

REVUE DE STATISTIQUE APPLIQUÉE

G. MÉLARD

R. ROY

Modèles de séries chronologiques avec seuils

Revue de statistique appliquée, tome 36, n° 4 (1988), p. 5-23.

http://www.numdam.org/item?id=RSA_1988__36_4_5_0

© Société française de statistique, 1988, 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/legal.php>). 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/>

MODÈLES DE SÉRIES CHRONOLOGIQUES AVEC SEUILS

G. MÉLARD(1) et R. ROY(2)

(1) Institut de Statistique – CP 210, Campus Plaine ULB,
Bd du Triomphe, B-1050 Bruxelles, Belgique.

(2) Département d'informatique et de recherche opérationnelle, Université de Montréal,
Case postale 6128, Succursale "A", Montréal P.Q., H3C 3J7, Canada.

RÉSUMÉ

Nous considérons les modèles ARMA avec seuils. Ils généralisent les modèles autorégressifs avec seuils qui ont fait l'objet d'une étude approfondie dans la littérature et qui sont souvent plus adéquats que des modèles linéaires. Toutefois la paramétrisation du terme indépendant est différente ici et nous justifions notre approche. Nous décrivons une méthode d'estimation des paramètres des modèles ARMA avec seuils. Après avoir illustré la méthode sur des données simulées, nous reprenons le célèbre exemple du nombre annuel de lynx canadiens pris au piège et nous envisageons l'application de la méthode à des données de pollution atmosphérique. La supériorité des modèles ARMA avec seuils ressort de l'étude.

ABSTRACT

Threshold ARMA models are considered, generalizing threshold autoregressive models which have been deeply studied in the literature and appear often to be more satisfactory than linear models. Parametrization of the independent term is however different here and justifications are given. A method for estimating the parameters of threshold ARMA models is described. The method is illustrated with artificial data. The well known series of the annual Canadian lynx trappings is analyzed again and atmospheric pollution data are also considered. It appears that threshold ARMA models perform better.

1. Introduction

Une grande partie de la littérature consacrée à l'analyse de séries chronologiques est focalisée sur les modèles linéaires et en particulier sur les modèles ARMA. Dans un modèle ARMA, la valeur en t de la variable étudiée, x_t , est exprimée comme combinaison linéaire de l'innovation en t , a_t , et de valeurs passées de la variable et de l'innovation :

$$x_t = \sum_{j=1}^p \phi(j)x_{t-j} + a_t - \sum_{k=1}^q \theta(k)a_{t-k}.$$

Les innovations a_t sont des variables aléatoires normales indépendantes, de moyennes nulles et de variances égales à σ^2 . Les coefficients $\phi(j)$ et $\theta(k)$ ainsi que

σ^2 sont invariants dans le temps.

Récemment, plusieurs auteurs ont critiqués les modèles linéaires qui donnent souvent une représentation trop schématique de la réalité. Ceci est particulièrement frappant dans les domaines de la physique, de la chimie ou de la biologie où des modèles théoriques existent, surtout en temps continu. Des modèles de séries chronologiques apparaissent comme approximation, afin de fournir une formulation statistique à partir de données, généralement discrètes dans le temps. Les modèles ARMA alors utilisés n'ont pas les mêmes caractéristiques que les modèles théoriques non linéaires qu'ils sont sensés approcher. En particulier, Tong [20] a étudié plusieurs problèmes de cette nature. Il a observé que des phénomènes comme l'existence d'un cycle limite, la non-réversibilité du temps et la dépendance entre l'amplitude et la fréquence ne peuvent pas être représentés par des modèles linéaires. Il a proposé une classe de modèles autorégressifs avec seuils qui permettent d'incorporer les phénomènes de non-linéarité. Considérons à titre d'exemple le modèle autorégressif (AR) d'ordre 1 défini par la relation :

$$x_t = \begin{cases} \phi'_1 x_{t-1} + a_t, & x_{t-1} < \alpha, \\ \phi''_1 x_{t-1} + a_t, & x_{t-1} \geq \alpha. \end{cases}$$

Tong [20] a étudié une classe très large de modèles AR avec seuils et a défini (Tong [20], p. 61) les modèles ARMA mais il n'a pas poursuivi leur étude. D'autres classes de modèles non linéaires sont les modèles bilinéaires (Granger et Andersen [7], Robinson [15]), les modèles autorégressifs exponentiels (Haggan et Ozaki [9]), et, plus généraux car contenant tous les autres comme cas particuliers, les modèles dépendant de l'état (Priestley [14]). Un exemple de modèle dépendant de l'état est donné par

$$x_t = \phi_1(X_{t-1})x_{t-1} + a_t - \theta_1(X_{t-1})a_{t-1}$$

où les coefficients $\phi_1(\cdot)$ et $\theta_1(\cdot)$ dépendent du vecteur d'état

$$X_{t-1} = (x_{t-1}, x_{t-2}, a_{t-1}, a_{t-2})'.$$

Dans cet article, nous proposons une nouvelle définition des modèles ARMA avec seuils qui donne lieu à une interprétation différente du terme indépendant. La différence entre notre définition et celle de Tong [20] est illustrée à l'aide de données simulées. Nous décrivons une méthode d'estimation des paramètres des modèles ARMA avec seuils. Après avoir illustré la méthode avec des données simulées, nous reprenons l'exemple du nombre annuel de lynx canadiens pris au piège et nous appliquons aussi la méthode à des données de pollution atmosphérique. La supériorité des modèles ARMA avec seuils ressort de l'étude.

2. Les modèles autorégressifs avec seuils

Les modèles autorégressifs non linéaires avec seuils ont été introduits par Tong [19] et ont fait l'objet d'études approfondies de Tong et Lim [21], Tong [20], Chan et Tong [5]. Ayant deux séries chronologiques

$$x = (x_t ; t = 1, \dots, N) \quad \text{et} \quad y = (y_t ; t = 1, \dots, N),$$

on dit que x suit un modèle TAR($\ell; p_1, \dots, p_\ell$) (“Threshold autoregressive model”) si x_t est régi par un modèle autorégressif d’ordre p_i et de coefficients

$$\underline{\phi}_i = (\phi_i(1), \dots, \phi_i(p_i))' \quad \text{et} \quad \theta_{0i},$$

c’est-à-dire

$$x_t = \theta_{0i} + \sum_{j=1}^{p_i} \phi_i(j)x_{t-j} + a_t, \quad (2.1)$$

pour tout t tel que $\alpha_{i-1} \leq y_t < \alpha_i$ ($i = 1, \dots, \ell$). Les α_i ($i = 0, 1, \dots, \ell$) sont appelés les seuils et sont tels que $\alpha_0 = -\infty < \alpha_1 < \dots < \alpha_{\ell-1} < \alpha_\ell = \infty$. Dans (2.1), les a_t sont des variables aléatoires normales indépendantes, de moyennes nulles et de variances σ_i^2 . Si $y_t = x_{t-b}$ pour $b > 0$, on dit que le modèle est de type SETAR (“Self-exciting Threshold Autoregressive Model”). Une condition nécessaire et suffisante d’ergodicité du processus existe si $\ell = 2$ et $p_1 = p_2 = 1$ (Petrucci et Woolford [13]). La distribution stationnaire de ce processus particulier est étudiée par Andel *et al.* [2]. Pour ℓ et $\underline{\alpha}_\ell = (\alpha_1, \dots, \alpha_{\ell-1})'$ fixés, l’estimation des paramètres $\underline{\phi} = (\underline{\phi}'_1, \dots, \underline{\phi}'_\ell)'$ et $\underline{\theta}_0 = (\theta_{01}, \dots, \theta_{0\ell})'$ est réalisée par la méthode du minimum d’entropie d’Akaike [1] qui consiste à minimiser par rapport à $\underline{p} = (p_1, \dots, p_\ell)'$

