

REVUE DE STATISTIQUE APPLIQUÉE

BENCHETRIT

PHAM-DINH-TUAN

Un essai d'analyse statistique des séries de données respiratoires

Revue de statistique appliquée, tome 22, n° 2 (1974), p. 51-68

http://www.numdam.org/item?id=RSA_1974__22_2_51_0

© Société française de statistique, 1974, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

UN ESSAI D'ANALYSE STATISTIQUE DES SÉRIES DE DONNÉES RESPIRATOIRES (1)

Mme BÉNCHETRIT et M. PHAM-DINH-TUAN



1. INTRODUCTION

Les données considérées dans ce texte sont les amplitudes et périodes cycle par cycle de la respiration enregistrée chez des sujets humains au repos. En effet, chaque cycle respiratoire est identifié par deux mesures ; une amplitude notée V qui représente le volume inspiré ou expiré et une durée notée T , qui est la période du cycle. Ces valeurs cycle par cycle de la respiration ont des variations non négligeables. Ceci nous a amené à entreprendre une étude statistique pour tenter de décrire le comportement respiratoire au repos. Sur des séries ininterrompue de 43 à 247 cycles, nous avons utilisé les méthodes statistiques pour étudier la dépendance sérielle des termes successifs dans les séries vectorielles de ces mesures et nous avons cherché à déterminer un modèle simple qui représenterait approximativement le phénomène.

2. PRESENTATION DES DONNEES

Nous avons 25 séries classées en 4 groupes, chaque groupe correspond à des séries relatives à un sujet humain donné. Ces séries sont de longueurs variables, qui contiennent entre 43 à 247 termes. Chaque terme correspond à deux mesures, à savoir le volume et la durée de chaque cycle respiratoire.

Pour avoir une idée des grandeurs de ces séries, nous avons reproduit dans la table 1, une série dans le groupe 3 identifiée par M.S.6 qui comportent 123 mesures.

Nous avons tracé les courbes représentatives de V_i et de T_i en fonction du numéro d'ordre i , et en même temps les courbes des moyennes mobiles sur 5 valeurs consécutives, c'est-à-dire les courbes des \bar{V}_i et \bar{T}_i où :

(1) article remis le 19/1/73, révisé le 25/5/73

$$\bar{V}_i = \frac{1}{5} \sum_{h=-2}^2 V_{i+h}$$

$$\bar{T}_i = \frac{1}{5} \sum_{h=-2}^2 T_{i+h}$$

L'examen des courbes en moyenne mobile ne révèle aucune tendance particulière des séries étudiées. (voir figures 1 et 2). Nous pouvons penser que ces séries ont la moyenne constante au cours du temps.

3. CORRELOGRAMME ET SPECTRE

Pour toute série chronologique vectorielle $X(j)$, $j = 1, \dots, N$, nous définissons les matrices de covariances empiriques d'ordre h ($h = 0, 1, 2, \dots$) par :

$$R^N(h) = \frac{1}{N-h} \sum_{j=1}^{N-h} [X(j) - \bar{X}]^t [X(j+h) - \bar{X}]$$

où

$$X = \frac{1}{N} \sum_{i=1}^N X_i$$

est la moyenne des observations.

Dans le cas présent, $X(j) = (V_j, T_j)$, nous avons :

$$R_{11}(h) = \frac{1}{N-h} \sum_{j=1}^{N-h} (V_j - \bar{V})(V_{j+h} - \bar{V})$$

$$R_{12}(h) = \frac{1}{N-h} \sum_{j=1}^{N-h} (V_j - \bar{V})(T_{j+h} - \bar{T})$$

$$R_{21}(h) = \frac{1}{N-h} \sum_{j=1}^{N-h} (T_j - \bar{T})(V_{j+h} - \bar{V})$$

$$R_{22}(h) = \frac{1}{N-h} \sum_{j=1}^{N-h} (T_j - \bar{T})(T_{j+h} - \bar{T})$$

