REVUE DE STATISTIQUE APPLIQUÉE

ALAIN MÉOT
BERNADETTE LECLERC

Voisinages a priori et analyses factorielles : illustration dans le cas de proximités géographiques

Revue de statistique appliquée, tome 45, n° 3 (1997), p. 25-44
<http://www.numdam.org/item?id=RSA_1997__45_3_25_0>

Rev. Statistique Appliquée, 1997, **XLV** (3), 25-44

VOISINAGES *A PRIORI* ET ANALYSES FACTORIELLES : ILLUSTRATION DANS LE CAS DE PROXIMITÉS GEOGRAPHIQUES

Alain Méot (1) et Bernadette Leclerc (2)

(1) UPRES-A 6024 CNRS, «Laboratoire de Psychologie Sociale de la Cognition».
Université Blaise Pascal-Clermont-Ferrand II.
34, avenue Carnot. 63037 Clermont-Ferrand.
(2) Unité SAD Armorique. I.N.R.A. 65, rue de Saint Brieuc. 35042 Rennes Cedex.

RÉSUMÉ

Diverses analyses factorielles ont été proposées pour analyser les structures présentées par une ou deux séries d'observations multivariées, conditionnellement à l'existence de graphes de voisinage. Tout aussi bien les indices cibles à maximiser que les graphes utilisés ont petit à petit acquis des caractères très généraux. Dans cet article, nous rappelons tout d'abord un résultat permettant d'utiliser n'importe quel graphe dans ce type d'analyses. Les liens avec des approches antérieures ou plus classiques sont soulignés.

L'extrême généralité de ce cadre autorise des pratiques exploratoires souples que nous illustrons lors du couplage de deux tableaux dont l'un n'est défini que sur une fraction des individus de l'autre. Les voisinages utilisés combinent des informations sur des proximités géographiques et sur des ressemblances définies sur la base de caractéristiques statistiques supplémentaires.

Mots-clés : *Autocorrélation, voisinages* a priori, *graphes, analyses factorielles.*

ABSTRACT

Several factor analyses have been proposed to study the structures shown by one or two data sets, conditionally of the existence of a neighbourhood graph. Both the maximised indices and the neighbourhood graphs have became more and more general. In this paper, we first recall a general result which allows to use every kind of graph to do this type of analyses. The links with previous or more classical approaches are briefly underlined.

The extreme broadness of this frame allows very flexible exploratory practices. This is illustrated by the study of the links between two data tables, one of them containing only a fraction of the statistical units of the other. The used neighbourhoods combine geographical proximities and informations about supplementary statistical characteristics.

Keywords : *Autocorrelation,* a priori *neighbourhood, graphs, factor analysis.*

1. Introduction. Bibliographie

Divers travaux d'analyses factorielles ont utilisé comme indices cibles à maximiser des quantités dérivées des numérateurs des indices d'autocorrélation de Geary et de Moran.

Les premiers d'entre eux, dus à Lebart (1969, 1973), ont consisté à rechercher des combinaisons linéaires des variables d'un tableau de données qui maximisent successivement le numérateur de l'indice d'autocorrélation de Geary (Geary, 1954), calculé pour un graphe de contiguïté binaire (un individu est voisin ou non d'un autre) et symétrique. L'approche sera ensuite étendue à diverses formes dérivées de l'indice pour prendre en compte la nature du tableau étudié, et les diverses pondérations individuelles qui peuvent en découler («type» analyse des correspondances, analyse des correspondances multiple, ...).

Ces travaux ont ensuite été généralisés par Le Foll (1982) aux cas de pondérations quelconques symétriques des couples d'individus. La mise en évidence d'une semi-norme pour l'expression générale du numérateur de l'indice de Geary a amené cet auteur à utiliser le formalisme du schéma de dualité pour exprimer ce type d'approches. Dans une autre perspective, Carlier (1985) a proposé d'utiliser une forme particulière de voisinages temporels amenant à représenter de manière optimale des trajectoires multivariées. Plus récemment Méot (1992) et Méot *et al.* (1993) ont montré que l'indice maximisé pouvait s'écrire comme le produit scalaire entre la variable d'intérêt et sa transformée par un opérateur symétrique semi-défini positif. Cette écriture a conduit à proposer un schéma de dualité «dissymétrique» pour formaliser ces analyses. Les propriétés extrémales d'autocorrélation présentées par les vecteurs propres de l'opérateur ont par ailleurs été exploitées dans le cadre d'analyses en composantes principales sur variables instrumentales (ACPVI) pour décomposer un tableau de données en divers modèles orthogonaux présentant des structures liées à des autocorrélations positive, négative ou à l'absence d'autocorrélation.

De leurs cotés, Royer (1984), Switzer et Green (1984) et Wartenberg (1985) ont proposé le même type d'approches sur la base du numérateur du I de Moran, autre indice très classique d'autocorrélation (Moran (1948)). Tout aussi bien les formalisations que les généralisations qui ont pu être proposées de ces analyses sont restées longtemps peu importantes. L'extension des propriétés du schéma de dualité dissymétrique donné par Méot *et al.*, 1993, lorsque sont utilisés de simples opérateurs symétriques (éventuellement de simple matrices symétriques «symétrisées» pour la norme utilisée dans l'espace des variables), en est cependant assez immédiate (Chessel et Sabatier (1994)). La forme quadratique maximisée a dans ce cas pour particularité de n'être pas obligatoirement positive. Notons que l'étude des propriétés d'un opérateur symétrique conduit aussi à des résultats généraux de maximisation de l'autocorrélation, au sens d'un indice du type du numérateur du I; ici aussi des stratégies de type ACPVI sur vecteurs propres associés à la matrice d'un graphe symétrique pourront donc être utilisées.

Thioulouse *et al.* (1996) ont par ailleurs proposé une pondération astucieuse amenant à une décomposition de la variance (inertie) comme somme d'un indice ayant la forme du numérateur de l'indice de Geary et d'un indice de la forme du numérateur de celui de Moran. Les deux types d'analyses précédentes se trouvent

ainsi reliées dans ce cas. Bien que non explicité dans leur texte, il est clair que cette décomposition pourrait être aisément étendue à des graphes pondérés symétriques.

L'avancée la plus décisive est cependant certainement celle proposée par Dolédec *et al.* (1996) dans le cadre de l'analyse d'un triplet d'analyse des correspondances (AFC). Il s'agit dans ce cas de coupler l'analyse de deux tableaux ne faisant pas référence aux mêmes individus (ni mêmes objets conceptuels dans ce cas) à l'aide d'une troisième table qui sert de lien entre ces tableaux. Cette analyse relève d'une approche générale permettant de coupler deux tableaux quelconques sous contrainte d'un voisinage lui aussi quelconque, en particulier non symétrique.

L'objet de cet article est double : (i) Donner les résultats essentiels qui permettent de mener ce dernier type de couplage en soulignant les liens existant avec des approches antérieures (ii) Développer une illustration concernant l'intégration de voisinages géographiques, qui furent à l'origine des premiers travaux, et restent importants pour prendre en compte la dimension spatiale en multivarié.

2. Théorie

2.1. Cadre général

Soit deux variables statistiques $\mathbf{x} = (x_1, \ldots x_n)$ et $\mathbf{y} = (y_1, ..., y_m)$ mesurées respectivement sur n et m individus et notées sous forme de vecteurs colonnes. Soit $\mathbf{W}$ une matrice à n lignes et m colonnes de terme général réel w_{ik} ($i = 1, \ldots, n$; $k = 1, .., m$). La quantité :

$$C_g(\mathbf{x}, \mathbf{y}) = \mathbf{x^t W y} = \Sigma_{i,k} w_{ik} x_i y_k$$

sera définie en tant que covariance généralisée entre les variables $\mathbf{x}$ et $\mathbf{y}$ (^{t} désigne la transposition d'un vecteur ou d'une matrice).

Les cas les plus courants de covariance généralisée sont ceux relatifs aux indices d'autocorrélation de Geary et de Moran (Geary 1954; Moran 1948) ou leurs équivalents géostatistiques, variogrammes et covariogrammes croisés observés (e.g. Matheron (1965), Journel et Huijbreghts (1978)). Dans ces cas, on a le même nombre de mesure ($m = n$) ainsi qu'un centrage préalable pour le second de ces indices. Dolédec *et al.* (1996) utilisent le cas général à partir d'une problématique écologique.

