ON THE LONG RANGE CORRELATION IN FBM-BASED SIGNALS WITH MIXED STATISTICS

A. SCIPIONI

Laboratoire d'Instrumentation Electronique de Nancy,
Université Henri Poincaré, Nancy-Université,
BP 239 Vandœuvre-lès-Nancy Cedex, F-54506, France
E-mail: angel.scipioni@iut-longwy.uhp-nancy.fr

P. RISCHETTE

Research center of the French Air Force,
Morpho-Analysis in Signal processing team,
BA701, F-13661 Salon air, France
E-mail: pascal.rischette@inet.air.defense.gouv.fr

The understanding of energy transfer mechanisms in a tokamak edge plasma is a major challenge for controlled fusion test reactors. High values of the Hurst exponent ($H > 0.5$) encountered in experimental probe data acquired in the scrape-off-layer (SOL) suggest the presence of long-range correlations favoring the hypothesis of an avalanche-type of radial transport. This communication aims at showing that this high value of Hurst coefficient does not necessarily imply the existence of long-time range correlations but it can be the witness of the presence of a particular behavior at small-time scales. Indeed, the development of a wavelet-based observer on synthetic signals, relying on fractional Brownian motions, has allowed the realization of a study model with mixed statistics. The associated time series, for which the H value is controlled, have been broken into blocks of variable length. Then, these different blocks have been scrambled randomly. Although potential long-range correlations have been thus destroyed, the wavelet-based estimator applied to these new synthetic signals is able to measure the original value of the Hurst parameter on a variable scale range. This approach highlights the persistence level on the scale range for several H and block size values. This technique reveals the reminiscent character of the synthetic process behavior appearing from small-time scales to long-time scales.

Keywords: Turbulence; Transport; Long-range correlation; Wavelet; Fractional Brownian motion; Fractional Gaussian noise; Mixed statistical models.

1. Context and motivation

Hot plasmas produced by tokamaks generate high turbulence signals for which models can be built on fractional Brownian motion (fBm) characterized by the Hurst exponent H. Many estimators of the parameter H have been proposed.[1] Some are based on the estimation of geometrical parameters (boxes method, mathematical morphology) while others implement temporal methods (maximum likelihood estimator). There are also frequential methods (spectral analysis), multi-scale approaches (Burlaga-Klein, variance) and especially the estimation by wavelet,[2] particularly well-suited to the process in power-law.[3] Veitch and Abry[4] have also shown their excellent statistical performance, their unbiased feature, and their very low variance. One of the key points to ensure maximum efficiency of this estimator is based on the choice of the appropriate wavelet. That is why we propose a study to find it in the case of the fBm. From this best wavelet and by using the associated estimator, this communication aims at showing that a high value of Hurst coefficient does not necessarily imply the existence of long-time range correlations but it can be the witness of the presence of a particular behavior at small-time scales.

2. Fractional Brownian motion and fractional Gaussian noise: definition and properties

2.1. *Fractional Brownian motion (fBm)*

The fBm is a $1/f$ stochastic process (*cf.* Fig. 1a) characterized by its Hurst exponent H ($0 < H < 1$). It results from the generalization of the classical Brownian motion ($H = 0.5$) described by Kolmogorov in 1939 and thoroughly studied by Mandelbrot and Ness in 1968.[5] It is denoted as $B_H(t)$ and defined by

$$\{B_H(t+\delta t) - B_H(t)\} \equiv \mathcal{N}\left(0, \sigma^2 \delta t^{2H}\right) \text{ with } B_H(0) = 0. \tag{1}$$

From Eq. (1), one can easily establish that $\{B_H(t)\} \equiv \mathcal{N}\left(0, \sigma^2 t^{2H}\right)$ and thus write that $B_H(t)$ is a centred Gaussian process. Its autocovariance function is expressed by the relation

$$\gamma_{B_H}(t-t') = \frac{\sigma^2}{2}\left(|t|^{2H} + |t'|^{2H} - |t-t'|^{2H}\right), \tag{2}$$

which indicates the *non-stationary* character of fBm process. The mean of its Wigner-Ville spectrum is $\overline{S_{B_H}}(f) \sim |f|^{-(2H+1)}$ and that is the reason why fBm is a $1/f$ process.[6]

The self-similarity relation of a Gaussian process $x(t)$ is written as

$$\mathbb{E}\left[(x(at))^n\right] = \mathbb{E}\left[a^{nH}(x(t))^n\right] < +\infty, \forall n \in \mathbb{N}, \forall a > 0. \tag{3}$$

It is verified for orders $n = 1$ and $n = 2$ and is equal to 0 and $\sigma^2 t^{2H}$, respectively. As all moments of a Gaussian process can be deduced from orders $n = 1$ and $n = 2$, the two relations which define them are sufficient to prove that fractional Brownian motion is statistically *self-similar*.

