

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

JOSÉ DELCOUR

Amélioration de l'algorithme de la transformation discrète de Fourier. Application au lissage des séries chronologiques

Journal de la société statistique de Paris, tome 128 (1987), p. 171-178

http://www.numdam.org/item?id=JSFS_1987__128__171_0

© Société de statistique de Paris, 1987, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

AMÉLIORATION DE L'ALGORITHME DE LA TRANSFORMATION DISCRÈTE DE FOURIER. APPLICATION AU LISSAGE DES SÉRIES CHRONOLOGIQUES

José DELCOUR

Ingénieur civil informaticien

(Travail présenté par le Professeur M.E. Dehousse, Directeur général de l'Institut supérieur de statistique de Lubumbashi, Zaïre)

On sait que l'analyse des séries temporelles peut être effectuée par la méthode des moyennes mobiles, de la droite de tendance, etc...

Dans le cas où les données sont distribuées uniformément sur l'axe des temps, l'application des séries de FOURIER peut aider à rejeter les mouvements irréguliers ou aléatoires de la série initiale conduisant ainsi à un meilleur ajustement de ces données et à en dégager l'existence de mouvements cycliques ou saisonniers.

Nous établissons de nouvelles relations permettant d'obtenir une transformation discrète de FOURIER et proposons un filtre digital plus efficace pour l'obtention d'un lissage rationnel dégageant le graphe périodique adouci.

Nous introduisons également une transformation des axes de référence.

Nous proposons un programme BASIC de traitement de séries chronologiques et l'appliquons à la recherche de la périodicité d'une fonction bruitée.

I. Établissement des formules relatives aux transformations directe et inverse de FOURIER dans le cas discret.

81

On sait que si une fonction $f(t)$, périodique de période $2l$, est développable en série de FOURIER, on peut écrire :

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi t}{l}\right) + b_k \sin\left(\frac{k\pi t}{l}\right) \quad (1)$$

$$\text{avec : } a_k = \frac{1}{l} \int_0^{2l} f(u) \cos\left(\frac{k\pi u}{l}\right) du$$

$$b_k = \frac{1}{l} \int_0^{2l} f(u) \sin\left(\frac{k\pi u}{l}\right) du$$

Dans le cas où $f(t)$ est une fonction discrète de la variable t distribuée uniformément sur l'axe des temps en $N+1$ couples de points (x_j, y_j) , les intégrales relatives aux coefficients a_k et b_k doivent être calculées par une méthode approchée.

La théorie de la transformation discrète de FOURIER utilise la méthode des rectangles. Nous appliquerons la formule des trapèzes conduisant ainsi à une meilleure approximation.

sachant que :

$$\int_0^N f(x) dx \approx \left(\frac{y_0 + y_N}{2} + \sum_{k=1}^{N-1} y_k \right)$$

on peut écrire :

$$a_k = \frac{1}{N} \left[y_0 + y_N + 2 \sum_{j=1}^{N-1} y_j \cos \left(\frac{2\pi k j}{N} \right) \right] \quad (k = 0, 1, 2, \dots, N)$$

$$b_k = \frac{1}{N} \left[\sum_{j=1}^{N-1} y_j \sin \left(\frac{2\pi k j}{N} \right) \right]$$

$$\text{avec : } 2l = N$$

$$x_j = j$$

On montre alors que, dans le cas discret, la relation (I) devient :

$$y_j = 2a + 2 \sum_{k=1}^{N-1} a_k \cos \left(\frac{2\pi k j}{N} \right) + b_k \sin \left(\frac{2\pi k j}{N} \right) \quad (j = 1, 2, \dots, N-1)$$

$$y_j = \frac{1}{N} \sum_{k=1}^{N-1} y_j \cos \left(\frac{2\pi k j}{N} \right)$$

$$y_N = \frac{1}{N} \sum_{k=1}^{N-1} y_j \sin \left(\frac{2\pi k j}{N} \right)$$

Définissant la transformation directe :

$$R_k^* = \frac{1}{2N} \left[y_0 + y_N + 2 \sum_{j=1}^{N-1} y_j \cos \left(\frac{2\pi k j}{N} \right) \right] \quad (1)$$

$$I_k^* = \frac{1}{N} \left[\sum_{j=1}^{N-1} y_j \sin \left(\frac{2\pi k j}{N} \right) \right] \quad (2)$$

$$(\text{avec : } a_k = 2R_k^* \quad \text{et} \quad b_k = 2I_k^*)$$

la transformation inverse peut s'écrire :