Considérons maintenant deux tableaux de données $\mathbf{X}$ et $\mathbf{Y}$ comprenant n (respectivement m) lignes et p (resp. q) colonnes correspondant aux mesures de p (resp. q) variables sur n (resp. m) individus. Les espaces de représentation des lignes de $\mathbf{X}$ et $\mathbf{Y}$, $\mathbb{R}^p$ et $\mathbb{R}^q$, sont munis de deux produits scalaires notés $\mathbf{Q_X}$ et $\mathbf{Q_Y}$.

Le tableau $\mathbf{X^t W Y}$, de taille $p \times q$, contient les covariances généralisées entre les variables de $\mathbf{X}$ (en lignes) et celles de $\mathbf{Y}$ (en colonnes).

L'analyse du triplet ($\mathbf{X^t W Y}$, $\mathbf{Q_Y}$, $\mathbf{Q_X}$) permet de construire, par projection en éléments supplémentaires des lignes de $\mathbf{X}$ et $\mathbf{Y}$ sur les axes et composantes principales, une combinaison linéaire des colonnes de $\mathbf{X}$, notée $\mathbf{L_X}$, et une combinaison linéaire des colonnes de $\mathbf{Y}$, notée $\mathbf{L_Y}$, de covariance généralisée maximum :

$$C_g(\mathbf{L_X}, \mathbf{L_Y}) = \mathbf{L_X^t W L_Y} = \Sigma_{i,k} w_{ik} \mathbf{L_X}_i \mathbf{L_Y}_k$$

En effet, les coordonnées factorielles lignes du triplet sont données par les $\mathbf{Q_Y}$-projections orthogonales des lignes de $\mathbf{X^tWY}$ sur les axes principaux :

$$\mathbf{L = X^tWYQ_Yu} \qquad \mathbf{u} \text{ étant un vecteur axial factoriel}$$

Elles maximisent successivement :

$$||\mathbf{L}||^2_{\mathbf{Q_X}} = \mathbf{L^tQ_XL} = \mathbf{u^tQ_YY^tW^tXQ_XX^tWYQ_Yu}$$

Les composantes principales normées étant égales à $\mathbf{v} = \mathbf{L}/\sqrt{\lambda} = \mathbf{X^tWY}$ $\mathbf{Q_Yu}/\sqrt{\lambda}$, où λ est la valeur propre correspondante, on a :

$$||\mathbf{L}||^2_{\mathbf{Q_X}} = \sqrt{\lambda}\ \mathbf{v^tQ_XX^tWYQ_Yu} = \sqrt{\lambda}\ \mathbf{L_X^tWL_Y} = \lambda$$

avec $\mathbf{L_X = XQ_Xv}$ et $\mathbf{L_Y = YQ_Yu}$.

Et donc $\mathbf{L_X^tWL_Y} = \sqrt{\lambda}$ est maximisé.

2.2. Liens avec d'autres approches

Le jeu sur les paramètres $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, $\mathbf{Q_X}$ et $\mathbf{Q_Y}$ introduit évidemment une diversité extrême de cas particuliers possibles qui prennent alors tout leur intérêt ... en tant que cas particuliers.

Si $\mathbf{Y = X}$, $\mathbf{Q_X = Q_Y}$, et $\mathbf{W}$ est associée à une pondération des lignes de $\mathbf{X}$, alors $\mathbf{L_X}$ est égale aux coordonnées factorielles lignes de l'ACP($\mathbf{X}$, $\mathbf{Q_X}$, $\mathbf{W}$). Développer par ce biais cette dernière approche est évidemment sans grand intérêt.

De même, si $\mathbf{Y = X}$, $\mathbf{Q_X = Q_Y}$ et $\mathbf{W}$ est associé à un opérateur symétrique (pour une pondération à définir par exemple), on retrouve les éléments de l'analyse d'un schéma de dualité dissymétrique. Là aussi, il est cependant à priori sans grand intérêt d'utiliser ce cadre pour réaliser ce type d'analyse.

Dans le cas où $\mathbf{Y = X}$, $\mathbf{Q_X = Q_Y}$ et $\mathbf{W}$ est un tableau quelconque à n lignes et n colonnes, cette formalisation permet d'aborder l'étude de l'autocorrélation multivariée dans le cas où le graphe de voisinage n'est pas symétrique. Ce cas est courant en univarié (e.g. Cliff et Ord (1973), Upton et Fingleton (1985)).

Si $\mathbf{Y}$ et $\mathbf{X}$ ont le même nombre de lignes, $\mathbf{Q_X \neq Q_Y}$ et $\mathbf{W}$ est associé à une pondération des lignes communes aux deux tableaux, on retrouve les approches de co-inertie (Chessel et Mercier (1993)).

En utilisant dans ce dernier cas une matrice $\mathbf{W}$ quelconque (symétrique, «symétrisée» pour une pondération individus ou général), on pourra aborder des analyses de co-structure spatiales (Chessel et Sabatier (1994), dans le cas de graphe de contiguïté).

Si $\mathbf{Y}$ et $\mathbf{X}$ n'ont pas le même nombre de lignes et $\mathbf{W}$ n'est pas carré, on pourra aborder l'étude des ressemblances multivariées sous contraintes de voisinages-liens quelconques. La définition de $\mathbf{Q_X}$ et $\mathbf{Q_Y}$ pourra se faire de diverses manières utilisant les propriétés de $\mathbf{X}$ et $\mathbf{Y}$ et des analyses à conduire. Notons enfin que se pose ici de

manière explicite ou non la question des relations entre les trois triplets statistiques $(\mathbf{X}, \mathbf{Q_X}, \mathbf{D_n})$, $(\mathbf{Y}, \mathbf{Q_Y}, \mathbf{D_m})$ et $(\mathbf{W}, \mathbf{D_m}, \mathbf{D_n})$ où $\mathbf{D_n}$ et $\mathbf{D_m}$ sont des métriques-pondérations dans les espaces de représentation des variables de $\mathbf{X}$ et $\mathbf{Y}$. Dolédec *et al.* ont choisi d'utiliser les métriques définies par l'AFC du tableau $\mathbf{W}$. L'illustration qui suit les définit sur la base des tableaux $\mathbf{X}$ et $\mathbf{Y}$.

2.3. Remarques

1) L'extrême diversité des approches permises par le jeu sur les divers paramètres des trois triplets a pour corollaire des pratiques de dépouillement non moins diverses. On pourra bien sûr cartographier les coordonnées factorielles lorsque le voisinage est géographique, représenter les coefficients des combinaisons linéaires utilisées ($\mathbf{Q_Y u}$ et $\mathbf{Q_X v}$), calculer des corrélations entre colonnes des divers tableaux et les coordonnées ($\mathbf{L_X}$ et $\mathbf{L_Y}$), positionner les axes des analyses séparées par rapport à ceux de l'analyse globale, ... On trouvera quelques exemples de ces pratiques dans l'illustration qui suit ainsi que dans l'article de Dolédec *et al.*

2) Notons enfin une propriété générale qui peut conduire à des pratiques faisant intervenir des choix sur une seule partie des structures présentées par la matrice $\mathbf{W}$.

Soient deux normes $\mathbf{D_n}$ et $\mathbf{D_m}$ dans $\mathbb{R}^n$ et $\mathbb{R}^m$. Les matrices :

$$\mathbf{S_m} = \mathbf{D_m^{-1} W^t D_n^{-1} W} \quad \text{et} \quad \mathbf{S_n} = \mathbf{D_n^{-1} W D_m^{-1} W^t}$$

sont respectivement $\mathbf{D_m}$ et $\mathbf{D_n}$-symétriques. Leurs vecteurs propres associés aux mêmes $\mathbf{k}$ valeurs propres non nulles constituent une suite de couples de vecteurs orthonormés de $\mathbb{R}^m$ et $\mathbb{R}^n$, $((\mathbf{g_1}, \mathbf{f_1}), (\mathbf{g_2}, \mathbf{f_2}), \ldots (\mathbf{g_k}, \mathbf{f_k}))$ qui maximisent successivement la covariance généralisée :

$$C_g(\mathbf{f_l}, \mathbf{g_l}) = \mathbf{f_l^t W g_1} = \sqrt{\lambda_l}$$

où λ_1 est la valeur propre associée au l-ième couple ($l = 1, \ldots, k$).

