4	20
		0	0	1	1	1	1	0	1	1	0	1	0	1	0	0	0	382	62	12	20
		0	1	1	1	1	0	0	1	1	1	1	0	1	1	0	0	272	66	6	27
		0	1	1	0	1	0	0	1	1	1	1	0	1	1	1	1	214	80	10	18
		0	0	1	0	0	1	0	1	1	1	0	0	1	1	1	0	105	85	12	10
		0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	148	100	17	5
		0	0	0	1	0	0	0	1	1	1	1	0	1	0	1	1	62	123	16	15

Table. 1 : Le tableau de fréquences F et le tableau descriptif X.

III. Technique de généralisation

Un "vrai" tableau de fréquences est un tableau croisant deux variables qualitatives. Il est communément traité par l'Analyse Factorielle des Correspondances (AFC) qui permet d'étudier l'écart à l'hypothèse d'indépendance. En AFC, lignes et colonnes sont traitées avec une complète symétrie, et sont représentées par leurs profils.

Pour étudier conjointement le tableau de fréquences et le tableau de variables, il est souhaitable de conserver le point de vue de l'AFC pour le tableau de fréquences et donc conserver les distances du χ^2 , entre lignes ou colonnes, pour pouvoir en restituer la structure.

Comme, l'AFM est une Analyse en Composantes Principales (ACP) particulière, la technique de généralisation sera alors basée sur l'équivalence entre l'AFC du tableau de fréquences F et l'ACP non normée d'un tableau transformé F' , ses lignes et ses colonnes étant munies de métriques adéquates. En effet, il est équivalent de réaliser l'AFC du tableau de terme général f_{ij} ou l'ACP non normée du tableau transformé de terme général $f_{ij}/f_i f_j$, les individus étant munis des poids f_i et les variables des poids f_j .

Cette équivalence nous permet d'appliquer donc l'AFM au tableau (table 2) composé de deux groupes de variables numériques définis par F' et X . Le tableau F' est f_i centré³ et a pour terme général $(f_{ij}/f_i f_j - 1)$ et le tableau X est centré et réduit pour les poids f_i . Nous notons $\bar{x}_k$ et s_k la moyenne et l'écart-type de la variable k calculés avec les poids f_i . Le terme général du tableau X est $(x_{ik} - \bar{x}_k)/s_k$. Les sites i sont munis des poids f_i et les espèces j ont le poids f_j . Nous notons $D = \text{diag}(f_i)$ et $M = \text{diag}(f_j)$ les métriques induites par les poids f_i et f_j sur les espaces R^I et R^J dans lesquels sont situés respectivement les variables et les sites. Les deux nuages de sites associés aux tableaux F' et X ont été centrés, pour éliminer l'influence de la position des deux barycentres et ne prendre en compte que la forme des nuages. Par ailleurs, cela permet d'interpréter le cosinus entre deux vecteurs représentant les variables de X comme un coefficient de corrélation, avec la métrique D .

IV. Influence des pondérations D et M et surpondération des groupes

L'AFM de groupes de variables numériques se fait en deux étapes. La première est une ACP de chacun des groupes de variables pris séparément. Quelle est de ce fait, l'influence des métriques introduites sur les différents nuages associés à chacun des deux tableaux F' et X ?

Pour le tableau de fréquences transformé F' , les métriques M et D n'ont d'autre effet que la conservation de la structure AFC du tableau de fréquences initial F .

Le tableau de variables descriptives n'est influencé que par la métrique D , qui donne à la distance d'un site à l'origine, une importance proportionnelle à son effectif exprimé dans F , et influe en particulier sur la liaison entre deux variables, puisque celle-ci s'exprime à travers une f_i corrélation.

La seconde étape de l'AFM est une ACP de l'ensemble des variables de tous les groupes. Les variables de chaque groupe sont pondérées, en divisant leur poids

3. On parlera de f_i centrage et de f_j corrélation pour le centrage et la corrélation induits par la métrique $D = \text{diag}(f_i)$.

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCE

initial par la première valeur propre obtenue dans chacune des ACP séparées, soit λ_1 pour le tableau F' et μ_1 pour le tableau X. Cette pondération a pour effet d'obtenir dans une direction quelconque pour chacun d'eux, une inertie maximale égale à 1, et ce dans le but d'équilibrer leur influence respective.

Pour un tableau de fréquences, le problème de la surpondération ou de la non surpondération va se poser avec une certaine acuité, compte tenu des particularités de ce type de tableau ; les valeurs propres obtenues dans une AFC sont toutes majorées par 1, et l'inertie associée au tableau dépend de la structure du tableau. Cette dernière se traduit plus particulièrement sur la première valeur propre, autrement dit, plus la structure est forte, plus la première valeur propre est grande.

