REVUE DE STATISTIQUE APPLIQUÉE

D. CHESSEL
J. D. LEBRETON
N. YOCCOZ

# Propriétés de l'analyse canonique des correspondances; une illustration en hydrobiologie

*Revue de statistique appliquée*, tome 35, n° 4 (1987), p. 55-71
<http://www.numdam.org/item?id=RSA_1987__35_4_55_0>

*Rev. Statistique Appliquée*, 1987, **XXXV** (4), 55-72

# PROPRIÉTÉS DE L'ANALYSE CANONIQUE DES CORRESPONDANCES; UNE ILLUSTRATION EN HYDROBIOLOGIE

D. CHESSEL

*Ecologie des eaux douces, UA 367, Université Lyon 1, 69622 Villeurbanne Cedex.*

J.D. LEBRETON

*CEPE/CNRS, BP 5051, 34033 Montpellier Cedex.*

N. YOCCOZ

*Laboratoire de Biométrie, UA 243, Université Lyon 1, 69622 Villeurbanne Cedex.*

RÉSUMÉ

L'analyse canonique des correspondances (ACC) est définie par TER BRAAK (1986) comme la recherche d'une combinaison linéaire de variables quantitatives (mesures de milieu) maximisant la variance des moyennes conditionnelles par colonnes (espèces) d'un tableau floro-faunistique. Nous précisons ici la définition de cette méthode d'analyse multivariée, en termes de schéma de dualité dans l'optique de TENENHAUS & YOUNG (1985), en étudions les propriétés optimales, et fournissons un algorithme plus efficace que celui de TER BRAAK. Une illustration porte sur la typologie des peuplements de Poissons du Doubs à l'aide de variables de milieu (données de VERNEAUX 1973).

***Mots clés :*** *Analyse des correspondances, Analyse canonique des correspondances, Contraintes linéaires, Analyse en composantes principales, Analyse canonique.*

ABSTRACT

Canonical Correspondence Analysis is defined by TER BRAAK (1986) as searching a linear combination of quantitative environmental variables which maximizes the variance of conditional means of columns (species) of a species-sample matrix. We precise here the definition of this multivariate method with reference to duality diagrams as viewed by TENENHAUS & YOUNG (1985), give a review of its optimal properties, and provide an algorithm more efficient than TER BRAAK's one. An example is given about the typology of fish species of the river Doubs in relation with environmental variables (data from VERNEAUX 1973).

***Key words :*** *Correspondence Analysis, Canonical Correspondence Analysis, Linear Constraints, Principal Component Analysis, Canonical Analysis.*

## Introduction

Dans un article récent TER BRAAK (1986) considère une méthode d'analyse des relations espèces-milieu qui traite un ensemble de données comportant n relevés écologiques. Chacun d'eux porte une mesure de l'abondance de t

espèces et une mesure de la valeur de p variables environnementales (quantitatives). L'algorithme proposé compte 7 pas :

1) Coder numériquement les échantillons (relevés) avec un code arbitraire non constant.
2) Calculer la moyenne conditionnelle par espèces.
3) Calculer la moyenne conditionnelle par relevé de ce code espèce (« reciprocal averaging »)
4) Faire la régression multiple pondérée (poids marginaux du tableau relevés-espèces) de cette nouvelle variable sur les p variables environnementales.
5) Calculer les valeurs prédites par cette régression.
6) Centrer et réduire les valeurs obtenues.
7) Réutiliser ce code en 1) et arrêter la boucle lorsque deux itérations successives donnent un résultat assez voisin.

TER BRAAK (1987) donne ensuite une propriété d'optimalité du code relevé ainsi obtenu en précisant qu'il s'agit d'une combinaison linéaire de variables de milieu maximisant la variance des moyennes conditionnelles par espèces. Nous désirons, dans cette note, replacer cette analyse appelée par TER BRAAK *analyse canonique des correspondances* (ACC) dans le modèle général du schéma de dualité afin d'en faciliter la programmation et l'interprétation, d'en présenter une autre propriété optimale et de préparer l'extension au cas des variables qualitatives.

## Une procédure

Considérons un couple de tableaux. Le premier est un tableau T relevant de l'analyse des correspondances (AFC). Il a n lignes et t colonnes. Son terme général positif ou nul est noté $t_{ij}$ ($1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant t$). Le second est un tableau X relevant de l'analyse en composantes principales (ACP sur variables quantitatives). Il a n lignes et p colonnes. Son terme général est $x_{ik}$ ($1 \leqslant i \leqslant n$, $1 \leqslant k \leqslant p$). On suppose de plus que n est strictement supérieur à p, ce qui est une contrainte impérative, et que t est supérieur à p, le cas contraire étant évoqué plus loin. La procédure de l'analyse étudiée comporte cinq étapes.

1) Comme en AFC on calcule les marges du tableau T, soit :

$$s = \sum_{i,j} t_{ij} \qquad t_{i.} = \sum_{j} t_{ij} \qquad t_{.j} = \sum_{i} t_{ij}$$

puis les fréquences conjointes et marginales, soit :

$$p_{ij} = t_{ij}/s \qquad p_{i.} = t_{i.}/s \qquad p_{.j} = t_{.j}/s$$

On notera $D_n$ la matrice diagonale des $P_{i.}$, $D_t$ la matrice diagonale des $p_{.j}$, $\mathbf{1}_n$ le vecteur à n composantes égales à 1, $\mathbf{1}_t$ le vecteur à t composantes égales à 1 et P le tableau de terme général $p_{ij}$.

2) Comme en ACP pondérée on calcule moyennes et variances du tableau X pour la pondération $D_n$, soit :

$$m_j = \sum_{i} p_{i.} x_{ij} \qquad s_j^2 = \sum_{i} p_{i.}(x_{ij} - m_j)^2$$

puis le tableau $X_0$ de terme général :

$x_{ij} - m_j$ (option centrage)

ou $(x_{ij} - m_j)/s_j$ (option normalisation)

et la matrice des covariances, soit :

$$C = X_0' D_n X_0$$

3) La matrice C, semi-définie positive, est diagonalisée. On conserve l'intégralité de ses valeurs propres non nulles qu'on suppose au nombre de f. Soient $\lambda_1, \lambda_2, \ldots, \lambda_f$ ces valeurs propres positives rangées par ordre décroissant, $\Lambda$ la matrice diagonale correspondante et U la matrice à p lignes et f colonnes contenant en colonnes les vecteurs propres orthonormés (métrique canonique) $u_1$, $u_2, \ldots, u_f$ associés, ce qui correspond à :

$$C U = U \Lambda \quad \text{et} \quad U' U = I_f$$

$U \Lambda^{-1/2}$ peut le cas échéant être remplacé à partir d'une autre décomposition de C, par exemple par une inverse généralisée de la transformation de CHOLESKI de C. Tout ce qui suit reste alors valable *mutatis mutandis*. $U \Lambda^{-1} U'$ est alors en particulier remplacée par $C^-$ inverse généralisée de C.

4) Le tableau $T_0$ défini par :

$$T_0 = D_t^{-1} P' X_0$$

comporte t lignes et p colonnes et est formé des moyennes conditionnelles des variables du tableau $X_0$ pour les distributions conditionnelles colonnes du tableau T. On diagonalise alors la matrice

$$A = \Lambda^{-1/2} U' T_0' D_t T_0 U \Lambda^{-1/2}$$

dont les valeurs propres sont $\delta_1, \delta_2, \ldots, \delta_f$ et les vecteurs propres orthonormés (métrique canonique) associés sont $w_1, w_2, \ldots, w_f$. Les q premières valeurs propres sont conservées dans la matrice diagonale $\Delta$ et les q premiers vecteurs propres sont conservés en colonnes dans la matrice W. Cette diagonalisation correspond à l'ACP du triplet $(T_0, U \Lambda^{-1} U', D_t)$.

