

Rythmes saisonniers et composantes stationnelles en milieu aquatique

I. — *Description d'un plan d'observation complet par projection de variables*

S. Dolédec, D. Chessel

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

RÉSUMÉ

L'article décrit l'analyse statistique d'un tableau de relevés de milieu multivarié (variables quantitatives) obtenu en visitant s stations à t dates d'échantillonnage. Pour obtenir une mesure précise de la part respective des rythmes saisonniers et des composantes stationnelles dans la variabilité des mesures on peut simplement utiliser l'analyse en composantes principales sur les tableaux issus du modèle linéaire dans une analyse de variance à deux facteurs sans répétition. Cette pratique concerne une situation expérimentale très fréquente en hydrobiologie. Elle introduit à l'usage des méthodes de description multivariée par projection sur un sous-espace choisi par l'expérimentateur.

MOTS-CLÉS : *Variables physico-chimiques - Structures spatio-temporelles -
Projection de variables - Composantes principales - Hydrobiologie.*

ABSTRACT

The statistical analysis of a multivariate table of environmental samples (quantitative variables) collected at s locations for t sampling dates, is described. To have a distinct view of the respective influence of the seasonal succession and the sample location on the variability of the measures, Principal Components Analysis was used on tables from the linear model of variance analysis in the two-way layout with one observation per cell. Such spatio-temporal experimental programmes are frequently used in hydrobiological studies; the proposed practice introduces the use of descriptive methods of multivariate analysis by projection on a subspace chosen by the researcher.

KEY-WORDS: *Physico-chemical variables - Spatio-temporal structure - Projection of variables -
Principal Components - Hydrobiology.*

INTRODUCTION

Trouver dans un tableau de données la part de variabilité qui dépend ou ne dépend pas d'un autre tableau de données est un problème statistique important et, en écologie, une question centrale : une bonne partie de la statistique multidimensionnelle utilisée dans l'étude des écosystèmes est consacrée aux analyses des couples formés par un tableau florofaunistique et un tableau de variables de milieu. En témoigne l'ouvrage de GITTINS (1985) centré sur l'analyse canonique et plus spécialement sur son rôle en écologie. De même l'ouvrage remarquable de GREENACRE (1984) sur l'analyse des correspondances souligne la part active prise par la

réflexion proprement écologique à la compréhension du rôle de cette technique statistique devenue un standard.

Les pratiques de couplage de tableaux, dans la perspective synécologique, sont nombreuses : des jalons solides ont été posés par l'introduction successive de l'analyse des gradients (WHITTAKER, 1967), l'analyse discriminante (GREEN, 1971), l'analyse des correspondances sur tableaux de profils écologiques (ROMANE, 1972), la régression multiple (PRODON & LEBRETON, 1981) et l'analyse canonique (CARLETON, 1984) sur facteurs d'analyse des correspondances, enfin l'analyse canonique des correspondances (TER BRAAK, 1986). En abordant l'étude des écosystèmes aquatiques, l'hydrobiologie a ajouté à cette question un niveau de difficulté considérable en voulant préciser ce qui dans l'abondance des taxons dépendait d'un environnement par essence dominé par les rythmes saisonniers (température de l'eau) et les variations plus ou moins érratiques du débit. La recherche des relations faune-milieu déplacée du domaine des typologies d'état à celui des typologies de fonctionnement (par exemple dans AMOROS & CHESSEL, 1985) place les besoins exprimés en analyse de données écologiques au niveau des développements les plus récents de la théorie statistique.

On remarque alors la convergence qui s'établit entre plusieurs problèmes tels que :

- Trouver la part d'un tableau faunistique qui dépend des mesures environnementales (approche des facteurs écologiques);
- Trouver la part d'un tableau de milieu pouvant être reconstituée par un tableau faunistique (approche des indicateurs biologiques);
- Trouver dans une chronique multivariée et multistationnelle ce qui ne dépend que du temps, ou que d'une typologie des stations, ou qui ne peut s'expliquer que par une interaction des deux éléments.
- Trouver dans un tableau faunistique ce qui ne dépend pas des conditions de récolte, de capture ou de piégeage (élimination des effets météorologiques), etc.

Ces besoins trouvent une large panoplie de méthodes adaptées à ces divers objectifs dans la littérature statistique, mais l'emploi effectif de ces outils passe par un effort de diffusion auquel voudrait contribuer notre travail. Si les plans d'observations sont diversifiés à l'extrême, la méthodologie est au contraire coordonnée par un support mathématique principalement centré sur la géométrie euclidienne (CAILLIEZ, 1984) et le schéma de dualité (cf. ESCOUFIER, 1987). Les statisticiens placent au centre de cette partie mathématique la notion de projecteurs sur un sous-espace qui permet à TAKEUCHI *et al.* (1982) de relier dans une approche unifiée l'analyse de variance traditionnelle, la régression multiple ou l'analyse canonique et nombre de méthodes moins connues ou décrites indépendamment les unes des autres : par exemple, CHESSEL *et al.* (1987) retrouve l'analyse canonique des correspondances de TER BRAAK (1986) comme projection sur le sous-espace engendré par les variables de milieu.

Le but de cette note est d'illustrer cette notion centrale de projecteurs dans un des cas les plus simples, celui des plans d'observations complets de la qualité physico-chimique des eaux continentales et de montrer qu'il n'est pas difficile d'intégrer à la description d'un ensemble de données écologiques des objectifs clairement exprimés.

LES TABLEAUX DE PHYSICO-CHIMIE

Un des problèmes d'analyse des données le plus fréquemment posé en hydrobiologie est celui de la description des tableaux de variables de milieu (physico-chimie de l'eau en particulier) issus d'un plan d'observation espace-temps. Chaque relevé appartient à une station i ($1 \leq i \leq s$) et renvoie à une date j ($1 \leq j \leq t$). Le plan est complet si chaque station est visitée à chaque date. Il y a alors st prélèvements : chacun d'entre eux donne une valeur pour chacun des p descripteurs utilisés.

Quand une seule période d'échantillonnage est envisagée les données sont traitées par l'analyse en composantes principales (*ACP*) normée qui fait un bilan des corrélations intervariables. Quand une seule station est concernée l'expérimentateur doit décrire une chronique multivariée : la reconstitution des données par une *ACP* (CARREL, 1986; CARREL *et al.*, 1986) peut être employée. Dans le cas général on est en présence d'un tableau à trois indices où z_{ijk} est la mesure de la variable k dans la station i et à la date j . Le traitement des données, du point de vue statistique, renvoie aux méthodes d'analyses multitableaux (ESCOUFIER, 1980). Avant d'examiner le rôle de ces méthodes (cf. FOUCART, 1984) nous avons choisi de décrire une des approches les plus simples.

Les représentations graphiques utilisées par les expérimentateurs définissent clairement le problème à résoudre. Indépendamment des résumés techniques que constituent les diagrammes ioniques de Kufferath (cf. par exemple ILTIS & LÉVÈQUE, 1982; LÉVÈQUE *et al.*, 1983; MAQUET, 1983) de Schoëller et Berkloff (cf. BOUZIDI, 1983) ou de Symoens (cf. ROSILLON, 1984), les représentations les plus communément rencontrées dans la littérature (postérieure à 1954) se résument à des collections ou superpositions, au sens graphique d'AUDA (1983), réalisées en fonction du temps, de l'espace ou des variables (fig. 1); les deux premiers cas, de loin les plus courants, permettent une étude par variable; le troisième cas peut déboucher sur une analyse de fonctionnement mais la superposition graphique reste limitée à 3 variables pour une bonne lisibilité.

Quelques calculs statistiques de base peuvent être effectués, en fonction des objectifs expérimentaux, et conduisent à la représentation graphique (ORTAL & POR, 1977; GOEDMAKER, 1981; JOP, 1981...) ou cartographique (DESCY & EMPAIN, 1984...) des moyennes par station ou à la représentation des moyennes par date (MINSHALL & ANDREWS, 1973; HALL *et al.*, 1977; MOHAMMAD, 1980; EDWARDS & BROOKER, 1984...). Notons enfin que les outils statistiques de description des tableaux de relevés de milieux que constituent les analyses multivariées ne sont pas la règle, même si quelques références peuvent être citées qu'il s'agisse d'Analyses en Composantes Principales (FEUILLADE, 1972; DECAMPS *et al.*, 1976; JORDANA & GARCIA, 1981; VESPINI, 1985...) ou d'Analyses Factorielles des Correspondances (VERNEAUX & REZZOUK, 1974; GROSS, 1975; ANGELIER *et al.*, 1978; GRÉGOIRE & CHAMPEAU, 1984...).

Les objectifs expérimentaux reviennent en fait à identifier les évolutions communes entre stations ou entre variables, reconnaître au contraire les distorsions localisées, séparer ce qui revient à la variabilité interstationnelle, ce qui appartient au rythme commun, ce qui est interaction ponctuelle (données aberrantes, effet multiplicatif, ...). La présente note a pour but de montrer qu'il est possible d'automatiser ce travail de recherche.

On utilisera pour illustrer la procédure un jeu de données numériques de faibles dimensions mais d'intérêt biologique non négligeable extrait de la thèse de PEGAZ-MAUCET (1980). La figure 2 donne le tableau des données brutes et le protocole expérimental. Le Méaudret, ruisseau du Vercors septentrional, affluent de la Bourne, reçoit les eaux résiduaires de deux communes (Autrans, Méaudre). Cinq stations y ont été choisies d'amont en aval. Une sixième station est située sur la Bourne peu avant son confluent et sert de témoin. Ces 6 stations ont été visitées à quatre reprises et 10 variables ont été mesurées en chacun des 24 prélèvements.

On trouvera une présentation des objectifs expérimentaux dans l'article de BOURNAUD *et al.* (1983). La figure 3 résume l'Analyse en Composantes Principales Normée (*ACPN*) de ce tableau telle qu'elle est habituellement pratiquée. Les quatre premiers facteurs sont consacrés à la description de corrélations inter-variables liées aux structures spatio-temporelles. Le premier axe factoriel (57 % d'inertie) prend en compte pH, conductivité, oxygène, DBO_5 , oxydabilité, ammoniaque et phosphates. Il définit le gradient de minéralisation et le niveau global de *pollution organique* maximal à la station 2 et au mois d'août. Cette pollution se traduit classiquement par une acidification (baisse

