

REVUE DE STATISTIQUE APPLIQUÉE

M. SIMIER

L. BLANC

F. PELLEGRIN

D. NANDRIS

**Approche simultanée de K couples de tableaux
: application à l'étude des relations pathologie
végétale - environnement**

Revue de statistique appliquée, tome 47, n° 1 (1999), p. 31-46

http://www.numdam.org/item?id=RSA_1999__47_1_31_0

© Société française de statistique, 1999, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

APPROCHE SIMULTANÉE DE K COUPLES DE TABLEAUX :

Application à l'étude des relations pathologie végétale – environnement

M. Simier (1), L. Blanc (2), F. Pellegrin (3) et D. Nandris (3)

(1) ORSTOM, Laboratoire Halieutique et Ecosystèmes Aquatiques,
B.P. 5045, 34032 Montpellier Cedex 1, France.

(2) Université Montpellier II, Laboratoire d'Hydrobiologie Marine et Continentale,
UMR 5556, Case 093, 34095 Montpellier, France.

(3) ORSTOM, Laboratoire de Phytopathologie Tropicale,
BP 5045, 34032 Montpellier Cedex 1, France

RÉSUMÉ

La question de l'approche simultanée de K couples de tableaux peut être traitée en combinant d'une part la logique des méthodes STATIS décrites par Lavit *et coll.* (1994) : rechercher un fond typologique stable à travers une série de K tableaux ayant au moins une dimension commune, et d'autre part la logique des couplages de co-inertie proposés par Dolédec et Chessel (1994) : trouver ce qui dans deux groupes de descripteurs engendre une typologie commune des objets décrits dans chacun des tableaux. La méthode STATICO est une analyse triadique partielle sur les K tableaux croisés de co-inertie calculés à chaque répétition. Sur la base de données épidémiologiques concernant les pathogènes fongiques du caféier, une application pratique de STATICO est proposée ici. Elle permet de caractériser la typologie des sites de culture en fonction de l'évolution temporelle de la liaison entre leurs caractéristiques phytosanitaires et environnementales. La méthode STATICO peut être mise en œuvre dans la programmathèque ADE-4, en libre diffusion sur Internet.

Mots-clés : analyse de données, K -tableaux, STATIS, analyse triadique partielle, co-inertie, couplage de tableaux, STATICO, pathosystème, pathogènes fongiques, caféier.

ABSTRACT

The question of simultaneous approach of K pairs of tables can be treated by combination of two methods : the logic of STATIS methods (Lavit *et coll.* 1994) which consists to find a stable typology through a set of K tables presenting at least a common dimension and the logic of cross tables of co-inertia analysis presented by Dolédec and Chessel (1994) which consists to find from two groups of descriptors what produces the common typology extracted from both tables. The STATICO method is a partial triadic analysis of K cross tables of co-inertia calculated at each repetition. We propose a practical application of STATICO by way of an example on epidemiological data about the fungal pathogens of coffee tree. It enables

to characterize the typology of coffee-growing sites in relation to temporal evolution of the link between their pathological and environmental features. The STATICO method can be performed in ADE-4 software, in freeware distribution on Internet.

Keywords : *data analysis, K-tables, STATIS, partial triadic analysis, co-inertia, cross tables, STATICO, pathosystem, fungal pathogens, coffee tree.*

Introduction

L'analyse factorielle de deux tableaux de données portant sur les mêmes objets donne lieu à deux typologies distinctes de ces objets, engendrées respectivement par chacun des groupes de descripteurs. Une manière de caractériser les relations existant entre les deux tableaux peut s'envisager selon Chessel et Mercier (1993) comme une recherche de la structure commune à ces deux typologies d'objets. C'est l'idée développée dans l'analyse de co-inertie proposée par Dolédec et Chessel (1994).

Lorsque cette double étude est renouvelée à plusieurs reprises, par exemple dans le temps ou dans l'espace, on peut chercher à caractériser la stabilité des relations observées entre les deux groupes de descripteurs. Une première voie de réponse est explorée dans les analyses de co-inertie inter-classes et intra-classes proposées par Franquet et Chessel (1994) et Franquet *et coll.* (1995). Nous utilisons ici l'extension de la méthode ACT-STATIS (Analyse Conjointe de Tableaux – Structuration des Tableaux A Trois Indices de la Statistique), décrite par Lavit *et coll.* (1994), aux opérateurs de co-inertie. Cette méthode, appelée STATICO (pour STATIS et CO-inertie), permet de traiter simultanément une série de K couples de tableaux par une analyse triadique partielle portant sur les tableaux croisés des K analyses de co-inertie. On combine ainsi la logique de STATIS (trouver ce qui dans plusieurs tableaux constitue le fond typologique commun, fond commun qui peut exister derrière de fortes particularités propres à chaque tableau) et la logique des couplages de co-inertie (trouver ce qui dans deux groupes de descripteurs engendre une typologie commune des objets décrits dans chacun des tableaux).

Une application de la méthode STATICO est présentée dans le cadre de recherches phytopathologiques menées par l'ORSTOM en Nouvelle-Calédonie en vue d'appréhender les relations entre l'état phytosanitaire de plantations de caféier et leur environnement (sol et climat). Dans cet exemple, STATICO permet de mettre en évidence une typologie originale des sites fondée sur l'évolution temporelle de la liaison entre leurs caractéristiques phytosanitaires et environnementales. Cette application est à l'origine de la méthode qui a été initiée par D. Chessel dans le cadre de la programmation ADE-4 (Chessel *et coll.* 1996) et présentée par Simier *et coll.*, 1996. Elle fait ici l'objet d'une première publication.

1. Les bases méthodologiques

1.1. L'analyse de co-inertie

L'analyse de co-inertie, proposée par Chessel et Mercier (1993) et Dolédec et Chessel (1994), est une méthode de couplage de tableaux (figure 1) qui s'appuie sur le schéma de dualité défini par Escoufier (1987) pour autoriser le couplage des

différents types d'analyses multivariées de base (ACP, AFC, ACM...). Elle recouvre ainsi l'analyse inter-batterie de Tucker (1958), l'analyse canonique sur variables qualitatives de Cazes (1980) et l'analyse des correspondances de tableaux de profils écologiques (Mercier *et coll.* 1992).