Nous avons opté pour la surpondération classique des deux groupes, ce qui permettait de conserver les propriétés de l'AFM pour la comparaison des groupes, à savoir un même intervalle de variation pour la mesure de la liaison entre groupes et facteurs. Dans les autres cas, nous avons un intervalle de variation qui est variable, ce qui rendait l'indice mesurant la liaison entre groupes et facteurs, difficilement interprétable.

En outre, cette solution permet justement de donner la même importance dans l'analyse globale, aux distances entre sites définis par les espèces et par les variables, ceci que la structure définie par la répartition des espèces soit forte ($\lambda_1 \simeq 1$) ou faible ($\lambda_1 \simeq 0$).

	Espèces	Variables	Poids
Sites	$\dots \frac{f_{ij}}{f_i f_j} - 1 \dots$	$\dots \frac{x_{ik} - \bar{x}_k}{s_k} \dots$	$f_{i.}$
Poids	$\dots f_j \dots$	$\dots 1 \dots$	

Table 2 : Le tableau de fréquences F' et le tableau descriptif X centrés pour les poids $f_{i.}$

V. Individus, Variables et Groupes

En AFM, on analyse trois types d'objets : les individus, les variables et les groupes. Ainsi, nous sommes amenés à étudier les caractéristiques des différents nuages les représentant.

V.1 Les individus

On cherche comme dans l'AFM classique, une représentation graphique du nuage des individus $N(I)$ caractérisés par l'ensemble des variables, et une représentation superposée des deux nuages de sites $N(I_1)$ et $N(I_2)$ caractérisés par la répartition des espèces et par les variables d'environnement. L'espace $R^{J \oplus K}$ de dimension $J+K$ contient les représentations par le nuage $N(I)$ des individus définis par l'ensemble des colonnes des tableaux X et F .

La projection de $N(I)$ sur le sous espace de $R^{J \oplus K}$ engendré par les J premières composantes, se déduit de $N(I_1)$ par une simple homothétie de rapport $1/\lambda_1$. De même, la projection de $N(I)$ sur le sous espace de $R^{J \oplus K}$ engendré par les K dernières composantes, se déduit de $N(I_2)$ par une homothétie de rapport $1/\mu_1$. Ces homothéties sont dues à la surpondération.

Le nuage des sites décrits par les espèces, $N(I_1)$ est f_i centré et la distance entre deux sites s'exprime par la distance du χ^2 et correspond au nuage des sites obtenu dans l'AFC du tableau de fréquences initial F .

Le nuage des sites décrits par les variables descriptives, $N(I_2)$ correspond aux lignes du tableau X est encore f_i centré et la distance entre deux sites est la distance euclidienne. Le nuage des sites décrits par les espèces et par les variables descriptives, $N(I)$ est la "réunion" de deux sous nuages f_i centrés est aussi f_i centré, et correspond aux lignes du tableau $F'UX$. Le carré de la distance entre deux sites s'écrit comme la somme des carrés pondérés des distances entre les sites dans chacun des sous nuages $N(I_1)$ et $N(I_2)$.

V.2 Les variables

Les variables (espèces et variables descriptives) associées aux deux groupes sont situées dans R^I , et génèrent de ce fait deux nuages de points, correspondant aux espèces, aux variables descriptives et un troisième nuage qui est la réunion des deux premiers.

Le nuage des espèces, $N(J)$ est f_j centré et l'interprétation pour ce nuage de variables est analogue à celle d'une AFC, puisque la distance entre deux espèces est la distance du χ^2 et leur liaison s'exprime sur un axe par leur proximité au sens euclidien. Le nuage des variables descriptives, $N(K)$ n'est pas centré et on retrouve pour celui-ci les interprétations de l'ACP pour les variables, avec entre autre, une interprétation des liaisons sur un plan, à l'aide d'un cercle des f_j corrélations. Le nuage des espèces et des variables descriptives, $N(JUK)$, est situé dans R^I est également non centré, puisqu'il est la réunion de deux sous nuages dont l'un n'est pas centré, soit $N(K)$. Il correspond aux colonnes du tableau $F'UX$.

Par ailleurs, la liaison entre une espèce j et une variable k est alors mesurée par la covariance entre le profil de l'espèce et la variable. Cette covariance

est délicate à interpréter, mais nous verrons que sur un facteur, grâce aux formules de transition, l'interprétation des positions relatives espèces-variables est simple.