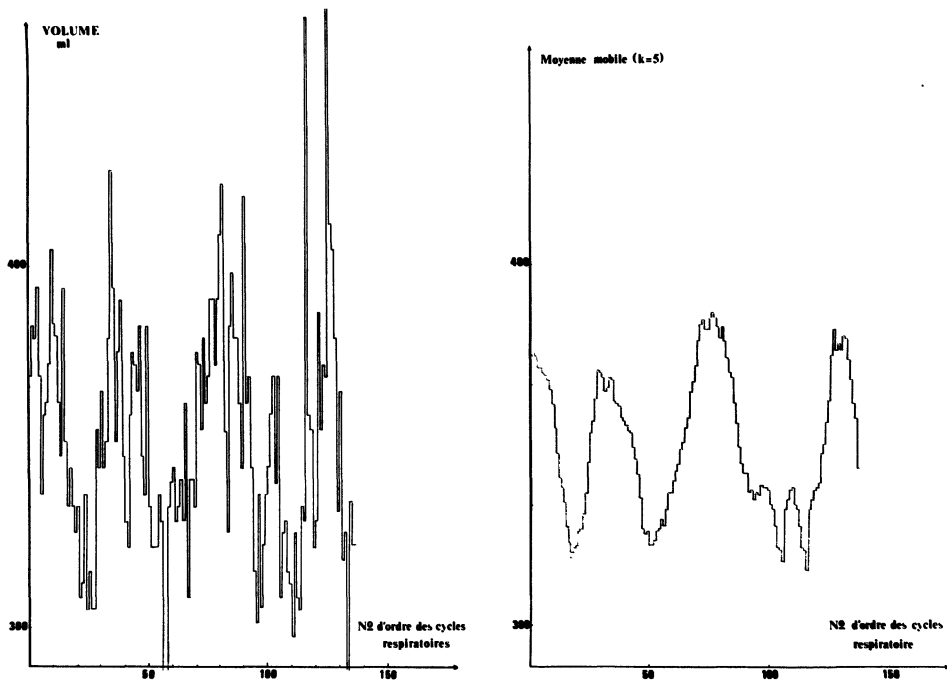


Figure 1 - M.S.6

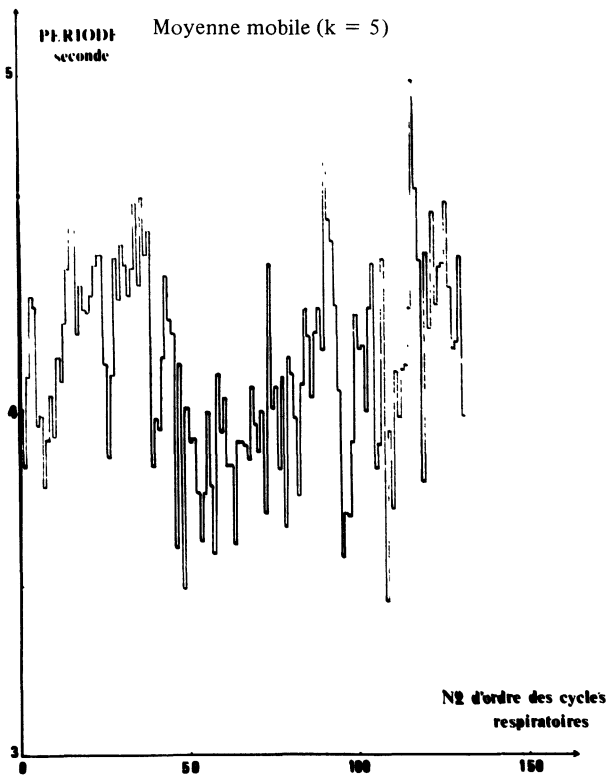


Figure 2 - M.S.6

Les coefficients de corrélation empiriques sont alors donnés par :

$$\rho_{11}(h) = \frac{R_{11}^N(h)}{R_{11}^N(0)}$$

$$\rho_{12}(h) = \frac{R_{12}^N(h)}{\sqrt{R_{11}^N(0) R_{11}^N(h)}}$$

$$\rho_{21}(h) = \frac{R_{21}^N(h)}{\sqrt{R_{11}^N(0) \cdot R_{22}^N(0)}}$$

$$\rho_{22}(h) = \frac{R_{22}^N(h)}{R_{22}^N(0)}$$

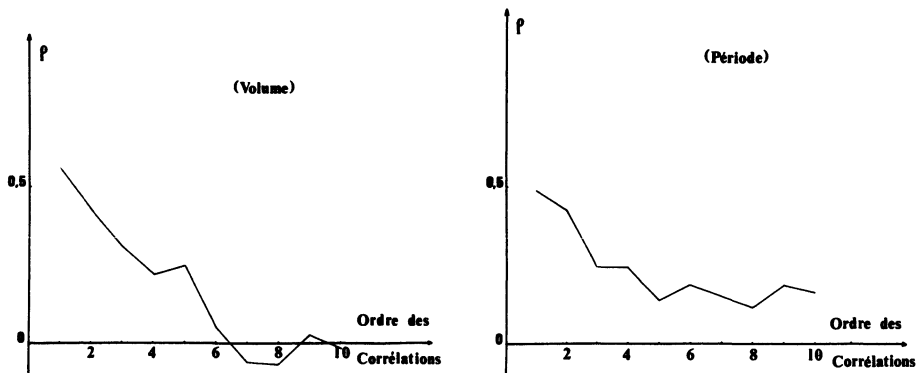


Figure 3 - Corrélogramme M.S.6

Les courbes des $\rho_{ij}^N(h)$ en fonction de h constituent les corrélogrammes (voir figure 3). Un examen rapide montre que les premières corrélations, à savoir $\rho_{11}^N(1)$ et $\rho_{22}^N(1)$ sont significativement différentes de 0 (sous l'hypothèse que les termes successifs de la série des observations sont indépendantes, ces coefficients peuvent être considérés en première approximation comme des corrélations empiriques ordinaires). Les observations ne sont donc pas sériellement indépendantes. De plus, les coefficients de corrélation d'ordre supérieur à 1 ne s'éteignent pas rapidement et restent souvent significatives jusqu'à un ordre élevé, ce qui suggère l'adéquation de ces séries à un modèle auto-régressif que nous allons étudier dans le paragraphe suivant.

L'étude des corrélations de différents ordres d'une série temporelle fournit des indications sur la dépendance des termes successifs de cette série. Elle doit être complétée par l'étude de la fonction matricielle de densité spectrale qui est la transformée de Fourier de la fonction des covariances (théoriques) :

$$f(\lambda) = \sum_{h=-\infty}^{\infty} R(h) e^{-ixh}$$

où $R(h)$ est la matrice des covariances (théoriques) d'ordre h . En pratique on travaille plutôt sur la densité spectrale normée, qui est la transformée de Fourier de la fonction des corrélations (théoriques) :

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\lambda h}$$

Ecrivons la matrice (hermitienne positive) $f(x)$ sous la forme :

$$f(\lambda) = \begin{pmatrix} I(\lambda) & C(\lambda) - iQ(\lambda) \\ C(\lambda) - iQ(\lambda) & J(\lambda) \end{pmatrix}$$

$I(\lambda)$ et $J(\lambda)$ sont appelés respectivement densité spectrale normée du processus $\{V_i\}$ et celle du processus $\{T_i\}$ et $C(\lambda)$ et $Q(\lambda)$ respectivement la densité co-spectrale et quadrature spectrale (normées) entre les processus $\{V_i\}$ et $\{T_i\}$

D'une manière générale, la densité spectrale (normée) mesure l'amplitude relative de la composante harmonique de fréquence $2\pi/\lambda$ de la série temporelle en question et la densité co-spectrale (resp. quadrature - spectrale) normée mesure la covariance relative entre les composantes harmoniques de même fréquence $2\pi/\lambda$ de deux séries temporelles, de même phase (resp. en quadrature - de phase).

