

Application de l'analyse spectrale par maximisation de l'entropie dans la description des séries brèves en écologie marine

Analyse spectrale
Entropie maximale
Séries chronologiques
Système pélagique
Pollution

Spectral analysis
Chronological series
Maximum entropy
Pelagical system
Pollution

Dominique Calmet^a, Sabine Charmasson^a, François Blanc^b

^a Commissariat à l'Énergie Atomique, Service de Protection Sanitaire, Centre d'Études Nucléaires, B.P. n° 6, 92260 Fontenay-aux-Roses, France.

^b Centre d'Océanologie de Marseille, U.R.A. 41, Campus de Luminy, Case 901, 13288 Marseille Cedex 9, France.

Reçu le 14/5/82, révisé le 8/6/83, accepté le 11/7/83.

RÉSUMÉ

Dans ce travail, l'analyse spectrale par maximisation de l'entropie (A.S.M.E.) est brièvement présentée, et ses propriétés sont comparées à celles de l'analyse spectrale suivant Fourier. Il ressort des conditions d'application que la première analyse est parfaitement adaptée aux séries chronologiques courtes très souvent recueillies en écologie marine. Le principal point faible de cette méthode réside dans la détermination objective du nombre M de coefficients définissant le processus autorégressif sous-jacent. Cependant, l'utilisation d'une valeur de M proche de $2N/\text{Log}2N$ et pour laquelle l'erreur finale prédite d'Akaike passe par un minimum, fournit des estimations spectrales satisfaisantes.

L'application de l'analyse spectrale par maximisation de l'entropie à des séries hydrologiques et biologiques, définies par $N = 58$ points, décrivant un système marin perturbé par des rejets anthropiques (émissaire urbain) permet de cerner l'impact temporel des eaux polluées sur le système pélagique concerné. Deux composantes sont clairement mises en évidence par l'A.S.M.E.; l'une décrit les évolutions tendanciennes journalières, l'autre, de périodicité plus courte, de l'ordre de six heures, correspond aux heures charnières de l'activité urbano-industrielle de l'agglomération marseillaise.

Oceanol. Acta, 1984, 7, 1, 77-86.

ABSTRACT

Maximum entropy spectral analysis applied to short series description in marine ecology

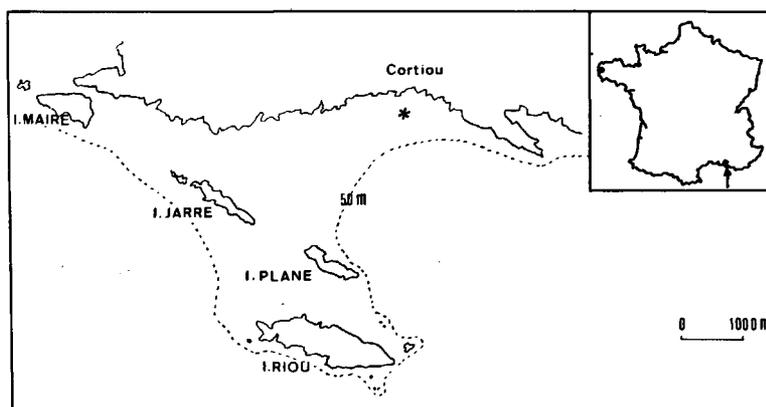
The mathematical theory of maximum entropy spectral analysis (MESA) is briefly explained and compared with the spectral analysis defined by Fourier (ASF). The conditions of application show that the former analysis is perfectly adapted to short chronological series, most often observed in marine ecology.

The main drawback of this method lies in the determination of the number "M" of coefficients that defines the underlying autoregressive model. However, with the use of an operator length close to $2N/\text{Log}2N$, for which the Akaike final prediction error reaches a minimum value, good spectral estimates are obtained. When applied to hydrological and biological series defined by $N = 58$ points and describing a marine system polluted by sewage (urban outfall of Marseilles), MESA makes it possible to assess the temporal impact of polluted waters on the local pelagical system. Two components are thrown into relief by the MESA exercise: one describes the daily trend; the other, characterized by a periodicity of about 6 hours, reflects the peak hours of urban and industrial activities in Marseilles.

Oceanol. Acta, 1984, 7, 1, 77-86.

INTRODUCTION

Le développement actuel en océanographie des enregistrements en continu de caractères physiques comme la vitesse et la direction des courants, la turbidité ou la température, de caractères chimiques comme la salinité et les différents sels minéraux et de paramètres biologiques comme le phytoplancton et le zooplancton, aboutit à des séries spatio-temporelles définies par plusieurs centaines de mesures ($N > 100$) (Glover, 1967; Cruzado, Kelley, 1973; Denman, Platt, 1975; Ibanez, 1975; Platt, Denman, 1975; Prieur, 1980; Colebrook, 1981). Cependant, aujourd'hui encore, nombreux sont les cas où les processus naturels sont décrits par l'intermédiaire de séries temporelles et/ou spatiales constituées d'un nombre restreint d'observations. Ainsi la description de cycles annuels de nombreux paramètres hydrologiques et biologiques repose sur des mesures mensuelles ou bimensuelles (Frontier, 1974; Dessier, Laurec, 1978), constituant au mieux des séries annuelles de 24 points, cas classique des programmes de surveillance de l'environnement. De même, les évolutions journalières de ces mêmes caractères sont souvent appréciées au travers de prélèvements horaires ou bihoraires sur un cycle de 24 heures, soit donc par l'intermédiaire de séries définies par 48 points. C'est ainsi que le groupe d'études des systèmes pélagiques du Centre d'Océanographie de Marseille a été amené à cerner l'influence des rejets d'eaux résiduaires de l'agglomération marseillaise sur les peuplements planctoniques littoraux. L'étude, conduite dans un cadre multivarié, repose sur l'analyse de 62 caractères écologiques mesurés sur des prélèvements d'eau effectués toutes les 25 minutes pendant 24 heures en un point fixe situé à environ 1 mille de l'émissaire de rejet (fig. 1).



L'analyse spectrale suivant Fourier constitue l'outil privilégié et justifié de l'étude des séries temporelles définies par un grand nombre de points; une nouvelle méthode dite d'analyse spectrale par maximisation de l'entropie ou de l'information, mieux adaptée aux séries courtes, a récemment été présentée (Burg, 1967; 1968; 1970; Ables, 1974). Son utilisation en géophysique et en astrophysique a clairement mis en évidence ses performances (Ulrych, 1972; Ponsonby, 1973; Wells, Chinnery, 1973), et en biologie, son application aux données de pêche a connu quelque succès (van Winkle *et al.*, 1979).

Le but du présent travail est de faire ressortir l'intérêt de l'application de l'A.S.M.E. à la description des séries écologiques limitées en nombre de points, ainsi que comme test d'hypothèse quant à la formulation correcte des différentes composantes évolutives des séries analysées.

Nous présenterons uniquement les résultats des deux types d'analyses spectrales appliquées à cinq paramètres représentatifs des différentes composantes temporelles mises en évidence sur l'ensemble des caractères mesurés dans cette étude (Calmet, Blanc, 1980), à savoir la salinité, l'oxygène, l'ammoniaque, les comptages phytoplanctoniques totaux et de *Rhizosolenia fragilissima*.

MÉTHODES D'ANALYSES

Les hypothèses

L'idée de base de la théorie statistique des séries temporelles en temps discret $x(t)$, obtenues suivant un pas d'échantillonnage constant au cours du temps Δt , est de considérer ces séries comme des réalisations représentatives d'un ensemble de variables aléatoires $\{X(t)\}$, $t = 0, \pm 1, \pm 2, \dots, \pm N$. Cet ensemble de variables aléatoires et leurs distributions de probabilité associées définissent un processus stochastique. L'approche mathématique de ce dernier nécessite le respect de différentes hypothèses, en particulier de linéarité, de normalité et de stationnarité. La notion de linéarité a bien été développée par Box et Jenkins (1970). Cette propriété est fondamentale notamment dans le cadre des processus autorégressifs. La normalité est une hypothèse également très importante, puisque les inférences statistiques classiques en dépendent. Les

Figure 1

Localisation du point de prélèvement eulérien situé à proximité de l'émissaire de Marseille-Cortiou (France), dont l'échantillonnage a été effectué, toutes les 25 minutes, pendant 24 heures du 17 mai 1978 (12h) au 18 mai 1978 (11 h 45).