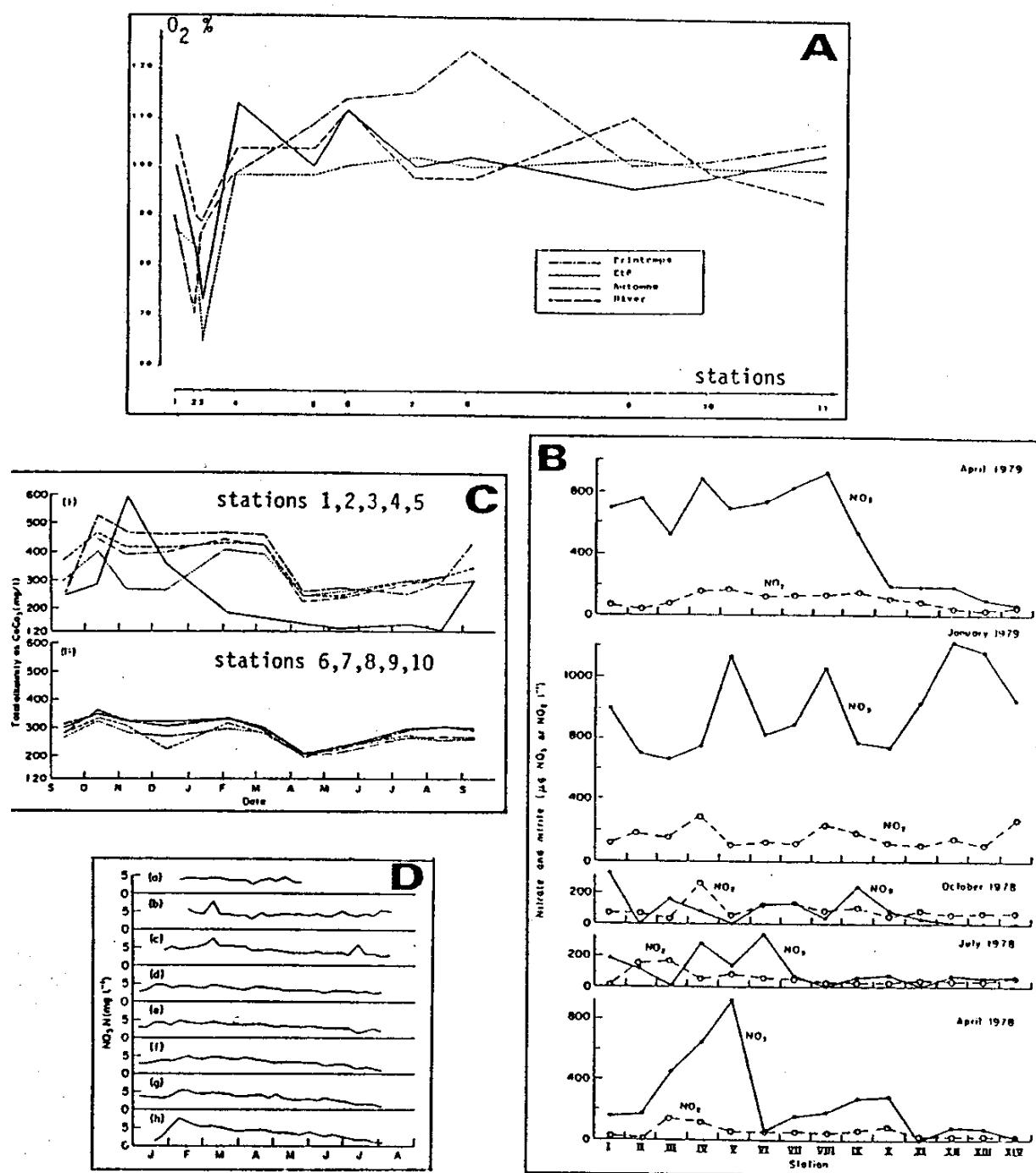


FIG. 1. — Représentations graphiques principales utilisées dans la description d'un plan d'échantillonnage espace-temps en milieu aquatique. **A** : Représentation date-station pour une variable (GIUDICELLI *et al.*, 1980). **B** : Représentation de 2 variables pour un profil-station, par date (SAAD & ABBAS, 1985). **C** : Représentation station-chronologie des prélevements (MINSHALL & ANDREWS, 1973). **D** : Représentation de la chronologie d'une variable pour chaque station (CASEY & LADLE, 1976). La diversité des figures utilisées dérive de la nécessité de représenter des données à trois indices (stations, dates, variables).

de pH), une diminution de la concentration en oxygène et par une augmentation de la DBO₅, de l'oxydabilité, des quantités de NH₄⁺ et PO₄³⁻ caractéristiques de la charge organique.

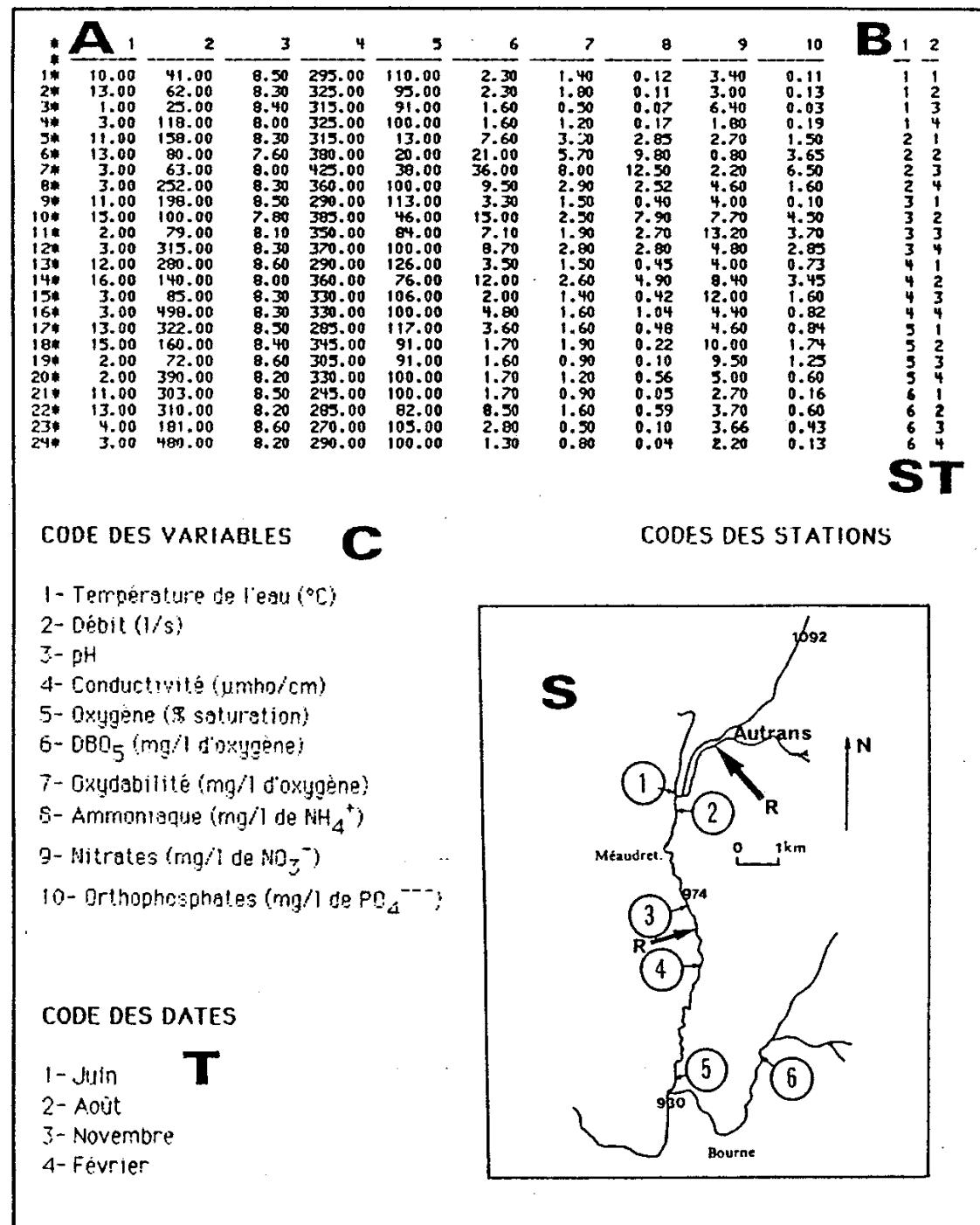


FIG. 2. — Exemple numérique d'un plan complet de mesures physico-chimiques. **A** : Tableau des données brutes (24 prélèvements x 10 variables). **B** : Tableau des variables expérimentales (S stations, T dates). **C** : Code des variables, **T** code des dates, **S** code des stations).

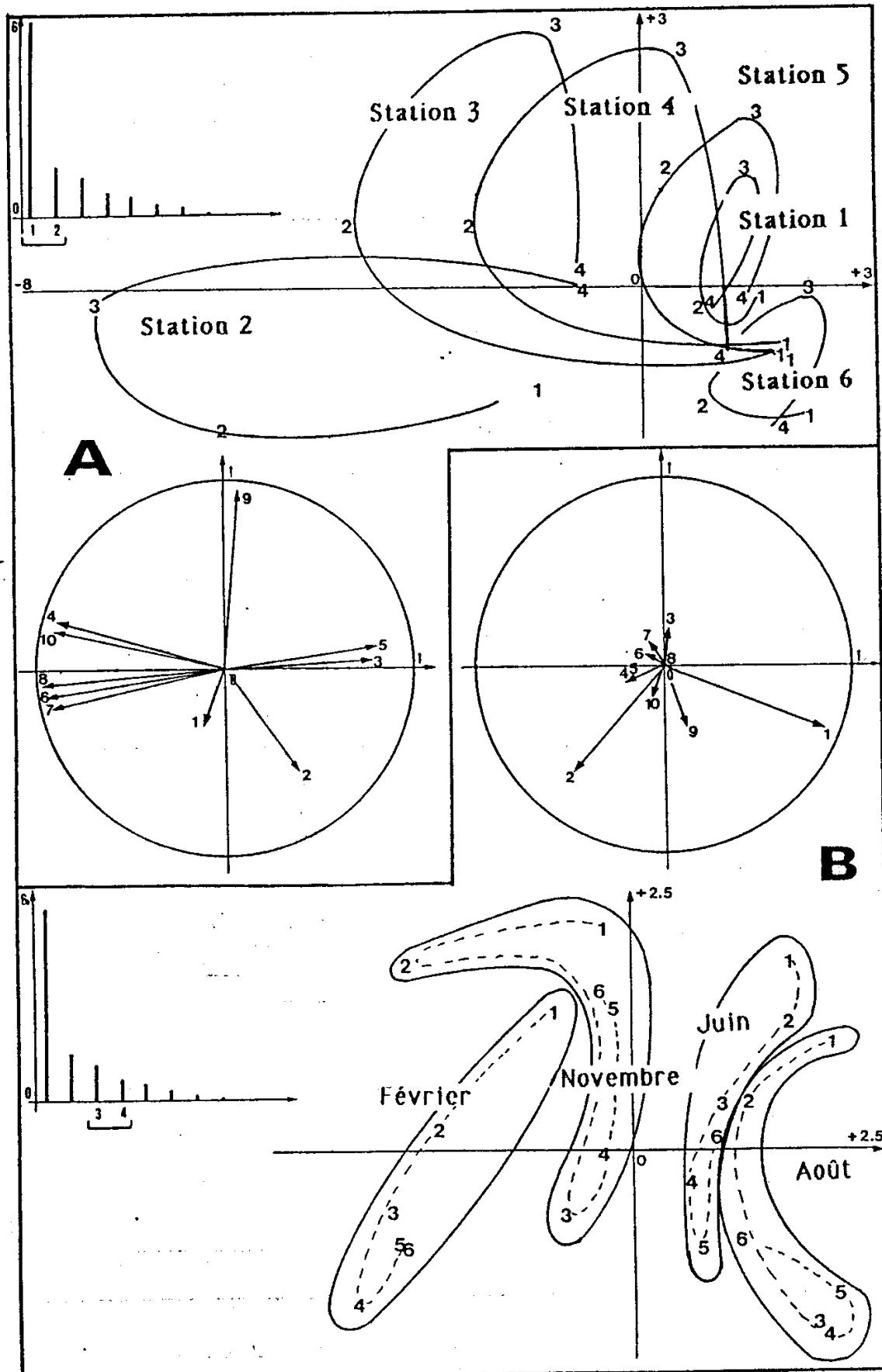


FIG. 3. — Analyse en Composantes Principales Normée du tableau des données brutes. A : plan d'observations complet par projection de variables. Acta Ecologica, Ecologia Generalis 8:403-426.

Le cours d'eau se restaure à partir de la station 3 jusqu'à la station 5, tandis que les stations 1 et 6 définissent un niveau témoin. Le rythme saisonnier de cette pollution se déforme d'une station à l'autre mais ne se confond pas avec celui de la température (axe 3). De même le gradient de restauration de la qualité de l'eau ne se confond pas avec le gradient amont-aval du débit (axe 4). L'analyse sans *a priori* décrit simultanément rythme saisonnier et typologie stationnelle intervenant dans le processus spatio-temporel généré par le déplacement de l'eau et l'évolution de la température externe. On peut examiner ce processus sous plusieurs angles au sens géométrique du terme.