5) Les produits de cette procédure sont les matrices B, D, R et E avec :

$$B = U \Lambda^{-1/2} W$$

$$D = U \Lambda^{1/2} W$$

$$R = X_0 B$$

$$E = D_t^{-1} P' R = T_0 B$$

B comportant p lignes et q colonnes est dite matrice des q premiers facteurs de l'ACC. D comportant p lignes et q colonnes est dite matrice des covariances variables-codages. R comportant n lignes et q colonnes contient les codages canoniques des individus. E comportant t lignes et q colonnes est la matrice des coordonnées factorielles des colonnes du tableau T. Cette procédure de calcul est justifiée par les propriétés suivantes exprimées dans le cadre du schéma de dualité, plus précisément dans la présentation récente de TENENHAUS & YOUNG (1985).

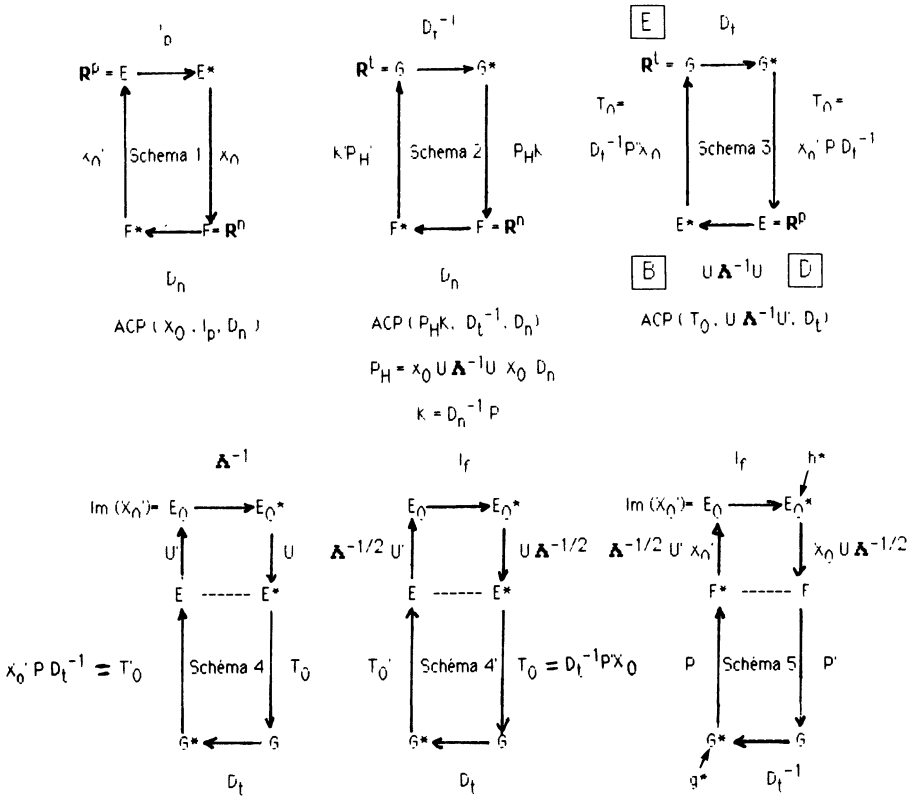
## L'ACC comme analyse discriminante

On notera E l'espace $\mathbf{R}^p$, E* son dual, F l'espace $\mathbf{R}^n$, F* son dual, G l'espace $\mathbf{R}^t$ et G* son dual. Les vecteurs, sauf indication contraire, sont exprimés par leurs composantes dans les bases canoniques. Les matrices, sauf indication contraire, définissent les applications linéaires en jeu par rapport aux bases canoniques. La première diagonalisation est celle du schéma 1 (voir tableau I), soit l'ACP pondérée du triplet $(X_0, I_p, D_n)$. L'image de $X_0$ est engendrée par les composantes principales ($D_n$-normées) consignées dans les colonnes de

$$H = X_0 U \Lambda^{-1/2}$$

TABLEAU I

Six schémas de dualité associés à l'ACC. 1 — ACP initiale, 2 — ACP d'un nuage projeté, 3 — ACP du tableau des moyennes conditionnelles, 4 et 4′ — Diagonalisation d'une matrice symétrique et maximisation de la variance des moyennes conditionnelles, 5 — Maximisation de la corrélation sous contrainte linéaire associée à une distribution bivariée. Les cinq derniers schémas ont en commun un opérateur « qui en fait le tour » ce qui permet de relier les différents modes d'approche, comme le font TENENHAUS & YOUNG (1985) pour l'AFC Multiples.



Le $D_n$-projecteur de F sur $Im(X_0)$, sous-espace engendré par ces composantes principales, a pour matrice (par rapport à la base canonique)

$$P_H = X_0 U \Lambda^{-1} U' X_0' D_n = HH' D_n$$

car $\quad P_H P_H = P_H,\ P_H H = H \quad$ et $\quad P_H' D_n P_H = P_H' D_n = D_n P_H$

Dans le cas particulier où C est de rang p, on a l'écriture classique du projecteur :

$$P_H = X_0 (X_0' D_n X_0)^{-1} X_0' D_n$$

Notons K le tableau des distributions conditionnelles lignes associés à T soit :

$$K = D_n^{-1} P$$

et considérons l'ACP du nuage des colonnes de K projeté sur $Im(X_0)$, suivant la méthode introduite par LAURO & D'AMBRA (1983, ACP par rapport à un espace de référence ou ACP d'un nuage projeté qui définit, par exemple, l'AFC non symétrique). On obtient le schéma 2 (cf. tableau I), c'est-à-dire l'ACP du triplet $(P_H K,\ D_t^{-1},\ D_n)$. La matrice à diagonaliser $K' P_H' D_n P_H K D_t^{-1}$ peut se réécrire, puisque $P_H' D_n P_H = D_n P_H$ :

$$D_t (D_t^{-1} P' X_0)\ (U \Lambda^{-1} U')\ (X_0' P D_t^{-1}) = D_t T_0 (U \Lambda^{-1} U')\ T_0'$$

qui renvoie au schéma 3. La diagonalisation de la matrice symétrique

$$\Lambda^{-1/2} U' X_0' P D_t^{-1} P' X_0 U \Lambda^{-1/2} = \Lambda^{-1/2} U' T_0' D_t T_0 U \Lambda^{-1/2} = A$$

a été retenue dans la procédure ci-dessus. Le schéma 3 est équivalent au schéma 4 où la matrice U' caractérise le projecteur sur $Im(X_0')$ par rapport à la base canonique de E et la base des axes principaux engendrant $Im(X_0')$.

$U \Lambda^{-1} U'$ est la métrique de MAHALANOBIS quand $X_0' D_n X_0$ est inversible; elle reste une norme sur $Im(X_0')$ dans le cas contraire. l'ACC est donc l'ACP du nuage des moyennes conditionnelles par colonne de T des variables centrées ou normalisées de X (tableau $T_0$), avec la pondération de l'AFC de T et la norme de MAHALANOBIS associée au tableau X. C'est donc une généralisation de l'analyse discriminante vue comme ACP du nuage des centres de gravités (CAILLIEZ & PAGES, 1976, p. 412) et se confond avec cette analyse quand le tableau T est formé par les indicatrices des modalités d'une variable qualitative.

