

Un modèle de régression linéaire généralisé dans le domaine des ondelettes pour des erreurs de mesure en $\frac{1}{f}$

Mohamed Jalal FADILI¹, Ed BULLMORE²

¹GREYC-ISMRA UMR 6072, Bd du Maréchal Juin, 14050 Caen Cedex, France

²Brain Mapping Unit, Addenbrooke's Hospital, University of Cambridge UK

Jalal.Fadili@greyc.ismra.fr, etb23@cam.ac.uk

Résumé – Les processus à longue mémoire peuvent sérieusement compromettre l'estimation des paramètres des modèles linéaires ainsi que les variances associées. Les modèles classiques ARMA peuvent fournir une approche adéquate au problème de courte dépendance pour les systèmes linéaires invariants dans le temps. Cependant, ils s'avèrent inefficaces pour des bruits invariants d'échelle. Ici, nous présentons une nouvelle approche de moindres-carrés pondérés dans le domaine des ondelettes (MCPO). Celle-ci fournit un estimateur du maximum de vraisemblance (MV) approché des paramètres du signal et du bruit. Les propriétés statistiques de cet estimateur, y compris une approximation des limites de Cramér-Rao, sont établies d'un point de vue théorique puis comparées à ses performances empiriques pour diverses situations de données simulées. En comparaison aux moindres carrés ordinaires (MCO) ou encore aux estimateurs type ARMA, on montre que le MCPO est plus efficace et permet un excellent contrôle des risques de 1^{ère} et 2^{nde} espèce. La méthode est ensuite illustrée sur des séquences neurophysiologiques d'IRMf.

Abstract – Long-memory noise can seriously confound estimation of linear regression model parameters and their standard errors. Classical ARMA methods can adequately address the problem of LTI, short-memory errors but may be inefficient in the context of scale invariant noise. Here we introduce a novel method, called wavelet-generalised least squares (WLS), which provides approximate Maximum Likelihood Estimates (MLE) of regression model parameters in the context of long-memory errors; as well as approximate MLE of the Hurst exponent characterising the correlational structure of the errors. Statistical properties of this estimator, including its approximate Cramér-Rao bounds, are derived theoretically and compared to its empirical performance on a range of simulated data. Compared to ordinary least squares and ARMA-based estimators, WLS is shown to be more efficient and to give excellent type I and type II error control. The method is finally illustrated on neurophysiological fMRI time series.

1 Introduction

On suppose que l'on dispose d'un vecteur de N observations $\{y_i\}$ avec les p variables explicatives \mathbf{x}_k formant la matrice de dessin \mathbf{X} . Le modèle de régression classique relie les observations \mathbf{y} aux explicatives \mathbf{x}_k via un modèle additif linéaire qui s'écrit:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma) \quad (1)$$

où ε est une composante stochastique modélisant l'incertitude sur les mesures, de moyenne nulle et de covariance Σ , β est un vecteur de paramètres p -dimensionnel et les observations sont d'énergie finie dans \mathbb{R}^N . Les approximations classiques postulent que les erreurs $\{\varepsilon_i\}$ sont indépendantes et normalement distribuées. Les effets de l'écartement de la normalité ont été l'un des principaux thèmes de recherche de la statistique robuste [1][2]. Il existe par ailleurs une littérature extrêmement riche concernant la déviation de l'indépendance, qui a été largement modélisée sous forme de processus à courte mémoire [3], ou encore par des approches dans l'espace des états [4]. Ici nous considérons le cas des modèles linéaires paramétriques où ε est un processus à mémoire longue stationnaire au sens large. Le problème de la régression linéaire dans le contexte d'erreurs à longue dépendance a déjà retenu l'attention de plusieurs auteurs [5][2][6]. Bien que plus efficace que les

MCO, l'estimateur des MCP requiert une estimation itérative de la matrice de covariance Σ , inconnue *a priori*, engendrant à chaque itération une inversion numériquement instable et exigeante en ressources [5][7]. L'objectif de ce papier est de proposer un estimateur dans le domaine des ondelettes pour les paramètres du modèle de régression linéaire avec des erreurs à mémoire longue. L'idée clé est que la projection sur une base d'ondelettes orthonormale agit comme une décomposition de Karhunen-Loève et joue donc le rôle d'un filtre de blanchiment adapté aux processus en $\frac{1}{f}$ [8]. L'estimateur MCPO n'a aucunement besoin d'une inversion explicite de la matrice de covariance et est par conséquent numériquement plus stable et plus rapide. L'algorithme proposé fournit à la fois les paramètres du signal (composante déterministe) et ceux du bruit. Les propriétés des estimées sont établies dans le sens du MV. Des simulations Monte-Carlo sont menées pour appuyer les résultats théoriques et comparer l'estimateur aux modèles de courte dépendance et des MCO en termes de sensibilité, d'efficacité et de contrôle du risque de 1^{ère} espèce.