$$y_j = R_0^* + \sum_{k=1}^{N-1} \left[R_k^* \cos \left(\frac{2\pi k j}{N} \right) + I_k^* \sin \left(\frac{2\pi k j}{N} \right) \right] \quad (j = 1, 2, \dots, N-1) \quad (3)$$

$$y_0 = \frac{1}{N} \cdot \sum_{k=0}^{N-1} \sum_{j=1}^N y_j \cdot \cos \left(\frac{2\pi k j}{N} \right) \quad (4)$$

$$y_N = \frac{1}{N} \cdot \sum_{k=0}^N \sum_{j=1}^N y_j \cdot \cos \left(\frac{2\pi k j}{N} \right) \quad (5)$$

Cherchons à présent à simplifier ces formules :

Comme

$$R_{N-k}^* = \frac{1}{2N} \cdot \left[y_0 + y_N + 2 \sum_{j=1}^{N-1} y_j \cdot \cos \left(\frac{2\pi(N-k) \cdot j}{N} \right) \right] = R_k^* \quad \text{et}$$

$$I_{N-k}^* = \frac{1}{N} \cdot \left[\sum_{j=1}^{N-1} y_j \cdot \sin \left(\frac{2\pi(N-k) \cdot j}{N} \right) \right] = I_k^*$$

on peut écrire :

$$y_j = R_0^* + \sum_{k=1}^{\frac{(N-1)}{2}} \left[R_k^* \cdot \cos \left(\frac{2\pi k j}{N} \right) + I_k^* \cdot \sin \left(\frac{2\pi k j}{N} \right) \right] +$$

$$\alpha \cdot \left(\frac{R_N^*}{2} \cos(\pi j) + \frac{I_N^*}{2} \sin(\pi j) \right) \quad (j = 1, 2, \dots, N-1) \quad (6)$$

avec $\alpha = \begin{cases} 1 & \text{si } N \text{ est pair} \\ 0 & \text{si } N \text{ est impair} \end{cases}$

La fonction $E(x)$ représente l'entier positif directement inférieur à x .

A l'aide d'un raisonnement analogue on peut aussi écrire :

$$y_0 = -\frac{y_N - y_0}{2} + R_0^* + 2 \sum_{k=1}^{E\left(\frac{N-1}{2}\right)} R_k^* + \alpha \frac{R_N^*}{2} \quad (7)$$

$$y_N = y_0 + 2 \cdot \left(\frac{y_N - y_0}{2}\right) \quad (8)$$

On remarquera que les formules (6), (7) et (8) sont des identités si on y remplace R^*_k et I^*_k par leurs valeurs tirées de (1) et (2). Les termes R^*_0 et $(R^*_{2k} + I^*_{2k})$ représentent respectivement la fondamentale et l'amplitude du k -ième harmonique de la série de FOURIER associée à la distribution discrète d'ordre N .

Notre but étant de réduire le « bruit » dû aux parasites relatifs à la distribution discrète, on va rechercher un filtre digital $f(k, l)$, fonction de deux variables k et l représentant respectivement le k -ième harmonique de la série et un harmonique de coupure d'ordre l , atténuant les harmoniques de la série de FOURIER.

Les formules (1) à (5) peuvent s'écrire comme suit :

Transformation directe :

$$R_k^* = \frac{f(l, k)}{2N} \left[y_0 + y_N + 2 \sum_{j=1}^{N-1} y_j \cos\left(\frac{2\pi k j}{N}\right) \right]$$

$$I_k^* = \frac{f(l, k)}{N} \left[\sum_{j=1}^{N-1} y_j \sin\left(\frac{2\pi k j}{N}\right) \right]$$

$$\left[k = 0, 1, \dots, E\left(\frac{N-1}{2}\right) \right]$$

Transformation inverse

$$y_j^* = R_0^* + \sum_{k=1}^{E\left(\frac{N-1}{2}\right)} \left[R_k^* \cdot \cos\left(\frac{2\pi k j}{N}\right) + I_k^* \cdot \sin\left(\frac{2\pi k j}{N}\right) \right] +$$

$$\alpha \cdot \left(\frac{R_N^*}{2} \cos(\pi j) + \frac{I_N^*}{2} \sin(\pi j) \right) \quad (j = 1, 2, \dots, N-1)$$