La matrice B contient donc les facteurs principaux de l'analyse du schéma 3, la matrice E contient les coordonnées factorielles de la même analyse et la matrice D contient les axes principaux $(U \Lambda^{-1} U')$-orthonormés dans l'image de $X_0'$. Les facteurs principaux, vecteurs de E*, ont des composantes dans la base canonique de E* qui peuvent être considérées comme des coefficients de combinaisons linéaires des variables de $X_0$. Les valeurs de ces dernières sont dans la matrice R. La matrice de dispersion associée s'écrit, puisque $X_0' D_n X_0 U = C U = U \Lambda$, et que $U' U = I$ :

$$D_R = R' D_n R = W' \Lambda^{-1/2} U' X_0' D_n X_0 U \Lambda^{-1/2} W = W' W = I_q$$

ce qui signifie qu'elles sont de variance unité et non corrélées deux à deux. Les facteurs principaux (cf. TENENHAUS & YOUNG op. cit.) maximisent successivement dans G* la quantité

$$\| T_0 (e^*) \|^2 = \| D_t^{-1} P' X_0 (e^*) \|^2 = e^{*\prime} X_0' P D_t^{-1} D_t D_t^{-1} P' X_0 e^*$$

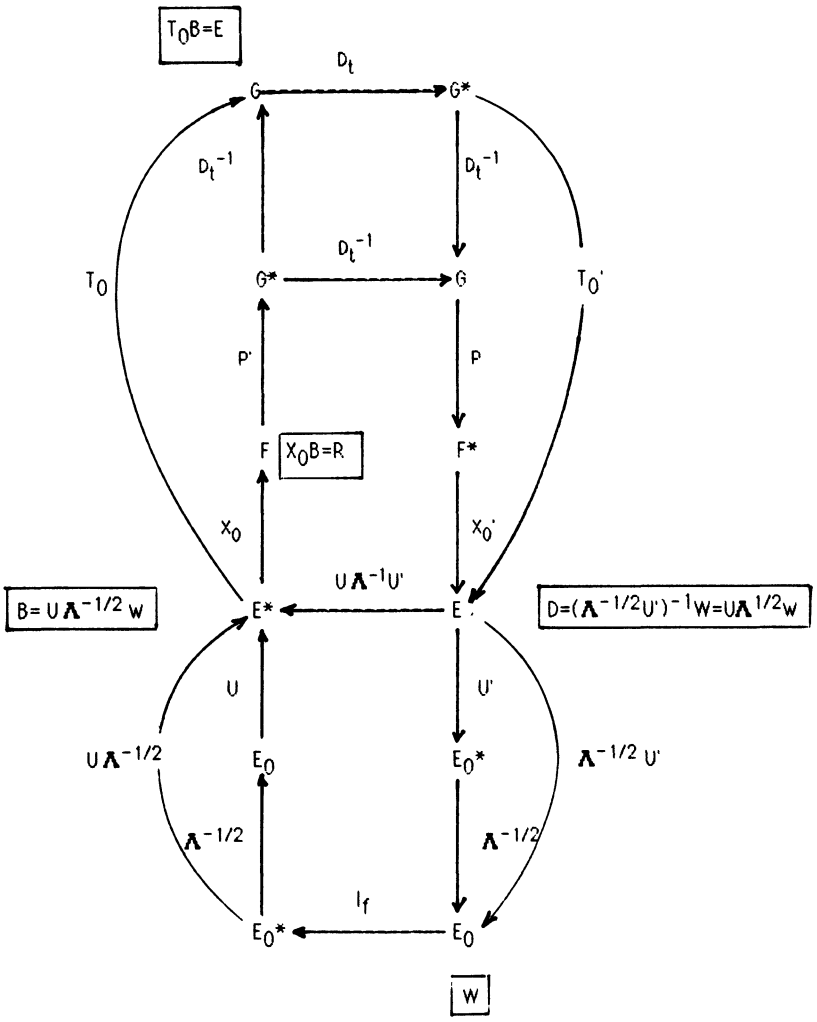
ce qui apparaît clairement sur le schéma 4' et permet de définir l'analyse considérée par la propriété suivante.

***Proposition 1 :***

Si on cherche une combinaison linéaire des variables de $X_0$ de variance unité qui maximise la variance des moyennes conditionnelles par colonnes du tableau T avec la pondération $D_t$ on prendra pour coefficients de cette combinaison linéaire les scalaires de la première colonne de la matrice B (premier facteur de l'analyse).

TABLEAU II

Schéma 6 résumant les schémas 3 à 5. On a placé en encadré les matrices dont les colonnes (axes factoriels, facteurs, composantes principales) correspondent à des vecteurs de l'espace considéré.



Les valeurs de cette combinaison linéaire sont dans la première colonne de R. Le maximum atteint est la première valeur propre $\delta_1$. Les moyennes conditionnelles par colonne de T sont dans la première colonne de E et forment les coordonnées factorielles sur le premier axe principal des colonnes de T. Si on cherche une nouvelle combinaison indépendante de la première pour la pondération $D_n$, on prendra le second facteur et ainsi de proche en proche. *L'analyse considérée est donc l'analyse canonique des correspondances (ACC) de TER BRAAK.*

Observons que la procédure de l'auteur (cf. introduction) cherche les solutions dans $\mathbf{R}^n$ de l'équation

$$X_0 C^{-1} X_0' D_n D_n^{-1} P D_t^{-1} P' f = \delta f$$

où

$$\delta = (f' D_n f)^{1/2}$$

ce qui renvoie à la diagonalisation dans la plus grande des 3 dimensions (n, t, p), à savoir n, ce qui est loin d'être optimal numériquement. Le schéma de synthèse consigné dans le tableau II résume l'ensemble de ces remarques.

## L'ACC comme analyse en composantes principales

Les combinaisons linéaires précitées peuvent être appelées *codages canoniques.* Le terme d'analyse canonique des correspondances ne souligne pas la stratégie d'ACP sur variables instrumentales, ou de projection sur un sous-espace de référence, ou de contrainte linéaire (cf. la synthèse de SABATIER 1983) essentiellement dissymétrique qui la génère. Ces réserves pourraient amener à modifier l'appellation choisie par son inventeur.

Les covariances variables-codages canoniques sont contenues dans la matrice :

$$X_0' D_n R = X_0' D_n X_0 U \Lambda^{-1/2} W = U \Lambda^{1/2} W = D$$

car les variables de $X_0$ sont $D_n$-centrées par l'hypothèse et que celles de R sont $D_n$-normées par construction. Dans l'option normalisation (cas des variables d'unités arbitraires) D contient des coefficients de corrélation variables-codages qui donne par représentation plane des cercles de corrélation comme en ACP ou en analyse discriminante. Remarquons de plus que la matrice D contient en colonne les axes factoriels du schéma 3 dans la base canonique car :

$$D' (U \Lambda^{-1} U') D = W' W = I_q$$

et

$$T_0' D_t T_0 (U \Lambda^{-1} U') D = U \Lambda^{1/2} \Lambda^{-1/2} U' T_0' D_t T_0 U \Lambda^{-1/2} W$$
$$= U \Lambda^{1/2} A W = U \Lambda^{1/2} W D = D \Delta$$

car $UU'$ est le $I_p$-projecteur sur $\mathrm{Im}(X_0')$ et laisse invariant les vecteurs de ce sous-espace. Il s'en suit que les projections des vecteurs de la base canonique de E sur les axes principaux du schéma 3 sont dans :

$$I_p (U \Lambda^{-1} U') D = U \Lambda^{-1/2} W = B$$

D'où la propriété suivante :

***Proposition 2 :***

On peut dépouiller l'ACC par la représentation simultanée sur les plans factoriels des trois éléments de l'analyse à savoir :

a) Les lignes des tableaux (individus) sont placés par les codages canoniques (colonnes de R), comme projection des lignes de $X_0$. Le nuage est $D_n$-normé (moyennes nulles, variances unitaires et covariances nulles).

b) Les variables du tableau X sont placées par les facteurs principaux (colonnes de B), comme projection des vecteurs de la base canonique de E. La représentation se fait par des vecteurs. Les composantes de ces vecteurs sont les coefficients des combinaisons linéaires définissant les codages canoniques.

c) Les colonnes du tableau T sont placées par les coordonnées factorielles (colonnes de E). Chaque colonne y figure au centre de gravité du nuage des points-individus pour la distribution conditionnelle qui lui correspond dans T. Le nuage est $D_t$-centré, non corrélé, de variances $\delta_k$.