Pour l'estimation de la densité spectrale correspondant au processus observé, nous utilisons la formule Bartlett (Hamnan [5], p. 278). Notons $I^N(\lambda)$, $J^N(\lambda)$ les estimateurs de la densité spectrale normée des séries des V_i et des T_j , on a :

$$I^N(\lambda) = \frac{1}{2\pi} \left[1 + 2 \sum_{h=1}^{M-1} \left(1 - \frac{h}{M}\right) \rho_{11}^N(h) \cos(\lambda h) \right]$$

$$J^N(\lambda) = \frac{1}{2\pi} \left[1 + 2 \sum_{h=1}^{M-1} \left(1 - \frac{h}{M}\right) \rho_{22}(h) \cos(\lambda h) \right]$$

Les co-spectre et quadrature-spectre estimés entre la série des V_j et celle des T_j sont :

$$C^N(\lambda) = \frac{1}{2\pi} \rho_{12}(0) + \sum_{h=1}^{M-1} [\rho_{12}(h) + \rho_{21}(h)] \left(1 - \frac{h}{M}\right) \cos \lambda h$$

$$Q^N(\lambda) = \frac{1}{2\pi} \sum_{h=1}^{M-1} [\rho_{12}(h) - \rho_{21}(h)] \left(1 - \frac{h}{M}\right) \cos(\lambda d).$$

Nous calculons ces fonctions pour des valeurs de $\lambda = \lambda_k = \pi/M$, $k = 0, 1, \dots, M$. La constante M détermine le biais et la variance des estimateurs spectraux dans les formules précédentes. En effet, la variance asymptotique de $I^N(\lambda)$ (ou $J^N(\lambda)$) est donnée par (Hannan [4]) :

$$\sigma^2 [I^N(\lambda)] \sim \frac{2}{3} \frac{M}{N} [I^N(\lambda)]^2$$

ce qui correspond à un coefficient de variation

$$\nu = \frac{\sigma^2 (I^N(\lambda))}{(I^N(\lambda))^2} \sim \frac{2}{3} \frac{M}{N}$$

Le biais de l'estimateur spectral dépend en général du degré de lissage de la fonction densité spectrale, mais on sait qu'une bande de l'ordre π/M ce qui correspond à une bande de $1/(2M)$ (en terme de fréquence) serait mélangée dans l'estimation de la densité spectrale au point λ . C'est pourquoi, nous calculons ces estimateurs à des points équidistants de π/M .

L'analyse spectrale a pour but de déceler les fréquences dominantes. L'examen des courbes de densité spectrale estimée (voir figures 4 et 5) ne fournit aucune évidence des composants harmoniques dans le domaine de fréquence plus grand que $1/60$ (c'est-à-dire des périodes plus grandes que 60 cycles respiratoires). Toutefois, les courbes de la densité spectrale estimée sont relativement grandes au voisinage de 0, et on voit apparaître dans le cas de la série D.P. 4 (voir figure 9 bis) qui comporte 247 observations, un pic à la fréquence environ $1/100$, sur la courbe de densité spectrale estimée. Il y a probablement une oscillation du rythme respiratoire de période assez grande, mais compte tenu des longueurs de nos séries, on ne peut pas donner une réponse sérieuse à cette question ni déterminer la période d'oscillation en question.

4. ADEQUATION DU MODELE AUTO-REGRESSIF

Les résultats de l'analyse précédents nous ont inspiré l'utilisation d'un modèle autofégressif pour décrire le comportement de ces séries des données respiratoires. Ce modèle permettrait de prédire le volume et la durée d'un cycle connaissant ceux des cycles précédents.

Le modèle auto-régressif général (d'ordre q) s'écrit :

$$X(j) - X = \sum_{k=1}^q A(k) [X(j-k) - X] + \epsilon(j)$$

où $A(1), \dots, A(q)$ sont des matrices des coefficients d'auto-régressifs et $\epsilon(j)$ est le vecteur (de composants $\epsilon_1(j), \epsilon_2(j)$) des résidus qui résume les effets du hasard. Ce sont des vecteurs aléatoires centrés indépendants de même loi, ayant une matrice de covariance G .

Si on examine les coefficients de corrélation empiriques croisés entre les V_j et les T_j , c'est-à-dire les coefficients $\rho_{12}(h)$, $h = 0, 1, \dots$, et $\rho_{21}(h)$, $h = 1, 2, \dots$, on constate que pour les séries vectorielles dont nous disposons, très souvent ces coefficients ne sont pas significativement différents de 0. Nous considérons le modèle le plus simple qui lie seulement les V_j (resp. les T_j) entre eux, à savoir :

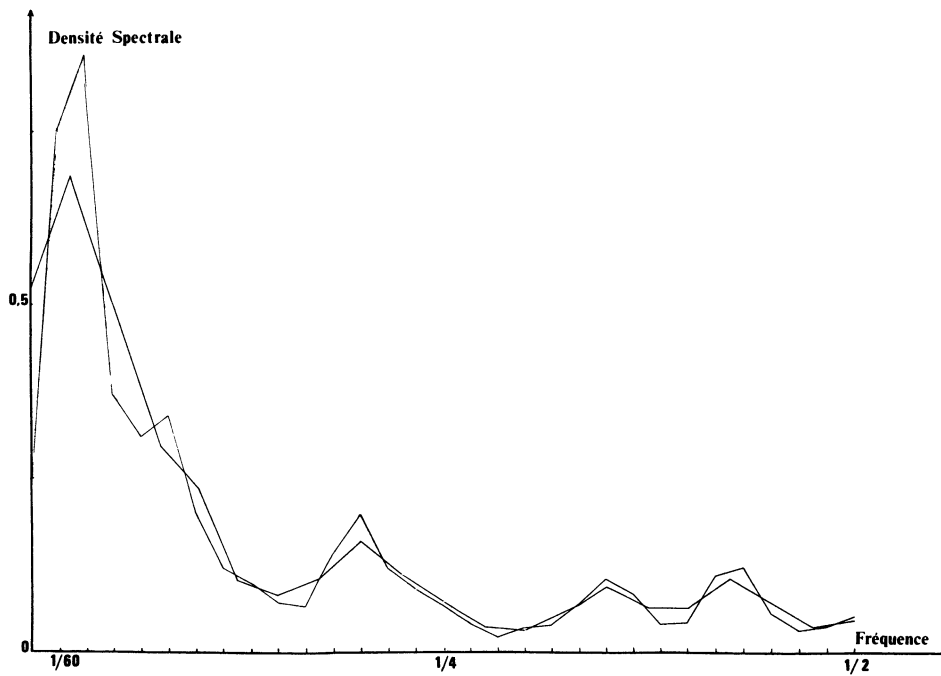


Figure 4 – Analyse spectrale (volume) M.S.6
(Les courbes correspondant à $M = 20,30$)