2.2. *Fractional Gaussian noise (fGn)*

The fGn, denoted as $G_{H,\delta t}(t)$, is the fBm increment process (*cf.* Fig. 1b) and is defined by

$$G_{H,\delta t}(t) = \frac{1}{\delta t}\left(B_H(t+\delta t) - B_H(t)\right). \tag{4}$$

We have $\{G_{H,\delta t}(t)\} \equiv \mathcal{N}\left(0, \sigma^2 \delta t^{2(H-1)}\right)$, which indicates, as for the fBm case, the centered Gaussian nature of the process. Its autocovariance function

$$\gamma_{G_{H,\delta t}}(\tau) = \frac{\sigma^2}{2\delta t^2}\left(|\tau+\delta t|^{2H} - 2|\tau|^{2H} + |\tau-\delta t|^{2H}\right) \tag{5}$$

highlights the *stationarity* property of the fGn process and reveals therefore its *non self-similar* nature. For $0 < H < 0.5$, $\gamma_{G_{H,\delta t}}(\tau)$ converges, and $G_{H,\delta t}(t)$ is called short-range correlation process. However, for $0.5 < H < 1$, $\gamma_{G_{H,\delta t}}(\tau)$ diverges, and $G_{H,\delta t}(t)$ is called long-range correlation process. Lastly, $H = 0.5$ is the non-correlated classical Brownian motion.

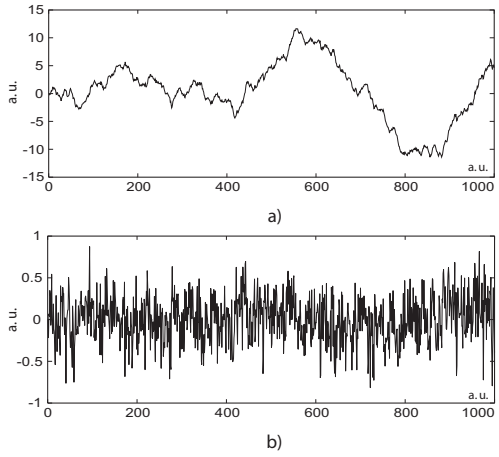

Fig. 1. a) fBm and b) fGn for $H = 0.7$.

3. Wavelet-based estimator and $1/f$ processes

Let x be the process to be analysed, one denotes as $d_x(j,k) = \langle x, \psi_{j,k} \rangle$ detail coefficients of the discrete wavelet decomposition, where $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k),\ j \in \{1, \ldots, N_j\}\}$ is the basis provided by the mother wavelet ψ_0 and used in multiresolution analysis (MRA). Wavelets associated to MRA is an effective tool for the study of $1/f$ processes.

Indeed, if there is a process with a power-law $S(\nu) \sim \sigma_x^2 |\nu|^{-\alpha}$ with $\sigma_x^2 = \frac{1}{\pi} k\Gamma(\alpha) \sin \frac{(\alpha+1)\pi}{2}$ (Γ : Eulerian integral of the second kind), then its autocovariance function is $\gamma(\tau) \sim k|\tau|^{-(1-\alpha)}$, with $\alpha = 2H + 1$. Wavelet basis is particularly well-suited to highlight the scale invariance property because it is built on the scale operator itself with

$$\psi_{j,0}(t) = 2^{-j/2}\psi_0(2^j t), \tag{6}$$

and

$$\begin{aligned} \gamma_x(j) &= \mathbb{E}\left[d_x^2(j)\right] \\ &= 2^{j\alpha}\sigma_x^2 I_{\alpha,\psi_0}, \end{aligned} \tag{7}$$

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

Doubly orthogonal structure of wavelet basis joined to the number of vanishing moments provides almost uncorrelated coefficients. Because of this, detail variance calculus with the expression

$$\mathbb{E}\left[d_x^2(j)\right] = \frac{1}{N_j} \sum_{k=1}^{N_j} d_x^2(j,k), \tag{8}$$

(N_j: number of points at scale j) allows having a good estimate of the slope α.

