

# Analyse statistique de l'évolution temporelle de paramètres hydrobiologiques dans les bassins du port du Havre

Analyse spectrale  
Désaisonnalisation  
Séries temporelles multivariées  
Eutrophisation  
Spectral analysis  
Deseasonalisation  
Multivariate time-series  
Nutrient enrichment

Alain Ménesguen<sup>a</sup>, Alain Laurec<sup>b</sup>

<sup>a</sup> Centre National pour l'Exploitation des Océans (CNEXO), Centre Océanologique de Bretagne, B.P. 337, 29273 Brest Cedex, France.

<sup>b</sup> Institut Scientifique et Technique des Pêches Maritimes, rue de l'Île d'Yeu, B.P. 1049, 44037 Nantes Cedex, France.

Reçu le 1/12/82, révisé le 10/9/83, accepté le 14/9/83.

## RÉSUMÉ

L'étude de la structure spatiale et de l'évolution temporelle de plusieurs paramètres hydrobiologiques dans un système de bassins portuaires semi-fermés a permis d'illustrer les possibilités analytiques d'une méthodologie de traitement statistique des séries temporelles multivariées. Cette technique se caractérise par l'extraction préalable des tendances et leur traitement multivariable séparé, puis par l'analyse multispectrale des signaux résiduels. Elle a l'avantage de clarifier l'analyse des traits majeurs du phénomène (les tendances), et de démasquer la structure spatio-temporelle fine au travers des signaux résiduels. L'application au Port Autonome du Havre a mis en évidence un enrichissement majeur en sels nutritifs d'une part, une origine partielle possible pour cet enrichissement d'autre part, sous forme de sources de perturbations caractérisées par leur contenu fréquentiel et la séquence de propagation associée.

*Oceanol. Acta*, 1984, 7, 1, 53-61.

## ABSTRACT

Statistical analysis of the temporal evolution of hydrobiological parameters in the flooding docks at Le Havre (France)

The study of the spatial structure and temporal evolution of several hydrobiological parameters in a flooding docks system provided a good illustration for the analytical power of a methodology dealing with the statistical processing of multivariate time-series. This technique can be characterized by prior extraction of the trends, treated separately as ordinary multivariate data, and by a consequent multi-spectral analysis of the residual signals. This procedure offers the advantage of clarifying the analysis of the major features of the phenomenon (*i.e.* the trends) and unmasking the subtle spatio-temporal structure through the residual signals. The application on Le Havre harbour showed a strong enrichment in nutrients and also a plausible but partial origin for this enrichment in the form of perturbations sources. These can be distinguished by their frequential signature and the associated path of propagation.

*Oceanol. Acta*, 1984, 7, 1, 53-61.

## INTRODUCTION

Depuis plus d'une décennie, l'abondance croissante des données a conduit l'écologie marine à faire usage de techniques mathématiques telles que l'analyse multivariable de grands tableaux d'une part, et l'analyse spectrale de séries temporelles uni- ou bivariées d'autre part. Un contrat passé au Centre National pour

l'Exploitation des Océans (CNEXO) par le Port Autonome du Havre (P.A.H.) pour l'étude de la variabilité spatiale et temporelle des phénomènes hydrobiologiques dans ses bassins à flot, a livré l'occasion de préciser, puis d'illustrer l'utilisation combinée de ces deux types de techniques.

## LES DONNÉES ET LEUR CONTEXTE BIBLIOGRAPHIQUE

Les données, fournies par le P.A.H., se présentaient en effet sous forme de 11 séries temporelles multivariées, chacune correspondant à une station de mesure particulière dans les bassins. Les paramètres mesurés, douze au total, correspondaient à des grandeurs physiques (température, salinité, pH, matières en suspension) et chimiques (phosphate, ammonium, nitrite, nitrate, azote organique total, sulfates, oxygène dissous et fluor). Quant à l'échantillonnage, il s'étalait au maximum sur une période de cinq ans, de 1975 à 1979, à raison d'une mesure par mois.

La relative complexité d'une telle base de données, où se mêlent des caractères spatiaux et des aspects temporels de différentes périodes, nous a semblé requérir une approche en trois phases. La première consiste naturellement en la recherche de grandes lignes de fonctionnement du système hydrobiologique telles qu'elles ont éventuellement été décrites dans des études antérieures : en l'absence de travaux traitant spécifiquement des phénomènes hydrobiologiques à l'intérieur des bassins du P.A.H., il a fallu faire appel à des études sur les milieux géographiquement voisins du P.A.H., et avec lesquels les bassins sont d'ailleurs en contact. Une première série de résultats peut être tirée de l'étude faite par Guillaud (1980) sur les données hydrobiologiques collectées en routine par le Réseau National d'Observation de la Qualité du Milieu Marin (R.N.O.) en huit stations de l'embouchure et de la baie de la Seine. Outre la position géographique du gradient de dilution de la Seine, ce travail met en évidence une évolution pluriannuelle de certains paramètres hydrobiologiques : de 1976 à 1979, un phénomène climatique général a entraîné une baisse de la salinité et de la température à l'intérieur de l'estuaire ; par ailleurs est intervenue une modification des apports de la Seine en sels nutritifs : l'ammonium a diminué tandis que les nitrates ont augmenté ; les phosphates sont restés stationnaires. A cette tendance pluriannuelle, se superpose naturellement un cycle hydrobiologique saisonnier, intimement lié au fonctionnement de l'écosystème planctonique ; ceci peut être retrouvé et précisé dans les nombreuses études écologiques réalisées dans les zones côtières avoisinant les sites d'implantation de centrales nucléaires, en particulier Paluel et Penly, situées non loin de la baie de Seine (cf. Anonyme, 1978 et 1980). Il apparaît en particulier que l'entrée des éléments minéraux dans les compartiments biologiques du cycle se fait assez régulièrement chaque année en mai-juin, lors de l'intense développement des algues unicellulaires pélagiques (« bloom » phytoplanctonique).

Étant donné la part importante prise par au moins trois composantes majeures du signal identifiées *a priori* (gradient géographique, variation pluriannuelle, cycle saisonnier), il devenait évident qu'une simple analyse multivariable de l'ensemble des données ne fournirait qu'une explication confuse du fonctionnement hydrobiologique des bassins. Une réflexion approfondie sur le traitement des diverses composantes du signal s'imposait, et c'est de ces considérations méthodologi-

ques, détaillées ci-après, que découleront la seconde et la troisième étape de ce travail : l'analyse multivariable séparée des moyennes annuelles et des termes mensuels, puis l'analyse multispectrale des signaux résiduels.