V.3 Les groupes

Dans l'AFM, l'un des objectifs est comparer globalement les groupes. Aussi compte tenu des choix qui ont été faits en matière de surpondération, nous avons repris les mêmes outils que ceux présentés dans l'AFM classique, ESCOFIER (1990). Les deux groupes sont donc représentés dans R^{I^2} par les opérateurs $W_1 D$ et $W_2 D$ avec $W_1 = {}^t F' M_1 F'$ et $W_2 = {}^t X M_2 X$, muni du produit scalaire $\langle A, B \rangle = \text{Trace } A {}^t B$; M_1 et M_2 sont les matrices diagonales des poids des variables pour les groupes 1 et 2. Dans notre cas, $M_1 = M$ et $M_2 = I$.

Dans cet espace, un groupe est représenté par sa coordonnée sur l'axe factoriel $z_s {}^t z_s D$ qui est l'inertie de la projection du nuage défini par le groupe sur la composante principale z_s de l'analyse globale de F'UX.

Deux groupes G_1 et G_2 sont proches si la distance $d^2(G_1, G_2)$ est faible et ceci correspond au cas où les distances entre individus sont semblables dans $N(I_1)$ et $N(I_2)$.

V.4 Les formules de transition

Soient :

- γ_s la valeur propre de rang s obtenue dans l'AFM de F'UX.
- $F_s(i)$ le facteur d'ordre s pour le site i obtenu dans l'AFM de F'UX.
- $G_s(j)$ le facteur d'ordre s pour l'espèce j obtenu dans l'AFM de F'UX.
- $H_s(k)$ le facteur d'ordre s pour la variable descriptive k obtenu dans l'AFM de F'UX.

Dans l'analyse globale, les sites sont munis des poids f_i , les espèces des poids f_j/λ_1 et les variables descriptives des poids $1/\mu_1$.

$$H_s(k) = \frac{1}{\sqrt{\gamma_s}} \sum_i f_i \frac{x_{ik}}{s_k} F_s(i) \quad k \in K$$

$$G_s(j) = \frac{1}{\sqrt{\gamma_s}} \sum_i \frac{f_{ij}}{f_j} F_s(i) \quad j \in J$$

$$F_s(i) = \frac{1}{\sqrt{\gamma_s}} \sum_j \frac{f_j}{\lambda_1} \left(\frac{f_{ij}}{f_j} - 1 \right) G_s(j) + \sum_k \frac{1}{\mu_1} \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) H_s(k) \quad i \in I$$

VI. Représentation des sites et des variables sur les facteurs et interprétation des résultats

VI.1 Les groupes et les facteurs

L'inertie du premier facteur vaut 1,95. Ceci indique déjà l'existence d'une structure commune entre les deux groupes, puisque la valeur maximum de cette inertie qui est 2, est atteinte lorsque les directions principales d'inertie des deux nuages se confondent, et que le lien facteur-groupe vaut 1. Ici, ce lien vaut respectivement 0,97 et 0,98 pour le premier et le deuxième groupe. La corrélation entre ce facteur et la projection associée du nuage $N(I_1)$ vaut 0,99 et vaut également 0,99 pour le nuage $N(I_2)$; le facteur commun est donc confirmé : il existe dans les deux nuages $N(I_1)$ et $N(I_2)$ une direction de dispersion commune. Ces directions sont très proches des premiers facteurs de F et de X. La f_1 corrélation vaut 0,97 pour le premier et 0,99 pour le second. L'inertie du deuxième facteur vaut 0,81 et le lien avec le premier groupe est de 0,79 et avec le second groupe, il n'est que de 0,020. Il s'agit donc, d'un facteur de répartition des espèces non lié aux variables descriptives, qui se confond presque avec le deuxième facteur de F (corrélation = 0,98). Les facteurs suivants sont également des facteurs spécifiques à la répartition des espèces sur les sites, situation qui était attendue, puisque dans les ACP séparées, le tableau X est quasi-monodimensionnel et le tableau F est multidimensionnel. Dans la suite, nous nous limiterons donc aux deux premiers facteurs.

1. Les variables descriptives (Fig. 1)

L'interprétation de la projection du nuage des variables descriptives se fait de manière analogue à celle de l'ACP. En effet, la coordonnée de la projection d'une variable k sur un facteur représente la f_k corrélation entre cette variable et le facteur. On note sur le premier facteur, une opposition très marquée entre l'indice de salinité (SALI) et la distance à la mer (DIST). La profondeur du substrat (SOIL) est très corrélée avec la distance à la mer et la pente du site (SLOP) l'est avec la salinité (SALI). Pour le deuxième facteur, les variables descriptives n'interviennent quasiment pas.