En effet ces vecteurs, qui sont reliés par les deux équations de passage suivantes :

$$\begin{cases} \mathbf{W g_l} = \sqrt{\lambda_l}\, \mathbf{D_n f_l} \\ \mathbf{W^t f_l} = \sqrt{\lambda_l}\, \mathbf{D_m g_l} \end{cases}$$

peuvent être considérés comme renvoyant aux facteurs et cofacteurs dans l'analyse du triplet $(\mathbf{W}, \mathbf{D_m^{-1}}, \mathbf{D_n^{-1}})$. Les coordonnées lignes de cette analyse sont égales à $\mathbf{W g_l}$ et sont successivement de $\mathbf{D_n^{-1}}$ -norme maximale :

$$||\mathbf{W g_l}||^2_{\mathbf{D_n^{-1}}} = \sqrt{\lambda_l}\ \mathbf{f_l^t W g_l} = \lambda_l$$

et donc $\quad C_g(\mathbf{f_l}, \mathbf{g_l}) = \mathbf{f_l^t W g_l} = \sqrt{\lambda_l}$ $\quad$ est maximisée

Cette propriété, utilisée de manière classique en AFC pour la recherche d'un couple de codes numériques lignes et colonnes qui sont de corrélation canonique maximale (e.g. Esteve (1978)), entraîne que le couplage des triplets $(\mathbf{X}, \mathbf{Q_X}, \mathbf{D_n})$ et

$(\mathbf{Y}, \mathbf{Q_Y}, \mathbf{D_m})$ sous la contrainte du graphe $\mathbf{W}$ dépend à la fois de la qualité de la représentation des variables de chacun des tableaux $\mathbf{X}$ et $\mathbf{Y}$ par $\mathbf{D_n}$ $(\mathbf{D_m})$-projections sur les vecteurs $\mathbf{f_l}$ et $\mathbf{g_l}$ et de la grandeur de la valeur propre associée. En effet, l'analyse du triplet $(\mathbf{W}, \mathbf{D_m^{-1}}, \mathbf{D_n^{-1}})$ conduit à une formule de reconstitution des données en fonction des facteurs $(\mathbf{g_l})$ et cofacteurs $(\mathbf{f_l})$ qui s'écrit de la manière suivante :

$$\mathbf{W} = \mathbf{D_n F}\, \Lambda^{1/2}\, \mathbf{G^t D_m}$$

où $\mathbf{F}$ et $\mathbf{G}$ sont les matrices contenant en colonnes les cofacteurs et facteurs et Λ est la matrice diagonale des valeurs propres correspondantes.

On a alors :

$$\mathbf{X^t W Y} = \mathbf{X^t D_n F}\, \Lambda^{1/2}\, \mathbf{G^t D_m Y}$$

Les matrices $\mathbf{F^t D_n X}$ et $\mathbf{G^t D_m Y}$ contiennent les coefficients des projections orthogonales des colonnes de $\mathbf{X}$ et $\mathbf{Y}$ sur les cofacteurs et facteurs.

Deux types de conséquences découlent de cette écriture.

(i) On pourra approximer $\mathbf{W}$ par quelques couples facteurs-cofacteurs qui auront été choisis selon leurs qualités de représentations des colonnes de $\mathbf{X}$ et $\mathbf{Y}$. Ceci permettra en particulier de se «débarrasser» des bruits que peut éventuellement contenir la matrice $\mathbf{W}$. Notons cependant que la matrice $\Lambda^{1/2}$ sert déjà de filtre implicite à ces bruits.

(ii) Le couplage des triplets associés à $\mathbf{X}$ et $\mathbf{Y}$ sous la contrainte du graphe $\mathbf{W}$ correspond aussi à un couplage type analyse de co-inertie des paramètres des modèles de régression linéaire des variables de ces tableaux sur les codes numériques de $\mathbb{R}^n$ et $\mathbb{R}^m$ qui maximisent successivement la covariance généralisée sous contraintes de $\mathbf{D_n}$ et $\mathbf{D_m}$-orthonormalités.

Ainsi, si l'on note $\mathbf{Z} = \mathbf{F^t D_n X}$ et $\mathbf{B} = \mathbf{G^t D_m Y}$, l'analyse du triplet $(\mathbf{X^t W Y}, \mathbf{Q_Y}, \mathbf{Q_X})$ est équivalente à celle du triplet $(\mathbf{Z^t}\, \Lambda^{1/2}\, \mathbf{B}, \mathbf{Q_Y}, \mathbf{Q_X})$. Les combinaisons linéaires $\mathbf{L_Z} = \mathbf{Z Q_X v_l} = \mathbf{F^t D_n L_X}$ et $\mathbf{L_B} = \mathbf{B Q_Y u_l} = \mathbf{G^t D_m L_Y}$ maximisent alors successivement la quantité $\mathrm{Cg}(\mathbf{L_Z}, \mathbf{L_B}) = \mathbf{L_Z^t}\, \Lambda^{1/2}\, \mathbf{L_B} = \Sigma_{i=1,k} \sqrt{\lambda_i}\, \mathbf{L_{Z}}_i \mathbf{L_{B}}_i$. Cette propriété permettra de juger s'il existe des ressemblances entre les paramètres des variables de $\mathbf{X}$ sur les codes $\mathbf{F}$ et ceux des colonnes de $\mathbf{Y}$ sur les codes $\mathbf{G}$. Elle pourra être utilisée comme un raccourci à l'analyse sur les tableaux bruts.

Notons enfin que la pondération $\Lambda^{1/2}$ assure ici aussi qu'interviendront plus fortement les codes les plus «importants» dans la représentation de la matrice $\mathbf{W}$; on pourra cependant essayer de la remplacer par une pondération uniforme pour avoir une co-inertie «non sélective» entre paramètres.

3) Les calculs et les figures de l'illustration qui suit ont été faits par l'intermédiaire du logiciel ADE (Thioulouse *et al.*, 1995), pour partie en utilisant des modules standards et pour partie en programmant certaines applications sous cet environnement.

3. Illustration : étude des relations entre deux tables partiellement appariées

Les sciences biologiques sont fortement génératrices d'observations localisées dans l'espace et (ou) le temps. L'intégration de voisinages *a priori* permet alors d'explorer certaines hypothèses concernant l'effet des proximités spatiales ou temporelles sur les ressemblances entre individus ou variables. Il est possible de définir des pondérations des couples d'individus tout aussi bien sur la base de proximités géographiques simples -qui traduiront par exemple l'effet de l'éloignement, de l'altitude, de «séparations» physiques entre les individus, ...- que sur des informations beaucoup plus fines ayant trait à une connaissance de ressemblances inter- individus basées sur des critères annexes à ceux que l'on souhaite décrire.

Ces données sont souvent «sales» (Legendre, 1993). La récolte des informations subit en effet de nombreuses contraintes qui vont de classiques questions de coûts des mesures à celles encore plus difficiles tenant à la nécessité de ne pas trop perturber l'environnement des objets que l'on souhaite décrire sous peine de changer la nature même de ces objets et de leurs interrelations. Les outils statistiques destinés à l'exploration des structures présentes dans les données doivent alors faire preuve d'une très grande adaptabilité qui nécessite en particulier la possibilité d'intégrer certaines hypothèses venues des champs biologiques mais non «prouvées» statistiquement.

L'illustration que nous proposons provient du champ agronomique et entre pleinement dans ces remarques. Il s'agit essentiellement de tirer le plus d'informations possible d'une expérience lourde ayant plusieurs objectifs hiérarchisés induisant des contraintes très fortes sur la qualité et la quantité d'information disponible pour les objectifs de niveaux «inférieurs». C'est bien évidemment l'un de ces niveaux que nous présentons ici. Notre approche permet de proposer une solution simple à un problème courant des données de terrain qui concerne l'existence d'une dissymétrie dans le nombre d'observations relatives à deux types d'informations inter-reliées.

