



Longue mémoire pour des séries irrégulièrement observées. Application aux Black-out électriques.

Hermann Courteille, Anne Philippe, Frédéric Héliodore, Serge Poullain,
François Gallon

► To cite this version:

Hermann Courteille, Anne Philippe, Frédéric Héliodore, Serge Poullain, François Gallon. Longue mémoire pour des séries irrégulièrement observées. Application aux Black-out électriques.. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386717

HAL Id: inria-00386717

<https://inria.hal.science/inria-00386717>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LONGUE MÉMOIRE POUR DES SÉRIES IRRÉGULIÈREMENT OBSERVÉES. APPLICATION AUX BLACK-OUT ÉLECTRIQUES.

Hermann Courteille^{(1) (2)}, Anne Philippe⁽²⁾, Frédéric Héliodore⁽¹⁾, Serge Poullain⁽¹⁾,
Francois Gallon⁽¹⁾

⁽¹⁾*AREVA TD, 1, place Jean Millier, 92084 Paris La Défense*

⁽²⁾*Laboratoire Jean Leray, 2 rue de la houssinière, 44000 Nantes*

Résumé

Cette note est dédiée à l'estimation du paramètre de longue mémoire dans les séries temporelles observées à des temps aléatoires. L'existence de corrélations à long terme dans la distribution des pannes électriques, i.e. "black-out", doit trouver son origine dans la politique de maintenance et de restauration des opérateurs de réseaux. Les travaux portent sur deux transformations des séries visant à préserver l'exposant de mémoire et à récupérer un échantillonnage régulier. A la suite, une méthode d'estimation prenant en compte les temps d'observation et les valeurs observées est dérivée. Le lien entre persistance ($H > 0.5$) et restauration des réseaux électriques est validé.

Abstract

This note deal with estimation of a long memory parameter when series are observed at random times. Long-term correlations in electric failure distribution could be explained by the maintenance and recovery policy of electrical system operator. This work focus on two transformations which preserve memory exponent and render sampling regular. Futhermore, an estimator whose entries are sampling times and sampled observations is presented. The links between persistancy ($H > 0.5$) and recovery of electric network is finally validated.

mots-clés : *Série temporelle, échantillonnage irrégulier, mémoire longue, black-out.*

1 Introduction

Les réseaux de transport électrique doivent être considérés comme des systèmes complexes en condition limite de fonctionnement. Les exemples récents de black-out majeurs ne peuvent laisser indifférent et il est nécessaire d'appréhender au mieux les mécanismes physiques régissant la ou les dynamiques de ces réseaux. Dobson et al [4] ont proposé un modèle de Criticalité Auto-Organisée (CAO) fondé sur la dynamique du Tas de Sable [1]. Celui ci permet de décrire l'évolution d'un réseau par une dynamique lente (croissance de la charge) et une dynamique rapide (phénomène d'avalanche ou black-out). On peut ainsi reproduire le comportement d'un réseau sur un horizon large et reconstituer un historique de pannes

en fonction de paramètres significatifs [3]. Cet historique de pannes doit être analysé avec toutes les précautions d’usage afin d’en extraire des informations physiques réellement invariantes.

La base de données présente des événements datés en jour T_n et dont l’amplitude est mesurée à travers différents indicateurs tels que le nombre de consommateurs perturbés par l’événement C_n , la puissance P_n non délivrée. On est donc en présence de séries irrégulièrement observées dont on veut évaluer la mémoire. Dans [4], l’échantillonnage est rendu régulier en introduisant des valeurs nulles hors des jours de black-out. Cet ajout n’est pas sans affecter la structure de covariance du processus car il crée en premier lieu de l’instationnarité.

Définition 1. Soit $\mathbf{X} = (\mathbf{X}_n)_{n \in \mathbb{N}}$ un processus à temps discret, stationnaire au second ordre de covariance $\sigma_X(h)$. Le processus $\mathbf{X}$ sera dit à *mémoire longue* si la suite des covariances $\sigma_X(h)$ est non sommable, $\sum |\sigma_X(k)| = \infty$ et usuellement de la forme :

$$\sigma_X(h) = L(h)h^{2H-2} \quad (1)$$

où $H \in]1/2, 1[$ et $L(h)$ une fonction à variation lente à l’infini au sens que $\frac{L(ah)}{L(h)} \xrightarrow{h \rightarrow \infty} 1$, pour tout a .