Ces remarques transcrivent les propriétés du schéma 6 (tableau II).

## L'ACC comme double codage sous contrainte linéaire

Considérons enfin f un codage normalisé des individus pour la pondération $D_n$ défini comme combinaison linéaire des variables du tableau $X_0$. Ecrivons $f = (f_1, \ldots, f_n)'$ comme vecteur de F dans la base canonique. Définissons de même g* un codage normalisé des colonnes de T, pour la pondération $D_t$

$$g^* = (g_1, \ldots, g_t)'$$

comme vecteur de G* dans la base canonique. La distribution de fréquences bivariée P définit la corrélation

$$R(f, g^*) = \sum_{i,j} p_{ij} f_i g_j = \langle P' f, g^* \rangle$$

avec la notation traditionnelle des crochets de dualité.

f s'écrit de manière unique comme combinaison linéaire $R\alpha$ des composantes principales du schéma 1 (base orthonormée de vecteurs propres de $Im(X_0)$), soit :

$$f = R\alpha = X_0 B \alpha = X_0 U \Lambda^{-1/2} W \alpha = X_0 U \Lambda^{-1/2} h^*$$

où la contrainte $f' D_n f = 1$ impose que $h^{*\prime} h^* = 1$; en effet :

$$f' D_n f = h^{*\prime} \Lambda^{-1/2} U' X_0' D_n X_0 U \Lambda^{-1/2} h^* = h^{*\prime} h^* = 1$$

La quantité R (f, g*) devient

$$\langle P' X_0 U \Lambda^{-1/2} h^*, g^* \rangle = \langle \Lambda^{-1/2} U' X_0' P g^*, h^* \rangle$$

soit exactement la quantité successivement maximisée par les couples (facteurs-cofacteurs) du schéma 5, les cofacteurs étant aux composantes principales ce que les facteurs sont aux axes principaux dans la présentation de TENENHAUS (1983). Comme le vecteur $\mathbf{1}_t$ de G* est dans le noyau de l'application $X_0' P$ car

$$P \mathbf{1}_t = D_n \mathbf{1}_n = (p_{.1}, \ldots, p_{.n})' \quad \text{et} \quad X_0' D_n \mathbf{1}_n = 0 \text{ par centrage,}$$

les cofacteurs du schéma 5 sont $D_t$-normés et $D_t$-orthogonaux à $\mathbf{1}_t$ et forment donc des codages normalisés. Dans la perspective du dépouillement des analyses par codage numérique ouverte par SAPORTA (1975), on peut donc affirmer que :

***Proposition 3 :***

Si on cherche un couple formé par une combinaison linéaire des variables de $X_0$ de variance unité et un codage numérique des colonnes de T normalisé (pour la distribution marginale) qui maximise la corrélation induite par le tableau P on prendra respectivement le premier codage canonique (première colonne de R) et la première coordonnée factorielle normalisée (première colonne de E divisée par $\delta_1^{1/2}$). Le maximum atteint est $\delta_1^{1/2}$. Si on cherche un nouveau couple de codages numériques non corrélés aux précédents qui maximise à nouveau la même quantité on prendra les éléments propres de rang 2, etc.

L'ACC est donc une AFC sous contraintes linéaires et se confond avec l'AFC (AFC du tableau P) si le tableau X est celui des indicatrices des individus (auquel cas X est la matrice identité d'ordre n). Lorsqu'il y a une seule variable dans X, l'ACC se réduit au calcul de la moyenne conditionnelle par colonne de T de cette variable normalisée pour la pondération marginale $D_n$. Dans le cas d'une variable qualitative, l'ACC est équivalente à l'AFC de P′ X. L'ACC apparaît de plus comme une *méthode de double codage (dual scaling) sous contrainte* et invite à définir des méthodes de double codage sous double contrainte. Notons que si $\delta_1$ est la première valeur propre de l'ACC de X et de T, si $\mu_1$ est la première valeur propre (non triviale) de l'AFC de T, si $R^2$ est le carré de la corrélation multiple entre la première coordonnée factorielle de l'AFC de T et le tableau X (régression $D_n$-pondérée, voir par exemple PRODON & LEBRETON 1981) on a :

$$\mu_1 R^2 < \delta_1 < \mu_1$$

et que le rapport $\delta_1/\mu_1$, toujours inférieur à l'unité, mesure correctement l'aptitude du tableau X à rendre compte de la structure interne du tableau T.

## Un exemple d'utilisation en hydrobiologie

La faune ichtyologique présente des caractéristiques « d'indicateur biologique, impitoyable, de la qualité de l'eau et, de manière plus générale, de la qualité du milieu aquatique, dans ses composantes physiques, chimiques et biologiques » (PHILIPPART & VRANKEN 1983). Les espèces de poissons se répartissent dans les divers biotopes d'un cours d'eau selon des règles précises qui forment un modèle de répartition connu sous le nom de typologie d'HUET (1949 a, b, 1954). La thèse de VERNEAUX (1973) contient un très important ensemble de données numériques, mesures de la faune et du milieu, qui nous permettent d'illustrer le fonctionnement de l'ACC à partir d'une structure écologique relativement bien connue.

Le tableau X est formé de n = 35 stations d'étude réparties sur la totalité du cours du Doubs. On a sélectionné p = 11 variables en répartissant l'information entre les caractères morphologiques de la rivière (distance à la source, altitude, pente, débit) et les mesures de la qualité de l'eau (pH, calcium,

## TABLEAU III

Tableau X comportant n = 35 stations et p = 11 variables de milieu. 1 — Distance à la source (km × 10); 2 — Altitude (m); 3 — Pente (‰ × 10); 4 — Débit moyen minimum ($m^3/s$ × 100); 5-pH (× 10); 6 — Dureté totale (mg/l de Calcium); 7 — Phosphates (mg/l × 100); 8 — Nitrates (mg/l × 100); 9 — Azote ammoniacal (mg/l × 100); 10 — Oxygène dissous (mg/l × 10); 11 — Demande biologique en oxygène $DBO_5$ (mg/l × 10). Données numériques extraites de VERNEAUX (1973).

| * | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1* | 3 | 934 | 480 | 84 | 79 | 45 | 1 | 20 | 0 | 122 | 27 |
| 2* | 22 | 932 | 30 | 100 | 80 | 40 | 2 | 20 | 10 | 103 | 19 |
| 3* | 102 | 914 | 37 | 180 | 83 | 52 | 5 | 22 | 5 | 105 | 35 |
| 4* | 185 | 854 | 32 | 253 | 80 | 72 | 10 | 21 | 0 | 110 | 13 |
| 5* | 215 | 849 | 23 | 264 | 81 | 84 | 38 | 52 | 20 | 80 | 62 |
| 6* | 324 | 846 | 32 | 286 | 79 | 60 | 20 | 15 | 0 | 102 | 53 |
| 7* | 268 | 841 | 66 | 400 | 81 | 88 | 7 | 15 | 0 | 111 | 22 |
| 8* | 491 | 792 | 25 | 130 | 81 | 94 | 20 | 41 | 12 | 70 | 81 |
| 9* | 705 | 752 | 12 | 480 | 80 | 90 | 30 | 82 | 12 | 72 | 52 |
| 10* | 834 | 746 | 0 | 0 | 78 | 98 | 90 | 310 | 20 | 62 | 102 |
| 11* | 990 | 617 | 99 | 1000 | 77 | 82 | 6 | 75 | 1 | 100 | 43 |
| 12* | 1061 | 607 | 0 | 0 | 88 | 105 | 58 | 70 | 25 | 93 | 23 |
| 13* | 1234 | 483 | 41 | 1990 | 81 | 96 | 30 | 160 | 0 | 115 | 27 |
| 14* | 1324 | 477 | 16 | 2000 | 79 | 86 | 4 | 50 | 0 | 122 | 30 |
| 15* | 1436 | 450 | 21 | 2110 | 81 | 98 | 6 | 52 | 0 | 124 | 24 |
| 16* | 1522 | 434 | 12 | 2120 | 83 | 98 | 27 | 123 | 0 | 123 | 38 |
| 17* | 1645 | 415 | 5 | 2300 | 86 | 86 | 40 | 100 | 0 | 117 | 21 |
| 18* | 1755 | 408 | 0 | 0 | 81 | 90 | 60 | 140 | 10 | 114 | 47 |
| 19* | 1818 | 386 | 0 | 0 | 80 | 90 | 50 | 171 | 5 | 101 | 42 |
| 20* | 1859 | 375 | 20 | 1610 | 80 | 88 | 20 | 200 | 5 | 103 | 27 |
| 21* | 1945 | 354 | 0 | 0 | 81 | 90 | 60 | 200 | 10 | 90 | 48 |
| 22* | 1985 | 348 | 5 | 2430 | 80 | 92 | 20 | 250 | 20 | 102 | 46 |
| 23* | 2110 | 332 | 8 | 2500 | 80 | 90 | 50 | 220 | 20 | 103 | 28 |
| 24* | 2246 | 310 | 5 | 2590 | 81 | 84 | 60 | 220 | 15 | 106 | 33 |
| 25* | 2477 | 286 | 8 | 2680 | 80 | 86 | 30 | 300 | 30 | 103 | 28 |
| 26* | 2812 | 262 | 10 | 2720 | 79 | 85 | 20 | 220 | 10 | 90 | 41 |
| 27* | 2940 | 254 | 14 | 2790 | 81 | 88 | 20 | 162 | 7 | 91 | 48 |
| 28* | 3043 | 246 | 12 | 2880 | 81 | 97 | 260 | 350 | 115 | 63 | 164 |
| 29* | 3147 | 241 | 3 | 2976 | 80 | 99 | 140 | 250 | 60 | 52 | 123 |
| 30* | 3278 | 231 | 5 | 3870 | 79 | 100 | 422 | 620 | 180 | 41 | 167 |
| 31* | 3579 | 214 | 5 | 3910 | 79 | 94 | 143 | 300 | 30 | 62 | 89 |
| 32* | 3732 | 206 | 12 | 3960 | 81 | 90 | 58 | 300 | 26 | 72 | 63 |
| 33* | 3947 | 195 | 3 | 4320 | 83 | 100 | 74 | 400 | 30 | 81 | 45 |
| 34* | 4220 | 183 | 6 | 6770 | 78 | 110 | 45 | 162 | 10 | 90 | 42 |
| 35* | 4530 | 172 | 2 | 6900 | 82 | 109 | 65 | 160 | 10 | 82 | 44 |

## TABLEAU IV

Tableau T comportant n = 35 stations et t = 27 espèces de Poissons.

| | | | | | |
|---|---|---|---|---|---|
| 1 **a** | Chabot | *Cottus gobio* L. | 2 **b** | Truite | *Salmo trutta fario* L. |
| 3 **c** | Vairon | *Phoxinus phoxinus* L. | 4 **d** | Loche | *Nemacheilus barbatulus* L. |
| 5 **e** | Ombre | *Thymallus thymallus* L. | 6 **f** | Blageon | *Telestes soufia agassizi* C. |
| 7 **g** | Hotu | *Chondrostoma nasus* L. | 8 **h** | Toxostome | *Chondrostoma toxostoma* Vallot |

| | | | | | |
|---|---|---|---|---|---|
| 9 **i** | Vandoise | *Leuciscus leuciscus* L. | 10 **j** | Chevaine | *Leuciscus cephalus cephalus* L. |
| 11 **k** | Barbeau | *Barbus barbus* L. | 12 **l** | Spirlin | *Spirlinus bipunctatus* Bloch |
| 13 **m** | Goujon | *Gobio gobio* L. | 14 **n** | Brochet | *Esox lucius* L. |
| 15 **o** | Perche | *Perca fluviatilis* L. | 16 **p** | Bouvière | *Rhodeus amarus* Bloch |
| 17 **q** | Perche soleil | *Lepomis gibbosus* L. | 18 **r** | Rotengle | *Scardinius erythrophtalmus* L. |
| 19 **s** | Carpe | *Cyprinus carpio* L. | 20 **t** | Tanche | *Tinca tinca* L. |
| 21 **u** | Brème | *Abramis brama* L. | 22 **v** | Poisson-Chat | *Ictalurus melas* Rafinesque |
| 23 **w** | Grémille | *Acerina cernua* L. | 24 **x** | Gardon | *Rutilus rutilus* L. |
| 25 **y** | B. bordelière | *Blicca bjoerkna* L. | 26 **z** | Ablette | *Alburnus alburnus* L. |
| 27 + | Anguille | *Anguilla anguilla* Shaw | | | |

Données numériques et nomenclature extraites de VERNEAUX (1973).

| | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | . | 3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 2 | . | 5 | 4 | 3 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 3 | . | 5 | 5 | 5 | . | . | . | . | . | . | . | . | . | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 4 | . | 4 | 5 | 5 | . | . | . | . | . | 1 | . | . | 1 | 2 | 2 | . | . | . | . | 1 | . | . | . | . | . | . | . |
| 5 | . | 2 | 3 | 2 | . | . | . | . | 5 | 2 | . | . | 2 | 4 | 4 | . | . | 2 | . | 3 | . | . | . | 5 | . | . | . |
| 6 | . | 3 | 4 | 5 | . | . | . | . | 1 | 2 | . | . | 1 | 1 | 1 | . | . | . | . | 2 | . | . | . | 1 | . | . | . |
| 7 | . | 5 | 4 | 5 | . | . | . | . | 1 | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 8 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 9 | . | . | 1 | 3 | . | . | . | . | . | 5 | . | . | . | . | . | . | . | . | . | 1 | . | . | . | 4 | . | . | . |
| 10 | . | . | . | . | . | . | . | . | 1 | 3 | . | . | 1 | 2 | 1 | . | . | 2 | 1 | 2 | 1 | . | . | 5 | 5 | . | . |
| 11 | . | 1 | 4 | 4 | . | . | . | . | 2 | 2 | . | . | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 12 | . | 1 | 3 | . | . | . | . | . | 1 | 1 | . | . | 1 | 3 | 3 | . | . | 3 | 1 | 2 | 2 | . | 1 | 5 | 3 | . | . |
| 13 | 1 | 3 | 4 | 1 | 1 | . | . | . | . | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 14 | 2 | 5 | 4 | 4 | 2 | . | . | . | . | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 15 | 2 | 5 | 5 | 2 | 3 | 2 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 16 | 3 | 5 | 5 | 4 | 4 | 3 | . | . | . | 1 | 1 | . | 1 | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 17 | 3 | 4 | 4 | 5 | 2 | 4 | . | . | 3 | 3 | 2 | . | 2 | . | . | . | . | . | . | 1 | . | . | . | . | . | . | . |
| 18 | . | . | . | . | . | . | . | . | 5 | 3 | . | . | 3 | 3 | 3 | . | 1 | 3 | 1 | 3 | 3 | . | 4 | 3 | 3 | . | . |
| 19 | . | . | . | . | . | . | . | . | 3 | 3 | . | . | 3 | 3 | 4 | . | 1 | 3 | 2 | 3 | 3 | . | 4 | 5 | 4 | . | . |
| 20 | 2 | 3 | 3 | 5 | . | 5 | . | 4 | 5 | 2 | 2 | 1 | 2 | 1 | 1 | . | 1 | . | 1 | 1 | . | . | . | 1 | . | . | . |
| 21 | . | . | . | . | . | . | . | 5 | 5 | 1 | . | . | 2 | 2 | 2 | . | 1 | 3 | 1 | 2 | 2 | . | 3 | 4 | . | 3 | . |
| 22 | 1 | 2 | 4 | 4 | 1 | 2 | 1 | 4 | 3 | 2 | 3 | 4 | 1 | 1 | 2 | 1 | 1 | . | 1 | 1 | . | . | . | 2 | . | 2 | 1 |
| 23 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 3 | 2 | 3 | 3 | 3 | 2 | 1 | 3 | 2 | 1 | . | 1 | 1 | . | . | 1 | 2 | . | 2 | 1 |
| 24 | . | . | 3 | 5 | . | 1 | 2 | 3 | 2 | 1 | 2 | 2 | 4 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | . | 1 | 5 | 1 | 3 | 1 |
| 25 | . | . | 1 | 2 | . | . | 2 | 2 | 2 | 3 | 4 | 3 | 4 | 2 | 2 | 3 | 2 | 2 | 1 | 4 | 1 | . | 2 | 5 | 2 | 5 | 2 |
| 26 | . | . | 1 | 1 | . | . | 2 | 2 | 2 | 2 | 4 | 2 | 5 | 3 | 3 | 3 | 2 | 2 | 2 | 4 | 3 | 1 | 3 | 5 | 3 | 5 | 2 |
| 27 | . | . | . | 1 | . | . | 3 | 2 | 3 | 4 | 5 | 1 | 5 | 3 | 4 | 3 | 3 | 2 | 3 | 4 | 4 | 2 | 4 | 5 | 4 | 5 | 2 |
| 28 | . | . | . | . | . | . | . | . | . | 1 | . | . | . | . | . | . | . | . | . | . | . | . | . | 1 | . | 2 | . |
| 29 | . | . | . | . | . | . | 1 | . | . | 2 | . | . | 1 | . | . | . | 1 | . | . | . | . | . | 2 | 2 | 1 | 5 | . |
| 30 | . | . | . | . | . | . | . | . | 1 | 1 | . | . | 2 | 1 | . | . | . | 1 | . | . | . | . | 1 | 1 | . | 3 | . |
| 31 | . | . | . | 1 | . | . | 1 | . | 1 | 2 | 2 | 1 | 3 | 2 | 1 | 2 | 2 | 1 | 1 | 3 | 2 | 1 | 4 | 4 | 2 | 5 | 2 |
| 32 | . | . | . | 1 | . | . | 1 | 1 | 2 | 3 | 4 | 1 | 4 | 4 | 1 | 3 | 3 | 1 | 2 | 5 | 3 | 2 | 5 | 5 | 4 | 5 | 3 |
| 33 | . | . | . | 1 | . | . | 1 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 4 | 4 | 2 | 4 | 4 | 3 | 3 | 5 | 5 | 5 | 5 | 4 |
| 34 | . | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 4 | 5 | 3 | 5 | 5 | 4 | 5 | 5 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 4 | 5 | 4 |
| 35 | . | . | . | . | . | . | 1 | 2 | 3 | 3 | 3 | 5 | 5 | 4 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

phosphates, nitrates, ammoniaque, oxygène et demande biologique en oxygène). Les données sont consignées au tableau III. La seule variable 3 (pente) a été transformée ($y = \text{Log}(x + 1)$), sa distribution étant très dissymétrique.

Le tableau T comporte les mêmes stations ($n = 35$) et une mesure d'abondance de $t = 27$ espèces de poissons. Le code espèce et les valeurs enregistrées par VERNEAUX sur une échelle propre à chaque taxon figurent au tableau 4.

La figure 1 contient les principaux éléments de la description du milieu destinée à « expliquer » la répartition des espèces. Toutes les variables, sauf le pH, participent à la définition du premier codage des relevés qui transcrit globalement la structure amont-aval de la rivière (altitude et pente décroissante, débit augmentant avec la distance à la source, charge progressive en éléments minéraux et organiques). Cette direction n'est cependant pas exactement le premier axe, ce qui souligne deux propriétés de la rivière étudiée, la première étant une rupture de pente (entre les stations 10 et 13, cf. le profil en long et les secteurs morphologiques décrits par VERNEAUX, op. cit. p. 58), la seconde concernant la présence sur le cours supérieur de sources de pollution importante (agglomération de Pontarlier, op. cit. p. 81) qui affaiblissent la corrélation traditionnelle entre la qualité biologique de l'eau et l'altitude. L'utilisation du second codage souligne un troisième élément qui participe à la perturbation du gradient amont-aval. En suivant les numéros d'ordre des stations numérotées de haut en bas, 5 stations (10, 12, 18, 19, 21) ne correspondent pas à une évolution continue des paramètres : il s'agit des retenues artificielles ou naturelles qui interrompent le cours d'eau. La pente et le débit y prennent des valeurs nulles rapprochant ces relevés de ceux de la partie inférieure.

Les deux codages permettent alors simplement de regrouper les stations par leur position spatiale et le résumé des codes proposés par l'analyse : on observe 7 groupes notés de A à G sur la figure 1. En termes d'inertie, en utilisant les deux premiers codages, on tient compte de 75 % de l'inertie du nuage des moyennes par espèces des variables de milieu.

La carte factorielle des espèces (Fig. 2) replace chaque taxon au centre de gravité de sa distribution sur la carte des relevés dont la légende est le cercle des corrélations (Fig. 1). La première indication est celle de la présence de trois groupes d'espèces. Réduire cependant l'information apportée par un taxon à sa position moyenne c'est, du point de vue biologique, éliminer la notion centrale d'amplitude d'habitat (cf. CHESSEL & coll. 1982). Les moyens graphiques actuels permettent de restituer la totalité de l'information suivant une idée de WHITTAKER (1967), mise en œuvre par AUDA (1983) et utilisée en AFC par RICHARDOT-COULET & coll. (1986). L'abondance de chaque taxon est cartographiée dans le plan des relevés, ce qui explicite grandement leur mode de dispersion. Les concordances entre position moyenne et amplitude de répartition induisent une partition en 8 groupes détaillée dans la figure 2. La logique et les résultats de l'analyse peuvent alors se résumer par la figure 3.