2. Les espèces (Fig. 2)

A un coefficient près, la pondération des sites par f_i fait d'une espèce j le barycentre des sites qui la contiennent (car $\sum_i f_i \cdot F_i(i) = 0$) ; la propriété barycentrique de l'AFC est donc conservée.

La liaison entre deux espèces se traduit en terme de distance du χ^2 (à un coefficient près). On aura donc une interprétation du nuage projeté des espèces $N(J)$ qui sera analogue à celle de l'AFC. Il y a une opposition entre les espèces qui se projettent du côté de la distance à la mer (DIST) et celles qui se projettent du côté de l'indice de salinité (SALI). On peut en conclure qu'une espèce qui se projette du même côté que SALI, et à l'opposé de DIST, est plus fréquente dans les sites salins. A l'opposé, une espèce qui est du même côté que DIST est plus fréquente dans les sites éloignés de la mer et donc moins salins.

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCE

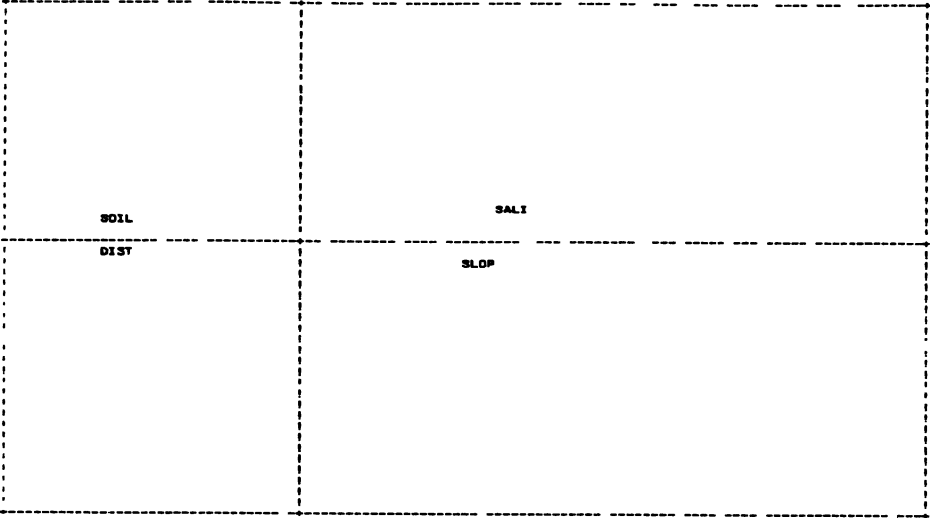


Fig. 1 : Représentation des variables descriptives sur le plan 1x2.

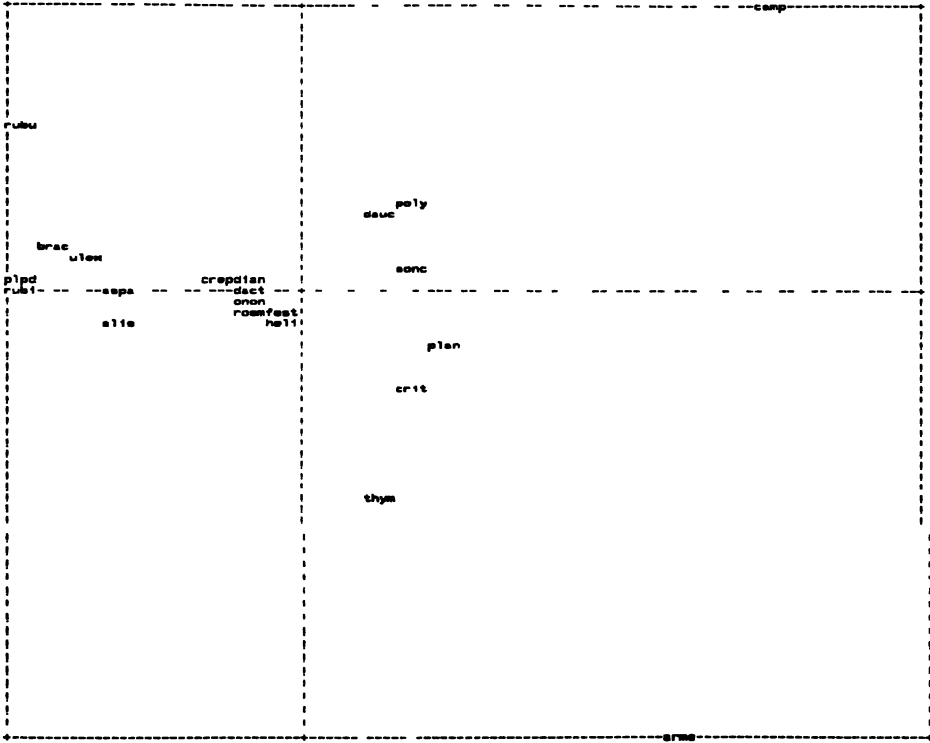


Fig. 2 : Représentation des espèces sur le plan 1x2.

