Belgian Journal of Operations Research, Statistics and Computer Science, Vol 25, n° 2-3.

RECURRENCE OU PERIODICITE

R. SNEYERS

Institut royal météorologique de Belgique
Avenue Circulaire 3
1180 Bruxelles
Belgique

ABSTRACT

The theory of autoregressive series and of random periodic series is briefly recalled. It is shown that the series of sample autocovariances enables to decide of the type of model to adjust to the original series. Two examples are considered: the series of daily averages of the athmospheric pressure at Uccle and the series of annual Wolf sunspot numbers.

1. *Introduction.*

L'une des applications les plus importantes de la statistique mathématique à la géophysique est la représentation des séries chronologiques d'observations au moyen de modèles stochastiques en vue de prévisions. Parmi ceux-ci deux types de modèles jouent un rôle majeur : les modèles récurrents ou autorégressifs et les modèles périodiques. Dans le premier cas, la série chronologique obéit à une équation récurrente non homogène dont le terme indépendant est une quantité aléatoire simple; dans le second, à une quantité aléatoire près, la série se représente à l'aide de la somme de fonctions sinusoïdales de périodes déterminées. Il va de soi que les deux modèles se distinguent nettement en ce qui concerne leur pouvoir de prédiction, le modèle récurrent n'autorisant que la prévision à courte échéance et le modèle périodique permettant au contraire la prévision à longue échéance.

L'objet de la présente note est de rappeler brièvement les propriétés des deux types de modèles et les moyens de les distinguer; de plus, deux exemples sont donnés qui illustrent l'un et l'autre modèle.

2. *Les séries aléatoires récurrentes.*

Une série chronologique $x_1, x_2, \ldots, x_n$ est dite aléatoire récurrente (autorégressive) si elle vérifie une équation de la forme :

$$x_{i+k} = a_0 + a_1 x_i + a_2 x_{i+1} + \ldots + a_k x_{i+k-1} + e_{i+k} \qquad (1)$$

où a_0 , $a_1, \ldots, a_k$ sont des constantes et où e_{i+k} sont des quantités aléatoires simples de moyennes nulles et de même répartition. De plus, à la condition que $a_1 \neq 0$, elle est dite d'ordre k .

Avec var $e_i = \sigma^2$ on a aussi : $\mathrm{cov}(x_i, e_j) = 0$ pour $i < j$ et $\mathrm{cov}(x_i, e_i) = \sigma^2$.

Toute solution de (1) peut se mettre sous la forme de la somme de solutions particulières de l'équation homogène :

$$x_{i+k} = a_1 x_i + a_2 x_{i+1} + \ldots + a_k x_{i+k-1} \qquad (2)$$

dont la solution générale est de la forme :

$$\Sigma \lambda_j x_i^{(j)} \quad \text{avec} \quad x_i^{(j)} = b_r^i, \; b_s^i \sin \alpha_s i, \; b_s^i \cos \alpha_s i$$

sachant que b_r et b_s $(\cos \alpha_s + \sqrt{-1} \sin \alpha_s)$ sont les racines réelles ou complexes de l'équation caractéristique associée :

$$z^k - a_k z^{k-1} - a_{k-1} z^{k-2} \ldots - a_1 = 0 \qquad (3)$$

On en déduit que si x_i est une série stationnaire, on doit avoir :

$|b_r| < 1$ et $|b_s| < 1$, donc aussi $|a_1| < 1$.

De plus, la série x_i est la somme de séries amorties de façon continue ou oscillante. Dans le second cas, on peut la considérer comme "pseudo-périodique".

Par ailleurs, si on pose $v_k = \mathrm{cov}(x_i, x_{i+k})$ pour $k > 0$, on établit que la série des autocovariances v_k vérifie l'équation homogène (2). Il s'ensuit que lim $v_k = 0$ lorsque k tend vers l'infini.