## ASPECTS MÉTHODOLOGIQUES

### Extraction des tendances

Nombreuses sont les méthodes permettant de séparer dans un signal, tendances et composantes résiduelles. L'aspect le plus important correspond cependant au modèle choisi pour décrire la tendance. Compte tenu de l'importance et du caractère répétitif des composantes saisonnières dans les données traitées, ces composantes doivent être en tout état de cause incluses dans la tendance. Les possibles variations pluriannuelles, compte tenu de la brièveté de la série, ne peuvent pas raisonnablement être analysées par le biais des techniques d'analyse spectrale. Elles ont donc été rattachées à la tendance qui regroupe une combinaison des évolutions pluriannuelles et saisonnières. Les évolutions pluriannuelles peuvent être traitées de diverses façons, les plus courantes passant par l'utilisation de polynômes de bas degré. Cette étude a préféré utiliser une description par paliers des évolutions pluriannuelles (Laurec, Le Gall, 1975). La tendance est alors définie pour le mois  $m$  de l'année  $a$  par la somme  $s_m + t_a$ , le  $s_m$  étant soumis à la classique contrainte

$$\sum_{m=1}^{12} s_m = 0, \text{ car ils définissent une simple modulation.}$$

Cela impose à la tendance des variations quelquefois brutales au passage d'une nouvelle année, mais offre des indices annuels directement interprétables.

Pour un signal mesuré pendant  $n$  années, l'estimation des  $n$  paramètres  $t_a$  et des 12 paramètres  $s_m$  se fait par un ajustement selon les moindres carrés. Ce procédé, non optimal dès lors que les résidus ne forment pas un bruit blanc (Anderson, 1971), a été utilisé par souci de simplicité. Il conduit, lorsque les séries sont sans lacune, au même résultat que des moyennes sur l'année pour estimer les  $t_a$  et des moyennes, sur l'ensemble des années, des déviations pour un mois donné par rapport à la moyenne annuelle pour estimer le  $s_m$ . Il présente l'avantage d'être utilisable même si des lacunes apparaissent.

### Traitement des séries bivariées

Les techniques courantes de traitement considèrent la série bivariée observée  $[(x_i, y_i), i = 1, I]$  comme une réalisation du processus bivarié  $(\tilde{x}, \tilde{y})$ . Parmi les hypothèses requises par ces techniques, et déjà commentées par Laurec (1982), on retiendra la notion de stationnarité : à l'ordre 1, elle implique que les processus sont dépourvus de tendance : à l'ordre 2, elle signifie que les autocovariances et covariances croisées entre les séries  $(\tilde{x}_i, \tilde{y}_i)$  et  $(\tilde{x}_{i+l}, \tilde{y}_{i+l})$  ne dépendent que du décalage  $l$ .

Après soustraction de la tendance estimée, la stationnarité d'ordre 1 peut être espérée. Le champ est alors ouvert aux techniques fondées sur les moments d'ordre 2 : étude des autocovariances et covariances croisées dans le domaine temporel, de leur transformées de Fourier (autospectres et spectres croisés) dans le domaine des fréquences. Si on appelle  $\tilde{s}_{xy}(f)$  la valeur complexe du spectre croisé à la fréquence  $f$ ,  $\tilde{r}_{xy}(f)$  et  $\tilde{q}_{xy}(f)$  les parties réelle et imaginaire de  $\tilde{s}_{xy}(f)$ , on utilisera :

la cohérence croisée  $\tilde{c}_o^2(f) = \tilde{s}_{xy}^2(f) / [\tilde{s}_{xx}(f) \cdot \tilde{s}_{yy}(f)]$ ,  
la phase  $\tilde{\Psi}(f) = \text{Arctg}[\tilde{q}_{xy}(f) / \tilde{r}_{xy}(f)]$ .

La cohérence, à l'instar d'une valeur absolue de corrélation, indique à chaque fréquence l'intensité d'un éventuel lien linéaire, tandis que la phase indique le décalage entre les deux variables. Alors que la corrélation simple représente une moyenne des liens sur toute la gamme des fréquences, l'étude dans le domaine des fréquences permet une étude plus fine du phénomène. La longueur  $I$  du signal limite cependant supérieurement les fréquences pour lesquelles une estimation est possible : toute estimation est impossible au-delà de la fréquence de Nyquist  $2/I$ .

L'estimation des autocorrélations et corrélations croisées se fait selon des méthodes désormais classiques. Toutefois, la présence de lacunes nécessite l'utilisation de procédés adaptés. Concrètement, il a été fait appel à la technique de Jones (1962; 1971), généralisée au cas multivarié par Laurec (1979; 1982).

### Traitement des séries multivariées

Ce qui est vrai pour une série uni- ou bivariée l'est aussi pour un signal multivarié. Comme l'a souligné la première partie, cela conduit dans un premier temps à analyser séparément tendances et résidus. A l'intérieur des tendances, on pourra poursuivre la séparation en traitant différemment composantes annuelles et saisonnières; quant aux signaux résiduels, on les analysera fréquence par fréquence.

#### Traitement des tendances

Étant donné le modèle additif adopté dans notre cas pour les tendances, il y a communauté de nature mathématique entre les termes annuels et les termes mensuels estimés; la méthode de traitement sera donc identique pour les composantes annuelles et saisonnières, et l'on pourra raisonner par exemple sur les secondes sans perte de généralité. Le problème qu'il faut alors résoudre est celui du passage d'un tableau de composantes saisonnières réellement tridimensionnel ( $k$  stations  $\times p$  paramètres hydrobiologiques  $\times 12$  mois) à un tableau bidimensionnel, traitable par les classiques méthodes d'analyse multivariée. Si l'on désire utiliser, comme dans notre cas, une méthode d'analyse en composantes principales (A.C.P.) centrées et réduites, dans laquelle les rôles de « variables » et d'« observations » ne sont pas symétriques, il faut de façon précise choisir les descripteurs qui seront considérés comme « variables » dans l'analyse. Parmi les nombreuses combinaisons possibles, restreignons-

nous à celles qui, conformément au désir de comprendre en premier lieu les interrelations entre paramètres hydrobiologiques, considèrent de toute façon ces paramètres hydrobiologiques comme « variables » de droit dans l'analyse : deux options sont alors particulièrement intéressantes à examiner. La première, considérant la variabilité spatiale comme suffisamment importante dans les bassins pour correspondre à une modification géographique des relations entre paramètres hydrobiologiques, placera aussi les stations en « variables », ce qui conduira à l'analyse d'un tableau de 12 observations et ( $k \times p$ ) variables. La seconde, que nous avons retenue dans ce travail, considère que la variabilité spatiale est minime, et incapable, en tout état de cause, de faire apparaître des différences entre les relations inter-paramètres hydrobiologiques aux diverses stations : seuls les  $p$  paramètres hydrobiologiques restent des « variables », et l'on a cette fois ( $12 \times k$ ) « observations ». A signaler enfin qu'à la différence de l'analyse spectrale, les techniques d'analyse multivariée de tableaux bidimensionnels préservent l'individualité des observations, dès lors qu'un usage approprié de la dualité « observations-variables » est fait.