The existence of this slope $\alpha = 2H+1$ reflects the self-affinity properties, which can be also obtained by the power spectral density but with a weak precision. This is depicted in Fig. 2 and Fig. 3 for a simulated classical Brownian motion $H = 0.5$.

The slope α can be obtained by a weighted linear regression in a representation $x_j = \log_2 2^j = j$ et $y_j = \log_2\left(\mathbb{E}\left[d_x^2(j)\right]\right)$ where $\mathbb{E}[y_j] = ax_j + b$.

The weighted linear regression is built on the variance[4] σ_j^2 of the y_j with

$$\widehat{a} = \frac{\sum y_j \left(S_0 x_j - S_1\right)/\sigma_j^2}{S_0 S_2 - S_1^2} \equiv \sum y_j w_j \tag{9}$$

where $S_0 = \sum \frac{1}{\sigma_j^2}$, $S_1 = \sum \frac{x_j}{\sigma_j^2}$ and $S_2 = \sum \frac{x_j^2}{\sigma_j^2}$, weights w_j verifying $\sum w_j = 0$ and $\sum j w_j = 1$. This estimator reaches the Cramer-Rao

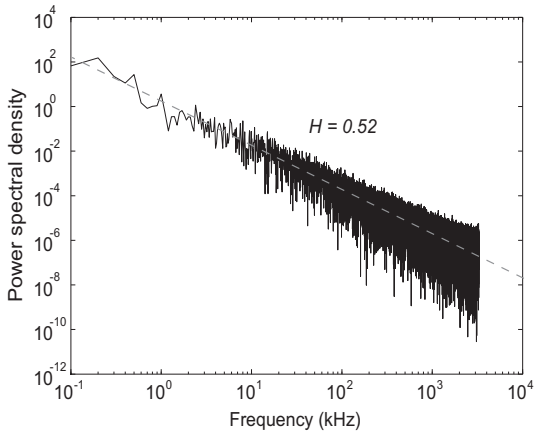

Fig. 2. Brownian motion $H = 0.5$: Hurst parameter from the power spectral density.

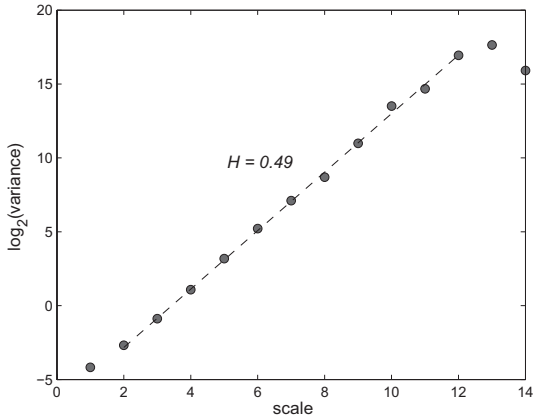

Fig. 3. Brownian motion $H = 0.5$: Hurst parameter from the discrete wavelet transform.

boundary relative to the estimation of a and therefore is consistent because $\lim_{N\to+\infty} \mathrm{Var}(\widehat{a}) = 0$. It is an interesting alternative to estimators based on the maximum likelihood principle. Indeed, performances of these estimators deteriorate when received data deviate from the reference model. The wavelet decomposition of a self-similar stochastic process with stationary increments, such as fBm, generates a new process which is *stationary*. Indeed, if $x_w(t) = \mathrm{CWT}[x(a,t)]$ where $x(t)$ is a self-similar stochastic process and CWT the continuous wavelet transform, then its autocovariance func-

tion is written as

$$\gamma_{x_w}(\tau) = -\sigma^2 a^{2H+1} \int |u|^{2H} \gamma_\psi \left(\frac{\tau}{a} - u\right) du, \tag{10}$$

proving that $x_w(t)$ is a stationary process.

The correlation range of a wavelet-based estimator is considerably reduced with regard to the initial process $x(t)$. Its origin divergence is proportional to $f^{-\alpha}$ and it is balanced by the behavior of $\Psi(f)$ to $|f|^m$. As

$$\begin{aligned} \lim_{\tau \to +\infty} \gamma_{x_w}(\tau) &= \lim_{f \to 0} S_{x_w}(f) \\ &\sim \tau^{2(H-M)}, \end{aligned} \tag{11}$$

the choice of M allows to adjust the correlation decrease and therefore to accelerate the estimator efficiency.

