

Mouvement Brownien fractionnaire : choix de l'ondelette dans l'estimation de l'exposant de Hurst

Angel SCIPIONI¹, Pascal RISCHETTE², Gérard BONHOMME¹

¹LPMIA, UMR 7040 du CNRS
Université Henri Poincaré, BP239, 54506 Vandoeuvre-lès-Nancy Cedex, France

²CRéA-MAS'Air
BA 701, 13661 Salon Air, France
Angel.Scipioni@iut-longwy.uhp-nancy.fr, Pascal.Rischette@air.defense.gouv.fr
Gerard.Bonhomme@lpmi.uhp-nancy.fr

Résumé – Cet article a pour objectif d'apporter des compléments de réponse à la question du choix de l'ondelette dans le cas de l'estimation du paramètre H du mouvement Brownien fractionnaire par une étude soumettant 270 fBm à 238 ondelettes. Après avoir rappelé les principales propriétés du fBm et du fGn, nous exposons les raisons pour lesquelles un estimateur à ondelettes est particulièrement bien adapté à l'analyse des processus en $1/f$ et décrivons les conditions de l'étude. Si les résultats dressent un classement des meilleures ondelettes, ils mettent également en évidence l'importance de la dimension fractale de l'ondelette utilisée.

Abstract – This article aims to bring complements of answer to the question of the wavelet choice in the case of H parameter estimates of a fractional Brownian motion. These response elements are obtained by a study subjecting 270 fBm to 238 wavelets. After having pointed out the principal properties of the fBm and fGn, we explain the reasons for which a wavelet-based estimator is particularly well adapted to the analysis of $1/f$ processes and describe the conditions of the study. If the results draw up a classification of the best wavelets, they also put in light the importance of the fractal dimension of the wavelet.

1 Introduction

Les plasmas chauds produits par les tokamaks sont des signaux de haute turbulence dont les modèles reposent sur des processus de type fBm. De nombreux estimateurs du paramètre H ont été proposés [1]. Certains reposent sur l'estimation de paramètres géométriques (méthode des boîtes, morphologie mathématique) alors que d'autres mettent en œuvre des méthodes temporelles (maximum de vraisemblance). Il existe également des méthodes fréquentielles (analyse spectrale), une approche multi-échelle (Burlaga-Klein, variance) et surtout l'estimation par ondelettes particulièrement adaptée aux processus en loi de puissance [2]. Veitch et Abry [3] ont d'ailleurs montré leur excellente performance statistique, leur absence de biais et leur très faible variance. Un des points essentiels pour garantir une efficacité maximale de ce type d'estimateur repose sur le choix de l'ondelette la plus appropriée. Or, les ondelettes englobent une telle diversité de caractéristiques, qu'aucune ne peut être retenue a priori. C'est la raison pour laquelle nous proposons une étude visant à apporter des éléments de réponse dans le cas du fBm.

2 fBm et fGn : définitions et propriétés

Le mouvement Brownien fractionnaire (fBm) est un processus stochastique caractérisé par son exposant de Hurst,

noté H et tel que $0 < H < 1$. Il résulte de la généralisation du mouvement Brownien classique ($H = 0,5$) décrite par Kolmogorov en 1939 et approfondie par Mandelbrot et Ness en 1968 [5]. Il est noté $B_H(t)$ et vérifie $\{B_H(t + \delta t) - B_H(t)\} \equiv N(0, \sigma^2 \delta t^{2H})$ et $B_H(0) = 0$. On peut facilement établir que $\{B_H(t)\} \equiv N(0, \sigma^2 t^{2H})$ et donc que $B_H(t)$ est un processus gaussien centré. Sa fonction d'autocovariance s'exprime par :

$$\gamma_{B_H}(t - t') = \frac{\sigma^2}{2} (|t|^{2H} + |t'|^{2H} - |t - t'|^{2H}) \quad (1)$$

et indique que le processus est non stationnaire. $B_H(t)$ est qualifié de processus en $1/f$ par le fait que la moyenne du spectre de Wigner-Ville [6] s'exprime sous la forme $\overline{S_{B_H}}(f) \sim |f|^{-(2H+1)}$.

La relation d'autosimilarité d'un processus gaussien $x(t)$ s'écrit :

$$\mathbb{E}[(x(at))^n] = \mathbb{E}[a^{nH}(x(t))^n] < +\infty, \forall n \in \mathbb{N}, \forall a > 0. \quad (2)$$

Elle est vérifiée aux ordres 1 et 2 et vaut, respectivement, 0 et $\sigma^2 a^{2H} t^{2H}$. Comme tous les moments d'un processus gaussien peuvent être déduits des ordres 1 et 2, les deux relations les caractérisant suffisent à prouver que le mouvement Brownien fractionnaire est statistiquement auto-similaire.