Une série à mémoire longue est donc caractérisée par une autocovariance qui décroît lentement par opposition à une décroissance exponentielle (série ARMA par exemple).

Dans la suite, nous nous intéressons à l’estimation du paramètre H lorsque l’échantillonnage est irrégulier. La série $\mathbf{X}$ est uniquement observée à des temps aléatoires $(T_n)_{n \in \mathbb{N}}$. On pose $T_0 = 0$. Le processus observé est alors $\mathbf{Y} = (\mathbf{Y}_n)_{n \in \mathbb{N}}$ défini par $Y_n = X_{T_n}$. L’information complète réside alors dans la donnée des couples $(T_n, Y_n)_{n \in \mathbb{N}}$.

2 Présentation des méthodes d’estimation

De nombreuses techniques ont été développées pour estimer le paramètre de longue mémoire (voir [2] pour une revue). Une grande partie de ces estimateurs sont construits comme des fonctionnelles du périodogramme ou de la suite des covariances empiriques qui ne peuvent pas être évaluée en présence de données irrégulièrement observées.

Dans ce contexte, deux stratégies sont possibles, soit on transforme les données pour se ramener à des données régulièrement observées et on peut alors utiliser les estimateurs classiques pour le paramètre H , soit on adapte les estimateurs classiques pour des données $(T_n, Y_n)_{n \in \mathbb{N}}$.

2.1 Mémoire du processus sous échantillonné $\mathbf{Y}$

On s’intéresse ici aux propriétés du processus $\mathbf{Y}$ et plus particulièrement à sa mémoire. Philippe et Viano [5] ont montré que le paramètre de longue mémoire de $\mathbf{X}$ peut être estimé directement à partir de la série régulièrement observée $\mathbf{Y}$.

Proposition 2 ([5]). *Soit $\mathbf{X}$ un processus à longue mémoire au sens (1). On suppose que la suite (T_n) est indépendante de $\mathbf{X}$ et les inter-temps $\Delta T = (T_{n+1} - T_n)$ sont i.i.d. Si $\mathbb{E}(\Delta T) < \infty$ alors la covariance du processus échantillonné $\mathbf{Y}$ satisfait:*

$$\sigma_Y(h) \underset{h \rightarrow \infty}{\sim} h^{2H-2} \mathbb{E}(\Delta T)^{2H-2} L(h)$$

Ainsi le processus $\mathbf{Y}$ reste un processus à longue mémoire et l'intensité est préservée. Le paramètre H peut donc être estimé à partir de $\mathbf{Y}$ en utilisant les estimateurs classiques.

Lorsque l'hypothèse $\mathbb{E}(\Delta T) < \infty$ n'est plus satisfaite, le sous échantillonnage peut alors supprimer la longue mémoire ou réduire son intensité (voir [5]). L'estimation du paramètre H ne peut donc plus s'effectuer uniquement à partir de la connaissance de la série $\mathbf{Y}$.

2.2 Agrégation temporelle

Pour se ramener à des données régulièrement observées, nous allons agréger les observations dans le temps. On somme le processus $\mathbf{X}$ sur des intervalles de temps disjoints de longueur Δ . Formellement, la série agrégée $\mathbf{Y}^{(\Delta)}$ est définie par:

$$Y_k^{(\Delta)} = \sum_{i=k\Delta}^{(k+1)\Delta-1} X_i$$

Proposition 3. *Lorsque $\mathbf{X}$ est à longue mémoire au sens (1) alors la covariance du processus $\mathbf{Y}^{(\Delta)}$ satisfait*

$$\sigma_{Y^{(\Delta)}}(h) \underset{h \rightarrow \infty}{\sim} C(\Delta) h^{2H-2} L(h)$$

Le processus agrégé est donc à longue mémoire et l'intensité de sa mémoire reste égale à celle de $\mathbf{X}$. On peut donc à nouveau estimer le paramètre H à partir de $\mathbf{Y}^{(\Delta)}$ et des estimateurs classiques.