$$AIC(\underline{p}; \ell, \underline{\alpha}) = \sum_{i=1}^{\ell} N_i \ln \hat{\sigma}_i^2(\hat{\theta}_{0i}, \hat{\phi}_i; p_i) + 2 \sum_{i=1}^{\ell} (p_i + 1), \quad (2.2)$$

où N_i est le nombre de valeurs de t telles que $\alpha_{i-1} \leq y_t < \alpha_i$ et $\hat{\sigma}_i^2(\hat{\theta}_{0i}, \hat{\phi}_i; p_i)$ est l’estimateur du maximum de vraisemblance de σ_i^2 obtenu en estimant les paramètres θ_{0i} et ϕ_i du modèle autorégressif d’ordre p_i défini par (2.1). On suppose ici que les x_t , $t < 1$, sont connus et fixés. La méthode utilisée par Tong fait intervenir une transformation de Householder pour les N_i observations relatives au $i^{\text{ième}}$ régime. La détermination de $\hat{\sigma}_i^2$, $\hat{\theta}_{0i}$ et $\hat{\phi}_i$ est donc effectuée de façon indépendante pour chacun des régimes. Le choix de ℓ , des seuils $\underline{\alpha}$ ainsi que de b dans le modèle SETAR est réalisé de façon heuristique : on fixe un maximum L pour ℓ , et pour chaque valeur de ℓ comprise entre 1 et L on choisit un ou plusieurs ensembles de $(\ell - 1)^{\text{uples}}$ $\underline{\alpha}$ pour lesquels on procède comme il a été indiqué ci-dessus.

Cette méthode d’estimation des paramètres de modèles avec seuils présente trois légers inconvénients auxquels nous avons cherché à remédier :

1. la méthode d’estimation des paramètres est limitée aux modèles autorégressifs et ne se généralise pas aux modèles ARMA : sans négliger l’intérêt des interprétations que permettent les modèles autorégressifs par la mise en évidence de cycles, des considérations de parcimonie conduisent parfois à des modèles moyenne mobile ou à des modèles mixtes autorégressif-moyenne mobile;
2. la méthode de choix des seuils procède par tâtonnement ou par balayage : il serait désirable d’estimer les seuils au moyen d’une méthode statistique;
3. les paramètres des ℓ modèles de type (2.1) doivent être indépendants les uns des autres ce qui peut conduire à un nombre excessif de paramètres : on

pourrait souhaiter que certains paramètres soient identiques pour les modèles de plusieurs régimes, par exemple que les variance σ_i^2 soient égales.

Dans un article récent, Chan et Tong [5] développent une méthode d'estimation du seuil d'un modèle SETAR. Dans ce travail, nous discutons l'estimation du seuil pour la classe plus générale des modèles ARMA.

3. Les modèles ARMA avec seuils

Le modèle qui sera traité dans cet article est le suivant :

$$x_t = \mu_i + \sum_{j=1}^p \phi_i(j)(x_{t-j} - \mu_{I(t-j)}) + a_t - \sum_{k=1}^q \theta_i(k)a_{t-k}, \quad (3.1)$$

si $\alpha_{i-1} \leq y_t < \alpha_i \quad (i = 1, \dots, \ell),$

où la fonction $I(t)$ est définie par

$$I(t) = i \quad \text{ssi} \quad \alpha_{i-1} \leq y_t < \alpha_i.$$

Il est essentiel que I_t soit fonctionnellement indépendant de x_t, x_{t+1}, \dots . Le bruit a_t est défini de la façon suivante. Soit $(\tilde{a}_t; t \in \mathbb{Z})$ un processus bruit blanc gaussien normé et soit $\sigma_i > 0 \quad (i = 1, \dots, \ell)$, le bruit a_t est défini par

$$a_t = \sigma_i \tilde{a}_t \quad \text{si} \quad \alpha_{i-1} \leq y_t < \alpha_i \quad (i = 1, \dots, \ell)$$

ou, plus simplement, $a_t = \sigma_{I(t)} \tilde{a}_t$. Posant $\tilde{x}_t = x_t - \mu_{I(t)}$, le modèle défini par l'équation

$$\tilde{x}_t = \sum_{j=1}^p \phi_i(j) \tilde{x}_{t-j} + a_t - \sum_{k=1}^q \theta_i(k) a_{t-k} \quad (3.2)$$

décrit le $i^{\text{ième}}$ régime du modèle avec seuils et est noté en abrégé R_i . On voit que \tilde{x}_t est obtenu au temps t par le modèle $R_{I(t)}$. Le modèle (3.1) ne sera considéré que pour $t \geq 1$. On supposera pour la commodité que les x_t pour $t \leq 0$ sont obtenus à partir de R_1 , mettons. À cette fin, on supposera que R_1 est un modèle stationnaire et inversible de façon à décrire la distribution liée des premiers x_t et à interpréter les premiers a_t comme les innovations du processus.

Pour la commodité, on a supposé que les ordres des modèles ARMA sont toujours p et q . Ceci n'est pas une restriction puisque, dans la méthode proposée, les paramètres $\phi_i(j)$ et $\theta_i(k)$ peuvent être fixés, par exemple à la valeur zéro, ou être identiques entre deux modèles. On suppose toutefois qu'il n'y a pas de relation entre les μ_i d'une part et les autres paramètres d'autre part, ni entre les σ_i d'une part et les autres paramètres d'autre part.

Une forme ressemblant à (3.1) a été envisagée par Tong [20] :

$$x_t = \theta_0^{(i)} + \sum_{j=1}^p \phi_i(j)x_{t-j} + a_t - \sum_{k=1}^q \theta_i(k)a_{t-k}. \quad (3.3)$$

D'une part Tong a limité son analyse aux modèles autorégressifs avec seuils c'est-à-dire le cas où $q = 0$, d'autre part on constate une différence dans la paramétrisation de la constante du modèle. Elle dépend seulement du régime où se trouve x en t dans le cas de (3.3) tandis qu'elle dépend des régimes où se trouve x en $t, t - 1, \dots, t - p$ dans le cas de (3.1) par l'intermédiaire des $\mu_{I(t-j)}$. Cette différence est plus importante qu'il peut paraître. Pour la mettre en évidence, émettons temporairement l'hypothèse suivante :

Hypothèse (I) : les a_s ($s \leq t$) sont conjointement distribués indépendamment de (y_1, y_2, \dots, y_t) . Cela signifie en particulier que le choix des modèles des différents régimes ne dépendent pas des a_s , excluant par exemple le cas des modèles de type SETAR où le sous-modèle en t est déterminé par x_{t-b} et par conséquent par certains des a_s pour $s \leq t - b$.