2 Modèle et algorithme

2.1 Longue dépendance et ondelettes

Nous considérons les processus avec une dépendance à long terme, une propriété de longue mémoire des processus stochastiques stationnaires. Sa forme la plus simple se caractérise par la divergence à l'origine de sa densité spectrale de puissance:

$$S(f) \underset{|f| \rightarrow 0}{\sim} \frac{\sigma^2 c_\gamma}{|f|^\gamma} \quad (2)$$

où l'exposant $\gamma = 2H - 1$ avec H l'exposant de Hurst. c_γ est sans dimension dépendant de γ . Des exemples répandus de tels processus sont le bruit Gaussien fractionnaire (fGn) et le processus ARFIMA. Il est largement connu que les coefficients d'ondelettes d'un processus en $\frac{1}{J}$ stationnaire ont une structure de corrélation qui décroît en $O\left(\left|2^j k - 2^{j'} k'\right|^{2(H-1)-2R}\right)$ [8], où R est le nombre de moments nuls de l'ondelette. Par conséquent, la corrélation résiduelle peut être raisonnablement négligée pour une ondelette de régularité suffisante. La matrice de covariance du processus ε est caractérisée par le paramètre de longue mémoire γ et la variance σ^2 , i.e., $\varepsilon \sim \mathcal{N}(0, \Sigma(\gamma))$. La projection du modèle Eq.1 dans une base d'ondelettes ortho-normale à l'échelle J donne naissance à un processus ε_w suivant une loi $\mathcal{N}(0, \Sigma_J(\gamma))$, où $\Sigma_J(\gamma)$ est une matrice diagonale dont la forme sera explicitée (en ignorant les corrélations résiduelles).

2.2 Maximum de vraisemblance approché

Il est important de noter que la corrélation résiduelle dans le processus projeté est négligée dans le calcul de la fonction de vraisemblance. Par conséquent, l'estimateur du maximum de vraisemblance (EMV) présenté est plus exactement un EMV approché. Il en découle que les propriétés asymptotiques de l'EMV exact ne s'appliquent pas directement à notre estimateur. En d'autres termes, les propriétés établies (e.g. les limites de Cramér-Rao) sont uniquement une approximation aux limites exactes. Le calcul des valeurs exactes entraîneraient une preuve difficile montrant la convergence uniforme de la fonction de vraisemblance approchée vers la vraie fonction de vraisemblance.

Sous hypothèse normale, l'EMV (le terme approché sera ignoré désormais) tente de maximiser la fonction de vraisemblance suivante:

$$LL(\gamma, \beta, \sigma^2) = -\frac{1}{2} \left[N \log(2\pi\sigma^2) + \log(\mathcal{S}_{a_J}) + \sum_j N/2^j \log(\mathcal{S}_{d_j}) \right] - \frac{1}{2\sigma^2} \left[\frac{\left(y_{J,0} - \sum_{l=1}^p x_{J,0}^l \beta_l\right)^2}{\mathcal{S}_{a_J}} + \sum_{j,k} \frac{\left(y_{j,k} - \sum_{l=1}^p x_{j,k}^l \beta_l\right)^2}{\mathcal{S}_{d_j}} \right] \quad (3)$$

\mathcal{S}_{d_j} et \mathcal{S}_{a_J} sont les variances des coefficients de détail et d'approximation de ε . Il est facile de montrer, pour $|\gamma| < 1$, que leurs formes analytiques prennent la forme suivante:

$$\frac{\mathcal{S}_{d_j}}{\sigma^2}(\gamma) = \frac{c_\gamma (2^j)^\gamma}{1 - \gamma} [2 - 2^\gamma], \quad \frac{\mathcal{S}_{a_J}}{\sigma^2}(\gamma) = \frac{c_\gamma (2^{J+1})^\gamma}{1 - \gamma} \quad (4)$$

2.3 L'algorithme MCPO

L'algorithme itératif des MCPO proposé pour résoudre cette équation est le suivant:

1. Choisir le type d'ondelettes, le mode de traitement des effets de bords et l'échelle J .
2. $t = 0$. Initialiser $\beta = \hat{\beta}^{(t)}$ par les MCO, calculer \mathbf{X}_w et \mathbf{y}_w les transformées d'ondelettes discrètes de \mathbf{X} et \mathbf{y} .
3. Soient $d_{j,k}^{(t)} = y_{j,k} - \sum_{m=1}^p x_{j,k}^m \beta_m^{(t)}$ les coefficients de détail des résidus et $a_{J,0}^{(t)} = y_{J,0} - \sum_{m=1}^p x_{J,0}^m \beta_m^{(t)}$ leur coefficient d'approximation. Un calcul direct à partir de l'Eq.3 donne une estimée de $\hat{\sigma}^{(t)2} = \frac{1}{N} \left[\frac{a_{J,0}^{(t)2}}{\mathcal{S}_{a_J}^{(t)}} + \sum_{j,k} \frac{d_{j,k}^{(t)2}}{\mathcal{S}_{d_j}^{(t)}} \right]$
4. Estimer $\hat{\gamma}^{(t)}$ en substituant σ^2 et β par leurs estimées dans l'Eq.3.
5. Construire $\hat{\Sigma}_J^{(t)}$. L'EMV de β est donné par la forme matricielle: $\hat{\beta}^{(t+1)} = \left(\mathbf{X}_w^T \hat{\Sigma}_J^{(t)-1} \mathbf{X}_w \right)^{-1} \mathbf{X}_w^T \hat{\Sigma}_J^{(t)-1} \mathbf{y}_w$. Il est évident que l'inversion de $\hat{\Sigma}_J^{(t)}$ est très rapide puisque celle-ci est diagonale d'où un gain précieux en complexité temporelle.
6. $t = t + 1$. Aller à l'étape 3 et répéter jusqu'à stabilisation des estimées.

Cette approche peut être vue comme une généralisation aux processus en $\frac{1}{J}$ de la méthodologie de Box-Jenkins pour les modèles ARMA.

2.4 Propriétés de l'estimateur des MCPO

On peut prouver que l'EMV de l'Eq.3 correspond à une solution unique et donc à un minimum global de la fonction de vraisemblance. Sous conditions de régularité, l'EMV existe, est convergent et efficace (tout du moins asymptotiquement). De plus, les corrélations résiduelles étant négligeables, le théorème de Gauss-Markov fournit une preuve formelle que notre estimateur est sans biais et à variance minimale. Par ailleurs, les estimées du MV tendent vers une loi normale avec une limite de Cramér-Rao (LCR) sur la matrice de variance-covariance donnée par la matrice d'information de Fisher. Compte tenu de notre remarque précédente, ces limites basses ne seront en fait qu'une approximation aux limites vraies puisque l'EMV est approché. Après un calcul long mais simple, on montre que les LCR approchées se mettent sous la forme:

$$\begin{aligned} V_{\hat{\beta}} &\geq \frac{N}{2N \left[(\log(\mathcal{S}_{a_J}))' \right]^2 + \sum_j N(j) \left[(\log(\mathcal{S}_{d_j}))' \right]^2 - 2 \left[\log(\mathcal{S}_{a_J})' + \sum_j N(j) \log(\mathcal{S}_{d_j})' \right]^2} \\ V_{\hat{\sigma}^2} &\geq \frac{2\sigma^2 \left[(\log(\mathcal{S}_{a_J}))' \right]^2 + \sum_j N(j) \left[(\log(\mathcal{S}_{d_j}))' \right]^2}{N \left[(\log(\mathcal{S}_{a_J}))' \right]^2 + \sum_j N(j) \left[(\log(\mathcal{S}_{d_j}))' \right]^2 - \left[\log(\mathcal{S}_{a_J})' + \sum_j N(j) \log(\mathcal{S}_{d_j})' \right]^2} \\ V_{\hat{\beta}} &\geq \text{diag} \left(\left(\mathbf{X}_w^t \Sigma_J^{-1} \mathbf{X}_w \right)^{-1} \right) \end{aligned} \quad (5)$$

Le symbole ' est celui de la dérivée par rapport à γ . En situation pratique, ces limites basses constituent une bonne

approximation et permettent de calculer les statistiques tests et les régions critiques ou de tester une hypothèse avec un très bon contrôle du taux de faux positifs (TFP).