$$y_0^* = -\frac{y_N - y_0}{2} + R_0^* + 2 \sum_{k=1}^{E\left(\frac{N-1}{2}\right)} R_k^* + \alpha \frac{R_N^*}{2}$$

$$y_N^* = y_0 + 2 \cdot \frac{y_N - y_0}{2}$$

Remarque : Si on impose la condition

$$l < E\left(\frac{N-1}{2}\right)$$

Ces dernières relations peuvent encore s'écrire :

$$y_j = R_0^* \sum_{k=1}^l \left[R_k^* \cdot \cos\left(\frac{2\pi kj}{N}\right) + I_k^* \cdot \sin\left(\frac{2\pi kj}{N}\right) \right]$$

$$y_0^* = -\frac{y_N - y_0}{2} + R_0^* + 2 \sum_{k=1}^{E\left(\frac{N-1}{2}\right)} R_k^*$$

$$y_N^* = y_0 + 2 \cdot \frac{y_N - y_0}{2}$$

L'utilisateur a la faculté de décider du genre de filtrage de sa série chronologique : plus l est grand, plus la série chronologique finale contiendra des informations irrégulières. Le programme proposé est conçu de manière interactive pour permettre à l'utilisateur de choisir au mieux le rang l de l'harmonique de coupure.

Remarquons enfin que le calcul des coefficients R_k^* et I_k^* de la transformation directe ne doit pas être entièrement refait pour chaque valeur de l . Notre programme tient compte de cette particularité.

II. Choix du filtre digital $f(l, k)$

Le filtre digital :

$$f^{(1)}(l, k) = \begin{cases} 1 & \text{si } k < l \\ 0 & \text{si } k \geq l \end{cases} \quad (9) \quad (k, l \simeq N)$$

présente l'inconvénient majeur de couper tous les harmoniques supérieurs au l -ième au lieu de les atténuer; l représentant l'ordre de l'harmonique de coupure.

De plus, si l est proche de $E(N/2)$, la série chronologique lissée diffère peu de la série chronologique initiale.

Des conclusions identiques peuvent être formulées pour le filtre :

$$f^{(2)}(l, k) = \begin{cases} 1 - \left(\frac{k}{l}\right)^2 & \text{si } k < l \\ 0 & \text{si } k \geq l \end{cases} \quad (10) \quad (k, l \simeq N)$$

Après divers essais, nous avons préféré utiliser le filtre digital suivant :

$$f(l, k) = \exp\left\{-\mu \left(\frac{k}{l}\right)^2\right\} \quad (11)$$

Pour un k -ième harmonique, ce dernier filtre lisse plus fortement la série chronologique initiale que les filtres (9) et (10).

Ce filtre ne restitue jamais la série chronologique initiale dans le cas où l est égal à la fonction $E(N/2)$.

Il est beaucoup plus sensible aux bruits et atténue différemment chacun des harmoniques.

Après expérimentation, nous avons pu obtenir une atténuation de 99 % du l -ième harmonique. Un calcul simple conduit à retenir $\mu = 4,6$.

On obtient dès lors :
$$f(l,k) = \exp \left\{ -4,6 \left(\frac{k}{l} \right)^2 \right\} \quad 1 \leq l \leq E(N/2) \quad (12)$$

La borne supérieure de variation au l -ième harmonique provient des conditions :
 $R^*_{N-k} = R^*_k$ et $I^*_{N-k} = I^*_k$

III. Transformation des axes de référence

La transformation d'axes effectuée est la suivante :
 Soit $y = ax + b$ l'équation de la droite de tendance avec

$$a = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

$$l = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

On définit une transformation $(t, y) \rightarrow (u, v)$ par :

$$v = y - (at + b)$$

$$c = \cos(\arctg a)$$

$$u = t/c$$

et la transformation inverse $(u, v) \rightarrow (t, y)$ par :

$$y = v + (at + b)$$

$$t = u \cdot c$$

Cette transformation est indispensable compte tenu de la nature périodique de la transformation de FOURIER. En effet, la série chronologique lissée doit vérifier la relation $f(x) = f(x + 2l)$; dans notre problème : $f(0) = f(N - 1)$. Un lissage dans les axes initiaux ne restituerait pas les données initiales et finales.

Pour bien mettre en évidence ce phénomène lié à la périodicité, nous avons considéré une série chronologique (1) à tendance croissante (figure 1a) que nous avons lissée à l'aide de notre algorithme sans effectuer de transformation des axes de référence. Cette même série a été lissée en effectuant la transformation suivant la droite de tendance (figure 1b). On constate directement, au vu des figures, que la série chronologique lissée dans les axes initiaux ne suit pas du tout l'allure de la série initiale au début et à la fin de celle-ci, alors que celle lissée dans le nouveau système d'axes restitue fidèlement l'allure générale de la série chronologique.