4. Wavelet choice for the estimation of H

In order to find the best wavelet for the estimation of H, we have used the algorithm of Meyer and Sellan[7] built on discrete wavelet basis and whose implementation has been discussed by Abry and Sellan.[8] We have generated 30 fBm occurrences of 2^{13} points for each H value and this, for $H \in \{0.1, 0.2, \ldots, 0.9\}$. We have submitted them to an estimator based on 95 discrete wavelets (DW) and 143 continuous and complex wavelets (CW) covering thus the main families, like Daubechies (`db`), symlets (`sym`), coiflets (`coif`), biorthogonal and reverse splines (`bior`, `rbio`), Meyer (`meyr`, `dmey`), Gauss (`gaus`,`cgau`), Morlet (`morl`, `cmor`), mexican hat (`mexh`), Shannon (`shan`) and β-spline with complex frequency (`fbsp`).

Results presented in Fig. 5 and Fig. 6 show clearly great behavioral differences between DW and CW. First of all, we note that DW measure H with a regular precision and they are more accurate when $H \geqslant 0.5$. On the other hand, CW, which have a random behavior, provide more stable and precise results when $H \leqslant 0.5$. However, for $H > 0.5$, their results are almost unusable as Fig. 7 confirms it.

In the case of usual families like `db`, `coif` and `sym`, DW demonstrate high stability whereas CW are stable only for the `coif` family but with unsatisfactory measures. By observing more specifically the case of the DW, we find that families `db` and `sym` deviate more and more from the reference parameter H when the wavelet order increases. This is especially true when H decreases. Other families `dmey`, `bior` and `rbio` offer a more variable behaviors. For example, the wavelet `bior3.1` provides a systematic bias which is never encountered in the case of the wavelet `bior1.3`.

It is by minimizing quadratic error in the least squares sense that we have found *the best wavelet:* `db2`. It is interesting to note the good adequacy between the fractal nature of this wavelet and the self-similar nature of an fBm (see Fig. 4).

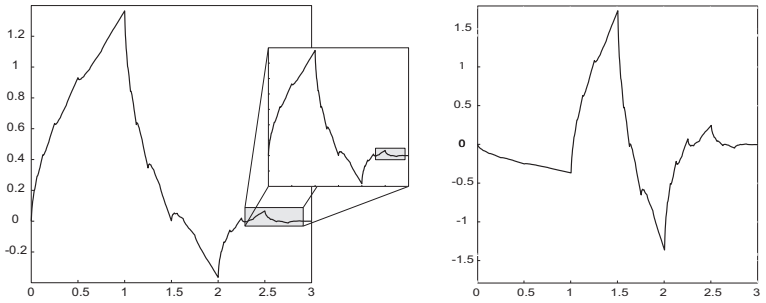

Fig. 4. Wavelet `db2`: a) scale function ϕ and b) wavelet function ψ.

Table 1. Estimation of H by DWT for different block sizes and different areas.

Scales		1 to 3		4 to BS		BS+1 to BS+3
block size (BS)	lag	area 1 (fBm)	lag	area 2 (Bm[a])	lag	area 3 (fGn)[b]
32	1	$H = 0.71$		–	23	$H^* = -0.25$
64	2	$H = 0.71$	12	$H = 0.48$	24	$H^* = -0.26$
128	3	$H = 0.71$	13	$H = 0.54$	25	$H^* = -0.26$
256	4	$H = 0.71$	14	$H = 0.51$	26	$H^* = -0.27$
512	5	$H = 0.71$	15	$H = 0.49$	27	$H^* = -0.28$
1024	6	$H = 0.71$	16	$H = 0.49$	28	$H^* = -0.31$
2048	7	$H = 0.71$	17	$H = 0.49$	29	$H^* = -0.33$
4096	8	$H = 0.71$	18	$H = 0.50$	30	$H^* = -0.40$
8192	9	$H = 0.71$	19	$H = 0.48$	31	$H^* = -0.37$
16384	10	$H = 0.72$	20	$H = 0.49$	32	$H^* = -0.35$
32768	11	$H = 0.72$	21	$H = 0.50$	33	$H^* = -0.27$
65536		–	22	$H = 0.50$		–

Note: [a] Bm: Brownian motion, [b] $H^* = H - 1$.