Dans cette recherche, une grande parcelle a été partitionnée en 94 placettes (figure 1) sur chacune desquelles on a observé, un jour par semaine durant les saisons de pâturage 1983 à 1985, le nombre de vaches dans l'activité de pâturage. Ce nombre est égal à la somme cumulée des animaux vus dans cette activité chaque quart d'heure de la journée de pâturage. De manière parallèle, la croissance de la phytomasse (exprimée en Kg de matière sèche par hectare) a été mesurée sur 7 transects. La position géographique de ceux-ci (figure 1) a été définie de manière à caractériser au mieux la diversité des zones floristiques présentes sur la parcelle. La lourdeur du dispositif, nécessitant la mise en défens (hors pâturage) de ces transects, explique que leur nombre ait été limité à 7. Les mesures de production ne sont pas réalisées avec la même fréquence que celles concernant le pâturage. La connaissance intuitive des rythmes de croissance en fonction de divers paramètres (météorologie, saisons, ...) conduit en effet à supposer l'homogénéité de la croissance durant certaines périodes, qui vont de 6 à 12 jours au printemps et de 15 à 22 jours en été et en automne. Les dates où ont été mesurées les occupations sont alors considérées avoir la même productivité végétale que la date la plus proche de mesure de la croissance présentant les mêmes conditions sur le plan de la production végétale. Enfin, en 1982 a été réalisé un relevé exhaustif de la flore sur toutes les placettes qui consiste en l'attribution d'une note de recouvrement (allant de 1 à 5 et traduisant diverses classes de recouvrement) aux

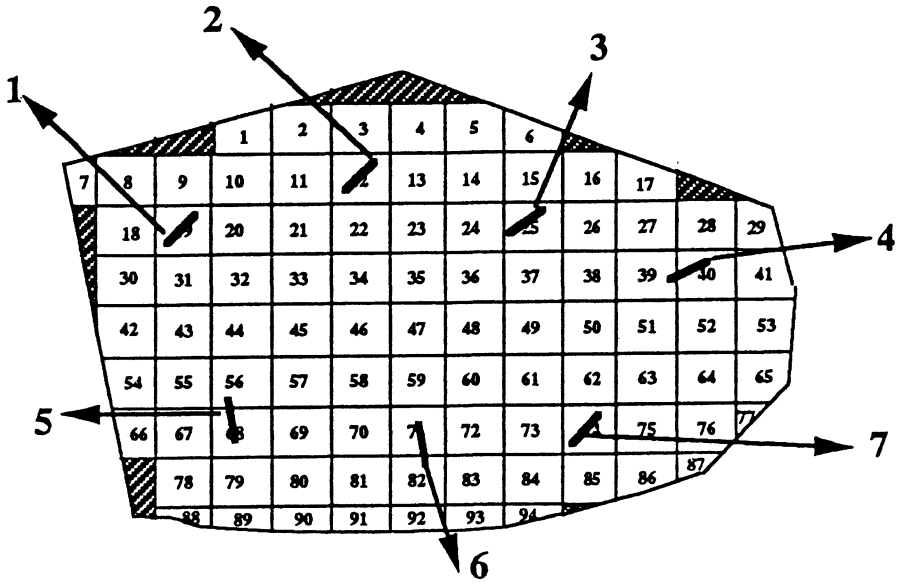

FIGURE 1

Schématisation de la surface utilisée lors des observations. Le quadrillage définit 94 placettes dont la numérotation fait référence aux numéros de lignes qui leurs sont affectées dans le tableau des occupations et dans celui des relevés de végétation. Les traits pleins présents sur certaines d'entre elles sont les transects utilisés pour mesurer la productivité végétale, le chiffre en bout de flèche désigne le numéro du transect correspondant ainsi que les numéros de lignes dans le tableau des mesures des croissances végétales.

principales espèces végétales présentes sur la placette. Ces données sont disponibles auprès du premier auteur.

L'objectif de ce travail est de chercher à comprendre comment les vaches utilisent l'espace de la parcelle face aux fluctuations spatiales et temporelles de l'offre alimentaire. La croissance renseigne sur l'aspect quantitatif de cette dernière alors que les recouvrements des principales espèces en donnent une idée plus qualitative. La dissymétrie existant entre les lieux où sont relevées les occupations et les données de croissance force à faire certaines hypothèses sur les ressemblances existant entre placettes sur le plan végétal. De mêmes que celles concernant l'homogénéité des productions végétales durant certaines périodes, ces hypothèses sont assumées sur le plan biologique lors de la définition du plan d'observation. Les solutions statistiques à l'exploration de la liaison entre les utilisations et la croissance, tout en gardant des possibilités critiques à leur égard, doivent donc tenter de les intégrer.

L'information utilisée est ainsi contenue dans trois tableaux correspondant aux dénombrements des animaux vus pâturant sur les placettes, aux mesures de croissance végétale sur les transects et enfin aux notes de recouvrement. Le premier contient donc 94 lignes et 51 colonnes correspondant aux placettes et jours d'observation (18 en 83, 17 en 84 et 16 en 85). Le second comprend 7 lignes correspondant aux transects et 51

colonnes relatives aux mêmes jours d'observation que pour le tableau des occupations. Enfin, le troisième, qui ne sera pas utilisé directement sous cette forme, comprend 94 lignes-placettes et 7 colonnes correspondant aux notes de présence des 6 principales espèces végétales (Dactyle, Agrostis, Fléole, Trèfle blanc, Pissenlit, Ray-grass) et de la terre nue.

Divers facteurs affectant le nombre d'animaux présents sur la parcelle au cours d'une journée (variation de la taille du troupeau, ouverture de parcelles contiguës), nous avons considéré le tableau centré-normé des occupations (avec la pondération 1/94). Les caractéristiques de croissance ont quant à elles été simplement centrées par dates (avec la pondération 1/7).

Par rapport aux notations du paragraphe 2.3, on a donc $\mathbf{D_n} = 1/94\,\mathbf{I_{94}}$, $\mathbf{D_m} = 1/7\,\mathbf{I_7}$, $\mathbf{X} = \mathbf{X}^*$ où $\mathbf{X}^*$ est le tableau normé des occupations et $\mathbf{Y} = \mathbf{Y}^\circ$ où $\mathbf{Y}^\circ$ est le tableau centré des notes de croissance sur les transects.

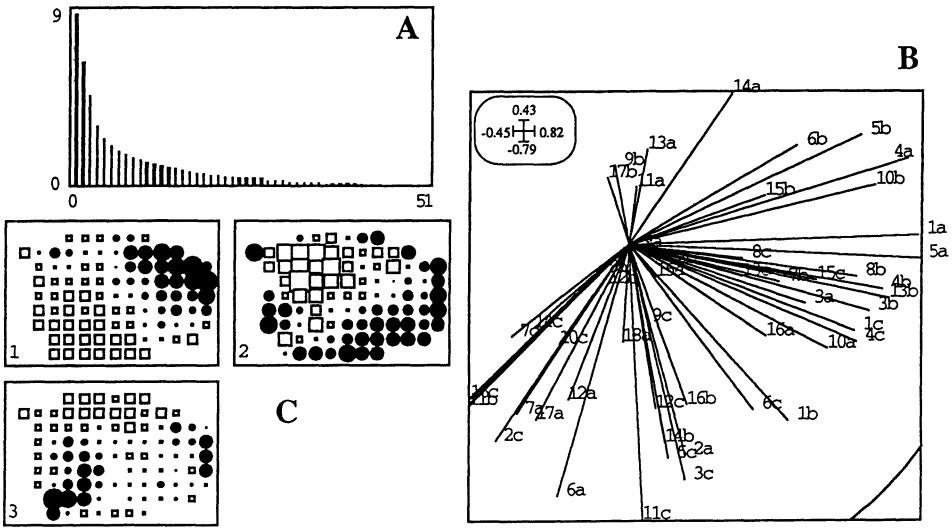

FIGURE 2

ACPN du tableau des occupations. A : Courbe des valeurs propres. B. Une partie du cercle des corrélations des dates avec les 2 premières composantes principales; les dates sont numérotées selon leur ordre d'apparition dans l'année, chaque année étant repérée par une voyelle (a = 83, b = 84, c = 85)*. C : Cartographie des coordonnées des placettes. Cercle = coordonnée positive; carré = coordonnée négative; cercles et carrés sont de taille proportionnelle à la valeur de la coordonnée. Les numéros des axes sont situés en bas à gauche de chaque fenêtre.*

Une première étape a été celle de la description des principales caractéristiques des deux tableaux à travers des analyses séparées. Les valeurs propres (figure 2A) de l'analyse en composantes principales normée (ACPN) des occupations montrent qu'il existe trois répartitions spatiales principales dont la première est tout à fait prédominante (figure 2C). Une assez forte hétérogénéité spatiale au sein des zones

qui sont opposées sur ces axes est cependant la règle générale. Les ressemblances temporelles (figure 2B) paraissent quant à elles être beaucoup plus liées à la particularité des dates d'observations qu'à l'existence d'effets saisonniers, annuels ou autres.

L'analyse en composantes principales centrée (ACPC) du tableau des croissances montre quant à elle une seule structure tout à fait dominante (figure 3A). Celle-ci repose principalement sur le transect n°3 situé au nord-est de la parcelle qui est en général plus productif que les autres (figure 3C). Certaines dates, parfois très rapprochées comme en 84, présentent plus particulièrement cette caractéristique (figure 3B), sans ici aussi mettre toutefois en évidence l'existence d'effets saisonniers ou annuels.