### Analyse spectrale multivariée des résidus

#### Principe

Des analyses multivariées sont concevables dans le domaine temporel. Cependant, l'analyse des séries uni- ou bivariées étant généralement plus aisément interprétable dans le domaine des fréquences, nous accorderons donc priorité, ainsi que Brillinger (1975), aux analyses multivariées dans le domaine des fréquences.

A chaque fréquence apparaît une matrice de spectres et spectres croisés, qui, en dépit de l'apparition de nombres complexes, joue en quelque sorte le rôle d'une matrice de variance-covariance dans les A.C.P. classiques. Par analogie, on peut songer à appliquer à cette matrice spectrale une procédure de diagonalisation. Cela est mathématiquement possible, les matrices de densités spectrales étant hermitiennes définies positives (en théorie, certains procédés d'estimation peuvent conduire à des matrices estimées ne respectant pas cette condition, mais on ne couvrira pas cet aspect). La diagonalisation définit une nouvelle base de « facteurs » orthogonaux (de cohérence nulle). Il est possible de décomposer les variables originelles par rapport à cette base. Les coordonnées par rapport aux facteurs seront en général des nombres complexes.

Brillinger (1975) offre un point de vue intéressant en montrant que la projection sur les  $k$  premières composantes,  $k$  étant un nombre fixé quelconque, offre une projection optimale sur un sous-espace de dimension  $k$ . L'optimalité s'entend au sens de « meilleur résumé possible » de la structure de la matrice multispectrale complète (théorème 9.2.3., p. 342).

Toutefois, la projection sur le ou les premiers facteurs à une fréquence donnée n'est pour Brillinger (1975) qu'un résultat intermédiaire, car il utilise l'ensemble des résultats aux différentes fréquences pour définir globalement dans le domaine temporel des séries

composantes principales. Ce point de vue, utile dans d'autres contextes, ne l'est pas si la priorité va à l'analyse fine des données, qui a tout à gagner à une séparation des fréquences (Laurec, 1979). Même si des phénomènes identifiés peuvent être rattachés aux premiers facteurs, qui se retrouvent globalement d'une fréquence à l'autre, rien ne garantit que le même phénomène sera attaché sur le premier facteur à toutes les fréquences. Le risque est grand de créer une série première composante au sens de Brillinger qui soit compositée.

Mathématiquement donc, l'approche suivie par Laurec (1979) n'est pas originale. Mais elle suggère, au contraire de Brillinger (1975) et d'études faites en océanographie dans un contexte voisin (Wallace, Dickinson, 1972), une interprétation séparée des résultats obtenus à chaque fréquence.

#### Interprétation des résultats à une fréquence donnée

Il est possible d'opérer une projection des variables sur un sous-espace défini par un nombre quelconque de facteurs. Vis-à-vis d'un facteur particulier, la coordonnée d'une variable sera généralement donnée comme on l'a dit par un nombre complexe. Ce nombre peut être décrit par parties réelle et imaginaire. Pour chaque facteur, le report des variables donnera donc naissance à une figure bidimensionnelle. La coordonnée complexe peut aussi être décrite par module et phase. La phase attachée à une variable particulière n'est pas interprétable en elle-même, car les facteurs sont définis à un coefficient de la forme  $e^{i\theta}$  prêt, comme dans les analyses factorielles classiques où les facteurs sont définis à un coefficient  $\pm 1$  prêt. En revanche, les éventuelles différences de phase d'une variable à l'autre sont riches d'informations, car elles peuvent traduire des délais dans les réactions des variables au facteur.

Le module, quant à lui, est directement lié à la cohérence entre une variable et un facteur. Concrètement, si avant la diagonalisation, les variables sont normées de façon à ce que leur autospectre soit égal à 1, le module donne directement la cohérence recherchée. Sur le diagramme plan attaché à chaque facteur (voir fig. 1), on pourra alors faire figurer un cercle de cohérence, analogue au cercle de corrélation des A.C.P.

#### Remarques pratiques

1) La possibilité de normer les spectres à chaque fréquence a été évoquée. En fait, trois procédures sont possibles :

- pratiquer les diagonalisations sur les matrices spectrales issues de l'analyse des données brutes ;
- ramener à 1 la variance globale des séries, ce qui ne garantit nullement l'égalité des autospectres à toutes les fréquences ;
- normer les autospectres à chaque fréquence.

Cette dernière solution paraît préférable pour que le résultat des analyses à une fréquence donnée soit aussi indépendant que possible des autres sources de variation.

2) Pas plus que dans les analyses multivariées classiques, il ne faut espérer systématiquement une identification d'un facteur de la diagonalisation à un

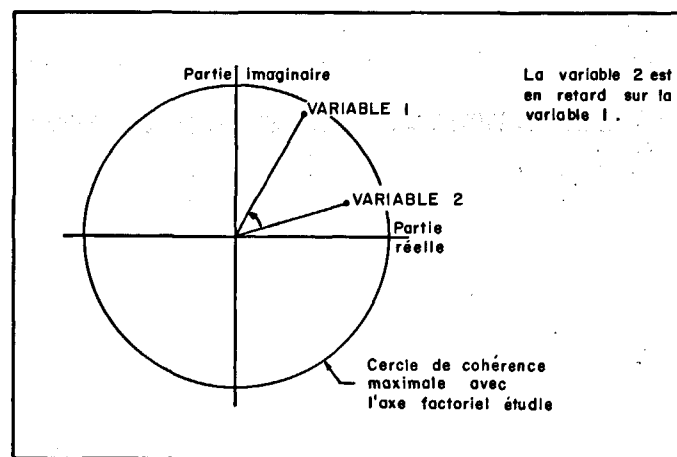


Figure 1

Visualisation des coordonnées complexes des différentes variables sur un axe issu de l'ACP de la matrice des spectres à une fréquence donnée.