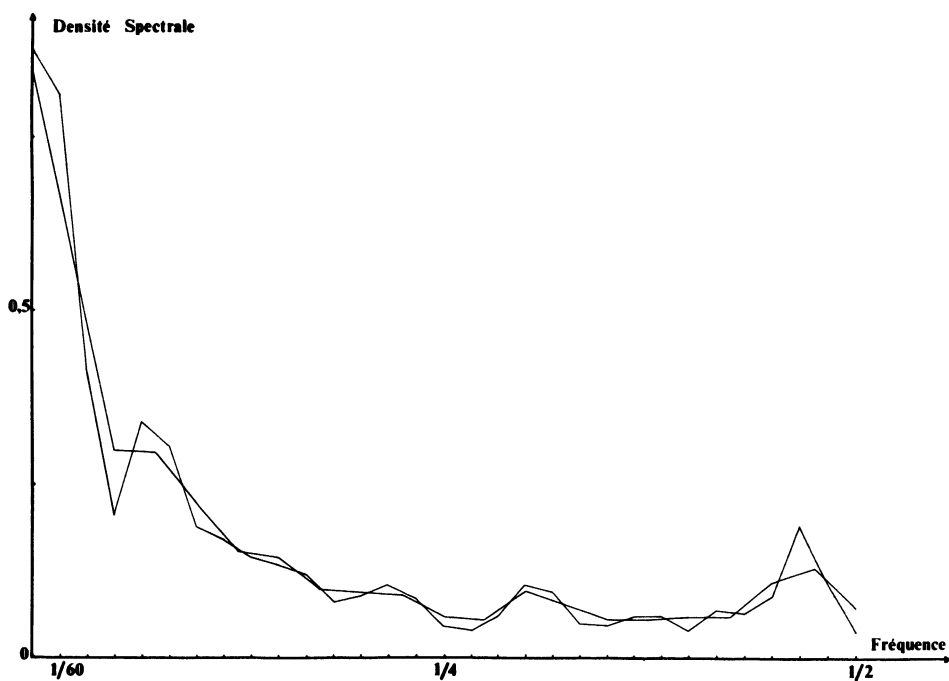


Figure 4 bis – Analyse spectrale (periode) M.S.6
(Les courbes correspondant à $M = 20,30$)

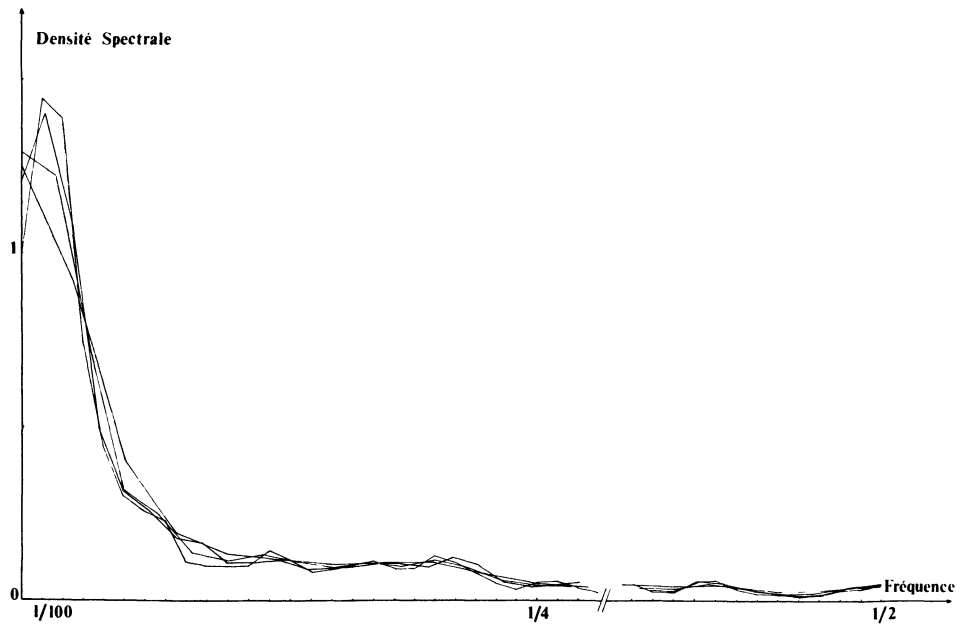


Figure 5 – Analyse spectrale (volume) D.P.4
 (Les courbes correspondent à $M = 20,30, 40,50$)

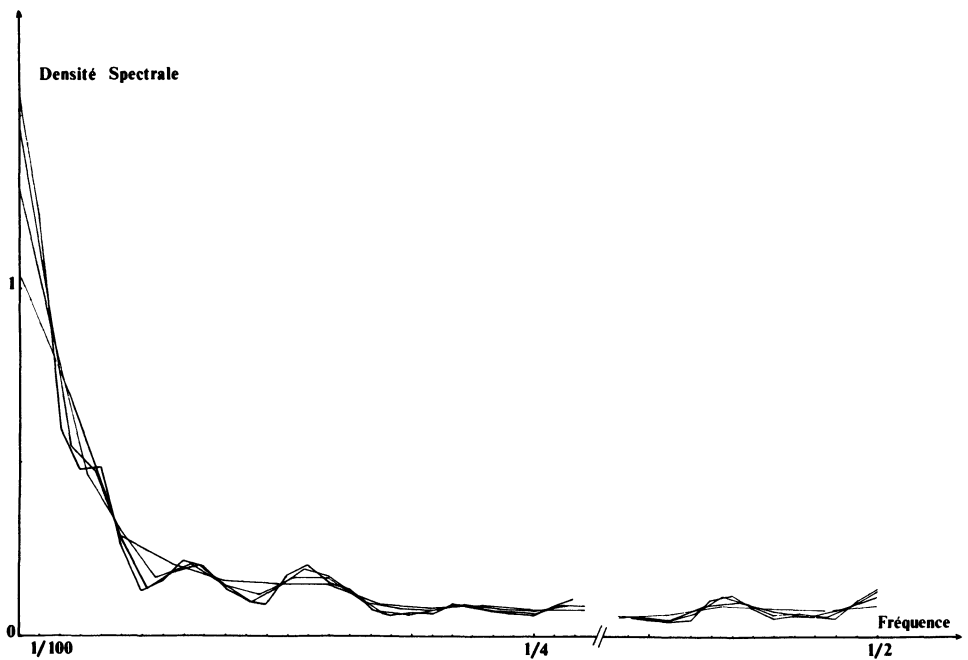


Figure 5 bis – Analyse spectrale (période) D.P.4
 (Les courbes correspondent à $M = 20,30, 40,50$)