Notons E_t l'opérateur d'espérance conditionnelle étant donné les sous-modèles utilisés aux temps $s \leq t$, c'est-à-dire étant donné $y^t = (y_t, y_{t-1}, \dots, y_1)$. Conditionnellement à y^t , les coefficients $\phi_i(j)$ et $\theta_i(k)$ sont des fonctions du temps $\phi_{I(t)}(j)$ et $\theta_{I(t)}(k)$ et (3.1) apparaît comme un modèle ARMA à coefficients dépendant du temps, aussi appelé modèle ARMA évolutif, qui a été étudié par Mélard [11]. En particulier, on peut écrire la décomposition de Wold-Cramér du processus (Cramér [6]) ou représentation moyenne mobile infinie. À cette fin, remplaçons dans (3.1) $x_{t-j} - \mu_{I(t-j)}$ par une expression similaire à (3.1) écrite au temps $t - j$ au lieu de t . En répétant indéfiniment cette substitution, on peut écrire (3.1) sous la forme

$$x_t - \mu_{I(t)} = \sum_{s=-\infty}^t \psi_{t,s} a_s. \quad (3.4)$$

Les coefficients $\psi_{t,s}$ dépendent des $\phi_{I(u)}(j)$ et des $\theta_{I(u)}(k)$ pour $u \leq t$. Ils peuvent être obtenus par application des relations de récurrence de Mélard ([11] p. 37). Par exemple, dans le cas où $p = 1$ et $q = 0$, il vient

$$\psi_{t,s} = \prod_{i=s+1}^t \phi_{I(i)}(1), \quad s \leq t - 1$$

et $\psi_{t,t} = 1$. Appliquons E_t aux deux membres de (3.4). Grâce à l'hypothèse (I), on a

$$E_t \{ \psi_{t,s} a_s \} = \psi_{t,s} E \{ a_s | y^t \} = \psi_{t,s} E \{ a_s \} = 0,$$

$1 \leq s \leq t$, et de même $E_t \{ \psi_{t,s} a_s \} = 0$ pour $s \leq 0$. Par conséquent

$$E_t \{ x_t - \mu_{I(t)} \} = 0, \quad t \geq 1.$$

Il apparaît donc que μ_i est, dans un certain sens, la moyenne des x_t qui se trouvent dans le $i^{\text{ème}}$ régime. Le modèle du régime i est donc caractérisé par un niveau moyen propre μ_i et par la fonction d'autocovariance associée aux coefficients, en supposant la stationnarité. Il est aisé de voir que $\theta_0^{(i)}$ ne jouit pas d'une interprétation similaire en tant que niveau moyen.

Pour montrer d'une façon empirique la différence entre (3.1) et (3.3), nous avons généré artificiellement une série de longueur 200 par chacun des deux

processus en utilisant des modèles AR(1) et MA(1). Le processus (3.1) est défini par

$$\begin{cases} \tilde{x}_t = 0,8\tilde{x}_{t-1} + a_t, & \text{si } y_t < 0, \\ \tilde{x}_t = a_t - 0,5a_{t-1}, & \text{si } y_t \geq 0, \end{cases} \quad (3.5)$$

où $\tilde{x}_t = x_t - \mu_{I(t)}$ tandis que le processus (3.3) est défini par

$$\begin{cases} x_t = \theta_0^{(1)} + 0,8x_{t-1} + a_t, & \text{si } y_t < 0, \\ x_t = \theta_0^{(2)} + a_t - 0,5a_{t-1}, & \text{si } y_t \geq 0, \end{cases} \quad (3.6)$$

pour $t \leq 2$. On a choisi $\sigma_1^2 = \sigma_2^2 = 1$. Dans chaque cas, le choix entre les deux modèles est basé sur le processus $\{y_t ; t \geq 1\}$ de type bruit blanc distribué uniformément sur $[-0,5 ; 0,5]$. Comme seuil, on a pris la valeur zéro et ainsi à chaque instant les deux modèles sont équiprobables. Les valeurs initiales sont $\tilde{x}_1 = x_1 = a_1$. Le tableau I montre les valeurs utilisées de a_t et y_t pour $t = 1, 2, \dots, 10$. Pour (3.5) on a employé $\mu_1 = 1$ et $\mu_2 = 5$ tandis que pour (3.6) on a choisi par tâtonnement $\theta_0^{(1)} = -1$ et $\theta_0^{(2)} = 5$ de façon à obtenir une allure similaire pour les deux séries. Le tableau I contient les premières valeurs des deux séries artificielles. Les figures 1(b) et 1(c) décrivent les deux séries. À l'œil, on distingue les graphiques des deux séries et cette différence est confirmée par les autocorrélations échantillonales $\hat{\rho}_k, k \geq 1$. En effet, on obtient

$$\begin{aligned} (3.5) : \quad & \hat{\rho}_1 = 0,06, \quad \hat{\rho}_2 = 0,04, \quad \hat{\rho}_3 = -0,02, \quad \dots \\ (3.6) : \quad & \hat{\rho}_1 = 0,36, \quad \hat{\rho}_2 = 0,18, \quad \hat{\rho}_3 = -0,02, \quad \dots \end{aligned}$$

Tableau I

Valeurs de a_t et y_t et des données générées selon les modèles (3.1) et (3.3) particularisés par (3.5) et (3.6) respectivement ($t = 1, \dots, 10$)

t	y_t	régime	a_t	x_t (3.5)	x_t (3.6)
1	-0,3137	1	0,0869
2	-0,2582	1	-0,7631	0,3064	-1,6936
3	0,0815	2	0,1809	5,5624	5,5624
4	-0,1883	1	-0,1120	1,3379	3,3379
5	-0,2338	1	-1,1685	0,1018	0,5018
6	-0,1279	1	-0,3256	-0,0441	-0,9241
7	-0,4532	1	0,0930	0,2577	-1,6463
8	0,3117	2	1,6009	6,5544	6,5544
9	-0,1919	1	-0,6344	1,6091	3,6091
10	0,4307	2	0,2902	5,6074	5,6074

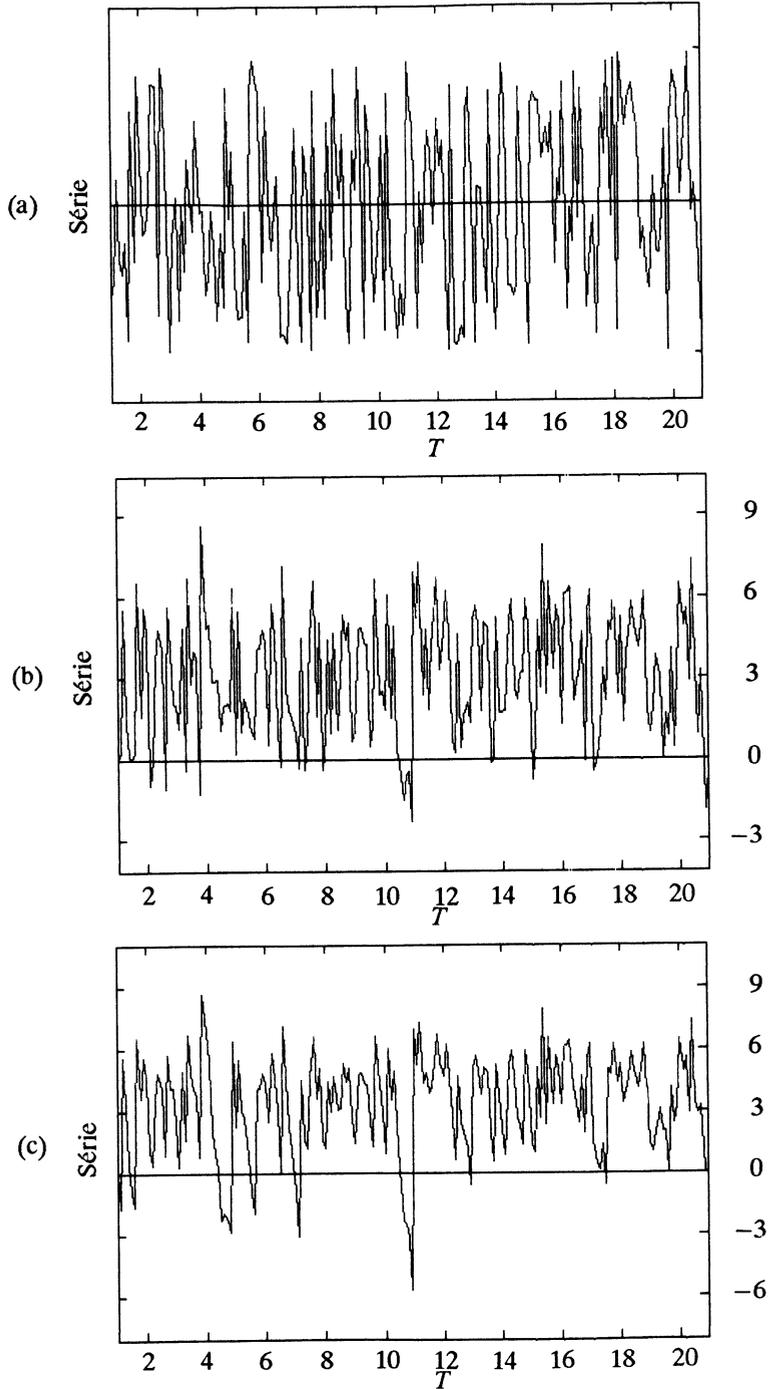


Figure 1 – Série des y_t (a) et des x_t générées selon les modèles (3.1) (b) et (3.3) (c) particularisés par (3.5) et (3.6) respectivement ($t = 1, \dots, 200$).