*Representation of the complex coordinates of the different variables on one axis produced by a principal components analysis of the spectral and co-spectral values at a given frequency.*

phénomène concret particulier. On peut craindre des facteurs issus d'influences combinées de plusieurs phénomènes. Si vis-à-vis de ces phénomènes, les déphasages des variables sont différents, le résultat « global » peut conduire à une structure difficilement interprétable.

3) Aux difficultés possibles d'interprétation inhérentes aux techniques factorielles, s'ajoutent celles propres aux analyses spectrales. C'est ainsi que des contaminations (aliasing) de fréquence à fréquence apparaissent. Il n'est donc pas utile d'opérer des analyses à toutes les fréquences ; on se contentera d'analyses suffisamment espacées pour limiter le phénomène de contamination, compte tenu des lissages opérés sur les spectres, et donc de l'ouverture des fenêtres spectrales utilisées.

En pratique par exemple, si  $M$  est l'ouverture de la fenêtre, on peut recommander de pratiquer les analyses multispectrales aux fréquences de base  $1/M$ ,  $2/M$ , etc. Par ailleurs, comme pour toute analyse spectrale, des problèmes de repliement des spectres sont à craindre, tout particulièrement à l'approche de la fréquence de Nyquist (cf Laurec, 1982).

Enfin, il faut signaler que, pour la fréquence nulle et la fréquence de Nyquist, les spectres croisés sont réels et l'on est ramené au cadre multivariable ordinaire. L'interprétation des résultats à ces fréquences peut toutefois être mise en cause.

4) Le traitement des séries de résidus par l'analyse spectrale vient d'emblée éliminer toute individualité des observations. Ceci est inéluctable, dès l'instant où cette technique d'analyse vise à estimer les propriétés du processus parental et non à décrire les particularités du signal. Ce trait caractéristique de l'analyse spectrale se retrouvera bien sûr à l'issue des analyses multivariées des matrices spectrales, où il n'existe plus de dualité « observations-variables », contrairement au cas des analyses multivariées des tendances.

Figure 2  
Position et nomenclature des stations de PAH (repérées par des nombres) et du RNO (repérées par des lettres).

Localization and appellation of the PAH (indicated by numbers) and the RNO (indicated by letters).

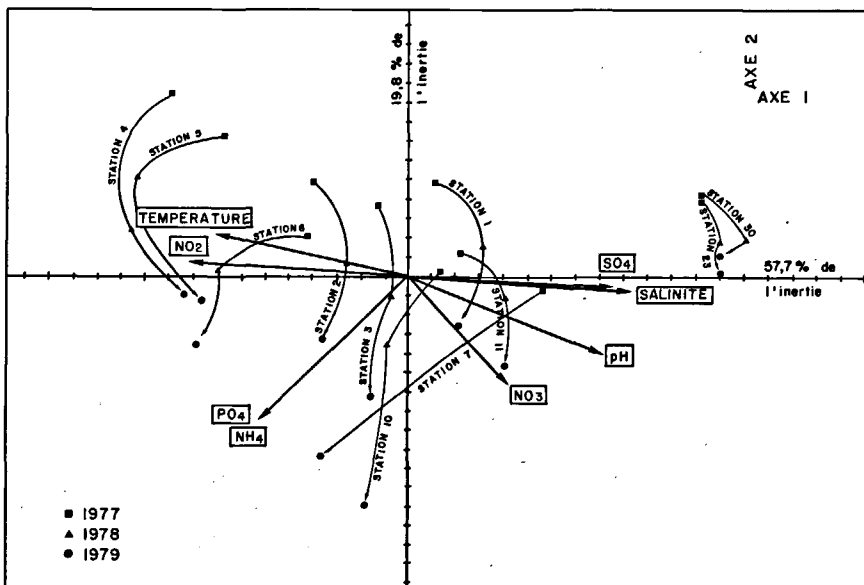
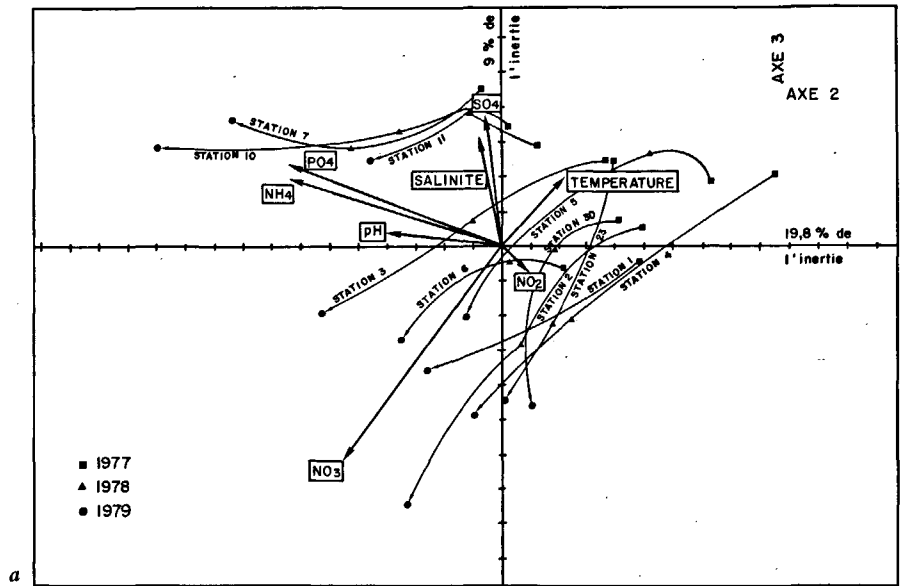
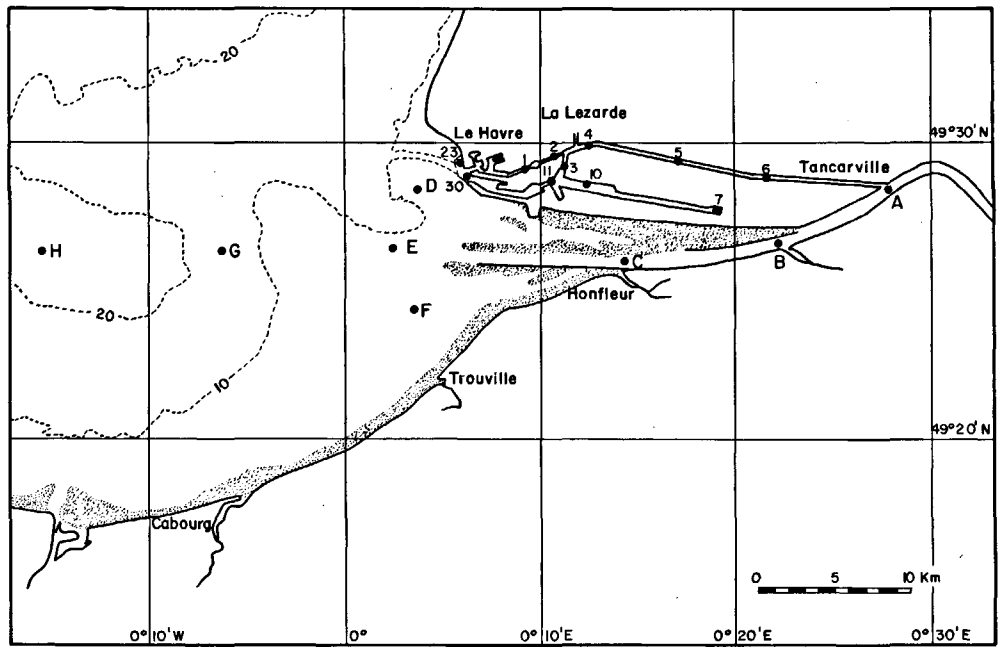