Table 1

Adéquation du modèle auto-régressif

$$(V_i - \bar{V}) = a(V_{i-1} - \bar{V}) + \epsilon_i$$

$$(T_i - \bar{T}) = b(T_i - \bar{T}) + \eta_i$$

Identification de la série	Taille N	Coeff. a	Variance de ϵ_i	Coeff. b	Variance de η_i
F.C.1	130	0.732*	3706	0,815*	0.204
F.C.2	115	0.342	3624	0.589	0.274
F.C.3	135	0.517	1577	0.688	0.164
F.C.4	155	0.555*	1043	0.538	0.163
F.C.5	100	0.573*	1791	0.678	0.183
F.C.6	170	0.580*	3407	0.668*	0.156
D.P.1	56	0.511	2101	0.381	0.038
D.P.2	43	0.570	1566	0.466	0.047
D.P.3	51	0.517	2089	0.650	0.029
D.P.4	247	0.532*	1658	0.728	0.037
D.P.5	91	0.748	1834	0.613	0.084
D.P.6	138	0.561*	2225	0.545	0.054
M.S.1	65	0.674	2613	0.347	0.071
M.S.2	78	0.655*	2912	0.266	0.098
M.S.3	80	0.517	3553	0.441*	0.077
M.S.4	59	0.646	2804	0.707	0.067
M.S.5	84	0.649	2001	0.470	0.056
M.S.6	137	0.562	1773	0.491*	0.064
M.S.7	126	0.549*	1969	0.460	0.074
M.S.8	134	0.738*	3978	0.581*	0.132
A.T.1	165	0.628*	749	0.454	0.047
A.T.2	141	0.627*	851	0.618*	0.034
A.T.3	163	0.645*	870	0.596	0.025
A.T.4	79	0.484	535	0.633	0.035
A.T.5	113	0.757	594	0.366	0.042

$$V_j - V = \sum_{k=1}^q a_k (V_{j-k} - V) + \epsilon_j$$

$$\text{(resp. } T_j - T = \sum_{k=1}^q b_k (V_{j-k} - V) + \eta_j \text{)}$$

Nous voulons tester le modèle d'ordre 1, qui s'écrit :

$$V_j - V = a(V_{j-1} - V) + \epsilon_j$$

$$T_j - T = b(T_{j-1} - T) + \eta_j$$

Pour effectuer le test, il faut fixer un ordre q maximum. Nous prenons $q = 3$ et nous utilisons le test de Quenouille qui est basé sur la statistique (Anderson [1], p. 219, Grenander & Rosenblatt [3] p. 109).

$$S = \frac{\{\rho^N(2) - [\rho^N(1)]^2\}^2 + \{\rho^N(3) - 2\rho^N(2)\rho^N(1) + [\rho^N(1)]^3\}^2}{\{1 - [\rho^N(1)]^2\}^2}$$

où $\rho^N(h)$, $h = 1, 2, \dots$, sont les coefficients de corrélation empirique d'ordre h (bien entendu, on calculera la statistique S pour la série des V_j et celle des T_j respectivement). Rappelons que le test de Quenouille dans le cas général est basé sur la statistique :

$$\sum_{j=q_0+1}^q [h_j^N / \hat{\sigma}^2]^2$$

où q_0 est l'ordre d'auto-régression du modèle à tester et q est l'ordre auto-régression maximum, $\hat{\sigma}^2$ l'estimateur de la variance σ^2 des résidus et

$$h_j^N = \sum_{k=0}^q \sum_{l=0}^{q_0} R^N(j-k-l) a_k^N a_l^N$$

où $a_0^N = -1$, a_k^N , $k = 1, \dots, q_0$ sont les estimateurs des coefficients a_k basé sur les covariances empiriques $R^N(h)$, $h = 0, 1, \dots, q_0$ (dans le cas $q_0 = 1$, on a $a_1^N = \rho^N(1)$). Il s'ensuit que h_j^N est un estimateur consistant de :

$$h_j = \sum_{k=0}^{q_0} \left[\sum_{l=0}^{q_0} R^N(j-k-l) a_k \right] a_l ; \quad a_0 = -1$$

qui est nul pour $j > q_0$ dans le cas où le modèle auto-régressif d'ordre q_0 car le crochet de l'expression précédente l'est.)

D'autre part, on démontre que $\sqrt{N} h^N / \sigma^2$ est asymptotiquement normale centrée réduite ; par conséquent sous l'hypothèse du modèle auto-régressif d'ordre 1, la variable aléatoire N.S. converge en loi quand $N \rightarrow \infty$, vers une variable du χ^2 à 2 degrés de liberté. Nous rejetons donc l'hypothèse du modèle auto-régressif d'ordre 1 au seuil α si

$$NS > \chi_2^2(\alpha)$$

où $\chi_2^2(\alpha)$ est telle que $P\{X > \chi_2^2(\alpha)\} = \alpha$ où X suit la loi du χ^2 à 2 degrés de liberté. Pour $\alpha = 5\%$, on a $\chi_2^2(\alpha) = 5.99$.

Parmi les 25 séries dont nous disposons, le test de Quenouille a rejeté l'hypothèse du modèle auto-régressif d'ordre 1 dans 12 cas pour la variable V et dans 7 cas pour la variable T (voir table 2 où un astérisque * correspond aux cas où le test rejette l'hypothèse).

On constate qu'en général les cas où l'hypothèse est rejetée, correspondant aux longues séries. Ceci peut s'expliquer de deux manières suivantes :

– Le test étant consistant, par conséquent, plus on dispose des observations, plus on a de chances de déceler la différence entre la réalité et le modèle présumé, ce dernier n'est probablement qu'une première approximation.

– Malgré toutes les précautions utiles concernant l'expérimentation, il est possible que la série observée ne corresponde pas tout à fait à un processus stationnaire et par conséquent pour les longues séries, le modèle auto-régressif ne convient plus.