3 Principaux résultats

3.1 Simulations

Compte tenu de la nature approché de l'EMV présenté, les simulations paraissent plus cruciales pour mieux comprendre les propriétés statistiques à nombre fini d'échantillons et la valeur de notre estimateur. Des signaux synthétiques ($N = 256, J = 8$) de moyenne nulle et de contenu fréquentiel différent sont contaminés par un bruit additif gaussien stationnaire en $\frac{1}{f}$ (ARFIMA ou fGn simulés par des méthodes spectrales [5]). Le contenu fréquentiel permet d'étudier l'interaction entre la complexité de la matrice de dessin, la longue dépendance du bruit et les statistiques des estimées, et leur impact sur les performances du MCPO.

3.1.1 Biais et variance

Pour chacune des 100 réalisations d'un bruit ARFIMA pour $H \in [0, 1]$, un signal périodique est ajouté ($T = 32, \beta \in [0, 1.4] \Rightarrow \text{RSB} \leq 1.5\text{dB}$). Afin d'étudier l'influence du choix de l'ondelette sur l'estimateur des MCPO, la DWT de chaque série simulée est calculée en utilisant des ondelettes de Daubechies de régularité croissante $R = \{1, 2, 4, 6, 8, 10\}$ (Haar, $R = 1$). Le biais et la RMSE du MCPO et du MCP (qui suppose la connaissance parfaite de Σ et donc de H) sont calculés. Ces résultats sont illustrés dans la Fig.1 pour les ondelettes D1, D4 et D10. Le biais est plus faible en moyenne pour l'ondelette la plus régulière ce qui ne risque pas de surprendre. Il reste tout de même négligeable pour toutes les ondelettes. La RMSE de la D10 est significativement plus faible que celle de la D1. Elle est en outre très proche de la LCR exacte calculée en supposant H connu. Dans ces simulations, il est difficile de noter une différence entre le biais et la RMSE obtenus par le MCPO pour les ondelettes D4 et D10. Ainsi, la D4 semble un bon choix de compromis entre les effets de bord et la régularité et sera utilisée par la suite. Tous ces résultats semblent plaider en faveur de notre estimateur comme étant non-biaisé et efficace.

3.1.2 Aspects inférentiels

En vue de quantifier les performances du MCPO en termes d'inférence statistique et de sa robustesse aux différents paramètres impliqués, nous le comparons aux estimateurs MCO et AR. Dans le contexte des modèles linéaires, il est bien connu que le MCO tend à sous-estimer la variance lorsque le bruit présente une corrélation non négligeable [3]. Par conséquent, sa statistique test est sur-estimée conduisant à une amplification du TFP. Des simulations sur la distribution de β sous hypothèse nulle (non incluses¹) ont clairement montré la normalité de la

1. Les résultats complets de ces simulation peuvent être obtenus sur requête.

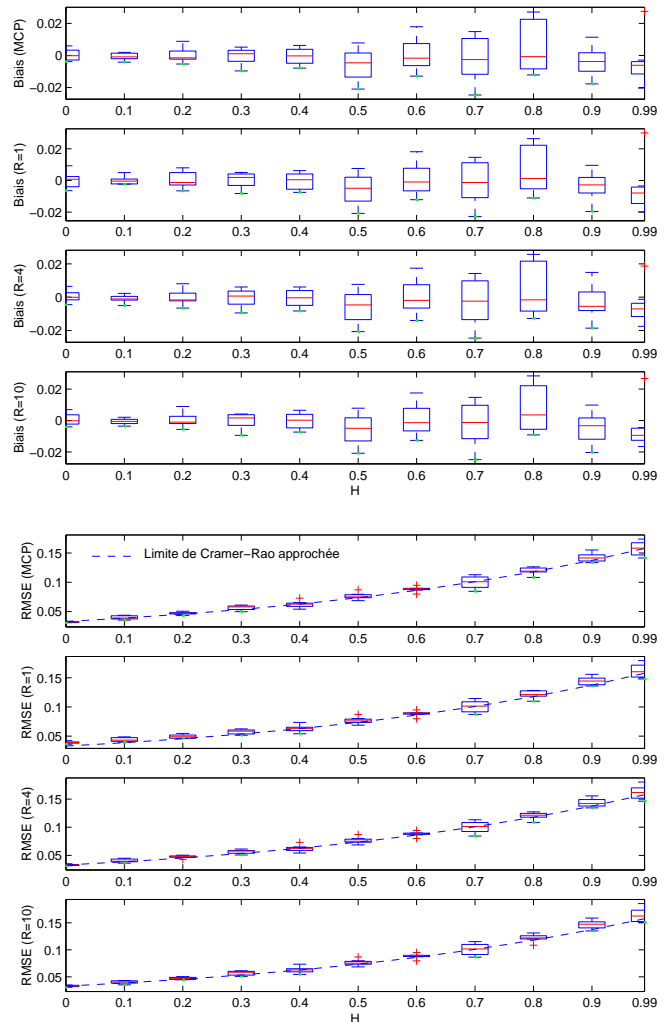


FIG. 1: Biais et variance sur β .