PROJECTION DE VARIABLES

La projection d'une variable est une opération de géométrie euclidienne justifiée dans l'annexe statistique. C'est ici une opération numérique très simple décrite dans la figure 4. On y considère une variable Z comptant 12 mesures réparties sur 3 stations et 4 dates. Une variable qualitative S associe à chaque valeur le numéro de la station correspondante (3 modalités); une variable qualitative T associe de même à chaque mesure le numéro de la date d'échantillonnage (4 modalités). On note z_i la $i^{\text{ème}}$ valeur de la variable Z , z est la moyenne des valeurs et $z_i - z$ est la valeur centrée pour la $i^{\text{ème}}$ observation. La variable centrée appelée X peut être considérée comme un tableau d'analyse de la variance à 2 facteurs sans répétition qui permet de calculer les moyennes respectives par dates et par stations : ce point de vue est illustré par le tableau I.

TABLEAU I. — 12 mesures forment une variable centrée (X) et un tableau d'analyse de variance à deux facteurs sans répétition. Calculs des tableaux de moyennes par station (X_1), de moyennes par date (X_2), des écarts observés ($X-X_1$ et $X-X_2$), du modèle linéaire additif (X_{12}) et des écarts correspondants ($X-X_{12}$). Ces opérations numériques sont écrites comme opérations géométriques dans la figure 4.

X						$X-X_2$							
Stations	Dates	1	2	3	4	moyennes		Stations	Dates	1	2	3	4
1		-35.5	-28.5	-26.5	-19.5	-27.5		1		-26	-28	-30	-26
2		+16.5	+23.5	+23.5	+28.5	+23		2		+26	+24	+20	+22
3		-9.5	+3.5	+13.5	+10.5	+4.5		3		0	+4	+10	+4
moyennes		-9.5	-0.5	+3.5	+6.5								
X_1						X_{12}							
Stations	Dates	1	2	3	4	Stations	Dates	1	2	3	4		
1		-27.5	-27.5	-27.5	-27.5	1		-37	-28	-24	-21		
2		+23	+23	+23	+23	2		+13.5	+22.5	+26.5	+29.5		
3		+4.5	+4.5	+4.5	+4.5	3		-5	+4	+8	+11		
X_2						$X-X_{12}$							
Stations	Dates	1	2	3	4	Stations	Dates	1	2	3	4		
1		-9.5	-0.5	+3.5	+6.5	1		+1.5	-0.5	-2.5	+1.5		
2		-9.5	-0.5	+3.5	+6.5	2		+3	+1	-3	-1		
3		-9.5	-0.5	+3.5	+6.5	3		-4.5	-0.5	+5.5	-0.5		
$X-X_1$													
Stations	Dates	1	2	3	4	Stations	Dates	1	2	3	4		
1		-8	-1	+1	+8	1		-8	-1	+1	+8		
2		-6.5	+0.5	+0.5	+5.5	2		-6.5	+0.5	+5.5	+0.5		
3		-14	-1	+9	+6	3		-14	-1	+9	+6		

On peut alors remplacer dans le tableau X chaque valeur par la moyenne de la station correspondante (tableau X_1). On obtient le tableau X_2 en faisant de même avec la moyenne des mesures par dates (tableau X_2). L'écart entre X et X_1 obtenu par les différences dans chaque case donne le

tableau $X-X_1$; la même définition donnant le tableau $X-X_2$. La somme des deux tableaux X_1 et X_2 donne X_{12} et l'écart à ce modèle additif donne $X-X_{12}$. Ces opérations numériques intervenant dans l'analyse de la variance de X est représentée dans la figure 4 comme une décomposition de vecteurs en composantes orthogonales. X devient un vecteur de l'espace vectoriel R^{12} dans lequel projeter la variable sur « le sous-espace station » consiste à remplacer l'observation centrée par la moyenne de la station correspondant à chaque mesure (variable X_1). Projeter la variable sur le sous-espace orthogonal au « sous-espace station » revient à remplacer l'observation centrée par la différence entre celle-ci et la moyenne de la station (variable $X-X_1$). De même on peut projeter sur le « sous-espace date » en utilisant la moyenne des observations de la même date (variable X_2), et sur son orthogonal en prenant les écarts (variable $X-X_2$). En remplaçant chaque valeur par la somme de la moyenne station et de la moyenne date (variable X_{12}) on fait une projection sur le sous-espace « station + date » et en prenant les écarts on effectue une projection sur le sous-espace orthogonal (variable $X-X_{12}$). Le théorème de Pythagore, qui justifie les équations de l'analyse de variance permet de décomposer cette dernière de plusieurs façons consignées dans la figure 4.

Si Z , au lieu d'être une variable unique, est maintenant un tableau comportant n lignes (relevés) et p colonnes (variables), les opérations précédentes peuvent porter sur chacune des colonnes de ce tableau et les projections s'appliquent au nuage des variables. L'Analyse en Composantes Principales de ce nuage projeté de variables renvoie, dans le contexte du schéma de dualité, à une analyse d'un nouveau nuage des individus, ce qui permet de faire une typologie de relevés avec intégration volontaire d'un point de vue. On peut alors décider de :

- ne tenir compte que des différences entre stations (analyse de X_1);
- ne pas tenir compte des différences entre stations (analyse de $Y_1 = X-X_1$);
- ne tenir compte que des différences entre dates (analyse de X_2);
- ne pas tenir compte des différences entre dates (analyse de $Y_2 = X-X_2$);
- tenir compte des effets additifs du temps et de l'espace (analyse de X_{12});
- éliminer ces effets et ne voir qu'une variabilité résiduelle (analyse de $Y_{12} = X-X_{12}$);

chacun des tableaux X_1 , X_2 , Y_1 , Y_2 , X_{12} , Y_{12} étant obtenu en reproduisant sur chaque variable indépendamment l'opération précédemment décrite. Des compléments d'information sont données dans l'annexe statistique.

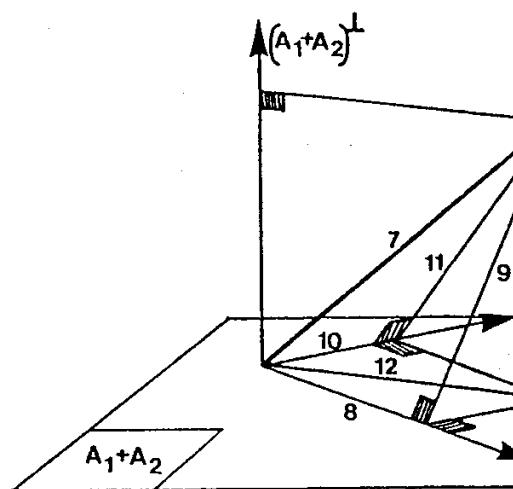
RÉSULTATS

La figure 5 résume l'ACP du tableau X_{12} sur l'exemple traité. Par définition, sur la carte factorielle des individus un prélèvement est positionné par la somme d'une composante numérique date et d'une composante numérique station. Les stations sont alors représentées par une évolution temporelle commune et l'analyse donne une lecture simplifiée des structures espace-temps vue par un modèle additif. On l'utilisera dès que la lecture en terme d'espace-temps des cartes factorielles de l'ACP classique apparaît confuse. Elle fournit rapidement une idée du rôle respectif de l'espace ou du temps, éventuellement de leur intervention simultanée comme c'est ici le cas.

FIG. 4. — Décomposition de la variance dans un plan complet à deux facteurs sans répétition et approche géométrique comme projection d'un vecteur sur des sous-espaces. Dans le tableau on trouve : 1 : numéro d'ordre du relevé, 2 : données brutes, 3 et 4 : variables expérimentales (stations, dates), 5 et 6 : tableaux disjonctifs associés aux variables expérimentales, 7 : variable centrée, 8 à 13 : décomposition par moyennes et écarts aux moyennes de la variable centrée. En bas les numéros désignent la représentation géométrique du vecteur dont les composantes figurent dans la colonne correspondante. A_1 , A_2 , $A_1 + A_2$ et leurs orthogonaux sont des sous-espaces de R^{12} .

On notera que la dimension de l'espace $A_1 + A_2$ ($s + t - 2 = 8$) est inférieure au nombre de variables et qu'on trouve ainsi 2 valeurs propres nulles.

DONNEES				PROJECTIONS								
1	2	3	4	5	6	7	8	9	10	11	12	13
	Z	S	T	A ₁	A ₂	X	X ₁	X-X ₁	X ₂	X-X ₂	X ₁₂	X-X ₁₂
1	21	1	1	100	1000	-35.5	-27.5	-8	-9.5	-26	-37	+1.5
2	28	1	2	100	0100	-28.5	-27.5	-1	-0.5	-28	-28	-0.5
3	30	1	3	100	0010	-26.5	-27.5	+1	+3.5	-30	-24	-2.5
4	37	1	4	100	0001	-19.5	-27.5	+8	+6.5	-26	-21	+1.5
5	73	2	1	010	1000	+16.5	+23	-6.5	-9.5	+26	+13.5	-3
6	80	2	2	010	0100	+23.5	+23	+0.5	-0.5	+24	+22.5	+1
7	80	2	3	010	0010	+23.5	+23	+0.5	+3.5	+20	+26.5	-3
8	85	2	4	010	0001	+28.5	+23	+5.5	+6.5	+22	+29.5	-1
9	47	3	1	001	1000	-9.5	+4.5	-14	-9.5	+0	-5	-4.5
10	60	3	2	001	0100	+3.5	+4.5	-1	-0.5	+4	+4	-0.5
11	70	3	3	001	0010	+13.5	+4.5	+9	+3.5	+10	+8	+5.5
12	67	3	4	001	0001	+10.5	+4.5	+6	+6.5	+4	+11	-0.5
MOY 56.5				0	0	0	0	0	0	0	0	0
VAR 478.25				478.25	435.17	43.08	36.25	442.0	471.42	6.83		
%				100	91	9	8	92	99	1		
Décomposition spatiale				91	9							
Décomposition temporelle							8	92				
Décomposition spatio-temporelle				91			8					1
Décomposition du modèle additif										99	1	



D'après BARTLETT (1934)
in SCHEFFE (1959 p. 43)

FIG. 4.