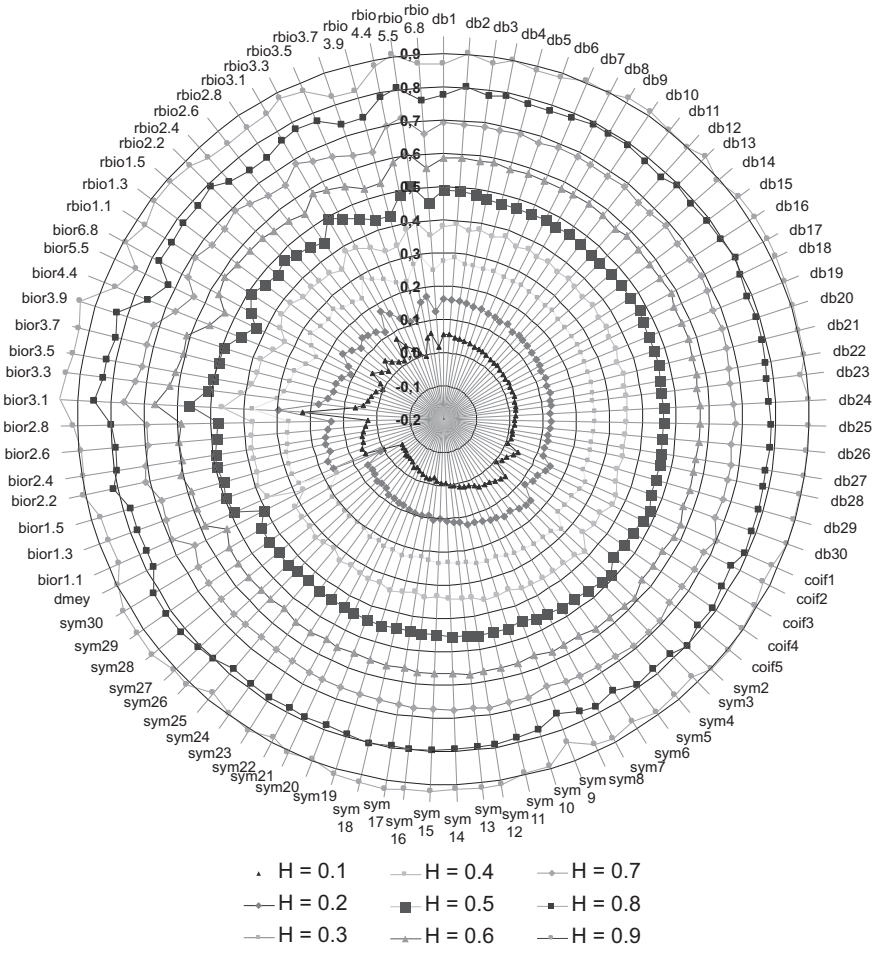

Fig. 5. Estimation of H with discrete wavelets for $H_{\mathrm{ref}} \in \{0.1, 0.2, \ldots, 0.9\}$.

5. Persistence of H in fBm with mixed statistics

In the next step of this work, we study modified fractional Brownian motion signals in order to understand the link between a high value of the Hurst parameter H and long-time correlation.

To this end we have generated an fBm of 2^{20} points with $H = 0.7$ according to values encountered in different tokamak edge plasmas. Time series has been broken into blocks of length M, the blocks scrambled randomly and this, for 12 values of M ranging from 2^5 to 2^{16}. This approach