On constate donc que la série engendrée par (3.5) a des propriétés similaires à celles d'un bruit blanc. En effet, on a $E(\tilde{x}_t) = 0$ pour tout $t \geq 1$ et $E(x_t) = \frac{1}{2}(1+5) = 3$.

De plus,

$$\begin{aligned} E(\tilde{x}_t^2) &= P(y_t < 0)E(0.8\tilde{x}_{t-1} + a_t)^2 + P(y_t \geq 0)E(a_t - 0.5a_{t-1})^2 \\ &= 0,32E(\tilde{x}_{t-1}^2) + 1,125. \end{aligned}$$

En régime stationnaire (pour t grand), $E(\tilde{x}_t^2) \simeq E(\tilde{x}_{t-1}^2)$, on trouve $E(\tilde{x}_t^2) \simeq 1,125/(1-0,32) = 1,6544$. Puisque, $E(x_t^2) = E(\tilde{x}_t^2) + 13$, on obtient $\text{var}(x_t) \simeq 5,6544$. Similairement, on calcule $E(x_t x_{t-1}) \simeq 9,4118$ ce qui donne comme coefficient d'autocorrélation d'ordre 1 la valeur de 0,0728.

4. Estimation des paramètres

Notons $\underline{\phi}$ la matrice $\ell \times p$ constituée des coefficients autorégressifs $\phi_i(j)$, $\underline{\theta}$ la matrice $\ell \times q$ constituée des coefficients moyenne mobile $\theta_i(k)$, $\underline{\mu} = (\underline{\mu}_1, \dots, \underline{\mu}_\ell)'$, $\underline{\sigma} = (\sigma_1, \dots, \sigma_\ell)'$ et $\underline{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_\ell)'$.

La méthode d'estimation que nous proposons est similaire à celle de Tong pour ce qui concerne ℓ, p et q . Il s'agit donc de minimiser

$$\text{AIC}(\ell, p, q) = -2 \ln L(\hat{\underline{\phi}}, \hat{\underline{\theta}}, \hat{\underline{\mu}}, \hat{\underline{\alpha}}, \hat{\underline{\sigma}}; \ell, p, q) + 2P, \quad (4.1)$$

où P est le nombre total de paramètres contenus dans $\underline{\phi}$, $\underline{\theta}$, $\underline{\mu}$, $\underline{\alpha}$ et $\underline{\sigma}$ et $L(\underline{\phi}, \underline{\theta}, \underline{\mu}, \underline{\alpha}, \underline{\sigma}; \ell, p, q)$ est la fonction de vraisemblance du modèle spécifié par ℓ , p et q , conditionnellement aux x_t , $t < 1$. On note $\hat{\underline{\phi}}, \hat{\underline{\theta}}, \hat{\underline{\mu}}, \hat{\underline{\alpha}}$ et $\hat{\underline{\sigma}}$ les estimateurs du maximum de vraisemblance de $\underline{\phi}$, $\underline{\theta}$, $\underline{\mu}$, $\underline{\alpha}$ et $\underline{\sigma}$, respectivement. Si on suppose que les ℓ paramètres σ_i sont indépendants, L s'écrit sous la forme

$$(2\pi)^{-N/2} \left(\prod_{t=1}^N \sigma_{I(t)}^{-1} \right) \exp \left\{ -\frac{1}{2} \sum_{t=1}^N \sigma_{I(t)}^{-2} a_t^2 \right\}, \quad (4.2)$$

où les a_t s'obtiennent par récurrence à partir de (3.1). Il s'ensuit que $-\ln L$ vaut, à une constante additive près,

$$\sum_{t=1}^N \ln \sigma_{I(t)} + \frac{1}{2} \sum_{t=1}^N \sigma_{I(t)}^{-2} a_t^2. \quad (4.3)$$

L'équation de vraisemblance pour σ_i , où les autres paramètres sont remplacés par leurs estimations, s'écrit donc :

$$\frac{N_i}{\sigma_i} - \frac{1}{\sigma_i^3} \sum_{t=1}^N \delta\{I(t) = i\} \hat{a}_i^2, \quad (4.4)$$

où $\delta\{\cdot\}$ est la fonction indicatrice et $N_i = \sum_{t=1}^N \delta\{I(t) = i\}$ est le nombre d'observations provenant du modèle i . La solution est donc

$$\hat{\sigma}_i^2 = \frac{1}{N_i} \sum_{t=1}^N \delta\{I(t) = i\} \hat{a}_i^2. \quad (4.5)$$

L'expression (4.3) s'écrit donc, à une constante additive près, indépendante des paramètres,

$$\sum_{i=1}^{\ell} N_i \ln \hat{\sigma}_i \quad (4.6)$$

comme dans (2.2). La différence réside dans la détermination des $\hat{a}_i = \hat{a}_i(\underline{\phi}, \underline{\theta}, \underline{\mu}, \underline{\alpha})$.

Supposons d'abord que $\underline{\alpha}$ soit fixé. Le processus x défini par (3.1) apparaît alors comme non stationnaire. On peut donc invoquer une méthode générale pour l'estimation des paramètres de modèles ARMA à coefficients dépendant du temps, qui se simplifie ici parce que la fonction de vraisemblance est conditionnelle par rapport aux x_t et a_t pour $t < 1$. Notons

$$\sigma = \left(\prod_{s=1}^N \sigma_{I(s)} \right)^{1/N} = \left(\prod_{i=1}^{\ell} \sigma_i^{N_i} \right)^{1/N} \quad (4.7)$$

$$g_i = \sigma_i / \sigma \quad (i = 1, \dots, \ell). \quad (4.8)$$

Étant donné (4.7), seuls $(\ell - 1)$ des g_i peuvent être des paramètres indépendants. Puisque σ est la moyenne géométrique des σ_i il résulte de (4.8) que la moyenne géométrique des y_i vaut 1. Par conséquent, en prenant $g_1, g_2, \dots, g_{\ell-1}$ comme paramètres, on a

$$g_{\ell} = \prod_{i=1}^{\ell-1} g_i^{-N_i/N_{\ell}}.$$

La fonction de vraisemblance s'écrit alors sous la forme

$$(2\pi)^{-N/2} \sigma^{-N} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^N g_{I(t)}^{-2} a_t^2(\underline{\phi}, \underline{\theta}, \underline{\mu}) \right\},$$

d'où il ressort que les estimateurs du maximum de vraisemblance peuvent s'obtenir en minimisant la somme des carrés

$$S(\underline{\phi}, \underline{\theta}, \underline{\mu}, \underline{g}) = \sum_{t=1}^N g_{I(t)}^{-2} a_t^2(\underline{\phi}, \underline{\theta}, \underline{\mu}), \quad (4.9)$$

où $\underline{g} = (g_1, \dots, g_{\ell})$, tandis que

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N g_{I(t)}^{-2} \hat{a}_t^2(\hat{\underline{\phi}}, \hat{\underline{\theta}}, \hat{\underline{\mu}}) \quad (4.10)$$

sert à déterminer les $\hat{\sigma}_i^2$ à l'aide de (4.8). À cause de la présence de $\underline{\theta}$, les a_t s'expriment de façon non linéaire par rapport aux paramètres. En revanche, si $q = 0$, et si l'on considère le modèle original de Tong (3.3), (4.9) s'écrit

$$\begin{aligned} & \sum_{t=1}^N g_{I(t)}^{-2} \left[x_t - \theta_0^{(I(t))} - \sum_{j=1}^p \phi_{I(t)}(j) x_{t-j} \right]^2 \\ &= \sum_{i=1}^{\ell} g_i^{-2} \sum_{t=1}^N \delta\{I(t) = i\} \left[x_t - \theta_0^{(i)} - \sum_{j=1}^p \phi_i(j) x_{t-j} \right]^2. \end{aligned} \quad (4.11)$$

