

SEPTIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 28 MAI au 2 JUIN 1979

MODELISATION D'UN ENREGISTREMENT DE BRUIT MARIN ET DETECTION

Denis de BRUCC

Université de Rouen - Laboratoire de Traitement de l'Information B.P 67 76130 Mt St AIGNAN.
Ce travail a été effectué dans le cadre d'un contrat Cethedec.

RESUME

L'identification part d'un modèle auto-regressif, modifié afin de tenir compte des propriétés statistiques particulières du bruit marin. L'énergie varie de façon importante : toutes les 2×10^{-2} secondes tandis que le spectre normalisé varie plus lentement. L'énergie $\sigma^2(L)$ du bruit pour la tranche L de temps, est calculée à partir des coefficients $A(L-1)$ du modèle autorégressif et ces coefficients s'obtiennent de façon récursive de tranche à tranche. Par contre, l'innovation $E(i)$ est fournie à chaque instant $i \Delta t$.

Les performances du procédé sont testées en simulation. Des enregistrements de bruit marin sont traités. Les propriétés de l'innovation sont fournies ainsi que celles des fluctuations énergétiques. Les coefficients $A(L)$ sont comparés avec ceux du modèle interne, (description auto-régressive à moyenne mobile).

SUMMARY

The identification starts from auto-regressive model, modified to take account of the special sea noise statistical properties. The energy changes abruptly over blocks of 2×10^{-2} sec., whereas the normalized spectra varies much slowly. The noise energy, $\sigma^2(L)$ of the block L , is computed with the use of the coefficients $A(L-1)$ for the auto regressive model of block $(L-1)$ and these coefficients are furnished by a block to block recursive method. On the other hand, the innovation $E(i)$ is provided at every moment $i \Delta t$.

First the performances of the procedure are analysed by simulation. Then the sea noise data are treated. The innovation properties are precised just as energy fluctuation ones. The $A(L)$ coefficients are compared with those of the internal model - autoregressive moving average model.



MODELISATION D'UN ENREGISTREMENT
DE BRUIT MARIN ET DETECTION

(1) Modèle A.R.M.A.

Le modèle autorégressif à moyenne mobile A.R.M.A. permet une première description des propriétés spectrales du bruit marin $Y(t); t \in [0, T]$.

Notons $(Y(i); i \in \mathbb{Z})$, l'enregistrement discrétisé suivant le temps de ce bruit. Pour intervalle de discrétisation, on a choisi $\Delta t = 1.25 \cdot 10^{-4}$ secondes.

Le modèle A.R.M.A. suppose la liaison statistique suivante entre ces valeurs $\forall i \in \mathbb{Z}$

$$Y(i) = A_1 Y(i-1) + \dots + A_p Y(i-p) + B_0 e(i) + B_1 e(i-1) + \dots + B_q e(i-q)$$

où A_1, \dots, A_p sont les p coefficients de la partie autorégressive et où B_1, \dots, B_q sont les q coefficients de la partie moyenne mobile. La suite $(e(i); i \in \mathbb{Z})$ constitue une famille orthonormale de variables aléatoires centrées.

On posera $\hat{Y}(i) \triangleq A_1 Y(i-1) + \dots + A_p Y(i-p)$ prévision linéaire du présent $Y(i)$, à partir du passé $Y(i-1), \dots, Y(i-p)$

La densité spectrale f de la solution stationnaire du modèle autorégressif vérifie

$$f(\omega) = \left| \frac{B_0 + B_1 e^{2i\pi\omega\Delta t} + B_2 e^{4i\pi\omega\Delta t} + \dots + B_q e^{2iq\pi\omega\Delta t}}{1 - A_1 e^{2i\pi\omega\Delta t} - A_2 e^{4i\pi\omega\Delta t} - \dots - A_p e^{2ip\pi\omega\Delta t}} \right|^2$$

Pour la solution stationnaire centrée, la covariance vérifie

$$\forall i \in \mathbb{Z} \quad \forall j \in \mathbb{Z} \quad \Gamma(i, i+j) \triangleq E(Y(i)Y(i+j)) = \Gamma(j)$$

approché par

$$\frac{1}{N-j} \sum_{l=j+1}^N Y(l) Y(l-j)$$

Les coefficients A_1, \dots, A_p du modèle A.R.M.A. vérifie l'équation

$$(A_1, A_2, \dots, A_p) \begin{pmatrix} \Gamma(q) & \Gamma(q+p-1) \\ \Gamma(q-p+1) & \Gamma(q) \end{pmatrix} = [\Gamma(q+1), \dots, \Gamma(q+p)]$$

Les régularités statistiques du bruit marin relèvent elles, entièrement de ce modèle ?