Location of the eulerian sampling station near the sewage outfall of Marseilles-Cortiou (France). Sampling was carried out every 25 minutes from May 17 1978 (noon) until May 1978 (11.45 a. m.).

stationnarités du premier et du second ordre seront atteintes lorsque les propriétés statistiques des données, moyenne et variance, seront rigoureusement indépendantes d'une translation sur l'échelle des temps. La stationnarité du premier ordre sera obtenue par la suppression de la dérive, ou tendance, des données sous la forme d'un polynôme ou de composantes périodiques de basses fréquences.

Les analyses spectrales

L'analyse des séries chronologiques repose sur un ensemble de méthodes mathématiques dont les analyses

spectrales représentent un mode de traitement particulièrement heuristique. Nous allons présenter sommairement deux de ces analyses en renvoyant le lecteur aux principaux travaux théoriques les concernant; par contre nous insisterons sur les contraintes respectives liées à l'emploi des deux méthodes.

Classiquement, l'étude mathématique des processus naturels définis par plusieurs centaines d'observations (N) repose en partie dans l'espace des temps sur l'utilisation des fonctions d'autocovariance et d'autocorrélation (Blackman, Tukey, 1958; Jenkins, Watts, 1968) et, dans l'espace des fréquences, sur le calcul des spectres de puissance suivant la méthode des périodogrammes (Jones, 1965) ou de la transformée de Fourier (A.S.F.) de la fonction d'autocorrélation (ρ_n). Ainsi, les estimations conventionnelles du spectre de puissance P_e , lissé par un opérateur linéaire ω_n , seront de la forme :

$$P_e = 1/(2N-1) \sum_{n=-N+1}^{N-1} \omega_n \rho_n e^{-i2\pi f n \Delta t}$$

avec $-1/2 \Delta t < f < 1/2 \Delta t$ et ρ_n la fonction d'autocorrélation et où

$$W = 1/(2N-1) \sum_{n=-N+1}^{N-1} \omega_n e^{-i2\pi f n \Delta t}$$

est appelée fenêtre spectrale.

La fenêtre spectrale vise à établir la plus grande stabilité des caractéristiques spectrales en limitant les phénomènes de contamination entre les fréquences voisines sans trop abaisser le niveau de résolution des spectres. L'utilisateur de l'A.S.F. est donc conduit à un premier choix : celui de ce que les anglo-saxons nomment « la charpente » de la fenêtre spectrale. Ce premier choix devra être complété par un second, de ce qu'il a été convenu d'appeler de façon toujours imagée : « l'ouverture » de la fenêtre spectrale. Cette double approche est le plus souvent menée empiriquement et dépendra donc du savoir-faire de l'utilisateur. Ainsi Laurec et Blanc (1974) déterminent expérimentalement que la fenêtre de Parzen semble la mieux adaptée à la nature des données océanographiques qu'ils ont à analyser. De même, l'ouverture de la fenêtre spectrale est le plus souvent choisie en opérant un lissage progressif des estimations spectrales. Jenkins et Watts (1968) mettent cependant en évidence les bons résultats obtenus pour une ouverture égale à $2N/3$.

De plus, les méthodes classiques d'analyse spectrale sont basées sur des hypothèses irréalistes concernant l'extrapolation des données en dehors de l'intervalle échantillonné. Ainsi, la méthode des périodogrammes (Jones, 1965) requiert une extension périodique des données avec une période égale à la longueur du domaine d'observation, tandis que l'approche par les autocorrélations (Blackman, Tukey, 1958) nécessite une extension avec des données nulles. Ces deux hypothèses très souvent irréalistes conduisent, dans le cas des séries brèves, à une limitation de la résolution des spectres associés, en particulier aux plus basses fréquences.

Le non respect des conditions d'application d'analyse, l'extrapolation des données en dehors de l'intervalle échantillonné discrétisé par un nombre trop faible

d'observations ($N < 50$) ou un mauvais choix de la charpente et de l'ouverture de la fenêtre spectrale, peuvent conduire à des interprétations erronées des spectres. Afin de se dégager de ces deux dernières contraintes, Burg (1967; 1972; 1975) proposa l'analyse spectrale par maximisation de l'entropie (A.S.M.E.).

Sachant que l'entropie d'une série temporelle de distribution gaussienne est proportionnelle à

$$J = \int_{-f_N}^{f_N} \text{Log } P_e(2\pi f \Delta t) df,$$

où $f_N = 1/2\Delta t$ représente la fréquence de Nyquist. La meilleure estimation du spectre doit se rapprocher de celle d'un « bruit blanc », c'est-à-dire d'un spectre où l'énergie sera distribuée uniformément sur chacune des fréquences estimées, ce qui correspond au cas où l'entropie est maximale, d'où le nom de la méthode. Burg proposa donc de choisir une estimation sous la forme d'une fonction ($P_e(2\pi f \Delta t)$) maximisant l'expression J soumise aux contraintes

$$\rho_n = \int_{-f_N}^{f_N} P_e(2\pi f \Delta t) e^{-i2\pi f n \Delta t} df$$

avec $n=0, 1, \dots, N-1$.

où ρ_n est la fonction d'autocorrélation associée à la série échantillonnée. Si l'on suppose alors que $F(2\pi f \Delta t)$ est la réponse dans le domaine des fréquences d'un filtre linéaire, et que $|F(2\pi f \Delta t)|^2 = P_e(2\pi f \Delta t)$, alors J correspond à la perte d'entropie qui résulte du passage d'un processus gaussien au travers du filtre (Bartlett, 1956). Choisir P_e maximal correspond à une hypothèse pessimiste, puisque cela revient à trouver le filtre qui réduit l'entropie et qui pourrait générer un processus à partir d'un bruit blanc, processus caractérisé par la fonction de corrélation observée (ρ_n) (Lacoss, 1971).

Par la suite, Akaike (1969a), Van den Bos (1971), Ulrych et Bishop (1975), ont démontré que le calcul du spectre par maximisation de l'entropie d'une série est équivalent à l'ajustement au sens des moindres carrés d'un modèle autorégressif. L'estimation spectrale est alors calculée à partir des coefficients de régression par de nouveaux algorithmes. Ce dernier point nous conduit à discuter de l'utilité du modèle autorégressif dans le cas des données océanographiques, de la validité des méthodes d'estimation des coefficients des modèles et, enfin, de la détermination du nombre de coefficients nécessaires et suffisants pour obtenir les meilleures estimations de la série initiale.

En nous plaçant dans un cadre résolument descriptif, il n'est pas interdit de représenter un processus stationnaire par une fonction linéaire simple de son propre passé, ce qui, par exemple, peut se traduire par un modèle autorégressif en domaine fini du type :