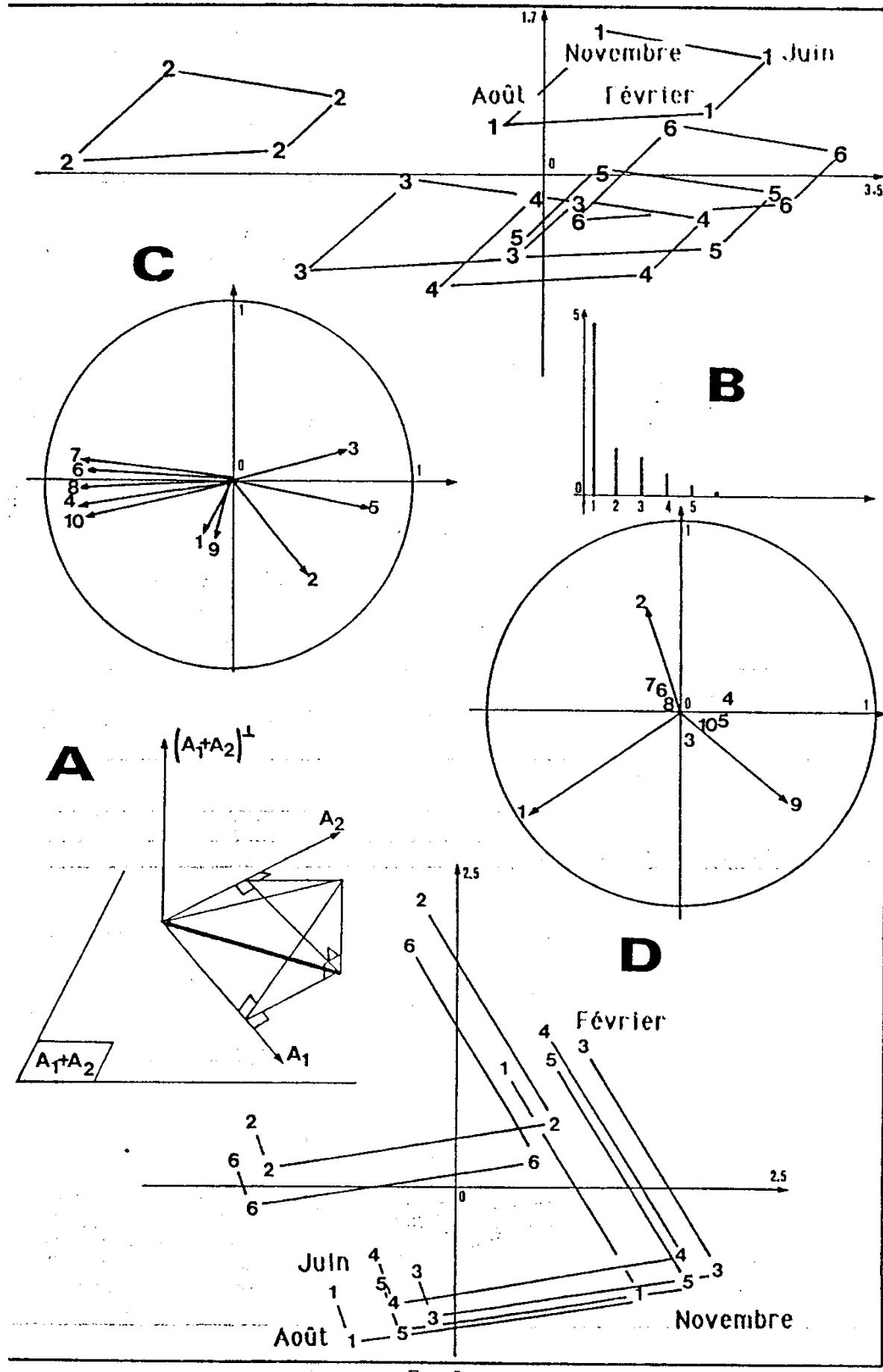


FIG. 5.

FIG. 5. — *ACP* du tableau X_{12} des valeurs $x_{ik} + x_{ik}$ prévues par le modèle linéaire. *A* : position géométrique des variables analysées (cf. fig. 4). *B* : Graphe des valeurs propres. *C* : Plan factoriel (individus) et carte factorielle des variables dans le sous-espace F_1-F_3 . *D* : Plan factoriel (individus) et carte factorielle des variables dans le sous-espace F_2-F_4 . L'ensemble donne une relecture simplifiée des structures spatio-temporelles par un modèle additif.

Le plan 1-3 est celui d'une typologie stationnelle correspondant à la pollution (station 2), à l'eutrophisation décroissante (stations 3, 4, 5) et au niveau témoin (station 6 et 1 avec un débit plus faible pour la dernière). Le plan 2-4 décrit un rythme saisonnier incluant le débit (étiage en novembre), la température et les

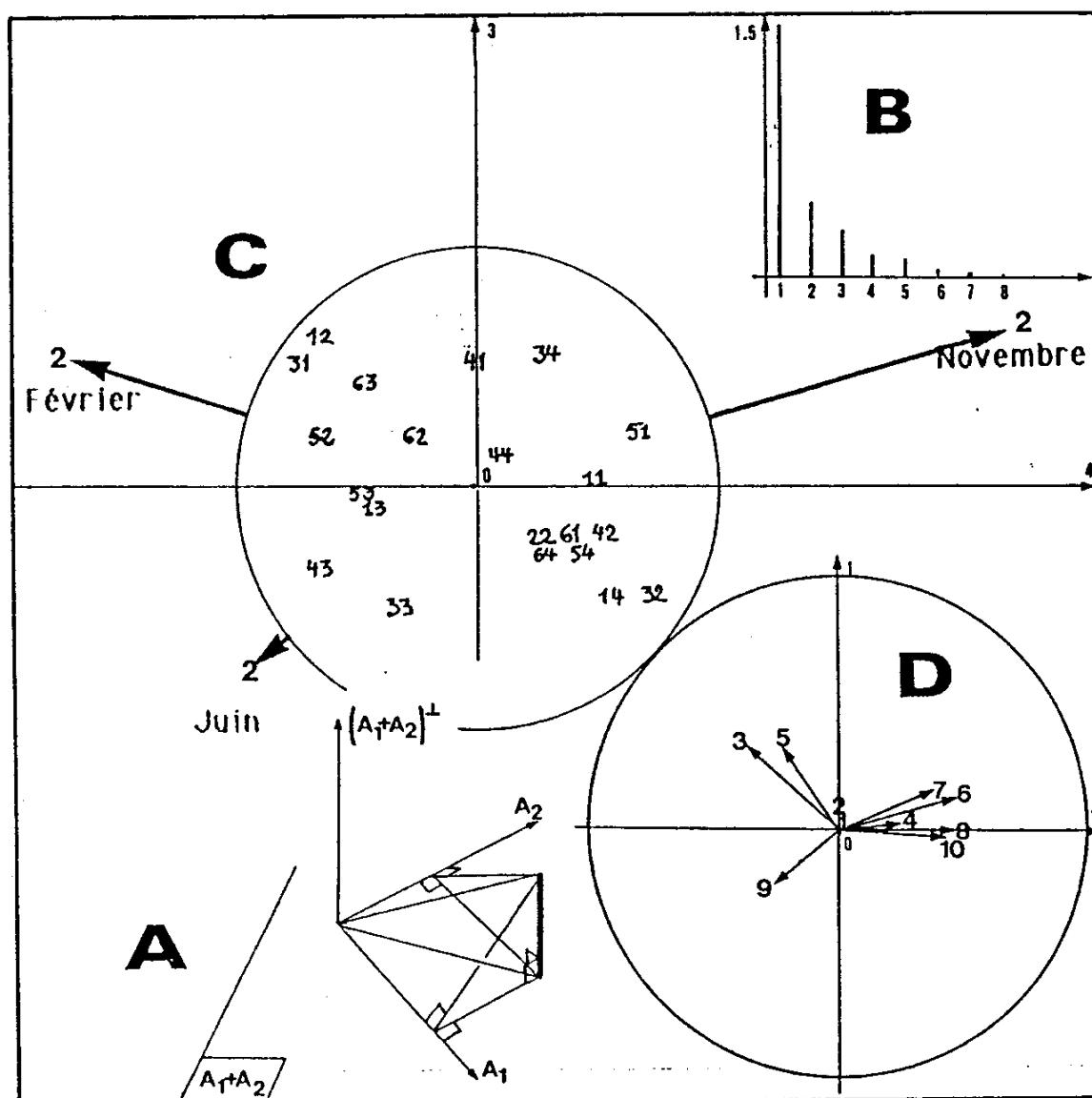


FIG. 6. — *ACP* du tableau Y_{12} des écarts au modèle linéaire. *A* : Position géométrique des variables analysées. *B* : Graphe des valeurs propres. *C* : Plan factoriel F_1-F_2 . *D* : Carte factorielle des variables F_1-F_2 . L'écart au modèle linéaire concerne la station 2.

Vol. 8, n° 3 - 1987

nitrates (maximum en novembre). Sur le troisième critère intervient une partition des stations explicite sur la figure 5; les quantités de nitrates sont pour, des raisons différentes, faibles en novembre aux stations 2 et 6.

Ce qui, dans les données, n'est pas pris en compte par un modèle additif est clairement exprimé dans l'ACP du tableau Y_{12} (fig. 6) consacrée à la mise en évidence du déphasage entre la source de pollution (station 2) et le processus d'oxydation de l'azote ammoniacal qui s'ensuit dans les stations aval. La quantité de nitrates diminue en février dans toutes les stations sauf la station 2, où la quantité d'oxygène n'est pas suffisante pour permettre cette oxydation.

La typologie stationnelle est très simplifiée à la figure 7 où est résumée l'ACP du tableau X_1 . Il y a alors $t-1$ soit trois valeurs propres non nulles. Le plan 1-2 sépare

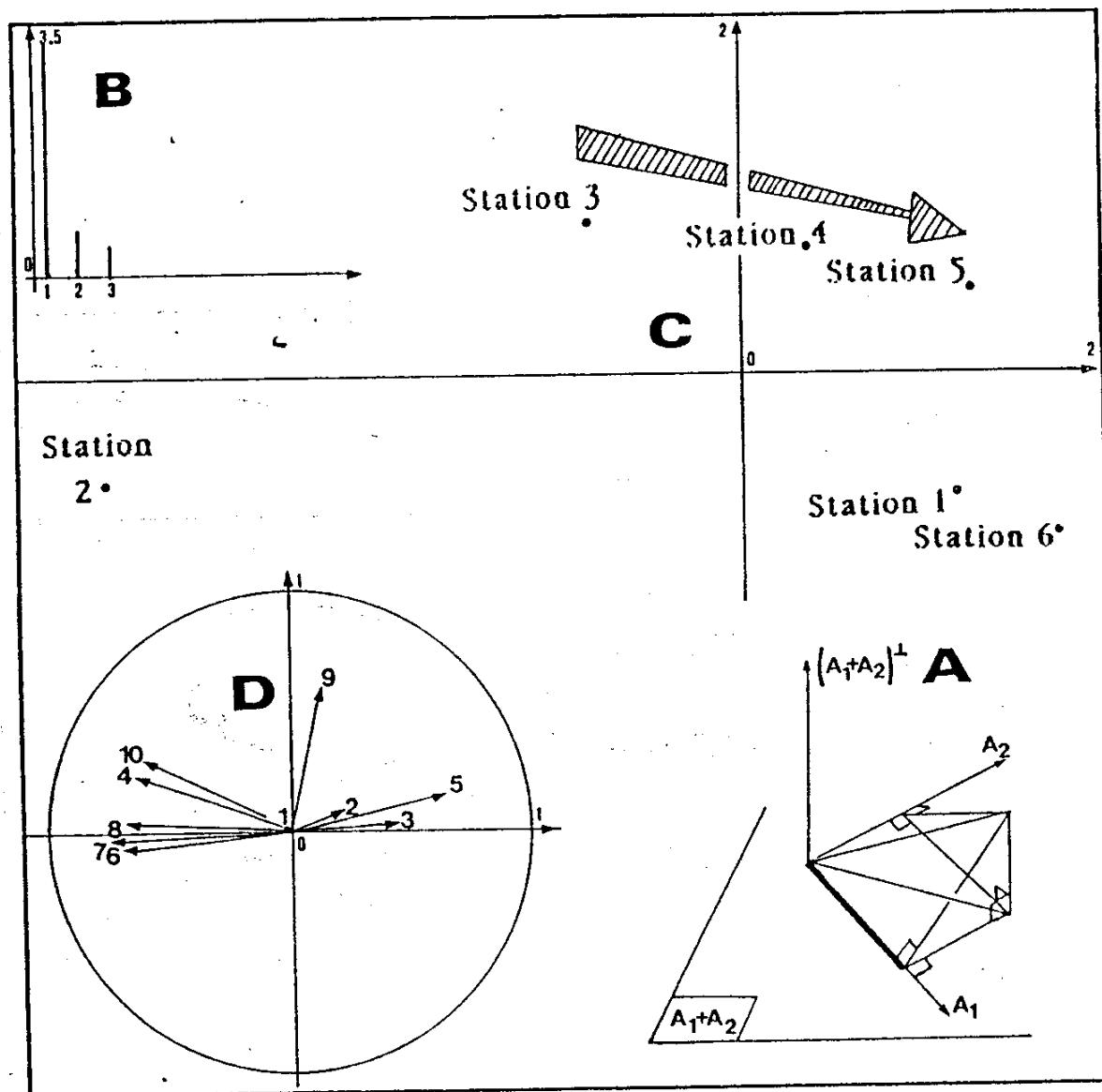


FIG. 7. — ACP du tableau X_1 (moyennes par stations). **A** : Position géométrique de l'analyse. **B** : Graphe des valeurs propres (le tableau est de rang 5 et deux valeurs propres sont très faibles). **C** : Carte factorielle F_1-F_2 (station 2 : pollution, stations 3, 4 et 5 : restauration, stations 1 et 6: témoins).

le point de pollution (station 2) et les stations témoins (1 à 6). Outre les paramètres indicateurs des teneurs en matière organique le processus d'auto-épuration concerne l'évolution amont-aval des nitrates dont le niveau moyen augmente avec la restauration du cours d'eau (stations 3 à 5).