Figure 3  
Représentation des situations annuelles moyennes aux stations du PAH dans le sous-espace des trois premiers axes de l'Analyse en Composantes Principales.  
Representation of the annual mean situations at the PAH stations in the subspace of the three first axes of the Principal Components Analysis.

## RÉSULTATS DE L'ÉTUDE DES BASSINS DU P.A.H.

### Estimation des tendances

Signalons d'abord que, dans le souci de replacer quantitativement l'évolution moyenne des bassins dans un contexte géographique plus large, nous avons adjoint aux données propres du P.A.H. un tableau strictement parallèle de mesures effectuées par le R.N.O. en ses huit stations du point d'appui « baie de Seine » : il s'agit des données dont l'exploitation par Guillaud (1980) a déjà été mentionnée. La figure 2 précise l'emplacement et la nomenclature des stations du P.A.H. et du R.N.O.

En second lieu, le modèle retenu pour la tendance, en particulier sa composante saisonnière de période 1 an, nous a conduit à éliminer d'emblée certains paramètres hydrobiologiques ne suivant pas, à première vue, ce type d'évolution. C'est le cas des matières en suspension, de l'azote organique total, de l'oxygène dissous et du fluor, qui ne présentent pas de cycle saisonnier marqué du fait de la prépondérance des phénomènes non biologiques dans leur dynamique propre. Par ailleurs, pour les huit paramètres restants, il a fallu éliminer les deux premières années d'étude (1975 et 1976) pour des raisons d'homogénéité de la base de données : seuls les paramètres physiques avaient en effet été mesurés lors des deux premières années. Finalement, pour chaque variable et à chaque station, il a fallu estimer trois termes annuels moyens (1977, 1978 et 1979) et douze termes mensuels moyens.

### Étude séparée des tendances

Ainsi qu'on peut le voir sur la figure 3, le système des trois premiers axes factoriels issus de l'AC.P. du tableau des  $(11 \times 3) \times 8$  moyennes annuelles permet de dégager une structure très forte (86% de la variance), où se distinguent nettement une composante géographique stable et prépondérante (d'où l'ordonnancement pérenne des stations sur l'axe 1) et une composante temporelle quasi-indépendante (d'où sa prise en compte presque

exclusive par l'axe 2), affectant parallèlement toutes les stations. On résumera l'analyse en disant que, bien que fortement caractérisée par leur position géographique sur le gradient d'influence de l'eau relativement douce, chaude et riche en nitrites introduite dans les bassins par la rivière la Lézarde (station 4), les stations du P.A.H. sont toutes affectées d'un enrichissement pluriannuel ininterrompu en nitrates, ammonium et phosphates, visualisé par un déplacement vers le pôle négatif du deuxième axe factoriel des points représentatifs d'une même station aux trois années successives. L'enrichissement en ammonium et phosphates est particulièrement prépondérant dans le bassin sud (stations 7, 10 et 11), là où justement se font d'importants rejets industriels de ces sels. La projection des points R.N.O. à titre illustratif (*cf* fig. 4) permet d'abord de retrouver la même tendance à l'enrichissement en nitrates sur les trois ans considérés, déjà signalée par Guillaud (1980); cet enrichissement est d'autant moins tamponné que l'on remonte dans l'estuaire (stations C, B, A). Elle permet par ailleurs de replacer la composante géographique de l'hydrologie des bassins dans le gradient de dilution Seine → baie de Seine. On remarquera que seules les stations proches des écluses (23, 30, 1 et 11), ainsi que celles du bassin Sud (10 et 7) se placent bien dans l'axe de dilution de la Seine matérialisé sur la figure 4 par l'alignement des stations A, B et C avec le pôle marin regroupant les stations D, E, F et G. Les autres s'écartent de cet axe de façon d'autant plus marquée que l'on s'approche de la station 4 : la raison de cette originalité du bassin nord réside dans sa richesse particulière en nitrites.

Une seconde analyse en composantes principales effectuée sur le tableau des  $(11 \times 12) \times 8$  termes mensuels met en évidence, ainsi qu'on peut le voir sur la figure 5, un remarquable cycle saisonnier : la structure annulaire visible sur le premier plan factoriel représente en effet 71% de la variance totale des termes mensuels. En dépit de l'absence de mesures biologiques, on peut néanmoins assurer, par comparaison avec les cycles mis en évidence lors des études de projet de centrales nucléaires de Paluel et Penly (Anonyme, 1978 et 1980),