Enfin, si on désigne par $\overline{v}_k$ les autocovariances empiriques de la série, c'est-à-dire :

$$\overline{v}_k = [\Sigma x_i x_{i+k} - (\Sigma x_i)(\Sigma x_{i+k})/(n-k)]/(n-k-1) \qquad (4)$$

on montre que la série des $\overline{v}_k$ vérifie une équation du type (1) où toutefois la quantité e_{i+k} n'est pas aléatoire simple. De ce fait, la série des $\overline{v}_k$ ne présente qu'approximativement les caractères pseudo-périodiques de la série des v_k mais apparaît habituellement comme une série persistante (non aléatoire simple).

3. *Les séries aléatoires périodiques.*

Une série chronologique $x_1, x_2, ..., x_n$ est dite aléatoire périodique si elle peut se mettre sous la forme :

$$x_i = a_0 + \sum_j (a_j \sin \alpha_j i + b_j \cos \alpha_j i) + e_i \qquad (1)$$

où a_0, a_j, b_j et α_j sont des constantes et où e_i sont des quantités aléatoires simples de moyennes nulles et de même répartition.

Avec $c_j = \sqrt{a_j^2 + b_j^2}$ et $b_j/a_j = \mathrm{tg}\,\varphi_j$, (1) peut encore s'écrire :

$$x_i = a_0 + \Sigma c_j \sin(\alpha_j i + \varphi_j) + e_i \qquad (2)$$

Avec var $e_i = \sigma^2$ on en tire immédiatement :

$$E(x_i) = a_0 \qquad v_0 = \mathrm{var}\, x_i = \Sigma c_j^2/2 + \sigma^2$$

$$v_k = \mathrm{cov}(x_i, x_{i+k}) = \Sigma (c_j^2/2) \cos \alpha_j k \qquad (3)$$

d'où il résulte que la série des autocovariances d'une série aléatoire périodique est une série périodique.

Si l'on définit la variance empirique $\overline{v}_0$ et les autocovariances empirique $\overline{v}_k$ à l'aide de la formule 2(4), il vient : $E(\overline{v}_0) \cong v_0$, $E(\overline{v}_k) \cong v_k$. De plus, si e_i est une variable normale, on a aussi :

$$\text{var } \overline{v}_0 = 2\sigma^4/(n\text{-}1) \quad , \quad \text{var } \overline{v}_k = \sigma^4/(n\text{-}k\text{-}1) \quad \text{et} \quad \text{cov}(\overline{v}_k \, , \, \overline{v}_l) = 0 \text{ pour } k \neq l > 0. \tag{4}$$

On en déduit que la série $\overline{v}_k$ est également une série aléatoire périodique ayant les mêmes périodes que la série originale, la variance de $\overline{v}_k$ étant toutefois croissante.

En particulier, si on pose $r_j^2 = \sigma^2/c_j^2$, il est clair que la périodicité apparaîtra d'autant mieux dans la série que r_j^2 est petit. Par ailleurs, le même rapport calculé pour la série $\overline{v}_k$ devient en vertu de (4) :

$$r_j^2(\overline{v}_k) = [\,\sigma^4/(n\text{-}k\text{-}1)]/(c_j^4/4) = \frac{4\,r_j^2}{n\text{-}k\text{-}1} \,.\, r_j^2 \tag{5}$$

ce qui montre que $r_j^2(\overline{v}_k) < r_j^2$ dès que $(n\text{-}k\text{-}1) > 4\,r_j^2$.

Il s'ensuit que la période apparaîtra mieux dans la série $\overline{v}_k$ dès que n est suffisamment grand.

Le calcul des autocovariances empiriques est donc un excellent moyen de détection des périodicités dans les séries chronologiques.

4. *Estimation de la récurrence ou de la composante périodique.*

L'estimation des constantes peut se faire dans les deux cas par la méthode des moindres carrés.