La station 2 présente de plus faibles valeurs en tant que secteur à charge polluante, les stations 1 et 6 en temps que secteurs propres exempts de pollution.

L'aspect temporel est remarquablement analysé par l'ACP du tableau Y_1 (tableau centré par stations) et deux types de rythmes saisonniers sont mis en évidence (fig. 8). Le premier concerne l'évolution du déséquilibre chimique afférant aux éléments azotés qui touche également les quantités de nitrites non prises en compte ici (PEGAZ-MAUCET, 1980). Le second est commandé par la température

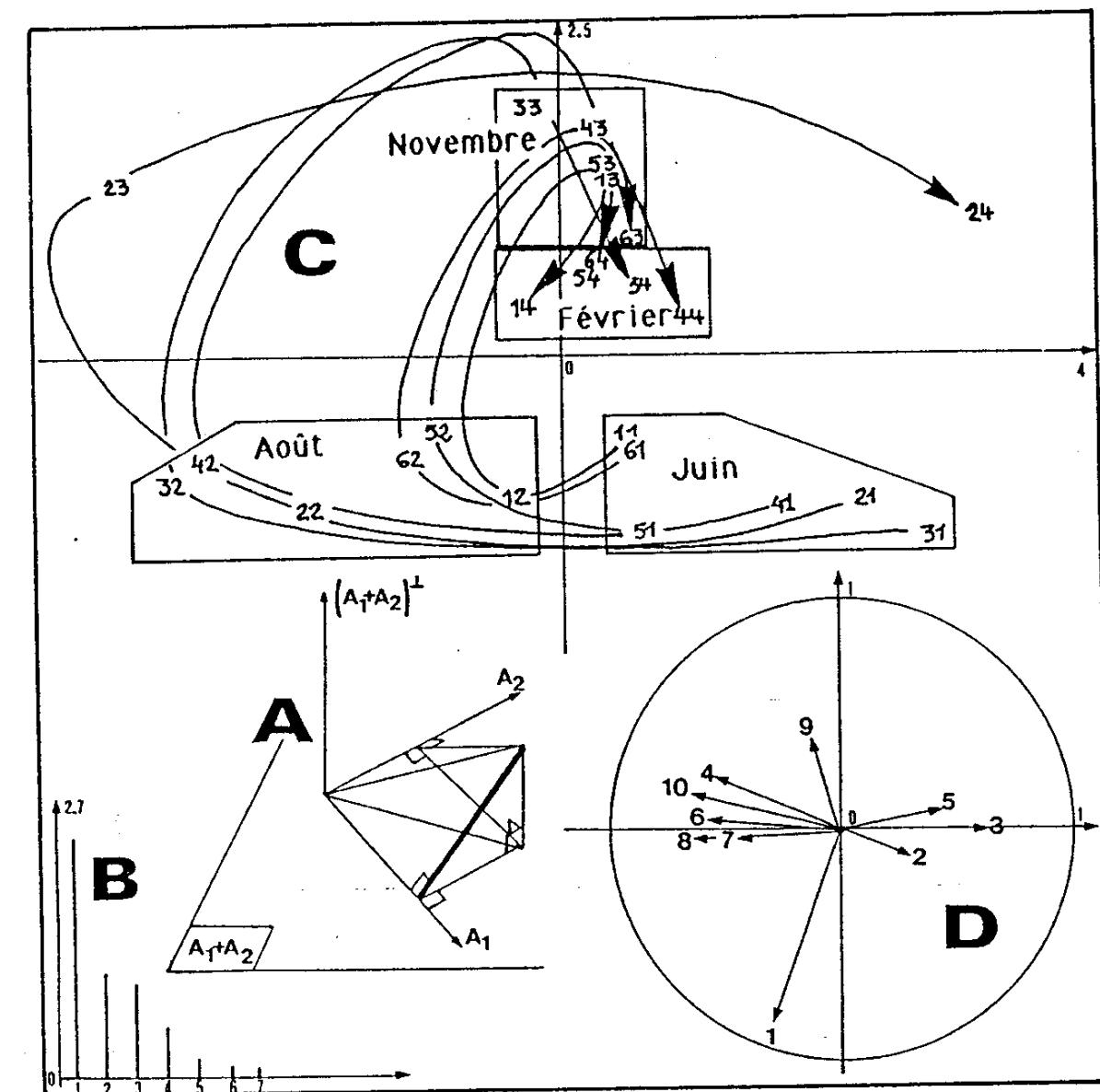


FIG. 8. — ACP du tableau Y_1 (écart aux moyennes par stations). **A** : Point de vue géométrique. **B** : Valeurs propres. **C** : Plan factoriel F_1-F_2 . **D** : Carte factorielle des variables (F_1-F_2). Cette analyse résume l'aspect temporel du phénomène après élimination de la typologie stationnelle. On peut lui préférer la figure 11.

et sépare saison chaude (juin et août) et froide (novembre et février). Les valeurs anormales de la station 2 auxquelles est consacrée la figure 6 sont à nouveau mises en avant, cette analyse tenant compte de l'effet date et de l'interaction.

Une image très simplifiée de la typologie saisonnière est obtenue par l'analyse du tableau X_2 centré par dates (fig. 9). Les mêmes phénomènes sont bien sûr en cause

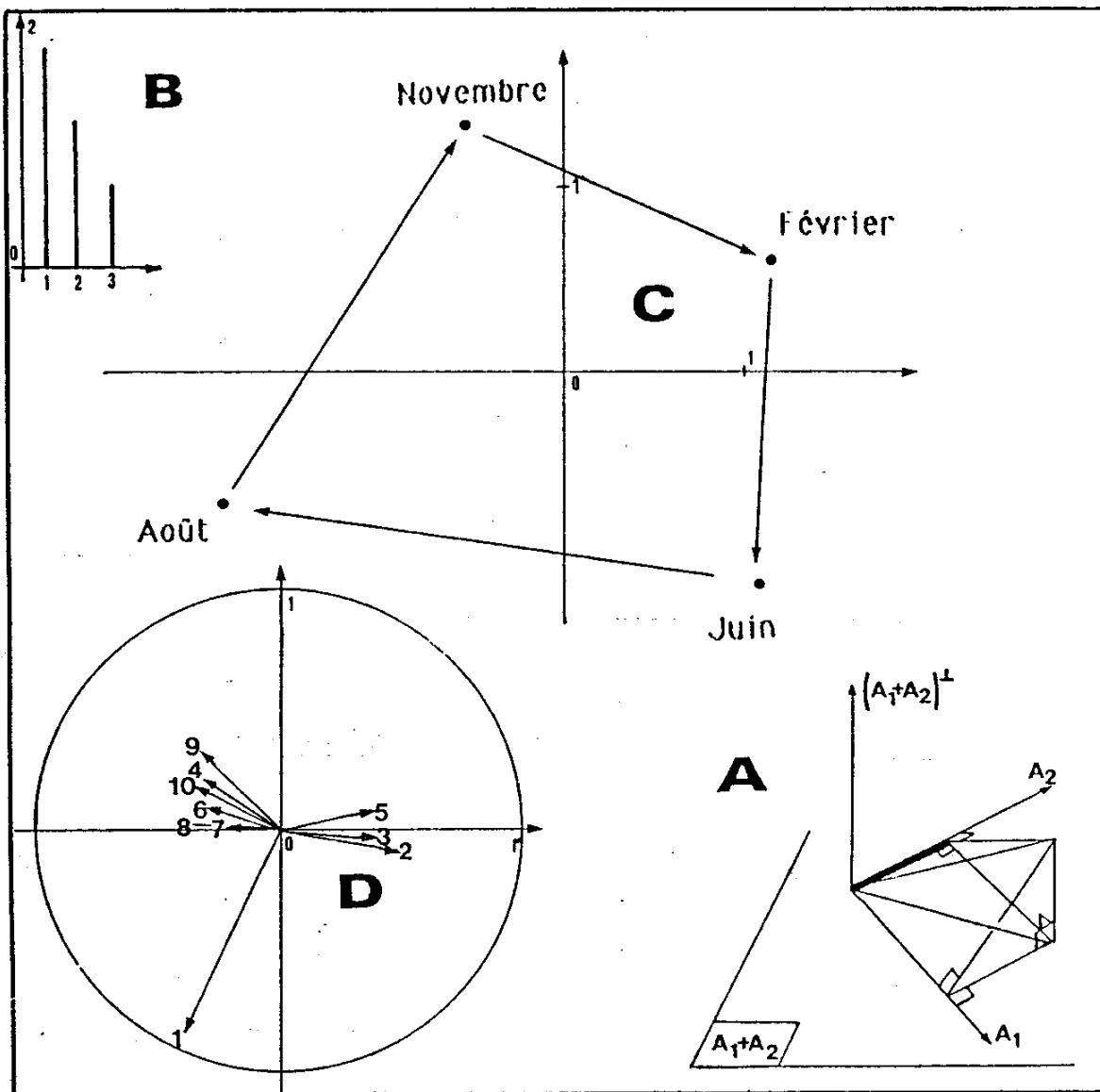


FIG. 9. — ACP du tableau X_2 (moyennes par dates). Cette analyse donne une image très simplifiée du rythme saisonnier qui peut être utile. On préférera ici la lecture de l'analyse du tableau Y_1 (fig. 8).

mais on notera qu'on obtient une meilleure approche des éléments temporels en étudiant les écarts aux moyennes par stations (tableau $X-X_1$) plutôt que les moyennes par dates : on conserve ainsi une image de la variabilité interne à la catégorie étudiée qui renforce la définition des positions relatives des éléments de cette catégorie (ici les dates).