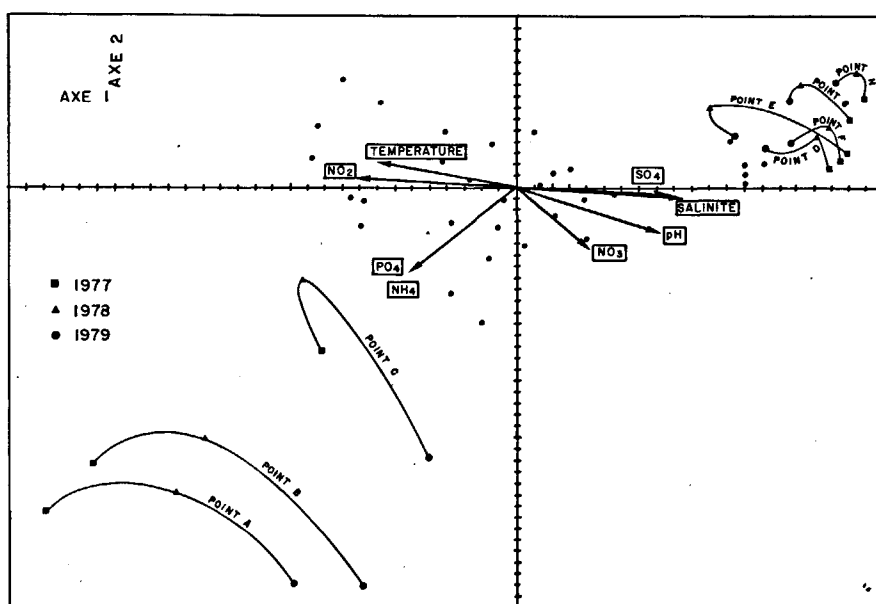


Figure 4

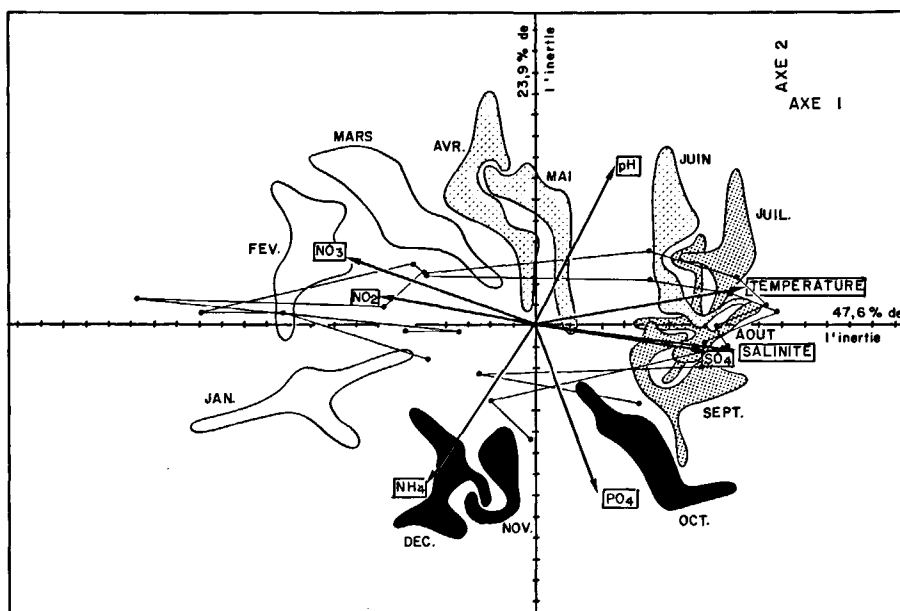
Représentation des situations annuelles moyennes aux stations RNO sous forme de « points-supplémentaires » dans le premier plan factoriel de l'analyse des situations aux stations du PAH.

Representation of the annual mean situations at the RNO stations, as additional (i. e. inactive) points in the first factorial plane of the analysis of the mean situations at the PAH stations.

Figure 5

Premier plan factoriel de l'ACP des termes mensuels de désaisonnalisation calculés aux stations du PAH.

First factorial plane of the Principal Components Analysis of the monthly additive factors produced by the deseasonalization of the signal at the PAH stations.



qu'il s'agit là d'un cycle planctonique tempéré classique. L'azote en particulier, essentiellement sous forme de nitrate en fin d'hiver, est intensément consommé par le phytoplancton en juin-juillet lors de sa floraison; cette dernière est repérable à l'augmentation bien visible du pH du fait de la rapide assimilation du  $\text{CO}_2$  dissous. En fin d'automne et en hiver, on retrouve bien les étapes successives de la reminéralisation de l'azote organique par les bactéries: maximum d'ammonium en novembre, de nitrite en janvier et de nitrate en février. Une comparaison avec les cycles visibles aux points R.N.O. de la baie de Seine et ceux déjà mentionnés pour Paluel et Penly (littoral de la Seine-Maritime) montre cependant une différence d'évolution entre la mer libre et les bassins du P.A.H.: on observe en effet en mer, en plus de la floraison estivale de juin-juillet déjà rappelée, une première floraison occasionnant en mars un pic de pH bien visible; cette floraison printanière est due principalement (Anonyme, 1978 et 1980) à la pullulation des diatomées *Skeletonema costatum*, *Thalassiosira* sp. et *Chaetoreros* sp., tandis que la floraison estivale est surtout due à *Asterionella japonica*, *Rhizosolenia* sp. et *Nitzschia* sp. En l'absence de suivi quantitatif de la composition du phytoplancton dans les bassins du P.A.H., on ne peut malheureusement ni vérifier une éventuelle similitude de peuplement estival entre les bassins et la mer, ni bien sûr rechercher une explication à l'absence apparente de la floraison printanière dans les bassins.

#### Étude des résidus

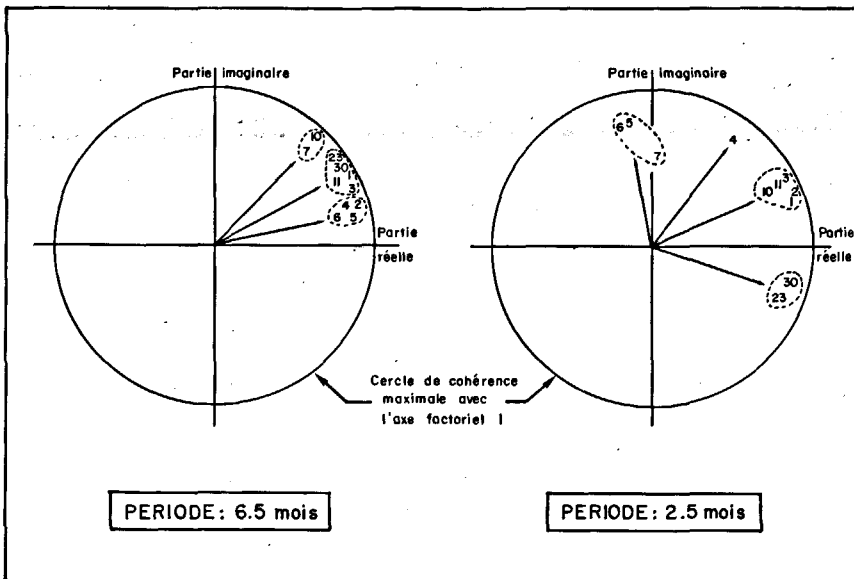
La périodicité mensuelle des mesures et la longueur totale des séries (3 ans) ne nous donnait accès qu'aux périodes allant de 2 mois à 1 an environ. De même que pour les tendances, une décision s'imposait quant aux descripteurs qui seraient considérés comme les « variables » du problème: paramètres hydrobiologiques, stations, ou les deux à la fois.