Si les paramètres des modèles des ℓ régimes sont fonctionnellement indépendants, minimiser (4.11) revient à résoudre ℓ problèmes de minimisation distincts qui portent chacun sur une somme de carrés d'une fonction linéaire par rapport aux paramètres. Notons que si $q = 0$ dans le modèle (3.1), à cause de la paramétrisation utilisée pour représenter la moyenne, le problème de minimisation porte sur l'ensemble des paramètres.

Pour simplifier la présentation nous avons supposé connus les x_t , $t < 1$, ce qui conduit à une méthode de type moindres carrés. Si l'on se souvient que conditionnellement à y^t le modèle ARMA avec seuils est un modèle ARMA évolutif dont les coefficients jusqu'en t dépendent de y_1, y_2, \dots, y_t on peut évaluer la fonction de vraisemblance exacte à condition de supposer, par exemple, que pour $t < 1$ le sous-modèle 1 est utilisé. Des méthodes basées sur la factorisation de Cholesky (Mélard [10]) ou le filtre de Kalman (Ansley et Kohn [3]; Mélard [11]) sont disponibles.

Pour des raisons de commodités, les exemples ont été traités par une méthode plus simple, similaire à la méthode des moindres carrés conditionnelle de Box et Jenkins [4]. Les $m = \max(p, q)$ premières observations sont employées pour amorcer les relations de récurrence (3.1). En supposant $a_t = 0$ pour $t \leq m$, on peut alors calculer a_t , $t = m + 1, \dots, N$. On compare plusieurs modèles à l'aide des critères d'information AIC et BIC (par exemple Hannan [8]), calculés au moyen des expressions suivantes

$$\text{AIC} = N [1 + \ln(2\pi) + \ln \hat{\sigma}^2] + 2P \quad (4.12)$$

$$\text{BIC} = N [1 + \ln(2\pi) + \ln \hat{\sigma}^2] + P \ln N \quad (4.13)$$

où P est le nombre total de paramètres. L'optimisation non linéaire a été réalisée notamment par l'algorithme de Davidon, Fletcher et Powell (par exemple Scales [17]).

Dans cet article nous n'étudions pas les propriétés asymptotiques des estimateurs considérés. La question n'est pas simple. Il faudrait préalablement établir une condition nécessaire et suffisante d'ergodicité pour le modèle (2.1), similaire à Petrucci et Woolford [13] avec $\ell = 2$ et $p_1 = p_2 = 1$. Dans le cas d'un processus ergodique satisfaisant (2.1) avec $\ell = 2$, Chan et Tong [5] ont établi un théorème de convergence presque sûre et de normalité asymptotique des estimateurs des moindres carrés conditionnels, dans le cas où les σ_i sont égaux. Petrucci [12] traite le modèle plus simple où $p_1 = p_2 = 1$ mais où les σ_i sont inégaux.

5. Estimation des seuils

Dans le paragraphe 4, on a supposé que les seuils contenus dans le vecteur α sont connus. Nous considérons ici l'estimation de ces seuils. Pour simplifier l'exposé, on se limite à un modèle à un seuil α c'est-à-dire $\ell = 2$. On suppose également que les valeurs de y ne sont pas aléatoires et sont toutes distinctes. Notons $(y_{(s)}; s = 1, \dots, N)$ la suite ordonnée par valeurs croissantes des $(y_t; t = 1, \dots, N)$. On note aussi $y_{(0)} = -\infty$ et $y_{(N+1)} = +\infty$. Définissons :

$$y_{(s+0,5)}^* = \frac{y_{(s)} + y_{(s+1)}}{2}, \quad s = 0, \dots, N. \quad (5.1)$$

Notons $T = \{1, \dots, N\}$. On définit

$$T_1(\alpha) = \{t \in T \mid y_t < \alpha\} \quad \text{et} \quad T_2(\alpha) = T \setminus T_1(\alpha) = \{t \in T \mid y_t \geq \alpha\}.$$

Il s'ensuit que pour tout $t \in T_1(\alpha)$, $a_t = a_t(\underline{\phi}, \underline{\theta}, \underline{\mu})$ se calcule comme

$$a_t = a_t^{(1)} = x_t - \mu_1 - \sum_{j=1}^p \phi_1(j) (x_{t-j} - \mu_{I(t-j)}) + \sum_{k=1}^q \theta_1(k) a_{t-k} \quad (5.2)$$

tandis que pour tout $t \in T_2(\alpha)$, a_t se calcule comme

$$a_t = a_t^{(2)} = x_t - \mu_2 - \sum_{j=1}^p \phi_2(j) (x_{t-j} - \mu_{I(t-j)}) + \sum_{k=1}^q \theta_2(k) a_{t-k}. \quad (5.3)$$

Considérons maintenant la somme des carrés $S(\alpha) = S(\alpha; \underline{\phi}, \underline{\theta}, \underline{\mu}, \underline{g})$ donnée par (4.9), comme fonction de α . On peut la décomposer sous la forme

$$S(\alpha) = \sum_{t \in T_1(\alpha)} g_1^{-2} (a_t^{(1)})^2 + \sum_{t \in T_2(\alpha)} g_2^{-2} (a_t^{(2)})^2. \quad (5.4)$$

Il est clair que $S(\alpha)$ reste constant quand α varie dans $]y_{(j)}; y_{(j+1)}]$ et vaut $S(y_{(j+0,5)}^*)$, pour tout $j = 0, 1, \dots, N$.

$S(\alpha)$ est une fonction en escalier dont les discontinuités surviennent en les $y_{(j)}$. Afin de pouvoir estimer α simultanément avec les autres paramètres du modèle il faudrait que $S(\alpha)$ soit continue par rapport à α . On va donc procéder à un lissage de $S(\alpha)$ afin d'éliminer les discontinuités. La méthode simple que nous préconisons se base sur la comparaison de (5.4) quand $\alpha \in]y_{(j-1)}; y_{(j)}]$ et quand $\alpha \in]y_{(j)}; y_{(j+1)}]$.