Ces arguments sont appuyés par le fait que si on découpe les longues séries en des séries plus courtes, et applique le test de Quenouille, on est amené à accepter l'hypothèse du modèle d'ordre 1 dans la plupart des cas.

Une manière alternative de procéder est de calculer les corrélations partielles respectivement d'ordre 1, 2, 3, ... Ce sont les corrélations entre V_j et V_{j-k} (ou T_j et T_{j-h}) ayant éliminé l'effet de $V_{j-1}, \dots, V_{j-h+1}$ (ou $T_{j-1}, \dots, T_{j-h+1}$), que l'on notera $\rho(h|1, \dots, h-1)$. Dans le cas d'un processus stationnaire, appelons $a(j|h)$, $j = 1, \dots, h$ les coefficients relatifs au modèle auto-régressif d'ordre h, définis par les relations :

$$V_j = \sum_{k=1}^h a(k|h) V_{j-k} + \epsilon_j(h)$$

où $\epsilon_j(h)$ est non corrélé avec V_{j-1}, \dots, V_{j-h} et est appelé résidu (autrement dit, on a fait une régression de V_j sur V_{j-1}, \dots, V_{j-h}), on peut alors montrer que le dernier coefficient $a(h|h)$ est égal à $\rho(h|1, \dots, h-1)$.

Pour estimer les coefficients $a(k|h)$, $k = 1, \dots, h$, nous nous servons du fait que ces coefficients satisfont aux équations de Yule-Walker :

$$\sum_{k=1}^k a(k, h) \rho(k-l) = l = 1, \dots, h$$

Si nous remplaçons dans ces équations les coefficients de corrélations théorique $\rho(k)$, $k = 1, \dots, h$ par les coefficients empiriques correspondantes $\rho_{(k)}^N$, $k = 1, \dots, h$ on obtient des estimateurs $a^N(k|h)$ de $a(k|h)$, $k = 1, \dots, h$.

Maintenant, si le processus est auto-régressif d'ordre q , les coefficients $a(h|h)$ pour $h > q$ sont nuls et on peut démontrer (Anderson [1], p. 123) que $\sqrt{N} a^N(h|h)$, $h > q$ sont asymptotiquement indépendantes, de loi normale centrée réduite. On peut appliquer ce résultat pour décider l'ordre d'auto-régression qui convient le mieux. On fixe d'abord un ordre maximum q_m et on détermine l'indice q à partir duquel les coefficients $a(h|h)$, $q < h \leq q_m$ ne sont pas significativement différents de 0. Pour un seuil de 5 %, cela revient à chercher le plus petit indice q tel que $|a^N(h|h)| < 2/\sqrt{N}$, $q < h \leq q_m$.

Si on applique la règle précédente pour les séries ou le test de-Quenouille a rejeté l'hypothèse du modèle d'ordre 1, on trouve que dans la plupart des cas l'ordre d'auto-régression trouvé est 2 ($q_m = 6$). Un exemple des calculs effectués est donné dans la table 3, les coefficients $a(k|h)$, $k = 1, \dots, h$ étant obtenu à l'aide de l'algorithme de Durbin (Box Jenkins [2], p. 82).

En conclusion, le modèle auto-régressif d'ordre 1 n'est probablement qu'une première approximation. Le processus en question est certainement plus complexe. Néanmoins, ce modèle a l'intérêt d'être simple et dans l'optique de la prédiction, il est tout à fait satisfaisant. En effet, si on compare les variances de $\epsilon(1)$ et $\epsilon(2)$ (ou de $\eta(1)$ et $\eta(2)$) on trouve un rapport assez proche de 1 (voir table 3). Par ailleurs, ce modèle correspond à l'idée de l'existence d'une mémoire qui conditionne les caractéristiques d'un cycle respiratoire à partir de celles des cycles précédents.

Table 2

Calcul des coefficients auto-régressifs de
différents ordres

Série AT.2 Variable V (141 observations)

Ordre	Var. résid. Variance		Coeff. d'auto-régressif				
1	0.607	0.627					
2	0.580	0.497	0.208				
3	0.574	0.477	0.160	0.096			
4	0.554	0.459	0.130	0.006	0.189		
5	0.554	0.459	0.130	0.006	0.189	-0.0008	
6	0.551	0.459	0.144	0.006	0.199	0.0033	-0.074

Série AT.2 Variable T (141 observations)

1	0.618	0.618					
2	0.591	0.490	0.206				
3	0.589	0.477	0.174	0.065			
4	0.588	0.476	0.170	0.054	0.023		
5	0.572	0.472	0.161	0.026	0.056	0.165	
6	0.572	0.469	0.162	0.026	0.059	0.158	0.016

5. ADEQUATION DU MODELE AUTO-REGRESSIF VECTORIEL D'ORDRE 1 -

Jusqu'à maintenant, nous avons étudié les séries des V_j et des T_j séparément au lieu de la série des vecteurs (V_j, T_j) . Il semble donc préférable d'adapter un modèle auto-régressif vectoriel qui fait intervenir les liaisons entre les V_j et T_j .

Le modèle auto-régressif vectoriel d'ordre 1 s'écrit :

$$V_j = a_{11} V_{j-1} + a_{12} T_{j-1} + \epsilon_j$$

$$T_j = a_{21} V_{j-1} + a_{22} T_{j-1} + \eta_j$$

La matrice A de termes généraux a_{ij} , $i = 1, 2, j = 1, 2$ peut être estimée par la matrice $\hat{A} = R^N (-1) [R(0)]^{-1}$ où $R^N(h)$ est la matrice de covariances empiriques d'ordre h .

D'autre part, le vecteur des résidus (ϵ_j, η_j) a une matrice de covariance G qui peut être estimée par :

$$R^N(0) - R^N(-1) [R^N(0)^{-1}] R^N(0)$$

Nous nous intéressons à tester la nullité des coefficients a_{12} et a_{21} dans le modèle précédent. Si ces coefficients ne sont pas significativement différents de 0, on peut alors admettre que les séries des V_j et des T_j suivent séparément des modèles auto-régressifs d'ordre 1.