Le premier choix revenait à privilégier les relations fonctionnelles entre paramètres hydrobiologiques en un point donné des bassins, et à effectuer donc une analyse

multispectrale des résidus des huit variables en chaque station prise séparément. Cette approche a en fait l'inconvénient d'ignorer les effets de l'advection et de la diffusion dans les perturbations locales des paramètres. Le second choix est donc préférable, qui met au contraire l'accent sur les relations temporelles entre stations pour un paramètre donné: on obtient ainsi une première appréciation du chemin suivi par des perturbations par rapport au niveau moyen du paramètre en question. Dans le cas du P.A.H., où il existe plusieurs sources potentielles de perturbations, différentes de surcroît d'un paramètre hydrobiologique à l'autre, étant donné la variété des contenus des rejets industriels, l'étude des relations temporelles entre stations aura de plus l'avantage de localiser, au travers des avances de phase, les sources majeures de perturbations de chaque paramètre. Quant au troisième choix, il nous a semblé alourdir inutilement le traitement, du fait même que nous avons considéré les résidus plutôt comme traceurs de l'écoulement que comme moyens d'investigation des relations fonctionnelles entre paramètres hydrobiologiques, abordées au travers des tendances.

Finalement, deux paramètres hydrobiologiques seront pris ici, du fait de la lisibilité relativement bonne de leur traitement, comme exemples de la richesse d'informations que peut fournir l'analyse multispectrale des perturbations par rapport aux tendances: il s'agira de la température puis des phosphates. Pour un paramètre hydrobiologique et à une fréquence donnée, l'analyse multivariable sera donc faite sur la matrice spectrale des signaux résiduels aux diverses stations du P.A.H., normée ainsi qu'il a été dit au 2<sup>e</sup> paragraphe. La notion de « variable » utilisée dans la description méthodologique du paragraphe précédent, s'appliquera donc ici aux stations. Par ailleurs, spectres et spectres croisés ont été lissés par une fenêtre de Parzen d'ouverture 16, ce qui, conjugué avec la brièveté des séries, n'autorisera qu'une séparation rudimentaire des fréquences.

Malgré le petit nombre de fréquences accessibles du fait de l'échantillonnage, il serait fastidieux d'examiner le

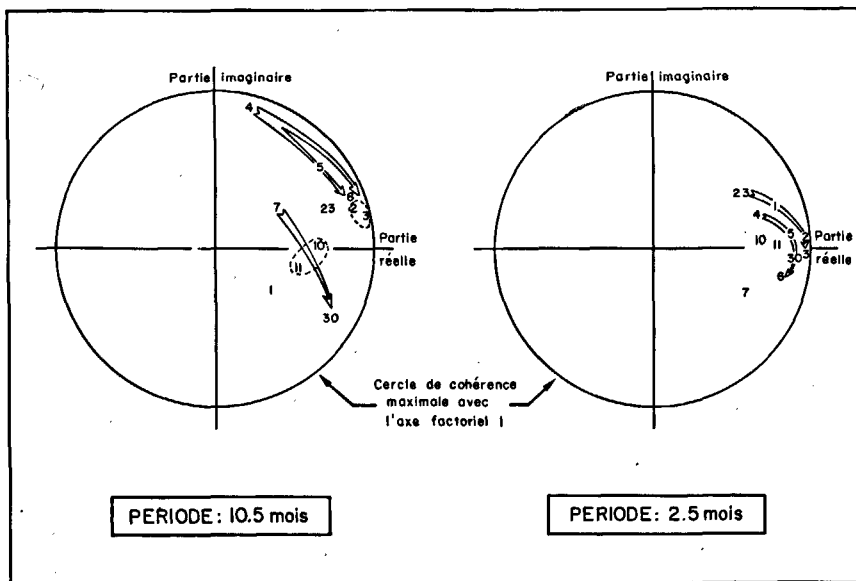


**Figure 6**  
 Origine et propagation des perturbations thermiques dans les bassins du PAH. Étude à deux fréquences particulières.  
*Origin and propagation of thermal perturbations in the PAH flooding docks. Study at two particular frequencies.*

résultat de l'analyse en composantes principales effectuée à chaque fréquence; de plus, certaines d'entre elles apparaissent difficilement interprétables. Un choix arbitraire de deux bandes de fréquences relativement bien structurées a donc été fait. Sur la figure 6 par exemple, sont visualisées les relations temporelles entre les perturbations thermiques aux différentes stations pour des périodes de 6 mois et plus d'une part, ainsi que pour des périodes de 2,5 mois environ d'autre part (du fait du phénomène de contamination, les résultats nominalement attachés à cette période de 2,5 mois sont vraisemblablement affectés par des fréquences supérieures, et notamment la première fréquence contaminante, associée à une période de 1,66 mois. Eu égard, par ailleurs, au lissage opéré par les fenêtres spectrales, on voit qu'une identification fine de la fréquence serait illusoire). Plusieurs sources de perturbations thermiques apparaissent, ainsi que des séquences de propagation associées, dont on peut partiellement identifier la cause grâce au descriptif des implantations industrielles et de leurs rejets moyens fourni par le P.A.H. Par exemple, pour des périodes de 6 mois et plus, les déphasages indiquent clairement que la source majeure de fluctuation se trouve dans le Canal Central Maritime

(stations 7 et 10): entre ces deux stations existe effectivement un rejet de réfrigération d'une usine d'engrais azotés dont on peut donc, malgré l'absence de mesures, supposer des variations lentes de débit de pompage. Vers les périodes analysables les plus courtes (2 mois environ), on peut suivre nettement la propagation de perturbations nées au niveau des stations 5 et 6, atteignant d'abord la station 4, puis l'ensemble des bassins et les stations hors bassin: l'origine en est probablement le seul rejet thermique recensé dans le canal de Tancarville causé par une usine de matières plastiques; il apparaît comme assez irrégulier puisque riche en « hautes » fréquences.