Ces deux sommes ont $N-1$ termes en commun : ce sont ceux qui correspondent à $t \in T_1(y_{(j-0,5)}^*) \cup T_2(y_{(j+0,5)}^*)$. La discontinuité en $y_{(j)}$ survient à cause du terme correspondant à t égal à τ , où $\tau = I(j)$. Si $\alpha \leq y_{(j)}$, ce terme vaut $S_2 = g_2^{-2} (a_\tau^{(2)})^2$ tandis que si $\alpha > y_{(j)}$, ce terme vaut $S_1 = g_1^{-2} (a_\tau^{(1)})^2$. Pour $\alpha \in]y_{(j-0,5)}^*; y_{(j+0,5)}^*]$

où $j \in \{1, \dots, N\}$, nous définissons $T(\alpha)$ comme suit

$$T(\alpha) = \begin{cases} S_2 + \frac{S_1 - S_2}{2} \left[\frac{\alpha - y_{(j-0,5)}^*}{y_{(j)} - y_{(j-0,5)}^*} \right]^2, & \alpha \leq y_{(j)}, \\ S_1 + \frac{S_1 - S_2}{2} \left[\frac{\alpha - y_{(j+0,5)}^*}{y_{(j)} - y_{(j+0,5)}^*} \right]^2, & \alpha > y_{(j)}. \end{cases} \quad (5.5)$$

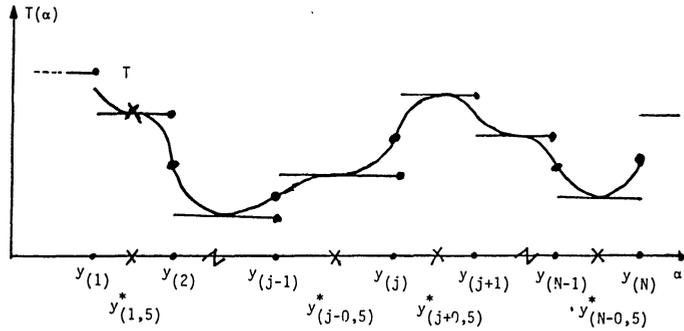


Figure 2 – Graphe de $T(\alpha)$, défini par (5.5)

Toutefois, si $\alpha \leq y_{(1)}$, on pose $T(\alpha) = S_2$ tandis que si $\alpha > y_{(N)}$ on pose $T(\alpha) = S_1$. On définit finalement

$$\tilde{S}(\alpha) = \sum_{t \in T_1(y_{(j-0,5)}^*)} g_1^{-2} (a_t^{(1)})^2 + \sum_{t \in T_2(y_{(j+0,5)}^*)} g_2^{-2} (a_t^{(2)})^2 + T(\alpha). \quad (5.6)$$

Il est évident que $\tilde{S}(\alpha)$ est une fonction continue de α dans $]y_{(1)}, y_{(N)}]$. En effet, on a remplacé S_2 ou S_1 par une fonction de α qui est quadratique par morceaux et qui vaut S_2 en $y_{(j-0,5)}^*$, S_1 en $y_{(j+0,5)}^*$ et $(S_1 + S_2)/2$ en $y_{(j)}$. La figure 2 illustre la forme de $T(\alpha)$.

6. Exemples

6.1. Les séries simulées

Nous reprenons ici les deux séries de 200 points générées au paragraphe 3. Nous avons considéré le modèle (3.1) qui est le modèle correct pour la première série mais qui n'est pas le modèle correct pour la deuxième.

Les paramètres ont été estimés d'abord à partir des 50 premières données puis sur l'ensemble des 200 données. Les résultats sont présentés au tableau II. On constate que les valeurs estimées des paramètres sont très proches des vraies valeurs pour la première série, avec la longueur 200. Pour la deuxième série les résultats sont mauvais car le modèle n'est pas correct.

Tableau II
Estimation des paramètres du modèle (3.5)

Série	N	$\phi_1(1)$	μ_1	$\theta_2(1)$	μ_2	α	$\sigma_1^2 = \sigma_2^2$
1	50	0,84	0,46	0,59	4,93	-0,019	1,19
	200	0,78	0,83	0,47	5,03	-0,008	1,03
2	50	1,07	2,57	0,40	4,98	-0,019	2,16
	200	1,11	2,87	0,35	5,04	-0,018	1,53
vraies valeurs		0,80	1	0,50	5	0	1

6.2. Les données de nombre de lynx

Les données de nombre annuel de lynx pris au piège dans le district de la rivière Mackenzie entre 1821 et 1934 ont fait l'objet de plusieurs études. Tong [20, p. 187] a obtenu, sur les 100 premières données en logarithmes, le modèle SETAR suivant :

$$x_t = \begin{cases} 0,80 + 1,07x_{t-1} - 0,21x_{t-2} + 0,17x_{t-3} - 0,45x_{t-4} \\ \quad + 0,22x_{t-5} - 0,03x_{t-6} + a_t, & x_{t-2} \leq 3,05, \\ \hat{\sigma}_1^2 = 0,0302, \\ 2,30 + 1,42x_{t-1} - 1,08x_{t-2} - 0,09x_{t-3} + a_t, & x_{t-2} > 3,05, \\ \hat{\sigma}_2^2 = 0,0564, \end{cases}$$

Nous avons réestimé un modèle similaire par la méthode décrite aux paragraphes 4 et 5, en prenant des valeurs initiales plausibles pour μ_1 , μ_2 et α et des valeurs initiales égales à 0, 1 pour les autres paramètres. Les résultats sont présentés dans le tableau III(a).

En reprenant l'optimisation à partir des valeurs finales du tableau III(a), on a obtenu après plusieurs itérations les résultats présentés dans le tableau III(b). En utilisant l'estimation de la matrice hessienne fournie par l'algorithme de Davidon, Fletcher et Powell, les erreurs-types de certains des paramètres sont grandes par rapport aux valeurs estimées. En éliminant du modèle ces paramètres et en reprenant l'estimation on trouve les résultats du tableau III(c). Ce modèle est encore très bon car il n'y a aucune autocorrélation résiduelle significativement différente de 0 au niveau de probabilité de 5 % et le BIC est plus petit que pour les modèles des tableaux III(a) et (b). Ceci tend à montrer que les modèles obtenus par Tong [20] comportent un nombre excessif de paramètres à cause du critère d'information AIC utilisé.

Les modèles utilisés jusqu'ici pour représenter la série de nombre de lynx différent par la formulation du terme indépendant et la méthode d'estimation. Nous considérons maintenant des modèles ARMA au lieu des modèles AR. Les méthodes d'identification de l'analyse des séries chronologiques classique ne pouvant plus s'appliquer nous avons adopté une approche par essai et erreur. Au modèle (c) du tableau III, on ajoute un terme moyenne mobile d'ordre 1. Les résultats sont présentés dans le tableau III (d). Ce modèle est supérieur aux modèles présentés jusqu'ici, même au sens de AIC. Les modèles ARMA avec seuils s'avèrent donc mériter plus d'attention.

Tableau III

Modèles avec seuil estimés sur les données de nombre de lynx

Notations : g_1 voir (4.8); $\hat{\sigma}^2$ voir (4.10); $S(\alpha)$ voir (5.4);

AIC voir (4.12); BIC voir (4.13).