Son principe de base est la recherche d'axes de co-inertie maximisant la covariance entre les coordonnées des projections des lignes de chacun des tableaux. Si $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ et $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$ sont les deux triplets statistiques, \mathbf{D}_p et \mathbf{D}_q étant les matrices des poids associés aux colonnes de \mathbf{X} et de \mathbf{Y} respectivement et \mathbf{D}_n étant la matrice des poids associées aux lignes de \mathbf{X} et de \mathbf{Y} , l'analyse de co-inertie est celle du triplet $(\mathbf{Y}^t \mathbf{D}_n \mathbf{X}, \mathbf{D}_p, \mathbf{D}_q)$. Les axes de co-inertie sont obtenus en diagonalisant :

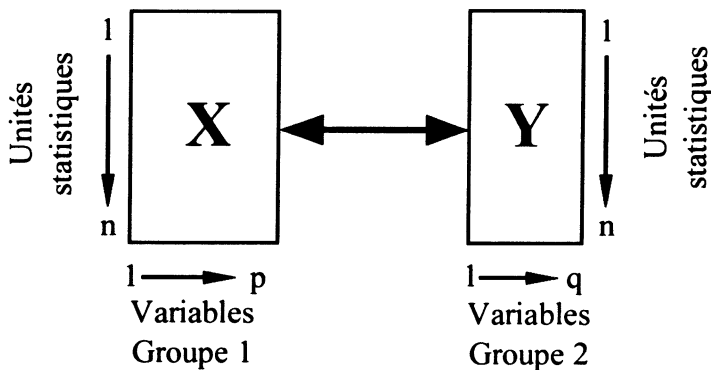
$$\mathbf{W} = \mathbf{D}_p^{1/2} \mathbf{X}^t \mathbf{D}_n \mathbf{Y} \mathbf{D}_q \mathbf{Y}^t \mathbf{D}_n \mathbf{X} \mathbf{D}_p^{1/2}$$


FIGURE 1

Application de l'analyse de co-inertie :
un couple de tableaux \mathbf{X} (n unités statistiques, p variables)
et \mathbf{Y} (n unités statistiques, q variables) appariés par les unités statistiques.

1.2. L'analyse triadique partielle

Envisagée à l'origine par Tucker (1966), l'analyse triadique partielle est également appelée PCA-SUP («*PCA of a derived two-way supermatrix*») par Kiers (1991), ou encore Pré-Statist par Leibovici (1993). Elle a été introduite en écologie par Thioulouse et Chessel (1987), et reprise plus récemment par Blanc *et coll.* (1998). Des exemples d'applications sont fournis par Dolédec (1988), Centofanti *et coll.* (1989), Aliaume *et coll.* (1993) et Blanc et Beaudou (1998). Son objet est l'analyse d'un «cube» de données, matrice à trois indices $X_{(i,j,k)}$ vue sous l'angle d'une série de K tableaux à deux indices $X_{(i,j)}$ (figure 2).

L'analyse triadique partielle (ATP) appartient à la famille des méthodes STATIS, décrites dans l'ouvrage de Lavit (1988) et faisant l'objet d'une synthèse récente dans Lavit *et coll.* (1994). Ces méthodes s'appuient sur les notions, définies par Escoufier (1973), de variance vectorielle d'un tableau $V_{av}(\mathbf{X}_k)$, de covariance

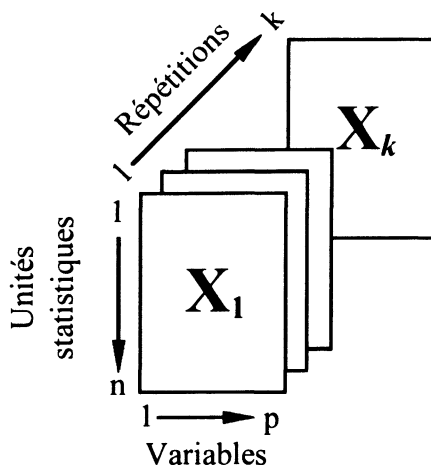


FIGURE 2

Application de l'analyse triadique partielle :
un cube de données \mathbf{X} (n unités statistiques, p variables, K répétitions),
présenté sous la forme d'une série de K tableaux (n, p) .

vectorielle entre tableaux $Covv(\mathbf{X}_k, \mathbf{X}_l)$, et de corrélation vectorielle entre tableaux $Rv(\mathbf{X}_k, \mathbf{X}_l)$. Elles sont fondées sur la recherche d'une matrice «compromis», qui exprime la structure commune aux K tableaux. Dans le cas particulier de l'ATP, les K tableaux \mathbf{X}_k partagent les mêmes dimensions lignes (n) et colonnes (p), et ils sont associés aux mêmes métriques \mathbf{D}_p et \mathbf{D}_n ; il est donc tout naturel de calculer un compromis de mêmes dimensions (n, p) qui n'est autre qu'une somme

pondérée des K tableaux : $\sum_{k=1}^K \alpha_k \mathbf{X}_k$. Les coefficients α_k sont recherchés tels que

$\left(\sum_{k=1}^K \alpha_k \mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_n \right)$ présente une inertie maximum sous la contrainte $\sum_{k=1}^K \alpha_k^2 = 1$.

L'analyse se déroule en trois temps, correspondant aux trois étapes définies par Lavit et coll., 1994 : **Interstructure**, **Compromis** et **Trajectoires**.

Première phase : Recherche de la moyenne (Interstructure).