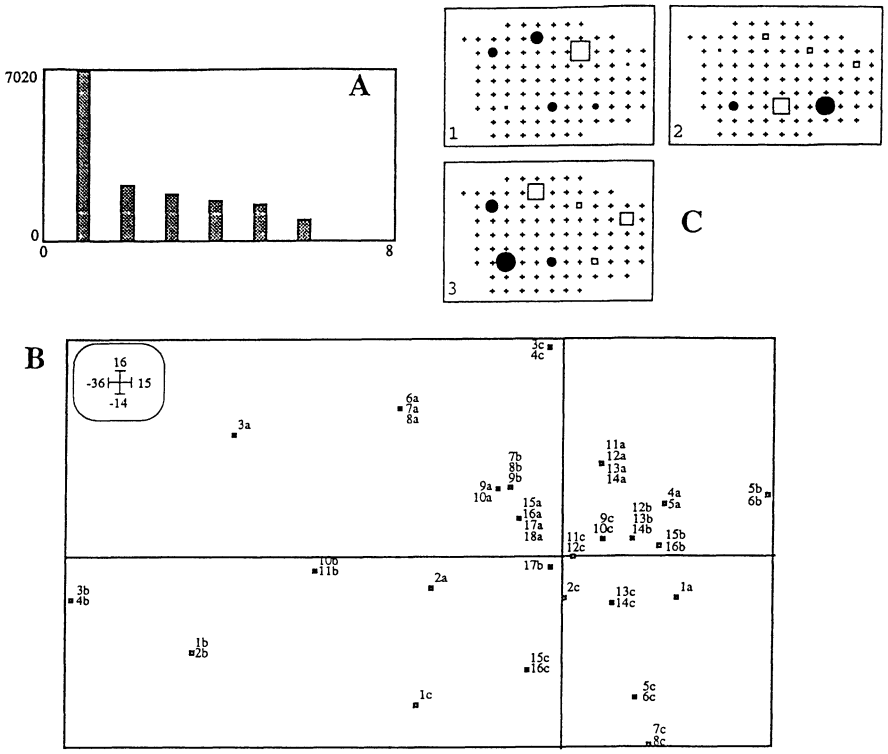

FIGURE 3

ACPC du tableau des croissances. A : Courbe des valeurs propres. B : Coordonnées des dates. C : Cartographie des coordonnées des placettes; mêmes représentations que sur la figure 2.

Une seconde étape a consisté à coupler directement le tableau des occupations avec celui des croissances végétales. Cette analyse (notée ANA1) correspond à la projection en éléments supplémentaires des lignes de $\mathbf{Y}$ et $\mathbf{X}$ sur les axes et composantes principales du triplet $(\mathbf{X}^t\mathbf{W}_1\mathbf{Y}, \mathbf{I}_{51}, \mathbf{I}_{51})$ où W1 est une matrice à

94 lignes et 7 colonnes telle que :

$\mathbf{W_{1}}_{ik} = 0$ si la placette i ne correspond pas au transect k
$\mathbf{W_{1}}_{ik} = 1$ sinon

pour $i = 1, \ldots, 94$ et $k = 1, \ldots 7$.

Cette analyse conduit à construire une série de couples de combinaisons linéaires des variables de $\mathbf{X}(\mathbf{L_X})$ et de $\mathbf{Y}(\mathbf{L_Y})$ qui maximisent successivement la quantité :

$$\mathrm{Cg}(ANA1) = \mathbf{L}_{\mathbf{X}}^{\mathbf{t}} \mathbf{W_1} \mathbf{L_Y} = \Sigma_{i,k} \mathrm{w}_{1ik} \mathbf{L}_{\mathbf{X}i} \mathbf{L}_{\mathbf{Y}k}.$$

Notons que ce couplage équivaut à celui du tableau à 7 lignes et 51 colonnes correspondant à la sélection dans le tableau des occupations des seules lignes associées aux transects avec le tableau des croissances. Les autres lignes du tableau des occupations n'interviennent en effet pas comme éléments actifs dans cette analyse (c'est-à-dire dans le calcul des covariances généralisées) et sont simplement projetées en lignes supplémentaires sur les axes principaux.

La construction de la covariance généralisée sur la base des seules placettes où existe un transect conduit à souligner fortement les structures locales communes aux deux tableaux. La première paire de coordonnées lignes met ainsi fortement en évidence la spécificité du transect 3 et, au niveau des occupations (figure 4C), des placettes immédiatement contiguës. Une croissance très importante sur ce transect alliée à une occupation privilégiée de la zone alentour sont effectivement des traits communs à nombre de dates (figure 4B). La particularité de cette structure prend place dans une opposition générale de l'est à l'ouest de la parcelle. Les coordonnées des placettes de ces deux parties de la parcelle sont toutefois extrêmement hétérogènes. La seconde paire de coordonnées souligne quant à elle plus particulièrement le transect 4 ainsi, qu'ici aussi au niveau des occupations, les placettes immédiatement adjacentes à ce transect. Apparaît ainsi une opposition entre le nord-est de la parcelle, pour lequel les coordonnées sont très homogènes, et le sud-ouest, où la variabilité est beaucoup plus importante. Comme le laisse supposer la faible valeur de la seconde valeur propre (figure 4A), cette seconde caractéristique ne concerne qu'un très petit nombre de dates.

La comparaison avec les résultats des analyses séparées montre de manière assez évidente que la dépendance de la covariance généralisée par rapport aux seuls transects où est mesurée la croissance conduit à exagérer l'impact des croissances les concernant sur la structure des occupations. En effet, alors que les premières coordonnées lignes relatives au tableau des croissances correspondent exactement à celles de l'ACPC de ce tableau, celles définies par les occupations sont assez lointaines de leur ACPN initiale. Tout en conservant un lien avec celles de l'ACPN, ces coordonnées soulignent beaucoup plus que celles-ci le côté local des occupations en opposant tout particulièrement les zones entourant certains transects. La forte corrélation (-0.57) existant entre les deux premières coordonnées relatives au tableau des occupations dans cette analyse souligne d'ailleurs le côté quelque peu artificiel de la séparation entre ces deux axes.

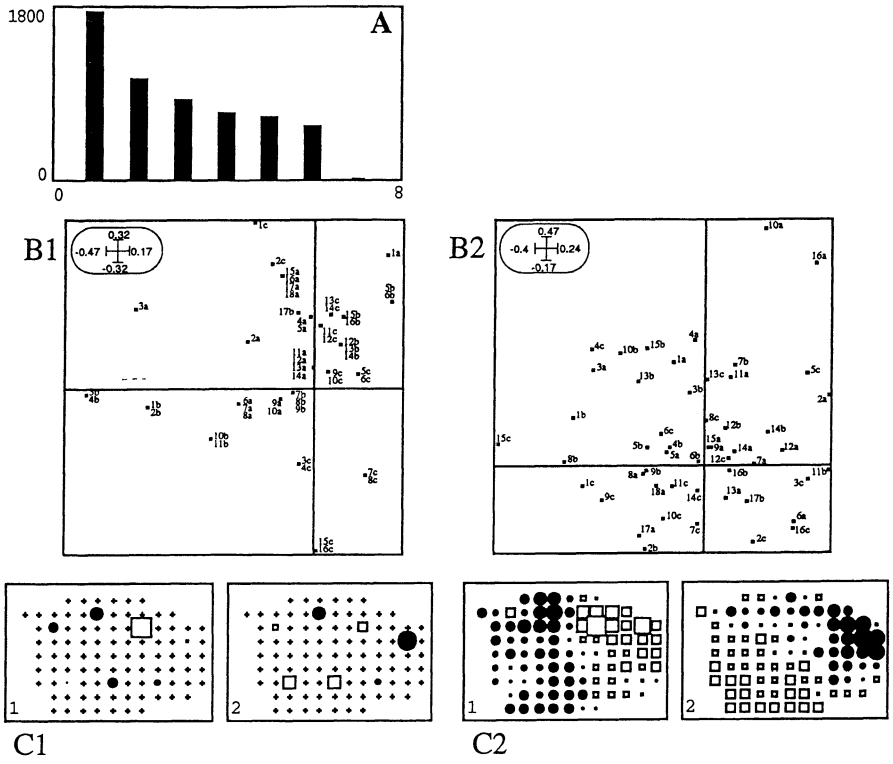

FIGURE 4

Analyse du triplet ($\mathbf{X^t W_1 Y}$, $\mathbf{I_{51}}$, $\mathbf{I_{51}}$) *(analyse ANA1). Voir le texte pour la construction de* $\mathbf{W_1}$. *A : Courbe des racines carrés des valeurs propres. B1 et B2 : Coefficients (facteurs) des combinaisons linéaires des variables pour les parties croissances (à gauche) et occupations (à droite). C1 et C2 : Coordonnées lignes pour les croissances (à gauche) et les occupations (à droite).*

L'existence d'une grande homogénéité d'occupation autour des transects tend cependant aussi à démontrer la présence d'une autocorrélation positive entre les placettes pour les caractéristiques de production. Cette autocorrélation ne pouvant être mesurée sur la base des seules variables de production, il est alors possible d'introduire certaines hypothèses la concernant dans l'analyse à travers l'utilisation de voisinages *a priori*.