3. Les variables descriptives et les espèces

L'espèce j et la variable k se projettent du même côté, si des sites qui ont une fréquence de j supérieure à la fréquence moyenne des sites, ont aussi une valeur supérieure à la valeur moyenne pour k . L'espèce "camp" et la variable SALI se projettent du même côté. On peut en conclure que cette espèce est beaucoup plus fréquente que la moyenne dans les sites qui ont un indice de salinité supérieur à la valeur moyenne.

VI.2 Représentation des sites

Divers points de vue pour les sites seront examinés : le premier concerne les sites vus au travers des deux groupes de variables et les suivants décrivent les sites vus au travers de chacun des deux groupes.

1. Représentation du nuage des sites vus au travers des 2 groupes (Fig. 3)

La représentation du nuage des sites $N(I)$ s'obtient en faisant l'ACP pondérée du tableau conjoint F'UX. Deux sites sont proches, s'ils sont en même temps proches au sens de la répartition sur les espèces et au sens des variables descriptives. Sur le facteur 1, il y a une opposition très nette entre les sites éloignés de la mer qui se projettent du côté de DIST et ceux qui sont les plus salins qui se projettent à l'autre extrémité de l'axe. Ils se projettent aussi du côté où les espèces qu'ils contiennent sont les plus fréquentes.

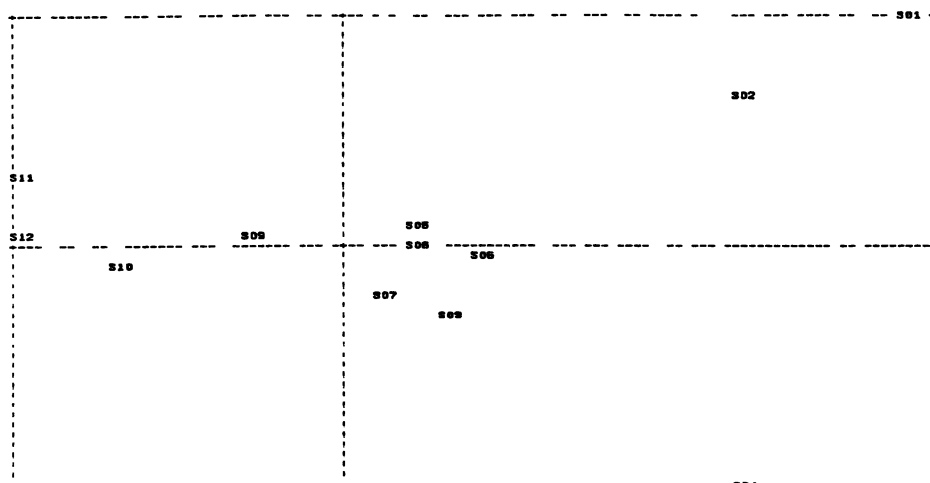


Fig. 3 : Représentation des sites au travers des espèces et des variables descriptives sur le plan 1x2

2. Représentation des sites vus au travers de leur répartition sur les espèces

Pour obtenir la représentation d'un site i_1 du nuage $N(I_1)$, il suffit de considérer le tableau $\tilde{F}' = F' \cup [0]$ en supplémentaire dans l'analyse globale, où 0 est le tableau de même taille que X , ayant tous ses termes nuls.

La coordonnée d'un site i_1 sur l'axe s obtenu dans l'ACP globale sera :

$$F_s(i_1) = \frac{1}{\lambda_1 \sqrt{\gamma_s}} \sum_j \frac{f_{ij}}{f_i} G_s(j)$$

ce qui fait de i_1 le barycentre des espèces qu'il contient à un coefficient près.

3. Représentation des sites vus au travers des variables descriptives (Fig. 4)

Pour obtenir la représentation des sites vus au travers des variables descriptives, on projette le tableau $\tilde{X} = [0, X]$ en supplémentaire dans l'ACP globale.