Le bruit Gaussien fractionnaire (fGn), noté $G_{H,\delta t}(t)$, est

le processus incrément du fBm et s'écrit :

$$G_{H,\delta t}(t) = \frac{1}{\delta t} (B_H(t + \delta t) - B_H(t)). \quad (3)$$

Nous avons donc $\{G_{H,\delta t}(t)\} \equiv N(0, \sigma^2 \delta t^{2(H-1)})$, qui indique, comme pour le fBm, le caractère gaussien centré du processus. Sa fonction d'autocovariance :

$$\gamma_{G_{H,\delta t}}(\tau) = \frac{\sigma^2}{2\delta t^2} (|\tau + \delta t|^{2H} - 2|\tau|^{2H} + |\tau - \delta t|^{2H}) \quad (4)$$

met en évidence la stationnarité du processus et révèle donc toute absence d'autosimilarité. Pour $0 < H < 0,5$, $\gamma_{G_{H,\delta t}}(\tau)$ converge et $G_{H,\delta t}(t)$ est qualifié de processus avec une dépendance à court terme alors que pour $0,5 < H < 1$, $\gamma_{G_{H,\delta t}}(\tau)$ diverge et $G_{H,\delta t}(t)$ est qualifié de processus avec une dépendance à long terme, $H = 0,5$ caractérisant le mouvement Brownien classique non corrélé.

3 Estimateur à ondelettes et processus en $1/f$

Soit x le processus à analyser, on désigne par $d_x(j, k) = \langle x, \psi_{j,k} \rangle$ les coefficients de détails de la décomposition en ondelettes discrètes, où $\{\psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k), j \in \{1, \dots, N_j\}\}$ constitue la base générée à partir de l'ondelette mère ψ_0 et utilisée en analyse multirésolution (AMR). Les ondelettes associées à l'AMR constituent un outil très efficace pour l'étude des processus en $1/f$. Examinons les principaux arguments qui défendent cette idée.

S'il existe un processus avec une loi de puissance du type $S(f) \sim \beta |f|^{-\alpha}$ où $\beta = \frac{1}{\pi} k \Gamma(\alpha) \sin(\frac{\alpha+1}{2}\pi)$ (Γ : euclérienne de seconde espèce), alors la fonction d'autocovariance de ce processus est de la forme $\gamma(\tau) \sim k |\tau|^{-(1-\alpha)}$, avec $\alpha = 2H + 1$. Or, une base d'ondelettes est particulièrement bien adaptée à la mise en évidence de la propriété d'invariance d'échelle puisqu'elle est elle-même construite sur l'opérateur de changement d'échelle avec :

$$\psi_{j,0}(t) = 2^{-j/2} \psi_0(2^j t) \text{ et } \gamma_x(j) = \mathbb{E}[d_x^2(j)] = 2^{j\alpha} \beta I_{\alpha, \psi_0}, \quad (5)$$

où $I_{\alpha, \psi_0} = \langle |f|^{-\alpha}, |\psi_0(f)|^2 \rangle$.

La structure doublement orthogonale de la base d'ondelettes associée au nombre de moments nuls ($M \geq 1$) produisent des coefficients $d_x(j, k)$ quasiment décorrés. De ce fait, l'estimation des variances dans l'espace des détails par :

$$\mathbb{E}[d_x^2(j)] = \frac{1}{N_j} \sum_{k=1}^{N_j} d_x^2(j, k) \quad (6)$$

(N_j : nombre de points à l'échelle j) permet une bonne mesure de la pente α . Cette pente peut être obtenue par régression linéaire dans une représentation $x_j = \log_2 2^j$ et $y_j = \log_2(\mathbb{E}[d_x^2(j)])$ où $\mathbb{E}[y_j] = ax_j + b$. Il est possible de construire une régression linéaire pondérée par la variance σ_j^2 des y_j [3] avec

$$\hat{a} = \frac{\sum y_j (S_0 x_j - S_1) / \sigma_j^2}{S_0 S_2 - S_1^2} \equiv \sum y_j w_j \quad (7)$$

où $S_0 = \sum \frac{1}{\sigma_j^2}$, $S_1 = \sum \frac{x_j}{\sigma_j^2}$ et $S_2 = \sum \frac{x_j^2}{\sigma_j^2}$, les pondérations w_j vérifiant $\sum w_j = 0$ et $\sum j w_j = 1$. Cet estimateur atteint la borne de Cramer-Rao relative à l'estimation de a et de plus, est consistant car $\lim_{N \rightarrow +\infty} \text{Var}(\hat{a}) = 0$. Il constitue une alternative intéressante aux estimateurs reposant sur le principe du maximum de vraisemblance. En effet, ces estimateurs voient leurs performances se détériorer lorsque les données reçues s'écartent du modèle référence.