(2) Propriétés particulières du bruit traité.

Le bruit $Y(t); t \in \mathbb{R}$ est centré. La stationnarité statistique de l'observation Y doit être remise en question : l'énergie $E(Y(t)^2)$ varie rapidement mais cependant sur des durées $K\Delta t$ avec K à préciser, l'énergie de l'erreur $Y(i) - \hat{Y}(i), i \in \mathbb{Z}$ est considérée comme constante et estimée par

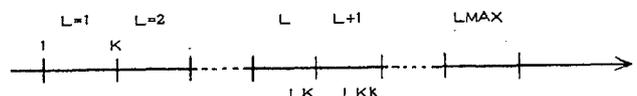
$$\bar{y}^2 \triangleq \frac{1}{K} \sum_{LK}^{LKK} (Y(i) - \hat{Y}(i))^2$$

avec $LKK \triangleq LK+K$ et LK , a priori, quelconque.

Pour avoir des effets de moyenne, on prend en statistique $K \geq 30$ et nous retenons la valeur $K = 160$ ce qui suppose l'énergie de l'erreur, stationnaire sur des durées de l'ordre de $K \times \Delta t = 2 \cdot 10^{-2}$ sec.

(3) Algorithme d'identification.

Les observations $(Y(i); i \in \mathbb{N})$ arrivent ainsi par paquets de K valeurs, paquets que nous indiquons par $L \in \mathbb{N}$ et l'estimation des A peut s'effectuer de façon récursive.



Etape L :

$$LK \triangleq (L-1)K+1$$

$$E(L) \triangleq Y(L) - \hat{Y}(L)$$

$$\bar{y}^2(L+1) \triangleq \frac{1}{K} \sum_{LK}^{LKK} E(i)^2$$

l'estimation linéaire \hat{Y} s'effectuant avec les coefficients A de l'étape antérieure L .