(a)		si $x_{t-2} < 3,22$
	$\begin{cases} x_t = \tilde{x}_t + 3,14 \\ \tilde{x}_t = 1,02\tilde{x}_{t-1} - 0,28\tilde{x}_{t-2} + 0,20\tilde{x}_{t-3} - 0,46\tilde{x}_{t-4} \\ \quad + 0,31\tilde{x}_{t-5} - 0,29\tilde{x}_{t-6} + 0,22\tilde{x}_{t-7} + a_t \end{cases}$	
		si $x_{t-2} \geq 3,22$
	$\begin{cases} x_t = \tilde{x}_t + 3,14 \\ \tilde{x}_t = 1,52\tilde{x}_{t-1} - 1,28\tilde{x}_{t-2} + 0,05\tilde{x}_{t-3} + a_t \end{cases}$	
	$g_1 = 0,873 \quad \hat{\sigma}^2 = 0,0457$ $S(3,22) = 3,521 \quad \text{AIC} = -10,6 \quad \text{BIC} = 28,2$	
(b)		si $x_{t-2} < 3,22$
	$\begin{cases} x_t = \tilde{x}_t + 3,18 \\ \tilde{x}_t = 1,03\tilde{x}_{t-1} - 0,29\tilde{x}_{t-2} + 0,21\tilde{x}_{t-3} - 0,47\tilde{x}_{t-4} \\ \quad + 0,33\tilde{x}_{t-5} - 0,31\tilde{x}_{t-6} + 0,24\tilde{x}_{t-7} + a_t \end{cases}$	
		si $x_{t-2} \geq 3,22$
	$\begin{cases} x_t = \tilde{x}_t + 3,20 \\ \tilde{x}_t = 1,53\tilde{x}_{t-1} - 1,41\tilde{x}_{t-2} + 0,05\tilde{x}_{t-3} + a_t \end{cases}$	
	$g_1 = 0,877 \quad \hat{\sigma}^2 = 0,0455$ $S(3,22) = 3,507 \quad \text{AIC} = -11,0 \quad \text{BIC} = 27,8$	
(c)		si $x_{t-2} < 3,24$
	$\begin{cases} x_t = \tilde{x}_t + 3,16 \\ \tilde{x}_t = 1,01\tilde{x}_{t-1} - 0,19\tilde{x}_{t-2} - 0,20\tilde{x}_{t-4} \end{cases}$	
		si $x_{t-2} \geq 3,24$
	$\begin{cases} x_t = \tilde{x}_t + 3,17 \\ \tilde{x}_t = 1,53\tilde{x}_{t-1} - 1,28\tilde{x}_{t-2} \end{cases}$	
	$g_1 = 0,893 \quad \hat{\sigma}^2 = 0,0466$ $S(3,24) = 3,958 \quad \text{AIC} = -12,3 \quad \text{BIC} = 13,5$	

(d)

$\begin{cases} x_t = \tilde{x}_t + 3,10 \\ \tilde{x}_t = 0,81\tilde{x}_{t-1} - 0,26\tilde{x}_{t-4} + a_t + 0,30a_{t-1} \end{cases}$	si $x_{t-2} < 3,24$
$\begin{cases} x_t = \tilde{x}_t + 3,21 \\ \tilde{x}_t = 1,21\tilde{x}_{t-1} - 1,06\tilde{x}_{t-2} + a_t - 0,88a_{t-1} \end{cases}$	si $x_{t-2} \geq 3,24$
$g_1 = 0,900 \quad \hat{\sigma}^2 = 0,0433$ $S(3,24) = 3,627 \quad AIC = -18,9 \quad BIC = 9,6$	

6.3. Des données de pollution au SO₂

Les modèles linéaires de type ARIMA sont fréquemment utilisés pour modéliser la pollution atmosphérique, voir par exemple Roy et Pellerin [16] ainsi que les références citées. Dans la suite, nous comparons les modèles ARIMA aux modèles ARMA avec seuils.

Nous avons analysé les données de la station de Warsage en Belgique pour l'année 1977. Il s'agit des moyennes journalières des mesures bihoraires de concentration de SO₂ exprimées en µg par m³ d'air. Les données météorologiques sont relatives à la station de Genk, située à environ 40 km de Warsage.

La méthode de Box et Jenkins [4] appliquée aux logarithmes des données, notées x_t , a conduit au modèle suivant

$$\tilde{x}_t = 0,52\tilde{x}_{t-1} + 0,16\tilde{x}_{t-3} + a_t,$$

où $\tilde{x}_t = x_t - 3,69$, $\hat{\sigma}^2 = 0,159$, AIC = 369, BIC = 384.

Notons y_t et z_t les valeurs de la vitesse du vent et de la température. Le tableau IV montre les corrélations entre $\{x_t\}$, d'une part, et $\{y_t\}$, $\{z_t\}$, $\{y_{t-1}\}$ et $\{z_{t-1}\}$, d'autre part. Nous avons choisi de construire pour $\{x_t\}$ un modèle avec seuil basé sur les vitesses du vent. Il semble raisonnable de ne pas utiliser de délai ou de prendre un délai de un jour au maximum. Nous avons donc essayé les deux possibilités.

Tableau IV

Corrélation entre $\{x_t\}$ (logarithmes des données de pollution) et $\{y_t\}$, $\{z_t\}$, $\{y_{t-1}\}$ et $\{z_{t-1}\}$

	y_t	z_t	y_{t-1}	z_{t-1}
x_t	-0,343	-0,267	-0,203	-0,270

Avec un délai d'un jour, on essaye d'abord un modèle autorégressif d'ordre 4, mettons. Les valeurs estimées sont reproduites dans le tableau V(a). Il n'y a qu'une autocorrélation résiduelle significative, d'ordre 6, ce qui semble indiquer que le modèle est acceptable quoique comportant trop de paramètres. En éliminant les paramètres non significatifs, on obtient un modèle plus simple repris dans le tableau V(b).

Nous avons essayé d'ajouter des termes de type moyenne mobile mais sans succès. Les résultats ne sont pas présentés.

L'analyse a ensuite été recommencée sans faire intervenir de délai dans l'action des variables météorologiques. Pour le modèle autorégressif d'ordre 4, on obtient les résultats du tableau V(c). En éliminant un à un les paramètres non significatifs, le modèle se réduit à un modèle autorégressif d'ordre 1 avec seuil, présenté dans le tableau V(d). Parmi les quatre modèles du tableau V, le meilleur au sens du BIC minimum est le modèle (b).

Tableau V

Modèles avec seuils sur la série de données de pollution
Notations : g_1 voir (4.8); $\hat{\sigma}^2$ voir (4.10); $S(\alpha)$ voir (5.4);
AIC voir (4.12); BIC voir (4.13).

(a)

	si $y_{t-1} < 7,82$
$\begin{cases} x_t = \tilde{x}_t + 3,71 \\ \tilde{x}_t = 0,78\tilde{x}_{t-1} - 0,14\tilde{x}_{t-2} + 0,13\tilde{x}_{t-3} + 0,02\tilde{x}_{t-4} + a_t \end{cases}$	
	si $y_{t-1} \geq 7,82$
$\begin{cases} x_t = \tilde{x}_t + 3,61 \\ \tilde{x}_t = 0,40\tilde{x}_{t-1} - 0,04\tilde{x}_{t-2} + 0,15\tilde{x}_{t-3} + 0,08\tilde{x}_{t-4} + a_t \end{cases}$	
$g_1 = 0,896$ $\hat{\sigma}^2 = 0,1526$ $S(7,82) = 53,1$ AIC = 362 BIC = 413	

(b)

	si $y_{t-1} < 7,83$
$\begin{cases} x_t = \tilde{x}_t + 3,75 \\ \tilde{x}_t = 0,77\tilde{x}_{t-1} + a_t \end{cases}$	
	si $y_{t-1} \geq 7,83$
$\begin{cases} x_t = \tilde{x}_t + 3,61 \\ \tilde{x}_t = 0,44\tilde{x}_{t-1} + a_t \end{cases}$	
$g_1 = 0,890$ $\hat{\sigma}^2 = 0,1559$ $S(7,83) = 55,7$ AIC = 364 BIC = 392	

(c)

	si $y_t < 7,89$
$\begin{cases} x_t = \tilde{x}_t + 3,88 \\ \tilde{x}_t = 0,73\tilde{x}_{t-1} - 0,18\tilde{x}_{t-2} + 0,16\tilde{x}_{t-3} + 0,04\tilde{x}_{t-4} + a_t \end{cases}$	
	si $y_t \geq 7,89$
$\begin{cases} x_t = \tilde{x}_t + 3,58 \\ \tilde{x}_t = 0,42\tilde{x}_{t-1} - 0,01\tilde{x}_{t-2} + 0,12\tilde{x}_{t-3} + 0,07\tilde{x}_{t-4} + a_t \end{cases}$	
$g_1 = 0,908 \quad \hat{\sigma}^2 = 0,1495$ $S(7,89) = 52,2 \quad AIC = 356 \quad BIC = 407$	