statistique test et la supériorité et la robustesse du MCPO comparé au MCO ou encore à un estimateur AR (sélection automatique de son ordre par le critère Bayésien de Schwartz). De la même manière, nous avons quantifié la performance relative des trois estimateurs en termes de contrôle du risque de 1^{ère} espèce sous $H_0 : \beta = 0$. Afin de quantifier l'effet de H , $M = 500$ répliques ont été tirées pour chaque valeur de $H \in [0.5, 1]$. Le risque nominal est fixé à $\alpha = 5\%$. Comme le montre la Fig.2 et conformément aux prévisions, le MCO est clairement sensible à l'exposant de Hurst. Le TFP observé se dégrade à mesure que H croit. Bien que meilleur que le MCO, le modèle AR engendre encore un TFP supérieur à celui du MCPO, particulièrement pour les valeurs intermédiaires et grande de H . Par ailleurs, l'implémentation de l'estimateur AR est compliquée par les instabilités numériques pour H grand, le rendant inadéquat pour ce type de processus.

Afin d'étudier le risque de 2nde espèce ou réciproquement la puissance de la statistique test, un signal périodique $T = 32$ d'amplitude $\beta \in [0, 1]$ ($\text{RSB} \leq 0\text{dB}$) est contaminé par $M = 500$ réalisations d'un bruit ARFIMA. Cette expérience est répétée pour chacun des H dans $[0, 1]$. La Fig.3 montre clairement que la puissance empirique (motifs) est très proche de la puissance théorique sous hypothèse normale (lignes), bien qu'il semble y avoir une légère variabilité lorsque H approche de la limite de la sta-

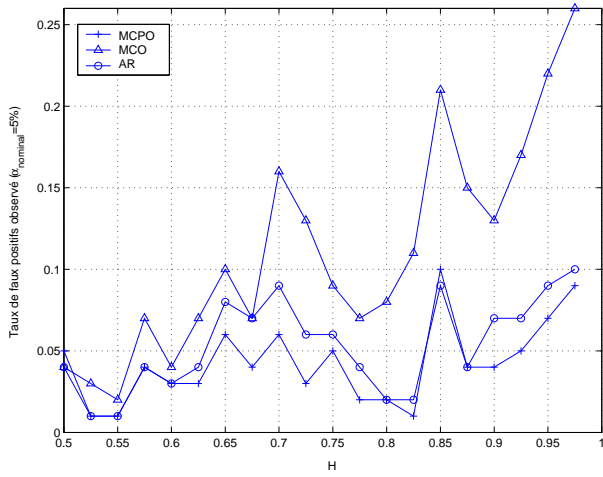


FIG. 2: Risque de 1^{re} espèce.

tionnarité. Ces études sur les aspects inférentiels révèlent que bien que les propriétés de l'EMV approché ne soient pas strictement vérifiées d'un point de vue théorique, les performances de l'estimateur restent très proches du problème idéalisé (estimateur MCP) où les paramètres du bruit sont supposés connus et non estimés.

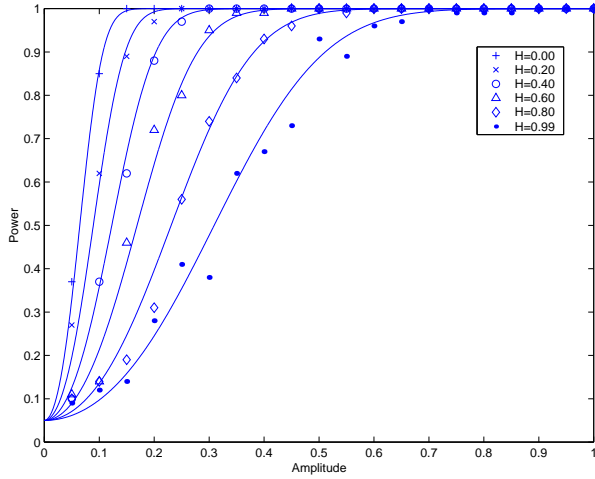


FIG. 3: Puissance du test.