MODELISATION D'UN ENREGISTREMENT
DE BRUIT MARIN ET DETECTION

ALPHA(L)=	0.843E+00	BETA(L)=-	0.164E+00	GM(1,1)=	0.30E+02	GM(1,2)=	0.72E+02	MDYE(L)=-	0.720E-03	SA42(L)=	0.1365E+00
A(1,.)=	0.84E+00	-0.16E+00	0.32E+00	-0.52E-01	0.29E-01	0.13E-01	-0.11E+00	-0.24E-01	-0.32E-01	0.28E-01	
	-0.25E-01	0.66E-01	-0.26E-01								
ALPHA(L)=	0.848E+00	BETA(L)=-	0.199E+00	GM(1,1)=	0.78E+02	GM(1,2)=	0.70E+02	MDYE(L)=-	0.229E-01	SA42(L)=	0.1545E+00
A(1,.)=	0.85E+00	-0.20E+00	0.35E+00	-0.55E-01	0.35E-01	-0.34E-02	-0.35E-01	-0.25E-01	-0.25E-01	0.34E-01	
	-0.36E-01	0.59E-01	-0.14E-01								
ALPHA(L)=	0.861E+00	BETA(L)=-	0.215E+00	GM(1,1)=	0.77E+02	GM(1,2)=	0.53E+02	MDYE(L)=-	0.433E-01	SA42(L)=	0.1490E+00
A(1,.)=	0.86E+00	-0.22E+00	0.35E+00	-0.74E-01	0.55E-01	-0.32E-01	-0.79E-01	-0.19E-01	-0.16E-01	0.37E-01	
	-0.39E-01	0.52E-01	-0.29E-01								
ALPHA(L)=	0.884E+00	BETA(L)=-	0.234E+00	GM(1,1)=	0.76E+02	GM(1,2)=	0.53E+02	MDYE(L)=-	0.337E-01	SA42(L)=	0.1535E+00
A(1,.)=	0.88E+00	-0.23E+00	0.35E+00	-0.70E-01	0.70E-01	-0.34E-01	-0.33E-01	-0.23E-02	-0.31E-01	0.31E-01	
	-0.40E-01	0.44E-01	0.15E-01								
ALPHA(L)=	0.826E+00	BETA(L)=-	0.181E+00	GM(1,1)=	0.73E+02	GM(1,2)=	0.54E+02	MDYE(L)=-	0.509E-01	SA42(L)=	0.5415E+00
A(1,.)=	0.83E+00	-0.18E+00	0.34E+00	-0.59E-01	0.59E-01	-0.53E-01	-0.53E-01	-0.32E-01	-0.89E-02	0.17E-01	
	-0.98E-02	0.41E-01	0.22E-01								
ALPHA(L)=	0.822E+00	BETA(L)=-	0.155E+00	GM(1,1)=	0.73E+02	GM(1,2)=	0.54E+02	MDYE(L)=-	0.919E-02	SA42(L)=	0.1421E+00
A(1,.)=	0.82E+00	-0.15E+00	0.32E+00	-0.73E-01	0.74E-01	-0.52E-01	-0.53E-01	-0.35E-01	-0.47E-02	0.23E-01	
	-0.20E-01	0.44E-01	0.19E-01								
ALPHA(L)=	0.837E+00	BETA(L)=-	0.176E+00	GM(1,1)=	0.73E+02	GM(1,2)=	0.55E+02	MDYE(L)=-	0.375E-01	SA42(L)=	0.1248E+00
A(1,.)=	0.84E+00	-0.18E+00	0.34E+00	-0.74E-01	0.84E-01	-0.53E-01	-0.49E-01	-0.55E-01	-0.75E-02	0.40E-01	
	-0.17E-01	0.39E-01	0.27E-01								
ALPHA(L)=	0.826E+00	BETA(L)=-	0.177E+00	GM(1,1)=	0.71E+02	GM(1,2)=	0.52E+02	MDYE(L)=-	0.177E-02	SA42(L)=	0.1927E+00
A(1,.)=	0.83E+00	-0.18E+00	0.36E+00	-0.11E+00	0.89E-01	-0.49E-01	-0.73E-01	-0.30E-01	-0.18E-01	0.42E-01	
	-0.18E-01	0.32E-01	0.24E-01								
ALPHA(L)=	0.834E+00	BETA(L)=-	0.171E+00	GM(1,1)=	0.70E+02	GM(1,2)=	0.52E+02	MDYE(L)=-	0.441E-01	SA42(L)=	0.1779E+00
A(1,.)=	0.83E+00	-0.17E+00	0.34E+00	-0.12E+00	0.11E+00	-0.54E-01	-0.59E-01	-0.39E-01	-0.25E-01	0.25E-01	
	0.59E-02	0.27E-01	0.29E-01								

VALEURS DE A POUR LA DERNIERE ITERATION										
0.83E+00	-0.13E+00	0.31E+00	-0.51E-01	0.19E-01	0.11E-01	-0.47E-01	-0.23E-01	-0.40E-01	0.20E-01	0.00
-0.71E-02	0.67E-01	-0.42E-02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 1 : Mer 4 - 5 p = 13 K = 160 ρ = 0.5

Les enregistrements ont eu lieu par petit fond de 16 mètres, à 9 mètres du fond. Une première bande comporte un enregistrement de bruit marin pour une mer de force 4 - 5 et une seconde bande provient d'un enregistrement par mer de force 2 avec émission à 100 mètres d'une fréquence pure de 1.500 Hz.

L'analyse statistique de ces deux types de données Y a été réalisée à l'Université de Rouen au moyen de la liaison informatique avec le C.I.R.C.E..