$$x_t - a_1 x_{t-1} - \dots - a_m x_{t-m} = \varepsilon_t(M),$$

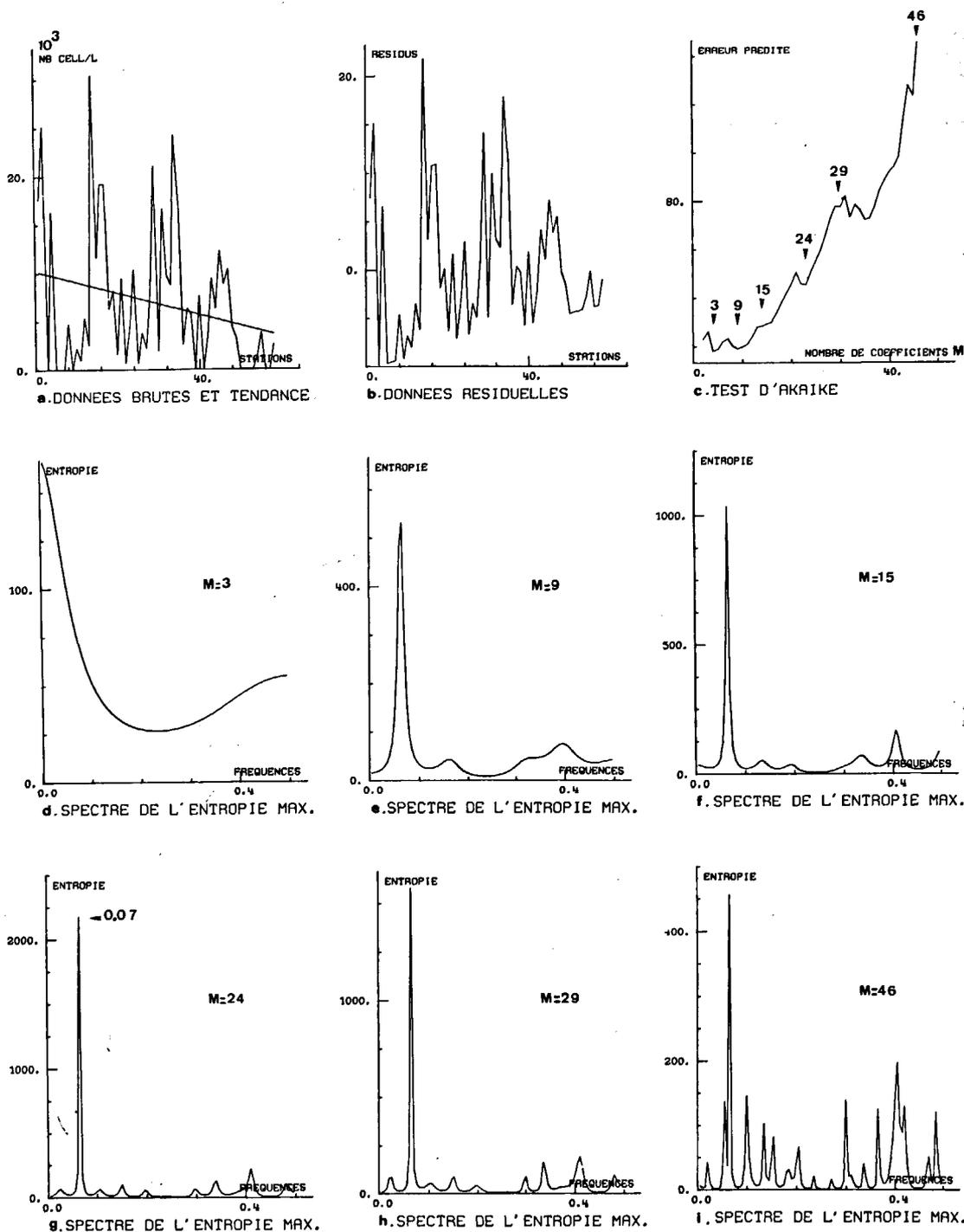
où ε_t représente l'erreur résiduelle de la série prédite à partir de la série mesurée. Cette valeur ε_t devra être minimisée. En ce qui concerne les méthodes d'estimation des coefficients a_1, \dots, a_m du modèle, nous renvoyons le

lecteur aux travaux de Durbin (1960), qui présente une étude comparative visant à établir la qualité des estimations et, en particulier, celle des moindres carrés pour les modèles autorégressifs simples. Reste l'estimation du nombre de coefficients M ; ce point est fondamental quant à la validité des résultats obtenus. En effet, une valeur trop faible du nombre de coefficients entraîne un spectre très lissé avec une distribution de l'entropie étalée sur de larges bandes de fréquences, tandis que des valeurs trop élevées renforcent exagérément l'entropie associée à certaines fréquences. Berryman (1978) passe en revue les différents tests possibles permettant de déterminer suivant des critères objectifs la valeur optimale de M , et propose, après une étude empirique, une valeur de $M = 2N/\text{Log } 2N$. Ulrych et Bishop (1975), puis Van Winkle *et al.* (1979) sont

amenés à proposer $M = N/2$ pour des valeurs de N comprises entre 20 et 40. Ils proposent également de procéder à des estimations successives du spectre en ajustant des modèles autorégressifs d'ordre croissant en leur associant un calcul de l'erreur prédite finale (E.P.F.), définie par Akaike (1969 *b*; 1970) comme étant la valeur moyenne du carré de la différence entre les valeurs mesurées et celles prédites. Pour notre part, après de nombreux tests, nous avons choisi les valeurs de M qui correspondent à une E.P.F. minimale pour des valeurs proches de $2N/\text{Log } 2N$ (fig. 2). L'A.S.M.E. ne nécessite donc aucune définition de fenêtre spectrale, et apparaît comme bien adaptée aux séries courtes alors qu'elle perd tout intérêt pour les séries longues. Kirk *et al.* (1979) font remarquer que les plus longues périodes prédites par l'A.S.F. sont égales

Figure 2

Détermination du nombre optimal, M , de coefficients sur la série résiduelle (b) des effectifs de *Rhizosolenia fragilissima* après suppression d'une tendance linéaire (a). Le tracé de l'erreur prédite finale d'Akaike en fonction de M (c) montre des discontinuités pour les valeurs de M égales à 3, 9, 15, 24 et celles proches de 30. Les estimations spectrales apparaissent lissées pour les valeurs les plus faibles de M (d, e, f) tandis que des pics parasites sont générés pour les valeurs les plus élevées (i). Le spectre défini par $M = 2N/\text{Log } 2N = 24$ (g) présente une entropie maximale par rapport à celui défini par $M = N/2 = 29$ (h).



Determination of the optimal number "M" of coefficients for the residual *Rhizosolenia fragilissima* series (b) after removal of the linear trend (a). The Akaike final prediction error plotted in terms of "M" shows some discontinuities when $M=3, 9, 15, 24$ and when "M" approximates 30 (c). Spectral estimates appear smoothed for the lowest M values (d, e, f) whereas spurious peaks appear for the highest M values (i). The spectrum computed with $M = 2N/\text{Log } 2N = 24$ (g) shows a higher entropy in comparison with that computed with $M = N/2 = 29$ (h).