On peut y voir une illustration étonnante des conceptions du milieu aquatique de HUET, dont les travaux ne contiennent aucune information numérique. Les termes d'usage courant établis par l'auteur permettent de légender

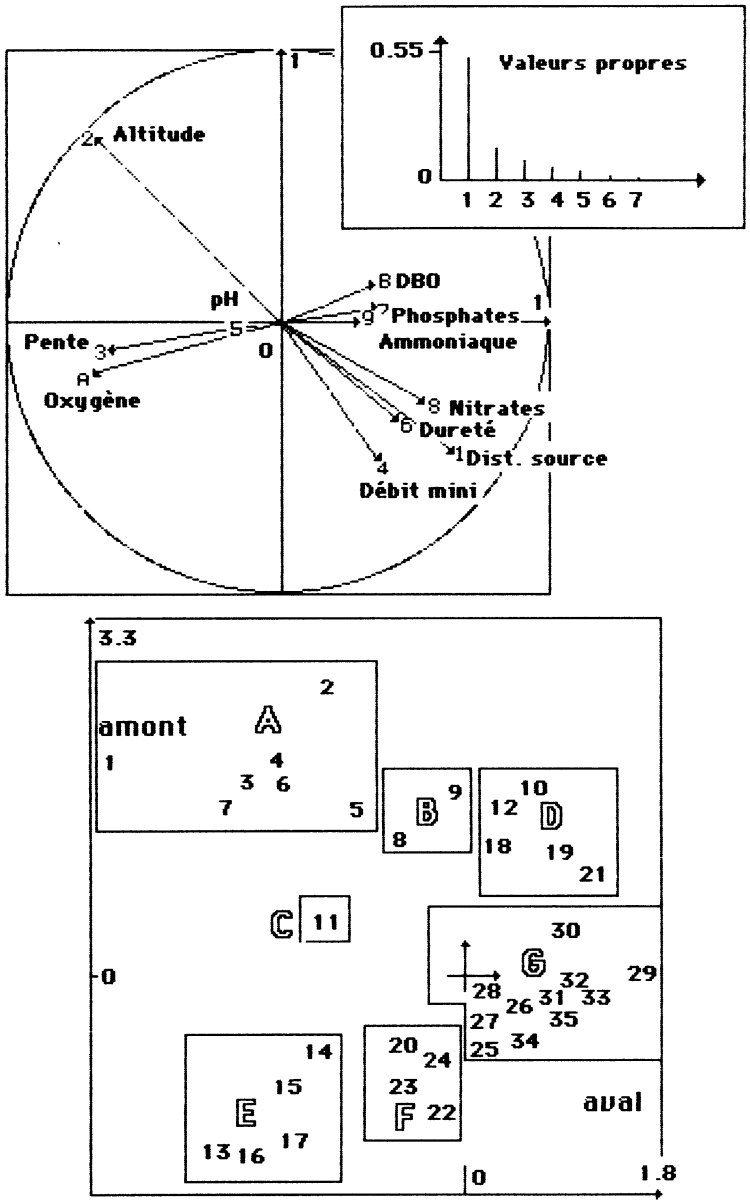



FIGURE 1

*Valeurs propres, cercle des corrélations et représentations des individus par codages canoniques en ACC. Le milieu aquatique est un gradient amont-aval perturbé par la présence de retenues (groupe D), les pollutions (groupe B) et les ruptures de pente. L'ACC donne les combinaisons de variables de milieu maximisant la dispersion entre les moyennes par espèces : à ce titre la figure indique d'abord une uniformisation progressive des communautés de haut en bas du cours d'eau.*

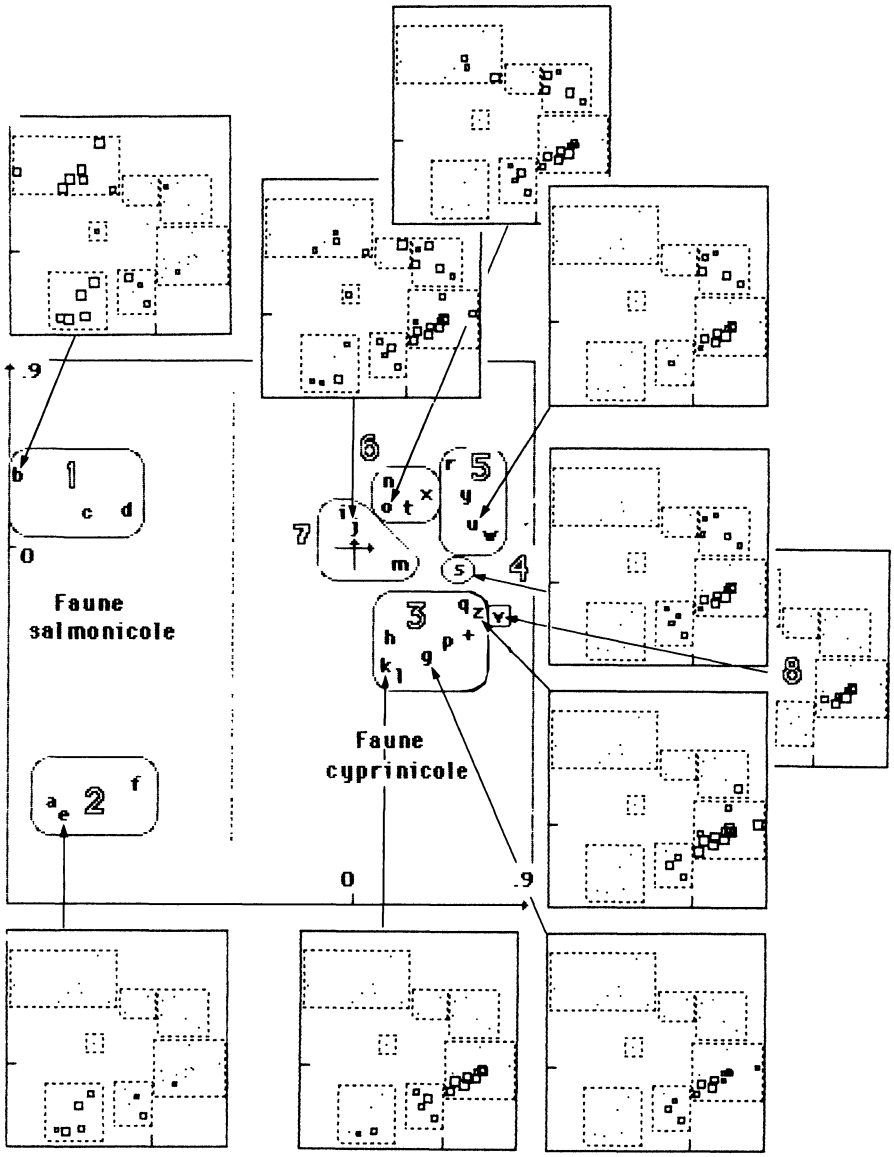
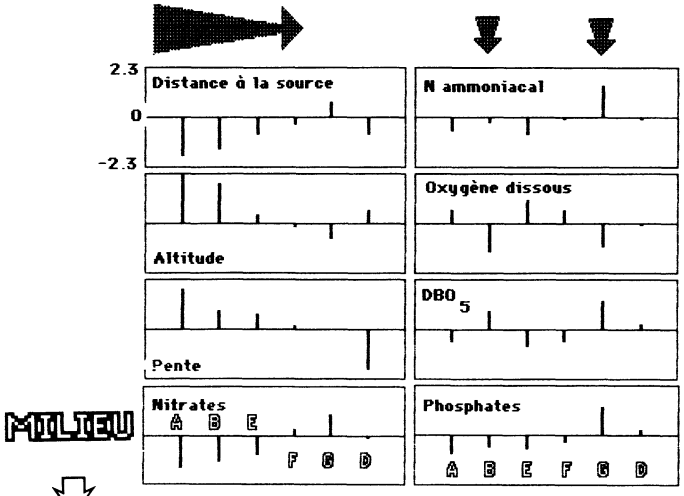