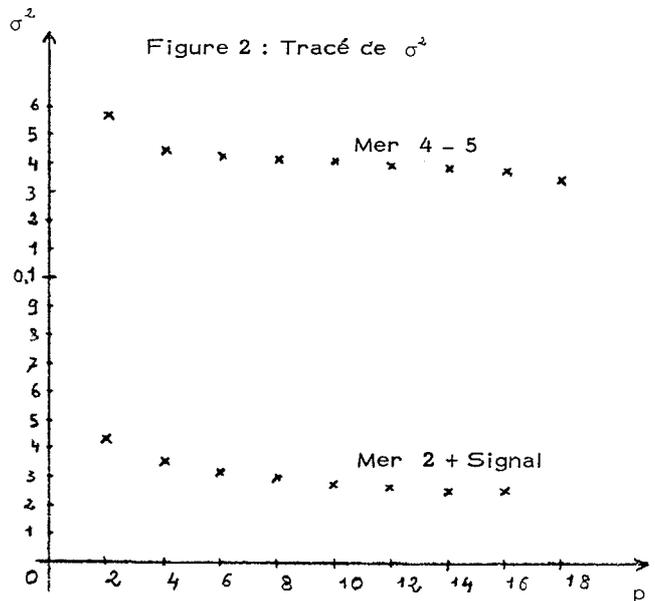
(6) Valeurs des A.

Le procédé itératif pour le calcul des A conduit rapidement à des valeurs stables. (voir Figure 1). Les valeurs initiales $A_1 = A_2 = \dots = A_p = 0$ sont prises nulles et dès le premier passage l'ordre de grandeur des valeurs limites est atteint. Ainsi 160 échantillons permettent d'approcher le spectre et 1600 valeurs assurent une grande stabilité du résultat.

La situation est la même pour $K = 80$ soit $K \Delta t = 10^{-2}$ secondes pour les deux enregistrements traités.

$$\text{L'estimation } \sigma^2 = \frac{\Delta}{N} \sum_1^N E(i)^2 \text{ de}$$

l'erreur moyenne entre $Y(i)$ et son estimation $\hat{Y}(i) = A_1 Y(i-1) + \dots + A_p Y(i-p)$ décroît avec la valeur de p - voir figure 2.



(7) Etude des stationnarités.

Pour les valeurs $p = 13$, $K = 160$, $\rho = 0.5$ et des données normalisées $Y(i) | i = 1, 9600$, on étudie l'évolution de $\gamma^2(L)$ figuré par le trait en fonction de $L = 9, 60$. Sur la même figure 3, on trace $\gamma^2(L/LMAX)$ figuré par p : les paramètres A sont dans ce cas, ceux de la dernière itération $L = LMAX$. En cas d'égalité numérique des valeurs, on ne dessine que le trait.

On conclut que la stationnarité des A, donc du spectre, est globalement respectée sur la seconde : 1 sec. = $50 \times 160 \times 1.25 \times 10^{-4}$ sec..