Le premier voisinage utilisé repose sur l'hypothèse simple selon laquelle plus une placette est géographiquement proche d'une autre et plus le lien entre celles-ci est élevé (et vice-versa). Nous avons donc utilisé l'analyse du triplet ($\mathbf{X^t W_2 Y}, \mathbf{I_{51}}, \mathbf{I_{51}}$) (notée ANA2), où la matrice des poids de voisinage $\mathbf{W_2}$ comprend toujours 94 lignes et 7 colonnes et est définie de la manière suivante, pour $i = 1, \ldots, 94$; $k = 1, \ldots, 7$ et $t = 7$:

(i) $\mathbf{W_{2}}_{ik} = (D_{ik})^{-1}/(\Sigma_{l=1,t}(D_{il})^{-1})$ si la placette i ne correspond pas au transect k. Le terme $\Sigma_{l=1,t}(D_{il})^{-1}$ est la somme des inverses des distances de la placette i aux transects l ($l = 1, \ldots, 7$).

(ii) $\mathbf{W_{2}}_{ik} = 1$ si la placette i correspond au transect k.

(iii) $\mathbf{W_{2}}_{ik} = 0$ si la placette i correspond à un transect différent de k.

où D_{ik} est la distance euclidienne calculée sur les coordonnées spatiales des placettes (qui sont égales aux numéros de ligne et de colonne de celles-ci; *cf.* figure 1).

Ces poids de voisinage dépendent uniquement de la distance mesurée sur la carte en deux dimensions. Dans (i) le numérateur assure que plus une parcelle est éloignée d'un transect et moins leurs ressemblances vont compter dans le calcul de la covariance généralisée. Le dénominateur conduit à des poids de voisinage pour une placette donnée dont la somme est égale à 1. Le calcul de WY est ainsi équivalent à assigner à la placette i (pour les observations de croissance) la moyenne des valeurs présentes sur les transects, moyenne pondérée par les inverses des distances de la placette aux transects. (ii) et (iii) reviennent à ce que l'occupation sur un transect ne dépende que de la croissance sur ce transect. Connaissant la vraie valeur de la croissance sur celui-ci, il est en effet inutile de chercher à «l'estimer» par une moyenne dépendant des distances entre lui-même et les autres transects.

Cette approche conduit à construire une série de couples de combinaisons linéaires des variables de $\mathbf{X}$ ($\mathbf{L_X}$) et de $\mathbf{Y}$ ($\mathbf{L_Y}$) qui maximisent successivement la quantité :

$$\mathrm{Cg\,(ANA2)} = \mathbf{L_X^t W_2 L_Y} = \Sigma_{i,k} \mathrm{w}_{2ik} \mathbf{L_X}_i \mathbf{L_Y}_k.$$

Le premier code lignes (figure 5C) relatif au tableau des occupations est pratiquement similaire à celui issu de l'ACPN. Il oppose globalement l'est à l'ouest de la parcelle en soulignant que les différences maximales prennent place entre le sud et le nord à l'intérieur de cette opposition. Les coordonnées correspondantes pour les variables de croissance opposent quant à elles le nord-est de la parcelle avec la totalité de l'ouest. Par rapport aux résultats des analyses précédentes, elles apparaissent beaucoup moins liées à la spécificité du transect n° 3 qui se singularise pour certaines dates par une croissance très élevée comparativement aux autres transects. La deuxième paire de coordonnées lignes de cette analyse, bien que nettement plus anecdotique d'après la courbe des valeurs propres (figure 5A), oppose le nord sans sa partie extrême est avec le sud. Comme dans toute les analyses précédentes, les dates sont extrêmement difficiles à regrouper sur la base d'autres critères (saison, année, ...) que ceux utilisés dans ces analyses.

Notons que ces coordonnées présentent des variations inter-placettes beaucoup plus faibles que celles issues des analyses précédentes (ACP et ANA1). Cette propriété souligne le lissage qui s'est opéré à travers l'utilisation du voisinage : chaque placette pour laquelle aucune mesure n'a été réalisée est renseignée par une somme pondérée des valeurs correspondant aux transects; les pondérations utilisées tiennent compte de la distance entre la placette et les transects.

Bien entendu, c'est maintenant le tableau des occupations qui est privilégié au cours de l'analyse : les coordonnées relatives à ce tableau sont en effet proches (et très proches pour la première) de celles issues de son ACPN, ce qui n'est pas le cas

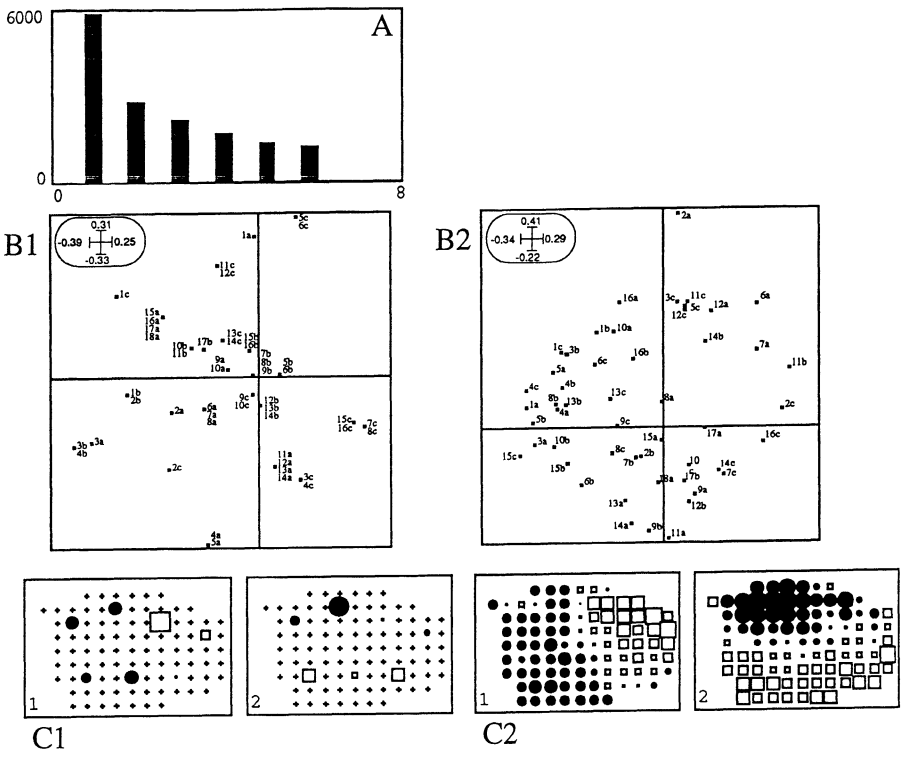

FIGURE 5

Analyse du triplet ($\mathbf{X^t W_2 Y}$, $\mathbf{I_{51}}$, $\mathbf{I_{51}}$) *(analyse ANA2). Voir le texte pour la définition de* $\mathbf{W_2}$. *A : Courbe des racines carrés des valeurs propres. B1 et B2 : Coefficients (facteurs) des combinaisons linéaires des variables pour les parties croissances (à gauche) et occupations (à droite). C1 et C2 : Coordonnées lignes pour les croissances (à gauche) et les occupations (à droite).*

pour le tableau des croissances. La forte homogénéité des zones mises en évidence par les coordonnées lignes des occupations (figure 5C2) permet de partitionner empiriquement la parcelle en deux grandes parties (figure 6, établie à partir de la consultation séparée des 2 coordonnées), ce qui paraît être en accord avec l'hypothèse biologique concernant l'existence de domaines vitaux journaliers comme modèle d'occupation privilégiée au sein d'une journée.

Ces domaines forment une partition de l'espace basée sur les exigences vitales de l'espèce- ici le pâturage. L'espace «utile» pour le pâturage apparaît ainsi restreint, dans la majorité des cas, à des zones sensiblement de même taille. Les domaines vitaux se chevauchent, mettant par là-même en évidence l'existence de la zone charnière AB qui apparaît comme un pivot dans la plupart des stratégies d'organisation journalière. Le lissage intervenu par l'intermédiaire des voisinages utilisés permet de bien mettre en évidence le rôle de cette zone et évite dans le même temps de la faire apparaître comme exclusive par rapport aux autres transects.