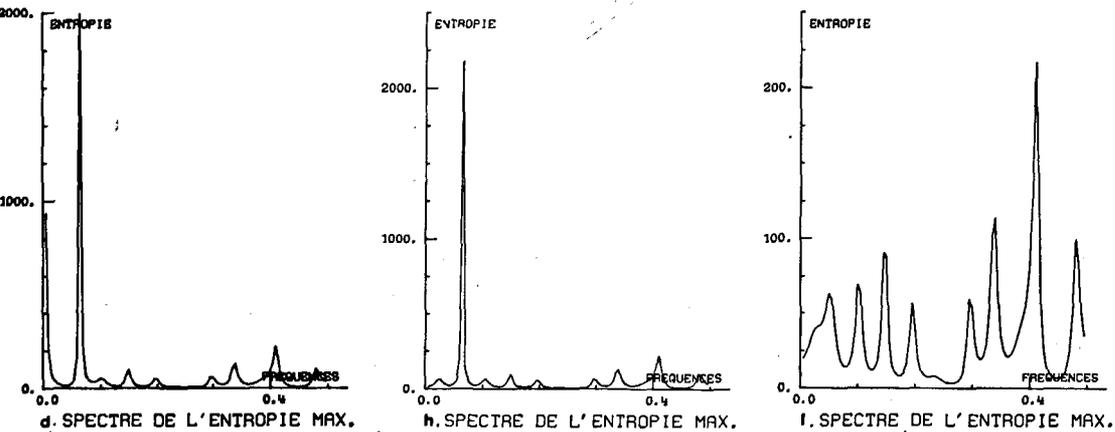
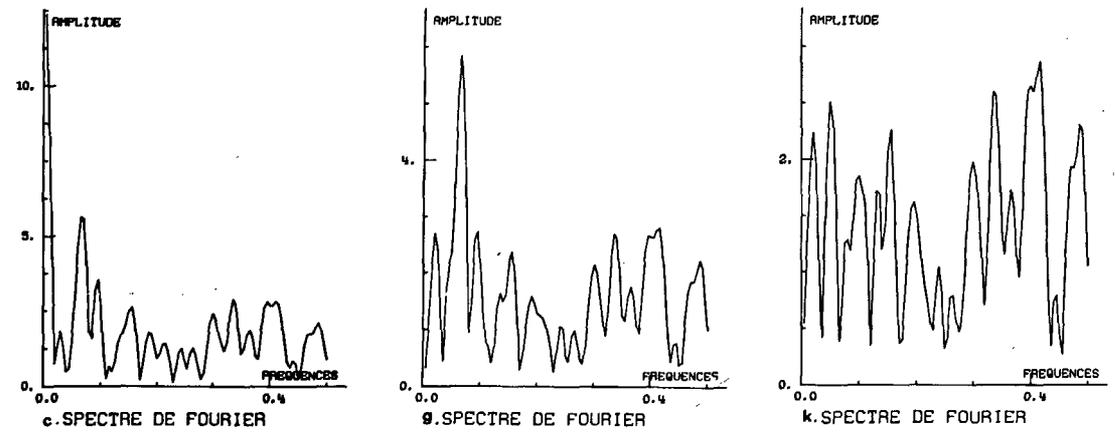
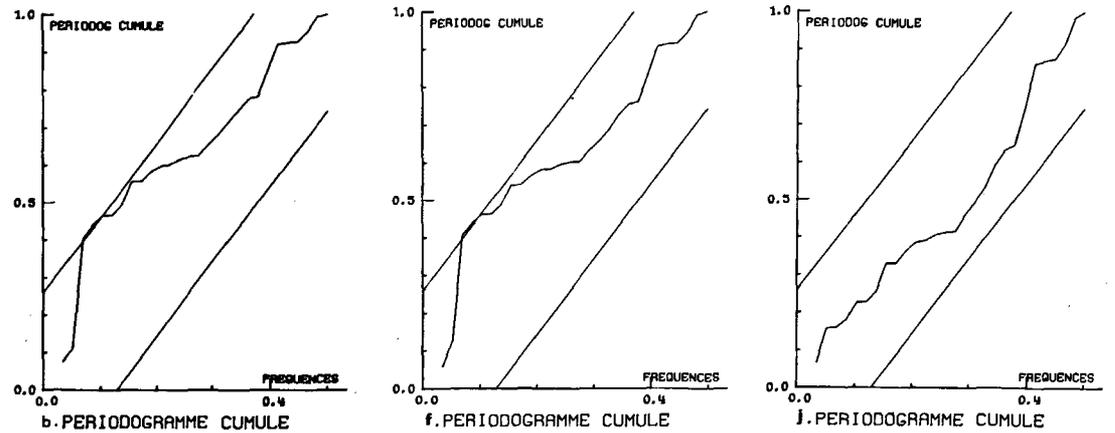
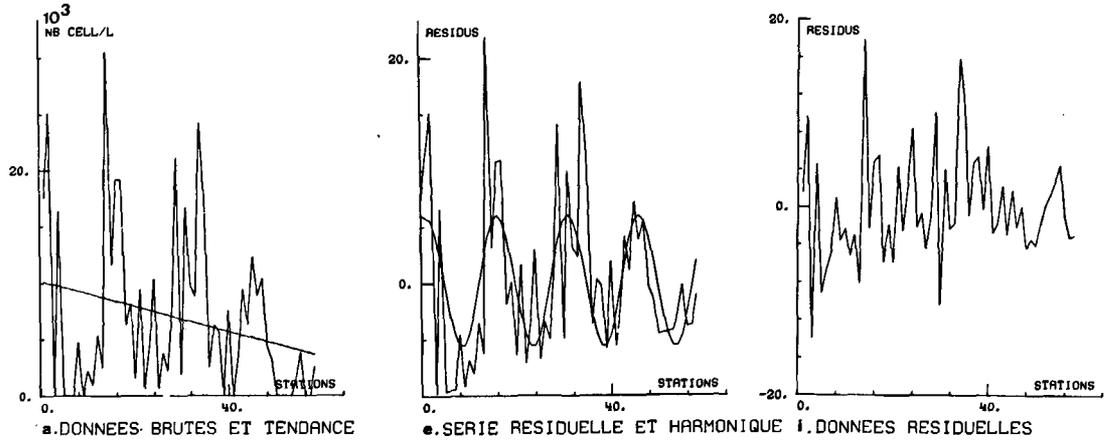


Figure 3
Opération de « blanchissage » sur la série de *Rhizosolenia fragilissima* (a) par élimination successive des composantes mises en évidence par les analyses spectrales (c, d). Les spectres suivant Fourier (g) et par maximum d'entropie (h) de la série résiduelle (e) obtenue par suppression d'une tendance linéaire (a) ne développent pas d'énergie à l'origine. Ils révèlent tous deux une importante distribution d'énergie pour des fréquences proches de 0,07, qui reste significative sur le périodogramme cumulé (b, f). La suppression d'une fonction sinusoïdale de la forme $y_n = 158 + 5,819 \sin(2\pi/15 x_n + 1,143)$ (e) conduit à une série résiduelle (i) dont le périodogramme cumulé (j) ainsi que les spectres (k, l) sont caractéristiques d'un bruit blanc, avec cependant une légère remontée d'énergie pour les fréquences supérieures à 0,3.

Whitening of the *Rhizosolenia fragilissima* data (a) through successive removal of the components demonstrated by spectral analyses (c, d). No peak appears at the origin of Fourier (g) and maximum entropy (h) spectra of the residual series (e) obtained after removal of the linear trend. Both reveal a statistically significant peak at frequency ≈ 0.07 (b, f). If a sine function: $y_n = 158 + 5.819 \sin(2\pi/15 x_n + 1.143)$ (e) is suppressed, a residual series (i) is obtained whose cumulative periodogram (j) and spectra (k, l) are typical of a white noise. None of the peaks is statistically significant; however, when frequencies are superior to 0.3, the peak amplitudes seem more important.

à la longueur des séries analysées, tandis que l'A.S.M.E. permet, dans certains cas, de prévoir des périodes supérieures à la longueur des séries analysées.

Quant à l'interprétation des estimations spectrales, il faut signaler que la distribution de l'énergie pour l'A.S.F. est proportionnelle à la hauteur des pics, tandis qu'elle est proportionnelle à la surface des pics pour l'A.S.M.E. Dans un but de simplification, les tracés chronologiques et les équations ne seront pas exprimés directement en fonction du temps mais en fonction du rang des prélèvements au cours du temps.

MÉTHODOLOGIE DE TRAITEMENT

Nous avons procédé à un « blanchissage » progressif des séries mesurées, méthode classique qui consiste à ramener les estimations spectrales à celles d'un bruit blanc par la suppression successive de chaque composante mise en évidence par les analyses spectrales. Nous avons également testé la significativité des différents pics par la méthode des périodogrammes cumulés de Fisher (Fuller, 1976) avec une limite de 0.95. Les différentes composantes significatives, qu'elles soient de basses ou de hautes fréquences, sont soustraites aux données initiales sous la forme de polynômes ou de fonctions sinusoidales, puis l'analyse est effectuée à nouveau sur les séries résiduelles (fig. 3).

Afin de limiter une remontée de l'énergie non significative liée aux plus hautes fréquences et de renforcer la distribution d'énergie aux basses fréquences, nous avons utilisé un filtre passe-bas dit des moyennes mobiles dont on contrôle l'action en repassant l'A.S.M.E. sur les séries lissées. Cette méthode, bien étudiée par Shiskin (1967), provoque pour un filtre, centré et équipondéré, de longueur p une réduction de la densité spectrale par absorption des harmoniques de fréquences supérieures à $1/p$ (fig. 4). Il est à remarquer que l'emploi d'un tel filtre fait perdre p points aux extrémités de la série, ce qui, dans le cadre des séries courtes, peut être un inconvénient qui conduit à la recherche d'un compromis entre l'effet de lissage et les conditions d'utilisation des analyses.