Utilisant les notations matricielles, sachant que θ' est le transposé de θ, on pose dans le cas de la récurrence :

$$x = \| \, x_{i+k} \, \| \quad , \quad u = \| 1 \; x_i \, x_{i+1} \cdots x_{i+k-1} \, \| \quad , \quad \theta' = \| \, a_0 \; a_1 \; \ldots \; a_k \| \, , \; e = \| \, e_{i+k} \, \|$$

tandis que dans le cas de la série aléatoire périodique, on fait :

$$x = \| \, x_i \, \| \quad , \quad u = \| \, 1 \; \sin \alpha_1 i \; \cos \alpha_1 i \; \sin \alpha_2 i \; \cos \alpha_2 i \ldots \| \quad , \quad \theta' = \| a_0 \; a_1 \; b_1 \; a_2 \; b_2 \; \cdots \| \, ,$$

$$e = \| \, e_i \, \|$$

d'où il résulte que les équations 2(1) et 3(1) se mettent sous la forme commune :

$$x = u\theta + e \tag{1}$$

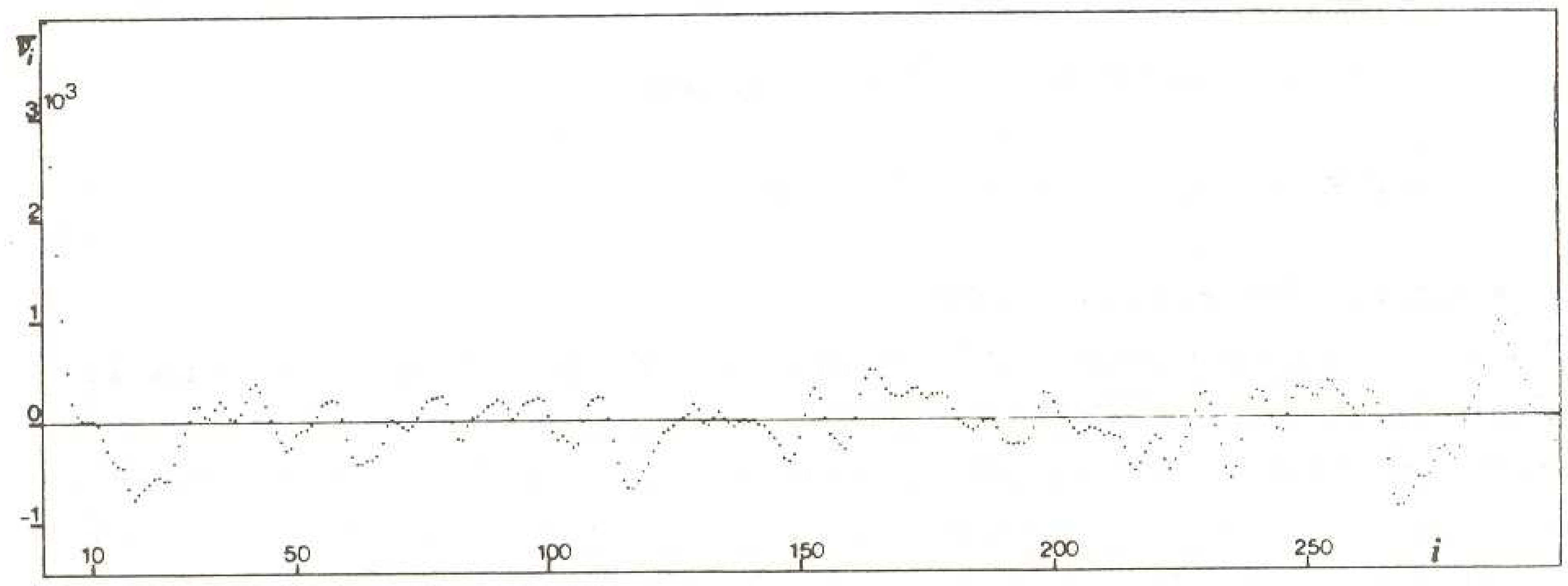

FIG. 1. — Autocovariances empiriques $\overline{v}_i$ pour i = 1,2,...,300, de la série des moyennes journalières de la pression atmosphérique à Uccle (d'après deux années d'observations).