2.3 Auto covariance par bloc

Partant des couples (Y_n, T_n) , nous prenons maintenant le parti, faute d'avoir suffisamment de couples observés distants d'un temps h pour calculer la covariance $\sigma_X(h)$, d'estimer une valeur moyenne de σ_X sur un voisinage de taille $2\Delta + 1$ autour de h (voir [6]). On pose pour tout h

$$R_\Delta(h) = \frac{1}{2\Delta + 1} \sum_{s=h-\Delta}^{h+\Delta} \sigma_X(s)$$

Si l'on suppose X à longue mémoire au sens (1), alors $R_\Delta(h)$ devient:

$$R_\Delta(h) = \frac{1}{2\Delta + 1} \sum_{s=h-\Delta}^{h+\Delta} L(s) s^{2H-2} \underset{h \rightarrow \infty}{\sim} L'(h) h^{2H-2} \quad (2)$$

Ainsi "l'agrégation des covariances" possède le même comportement asymptotique que la suite des covariances.

Un estimateur naturel de $R_\Delta(h)$ est :

$$\widehat{R}_\Delta(h) = \frac{1}{N_h} \sum_{T_i, T_j \in I_h} X_{T_j} X_{T_i}$$

où $I_h = [h - \Delta, h + \Delta]$ et $N_h = \text{card}\{(i, j) : T_i - T_j \in I_h\}$

A partir du comportement de R_Δ donné en (2), on peut estimer par une régression OLS de $\log(R_\Delta(h))$ versus $\log h$, le paramètre H .

Remarque 4. Δ doit faire l'objet d'un compromis biais/variance d'estimation et la plage de régression est à calibrer sur des benchmarks fractionnaires (FARIMA).

3 Paramètre de longue mémoire et restauration des réseaux électriques

3.1 Effet du sous échantillonnage sur un benchmark

Nous illustrons ici la méthode d'estimation directe présentée Section 2.1 sur un modèle FARIMA sous-échantillonné à des temps géométriques, figure(1). Nous voyons alors que la série échantillonnée $\mathbf{Y}$ et la série originale $\mathbf{X}$ décroissent hyperboliquement à la même vitesse et à la constante multiplicative près ($\mathbb{E}(\Delta T)^{2H-2}$). En faisant varier $\mathbb{E}(\Delta T)$ de 2 à 15, un biais apparaît: le sous-échantillonnage altère la structure de mémoire à court terme et, en conséquence l'estimateur utilisé FEXP décrit dans [2].

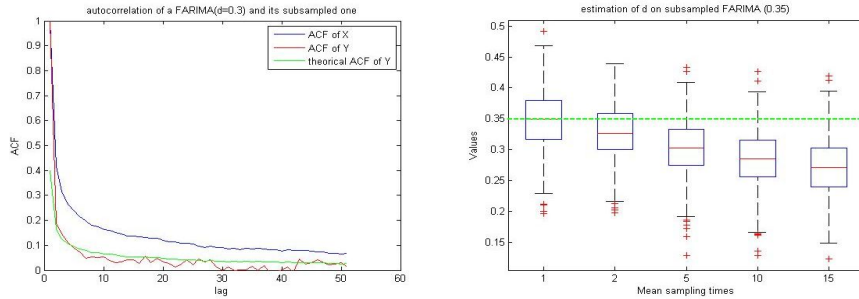


Figure 1: Sur la figure de gauche, fonction d'autocorrélation empirique $\widehat{\sigma}_X(h)$ d'un FARIMA($H = 0.8$), de son sous-échantillonné $\widehat{\sigma}_Y$ ($N = 5000$) à des temps géométriques de moyenne 10 et sa fonction d'autocorrélation théorique. A droite, les "boxplot" de l'estimation sur 500 FARIMA($H = 0.8$) pour les temps moyens d'échantillonnage $\mathbb{E}(\Delta T) = 1, 2, 5, 10, 15$.

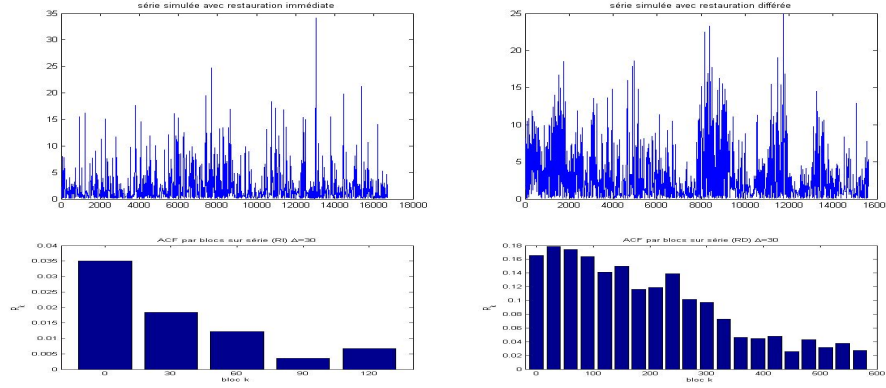


Figure 2: En haut à gauche série de pannes avec restauration immédiate (RI), à droite, avec restauration différée (RD). En dessous les graphes respectifs des ACF par bloc selon la méthode décrite Section 2.3 et sur lesquels nous estimons, par régression log-log, le coefficient H .