De manière parfaitement symétrique, ce qui est conforme à la nature du calcul, l'image la plus précise de la typologie stationnelle est peut-être obtenue par l'ACP du tableau Y_2 des écarts aux moyennes par dates. Tous les éléments précités sont présents (pollution, restauration, témoins) et bien que le temps soit en principe éliminé (par centrage) on obtient de cette manière une typologie des stations qui est

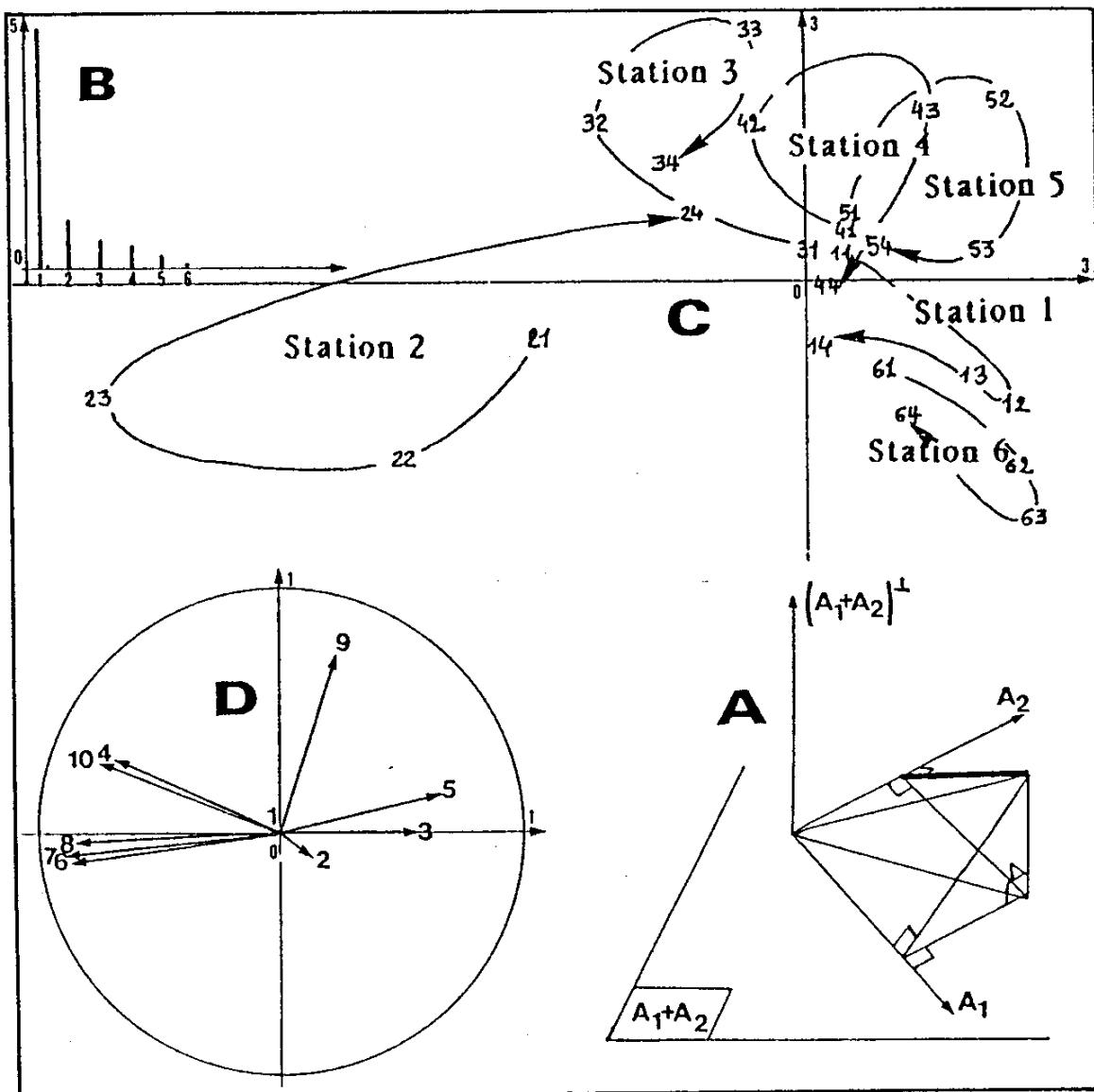


FIG. 10. — ACP du tableau Y_2 (écart aux moyennes par dates). Le plan factoriel est à comparer avec les figures 3 A, 5 C, 7 C. Cette analyse fait la meilleure description de la typologie interstationnelle en éliminant les éléments saisonniers communs; elle conserve cependant le fonctionnement spécifique de chaque station (déphasage).

une typologie de fonctionnement conforme aux objectifs actuels de l'écologie factorielle. Retenons qu'en terme de géométrie euclidienne la projection sur l'orthogonal du « sous-espace temps » l'emporte sur la projection sur le « sous-espace station », comme la projection sur l'orthogonal du « sous-espace station » l'emporte sur la projection sur le « sous-espace temps ».

DÉCOMPOSITION DE LA VARIANCE OU NORMALISATION CONDITIONNELLE

L'*ACP* retenue (analyse générale de LEBART *et al.*, 1982, p. 275) sur les tableaux des variables projetées présente, quant aux aides à l'interprétation, un avantage et un inconvénient. L'avantage réside dans la décomposition de la variance associée au théorème de Pythagore. Une variable normalisée Z est un vecteur unitaire de R^n . Géométriquement sa longueur (ou son carré) égale 1, statistiquement la variance de ses composantes égale l'unité. Le vecteur projeté sur un sous-espace a pour norme (carré de sa longueur) la variance de ses composantes si on se trouve dans l'orthogonal du sous-espace engendré par 1_n (centrage). Le rapport de la seconde variance sur la première est le pourcentage de variance extrait par la projection. La procédure choisie contient deux projections successives. La première est le point de vue de l'analyse, la seconde est une réduction de dimensions propre au modèle euclidien de l'*ACP*. On peut donc connaître dans chaque cas, en partant d'une variance unité, la quantité prise en compte par l'analyse choisie, puis la proportion de cette quantité prise en compte par la projection sur les cartes factorielles.

Dans un premier temps on pourra donc étudier plusieurs décompositions de la variance entre analyses, ce qui revient à l'usage descriptif des analyses de variances variable par variable. Chacune de ces décompositions est associée à un ensemble de projections sur des espaces orthogonaux, donc aux ensembles respectifs des tableaux X_1 et Y_1 , X_2 et Y_2 , X_{12} et Y_{12} , X_1 , X_2 et Y_{12} . Dans l'exemple traité on trouve, pour le dernier cas cité, les valeurs consignées au tableau II.

TABLEAU II. — *Exemple de décomposition de la variance de chaque variable normalisée (variance totale unité) entre trois sous-espaces formant une somme directe orthogonale. La représentation sur un graphique triangulaire résume l'importance des différents points de vue. L'analyse de variance univariée en jeu sur chaque ligne du tableau pourrait faire l'objet d'un test d'hypothèse en cas, au demeurant rare, de nécessité.*

Variables	Stations analyse X_1	Dates analyse X_2	Interaction analyse Y_{12}	total
1-Température	1	97	2	100
2-Débit	38	51	11	100
3-pH	25	40	35	100
4-Conductivité	55	31	14	100
5-Oxygène	52	18	30	100
6-DBO ₅	53	12	35	100
7-Oxydabilité	68	6	26	100
8-Ammoniaque	55	14	31	100
9-Nitrates	44	29	27	100
10-Phosphates	50	21	29	100
Total	441	319	240	1000

On pourra ensuite étudier plusieurs décompositions de la variance projetée propres à chacune des analyses. Rappelons que cette variance projetée est simplement le carré de la coordonnée factorielle correspondante (cf. LEGENDRE & LEGENDRE, 1979, p. 109). On obtient donc facilement une interprétation numérique de la prise

en compte de chaque variable par une analyse donnée. Par exemple le tableau III restitue cette approche dans l'analyse Y_{12} .

TABLEAU III. — *Exemple de décomposition de la variance de chaque variable normalisée (variance initiale unité) dans une ACP à nuage projeté. Toutes les valeurs sont multipliées par 100. On peut lire pour le caractère 9 que la projection choisie extrait 72 % de la variance parmi lesquels 61 % sont à nouveau extraits dans la projection sur le deuxième axe principal. Ces quantités ne sont que des indications numériques utiles pour le dépouillement.*

Variables	Variance initiale	Variance projetée	Axe 1	Axe2	Résiduelle
1-Température	100	3	0	0	3
2-Débit	100	49	4	1	44
3-pH	100	60	37	0	23
4-Conductivité	100	69	52	11	6
5-Oxygène	100	82	51	3	28
6-DBO ₅	100	88	81	1	6
7-Oxydabilité	100	94	86	1	7
8-Ammoniaque	100	86	83	0	3
9-Nitrates	100	72	8	61	3
10-Phosphates	100	79	64	10	15

On y observe que plus de 80 % de la variance des caractères DBO_5 , oxydabilité et ammoniaque (pollution) est utilisée dans la figure 10 mais que le dépouillement de cette analyse est resté inachevé.

Notons enfin que la combinaison des deux types de décomposition est légitime car la projection dans le sous-espace des axes principaux d'un caractère projeté est elle-même une projection de ce caractère (théorèmes des 3 perpendiculaires). Le rôle des variables dans la description des composantes stationnelles et des rythmes saisonniers peut être ainsi approché avec précision.

L'inconvénient précité est associé à cet avantage : en respectant la décomposition de la variance (aides numériques à l'interprétation), les coordonnées des variables représentées sur les différentes figures qui précèdent ne sont pas des coefficients de corrélation entre facteurs (coordonnées lignes) et caractères mais simplement des covariances. La représentation d'un cercle (conforme à la pratique des cercles de corrélation) y est donc abusive. Or une faible part de variance expliquée, traduite par des points-variables peu éloignés de l'origine par rapport au cercle unité, pourrait correspondre dans certains cas à des faits biologiques importants : c'est le cas lorsqu'on cherche à préciser des différences subtiles entre stations vues à travers des variations saisonnières très fortes. Pour retrouver des cercles de corrélation dans les cartes factorielles des variables (aides graphiques à l'interprétation), on pourrait exécuter une ACP normée sur les tableaux projetés. L'expérience montre qu'il vaut mieux utiliser une normalisation conditionnelle.

En effet, le problème se pose essentiellement pour les analyses du type Y_1 (écart aux moyennes par station) et Y_2 (écart aux moyennes par date). On peut calculer dans ce cas deux nouveaux tableaux T_1 et T_2 en substituant à x_{ijk} les quantités

$$(x_{ijk} - x_{i,k})/s_{i,k} \text{ (écart normalisé à la moyenne pour la station } i\text{)}$$

et $(x_{ijk} - x_{.jk})/s_{.jk} \text{ (écart normalisé à la moyenne pour la date } j\text{)}$