La décomposition en ondelettes d'un processus stochastique autosimilaire à accroissements stationnaires, tels que les fBm, génère un nouveau processus qui lui est stationnaire. En effet, soit $x_w(t) = CWT[x(a, t)]$ où $x(t)$ est un processus stochastique autosimilaire, sa fonction d'autocovariance s'écrit :

$$\gamma_{x_w}(\tau) = -\sigma^2 a^{2H+1} \int |u|^{2H} \gamma_\psi\left(\frac{\tau}{a} - u\right) du \quad (8)$$

et donc $x_w(t)$ est bien un processus stationnaire.

Un estimateur à ondelettes voit sa portée de corrélation considérablement réduite par rapport au processus $x(t)$ initial. Sa divergence à l'origine en $f^{-\alpha}$ est en effet compensée par le comportement de $\Psi(f)$ en $|f|^m$. Comme

$$\begin{aligned} \lim_{\tau \rightarrow +\infty} \gamma_{x_w}(\tau) &= \lim_{f \rightarrow 0} S_{x_w}(f) \\ &\sim a^{2M-1} f^{2(M-H)-1} \sim \tau^{2(H-M)}, \end{aligned} \quad (9)$$

le choix de M permet de régler la diminution de corrélation et, par conséquent, d'accélérer l'efficacité de l'estimateur.

4 Choix de l'ondelette pour la mesure de H

Dans toute application, l'analyse en ondelette est confrontée au problème récurrent du choix de l'ondelette optimale. Dans le cas du fBm, ce choix doit tenir compte de contraintes antagonistes comme, par exemple, le nombre de moments (M). En effet et comme l'indique (10), une valeur de M élevée favorise l'efficacité de l'estimateur alors qu'une valeur faible minimise la propagation due aux effets de bord et évite également de trop délocaliser l'information. Ces considérations nous ont conduit à soumettre des fBm à une grande variété d'ondelettes de manière à dresser un panorama de leur capacité à estimer H et d'en déduire quelques ondelettes particulièrement bien adaptées.

Différentes méthodes de synthèse des fBm existent [9]. Deux d'entre elles ont été mises en œuvre : celle basée sur la décomposition de Cholesky et celle proposée par Meyer et Sellan [7] reposant sur une base d'ondelettes discrètes et dont l'implémentation a été examinée par Abry et Sellan [8]. Les deux résultats obtenus, très peu différents l'un de l'autre, ne permettent pas d'établir une corrélation évidente entre la méthode de synthèse des fBm et la précision des valeurs de l'exposant de Hurst. Grâce à la seconde méthode, choisie a priori, nous avons généré, sous Matlab, 30 occurrences de fBm pour $H \in \{0, 1; 0, 2; \dots; 0, 9\}$. Nous les avons soumis à un estimateur reposant sur 95 ondelettes discrètes (OD) et 143 ondelettes continues et com-

plexes continues (OC), couvrant ainsi les principales familles existantes. que sont Daubechies (**db**), symlets (**sym**), coiflets (**coif**), splines biorthogonales et inverses (**bior** et **rbio**), Meyer (**meyr** et **dmey**), Gauss (**gaus** et **cgau**), Morlet (**mor1** et **cmor**), chapeau mexicain (**mexh**), Shannon (**shan**) et B-spline à fréquence complexe (**fbsp**).

Plusieurs longueurs d'échantillons ont été étudiées. Il s'avère que $N = 2^{13}$ points constitue une taille d'échantillon suffisante, les résultats ne variant guère pour un nombre de points plus élevé. En revanche, pour des longueurs de $N < 2^{12}$ points, le comportement des ondelettes est fortement perturbé (même si le haut du classement présenté ci-dessous reste le même), donnant des résultats instables directement liés à la statistique du signal qui ne peut pas s'exprimer pleinement sur des listes aussi petites.

La valeur de H est obtenue par la mise en œuvre de la régression linéaire pondérée décrite au paragraphe précédent. Afin de limiter l'influence des effets de bord, nous avons employé la technique du mirroring et appliqué la régression à une plage d'échelles automatiquement adaptée à la taille de l'échantillon, à celle du support de l'ondelette et à une variance $\mathbb{E}[d_x^2(j)]$ calculée sur un nombre significatif de coefficients. Ainsi, pour un échantillon de 2^{13} points et une ondelette **db2**, $j \in \{3, 4, \dots, 9\}$.