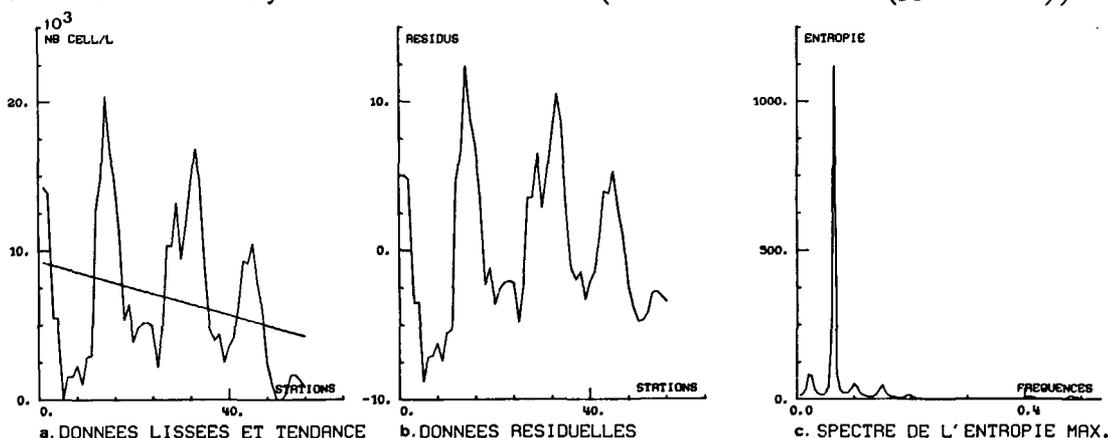


Figure 4

Effet du lissage par une moyenne mobile sur trois points : l'analyse spectrale par maximum d'entropie de la série résiduelle (b) obtenue par suppression d'une tendance linéaire sur la série lissée de *Rhizosolenia fragilissima* (a) conduit à un spectre (c) dont l'énergie associée aux fréquences supérieures à 0,3 a bien été réduite en comparaison du tracé de la figure 3 (h).

Smoothing effect of a running mean filter of order 3: MESA of the residual series (b), obtained by removal of the linear trend of the smoothed Rhizosolenia fragilissima series (a), gives a spectrum whose peaks associated with frequencies above 0.3 have been reduced in comparison with Figure 3 (h).

RÉSULTATS ET DISCUSSION

La comparaison des deux types d'analyse spectrale porte sur cinq paramètres distincts : salinité, teneurs en oxygène et en ammoniacque, peuplement phytoplanctonique total et effectif d'une espèce représentative de celui-ci : *Rhizosolenia fragilissima*. L'influence dominante des rejets d'eaux résiduaires est signalée par les valeurs d'ammoniacque les plus élevées tandis que les mesures les plus fortes des autres caractères traduisent un retour aux eaux marines les moins perturbées. Nous avons montré par un traitement de données plus général (Calmet, Blanc, 1980; Calmet *et al.*, 1982) qu'il existait une alternance régulière de ces deux types de structure, liée aux passages cycliques d'eaux polluées au point d'échantillonnage.

La difficulté de vérifier les propriétés de stationnarité des séries brèves, nous a conduit à utiliser les analyses spectrales sur les données brutes, afin de mettre en évidence l'existence d'une éventuelle tendance, parfois délicate à évaluer *a priori* sur les séries courtes. Dans ce cas, la comparaison de ces estimations spectrales avec celles associées aux séries résiduelles obtenues par suppression des différents types de tendance permet, en outre, de constater la plus ou moins bonne élimination de l'énergie à ces basses fréquences.

Les analyses spectrales portant sur ces cinq caractères révèlent l'existence commune de deux événements majeurs à des fréquences voisines de $f=0,020$ (période de 21 h) et $f=0,075$ (période de 5 h 40), et d'un nombre variable d'événements mineurs non significatifs à de plus hautes fréquences. La première composante représente la tendance de chacune des séries; elle est nettement mise en évidence par les deux types d'analyses spectrales. Elle peut être décrite sous la forme d'une fonction linéaire décroissante ($y_n = -113,158 x_n + 10193,158$) comme pour la série exprimant l'évolution de *Rhizosolenia fragilissima* (fig. 3). Cependant, comme le montre l'étude complète de la série associée à la salinité (fig. 5), l'évolution journalière peut être décrite sous la forme plus complexe d'une fonction périodique :

$$\left(y_n = 36,085 - 0,788 \sin\left(\frac{2\pi}{56}x_n + 1,399\right) \right),$$

cas de l'ammoniaque également :

$$\left(y_n = 32,849 + 13,826 \sin\left(\frac{2\pi}{56}x_n + 0,7411\right) \right),$$

ou sous la forme d'un polynôme du deuxième degré comme pour l'oxygène :

$$(y_n = 0,0004 x_n^2 - 0,0281 x_n + 6,048),$$

et du troisième degré pour le phytoplancton total ($y_n = -2,1206 x_n^3 + 139,19 x_n^2 - 2419,4 x_n + 67249,6$). Toutefois les tendances n'évoluent pas systématiquement dans le même sens ; ainsi, les mesures de salinité et de phytoplancton total décrivent une hausse nocturne tandis que les teneurs en ammoniaque augmentent en période diurne. La suppression des tendances associées à chaque série de mesures permet d'analyser des séries résiduelles stationnaires au premier ordre (fig. 6).

Les analyses spectrales suivant Fourier et par calcul de l'entropie maximale des séries résiduelles précédemment obtenues mettent bien en évidence la supériorité de définition et de résolution de cette dernière méthode, qui offre des spectres nettement moins bruités. Les estimations spectrales des séries résiduelles révèlent que les fréquences associées aux tendances ne développent plus d'énergie, tandis que celles caractérisant les rythmicités centrées sur 6 h sont prépondérantes. Cette dernière composante est à son tour traduite sous forme de fonctions sinusoïdales qui sont extraites, point à point, des séries résiduelles. Afin de vérifier que les séries résultantes de cette double opération correspondent à un bruit blanc, les analyses spectrales sont repassées une dernière fois (fig. 3). Une dernière approche, qui consiste à analyser les résultats de la « formalisation » mathématique des deux composantes précédemment décrites, a ensuite été menée. La pureté des spectres concernant la série théorique associée à l'évolution observée de la salinité met bien en évidence la finesse du pouvoir de résolution de l'A.S.M.E. (fig. 7). Comme pour les tendances, il apparaît, sur la composante de 6 h, une évolution en opposition de phase entre l'ammoniaque d'une part et la salinité (fig. 8), le phytoplancton total et *Rhizosolenia fragilissima* d'autre part. A ce propos une analyse spectrale croisée par maximisation de l'entropie apparaît nécessaire et demande à être développée puisque, comme le