La matrice des produits scalaires entre tableaux, ou covariances vectorielles $Covv(X_k, X_l)$, de terme général $Trace(X_k^t \mathbf{D}_n X_l \mathbf{D}_p)$, est calculée puis diagonalisée. On retient seulement le premier vecteur propre en raison de sa fonction particulière de définition d'un élément commun aux K tableaux : ses composantes normées fournissent en effet les coefficients α_k qui pondèrent la participation des tableaux au

compromis $\sum_{k=1}^K \alpha_k \mathbf{X}_k$. Contrairement à l'analyse STATIS où les α_k , calculés à partir d'une série d'opérateurs, sont toujours de même signe, en analyse triadique partielle

on peut obtenir dans certains cas particuliers des coefficients α_k de signes différents. Cela rend délicate l'interprétation du compromis et doit conduire l'utilisateur à envisager une autre approche méthodologique, comme une analyse STATIS sur les opérateurs.

Deuxième phase : Analyse de la moyenne (Compromis).

Dans une seconde étape, l'analyse du tableau compromis $\sum_{k=1}^K \alpha_k \mathbf{X}_k$ fournit des axes principaux sur lesquels se projettent ses lignes, et des composantes principales sur lesquelles se projettent ses colonnes. À ce niveau s'exprime clairement le fond typologique commun aux K tableaux.

Troisième phase : Reproductibilité de la moyenne (Trajectoires).

Chacune des n lignes de X_k est un point dans l'espace de ses p colonnes \mathfrak{R}_p et peut être projetée comme individu supplémentaire sur les axes principaux du tableau compromis. Il en va de même, dans l'espace des lignes \mathfrak{R}_n , pour les colonnes des X_k que l'on peut projeter en variables supplémentaires sur les composantes principales du compromis. Selon le même principe, les axes et composantes principales obtenus en réalisant séparément les K analyses se projettent dans les mêmes espaces, respectivement \mathfrak{R}_p et \mathfrak{R}_n et par conséquent sur l'analyse du compromis. Ces projections en éléments supplémentaires permettent de juger de la variabilité de la structure moyenne décrite par le compromis à travers les K tableaux d'origine.

2. Généralisation de l'analyse triadique partielle aux tableaux croisés de co-inertie : la méthode STATICO

Dans la question qui nous intéresse ici, l'étude k est formée d'un tableau X_k , donnant les valeurs prises par p variables du groupe 1 pour n_k unités statistiques, et d'un tableau Y_k donnant les valeurs de q variables du groupe 2 pour les mêmes unités statistiques. Cette double étude est renouvelée à K reprises. D'une répétition à l'autre, la liste des variables est constante tandis que la liste des unités statistiques peut varier (figure 3A). On posera : $X^t = (X_1^t, \dots, X_K^t)$; $Y^t = (Y_1^t, \dots, Y_K^t)$. Dans le cas d'un échantillonnage complet de toutes les unités statistiques à toutes les répétitions, on obtient un couple de cubes de données (figure 3B).

Chacun de ces tableaux s'insère dans un schéma de dualité (Escoufier 1987), respectivement $(\mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_{n_k})$ et $(\mathbf{Y}_k, \mathbf{D}_q, \mathbf{D}_{n_k})$, d'un type quelconque (figure 4). Les métriques diagonales \mathbf{D}_p et \mathbf{D}_q sont fixées et indépendantes de la répétition. La métrique des poids des unités \mathbf{D}_{n_k} est commune aux deux schémas, ce qui autorise l'analyse de co-inertie des deux tableaux. Les k schémas de co-inertie partagent la même dimension des tableaux et les mêmes métriques. À la répétition k , le tableau X_k inséré dans le schéma $(\mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_{n_k})$ et le tableau Y_k inséré dans le schéma $(\mathbf{Y}_k, \mathbf{D}_q, \mathbf{D}_{n_k})$ donnent le schéma de co-inertie $(\mathbf{Y}_k^t \mathbf{D}_{n_k} \mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_q)$.

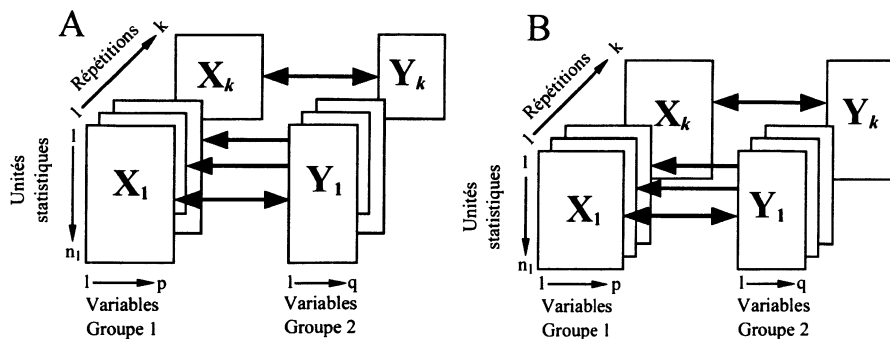


FIGURE 3

Situation expérimentale des K couples de tableaux. Deux séries de K tableaux de données X_k (n_k unités statistiques, p variables du groupe 1, k répétitions) et Y_k (n_k unités statistiques, q variables du groupe 2, k répétitions).

À chaque répétition k , le tableau X_k (n_k, p) et le tableau Y_k (n_k, q) s'associent par leur dimension commune n_k .

A : la liste des unités statistiques varie d'une répétition à l'autre;

B : la liste des unités statistiques est la même pour toutes les répétitions, on obtient un «double-cube» de données.

Notons $Z_k = Y_k^t D_{n_k} X_k$. Selon la définition de l'analyse triadique partielle donnée ci-dessus, trouver un compromis de co-inertie revient à trouver des nombres α_k tels que $\left(\sum_{k=1}^K \alpha_k Z_k, D_p, D_q \right)$ présente une inertie maximum sous la contrainte $\sum_{k=1}^K \alpha_k^2 = 1$.

En pratique, le compromis de l'analyse triadique partielle est ici une analyse de co-inertie fictive dont le tableau croisé, de la forme $Z = Y_c^t D X_c$, est une moyenne pondérée de K tableaux croisés. Ses colonnes sont les colonnes de X et ses lignes sont les colonnes de Y . Son analyse fournit des valeurs propres, une ordination de ses colonnes (donc des p variables de X) et une ordination de ses lignes (donc des q variables de Y).