MODELISATION D'UN ENREGISTREMENT
DE BRUIT MARIN ET DETECTION

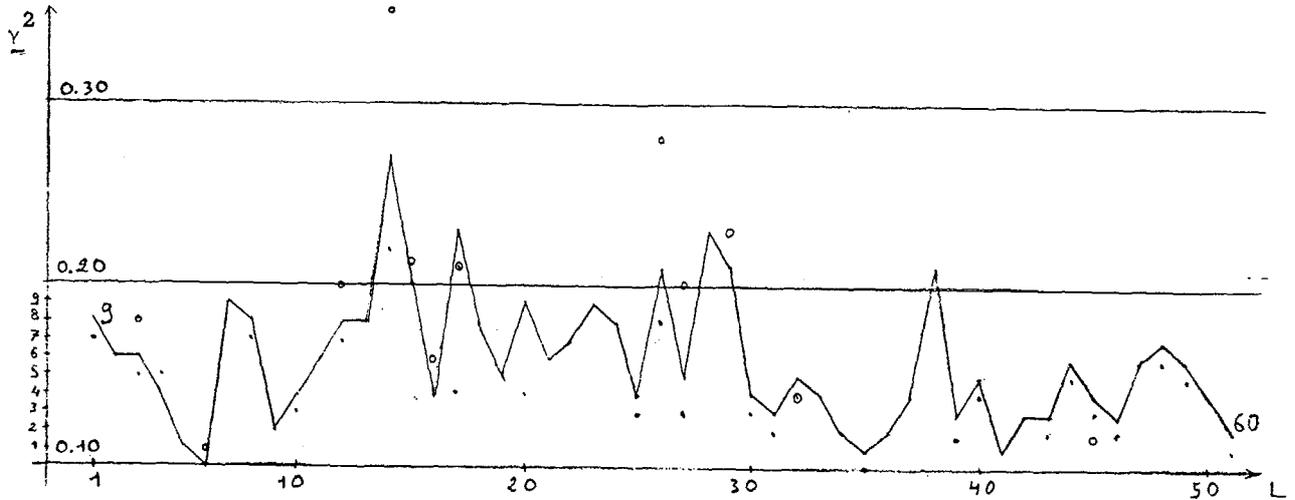


Figure 3 : Tracé de $\gamma^2(L)$ en fonction de L - Mer 4-5 - K = 160 - $\rho = 0.5$ - γ normalisé.
o Tracé de $\gamma^2(L/LMAX)$. $\rho = 0.01$

Mais le modèle, de plus mauvais :

$\gamma^2(L/LMAX) \geq \gamma^2(L)$ pour L petit, devient meilleur :
 $\gamma^2(L/LMAX) \leq \gamma^2(L)$ lorsque L se rapproche de LMAX.

Ce résultat est confirmé en prenant un grand amortissement $\rho = 0.01$ puisque, dans ce cas, l'erreur $\gamma^2(L)$ figuré par . est systématiquement plus faible.

Ainsi le modèle est plus adéquat en donnant au passé lointain une influence faible.

Sur la figure 3, on observe de plus, de grandes fluctuations des $\gamma^2(L)$ en fonction de L . Ces fluctuations proviennent-elles de l'imprécision statistique ?

Supposons que $E(t)$, $t = 1, 9600$ soit une suite de variables aléatoires gaussiennes centrées de variance σ^2 et indépendantes deux à deux. Alors $K \gamma^2(L) = \sum_{L=K}^{L+K} E(t)^2$ suivrait une loi du χ^2 à K degrés

de liberté et grâce au théorème central limite, $\gamma^2(L)$ est approximativement gaussienne, de moyenne σ^2 et de variance $\frac{2}{K} \sigma^4$.

La méthode des intervalles de confiance permet de tester la stationnarité éventuelle de l'énergie.

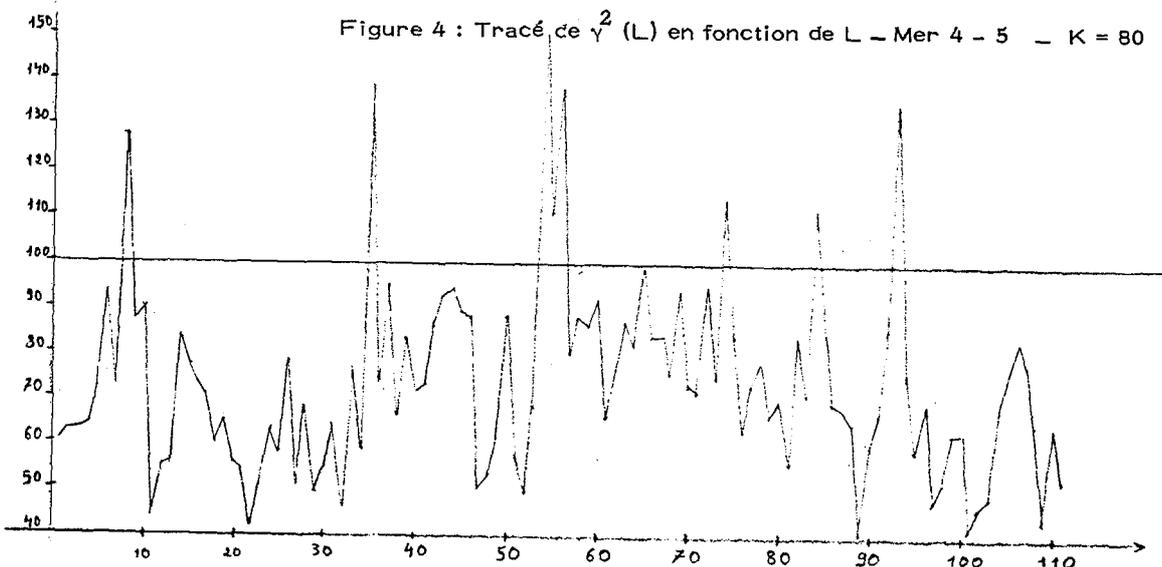
$$\text{Pour } \alpha = 1.96 \text{ P} \left(\left| \frac{\gamma^2 - \sigma^2}{\left(\frac{2}{K} \sigma^4 \right)^{1/2}} \right| > \alpha \right) = 0,05$$