L'estimation de θ s'obtient en minimisant la forme quadratique :

$$e'e = (x - u\theta)'(x - u\theta) \tag{2}$$

ce qui donne l'estimation :

$$\hat{\theta} = (u'u)^{-1} u'x \quad \text{et} \quad \text{var}\,\hat{\theta} = (u'u)^{-1} \sigma^2 \tag{3}$$

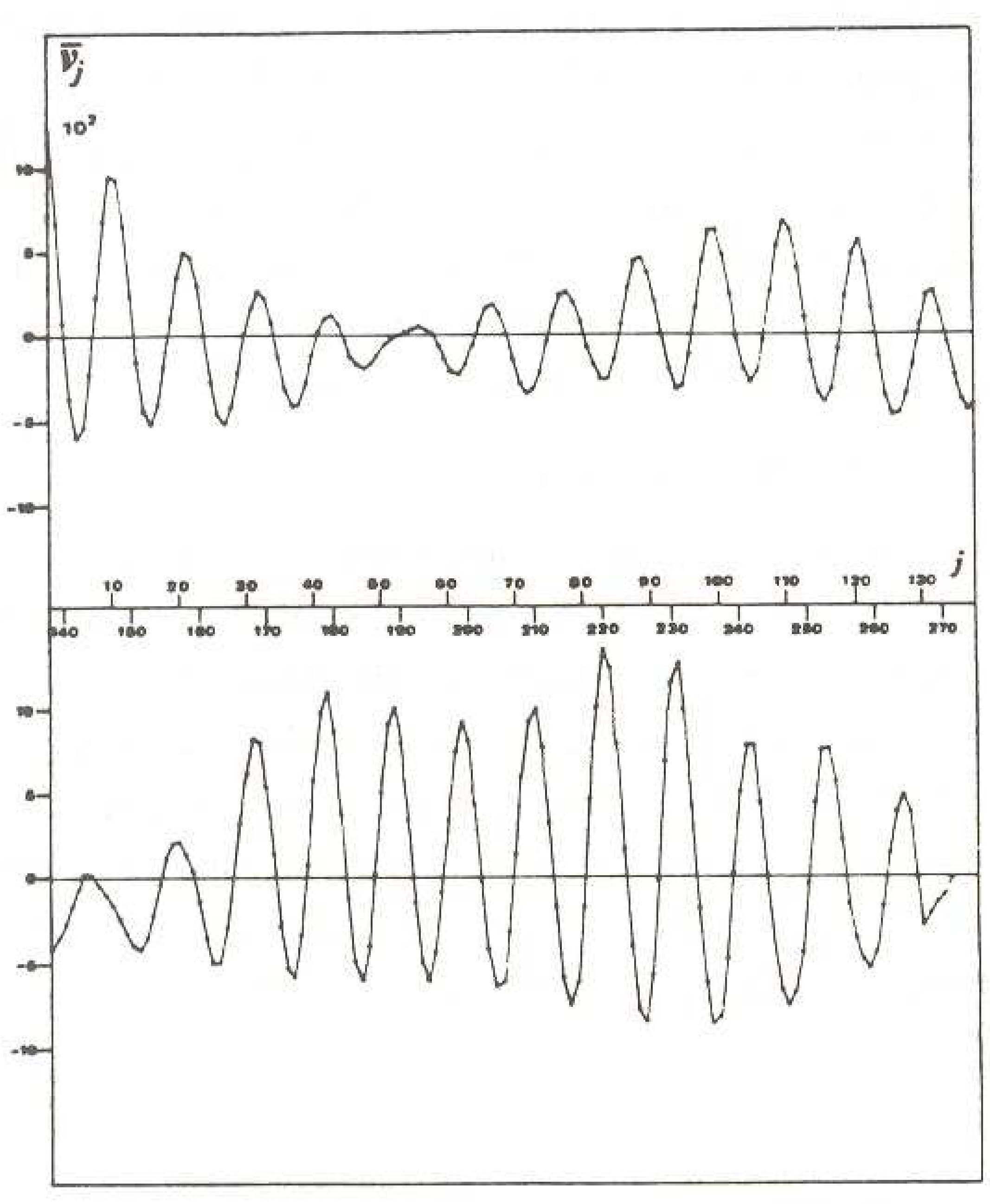

Fig.2 – Autocovariances empiriques $\overline{v}_j$ pour j = 1, 2, ..., 271 de la série des nombres de Wolf de 1700 à 1972.

De plus, on peut estimer σ^2 à l'aide de la relation :

$$\hat{\sigma}^2 = e'e/(n-n_1)$$

où n_1 est le nombre de paramètres estimés.