Or nous savons que l'estimateur \hat{A} de la matrice A défini précédemment est presque sûrement convergent et que $\sqrt{N}(\hat{A} - A)$ est asymptotiquement gaussien dont la structure de covariance est donnée par (Hannan [5]) :

$$\text{cov}(\hat{a}_{ij}, \hat{a}_{kl}) \sim \frac{1}{N} G_{ik} [(R(0))^{-1}]_{jl}$$

Cette formule permet de calculer la variance des estimateurs \hat{a}_{ij} de a_{ij} ($i = 1, 2 ; j = 1, 2$) et par conséquent de tester la nullité de ces coefficients.

Les calculs effectués sont résumés dans la table 4. Nous voyons que pour la plupart des séries étudiées, les coefficients a_{12} et a_{21} ne sont pas significativement différents de 0. (Dans la table un * qui suit l'estimation d'un coefficient a_{ij} indique que ce coefficient est significativement différent de 0). Nous avons d'autre part calculé les coefficients de corrélation estimés entre ϵ_j et η_j . Si on assimile ce coefficient comme un coefficient de corrélation empirique ordinaire, ce qui est une première approximation, on voit que ce coefficient est parfois significativement différent de 0 et parfois non. Nous ne pouvons donc pas répondre sérieusement à la question de l'indépendance entre ϵ_i et η_i .

Table 3

Estimation des coefficients du modèle

$$(V_i - V) = a_{11} (V_{i-1} - V) + a_{12} (T_{i-1} - T) + \epsilon_i$$

$$(T_i - T) = a_{21} (V_{i-1} - V) + a_{22} (T_{i-1} - T) + \eta_i$$

Identification de la série	Taille	a_{11}	a_{12}	a_{21}	a_{22}	corr (ϵ_i, η_i)
F.C.1	130	0.395*	49,2*	$0,580 \times 10^{-3}$	0,762*	0,480*
F.C.2	115	0.278*	12,6	$0,365 \times 10^{-3}$	0,571*	0,450*
F.C.3	135	0.413*	32,2	$0,112 \times 10^{-3}$	0,683*	0,130
F.C.4	155	0.538*	36,5	$0,131 \times 10^{-2}$	0,499*	0,270
F.C.5	100	0.412*	25,1	$-0,146 \times 10^{-3}$	0,685*	0,380*
F.C.6	170	0.503*	17,7	$0,534 \times 10^{-3}$	0,625*	0,466*
D.P.1	56	0.500*	25,2	$-0,738 \times 10^{-3}$	0,402*	0,190
D.P.2	43	0.574*	- 6,0	$0,575 \times 10^{-3}$	0,450*	0,060
D.P.3	51	0,409*	62,9*	$-0,447 \times 10^{-3}$	0,650*	0,200
D.P.	247	0,471*	28,4*	$-0,593 \times 10^{-3}$	0,765*	0,285*
D.P.5	91	0,709*	30,3*	$-0,208 \times 10^{-5}$	0,613*	0,050*
D.P.6	138	0.477*	35,8*	$-0,312 \times 10^{-5}$	0,575*	0,041
M.S.1	65					
M.S.2	78	0.693*	-55,3*	$-0,535 \times 10^{-3}$	0,283*	0,395*
M.S.3	80	0.569*	-28,4	$-0,174 \times 10^{-3}$	0,458*	0,510*
M.S.4	59	0.531*	33,3	$0,576 \times 10^{-3}$	0,635*	0,479*
M.S.5	84	0.697*	-36,2	$0,149 \times 10^{-3}$	0,479*	0,447*
M.S.6	137	0.569*	3,24	$0,166 \times 10^{-3}$	0,481*	0,360*
M.S.7	126					
M.S.8	134					
A.T.1	165	0.594*	-14,2	$-0,181 \times 10^{-2}$	0,364*	-0,10
A.T.2	141	0,545*	-36,1*	$-0,119 \times 10^{-2}$	0,548*	-0,024
A.T.3	79	0,645*	- 6,84	$0,152 \times 10^{-2}$	0,595*	0,016
A.T.4	79	0,469*	-23,8*	$-0,135 \times 10^{-2}$	0,623*	0,240*
A.T.5	113	0,753*	- 7,01	$-0,340 \times 10^{-3}$	0,360*	-0,022

6. COMPARAISON DES COEFFICIENTS D'AUTO-REGRESSIFS POUR LES DIFFERENTES SERIES OBSERVEES —

Dans le paragraphe précédent, nous avons essayé d'adapter un modèle auto-régressif d'ordre 1 aux différentes séries mesurées. Une question qui se pose naturellement est de savoir si les coefficients correspondants sont les mêmes. Dans ce cas, il est probable que ces coefficients aient une signification intrinsèque

Nous commençons d'abord à tester si les coefficients correspondants à des séries relatives à un sujet donné sont les mêmes. Or, nous savons que pour le processus auto-régressif d'ordre 1, l'estimateur du coefficient d'auto-régression qui est aussi le premier coefficient de corrélation empirique $\rho^N(1)$ a pour espérance mathématique $\rho(1)$ et pour variance :

$$\sigma^2(\rho^N(1)) \sim \frac{1}{N(1 - \rho^2(1))}$$

Supposons maintenant que nous disposons k séries de mesures relatives à un sujet, de tailles respectives N_1, \dots, N_k . A la i -ième série, nous calculons le premier coefficient de corrélation empirique que nous noterons maintenant x_i qui a pour moyenne ρ_i et de variance $N_i^{-1}(1 - \rho_i^2)$. Or dans le cas où les variances de x_i sont connues, soit σ_i , le test optimal pour tester l'égalité des moyennes est basé sur la statistique

$$\chi^2 = \sum_{i=1}^k t_i (x_i - \bar{x})^2$$

où

$$t_i = \sigma_i^{-2}$$

$$\bar{x} = \frac{\sum_{i=1}^k (t_i / \sum_{j=1}^k t_j) x_i}{\sum_{i=1}^k t_i}$$

Sous l'hypothèse d'égalité des moyennes la statistique χ^2 suit une loi du χ^2 à $k - 1$ degrés de liberté, le test rejette donc l'hypothèse si $\chi^2 > \chi_{k-1}^2(\alpha)$.