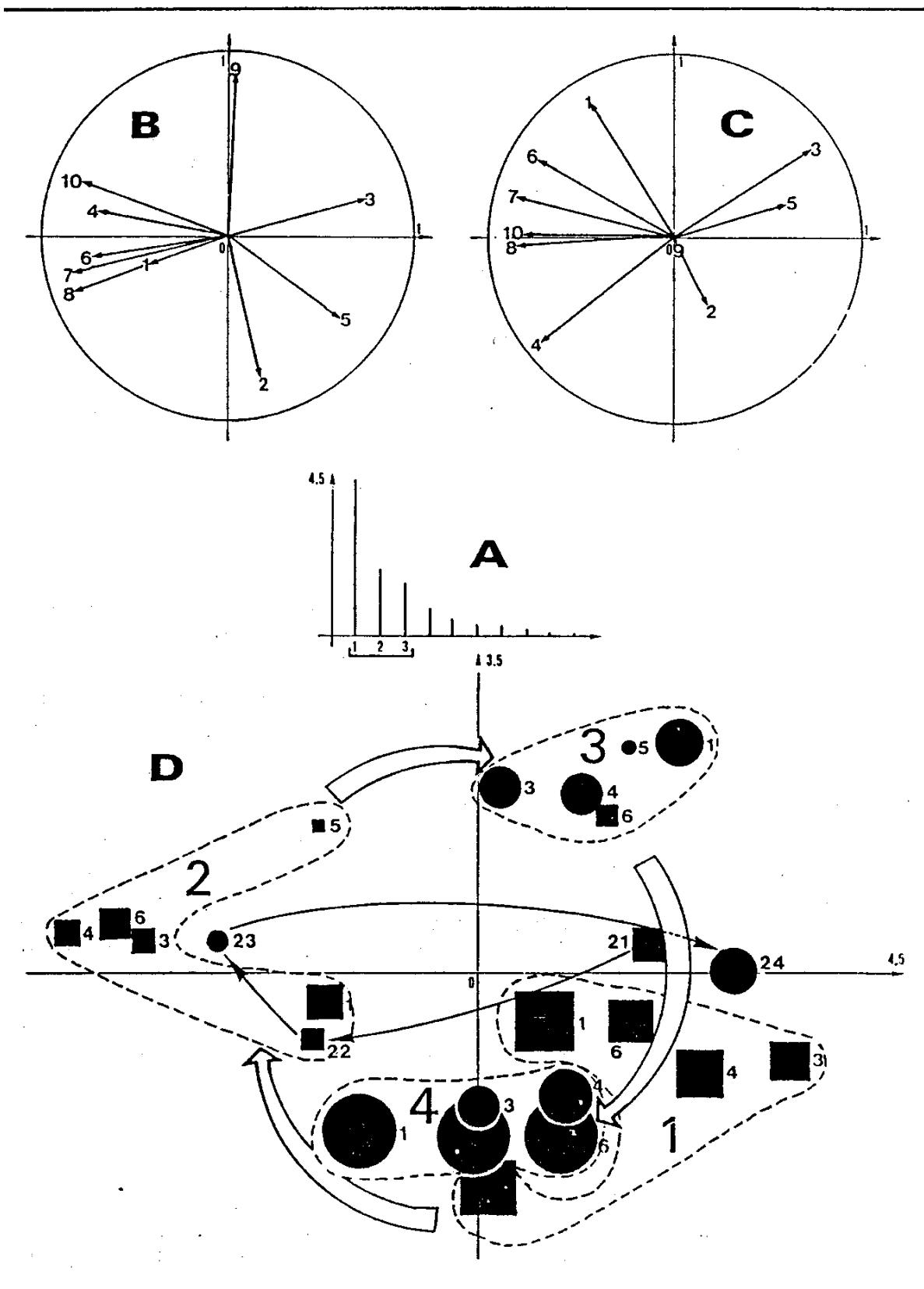


FIG. 11.

Acta Ecologica/Ecologia Generalis

où les quantités $s_{i,k}$ et s_{jk} sont les écarts types de la variable k respectivement calculés sur les valeurs prises à la station i ou à la date j . « Cette standardisation par date élimine la tendance évolutive des moyennes et des variances des caractères au cours du temps » indique PERNIN (1986, p. 163) qui utilise ce tableau (méthode LONGI, PONTIER et PERNIN, 1986) pour une analyse canonique entre le sous-espace qu'il engendre et les sous-espaces A_1 et A_2 définis plus haut, ceci dans la perspective décrite par PAGES et TOMASSONE (1979).

Pour une analyse plus spécifique visant à éliminer l'effet-station, respectivement l'effet-date, on exécutera une *ACP* (analyse générale), éventuellement avec un programme d'*ACP* centrée, du tableau T_1 respectivement T_2 . La figure 11 illustre l'analyse du tableau T_1 de l'exemple traité. On obtient un résumé efficace du processus temporel multivarié construit sur les trois composantes auto-épuration, débit et température qui impose une lecture à trois dimensions géométriques. Cette lecture est facilitée par la représentation graphique des valeurs du facteur 3 dans le plan factoriel 1-2 (cf. BENZECRI *et al.*, 1973, p. 248).

L'attention est attirée sur le rôle du débit dans le phénomène d'auto-épuration par le biais de l'oxygénation (proche de la sursaturation en février et juin) et de la dilution de la charge ionique. Le déphasage de la station 2, déjà mentionné, est également perçu avec précision.

CONCLUSION

L'intérêt principal des indications méthodologiques faites dans cette note réside dans l'interaction entre la facilité de mise en œuvre d'une part et le cadre de présentation utilisé d'autre part. Une partie importante des mesures physico-chimiques en milieu aquatique s'inscrit dans un plan d'observations espace-temps : nous avons montré qu'avec un programme ordinaire d'*ACP* et quelques manipulations préliminaires simples (moyennes, variances, centrage ou normalisation) on peut aisément agrandir les possibilités d'examen de telles mesures.

Nous avons choisi d'insérer ces pratiques dans le modèle de la géométrie euclidienne, ce qui n'est pas indispensable à leur usage, pour souligner qu'elles sont un cas particulier de l'analyse en composantes principales sur variables instrumentales de RAO (1964), de l'analyse des nuages projetés au sens de LAURO et D'AMBRA (1983), de l'analyse des covariances partielles de LEBART *et al.* (1982, p. 300), de l'analyse en composantes explicatives d'OBADIA (1978), de l'analyse des redondances de WOLLENBERG (1977) et des analyses sous contraintes linéaires de SCHEKTMAN (1978). SABATIER (1983) a établi les relations entre ces diverses approches.

FIG. 11. — *ACP* du tableau T_1 (écart normalisés aux moyennes par stations). **A** : Valeurs propres (on utilise 3 facteurs). **B** : Cercle des corrélations (carte factorielle des variables) F_1 - F_2 . **C** : Cercle des corrélations F_1 - F_3 . **D** : Carte factorielle F_1 - F_2 et valeurs du troisième facteur (carrés : valeurs positives; cercles : valeurs négatives; la surface du symbole est proportionnelle à la valeur absolue; indications de Y. AUDA, comm. pers.), 1 : juin, 2 : août, 3 : novembre, 4 : février. Les petits chiffres désignent la station (1 à 6). Cette analyse n'est pas l'*ACP* d'un nuage projeté mais ses cartes factorielles variables sont des cercles de corrélation comme en *ACP* normée (fig. 3).

Ces points de vue intègrent tous dans l'approche des données ce qui reste implicite dans la pratique courante, à savoir l'information complémentaire que l'expérience associe aux marges du tableau examiné, c'est-à-dire, dans le cas présenté, où et quand est faite la mesure. On peut choisir de ne voir dans les cartes factorielles que ce qui en dépend : on impose alors des contraintes à la structure de l'image produite. Ceci facilite et nuance la lecture des produits de l'analyse multivariée. L'analyse Y_1 est une analyse des covariances partielles en ce sens qu'elle ne contient que de l'information commune aux stations (c'est-à-dire l'approche globale du rythme saisonnier). Mais le fait que la variable dont on veut éliminer l'effet ne soit qu'une indication qualitative (numéro de la station) rend transparente l'opération numérique qui produit une image si différente des cartes de l'analyse sans contrainte. Le cas particulier qui nous intéresse ici n'est particulier que dans sa mise en œuvre. Si on désire ne voir dans un tableau faunistique que la structure indépendante des conditions météorologiques de piégeage des insectes, si ces conditions sont appréciées par un tableau de variables quantitatives, il conviendra d'utiliser un logiciel plus complexe, mais le raisonnement sera le même.

De même l'analyse X_{12} est une analyse en composantes explicatives du tableau physico-chimique sur le tableau formé par les deux variables qualitatives S et T . La reconstitution de la valeur d'un paramètre par une combinaison espace-temps est ici particulière en ce sens que le plan est complet et l'approximation est une formation simple des moyennes. Quand on cherchera à reconstituer l'évolution de la composition chimique des eaux par l'ensemble des débits des affluents il faudra à nouveau utiliser un programme spécifique mais un principe identique.

L'écart qui s'est creusé entre les propositions des théoriciens de l'analyse des données et la part effectivement utilisée par les expérimentateurs est devenu considérable. La présente note est un premier jalon pour inviter les écologues à utiliser des moyens d'analyse encore peu diffusés. Nous aborderons dans ce but l'examen des plans incomplets : si la procédure est sensiblement modifiée, l'utilisation effective ne présente pas de difficultés supplémentaires. L'emploi des composantes explicatives devrait faciliter l'examen d'un couple de deux tableaux (explication d'un ensemble de variables de milieu par un autre, influence des variables de milieu sur la composition faunistique). L'évaluation du rôle des méthodes multitableaux enfin fait partie de cette démarche.

ANNEXE STATISTIQUE : PROJECTION DE VARIABLES

Considérons une variable Z enregistrée dans s stations à t dates d'échantillonnage. On notera z_{ij} sa valeur à la station i et à la date j et m_z sa moyenne. Soit x_{ij} la quantité $z_{ij} - m_z$, valeur de la variable centrée correspondante. Notons X cette variable. Les valeurs x_{ij} forment un vecteur de R^n ($n = st$) considéré comme espace euclidien muni de la métrique des poids

$$D_n = (1/n) I_n$$

où I_n est la matrice identité. X est alors perpendiculaire au vecteur I_n dont les composantes sont toutes égales à 1.

Les variables S et T indiquant le numéro des stations et des dates se décomposent respectivement en tableaux disjonctifs complets par l'intermédiaire des indicatrices des

modalités. Les colonnes de ces tableaux engendrent deux sous-espaces vectoriels B_1 et B_2 dont l'intersection contient le vecteur 1_n . La décomposition en sous-espaces orthogonaux (CAILLEZ et PAGES, 1976, p. 359 ou BENZECRI, 1973, p. 180) s'écrit en général :

$$R^n = (B_1 \cap B_2) \oplus^\perp (B_1 \cap B_2^\perp) \oplus^\perp (B_1^\perp \cap B_2) \oplus^\perp (B_{11} \oplus B_{22}) \oplus^\perp (B_1 + B_2)^\perp$$

Dans le cas particulier qui nous intéresse $B_1 \cap B_2$ se réduit au sous-espace engendré par 1_n et $B_{11} \oplus B_{22}$ au vecteur nul. Si on note A_1 l'espace $B_1 \cap B_2^\perp$ et A_2 l'espace $B_1^\perp \cap B_2$ on a :

$$R^n = [1_n] \oplus^\perp A_1 \oplus^\perp A_2 \oplus^\perp (A_1 + A_2)^\perp$$

Les dimensions des espaces vectoriels en jeu étant :

$$st = 1 + (s - 1) + (t - 1) + (s - 1)(t - 1)$$

La décomposition numérique associée s'écrit :

$$z_{ij} = m_z + x_i + x_j + (x_{ij} - x_i - x_j)$$

où x_i est la moyenne par date (pour la station i) et x_j la moyenne par station (pour la date j) de la variable X . On projette ainsi respectivement sur A_1 (fig. 4, col. 8), sur A_1^\perp (col. 9), sur A_2 et A_2^\perp (col. 10 et 11), sur $A_1 + A_2$ et $(A_1 + A_2)^\perp$ (col. 12 et 13). Les décompositions traditionnelles de la variance découlent du théorème de Pythagore.

Dans le cas multivarié où z_{ijk} désigne la $k^{\text{ème}}$ colonne (variable) du tableau ($1 \leq k \leq p$) on normalise au préalable les données pour éliminer les différences d'unités de mesures. Soit alors x_{ijk} la valeur centrée (moyenne 0) et réduite (variance 1) de la variable k à la date i et à la station j , $x_{i,k}$ et $x_{.jk}$ les moyennes respectives par dates et par stations des variables normalisées. On peut constituer successivement 6 nouveaux tableaux de mêmes dimensions (n lignes et p colonnes) en substituant respectivement à x_{ijk} les quantités :

- tableau X_1 : $x_{i,k}$ (moyenne pour la station)
- tableau Y_1 : $x_{ijk} - x_{i,k}$ (écart à la moyenne pour la station)
- tableau X_2 : $x_{.jk}$ (moyenne pour la date)
- tableau Y_2 : $x_{ijk} - x_{.jk}$ (écart à la moyenne pour la date)
- tableau X_{12} : $x_{i,k} + x_{.jk}$ (prévision linéaire additive)
- tableau Y_{12} : $x_{ijk} - x_{i,k} - x_{.jk}$ (écart au modèle linéaire)

Chacun de ces tableaux est soumis à une ACP du type analyse générale de LEBART *et al.* (1982, p. 275) qu'on peut exécuter avec un programme d' ACP centrée qui renouvelera un centrage inutile. Dans R^n chacune des analyses exécute, à l'intérieur du sous-espace de référence qui la détermine, une projection sur le sous-espace des composantes principales. Elle se définit donc comme une projection de variables normalisées composée d'une projection imposée (point de vue de l'analyse) suivie d'une projection optimale (propriété de l' ACP).