3.2 Données réelles

Les séries neurophysiologiques mesurées en Imagerie par Résonance Magnétique fonctionnelle (IRMf) révèlent l'activité cérébrale lors de l'accomplissement d'une tâche. Le bruit IRMf peut montrer une forte corrélation temporelle pour des raisons physiques et physiologiques. Des séries d'IRMf ont été acquises sur 7 sujets volontaires sains. Ces derniers étaient en position allongée sans accomplir de tâche quelconque (i.e. hypothèse nulle $\beta = 0$). Les séries contenaient 64 à 512 volumes temporels. Ses séries ont été ensuite analysées en utilisant les trois estimateurs précédents avec trois matrices de dessin de complexité différente. Les TFPs moyens observés et leurs écarts types en fonction du risque de 1^{re} espèce nominal sont portées en Tab.1. On peut ainsi apprécier la supériorité du MCPO au MCO et à l'estimateur AR. Le MCO est inadéquat pour une majeure partie des jeux de données particulièrement lorsque le fondamental de la matrice de dessin est aux fréquences basses. L'estimateur AR possède une plus grande

Nominal	MCPO	MCP	AR
0.001	0.0008 (0.0004)	0.0035 (0.0027)	0.0015 (0.0014)
0.002	0.0017 (0.0007)	0.0069 (0.0053)	0.0030 (0.0027)
0.003	0.0025 (0.0010)	0.0103 (0.0079)	0.0045 (0.0041)
0.004	0.0033 (0.0014)	0.0136 (0.0105)	0.0060 (0.0054)
0.005	0.0042 (0.0017)	0.0170 (0.0131)	0.0075 (0.0067)
0.006	0.0050 (0.0021)	0.0203 (0.0156)	0.0090 (0.0080)
0.007	0.0058 (0.0024)	0.0236 (0.0181)	0.0105 (0.0093)
0.008	0.0067 (0.0027)	0.0269 (0.0206)	0.0119 (0.0106)
0.009	0.0075 (0.0030)	0.0301 (0.0230)	0.0134 (0.0118)
0.010	0.0083 (0.0034)	0.0333 (0.0254)	0.0149 (0.0131)

TAB. 1: Risque de 1^{ère} espèce nominal et observé.

variabilité et une plus grande sensibilité à la structure de \mathbf{X} . Ainsi, le risque observé excède le risque nominal dans 8 jeux de données. Ces résultats confortent clairement la validité du modèle de bruit en $\frac{1}{f}$ pour les séries d'IRMf.

4 Conclusion

Il ressort de ces simulations que notre estimateur conforte largement les prévisions théoriques. La variance empirique de l'estimateur reste très comparable aux limites de Cramér-Rao. Les performances sont peu sensibles aux paramètres du bruit (à RSB constant) et à la complexité de \mathbf{X} , ce qui n'est nullement le cas pour les MCO et les modèles ARMA. Ce gain s'effectue sans pénalisation algorithmique puisque la procédure converge généralement au bout de 4 ou 5 itérations. L'application à des séries temporelles neurophysiologiques en IRMf, susceptibles d'être contaminées par un bruit en $\frac{1}{f}$ sont très encourageants. Ils montrent une adéquation avec des résultats connus des praticiens. Finalement, le modèle proposé peut être affiné sur plusieurs plans notamment afin d'intégrer une source de bruit blanche supplémentaire ou encore pour traiter les processus localement stationnaires.

Références

- [1] P. Huber, *Robust statistics*, New York, Wiley, 1981.
- [2] H.L. Koul, "Asymptotics of r-, md- and lad- estimators in linear regression models with long range dependent errors," *Probability Theory and Related Fields*, vol. 95, pp. 535–553, 1993.
- [3] G. E. P. Box and G. M. Jenkins, *Time series analysis: forecasting and control*, San Francisco, Holden-Day, Revised edition, 1976.
- [4] R.H. Jones, *Longitudinal data with serial correlation: A state-space approach*, Monographs on statistics and applied probability 47. London, Chapman & Hall, 1993.
- [5] J. Beran, *Statistics for long-memory processes*, Monographs on statistics and applied probability 61. London, Chapman & Hall, 1994.
- [6] P. M. Robinson and F. J. Hidalgo, "Time series regression with long range dependence," *London, Suntory and Toyota International Centres for Economics and Related Disciplines*, 1997.
- [7] A. Azzalini, *Statistical inference based on the likelihood*, Monographs on statistics and applied probability 68. London, Chapman & Hall, 1996.
- [8] G. W. Wornell, "A karhunen-loève-like expansion for 1/f process via wavelets," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 859–861, 1990.