La projection d'un site i_2 du nuage N_{I_2} , sur l'axe s s'écrit :

$$F_s(i_2) = \frac{1}{\mu_1 \sqrt{\gamma_s}} \sum_k \frac{x_{ik} - \bar{x}_k}{s_k} H_s(k)$$

Représentation simultanée des sites pour les trois points de vue (Fig. 4)

Sur le facteur 1, les sites décrits par les espèces (qui sont notés S11 par exemple pour le site 1 du groupe 1) et ceux décrits par les variables descriptives sont quasi confondus, puisque c'est un facteur commun. Notons que le site 6 est très salin, mais a une répartition sur les espèces un peu moins marquée, que ne le laisse attendre les variables descriptives. En particulier, les sites 2 et 4 sont tous deux très salins, mais ont des répartitions sur les espèces qui sont très différentes. Cette différence s'explique par la présence de l'espèce "camp" pour le site 1 qui y est fréquente, alors que pour le site 4, elle est due à la présence de l'espèce "arme". On note S01 le site 1 vu au travers des deux groupes, S011 le site 1 vu au travers des espèces (groupe 1), et S012 le site 1 vu au travers des variables descriptives (groupe 2).

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCE

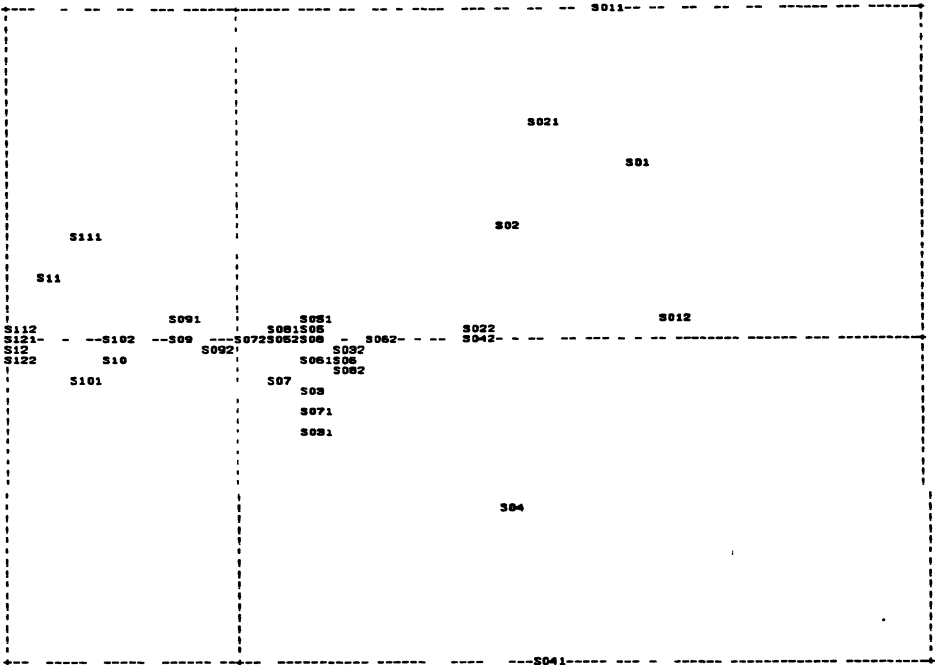


Fig. 4 : Représentation des sites vus au travers des trois points de vue sur le plan 1x2