(d)

	si $y_t < 7,64$
$\begin{cases} x_t = \tilde{x}_t + 3,83 \\ \tilde{x}_t = 0,65\tilde{x}_{t-1} + a_t \end{cases}$	
	si $y_t \geq 7,64$
$\begin{cases} x_t = \tilde{x}_t + 3,61 \\ \tilde{x}_t = 0,49\tilde{x}_{t-1} + a_t \end{cases}$	
$g_1 = 0,945 \quad \hat{\sigma}^2 = 0,1586$ $S(7,64) = 56,8 \quad AIC = 372 \quad BIC = 399$	

7. Conclusions

Cet article avait pour but d'étudier les modèles ARMA avec seuil qui avaient seulement été définis par Tong [20] dans sa monographie sur les modèles autorégressifs avec seuil. Nous nous sommes d'abord heurtés au choix d'une paramétrisation pour le terme indépendant qui a une incidence sur l'interprétation du modèle. Il faut noter que la paramétrisation choisie ici n'est pas compatible avec la méthode d'estimation rapide proposée par Tong. Notre analyse a montré que cette méthode n'est pas applicable dans le cas ARMA. Il a donc fallu recourir à l'optimisation non linéaire, avec les inconvénients que comporte cette méthode. En revanche, elle permet de la souplesse dans la modélisation, en particulier des restrictions sur les paramètres dont nous avons profité par souci de parcimonie. En employant BIC, un critère d'information qui pénalise les paramètres inutiles plus que ne le fait AIC, nous avons d'ailleurs observé que les modèles retenus par Tong sont surparamétrés. Ceci rejoint des conclusions tirées par Teräsvirta et Luukkonen [18]. Parfois, mais pas toujours, l'ajout de termes de type moyenne mobile améliore l'ajustement de façon appréciable.

Nous devons également tirer certaines conclusions négatives. En principe l'estimation du seuil simultanément avec les autres paramètres devrait être plus commode que la méthode préconisée par Tong. Nous avons observé que la somme des carrés des erreurs est une fonction du seuil qui présente parfois plusieurs extrema locaux. Il faut donc essayer plusieurs valeurs initiales dans la méthode d'optimisation

et même plusieurs algorithmes si l'on veut s'assurer de la validité de l'estimation obtenue. Une autre difficulté que nous avons rencontrée est l'absence de méthodes d'identification. Il n'existe pas actuellement de généralisation des autocorrélations et autocorrélations partielles qui puissent guider l'utilisateur dans le choix des modèles ARMA. Nous avons dès lors procédé par tâtonnement, avec beaucoup de satisfaction. Cela signifie que nous avons d'abord essayé des modèles autorégressifs d'ordre assez élevé (5 à 7) que nous avons simplifié par élimination des paramètres non significativement différents de 0 au niveau de signification de 5 %, en veillant à ce qu'il n'y ait pas d'autocorrélation résiduelle significative. Dès qu'un modèle stable a été obtenu, nous avons ajouté des coefficients du type moyenne mobile et ainsi de suite jusqu'à l'obtention d'un modèle satisfaisant. Enfin, nous n'avons pas abordé non plus dans notre étude les propriétés de cycle limite des modèles ARMA.

Remerciements

Ce travail est la continuation d'un travail de fin d'études entrepris par Chantal Van Dyck. Elle a écrit une première version des programmes permettant l'estimation des paramètres de modèles simples. Elle a également procédé au traitement préliminaire des données de pollution. Nous l'en remercions. Ce projet a bénéficié du soutien de la coopération scientifique entre la Province de Québec et la Communauté française de Belgique et de la coopération scientifique entre la France et la Communauté française de Belgique. Les commentaires de Christian Gouriéroux sur une première version de ce texte ont été très appréciés. Nous remercions les deux referees anonymes dont les commentaires ont contribué à améliorer notablement l'article.

Références

- [1] H. AKAIKE, (1978). — On the likelihood of a time series model, *Statistician*, 27, 215–235.
- [2] J. ANDĚL, I. NETUKA et K. ZVÁRA (1984). — On threshold autoregressive processes, *Kybernetika*, 20, 89–106.
- [3] C.F. ANSLEY et R. KOHN (1983). — Exact likelihood of vector autoregressive moving-average process with missing or aggregated data, *Biometrika*, 70, 275–278.
- [4] G.E.P. BOX et G.M. JENKINS (1976). — *Time series analysis, forecasting and control*, Holden-Day, San Francisco, (revised edition).
- [5] K.S. CHAN et H. TONG (1986). — On estimating thresholds in autoregressive models, *J. Time Ser. Anal.*, 7, 179–190.
- [6] H. CRAMÉR (1961). — On some classes of nonstationary stochastic processes, In *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, Vol. 2, University of California Press, Berkeley and Los Angeles, pp. 57–78.
- [7] C.W.J. GRANGER et A.P. ANDERSEN (1978). — *An introduction to bilinear time series models*, Vandenhoeck and Ruprecht, Göttingen.

- [8] E.J. HANNAN (1980). — The estimation of the order of an ARMA process, *Ann. Statist.*, 8, 1071-1081.
- [9] V. HANNAN et T. OZAKI (1980). — Amplitude-dependent exponential AR model fitting for non-linear random vibrations, In O.D. Anderson (Ed.) *Time Series*, North-Holland, Amsterdam, pp 57-71.
- [10] G. MÉLARD (1982). — The likelihood function of a time-dependent ARMA model, In O.D. Anderson and M.R. Perryman (Eds.) *Applied Time Series Analysis*, Proceedings of the International Conference held at Houston, Texas, August, 1981, North-Holland, Amsterdam, pp 229-239.
- [11] G. MÉLARD (1985). — *Analyse de données chronologiques*, Coll. Séminaire de Mathématiques Supérieures de l'Université de Montréal n° 89, Presses de l'Université de Montréal, Montréal.
- [12] J.D. PETRUCCELLI (1986). — On the consistency of least squares estimators for a threshold AR(1) model, *J. Time Ser. Anal.*, 7, 269-278.
- [13] J.D. PETRUCCELLI et S.W. WOOLFORD (1984). — A threshold AR(1) model, *J. Appl. Prob.*, 21, 270-286.
- [14] M.B. PRIESTLEY (1980). — State-dependent models : a general approach to nonlinear time series, *J. Time Ser. Anal.*, 1, 47-71.
- [15] P.M. ROBINSON (1977). — The estimation of a non-linear moving average model, *Stochastic Processes and their Applications*, 5, 81-90.
- [16] R. ROY et J. PELLERIN (1982). — On long term air quality trends and intervention analysis, *Atmospheric Environment*, 16, 161-169.
- [17] L.E. SCALES (1985). — *Introduction to non-linear optimization*, MacMillan, London.
- [18] T. TERASVIRTA et R. LUUKKONEN (1985). — Choosing between linear and threshold autoregressive models, In O.D. Anderson (Ed.), *Time Series Analysis : Theory and Practice*, 7, North-Holland, Amsterdam, pp. 129-137.
- [19] H. TONG (1978). — On a threshold model, In C.H. Chen (Ed.), *Pattern recognition and signal processing*, Sythoff and Noordhoff, Alphen aan den Rijn, pp 575-586.
- [20] H. TONG (1983). — *Threshold models in non-linear time series analysis*, Lecture Notes in Statistics 21, Springer-Verlag, New York.
- [21] H. TONG et K.S. LIM (1980). — Threshold autoregression, limit cycles and cyclical data (with discussion), *J. Roy. Statist. Soc. Ser. B*, 42, 245-292.