Dans le cas qui nous intéresse, les t_i sont inconnues et données par $N_i(1 - \rho_i)^{-1}$, mais on peut remplacer le paramètre inconnu ρ_i par un estimateur constant, cela ne change pas la loi asymptotique de χ^2 . Comme sous l'hypothèse à tester toutes les ρ_i sont égales et admettent \bar{x} comme estimateur constant, nous utilisons la statistique de test :

$$\chi^2 = \frac{1}{(1 - \bar{x}^2)} \sum_{i=1}^k N_i (x_i - \bar{x})^2$$

ou

$$\bar{x} = \frac{1}{\sum_{i=1}^k N_i} \sum_{i=1}^k N_i x_i$$

Dans la table 5, nous avons calculé ce χ^2 pour chaque groupe de séries relatives à une personne. On voit que sauf pour les séries des V_j du 1^{er} groupe, les χ^2 calculés ne sont pas significatifs au niveau de 5 % ; et qu'au niveau de 1 %, tous les χ^2 calculés ne sont pas significatifs. Nous pouvons donc conclure que les coefficients auto-régressifs sont les mêmes pour l'ensemble des séries relatives à un sujet.

Nous allons maintenant comparer les coefficients auto-régressifs relatifs aux différents sujets. Pour le j -ième sujet, nous disposons k_j séries de mesures de tailles respectives $N_{1j}, \dots, N_{k_j j}$, chacune donne lieu à un coefficient auto-régressif estimé que nous noterons maintenant $x_{1j}, \dots, x_{k_j j}$. Nous admettons l'hypothèse que les coefficients auto-régressifs sont les mêmes pour les différentes séries correspondant à un sujet, cela veut dire que les variables aléatoires $(x_{1j}, \dots, x_{k_j j})$ ont même moyenne, soit ρ_j . L'hypothèse que l'on veut tester est l'égalité des ρ_j . C'est donc un problème de test d'hypothèse linéaire à ceci près que les x_{ij} n'ont pas la même variance. Pour cela nous considérons le vecteur aléatoire Y de composants $Y_{ij} = \sqrt{N_{ij}} X_{ij}$, $i = 1, \dots, k_j, j = 1, \dots, p$ à valeurs dans R^E où :

$$E = \{(1, 1), \dots, (k_1, 1)\} \cup \dots \cup \{(1, p), \dots, (k_p, p)\}.$$

Le vecteur moyenne EY de Y appartient au sous-espace

$$H = \{m \in R^E, m_{ij} = \sqrt{N_{ij}} \rho_j\}$$

et l'hypothèse linéaire que l'on peut tester est défini par le sous-espace

$$H_0 = \{m \in R^E ; m_{ij} = \sqrt{N_{ij}} \rho_j\}$$

Un calcul élémentaire montre que la projection de Y sur H a pour composants $\sqrt{N_{ij}} \bar{x}_j$, $i = 1, \dots, k_j, j = 1, \dots, p$, où :

$$\bar{x}_j = \frac{1}{N_j} \sum_{i=1}^{k_j} N_{ij} x_{ij} \quad (N = \sum_{j=1}^p N_j).$$

De même la projection de Y sur H_0 a pour composants :

$$\sqrt{N_{ij}} \bar{x}, i = 1, \dots, k_j, j = 1, \dots, p \text{ où :}$$

$$\bar{x} = \frac{1}{N} \sum_{j=1}^p N_j \bar{x}_j \quad (N = \sum_{j=1}^p N_j).$$

Le test d'égalité des ρ_i est basé sur le rapport F dont la norme carrée de la projection de Y sur H-H₀, divisé par sa dimension et celle de Y sur R^E-H, divisée par sa dimension. Par conséquent, le numérateur de F est

$$\sum_{j=1}^k N_j (\bar{x}_j - \bar{x})^2 / (p - 1)$$

et le dénominateur est :

$$\sum_{j=1}^k \sigma_j^2 / \left(\sum_{j=1}^p k_j - p \right)$$

où :

$$\sigma^2 = \sum_{i=1}^{k_j} N_{ij} (x_{ij} - \bar{x}_j)^2$$

En appliquant ce test, on voit que pour les 4 sujets étudiés, le rapport F n'est pas significatif en ce qui concerne la variable V et quant à la variable T, ce rapport est significatif au seuil 5 % mais n'est pas significatif au seuil 1%. La discordance entre la série des V_j et celle des T_j semble anormale. Toutefois, on peut remarquer que dans le test du modèle auto-régressif d'ordre 1, pour la série de V_j, assez souvent ce modèle est rejeté ce qui n'est pas le cas de la série des T_j.

Table 4
Comparaison des coefficients d'auto-régressif

Personne	Nb. de séries	Nb. Tot. d'obs.	Variable V			Variable T		
			Moy. pond	Disper-sion	χ^2	Moy. pond	Disper-sion	χ^2
F.C.	6	805	0,568	9,551	14,1	0,660	6,158	10,91
D.T.	6	626	0,569	3,421	5,06	0,615	4,564	7,34
M.S.	8	763	0,621	4,208	6,85	0,475	9,316	12,03
A.T.	5	601	0,637	3,512	5,91	0,530	6,637	9,23

25	2885	0,568	20,69	0,570a	26,67
(Tot.)	(Total)	(Moy. pond)	(Total)	(Moy. pond)	(Total)

Rapport F pour la variable V : 0,899
 „ „ T : 4,13

7. CONCLUSION

Il est important de noter que ce qui précède n'est qu'un essai d'analyse statistique d'un phénomène très complexe. Il en ressort que sa description à l'aide de notre modèle est nécessairement incomplète et les résultats obtenus ne peuvent pas être considérés comme définitifs. Toutefois, ce modèle est adéquat, en première approximation pour décrire le comportement respiratoire et sera utile pour la prédiction des caractéristiques d'un cycle respiratoire à partir de ceux des cycles précédents.

BIBLIOGRAPHIE

- [1] ANDERSON T.W – The Statistical analysis of time series. Wiley New-York (1971).
- [2] BOX G.P.E., ROSENBLATT M. – Time series analysis, forecasting and control. Holden day - San Fransisco (1970).
- [3] GRENDER U., ROSENBLATT M. – Statistical analysis of stationary time series. Wiley, New-York (1957).
- [4] HANNAN E.J. – Time serie analysis. Methuen. London (1960).
- [5] HANNAN E.J. – Multiple time series. Wiley, New-York (1970).