De même, sur la figure 7, plusieurs sources de perturbations de la teneur en phosphates peuvent être identifiées. Deux pôles générateurs de perturbations à moyennes et longues périodes (entre 3 et 12 mois) semblent s'individualiser sur le schéma de gauche: la zone de la station 4, donc probablement les apports de la Lézarde, dont on peut suivre l'écoulement vers les stations 5 et 6, et la zone de la station 7, donc probablement le rejet de phosphogypse existant dans cette zone, ainsi que les eaux de lavage des appontements où est déchargé le minerai de phosphate. On remarquera



**Figure 7**  
 Origine et propagation des perturbations du contenu en phosphates des bassins du PAH. Étude à deux fréquences particulières.  
*Origin and propagation of perturbations of the PO<sub>4</sub> content of the PAH flooding docks. Study at two particular frequencies.*



la relative indépendance de ces deux sources, la première étant fortement cohérente avec l'axe 1, la seconde l'étant plutôt avec un axe d'ordre plus élevé. Si les apports de la Lézarde présentent encore une composante « hautes » fréquences, on peut voir sur le schéma de droite que ce sont cette fois les entrées d'eau marine aux écluses qui sont les principales causes de rapides fluctuations dans le sud des bassins (séquences 23 → 1 → 2 → 3, 11 → 7). La station 30 pose toutefois un problème d'interprétation, qui vient rappeler la prudence nécessaire à toute analyse au voisinage de la fréquence de Nyquist.

## CONCLUSIONS

L'analyse des traits majeurs de l'hydrobiologie des bassins du Port Autonome du Havre a fourni l'occasion de rappeler et d'illustrer quelques considérations fondamentales en analyse des séries temporelles multivariées :

— ainsi que le recommandaient déjà Jenkins et Watts (1968) et Legendre et Legendre (1979 *a* et *b*), l'analyse des séries temporelles uni- ou multivariées a tout à gagner de l'extraction préalable de la tendance et de tout phénomène strictement périodique dont l'existence est connue *a priori*; « ossature » du signal, la tendance doit toujours faire l'objet d'une étude séparée appropriée et les techniques classiques d'analyse multivariée peuvent alors s'avérer très adaptées à l'étude synthétique des composantes de la tendance ;

— afin d'exploiter toute la richesse des séries temporelles, l'analyse des résidus du signal par rapport à la tendance devra être effectuée bande de fréquences par bande de fréquences, ce qui est le but de l'analyse spectrale classique. Afin cependant de sauvegarder l'appréciation

synthétique des relations existant entre les différentes séries temporelles d'un signal multivariable, l'analyse spectrale peut être utilement complétée par une technique d'analyse multivariable des matrices de spectres et cospectres dans chaque bande de fréquences (Laurec, 1979) ;

— enfin, il n'est pas inutile de rappeler, avec Jenkins et Watts (1968), combien la finesse souhaitée, en particulier au niveau de l'analyse temporelle, conditionne la stratégie d'échantillonnage : le pas temporel d'échantillonnage fixera la fréquence de Nyquist, au-delà de laquelle toute étude du signal sera impossible ; la longueur totale du signal fixera la finesse de séparation des fréquences.

L'application de cette stratégie d'étude aux données du P.A.H. a permis d'abord, par l'étude des composantes de la tendance, de dégager un gradient hydrobiologique pérenne, un enrichissement continu en sels nutritifs de tous les bassins de 1977 à 1979 et un cycle saisonnier classique comparable à celui des eaux marines avoisinantes. Dans un deuxième temps, l'analyse des résidus a permis de tracer une première esquisse de l'origine et de la propagation de perturbations de différentes fréquences affectant les niveaux moyens de quelques paramètres. Finalement, on peut dire que l'ensemble de la méthode a permis de dresser une première image des relations spatio-temporelles liant les diverses variables hydrobiologiques dans ces bassins.

## Remerciements

Cette étude a été subventionnée par le Port Autonome du Havre (contrat CNEXO-PAH 79-6112-F).

Par ailleurs, nous remercions vivement M<sup>me</sup> Marguerite Breton (CNEXO-COB) pour la part importante qu'elle a prise dans les aspects informatiques de ce travail.

## RÉFÉRENCES

- Anderson T. W., 1971. *The statistical analysis of time series*, Wiley and sons, New York, 374 p.
- Anonyme, 1978. Site EDF de Paluel. Étude écologique de projet 2<sup>e</sup> cycle, Rapport général CNEXO-COB-ELGMM, Lab. Municipal de Rouen, 434 p.
- Anonyme, 1980. Site EDF du Penly. Étude écologique de projet 1<sup>er</sup> cycle, Rapport général CNEXO-COB-ELGMM.
- Brillinger D. R., 1975. *Time series data analysis and theory*, International series in decision processes, Holt Rinehart, Winston, 500 p.
- Guillaud J. F., 1980. *Étude hydrobiologique de l'estuaire de la Seine (étude des sels nutritifs)*, Agence Financière de Bassin Seine-Normandie CNEXO-COB-ELGMM, 98 p.
- Jenkins G. M., Watts D. G., 1968. *Spectral analysis and its applications*, Holden Day, San Francisco, 525 p.
- Jones R. H., 1962. Spectral analysis with regularly missed observations, *Annal. Math. Stat.*, **33**, 455-461.
- Jones R. H., 1971. Spectrum estimation with missing observations, *Annal. Inst. Math. Stat.*, **23**, 387-398.
- Laurec A., 1979. Analyse des données et modèles prévisionnels en écologie marine, *Thèse Doct. État, Univ. Aix-Marseille II*, 405 p.
- Laurec A., 1982. Traitement des signaux quantitatifs et implications dans l'échantillonnage, in: *Stratégies d'échantillonnage en écologie*, édité par S. Frontier, Collection d'Écologie n° 17, Masson, Paris et Presses de l'Université Laval, Québec, 217-270.
- Laurec A., Le Gall J. Y., 1975. De-seasonalizing of the abundance index of a species. Application to the albacore (*Thunnus alalunga*) monthly catch per unit of effort (c.p.u.e.) by the Atlantic Japanese longline fishery, *Bull. Far. Seas Res. Lab.*, **12**, 145-169.
- Legendre L., Legendre P., 1979 *a*. *Écologie numérique. Tome 1 : Le traitement multiple des données écologiques*, Coll. Écologie n° 12, Masson, Paris et Presses de l'Université, Laval, Québec, 197 p.
- Legendre L., Legendre P., 1979 *b*. *Écologie numérique. Tome 2 : La structure des données écologiques*, Coll. Écologie n° 13, Masson, Paris et Presses de l'Université Laval, Québec, 254 p.
- Wallace J. M., Dickinson R. E., 1972. Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations, *J. Appl. Meteorol.*, **11**, 6, 887-892.