IV. Programmation en langage BASIC du lissage des séries chronologiques

Ce problème a été traduit en BASIC Microsoft (version 4.51). L'ordinateur utilisé est un TRS80 modèle I dont la fréquence d'horloge a été portée à 2.62 MHz.

1. Cette série chronologique représente la variation relative de la hauteur virtuelle minimum (exprimée en km) de la région F de la couche ionosphérique pour une période de 59 jours, du 1^{er} janvier 1983 au 28 février 1983.

Figure 1a : Lissage dans les axes initiaux

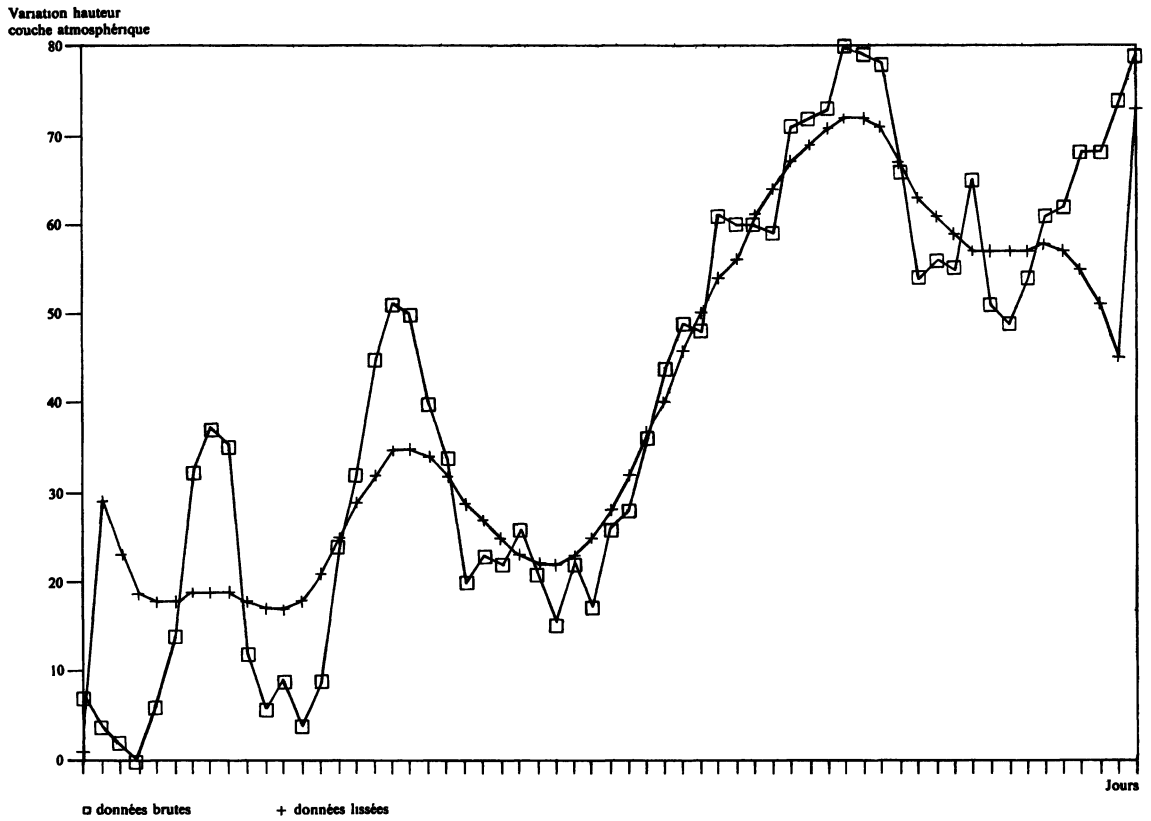
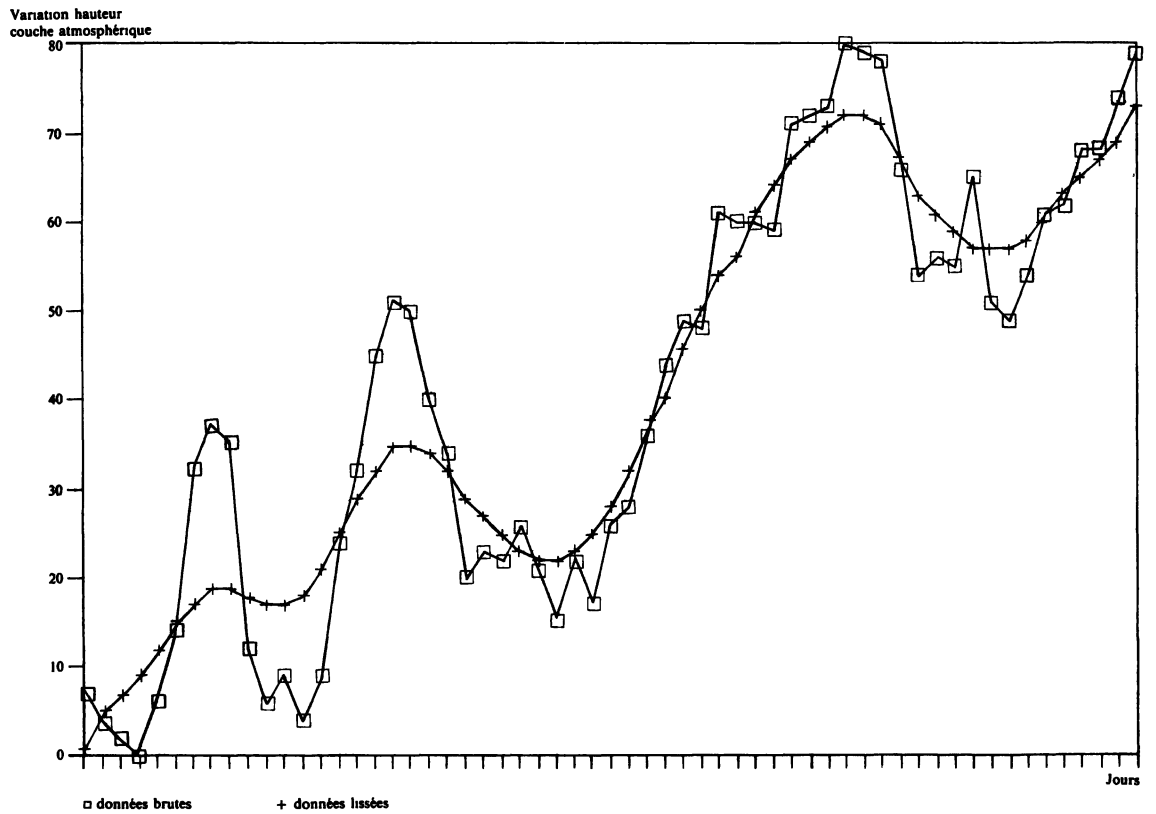