Comme dans toute analyse triadique partielle, on peut projeter, selon le principe des points supplémentaires, les éléments des K analyses séparées de co-inertie sur les axes et les composantes du compromis : les colonnes, les lignes, ainsi que les axes et composantes principales de Z_k ($1 \leq k \leq K$).

Du fait de la co-inertie, une dimension nouvelle intervient ici : celle des unités statistiques. On peut en effet projeter en lignes supplémentaires sur les axes d'inertie du compromis l'ensemble des lignes de X (unités statistiques) et sur les composantes du compromis l'ensemble des lignes de Y (unités statistiques également). En normalisant globalement leurs coordonnées pour retenir dans la co-inertie la partie «corrélation», on peut apparier les projections des unités statistiques issues respectivement des espaces des lignes des K -tableaux X et Y . La représentation de ce double nuage de points multifenêtrée par tableau permet de juger de l'évolution sur l'ensemble des tableaux de la réalisation de la co-structure commune. Il est

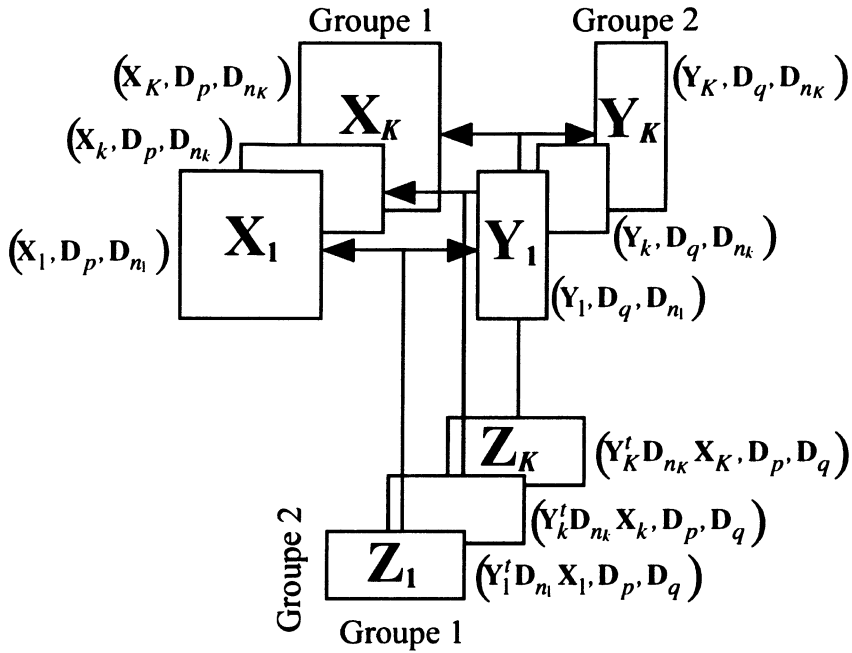


FIGURE 4

Insertion de la problématique des K couples de tableaux dans le schéma de dualité. L'analyse des K-tableaux croisés par l'analyse triadique partielle s'articule avec les analyses des K-tableaux X et Y par STATIS.

également possible de normer par tableau les coordonnées des lignes supplémentaires pour exprimer la corrélation instantanée; ce n'est pas l'option retenue ici.

3. Un exemple d'application : pathologie végétale & environnement

3.1 Les données

Les recherches phytopathologiques menées depuis 1992 par l'ORSTOM en Nouvelle-Calédonie ont pour but d'analyser le fonctionnement d'un pathosystème multiple associant le caféier (*Coffea arabica*), ses principaux pathogènes fongiques (*Hemileia vastatrix*, *Colletotrichum gloeosporioides*, *Cercospora coffeicola*) et l'environnement.

Seules les données concernant le cycle cultural de l'année 1994 sont retenues pour la présente étude. Le dispositif expérimental est constitué de 15 parcelles traditionnelles de caféier choisies dans des contextes écologiques contrastés. À intervalles réguliers (8 observations par an), une enquête épidémiologique est réalisée dans ces sites afin de procéder au relevé détaillé (feuille après feuille) de l'état sanitaire des caféiers. En complément, la caractérisation de l'environnement édaphique de

chaque site est effectuée dès le début du suivi, alors que les paramètres climatologiques sont enregistrés en continu tout au long de l'année. Ainsi, au cours du cycle cultural annuel, on obtient un total de 120 (8*15) relevés de la pathologie et de l'environnement.

Les variables pathologiques sont au nombre de six. Trois sont des notes sanitaires qui expriment, pour chacune des trois maladies, la sévérité des infections des feuilles de caféier :

- l'antracnose, maladie fréquente mais peu dommageable (**ANTHR**),
- la cercosporiose, maladie rare mais sévère (**CERCO**),
- la rouille, la plus répandue et la plus dommageable des trois (**ROUIL**).

Les trois autres variables sont différents indicateurs de l'état du feuillage :

- le pourcentage cumulé de feuilles devenues malades (tous pathogènes confondus) durant l'enquête (**%CFneoMAL**),
- le pourcentage cumulé de feuilles malades tombées (**%CFTMal**),
- le pourcentage cumulé de feuilles saines tombées (**%CFTSain**).

Les variables de milieu sont au nombre de huit. Deux d'entre elles, liées aux conditions climatiques, résultent d'une discrimination statistique préalable :

- pluviométrie cumulée (**pluv**),
- température minimale (**tmin**).

Les six autres variables caractérisent le contexte édaphique des sites. Elles sont déterminées au début de l'étude et sont considérées comme stables dans le temps :

- fertilité du sol (**fert**),
- structure du sol (**stru**),
- capacité de rétention en eau du sol (**capa**),
- pH moyen du sol (**pH**),
- altitude (**alti**),
- degré d'ombrage (**ombr**).