La procédure complète est donc très simple. Elle comporte une normalisation initiale, la constitution de six tableaux dérivés et l' ACP d'un ou plusieurs d'entre eux.

RÉFÉRENCES CITÉES

- AMOROS C. & CHESEL D., 1985. — Les peuplements de Cladocères (Crustacés) descripteurs de fonctionnement hydrologique des bras-morts fluviaux. *Annls Limnol.*, 21, 3, 227-240.
- ANGELIER E., BORDES J. M., LUCHETTA J. C. & ROCHARD M., 1978. — Analyse statistique des paramètres physico-chimiques de la rivière Lot. *Annls Limnol.*, 14, 1-2, 39-57.
- AUDA Y., 1983. — *Rôle des méthodes graphiques en analyses de données : application au dépouillement des enquêtes écologiques*. Thèse de 3^e cycle. Université de Lyon I, 127 p.
- BARTLETT M. S., 1934. — The vector representation of a sample. *Proc. Cambridge Philos. Soc.*, 30, 327-340.
- BENZECRI J. P., 1973. — *Analyse des données. 2. L'analyse des correspondances*. Bordas, Paris, 620 p.
- BOURNAUD M., MAUCET D. & CHAVANON G., 1983. — Structure des peuplements de macro-invertébrés benthiques en place et en dérive. *Acta Ecologica, Ecol. Gener.*, 4, 1, 39-57.
- BOUZIDI A., 1983. — *Cours d'eau temporaires du massif des Maures : Hydrologie, hydrochimie, organisation et évolution de leurs peuplements*. Thèse de 3^e cycle, Faculté St-Jérôme, Université de Aix-Marseille III, 179 p.
- CAILLEZ F. & PAGES J. P., 1976. — *Introduction à l'analyse des données*. SMASH, Paris, 616 p.
- CAILLEZ F., 1984. — *Analyse des données*. Les presses de l'Université de Montréal, 103 p.
- CARLETON T. J., 1984. — Residual ordination analysis, a method for exploring vegetation environment relationships. *Ecology*, 65, 469-477.
- CARREL G., 1986. — *Caractérisation physico-chimique du Haut-Rhône français et des ses annexes : incidences sur la croissance des populations d'alevins*. Thèse de doctorat, Université de Lyon I, 185 p.
- CARREL G., BARTHÉLÉMY D., AUDA Y. & CHESEL D., 1986. — Approche graphique de l'analyse en composantes principales normée : utilisation en hydrobiologie. *Acta Ecologica, Ecol. Gener.*, 7, 2, 189-203.
- CASEY H. & LADLE M., 1976. — Chemistry and biology of the South Winterbourne, Dorset England. *Freshwat. Biol.*, 6, 1, 1-12.
- CHESEL D., LEBRETON J. D. & YOCOZ N., 1987. — Propriétés de l'analyse canonique des correspondances : une illustration en hydrobiologie. *Rev. Stat. Appli.* (sous presse).
- DECAMPS H., MASSIO J. C. & DARIOS J. C., 1976. — Variations des teneurs en matières minérales et organiques transportées dans une rivière canalisée, le Lot. *Annls Limnol.*, 12, 3, 215-237.
- DESCY J. P. & EMPAIN A., 1984. — 1. Meuse. In: WHITTON B. A., *Ecology of European Rivers*, 1-23.
- EDWARDS R. W. & BROOKER M., 1984. — 3. Wye. In: WHITTON B. A., *Ecology of European Rivers*, 51-82.
- ESCOUFIER Y., 1980. — L'analyse conjointe de plusieurs matrices de données. In: JOLIVET et coll., eds., *Biométrie et Temps*, 59-77.
- ESCOUFIER Y., 1987. — The duality diagramm: a means for better practical applications. In: LEGENDRE P. & LEGENDRE L., eds., *Development in numerical ecology*, Nato advanced Study Institute Serie G (Ecological Sciences), Springer-Verlag, Berlin (sous presse).
- FEUILLADE J., 1972. — Étude de la physico-chimie d'un réseau de rivières du Morvan. *Ann. Hydrobiol.*, 3, 1, 47-57.
- FOUCART T., 1984. — *Analyse factorielle de tableaux multiples. Méthodes et programmes*. Masson, Paris, 187 p.
- GITTINS R., 1985. — *Canonical Analysis. A review with Application in Ecology*. Springer-Verlag, Berlin, 350 p.
- GIUDICELLI J., DIA A. & LEGIER P., 1980. — Étude hydrobiologique d'une rivière de région méditerranéenne, l'Argens Var, France. Habitats, hydrochimie, distribution de la faune benthique. *Bidjr. Dierkd.*, 50, 303-341.
- GOEDMAKERS A., 1981. — Population dynamics of 3 Gammarid species Crustacea, Amphipoda, in a french chalk stream. *Academisch proefschrift*, Amsterdam, 190 p.
- GREEN R. H., 1971. — A multivariate statistical approach to the Hutchinsonian niche: Bivalve molluscs of Central Canada. *Ecology*, 52, 543-556.
- GREENACRE H. J., 1984. — *Theory and applications of correspondence analysis*. Academic Press, Londres, 364 p.

- GREGOIRE A. & CHAMPEAU A. M., 1982. — Physico-chemical parameters of a regulated river: the Verdon France. In: LILLEHAMMER A. & SALTVEIT S. J., *Regulated Rivers*, 41-52.
- GROSS F., 1975. — *Étude écologique d'un ruisseau de l'entre-deux-mers*. Thèse de 3^e cycle, Université de Bordeaux I, 164 p.
- HALL A., VALENTE I. M. & DAVIES B. R., 1977. — Cabo Bassa: Some preliminary physico-chemical and zooplankton pre-impoundment survey results. *Hydrobiologia*, 50, 17-25.
- ILTIS A. & LEVEQUE C., 1982. — Caractéristiques physico-chimiques des rivières de Côte-d'Ivoire. *Rev. Hydrobiol. trop.*, 15, 2, 115-130.
- JOP K., 1981. — Ecology of the forest stream Lane Bloto in the Niepotomice Forest. 1. Chemism of water and bottom sediments and its changes under the influence of industrial pollution. *Acta Hydrobiol.*, 23, 2, 107-123.
- JORDANA C. & GARCIA L. R., 1981. — Principales tendencias de variación en los ecosistemas loticos de Sierra Nevada. *Actas del 1^{er} Congreso español de Limnología*, I, FORNELLS N. P., éd., 213-223.
- LAURO N. & D'AMBRA L., 1983. — L'analyse non symétrique des correspondances. In: DIDAY E. & coll., eds., *Data Analysis and Informatics*, vol. III, Elsevier Science Publishers, North-Holland, 433-446.
- LEBART L., MORINEAU A. & FENELON J. P., 1982. — *Traitements des données statistiques. Méthodes et Programmes*. Dunod, 2^e édition, Paris, 518 p.
- LEGENDRE P. & LEGENDRE L., 1979. — *Écologie numérique. 2. La structure des données écologiques*. Masson, Paris, 247 p.
- LEVEQUE C., DEJOUX C. & ILTIS A., 1983. — Limnologie du fleuve Bandama, Côte-d'Ivoire. *Hydrobiologia*, 100, 113-141.
- MAQUET B., 1983. — Caractéristiques chimiques et biologiques des eaux de la vallée du Samson. *Annls Soc. r. Zool. Belg.*, 113, 1, 3-18.
- MINSHALL G. W. & ANDREWS D. A., 1973. — An ecological investigation of the Portneuf River, Idaho: a semiarid-land stream subjected to pollution. *Freshwat. Biol.*, 3, 1, 1-30.
- MOHAMMAD M. B. M., 1980. — A hydrobiological survey of a polluted canal. *Hydrobiologia*, 74, 179-186.
- OBADIA J., 1978. — L'analyse en composantes explicatives. *Rev. Stat. Appl.*, 24, 4, 5-28.
- ORTAL R. & POR F. D., 1977. — Effect of hydrological changes on aquatic communities in the Lower Jordan River. *Verh. Internat. Verein. Limnol.*, 20, 3, 1543-1551.
- PAGES T. P. & TOMASSONE R., 1979. — Analyse des données et modélisation. In: MALOINE S. A., ed., *Elaboration et justification des modèles : applications en Biologie*, tome II, 407-439.
- PEGAZ-MAUCET D., 1980. — *Impact d'une perturbation d'origine organique sur la dérive des macro-invertébrés d'un cours d'eau. Comparaison avec le benthos*. Thèse de 3^e cycle, Université Lyon I, 130 p.
- PERNIN M. O., 1986. — *Contribution à la méthodologie d'analyse de données longitudinales. Exemple de la croissance chez l'être humain (Auxologie)*. Thèse de doctorat, Université de Lyon I, 264 p.
- PONTIER J. & PERRIN M. O., 1986. — Multivariate and longitudinal data on growing children solved by Longi. *Proc. of the 3rd Internat. Symposium on Data Analysis*, Bruxelles, Plenum Press, London, sous presse.
- 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. *Oikos*, 37, 21-38.
- RAO C. R., 1964. — The use and interpretation of principal component analysis in applied research. *San khya*, A, 26, 329-359.
- ROMANE F., 1972. — Utilisation de l'analyse multivariable en phytoécologie. *Inv. Pesq.*, 36, 1, 131-139.
- ROSILLON D., 1984. — *Dynamique des populations de macroinvertébrés benthiques d'une rivière salmonicole le Samson : approche des facteurs régulateurs*. Thèse Doct. ès Sciences, Fac. Sci. Namur, 199 p.
- SAAD A. H. S. & ABBAS M. M., 1985. — Limnological investigations of the Rosetta branch of the Nile. 2. Seasonal variations of nutrients. *Freshwat. Biol.*, 15, 6, 655-660.
- SABATIER R., 1983. — *Approximation d'un tableau de données. Application à la reconstitution des paléoclimats*. Thèse de 3^e cycle, Montpellier, 184 p.
- SCHEFFE H., 1959. — *The analysis of Variance*. Wiley, New York, 477 p.

- SCHEKTMAN Y., 1978. — *Contribution à la mesure en facteurs dans les sciences expérimentales et à la mise en œuvre automatique des calculs statistiques*. Thèse d'état, U. P. S., Toulouse.
- TAKEUCHI K., YANAI H. & MUKHERJEE B. N., 1982. — *The foundation of multivariate analysis. A unified approach by means of projection onto linear subspaces*. Wiley, eds, 2nd ed., New York, 458 p.
- TER BRAAK C. J. F., 1986. — Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 5, 1167-1179.
- VERNEAUX J. & REZZOUK M., 1974. — Les structures d'un grand cours d'eau à Salmonidés : la Loue, Massif du Jura. Essai typologique et problème des relations entre espèces et milieu. *Annls Limnol.*, 10, 2, 131-162.
- VESPINI F., 1985. — *Contribution à l'étude hydrobiologique du Buech, rivière non aménagée de Haute-Provence. Suivi d'une charge organique faible et mesure de ses effets sur les composantes physiques, chimiques et biologiques de l'écosystème*. Thèse 3^e cycle, Université de Aix-Marseille I, 148 p.
- WHITTAKER R. H., 1967. — Gradient analysis of vegetation. *Biological Reviews*, 42, 207-264.
- WOLLENBERG A. L., 1977. — Redundancy analysis, an alternative for canonical correlation analysis. *Psychometrika*, 42, 2, 207-219.