Figure 1b : Lissage dans les nouveaux axes



Le programme (1) effectue les opérations suivantes :

- a) lecture des données;
- b) calcul des coefficients de la droite de régression;
- c) transformation des coordonnées initiales suivant l'axe de régression;
- d) calcul des coefficients R^* et I^* ; filtrage de ces coefficients;
- e) calcul de la transformation inverse;
- f) impression des résultats soit sous forme graphique soit sous forme de tableau.

Le programme occupe 15,2 Kmots en mémoire centrale en tenant compte de l'initialisation des tableaux et des variables.

De part la nature compilée du programme, nous avons dû effectuer une déclaration statique des tableaux à 1 000 éléments maximum.

La seule contrainte de calcul réside dans la taille de la mémoire centrale susceptible d'accueillir les tableaux.

Nous aurions pu utiliser la technique de la dichotomie (P-fold condensation) pour réduire le temps de calcul des termes

$$\sum_{j=1}^{N-1} y_j \cos\left(\frac{2\pi kj}{N}\right); \sum_{j=1}^{N-1} y_j \sin\left(\frac{2\pi kj}{N}\right)$$

$$\sum_{k=1}^E \left[R_k^* \cos\left(\frac{2\pi kj}{N}\right) + I_k^* \sin\left(\frac{2\pi kj}{N}\right) \right]$$

Nous avons préféré programmer les équations fournies par notre algorithme.

V. Application au lissage d'une série chronologique

A titre d'application de notre méthode, nous avons recherché une fonction périodique à tendance croissante. Après divers essais, nous avons retenu la suivante : $f(x) = x/2 + 2 \cdot \sin(2x)$ et considéré un échantillon de 200 valeurs de celle-ci dont les abscisses sont distribuées uniformément sur l'intervalle $[0, 199]$.