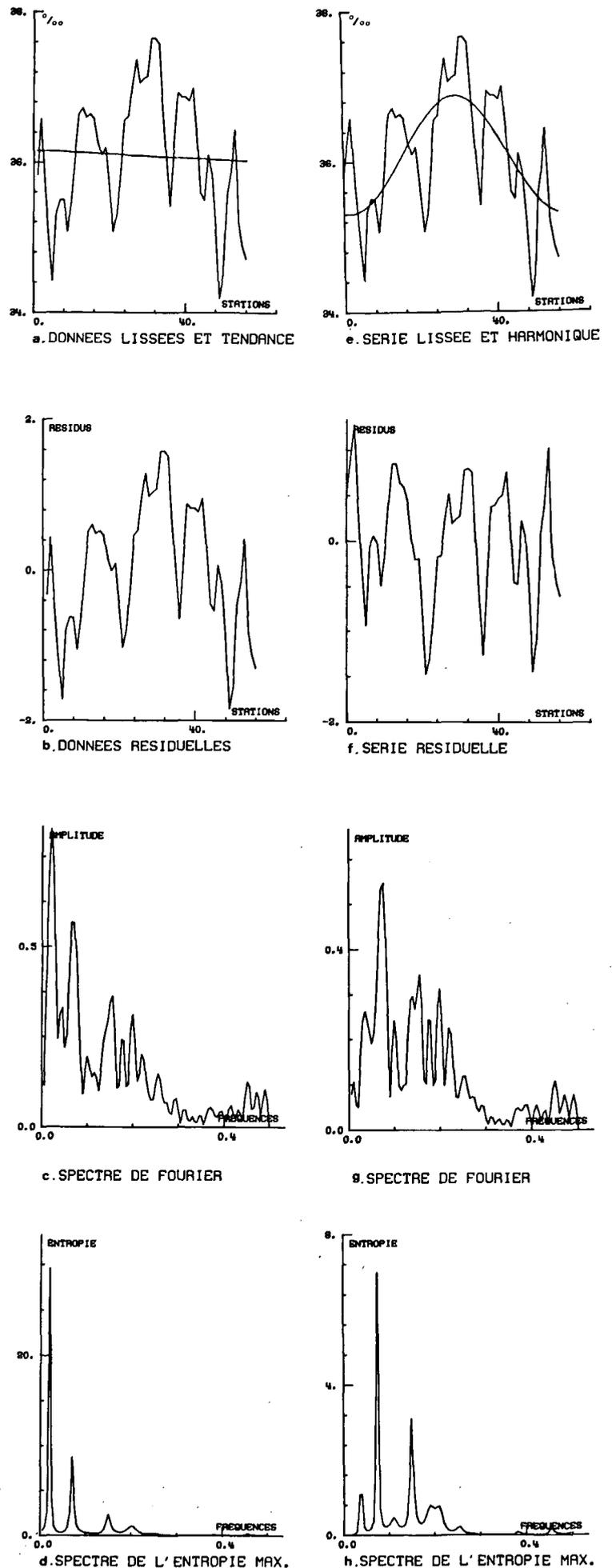


Figure 5

L'analyse spectrale comme critère de choix de la représentation mathématique de la tendance : la suppression d'une tendance linéaire sur la série de la salinité lissée par une moyenne mobile sur trois points (a), fournit une série résiduelle (b) dont les spectres (c, d) distribuent une part importante d'énergie aux plus basses fréquences, ce qui traduit une mauvaise élimination de l'effet de tendance. Par contre, les spectres (g, h) associés à la série résiduelle (f) obtenue après élimination d'une tendance sous la forme d'une fonction sinusoïdale $y_n = 36,085 - 0,788 \sin(2\pi/56 x_n + 1,399)$ (e) perdent toute distribution d'énergie à l'origine. Une composante périodique significative pour des fréquences centrées sur 0,07 est mise en évidence. Il est à remarquer ici encore que le spectre de l'entropie maximale (h) est beaucoup moins bruité que celui obtenu suivant Fourier (g).

Spectral analysis as a criterion for the selection of the mathematical formulation of the trend: removal of the linear trend of the salinity series smoothed by a running mean filter of order 3 (a) produces a residual series (b) whose spectra (c, d) show a dominant peak at the lowest frequencies. This phenomenon demonstrates bad removal of the trend effect. On the other hand, g and h spectra associated with the f residual series obtained by removal of a sine trend: $y_n = 36,085 - 0,788 \sin(2\pi/56 x_n + 1,399)$ (e) do not demonstrate any peak at their origin. A statistically significant peak appears for frequencies about 0,07. Here again, the maximum entropy spectrum is much more readable in comparison with Fourier's.

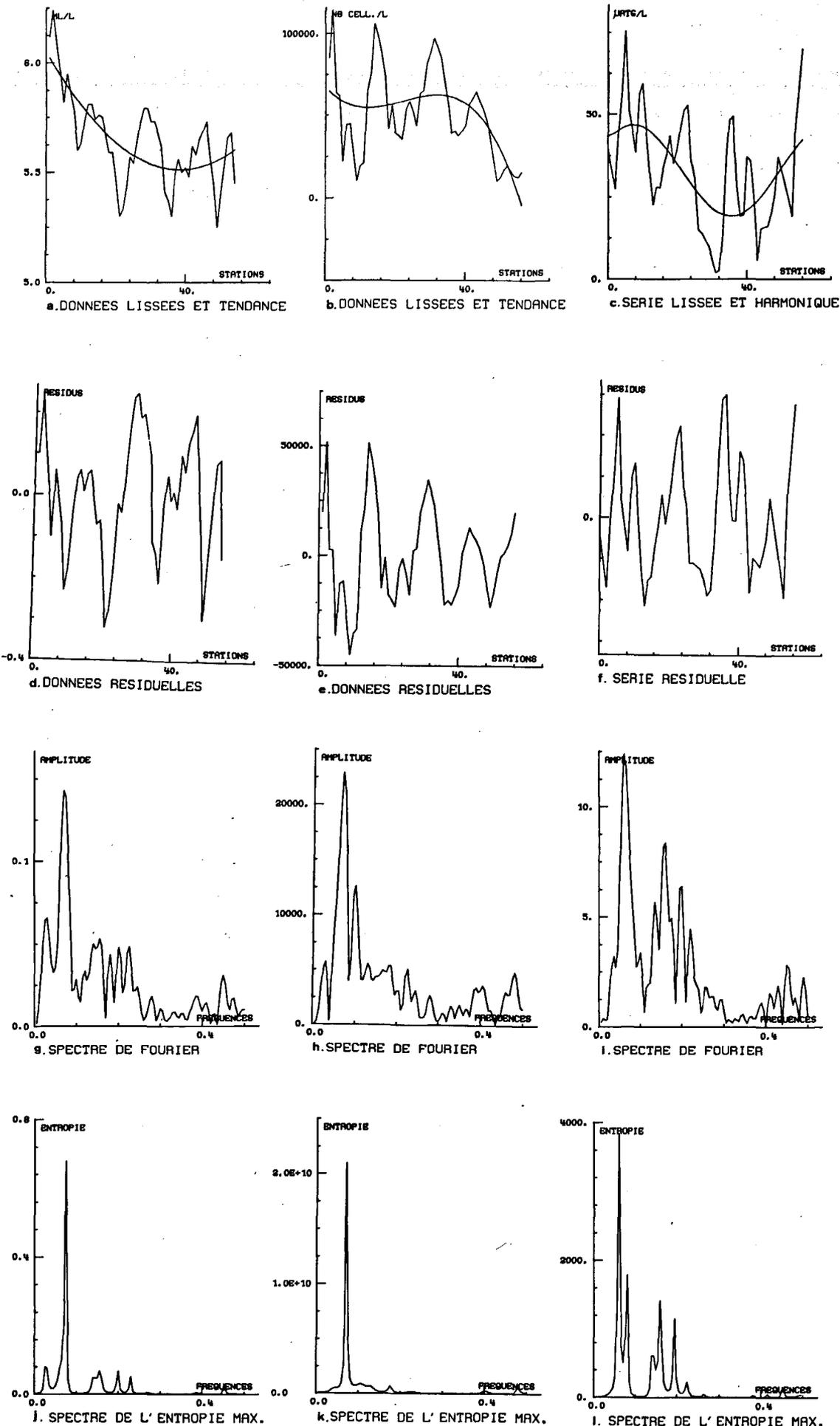


Figure 6

Les spectres (g à l) associés aux séries résiduelles de l'oxygène (d), des effectifs phytoplanctoniques totaux (e) et de l'ammoniaque (f) distribuent peu d'énergie au voisinage de l'origine ce qui traduit une bonne élimination de la tendance. Celle-ci a été traduite sous la forme d'un polynôme du second degré :

$$y_n = .0004 x_n^2 - .0281 x_n + 6,048,$$

pour l'oxygène (a), d'un polynôme du troisième degré :

$$y_n = -2,1206 x_n^3 + 139,19 x_n^2 - 2419,4 x_n + 67249,6,$$

pour le phytoplancton total (b) et d'une fonction sinusoïdale :

$$y_n = 32,849 + 13,826 \sin(2\pi/56 x_n + 0,7411)$$

pour l'ammoniaque (c).