3.2 L'approche choisie

Dans cette situation expérimentale, l'approche la plus accessible consiste à moyenner sur une année les valeurs des variables climatiques et épidémiologiques et à coupler les paramètres de milieu avec les moyennes annuelles des paramètres pathologiques pour faire un bilan en terme de co-structure entre moyennes. C'est la voie décrite dans Franquet et Chessel (1994) et explorée initialement par Lamouroux *et coll.* (1995).

Cette démarche présente l'inconvénient de ne pas prendre en compte une des caractéristiques majeures d'un développement épidémique : son évolution temporelle. On perd de ce fait une part de l'information sur les liens entre les dynamiques d'évolution des maladies et les causes environnementales qu'on leur suppose. Aussi, le but de cette nouvelle approche est-il d'explorer ce que l'on obtient en postulant que

le couplage entre ces deux ensembles de variables se fait à chaque date et de définir une co-structure moyenne (Pellegrin *et coll.* 1995) plutôt qu'une co-structure entre moyennes.

Le tableau des relevés de variables pathologiques, de dimension 120-6 est organisé sous la forme de huit tableaux X_k de dimension (15,6) indicés par le temps. Les variables physico-chimiques caractéristiques des sites et stables dans le temps sont dupliquées, de manière à construire à chaque date un tableau complet de variables environnementales Y_k de dimension (15,8) regroupant physico-chimie et climatologie. Les variables sont centrées par tableau, puis normalisées globalement (Bouroche 1975), transformation qui permet de prendre en compte la variance intra-tableau. À une date k donnée, le tableau de pathologie X_k et le tableau de milieu Y_k portent sur les mêmes n_k sites. La recherche de la typologie des sites commune aux deux ensembles de descripteurs passe par l'analyse du tableau croisé de co-inertie $Z_k = Y_k^t D_{n_k} X_k$ comportant 8 lignes (les variables environnementales) et 6 colonnes (les variables pathologiques). Nous nous trouvons donc en présence d'une série de 8 tableaux croisés Z_k sur lesquels nous effectuons une analyse triadique partielle.

3.3 Les résultats

3.3.1 Recherche de la moyenne

L'analyse triadique partielle des 8 tableaux croisés passe tout d'abord par la diagonalisation de la matrice des covariances vectorielles entre tableaux. De cette étape, nous ne retenons que les composantes du premier vecteur propre normé qui prend en compte 86% de l'inertie (figure 5A). Elles fournissent les coefficients de pondération α_k affectés à chacun des tableaux dans le calcul du compromis.

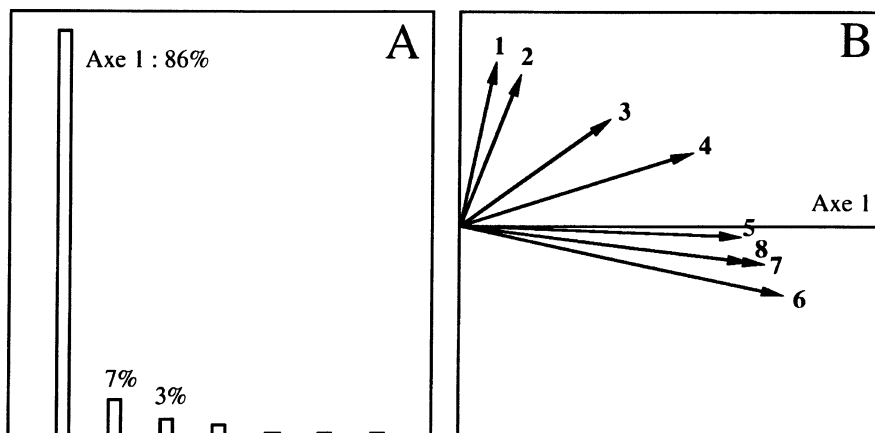


FIGURE 5

Analyse triadique partielle des co-inerties pathologie-milieu : interstructure.

A : histogramme des valeurs propres ; B : projection des 8 tableaux sur les axes 1 et 2.

Cette première étape ou interstructure permet d'observer la dynamique d'évolution temporelle de la co-structure pathologie-milieu (figure 5B) qui est ici étroitement liée au développement des maladies durant le cycle cultural (tableau 1). En effet, la variance vectorielle des deux premiers tableaux est faible : les maladies sont encore peu développées et ces deux dates participent très peu à la définition du compromis. Le troisième tableau amorce la mise en place d'une co-structure stable qui se reproduit ensuite du cinquième au huitième tableau. Ainsi le compromis que nous analyserons ci-dessous reflète majoritairement la co-structure observée lors des 4 ou 5 derniers relevés.

TABLEAU 1

Analyse triadique partielle des co-inerties pathologie-milieu : indices typologiques (en première colonne, le numéro du tableau; en deuxième colonne le coefficient α_k qui pondère la participation du tableau k à la définition du compromis; en troisième colonne la variance vectorielle du tableau qui mesure son inertie interne).

Tableau	Coefficient α_k	Variance vectorielle
1	0.0559	0.000567
2	0.0922	0.000491
3	0.2276	0.000942
4	0.3489	0.001716
5	0.4233	0.002288
6	0.4856	0.003198
7	0.4585	0.002703
8	0.4354	0.002562

3.3.2 Analyse de la moyenne

L'analyse du tableau compromis définit des valeurs propres (figure 6A). Nous choisissons d'expliciter les deux premières, qui prennent en compte respectivement 70 et 22% de l'information. L'ordination des colonnes et des lignes sur les axes du tableau compromis (respectivement variables de pathologie et variables de milieu moyennées sur toutes les dates) donne une représentation synthétique (figures 6B et 6C) de la co-structure pathologie-milieu qui constitue le fond commun des 8 tableaux de co-inertie.

Sur le premier axe, l'intensité croissante de l'attaque de la rouille s'exprime en valeurs positives, à l'opposé des vecteurs représentant les deux autres maladies. L'indice de défoliation pathologique (%CFTMal) apparaît fortement lié au niveau d'importance de la rouille. Le second axe sépare anthracnose et cercosporiose, mais il est avant tout lié à l'indice de défoliation physiologique (%CFTSain) qui reflète un éventuel stress des arbres. Nous considérerons donc qu'en terme de concordance pathologie-milieu, l'essentiel de l'information se retrouve sur le premier axe.