Nous avons ensuite modifié chacune de ces valeurs à l'aide de la fonction de bruitage suivante : $y_j^* = y_j \cdot (1 + c \cdot d)$ $j = 1, 2, \dots, 200$

c est un nombre aléatoire réel appartenant à l'intervalle $]0, 1[$ et fourni par un générateur de nombres aléatoires défini par la suite : $x_k = (24298x_{k-1} + 99991)$ modulo 199017 $k = 1, 2, 3, \dots$

$$x_0 \in 0, 199017$$

$$c = x_k \cdot 398034 - 1$$

Chaque nouvelle valeur de y_j appartient à l'intervalle

$$]y_j - c \cdot d, y_j + c \cdot d[.$$

Après expérimentation, nous avons donné au coefficient d respectivement les valeurs : 0.1, 0.5, 1.0 et 2.0. L'échantillon initial génère alors 4 échantillons de points dont les intervalles de variation de chaque ordonnée sont :

$$y_j^* \in]0.9 \cdot y_j, 1.1 \cdot y_j[\quad d = 0.1;$$

$$y_j^* \in]0.5 \cdot y_j, 1.5 \cdot y_j[\quad d = 0.5;$$

$$y_j^* \in]0.2 \cdot y_j, 3.0 \cdot y_j[\quad d = 1.0;$$

$$y_j^* \in]-y_j, 3.0 \cdot y_j[\quad d = 2.0.$$

$$j = 1, 2, \dots, 200$$

Après application de notre algorithme de lissage à ces quatre ensembles de points, nous constatons que la périodicité ainsi que l'allure générale de la courbe initiale ont bien été restituées.

VI. Conclusions

Nous avons montré comment, en améliorant l'algorithme de la transformation discrète de FOURIER et en introduisant une transformation des axes de référence, nous éliminons les problèmes liés à la tendance de la série chronologique.

De plus, à cause de cette transformation des axes de référence, la tendance de la série chronologique analysée peut être soit croissante, soit décroissante, l'algorithme de la transformation discrète se limitant alors uniquement à l'étude de fonctions périodiques à tendance nulle, donc celles pour lesquelles $f(x) = f(x + l)$.

La limitation de l'influence du bruit perturbateur dans les séries chronologiques est grandement facilitée par le calcul automatique des intégrales de la série de FOURIER et grâce au choix de notre filtre. Pour l'exemple traité, il aurait fallu calculer 300 intégrales (200 pour la transformation directe et 100 pour la transformation inverse). Grâce à l'utilisation de notre filtre, nous obtenons des résultats très satisfaisants en n'en retenant que 60 comme le démontrent nos graphiques. Notre programme est par ailleurs conçu de manière interactive : c'est l'utilisateur qui décide du genre de filtrage de sa série chronologique en faisant varier le paramètre k du filtre et qui détermine le nombre de termes à retenir pour faire ressortir incontestablement le caractère périodique du phénomène.

Le filtre digital proposé lisse plus fortement et différemment chaque harmonique relatif à la série initiale, ce que les autres filtres ne font pas.

Enfin, si le temps de calcul est un facteur prédominant, on peut introduire dans le programme proposé l'algorithme FFT (Fast FOURIER Transform) ou la technique de la dichotomie (P-fold condensation) et, éventuellement, programmer l'algorithme dans un autre langage évolué.

Remerciements

Nous tenons à remercier M. le Professeur Dr M.E. Dehousse pour les suggestions qu'il a formulé tout au long de ce travail.

BIBLIOGRAPHIE

- [1] ABRAMOWITZ M. & SEGUN I.A. Handbook of mathematical functions. Dover Publications, Inc. 1968, p. 885, 6.
- [2] AUBANEL L.E. & OLDHAM K.B. FOURIER smoothing without the Fast FOURIER Transform. Byte vol. X, n° 2, 1985, pp. 207-218.
- [3] BESSON R. Le traitement numérique des signaux : I. De l'analogique au numérique. Toute l'électronique, n° 497, pp. 47-51, 10/1984.
- [4] BRIGHAM E.O. The Fast FOURIER Transform. New York. Prentice Hall. 1974.
- [5] KNUTH D.E. The Art of computer programming. Vol. III. Addison-Wilsey Publishing Company, Inc. Reading Mass. 1970.
- [6] MONJALLON A. Introduction à la méthode statistique. Tome II. Paris. Vuibert. 1969.