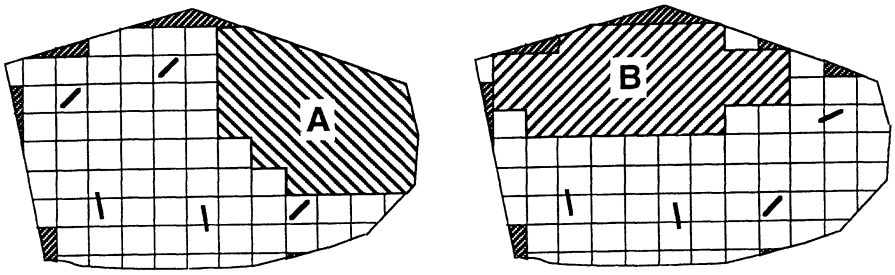

FIGURE 6

Partition en 2 espaces principaux à la base de la majorité des stratégies d'utilisation journalières. Cette partition est construite à partir des résultats de l'analyse ANA2.

Cette partition en 2 espaces principaux d'occupation liés par une zone charnière trouve au moyen de ces analyses une explication assez claire dans les conditions instantanées de production de biomasse. L'hypothèse utilisée de similarités élevées entre les parcelles les plus proches sur le plan spatial est cependant bien pauvre au regard des informations possédées à propos des caractéristiques permanentes du milieu végétal. Nous avons donc cherché à voir si ces structures étaient toujours apparentes en introduisant des poids de voisinage dépendant de ces caractéristiques. Une dernière analyse a donc été celle du triplet ($\mathbf{X^t W_3 Y}$, $\mathbf{I_{51}}$, $\mathbf{I_{51}}$), dans lequel la matrice $\mathbf{W_3}$ est définie de la manière suivante pour $i = 1, \ldots, 94$; $k = 1, \ldots 7$ et $t = 7$:

(i) $\mathbf{W_{3}}_{ik} = (V_{\mathrm{h}(ik)})^{-1} / \left(\Sigma_{l=1,t}(V_{\mathrm{h}(il)})^{-1}\right)$ si la placette i ne correspond pas au transect k.

(ii) $\mathbf{W_{3}}_{ik} = 1$ si la placette i correspond au transect k.

(iii) $\mathbf{W_{3}}_{ik} = 0$ si la placette i correspond à un transect différent de k.

où $\mathrm{V_{h(ik)}}$ est égal au variogramme multivarié (e.g. Bourgault et Marcotte (1991)) observé pour les espèces végétales ayant fait l'objet de mesures de recouvrement. Ce variogramme multivarié observé est égal à la moyenne des distances entre placettes calculées sur la base des notes de recouvrement, pour une classe de distance géographique donnée. Les classes de distance géographique sont définies à partir d'une relation de la reine (une placette est voisine des placettes immédiatement adjacentes sur les mêmes ligne, colonne ou diagonales) et des puissances booléennes successives relatives à la matrice de cette relation qui définissent des chemins comprenant 2, 3, ... voisins entre placettes. Les dernières classes (à partir d'une séparation de 10 placettes entre 2 «voisins») sont regroupées pour conserver un nombre d'individus raisonnable par rapport à ceux des autres classes.

La définition des poids de voisinage contenus dans $\mathbf{W_3}$ relève de la même idée que ceux définis dans $\mathbf{W_2}$. La figure 7D montre que le numérateur du (i) amène à affecter des poids de plus en plus faibles aux couples de placettes au fur et à mesure qu'elles s'éloignent. Contrairement à l'analyse précédente existe cependant une stabilisation de ces poids au niveau du plateau du variogramme. Le dénominateur conduit comme pour $\mathbf{W_2}$ à des poids de voisinage pour une placette donnée dont la somme est égale à 1. Le calcul de $\mathbf{WY}$ est ainsi équivalent à assigner à la placette

i (pour les observations de croissance) la moyenne des valeurs présentes sur les transects, moyenne pondérée par les inverses des variogrammes de la placette aux transects. (ii) et (iii) reviennent comme pour les poids précédents à ce que l'occupation sur un transect ne dépende que de la croissance sur ce transect.

Cette analyse conduit à construire une série de couples de combinaisons linéaires des variables de $\mathbf{X}(\mathbf{L_X})$ et de $\mathbf{Y}(\mathbf{L_Y})$ qui maximisent successivement la quantité :

$$\mathrm{Cg(ANA3)} = \mathbf{L_X^t W_3 L_Y} = \Sigma_{i,k} w_{3ik} \mathbf{L_{X}}_i \mathbf{L_{Y}}_k.$$

L'utilisation de cette pondération tend à valider les résultats de l'analyse précédente. La première paire d'axes y est quasiment similaire (figure 7) confirmant la très forte spécificité du nord-est de la parcelle qui supporte à la fois les meilleures

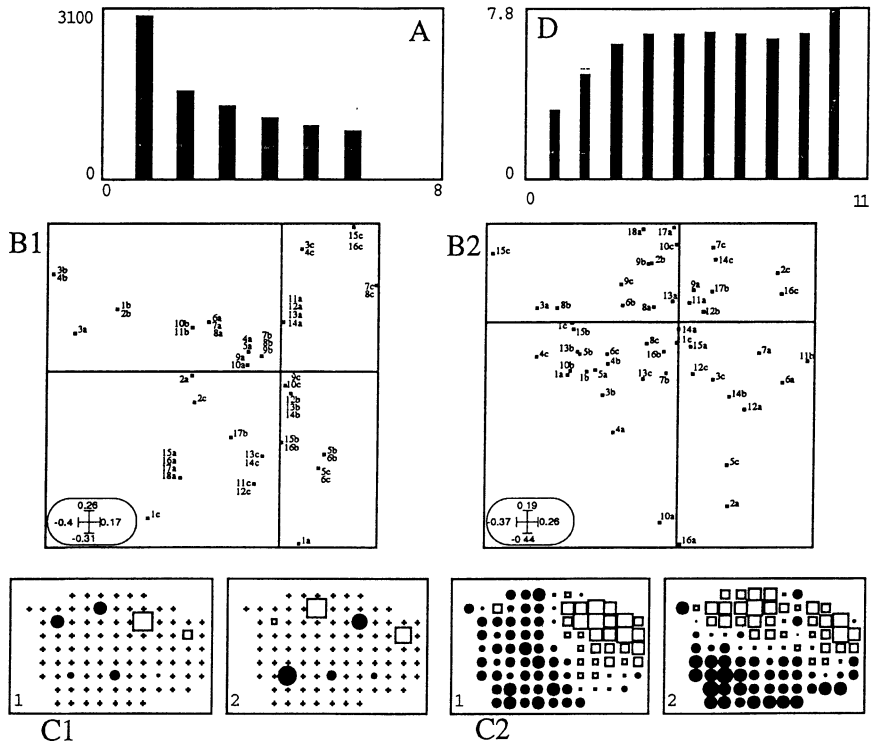

FIGURE 7

Analyse du triplet $(\mathbf{X^t W_3 Y}, \mathbf{I_{51}}, \mathbf{I_{51}})$ *(analyse ANA3). Voir le texte pour la définition de* $\mathbf{W_3}$. *A : Courbe des racines carrés des valeurs propres. B1 et B2 : Coefficients (facteurs) des combinaisons linéaires des variables pour les parties croissances (à gauche) et occupations (à droite). C1 et C2 : Coordonnées lignes pour les croissances (à gauche) et les occupations (à droite). D : Les variogrammes multivariés pour des classes de distances géographique entre voisins variant de 1 à 10 placettes de séparation.*

productions fourragères, les plus forts taux de pâturage et sa qualité de zone charnière entre les domaines vitaux. La deuxième paire d'axes est plus intéressante dans la mesure où elle souligne un nombre restreint de dates où la zone pivot correspondant au transect 3 n'est pas utilisée; pour ces dates c'est l'un ou (et) l'autre des domaines A et B qui est occupé, mais amputé des placettes adjacentes au transect 3. Les différences moyennes dans la composition floristique de zones aussi éloignées (figure 7D) explique ce phénomène, qui «colle» bien avec un certain nombre de dates. Notons enfin que quelques rares dates où c'est le sud-est de la parcelle qui est occupée sont aussi mises en évidence sur cette paire d'axes.