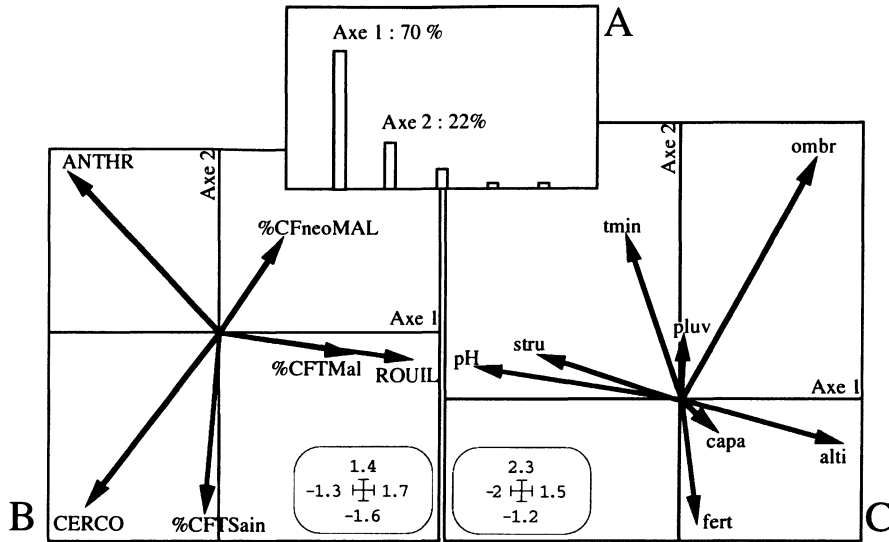


FIGURE 6

Analyse triadique partielle des co-inerties pathologie-milieu : étude du compromis.

A : Histogramme des valeurs propres; B : Projection des colonnes du tableau croisé compromis (variables de pathologie) sur les axes 1 et 2; C : Projection des lignes du tableau croisé compromis (variables de milieu) sur les axes 1 et 2.

Les variables environnementales s'expriment dans cette analyse comme facteurs favorisant ou limitant les maladies. À titre d'exemple, on retiendra que la rouille s'exprime préférentiellement dans des sites caractérisés par des sols mal structurés, un pH faible, des températures minimales relativement basses, un taux d'ombrage important et une altitude élevée.

3.3.3 Reproductibilité de la moyenne

3.3.3.1 Variables pathologiques et environnementales

Les lignes et colonnes de chacune des analyses de co-inertie séparées (respectivement variables environnementales et variables pathologiques mesurées à chaque visite) peuvent se projeter sur le référentiel commun formé par les axes du compromis (figure 7) et permettent de préciser la typologie moyenne.

Du point de vue des variables pathologiques (figure 7A), cette représentation met en évidence les dynamiques largement indépendantes des trois maladies (rouille, anthracnose et cercosporiose).

Du point de vue des variables environnementales (figure 7B), dont six sur huit sont stables dans le temps, on pourrait s'attendre à ce que leurs coordonnées dans le

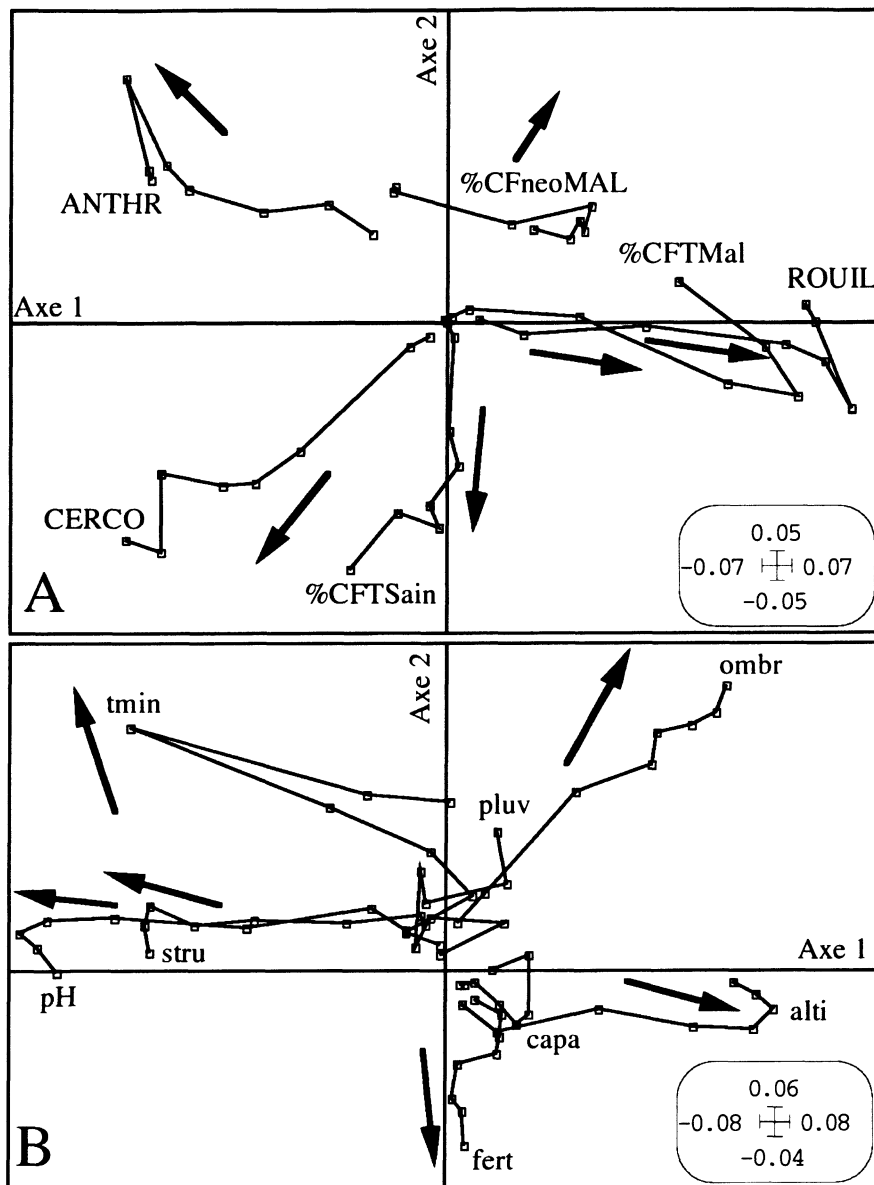


FIGURE 7

*Analyse triadique partielle des co-inerties pathologie-milieu :
étude de la reproductibilité de la moyenne.*