3.2 Application aux blackouts

Le modèle de criticalité auto-organisée permettant de reproduire le comportement d'un réseau de transport met en présence deux forces antagonistes :

- l'évolution de la demande électrique (charge), c'est à dire la dynamique lente
- la restauration (ou amélioration) du réseau électrique suite à un évènement de type black-out. La dynamique rapide représente la mise en défaut des lignes conduisant au black-out.

La gestion des flux électriques dans le réseau doit répondre à une contrainte d'équilibre entre puissance produite et puissance consommée et est conditionnée par les limites physiques des lignes et des générateurs. L'environnement aléatoire (évènement climatique, erreur humaine, ...) est pris en compte dans la mise en défauts des lignes.

L'action de la restauration doit alors être analysée précisément car celle ci est une variable à part entière dans le modèle CAO. En d'autres termes, il importe de savoir si le fait de réparer peut avoir des conséquences sur le comportement du réseau, à savoir sur la distribution de pannes ou black-out. Deux cas, simulés sur un horizon temporel de 50 ans, sont présentés sur la figure 1 avec les conditions suivantes, après black-out :

- restauration immédiate (si black-out, amélioration des capacités des lignes de transmission), RI;
- restauration différée (si 8 black-out, amélioration des capacités des lignes de transmission avec un délai de 1 an), RD.

Série	$\mathbb{E}[\Delta T]$	Section 2.1	Section 2.2, $\Delta = 30$	Section 2.3, $\Delta = 10$
(RI)	=12.1	$\hat{H} = 0.64$	$\hat{H} = 0.63$	$\hat{H} = 0.68$
(RD)	8.3	$\hat{H} = 0.84$	$\hat{H} = 0.83$	$\hat{H} = 0.82$

Table 1: Estimation du paramètre de mémoire sur les deux séries simulées dans le cas de la restauration immédiate (RI) et le cas de la restauration différée (RD) avec les 3 méthodes présentées Section 2. La méthode présentée Section 2.1 s'applique car la moyenne des intertemps entre black-out $\mathbb{E}[\Delta T]$ est fini.

L'estimation de H par les trois méthodes est résumée sur le tableau 1. Les méthodes proposées amènent à des valeurs similaires quant à l'estimation du paramètre de Hurst, ce qui permet de s'affranchir du problème de l'échantillonnage. Le choix de la plage de variation dans la technique de l'auto-covariance par bloc peut s'avérer délicat. Elles amènent aussi à des valeurs significatives quant à l'influence du paramètre "restauration". L'apparition de corrélations à long terme sur la distribution des black-out lors de la restauration différée permet d'entrevoir des stratégies de maintenance pour la gestion du risque.

References

- [1] P. Bak, C. Tang, K. Winsenfeld, "Self-Organised Criticality : An expansion of 1/f noise", *Physical Review Letter*, vol. 59, N° 4, pp.381-384, 1987
- [2] J-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, M.S Taqqu, S. Stoev, "Semi-parametric estimation of the long-range dependence parameter : a survey", *in Long range dependence, theory and applications*, Birkhäuser, 2003
- [3] Blackout Team, "Modèle SPFM : Statistical Power Flow Model", Communication interne AREVA-TD
- [4] B.A. Carreras, D.E. Newman, I. Dobson, A.B. Poole, "Initial Evidence for Self-Organized Criticality in Electric Power System Blackouts", *Proceeding of Hawaiï International Conference on System Sciences*, January 4-7, 2000, Maui, Hawaiï
- [5] A. Philippe, M. C. Viano, "Random sampling of long-memory stationary processes", *Preprint arXiv:0802.0817*, octobre 2008
- [6] W. Mayo, "Spectrum measurements with laser velocimeters", *Proc. Dynamic Flow Conf.*, pp.851-68, 1978