Comme l'indiquent clairement les figures 1 et 2, de grandes différences comportementales existent entre les OD et les OC. Outre le fait que les OD mesurent H de manière assez régulière, on constate qu'elles sont plus précises lorsque $H \geq 0,5$. A contrario, les OC, qui ont globalement un comportement aléatoire, fournissent des résultats plus stables et fidèles lorsque $H \leq 0,5$. Cependant, pour $H > 0,5$, leurs résultats sont quasiment inexploitable. Dans le cas des familles communes **db**, **coif** et **sym**, les OD font preuve d'une grande stabilité alors que pour les OC, seule la famille **coif** est stable mais avec des mesures insatisfaisantes. En observant plus précisément le cas des OD, on s'aperçoit que les familles **db** et **sym** ont tendance à s'éloigner de plus en plus du paramètre de référence H lorsque l'ordre de l'ondelette augmente. Ceci est d'autant plus vrai que H diminue. Les autres familles **dmey**, **bior** et **rbio** varient plus dans leur comportement. Notons, par exemple, que l'ondelette **bior3.1** fournit un biais systématique alors que ce n'est jamais le cas pour l'ondelette **bior1.3**.

Nous proposons un classement des OD uniquement, aucune OC n'étant satisfaisante sur l'ensemble des valeurs de H , construit autour de l'évaluation du critère :

$$c_i = \mathbb{E} \left[\left(\mathbb{E} \left[\widehat{H}_{i,j}^2 \right] \cdot |H_j - \mathbb{E} \left[\widehat{H}_{i,j} \right]| \right)^2 \right], \quad (11)$$

avec

$$\begin{cases} H : \text{exposant de Hurst de référence,} \\ \widehat{H} : \text{exposant de Hurst estimé,} \\ i : \text{numéro de l'ondelette, } i \in \{1, 2, \dots, 95\}, \\ j : \text{indice de l'exposant de Hurst de référence,} \\ j \in \{0, 1; 0, 2; \dots, 0, 9\}. \end{cases}$$

Le critère c_i de discrimination des ondelettes s'articule

autour du produit $c_{i,j}$ de deux mesures effectuées pour chaque valeur H_j et chaque ondelette i . Le premier terme calcule, sur l'ensemble des occurrences du signal, la variance $\mathbb{E} \left[\widehat{H}_{i,j}^2 \right]$ des coefficients de Hurst estimés. Néanmoins, la mesure de la dispersion autour de la moyenne n'est pas indicatrice d'un éventuel biais entre cette moyenne et la valeur de référence. Ce biais est donc calculé par le second terme. Faisant le choix de privilégier les ondelettes les plus robustes aux différentes valeurs de H , nous attribuons comme valeur finale du critère c_i le résultat de la variance des $c_{i,j}$.

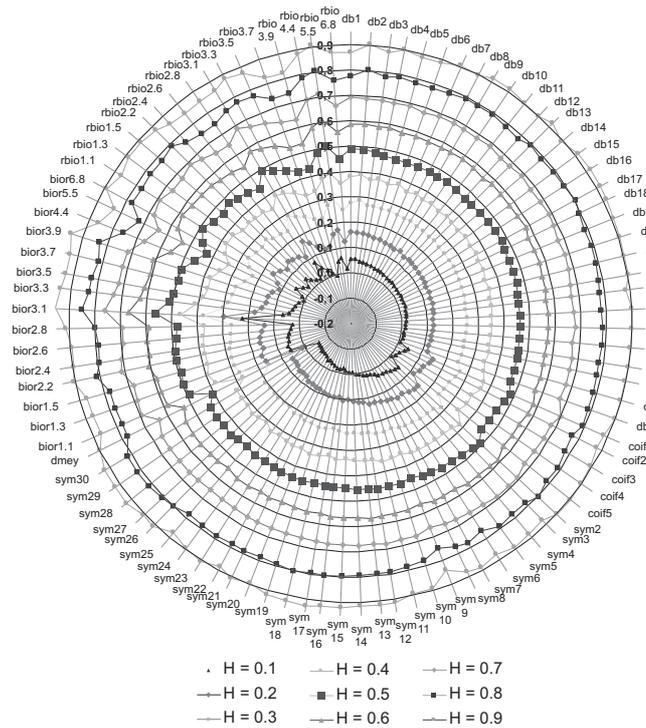


FIG. 1 – Estimation de H par ondelettes discrètes pour $H = 0, 1; 0, 2; \dots; 0, 9$.

TAB. 1 – Classement des 5 meilleures ondelettes discrètes

Rang	1	1	3	4	5
Ondelette	db2	sym2	bior2.2	db9	bior2.4
$10^{11} \cdot c_i$	1,33	1,33	1,64	1,74	1,95

Le tableau 1 présente le début du classement. Il montre que les meilleurs résultats sont obtenus pour les ondelettes **db2** et **sym2** et ont pour origine la bonne adéquation entre leur caractéristique fractale et la nature autosimilaire du fBm. Si la rugosité de l'ondelette ψ , au sens de la régularité lipschitzienne, est suffisante pour mesurer H dans le cas d'un fGn, elle ne l'est plus pour un fBm. Il est donc nécessaire de choisir une ondelette présentant une composante fractale comme le montre la figure 3.