At their origin spectra g to l associated with the residual series d, e, f (corresponding to oxygen, total phytoplankton and ammonia respectively) show little amplitude, which gives evidence of a good removal of trend. The trend is represented by a quadratic polynomial for oxygen (a):

$$y_n = 0,0004 x_n^2 - 0,0281 x_n + 6,048,$$

by a cubic polynomial for total phytoplankton (b):

$$y_n = -2,1206 x_n^3 + 139,19 x_n^2 - 2419,4 x_n + 67249,6$$

and by a sine function for ammonia (c):

$$y_n = 32,849 + 13,826 \sin(2\pi/56 x_n + 0,7411).$$

précisent Ulrych et Jensen (1974), les avantages acquis par l'A.S.M.E. sont conservés dans l'estimation des spectres croisés.

La signification écologique des deux composantes mises en évidence sur l'ensemble des séries est à rechercher dans la dynamique qui régit le mélange, au point d'échantillonnage, de deux masses d'eaux d'origines différentes. Ainsi, les eaux de la dérive liguro-provençale sont localement polluées par les eaux vannes de l'émissaire de Marseille-Cortiou, dont les volumes rejetés suivent le rythme de l'activité domestico-industrielle de l'agglomération marseillaise. Les tendances peuvent être expliquées par une forte diminution des rejets en phase nocturne, ce qui se traduit par une baisse des teneurs en divers polluants comme l'ammoniaque, et corrélativement par une augmentation de la salinité et des effectifs phytoplanctoniques. Ce schéma s'inverse pendant la période diurne. La dérive de la concentration en oxygène apparaît être la résultante particulière d'une combinaison complexe de différents processus, où interviennent à la fois les arrivées massives d'eaux polluées chaudes déficitaires en oxygène, la limitation de l'activité photosynthétique par une baisse diurne des effectifs phytoplanctoniques et un arrêt nocturne de la photosynthèse.

L'opposition des deux masses d'eaux décrites sur les tendances journalières se retrouve pour la composante de six heures qui correspond, compte tenu du temps de transit des eaux vannes au sein du réseau de collecte, aux augmentations relatives des rejets pour des heures de pointe centrées sur 6, 12, 18 et 24 heures. Celles-ci sont le reflet des variations d'activité de la population urbaine puisqu'elles correspondent respectivement au lever, déjeuner, arrêt du travail en fin d'après-midi et au coucher.

CONCLUSION

L'analyse spectrale par maximisation de l'entropie est, au niveau théorique, l'analyse qui se justifie dans le cadre de l'étude de séries limitées en nombre de points soit, comme nous l'avons vu, par des contraintes d'échantillonnage, soit par la recherche de segments stationnaires par découpage de séries plus longues, puisqu'à la différence de l'analyse spectrale suivant Fourier, elle ne nécessite aucune hypothèse sur les données en dehors de l'intervalle de la longueur de la série, ni la définition d'une fenêtre spectrale (charpente et ouverture) procurant des lissages plus ou moins importants. De plus, comme nous l'avons montré sur les séries modélisées, l'A.S.M.E. ne présente pas de contamination entre des fréquences voisines et offre une meilleure résolution spectrale. Ces deux dernières propriétés conduisent à une distribution d'énergie moins « bruitée » qui permet une interprétation immédiate des estimations spectrales.

Au niveau pratique, les principaux points faibles de cette analyse résident d'une part, dans l'absence d'un test statistique sur les estimations spectrales intégrées fournies par l'A.S.M.E. équivalent à celui de Fisher sur les périodogrammes cumulés. D'autre part, le choix,

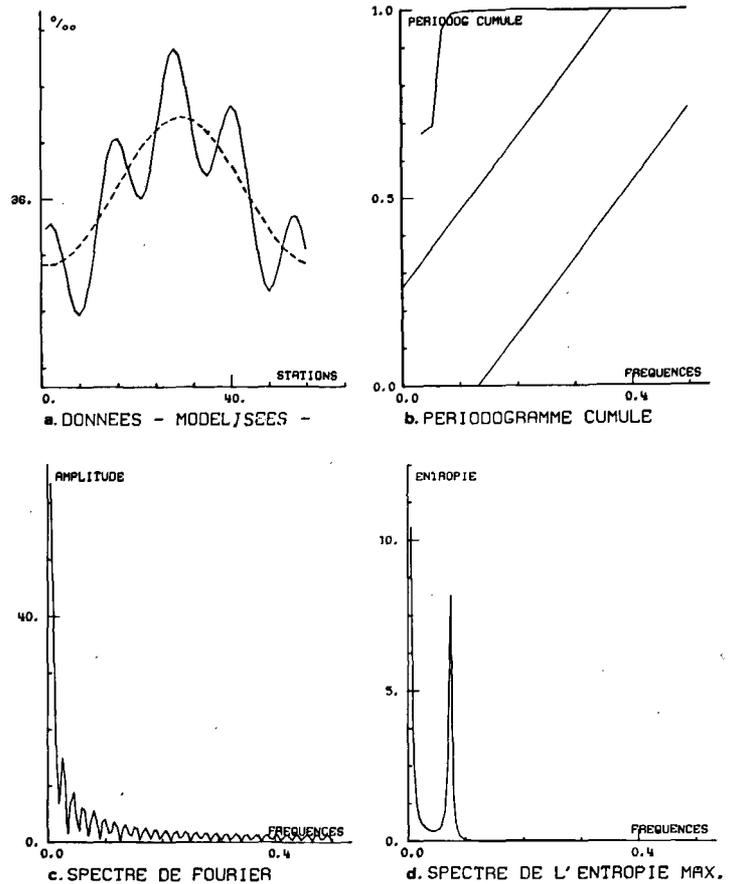


Figure 7

Analyses spectrales de la série « modélisée » de la salinité (a) sous la forme d'une somme de deux fonctions sinusoidales; l'une $y_1 = 36.085 - 0.788 \sin(2\pi/56 x_n + 1.399)$ décrit la tendance (---), l'autre $y_2 = -0.0299 + 0.512 \sin(2\pi/13 x_n + 0.6582)$ représente la composante de périodicité proche de 6 h (—). L'analyse spectrale par maximum d'entropie (d) apparaît nettement plus performante dans la détermination des deux composantes injectées dans le modèle.

Spectral analysis of the salinity modelled series obtained by addition of two sine functions: $y_1 = 36.085 - 0.788 \sin(2\pi/56 x_n + 1.399)$ describes the trend (---), and $y_2 = -0.0299 + 0.512 \sin(2\pi/13 x_n + 0.6582)$ represents the 6-hour periodic component (—).

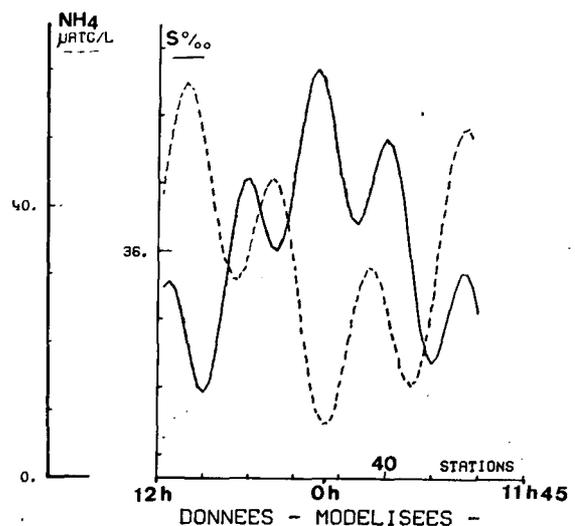


Figure 8

Superposition de deux séries modélisées correspondant à la salinité (—) et à l'ammoniaque (---). L'évolution en opposition de phase des deux composantes (tendance journalière et rythmicité de 6 h) est clairement mise en évidence.

Graphs of two modelled series corresponding to salinity (—) and ammonia (---). The evolution in opposition of phase of the two components (daily trend and 6-hour periodicity) is clearly demonstrated.

suivant des critères objectifs, du nombre de coefficients minimum nécessaire au meilleur ajustement n'est pas définitivement résolu. Cependant, comme il a été montré dans cette étude, le choix d'une valeur de M proche de $2N/\text{Log}2N$, déterminé par une valeur minimale de l'erreur prédite finale d'Akaike, permet d'aboutir à des estimations spectrales satisfaisantes.