*A : trajectoires des variables de pathologie (colonnes des tableaux croisés)
sur le plan 1-2 du compromis;*

*B : trajectoires des variables de milieu (lignes des tableaux croisés)
sur le plan 1-2 du compromis.*

compromis le soient également. Ce serait effectivement le cas si l'on se limitait à la seule analyse du tableau Y des relevés de milieu. Mais les covariances entre variables de milieu et variables pathologiques, qui s'expriment dans les tableaux croisés de co-inertie, augmentent avec le développement des maladies et induisent une nouvelle image de la croissance au cours du temps de la liaison entre ces deux groupes de variables.

3.3.3.2 Sites

À chaque date, la co-inertie pathologie-environnement s'exprime par une double typologie des sites, issue d'une part des descripteurs pathologiques et d'autre part des descripteurs environnementaux. Après normalisation globale des coordonnées des sites, ces deux typologies peuvent être mises en relation graphiquement sur le référentiel commun que constituent une fois encore les axes du compromis de co-inertie (figure 8). Ceci donne une image particulièrement explicite de la mise en place progressive de la co-structure entre la variabilité «pathologique» qui augmente au cours du temps et la variabilité «mésologique» qui reste stable. Aux dates 1 et 2, la variabilité «pathologique» est très faible et il n'y a aucune correspondance. La co-structure s'installe en 3 et 4. Elle est bien en place de 5 à 7 et se dégrade légèrement en 8. Pour ces dates, la représentation suggère une partition des sites en trois groupes (identifiés ici par des cercles) définis conjointement par la nature et la gravité de l'attaque des pathogènes et par leurs caractéristiques environnementales.

Conclusion

Dans le cadre des investigations épidémiologiques ayant servi de support à cette étude, l'usage de la méthode STATICO en terme de moyenne de co-structures a révélé l'existence de régulations originales exercées par l'environnement sur la dynamique des maladies du caféier. Les tendances ainsi identifiées ont été mises à profit pour modéliser ces maladies puis élaborer un outil de prédiction de leur développement (Nandris *et coll.* 1997).

Sur un plan plus général, la flexibilité de STATICO permet de s'adapter à une grande variété de problématiques. Face à toute situation expérimentale pouvant se décrire sous la forme d'une série de K couples de tableaux, STATICO permet de proposer une vision synthétique de la co-structure moyenne sous la forme d'un compromis de co-inertie et d'étudier la reproductibilité de cette co-structure dans chacun des K couples de tableaux. Enfin, les nombreuses aides graphiques à l'interprétation, tirant parti entre autres du multifenêtrage par tableau, sont autant d'outils d'exploration puis d'expression synthétique des résultats.

La méthode STATICO est mise en œuvre dans le cadre de la programmatisation ADE-4 (Thioulose *et coll.* 1997), fonctionnant sous Macintosh OS et PC Windows 95 ou NT. L'ensemble des programmes, leur interface utilisateur et leur documentation sont librement accessibles sur Internet (<http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html>).

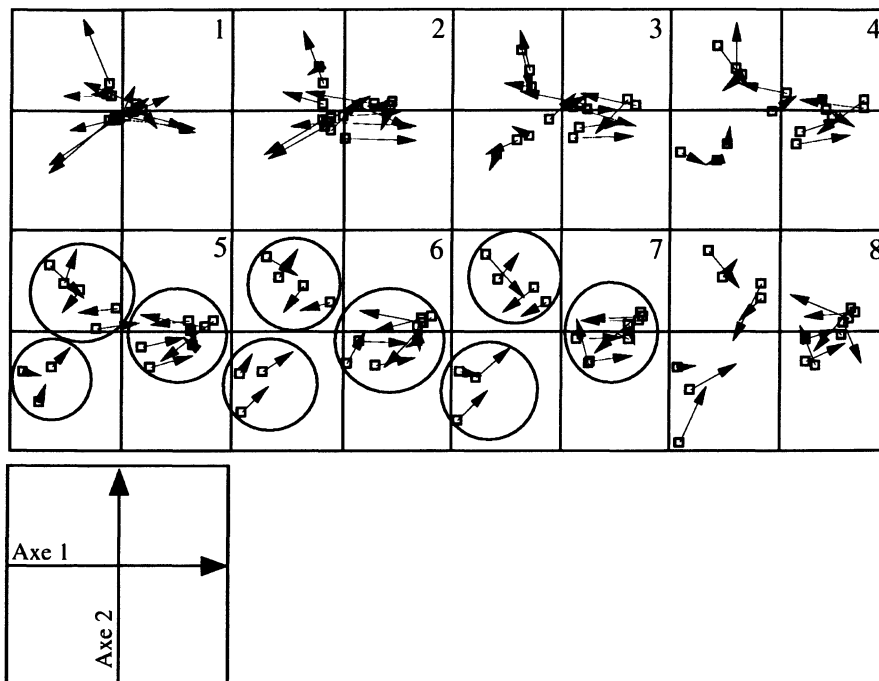


FIGURE 8

*Analyse triadique partielle des co-inerties pathologie-milieu :
double représentation des nuages de co-inertie associés aux relevés,
sur les axes 1 et 2 du compromis. A chaque date (fenêtres 1 à 8),
les 15 couples de points-sites sont projetés simultanément.*

*Pour un relevé donné, la position du point de vue pathologique (marquée par un carré)
et la position du point de vue environnemental (marquée par l'extrémité de la flèche)
sont d'autant plus proches que la co-variation pathologie-milieu est forte.*

Remerciements

Nous tenons à remercier Daniel Chessel pour son aide précieuse dans la conception et la mise en œuvre de cette méthode.