FIGURE 2

*Carte factorielle des espèces en ACC. Chaque taxon est au barycentre de sa distribution sur les relevés dans la figure 1. Pour préciser la dispersion autour du point moyen l'abondance de chaque taxon est représentée en chaque point par la surface d'un carré. Les espèces présentant une répartition très voisine sont regroupées et l'ensemble des taxons est partitionné en 8 groupes. Cette pratique est justifiée par la correspondance entre maximisation de la variance inter (utilisation du barycentre) et minimisation de la variance intra (utilisation des distributions) pour des codes relevés non corrélés de variance unité. Le dépouillement graphique strictement inféodé au principe de fonctionnement facilite le dialogue avec les utilisateurs non statisticiens.*




| | A+C | B | E | F | G | D |
|---|---|---|---|---|---|---|
| ESPACE | 1-7 11 | 8-9 | 13-17 | 20 22-24 | 25-35 | 10 12 18 19 21 |
| | Zone à Truite | pollution | Zone à Ombre | Zone à Barbeau | Zone à Brème | Milieu lentique |
| FAUNE | Zone salmonicole | | | Zone cyprinicole | | |
| 1 | 3.58 | 0.67 | 4.00 | 3.00 | 0.36 | 0.27 |
| 2 | — | — | 2.13 | 1.25 | 0.06 | — |
| 3 | — | — | — | 1.78 | 2.27 | 0.28 |
| 4 | — | — | — | 1.00 | 1.91 | 1.20 |
| 5 | 0.34 | — | — | 0.31 | 2.48 | 2.60 |
| 8 | — | — | — | — | — | 1.64 |
| 6 | 0.84 | 0.63 | 0.10 | 1.63 | 2.82 | 3.00 |
| 7 | 0.92 | 0.83 | 0.80 | 2.42 | 2.61 | 2.40 |
| Richesse | 6.0 | 2.5 | 7.8 | 21.3 | 22.4 | 13 |

FIGURE 3

*Réécriture simplifiée des données traitées dans le schéma de fonctionnement de l'ACC. En haut sont représentées les moyennes par groupe de relevés des variables de milieu normalisées. La flèche horizontale indique le gradient amont-aval et les flèches verticales les deux perturbations (pollution et rupture du cours d'eau par les retenues). Au centre sont indiqués les groupes de relevés et les qualificatifs classiques des zones utilisés depuis les travaux de HUET. En bas sont éditées les abondances moyennes par groupe d'espèces et de relevés dans la réorganisation du tableau faunistique rendant compte des structures observées. La richesse est le nombre moyen de taxons présents par groupe de relevés. Ces observations justifient l'emploi d'un code d'abondance propre à chaque espèce et replacé sur une échelle commune (0-5) introduit par VERNEAUX.*

la réécriture simplifiée du tableau qu'autorise l'ACC : coupure principale entre zone salmonicole et zone cyprinicole, partage de la première entre zone à Truite et zone à Ombre, partage de la seconde entre zone à Barbeau et zone à Brème, illustration de la « règle des pentes » (HUET op. cit.). Les données de VERNEAUX permettent de compléter la typologie en replaçant les milieux lentiques (eaux calmes) que traverse la rivière à la fin du gradient, ce qui confirme la prééminence de la vitesse du courant comme facteur écologique décisif du contenu des peuplements, d'autant plus que la pollution en zone salmonicole se traduit strictement par un appauvrissement de la faune sans substitution d'espèces.

## Perspectives

L'ACC marque l'introduction en écologie de méthodes linéaires spécifiques de couplage de tableaux écologiques. L'impact de ces nouvelles méthodes en écologie factorielle sera vraisemblablement considérable. Notons cependant que l'ACC, comme méthode de couplage entre un tableau d'AFC et un tableau d'ACP est l'une des trois stratégies possibles. La première, ici décrite, consiste à projeter le tableau des distributions conditionnelles colonnes de T sur le sous-espace engendré par $X_0$. La seconde consiste, symétriquement à projeter les colonnes de $X_0$ sur le sous-espace engendré par ces distributions conditionnelles et la troisième à faire l'analyse canonique entre ces deux sous-espaces. Il conviendra d'évaluer la valeur relative de ces pratiques en fonction de divers objectifs expérimentaux.

L'extension enfin aux variables qualitatives si elle peut se contenter en première approche de l'utilisation dans la même procédure des variables indicatrices des modalités (tableau disjonctif complet) sollicite des aides au dépouillement spécifiques que la synthèse de TENENHAUS & YOUNG permet d'établir.

La présentation d'une analyse par l'un de ses schémas de dualité qui induit les autres tend, avec l'enrichissement de l'ensemble des variantes disponibles, à devenir le seul moyen pratique d'identification. Notons dans cette optique que l'ACC dans le cas où t est sensiblement inférieur à p se fera simplement par la diagonalisation de la matrice symétrique issue du schéma 3, soit :

$$D_t^{-1/2}\, p'\, X_0\, (U\, \Lambda^{-1}\, U')\, X_0'\, p\, D_t^{-1/2}$$

ce qui fait également du schéma de dualité l'outil de base pour l'écriture des programmes.

## Références

AUDA Y. (1983). — Rôle des méthodes graphiques en analyse de données : application au dépouillement des enquêtes écologiques. Thèse de 3e cycle, Université Lyon 1, 127 pp.

CAILLIEZ F., PAGES J.P. (1976). — Introduction à l'analyse des données. SMASH, 9, rue Duban, 75016 Paris, 616 pp.

CHESSEL D., LEBRETON J.D., PRODON R. (1982). — Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés : cas d'un gradient simple, *C.R. Acad. Sc. Paris,* D, 295, 63-88.

HUET M. (1949 a). — Appréciation de la valeur piscicole des eaux douces. *Travaux de la Station de Recherches de GROENENDAAL,* série D, n° 10, 75 pp.

HUET M. (1949 b). — Aperçu des relations entre la pente et les populations piscicoles des eaux courantes. *Revue Suisse d'Hydrologie,* XI, 3-4, 332-351.

HUET M. (1954). — Biologie, profils en long et en travers des eaux courantes. *Bulletin Français de pisciculture,* 175, 41-53.

LAURO N., D'AMBRA L. (1983). — L'analyse non symétrique des correspondances. *in* Data analysis and Informatics, Vol. III, DIDAY E. et coll. Eds, ELSEVIA, North-Holland, 433-446.

PHILIPPART J.C., VRANKEN M. (1983). — Atlas des Poissons de Wallonie. Cahiers d'éthologie appliquée, Vol. 3, Supplément 1-2, 395 pp.

PRODON R., LEBRETON J.D. (1981). — Breeding avifauna of a Mediterranean succession : the holm oak and cork oak series in the eastern Pyrénées, 1 : Analysis and modelling of the structure gradient. *Oïkos,* 37, 21-38.

RICHARDOT-COULET M., CHESSEL D., BOURNAUD M. (1986). — Typological value of the benthos of old beds of a large river. Methodological approach. *Archiv. Hydrobiol.,* 107, 3, 363-383.

SABATIER R. (1983). — Approximations d'un tableau de données. Application à la reconstitution des paléoclimats. Thèse de 3$^{e}$ cycle, Montpellier, 184 pp.

SAPORTA G. (1975). — Liaisons entre plusieurs ensembles de variables et codage des données qualitatives. Thèse de 3$^{e}$ cycle, Paris VI, 102 pp.

TENENHAUS M. (1983). — L'analyse des données qualitatives par des méthodes de codage optimal. Les cahiers de Recherche du CESA, 78350 Jouy-en-Josas, CR 229, multigraph., 210 pp.

TENENHAUS M., YOUNG F.W. (1985). — An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika,* 50, 1, 91-119.

TER BRAAK C.J.F. (1986). — Canonical correspondence analysis : a new eingenvector technique for multivariate direct gradient analysis. *Ecology,* 67 (5), 1167-1179.

TER BRAAK C.J.F. (1987). — The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio,* 69, 69-77.

VERNEAUX J. (1973). — Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon, 257 pp.

WHITTAKER R.H. (1967). — Gradient analysis of vegetation. *Biol. Rev.,* 49, 207-264.

## Remerciements

Nous remercions vivement les deux referees de la revue pour leurs critiques et suggestions, G. CARREL et M. BOURNAUD pour leurs remarques sur le problème abordé et l'exemple traité.