d'où $0,82 \gamma^2 \leq \sigma^2 \leq 1,28 \gamma^2$ avec une probabilité de 95 % .

Les fluctuations de $\gamma^2(L)$ dépassent les limites acceptables et il n'y a pas stabilité énergétique sur 1 seconde, et même sur des périodes beaucoup plus courtes - 10^{-2} secondes.

Pour $K = 80$, on voit mieux apparaître une continuité des valeurs $\gamma^2(L)$ en fonction de $L = 4, 120$ mais il reste encore des fluctuations brutales: Figure 4.

Figure 4 : Tracé de $\gamma^2(L)$ en fonction de L - Mer 4 - 5 - K = 80 - $\rho = 0.01$





MODELISATION D'UN ENREGISTREMENT
DE BRUIT MARIN ET DETECTION

Mer 4 - 5																					
$p = 10$	100	-1	-1	0	0	1	1	2	0	-1	-5	-3	-2	-1	2	4	4	4	3	5	1
12	100	-1	-1	0	0	1	1	2	1	1	-1	-1	-4	-2	2	3	2	3	1	3	0
14	100	-1	0	0	1	1	1	2	1	1	0	0	-1	-1	0	2	2	2	1	3	-1
16	100	-1	-1	0	0	1	1	1	1	1	1	0	0	-2	-1	0	3	3	0	2	-1
Mer 2 + Signal																					
$p = 10$	100	-10	1	7	1	-7	-9	2	10	9	-1	2	3	-6	-7	0	4	8	6	7	7
12	100	-11	-1	5	0	-5	-7	1	10	7	0	-10	4	-3	-9	-1	3	7	5	7	6
14	100	-11	-1	4	1	-5	-6	1	8	7	-3	-7	-4	5	-7	-4	4	7	6	8	6
16	100	-11	-1	4	1	-3	-4	2	6	5	-3	-8	-2	1	4	-6	1	10	7	7	6

Figure 5 : Fonction de Corrélation de l'erreur $E(I)$ normalisée à 100 - $K = 80$ $\rho = 0.01$

(8) Blanchiment des données.

La corrélation $\Gamma(J)$ du résidu $E(I)$ est faible - figure 5. On préfère ce critère de blanchiment au tracé du spectre. Le spectre $f(\nu)$ associé à une covariance $\Gamma(J)$ s'écrit

$f(\nu) = 1 + 2 \sum \Gamma(J) \cos(2\pi \nu J \Delta t)$, estimé par l'expression numérique :

$$f(\nu) = \left| \frac{1}{N} \sum_I \text{Pond}(I) \times E(I) \times \exp(2i\pi \nu J \Delta t) \right|^2$$