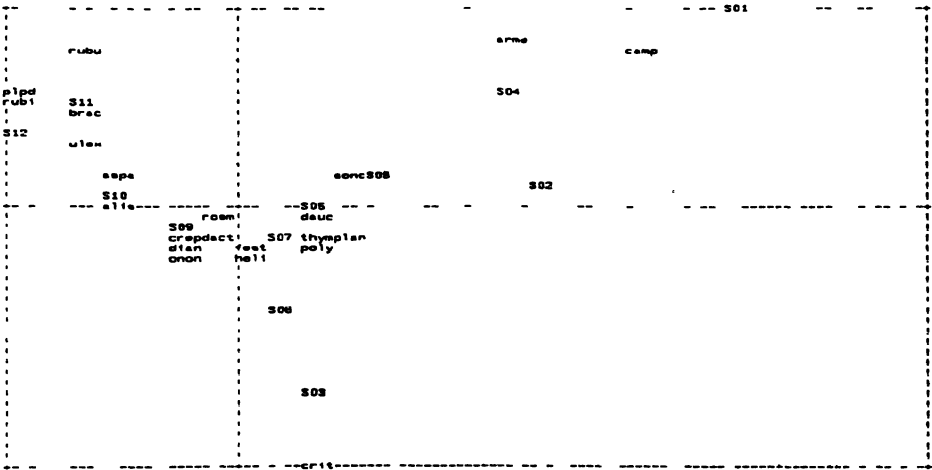


Fig. 5 : Représentation des espèces et des sites sur le plan 1x2 de l'ACC

ANALYSE FACTORIELLE MULTIPLE DE TABLEAUX DE FRÉQUENCE

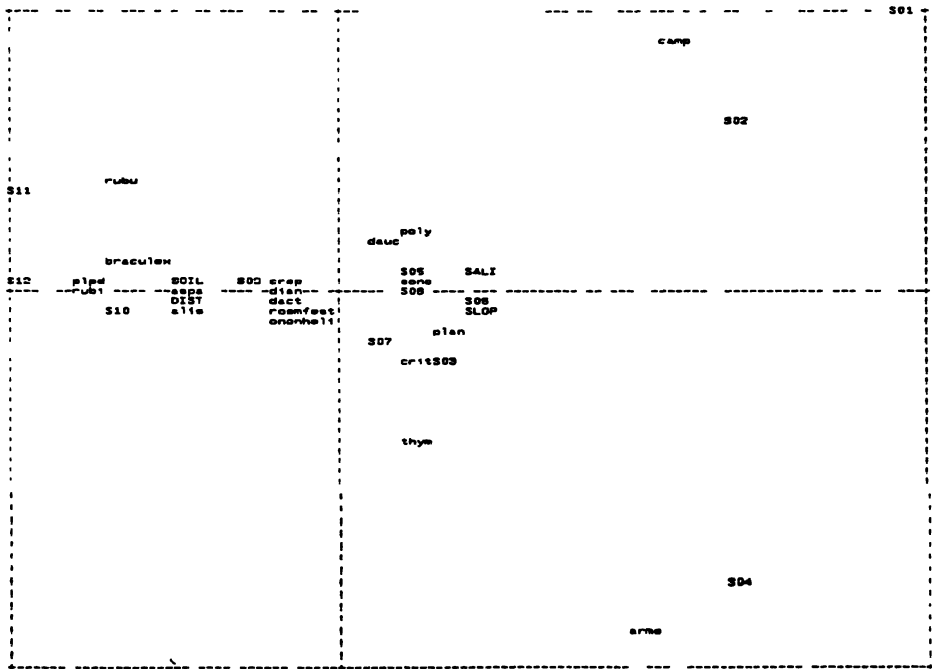


Fig. 6 : Représentation des espèces, des variables descriptives et des sites sur le plan 1x2 de l'AFM

VII. Comparaison avec l'Analyse Canonique des Correspondances (ACC)

L'ACC, Ter Braak (1986), permet d'étudier un tableau de fréquences et un tableau de variables numériques, mais ne prend en compte que la part de la structure associée au tableau de fréquence, qui est expliquée par les variables de X. L'ACC consiste à faire l'ACP des projections des colonnes du tableau F' présenté dans notre méthode, sur le sous espace engendré par les variables de X, avec les métriques D et M définies précédemment. L'ACC propose donc une approche non symétrique. L'AFM, quant à elle, permet une étude simultanée et non privilégiée des structures associées au tableau de fréquences et au tableau numérique. En plus, elle permet de détecter l'existence de structures communes et de structures spécifiques. L'AFM doit être un préalable à une ACC qui pourra être faite pour une analyse plus fine surtout si des structures communes ont été décelées.

L'application de l'AFM et l'ACC sur l'exemple présenté (Fig. 5 et 6), nous permet de trouver un premier facteur qui est le même dans les deux analyses. Ceci s'explique puisque le premier facteur de l'AFM est un facteur absolument commun aux deux groupes. Par contre les deuxième facteurs de l'AFM et de l'ACC sont très différents. Ceci est dû au fait que le deuxième facteur de l'AFM est un facteur de répartition des espèces qui est très peu lié aux variables descriptives, et qui ne peut apparaître dans l'ACC.

VIII. Généralisation à l'étude de plusieurs tableaux

Dans le cas d'un tableau de fréquences et de plusieurs tableaux descriptifs, la méthode présentée peut encore s'appliquer. Dans le cas où le tableau descriptif est constitué de variables qualitatives, les résultats utilisés dans l'AFM classique sont encore utilisables, à savoir l'équivalence entre l'ACM et l'ACP pondérée des indicatrices. Par ailleurs, pondérer les individus par f_i pour des variables qualitatives, ne pose pas de problème majeur.