Références bibliographiques

ALIAUME C., MONTEIRO C., LOUIS M., LAM HOAI T. et LASSERRE G. (1993). Organisation spatio-temporelle des peuplements ichtyologiques dans deux lagunes côtières : au Portugal et en Guadeloupe. *Oceanologica acta* : 16, 291-301.

- BLANCL., CHESSEL D. et DOLEDEC S. (1998). Stabilité temporelle des structures spatiales. I – Analyses d’une série de tableaux de relevés faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture* : 348, 1-21.
- BLANC L. et BEAUDOU D. (1998). Stabilité temporelle des structures spatiales. II – Analyse des peuplements piscicoles des régions Languedoc-Roussillon et Provence-Alpes-Côte d’azur. *Bulletin Français de la Pêche et de la Pisciculture* : 348, 23-45.
- BOUROCHE J.M. (1975). Analyse des données ternaires : la double analyse en composantes principales. Thèse de 3^e cycle, Université de Paris VI. 1–57 + annexes.
- CAZES P. (1980). L’analyse de certains tableaux rectangulaires décomposé en blocs : généralisation des propriétés rencontrées dans l’étude des correspondances multiples. I. Définitions et applications à l’analyse canonique des variables qualitatives. II Questionnaires : variantes des codages et nouveaux calculs de contributions. *Les Cahiers de l’analyse des Données* : 5, 145–161 et 387–406.
- CENTOFANTI M., CHESSEL D. et DOLEDEC S. (1989). Stabilité d’une structure spatiale et compromis d’une analyse statistique multi-tableaux : Application à la physico-chimie d’un lac réservoir. *Revue des Sciences de l’Eau* : 2, 1, 71–93.
- CHESSEL D. et MERCIER P. (1993). Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement*. Lebreton J.D. et Asselain B. (Eds.) Masson, Paris. 15–44.
- CHESSEL D., SIMIER M. et HANAFI M. (1996). STATICO : STATIS et Co-inertie. *Documentation thématique ADE-4, Vol. 5, section 8*, Université Claude-Bernard, Lyon 1. 1-41 (<ftp://pbil.univ-lyon1.fr/pub/mac/ADE/ADE4/DocThemPDF/Thema58.pdf>).
- DOLEDEC S. (1988). Les analyses multi-tableaux en écologie factorielle. II Stratification longitudinale de l’Ardèche à partir des descripteurs physico-chimiques. *Acta Œcologica, Œcologia Generalis* : 9, 119–135.
- DOLEDEC S. et CHESSEL D. (1994). Co-inertia analysis : an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.
- ESCOUFIER Y. (1973). Le traitement des variables vectorielles. *Biometrics* : 29, 750–760.
- ESCOUFIER Y. (1987). The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre P. et Legendre L. (Eds.) NATO advanced Institute, Serie G, Springer Verlag, Berlin. 139–156.
- FRANQUET E. et CHESSEL D. (1994). Approche statistique des composantes spatiales et temporelles de la relation faune-milieu. *C.R. Acad. Sci. Paris, Sciences de la vie* : 317, 202–206.
- FRANQUET E., DOLEDEC S. et CHESSEL D. (1995). Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiologia* : 300–301, 425–431.
- KIERS H.A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika* : 56, 449–470.

- LAMOUREUX N., PELLEGRIN F., NANDRIS D. et KOHLER F. (1995). The *Coffea arabica* fungal pathosystem in New Caledonia : interactions at two different spatial scales. *Journal of Phytopathology* : 143, 403–413.
- LAVIT CH. (1988). *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1–240.
- LAVIT CH., ESCOUFIER Y., SABATIER R. et TRAISSAC P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis* : 18, 97–119.
- LEIBOVICI D. (1993). Facteurs à mesures répétées et analyses factorielles : application à un suivi épidémiologique. Thèse de doctorat, Université Montpellier II. 1–282.
- MERCIER P., CHESSEL D. et DOLEDEC S. (1992). Complete correspondence analysis of an ecological profile data table : a central ordination method. *Acta Œcologica* : 13, 25–44.
- NANDRIS D., KOHLER F., MONIMEAU L. et PELLEGRIN F. (1997). Lutte intégrée contre les ravageurs (IPM) et approche intégrée du pathosystème *Coffea arabica*. Proceedings du 17th Colloque international de l'ASIC sur le Café. Nairobi, Kenya, 20-25 juillet 1997 : Vol II, 83–95.
- PELLEGRIN F., NANDRIS D., WESTRELIN S. et KOHLER F. (1995). Situation pathologique des arabica en Nouvelle-Calédonie, corrélations entre pathogénèse et environnement. Proceedings du 16ème Colloque international de l'ASIC sur le Café, Kyoto, Japon, 9-14 avril 1995 : Vol. II, 690–698.
- SIMIER M., HANAFIM. et CHESSEL D. (1996). Approche simultanée de K couples de tableaux. Recueil des résumés des communications des XXVIII^{èmes} Journées de statistique, Université Laval, Québec, Canada, 27-30 Mai 1996. 673–676.
- THIOULOUSE J. et CHESSEL D. (1987). Les analyses multi-tableaux en écologie factorielle. I. – De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Œcologica, Œcologia Generalis* : 8, 4, 463–480.
- THIOULOUSE J., CHESSEL D., DOLEDEC S. et OLIVIER J.M. (1997). ADE-4 : a multivariate analysis and graphical display software. *Statistics and Computing* : 7, 75–83.
- TUCKER L.R. (1958). An inter-battery method of factor analysis. *Psychometrika* : 23, 111–136.
- TUCKER L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* : 31, 279–311.