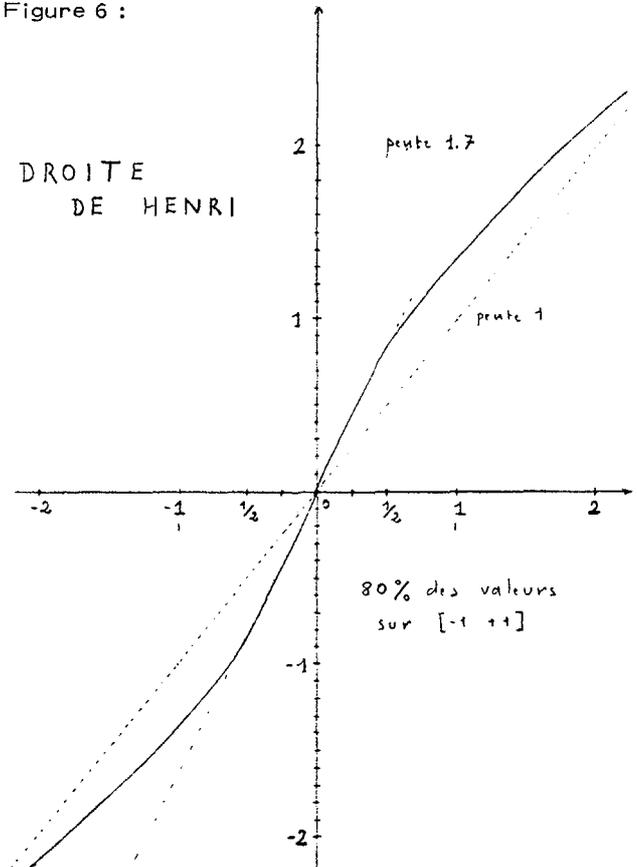
où Pond est une fonction de pondération destinée à diminuer les effets de bord.

(9) Test de Caractère Gaussien.

On utilise le procédé de la droite de Henri: On note F^* , la fonction de répartition expérimentale et Φ la fonction de répartition théorique hypothétique. Si la situation expérimentale caractérisée par F^* provient de la loi théorique Φ alors $x \rightsquigarrow \Phi^{-1}(F^*(x))$ vaut x aux erreurs statistiques près. Le test, Φ loi gaussienne, conduit à un échec (figure 6). ! L'erreur $E(I)I = 1,9600$ ne suit pas une loi gaussienne.

On aboutit à la même conclusion pour $\rho = 0.5$ $\rho = 0.01$ $p = 1,16$ $K = 80,160$ Mer 4-5, ou 2 avec signal.

Figure 6 :



(10) Modèle A.R ou Modèle A.R.M.A..

Les valeurs de A vérifiant l'équation vectorielle $A \mathcal{H} = S$ dépendent du paramètre q ! Le modèle A.R. est insuffisant pour expliquer le bruit marin.

Sous toute réserve, pour une Mer 4-5, le modèle A.R.M.A. pourrait être $p = 13$, $q = 5$ et pour une mer 2 avec signal sinusoïdal, le modèle A.R.M.A. pourrait être $p = 12$ $q = 2$.

La valeur de p choisie provient de l'étude de la variance estimée $\frac{1}{N} \sum E(I)^2$ qui décroît en fonction de p et la valeur de q choisie, de la constance des $A^{p,q}$ pour q passant de q à $q+1$.

MODELISATION D'UN ENREGISTREMENT
DE BRUIT MARIN ET DETECTION

(11) Rôle de γ^2 pour la détection.

On part des LMAX échantillons $Y(i), i = 1, \dots, K$, LKK de taille K, supposés gaussiens centrés et de matrice de covariance Γ_0 sous H_0 et Γ_1 sous H_1 . Les coefficients A déterminent le spectre $\rho(\omega)$ puis la matrice de covariance Γ .

Pour décrire la détection, on introduit K variables aléatoires $Y \triangleq (Y_1, \dots, Y_K)$ et les échantillons sont considérés comme des réalisations expérimentales de la variable vectorielle Y .

Le blanchiment parfait consiste à introduire une matrice B , carrée $K \times K$, inversible associée à l'orthonormalisation de Gram-Schmidt

$$Y_1 = B_{1,1} e_1$$

$$Y_2 = B_{1,2} e_1 + B_{2,2} e_2$$

$$Y_K = B_{1,K} e_1 + \dots + B_{K,K} e_K$$

d'où $(Y_1, Y_2, \dots, Y_K) = (e_1, e_2, \dots, e_K) B$

et $\Gamma_0 = E_0(Y^t Y) = B^t B$.