En nous plaçant dans un cadre descriptif et/ou d'études de paramètres évoluant suivant des modèles autorégressifs plus ou moins complexes, nous pouvons affirmer l'intérêt de l'A.S.M.E. en écologie marine. Ainsi, dans le cas de l'étude, par l'intermédiaire de caractères tant hydrologiques que biologiques, des perturbations

cycliques d'un écosystème pélagique néritique dues à des rejets anthropiques marqués, nous avons constaté l'efficacité de l'A.S.M.E. Elle a permis de décrire clairement aussi bien les composantes tendanciennes journalières que celles plus courtes, de l'ordre de 6 h, de l'ensemble de ces caractères. La comparaison des tracés graphiques a permis une esquisse des relations intervariables : évolutions en phase ou en opposition de phase. Il serait maintenant souhaitable d'appliquer l'analyse spectrale croisée par maximisation de l'entropie, afin d'effectuer une approche bivariable qui permettrait d'explicitier plus finement les relations intervariables (Strand, 1977).

RÉFÉRENCES

- Ables J.G., 1974. Maximum entropy spectral analysis, *Astron. Astrophys., suppl.*, **15**, 383-393.
- Akaike H., 1969 a. Power spectrum estimation through autoregressive model fitting, *Annal. Inst. Stat. Math.*, **21**, 407-419.
- Akaike H., 1969 b. Fitting autoregressive models for prediction, *Annal. Inst. Stat. Math.*, **21**, 243-247.
- Akaike H., 1970. Statistical predictor identification, *Annal. Inst. Stat. Math.*, **22**, 203-217.
- Bartlett M.S., 1956. *An introduction to stochastic process*, New York, Cambridge Univ. Press.
- Berryman J.G., 1978. Choice of operator length for maximum entropy spectral analysis, *Geophysics*, **43**, 7, 1384-1391.
- Blackman R. B., Tukey J. W., 1958. *The measurement of power spectra*, Dover, New York, 190 p.
- Box G. E., Jenkins G. H., 1970. *Time series analysis: forecasting and control*, Holden Day, San Francisco, 575 p.
- Burg J. P., 1967. Maximum entropy spectral analysis, paper presented at the 37th annual International Meeting Soc. Explor. Geophys., Oklahoma City, Oklahoma, October 31, 1967.
- Burg J. P., 1968. A new analysis technique for time series data, paper presented at the Advanced Study Institute on signal processing, NATO, Enschede, Netherlands.
- Burg J. P., 1970. New concepts in power spectra estimation, paper presented at 37th annual International S.E.G. Meeting, New Orleans, Louisiana.
- Burg J. P., 1972. The relationship between maximum entropy spectra and maximum likelihood spectra, *Geophysics*, **37**, 375-376.
- Burg J. P., 1975. Maximum entropy spectral analysis, *Doct. Dissert., Stanford Univ., Dep. Geophysics, May 1975*, 123 p.
- Calmet D., Blanc F., 1980. Échantillonnages et traitements des données (écosystèmes pélagiques perturbés), *V^e Journées Études Pollution, CIESM, Cagliari*, 117-126.
- Calmet D., Charmasson S., Blanc F., 1983. Effets des perturbations spatio-temporelles anthropiques sur les écosystèmes pélagiques (émissaire de l'agglomération marseillaise), *Oceanol. Acta, Actes 17^e Symposium Européen de Biologie marine, Brest, 27 septembre-1^{er} octobre 1982*, sous presse.
- Colebrook J.M., 1981. Continuous plankton records: persistence in time series of annual means of abundance of zooplankton, *Mar. Biol.*, **61**, 143-149.
- Cruzado A., Kelley J. C., 1973. Continuous measurements of nutrient concentrations and phytoplankton density in the surface water of the Western Mediterranean, winter 1970, *Thalassia Jugosl.*, **9**, 18, 19-24.
- Denman K. L., Platt T., 1975. Spectral analysis in ecology, *Annal. Rev. Ecol. Syst.*, **6**, 189-210.
- Dessier A., Laurec A., 1978. Le cycle annuel du zooplancton à Pointe-Noire (R.P. Congo). Description mathématique, *Oceanol. Acta*, **1**, 3, 285-304.
- Durbin J., 1960. The fitting of time series models, *Rev. Int. Inst. Stat.*, **28**, 233-244.
- Frontier S., 1974. Contribution à la connaissance d'un écosystème néritique tropical : étude descriptive et statistique du peuplement zooplanctonique de la région de Nosy-Bé (Madagascar), *Thèse Doct. État, Aix-Marseille II*, 268 p.
- Fuller W., 1976. *Introduction to statistical time series*, John Wiley and Sons Inc., New York, 470 p.
- Glover R.S., 1967. The continuous plankton recorder survey of the North Atlantic, *Symp. Zool. Soc. Lond.*, **19**, 189-210.
- Ibanez F., 1975. Contribution à l'analyse mathématique des événements en écologie planctonique : optimisations méthodologiques ; étude expérimentale en continu à petite échelle de l'hétérogénéité du plancton côtier, *Thèse Doct. État, Univ. Paris VI*, 154 p.
- Jenkins G. M., Watts D. G., 1968. *Spectral analysis and its applications*, Holden Day, San Francisco, 525 p.
- Jones R.M., 1965. A reappraisal of the periodogram in spectral analysis, *Technometrics*, **7**, 53, 531-542.
- Kirk B. L., Rust B. W., Van Winkle W., 1979. *Time series analysis by the maximum entropy method*, Oak Ridge National Laboratory, Oak Ridge, Tenn. ORNL 5332, 152 p.
- Lacoss R.T., 1971. Data adaptive spectral analysis method, *Geophysics*, **36**, 4, 661-675.
- Laurec A., Blanc F., 1974. Applications de l'analyse spectrale à la description séquentielle des paramètres thermo-halins dans une région d'upwelling, *Tethys*, **6**, 439-452.
- Platt T., Denman K.L., 1975. Coherences in the horizontal distributions of phytoplankton and temperature in the upper Ocean, *Mém. Soc. R. Sci. Liège, 6^e sér.*, **7**, 19-30.
- Ponsonby J.E.B., 1973. An entropy measure for partially polarized radiation and its application to estimating radio sky polarisation distributions from incomplete "aperture synthesis" data by the maximum entropy method, *Mon. Nat. R. Astron. Soc.*, **163**, 369-380.
- Prieur L., 1980. Hétérogénéité spatio-temporelle dans le bassin liguro-provençal, *XXVII^e Congrès Assemblée Plénière CIESM, Cagliari, 9-18 octobre 1980*.
- Shiskin J., 1967. The X-11 variant of census method II to seasonal assessment program, Technical paper n° 15 (revised), Washington, bureau of census.
- Strand O.N., 1977. Multichannel complex maximum entropy (autoregressive) spectral analysis, *I.E.E.E. Trans. Auto. Control.*, **AC22**, 634-640.
- Ulrych T. J., 1972. Maximum entropy power spectrum of long period geomagnetic reversals, *Nature*, **235**, 218-219.
- Ulrych T. J., Jensen O., 1974. Cross spectral analysis using maximum entropy, *Geophysics*, **39**, 3, 353-354.
- Ulrych T. J., Bishop T.N., 1975. Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space Phys.*, **13**, 1, 183-200.
- Van den Bos A., 1971. Alternative interpretation of maximum entropy spectral analysis, *I.E.E.E. Trans. Inform. Theory*, **IT**, **17**, 493-494.
- Van Winkle W., Kirk B. L., Rust B. W., 1979. Periodicities in Atlantic coast striped bass (*Morone saxatilis*) commercial fisheries data, *Fish. Res. Board Can.*, **36**, 54-62.
- Wells F.J., Chinnery M.A., 1973. On the separation of spectral components of polar motion, *Geophys. J. R. Astron. Soc.*, **34**, 179-192.