Ces structures sont rencontrées de manière plus ou moins stricte pour une bonne partie des dates. Pas plus que dans les analyses précédentes cependant, on ne peut mettre en évidence de regroupements sur une base calendaire. Ceci souligne l'instabilité temporelle des patterns spatiaux observés et soulève d'autres hypothèses concernant l'influence de variables non intégrées dans l'analyse telles que la taille du troupeau et la surface totale, qui peuvent fluctuer au cours du temps.

On pourra retenir de ces diverses analyses les points suivants : (i) Une utilisation généralement ciblée sur certaines zones. (ii) La zone la plus utilisée est celle entourant le transect 3, qui s'étend en général pour former les domaines vitaux A ou B. (iii) Ces domaines vitaux peuvent cependant pour certaines dates ne pas contenir les alentours du transect 3. (iv) L'explication de l'utilisation des domaines vitaux «étendus», c'est-à-dire comprenant le transect 3, est à chercher principalement dans la productivité végétale forte de celui-ci et de la zone alentour (v) L'utilisation restreinte de ces domaines, c'est-à-dire n'intégrant pas le transect 3, n'est apparente que dans la dernière analyse, c'est-à-dire lorsque est introduite une information concernant les compositions floristiques des placettes. Ce phénomène souligne des schémas d'occupation qui dépendent plus que les autres de facteurs qualitatifs (les compositions floristiques) au détriment de la stricte croissance. Ces shémas concernent cependant un nombre restreint de dates.

4. Discussion et conclusion

La généralisation d'analyses utilisant des voisinages binaires traduisant une structure de contiguïté à des pondérations quelconques, symétriques ou non des couples d'individus offre un cadre souple à l'exploration d'ensembles de données multivariables à propos desquelles on possède une information *a priori* sur les ressemblances inter-individuelles. L'illustration choisie souligne combien de nombreux champs sont demandeurs, pour des raisons souvent impératives, de cette souplesse. Notons que cette analyse aurait pu être conduite aussi sur la base d'un graphe symétrique en ajoutant au tableau $\mathbf{Y}$ autant de lignes nulles que de parcelles ne correspondant pas aux transects. Les graphes utilisés auraient alors été «carré» et symétriques; toutes les intersections de lignes et colonnes ne correspondant pas pour l'une des deux au moins à un transect auraient été simplement nulles (les autres contenant les poids utilisés ci-dessus).

Les pondérations choisies, dont la variété peut être infinie, doivent être mûrement réfléchies pour éviter les conclusions absurdes. Comme le note Cliff et Ord (Ord 1975; Cliff et Ord 1973, p. 137) à propos des indices d'autocorrélation, ces

poids traduisent un *a priori* sur le «degré d'interaction possible» entre les individus i et k et font donc référence à un modèle d'autocorrélation. Ce modèle d'autocorrélation pourra être testé dans un premier temps par l'analyse des corrélogrammes. Notons cependant que contrairement aux cas où l'on a les mêmes nombres d'individus pour les deux variables (ou tableaux de données), il n'existe pas de tests généraux pour juger de la signification de la covariance généralisée. Ceux-ci restent à développer.

Les deux voies les plus naturelles sont celles de poids binaires et poids construits sur la base des distances géographiques usuelles. Les poids binaires ont l'intérêt d'avoir de nombreuses relations avec les indices statistiques totaux (variance et covariance), cependant l'analyse ne porte que sur une classe de distance, ce qui en limite l'intérêt à la stricte contiguïté. Les distances «géographiques» permettent d'introduire une continuité forte dans les ressemblances. Notons enfin qu'une attention particulière devra être porté à l'introduction de poids optimaux destinés à estimer les valeurs en une localisation donnée sur la base des autres observations (e.g le krigeage).

Sur le plan statistique, il est évident que la grande généralité de ces résultats conduit à la cohabitation de plusieurs points de vue possibles sur ces approches. Chaque cas particulier demande alors à être, dans la mesure où cela n'a pas déjà été fait, travaillé en détail. La matrice centrale, **W**, pourra en particulier être analysée séparément pour tenter de ne travailler que sur une partie des structures qu'elle contient (§ 2.3). En tous les cas, le travail sur des modèles des données, via le calcul de sommes pondérées sur les voisins, tout en introduisant à la fois beaucoup de finesse et de généralité, conduit aussi à une dimension modélisatrice très importante qui doit être fortement réfléchie avant toute mise en œuvre.

Références bibliographiques

BOURGAULT G. et MARCOTTE D. (1991). Multivariable variogram and its application to the linear model of corregionalisation. *Mathematical Geology*, 23 : 899-928.

CARLIER A. (1985). Applications de l'analyse factorielle des évolutions et de l'analyse intra- périodes. *Statistiques et analyse des données*, 10(1) : 27-53.

CHESSEL D. et MERCIER P. (1993). Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement* (J. D. Lebreton et B. Asselain, eds) Masson, Paris. pp. 15-43.

CHESSEL D. et SABATIER R. (1994). Couplage de triplets statistiques et graphes de voisinage. In : *Biométrie et analyse des données spatio-temporelles* (B. Asselain et Coll. , eds), Société Française de Biométrie, ENSA, Rennes, 28-37.

CLIFF A.D. et ORD J.K. (1973). *Spatial autocorrelation.* Chorley et Harvey (edts), 178p.

DOLEDEC S., CHESSEL D., TER BRAAK C.J.F. ET CHAMPELY S. (1996). Matching species traits to environmental variables : a new three-table ordination method. *Environmental and Ecological Statistics*, 3, 143-146.

ESTEVE J. (1978). Les méthodes d'ordination : éléments pour une discussion. In *Biométrie et Ecologie*, Legay J.-M. et Tomassone R. (edts), Soc. Fr. de Biométrie.

GEARY R.C. (1954). The contiguity ratio and statistical mapping. *The incorporated statistician*, 5 (3) : 115-145.

JOURNEL A.G. et HUIJBREGHTS C.J. (1978). *Mining geostatistics*. Academic Press, Londres.

LEBART L. (1969). Analyse statistique de la contiguïté. *Publication de l'ISUP*, XVIII : 81- 112.

LEBART L. (1973). *Description statistique de certaines relations binaires (analyse des correspondances locales)*. Extrait d'un rapport CREDOC-CORDES : 133-177.

LE FOLL, Y. (1982), «Pondération des Distances en Analyse Factorielle», *Statistiques et Analyse des données*, 7, 13-31.

LEGENDRE P. (1993). Real data are messy. *Statistics and Computing*, 3, 197-199.

MATHERON (1965). *Les variables régionalisées et leur estimation. Une application des fonctions aléatoires aux sciences de la nature.* Masson, Paris.

MEOT A. (1992). *Explicitation de contraintes de voisinage en analyse multivariée. Applications dans le cadre de problématiques agronomiques.* Thèse de 3ème cycle, Université Lyon I, 192p.

MEOT A., CHESSEL D. et SABATIER R. (1993). Opérateurs de voisinage et analyse des données spatio-temporelles. In : *Biométrie et Environnement* (J. D. Lebreton and B. Asselain, eds) Masson, Paris. pp. 45-71.

MORAN P.A.P. (1948). The Interpretation of Statistical Maps. *Journal of The Royal Statistical Society*, 10B, 243-251.

ORD, J.K. (1975). Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70, 120-126.

ROYER J.-J. (1984). Proximity analysis : a method for multivariate geodata processing. Application to geochemical processing. *Sciences de la terre; informatique géologique*, 20(1) : 223-243. Université de Nancy.

SWITZER P. et GREEN A.A. (1984). Min/Max autocorrelation factors for multivariate spatial imagery. *Technical report N°6*, Department of Statistics, Stanford University, 11p.

THIOULOUSE J., DOLEDEC D., CHESSEL D. et OLIVIER J.M. (1995). ADE software : multivariate analysis and graphical display of environmental data. In : Proceedings of the 4th International Software Exhibition For Environmental Science and Engineering, pp. 57-62.

THIOULOUSE J., CHESSEL D. et CHAMPELY S. (1996). Multivariate analysis of spatial patterns : a unified approach to local and global structures. *Environmental and Ecological Statistics* **2**, 1-14.

UPTON G.J.G. et FINGLETON B. (1985). *Spatial analysis by example. Vol. 1 : Point pattern and quantitative data* (2nd ed.). John Wiley and Sons, Chichester & New-York

WARTENBERG, D. (1985). «Multivariate spatial correlation : a method for exploratory geographical analysis», *Geographical analysis*, 17 (4), 263-283.

Remerciements : Nous remercions les deux referees de la revue dont les critiques et suggestions ont permis d'améliorer grandement la qualité de notre texte.