On peut aussi envisager l'étude des liaisons entre plusieurs tableaux de fréquences et éventuellement d'autres tableaux descriptifs. Ce cas de figure se présente pour des tableaux de fréquences donnant des répartitions d'une population à des époques différentes pour la même variable, ou pour des tableaux de fréquences associés à des variables différentes...etc. Dans ce cas, la méthode que nous proposons ne peut s'appliquer que si les tableaux de fréquences ont les mêmes marges sur les lignes. Dans le cas contraire, il faudra se départir de la structure AFC de chacun des tableaux de fréquences, puisque les marges sur les lignes sont différentes ; on ne pourra pas restituer simultanément ces structures AFC, moyennant la transformation des tableaux de fréquences et l'introduction d'une même métrique sur les lignes. On peut alors soit se ramener à des tableaux de pourcentages, ou alors appliquer des méthodes ne faisant pas intervenir les marges sur les lignes, comme l'analyse par sous tableaux.

IX. Conclusion

Avec l'extension de l'AFM aux tableaux de fréquences, on peut traiter de façon simultanée et symétrique plusieurs tableaux de données de nature différente : ceux-ci peuvent être de type numérique, qualitatif ou être de type fréquence.

Par ailleurs, l'AFM et l'ACC sont deux méthodes complémentaires. On appliquera l'une ou l'autre des deux méthodes, selon que l'on privilégie une approche symétrique et dans ce cas, on effectuera une AFM, ou non symétrique et là, on utilisera l'ACC qui est plus adaptée.

D'autre part, si avec l'AFM, nous avons la possibilité de détecter des structures communes ou spécifiques, l'ACC viendra utilement affiner les résultats trouvés, essentiellement dans le cas où il existe des structures communes à au moins deux groupes.

TEXTES CITÉS EN RÉFÉRENCE

- ABDESSEMED L., ESCOFIER B. (1992) *Généralisation de l'analyse factorielle multiple aux tableaux de fréquences et comparaison avec l'analyse canonique des correspondances*, Bruxelles, XXIV^{es} Journées de Statistique de l'ASU, Mai.
- BENZÉCRI J.P. (1973) *L'analyse des données : II L'analyse des correspondances*, Paris, Dunod.
- DROUET D., ESCOFIER B. (1983) "Comparaison de plusieurs tableaux de fréquences", *Cahiers de l'analyse des données*, n° 4. Vol. VIII
- ESCOFIER B., PAGES J. (1990) *Analyses factorielles simples et multiples*, Paris, Dunod.
- GREENACRE M.J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press.
- LEBRETON J.D., CHESSEL D., PRODON R., YOCOZ N. (1988) "L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. Variables de milieu quantitatives", *Acta oecologica*, Oecol. Gen., Vol.9 n°1.
- LEBRETON J.D., CHESSEL D., RICHARDOT-COULET M., YOCOZ N. (1988) "L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. Variables de milieu quantitatives", *Acta oecologica*, Oecol. Gen., Vol.9 n°2.
- LEBRETON J.D., SABATIER R., BANCO G., BACOU A.M. (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*, Devillers and Karcher.
- TER BRAAK C.J.F. (1986) *Canonical Correspondence Analysis : a new Eigenvector Technique for Multivariate Direct Gradient Analysis*, Ecology.
- TER BRAAK C.J.F. (1987) *The Analysis of Vegetation-environment Relationships by Canonical Correspondence Analysis*, Ecology.
- TER BRAAK C.J.F. (1988) *A Classification and Related Methods of Data Analysis*, Bock H.H.

ANNEXE

Abréviations et noms des espèces

camp	:	Camphorosma monspeliaca
poly	:	Polycarpon polycarpoides
dauc	:	Daucus carota
sonc	:	Sonchus tenerrimus
plan	:	Plantago subulata
thym	:	Thymelea hirsuta
crit	:	Crithmum maritimum
fest	:	Festuca arvernensis
heli	:	Helichrysum stoechas
crep	:	Crepis bulbosa
dact	:	Dactylis glomerata
arme	:	Armeria ruscinonensis
rosm	:	Rosmarinus officinalis
onon	:	Ononis spinosa
ulex	:	Ulex parviflorus
aspa	:	Asparagus acutifolius
dian	:	Dianthus pyrenaicus
alis	:	Alyssum maritimum
brac	:	Brachypodium retusum
rubu	:	Rubus sp.
plpd	:	Polypodium australe
rubi	:	Rubia peregrina