Dans le cas d'une série aléatoire récurrente, l'estimation peut se faire en ajustant successivement des récurrences d'ordre croissant 1, 2, ..., k jusqu'à l'obtention d'un résidu aléatoire simple. Pour une série aléatoire périodique, une première estimation des α_j peut être déduite du graphique des autocovariances empiriques $\overline{v}_k$, l'estimation finale étant ensuite obtenue en retenant celle qui donne la plus grande valeur à c_j dans un intervalle $(\alpha_{j_1}, \alpha_{j_2})$ entourant la première estimation de α_j .

5. *Exemples.*

a) La série des moyennes journalières de la pression atmosphérique à Uccle (deux années d'observations).

La figure 1 donne la représentation graphique des autocovariances empiriques $\overline{v}_k$ pour $k = 1, 2, ..., 300$.

Il apparaît que la série est persistante, mais ne présente aucune périodicité systématique. Des ajustements successifs montrent que la série de pressions obéit à une récurrence aléatoire d'ordre 2. La pseudo période des solutions de l'équation récurrente homogène correspondante est 16,7 j avec une erreur type d'estimation de 3,4 j. Cette pseudo période est en bon accord avec le phénomène de vacillation que l'on observe dans l'atmosphère. De plus, le modèle récurrent a permis de faire une bonne prévision statistique des extrêmes mensuels des moyennes journalières de la pression, ceux observés durant une période de trente années se révélant en bon accord avec la répartition théorique.

b) Le nombre de Wolf de taches solaires.

La figure 2 donne les autocovariances empiriques $\overline{v}_k$ des nombres de Wolf moyens annuels observés de 1700 à 1972 pour $k = 1, 2, ..., 271$. Le graphique révèle l'existence d'une périodicité de l'ordre de 10,5 ans dont l'amplitude varie elle-même avec une période de l'ordre de 200 ans. Des essais successifs par moindres carrés du modèle qu'on en déduit permet de trouver les estimations $\hat{T} = 10{,}494$ et $\hat{T}' = 200{,}5$. Le phénomène des taches solaires apparaît donc comme dû principalement à un phénomène de battement entre deux ondes de périodes voisines de 10,5 ans. Le résidu des périodicités estimées a été traité de la même manière et en continuant de la sorte d'autres périodicités ont été tirées de la série.

Au total 26 fonctions périodiques de la forme $a_j \sin \alpha_j i + b_j \cos \alpha_j i$ ont été retirées avant que la sélection n'ait été arrêtée par un test de signification qui ne sera pas détaillé ici.

Une comparaison a été faite entre la prédiction des nombres de Wolf qu'on peut en déduire et les nombres observés de 1973 à 1983. La décroissance du nombre de Wolf jusqu'en 1976 a été prévue de façon acceptable de même qu'un accroissement de ce nombre au delà, mais l'écart entre les valeurs maximales observées et la prédiction est trop grande pour rendre le modèle admissible. En réalité, la prédiction fournie par un modèle formé des premières sinusoïdes suggérées par la série des autocovariances et un résidu autorégressif du second ordre apparaît par contre comme acceptable. Le phénomène serait donc du à la combinaison d'un effet périodique et d'un effet récurrent.

Références.

Anderson, T.W., 1971 : The Statistical Analysis of Time Series, Wiley

Kendall, M.G., and Stuart, A., 1963 : The Advanced Theory of Statistics, 3 vols, Griffin.

Kendall, Sir Maurice, 1975 : Time-Series, Griffin.

Sneyers, R., 1975 : Sur l'analyse statistique des séries d'observations, O.M.M. Note technique n° 143.

Sneyers, R., 1975 : Les séries aléatoires récurrentes, Application aux moyennes journalières de la pression à Uccle, in Hommage à-Hulde aan Jacques Van Mieghem, Inst. R. Mét. Pub. A, n° 91.

Sneyers, R., 1976 : Application of Least Squares to the Search for Periodicities, J. Appl. Met., Vol. 15, No 4, pp. 387-393.

Sneyers, R. et P. Cugnon , 1984 : On the predictability of the Wolf sunspot number, European Geophysical Society, 10th Annual Meeting, Louvain la Neuve, 30 July- 3 August 1984.