Par suite $E_1(e^t e) = B^{-1t} \Gamma_1 B^{-1} \triangleq \Delta$

Soit $\Delta = \sum_{i=1}^K \alpha_i (e_i > \varphi_i)$, la décomposition canonique de la matrice Δ symétrique de type positif à l'aide des valeurs propres α_i et des vecteurs propres $\varphi_i \quad i = 1, K$.

On a effectué un nouveau changement de base caractérisée par une matrice unitaire soit

$(f_1, f_2, \dots, f_K) \triangleq (e_1, e_2, \dots, e_K) U$ d'où

$$E_0(f^t f) = U^t E_0(e^t e) U = I \text{ identité}$$

$$E_1(f^t f) = U^t E_1(e^t e) U = U^t \Delta U$$

$$= \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & & \\ \dots & & \dots & \\ 0 & 0 & & \alpha_K \end{bmatrix}$$

Les variables aléatoires f_1, \dots, f_K sont gaussiennes centrées indépendantes de variance $1, \dots, 1$ sous H_0 et de variance $\alpha_1, \dots, \alpha_K$ sous H_1 .

La densité de Radon - Nikodym de la variable K dimensionnelle, s'écrit aisément et le résumé exhaustif vaut :

$$RE(f_1, \dots, f_K) = \text{Log} \prod_i \frac{1}{\sqrt{2\pi} \alpha_i} \exp\left(-\frac{f_i^2}{2 \alpha_i}\right)$$

$$- \text{Log} \prod_i \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f_i^2}{2}\right) =$$

$$- \sum_i \text{Log} \alpha_i - \frac{1}{2} \sum_i f_i^2 \left(\frac{1}{\alpha_i} - 1\right) =$$

$$- \text{Log} (\det \Delta) + \frac{1}{2} \sum_i f_i^2 \left(\frac{\alpha_i^{-1} - 1}{\alpha_i}\right)$$

La détection optimale consiste, partant de $Y(i) \quad i = 1, \dots, K$, à calculer $(Y(LK), \dots, Y(LKK)) B^{-1} U = (f_1, \dots, f_K)$ puis le RE (f_1, \dots, f_K) qui est quadratique des observations. Quelles sont les propriétés du Résumé Exhaustif ?

$$E_0(RE) = -\text{Log}(\det \Delta) + 1/2 \text{Trace} \left(\frac{\Delta^{-1} - I}{\Delta} \right)$$

$$E_1(RE) = -\text{Log}(\det \Delta) + 1/2 \text{Trace} (\Delta^{-1} - I).$$

Dans le cas particulier où $\Gamma_1 = \lambda \Gamma_0$ alors $\Delta = \lambda I$ et toute base $\varphi_i \quad i = 1, K$ est propre d'où $\alpha_i = \lambda \quad i = 1, K$. Il vient :

$$RE(e_1, \dots, e_K) = -K \text{Log} \lambda + \frac{1}{2} \frac{\lambda^{-1} - 1}{\lambda} \sum_{LK} E(i)^2$$

Ainsi $\gamma^2 = \frac{1}{K} \sum_{LK} E(i)^2$ est un résumé exhaustif pour

la détection entre deux hypothèses H_0, H_1 associées aux mêmes coefficients A_1, \dots, A_p . Il est sous-optimal si l'on veut faire une détection entre deux spectres différents.

BIBLIOGRAPHIE

Box and Jenkins : Time Series Analysis Holden Day

P. Faure : Réalisations markoviennes de processus stationnaires Rapport IRIA n°13 Mars 1973.

C.A. Darmon : Deux Méthodes d'estimation, identification de systèmes linéaires bruités multidimensionnels. Revue du Cethedec N° 49 1976 p. 1 - 31.

G. Favier : Identification d'une représentation gaussienne markovienne minimale à l'aide d'algorithmes de réalisation stochastique. Revue du Cethedec N° 51 1977 p 13 - 42

I.E.E.E. Automatic Control décembre 1974 volume AC 19 n° 6.