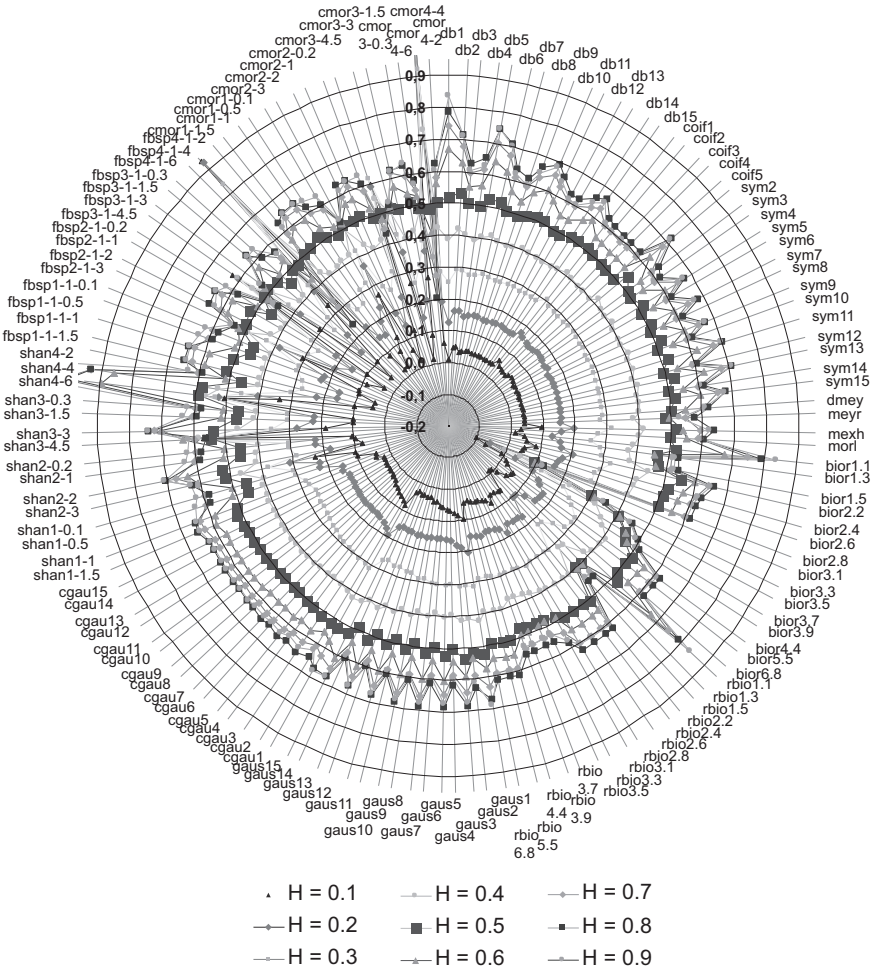

Fig. 6. Estimation of H with continuous wavelets for $H_{\text{ref}} \in \{0.1, 0.2, \ldots, 0.9\}$.

provides a simple model of data with mixed statistics.

The scrambled and unscrambled data were then used to calculate H parameter by using the wavelet-based estimator. The shuffling of blocks of length M should destroy any correlation existing for time scales larger than M.

Results are depicted in Fig. 8. It shows clearly three areas *for all block sizes*:

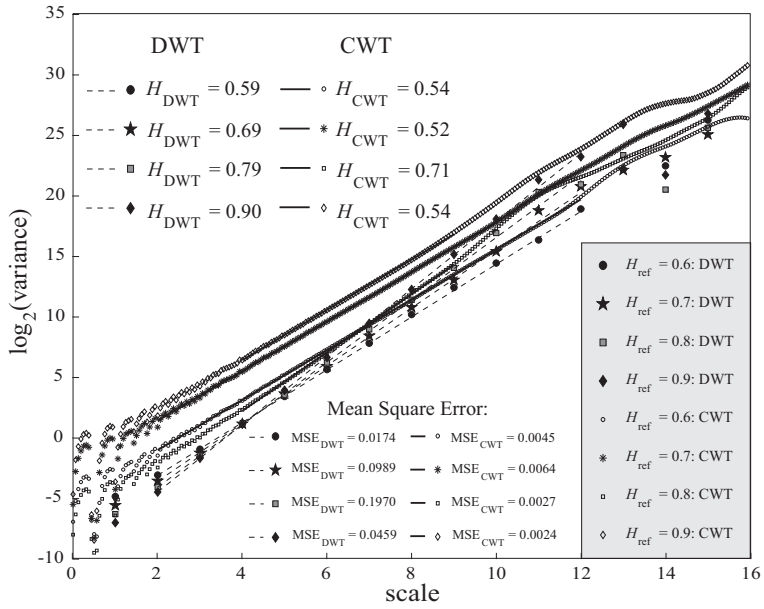

Fig. 7. Estimation of H with a continuous and a discrete `db2` wavelet transform for $H_{\text{ref}} \in \{0.6, 0.7, 0.8, 0.9\}$.

- a first one at small-time scales (1-3) where one measures $H = 0.71$ corresponding to the original fBm value $H = 0.7$ (see area 1 in Table 1),
- a second range varying from ~ 4 to the scale corresponding to the block size, where the behavior is the one of the classical Brownian motion $H = 0.5$ (see area 2 in Table 1),
- a last range of always three scales (see area 3 in Table 1), beginning at the block size scale, and where one measures

$$-0.25 \leqslant H_{\text{fGn}} = H_{\text{fBm}} - 1 \leqslant -0.40$$
$$\Rightarrow 0.60 \leqslant H \leqslant 0.75.$$

The value of H results of an *fGn behavior* but not any fGn: the one of *the increments of the fBm signal*. This is the indication of the persistence of H at long-time scales.

6. Conclusion

This study shows that discrete wavelets are good estimators of the Hurst exponent H and especially the 2-order of Daubechies wavelet. We have

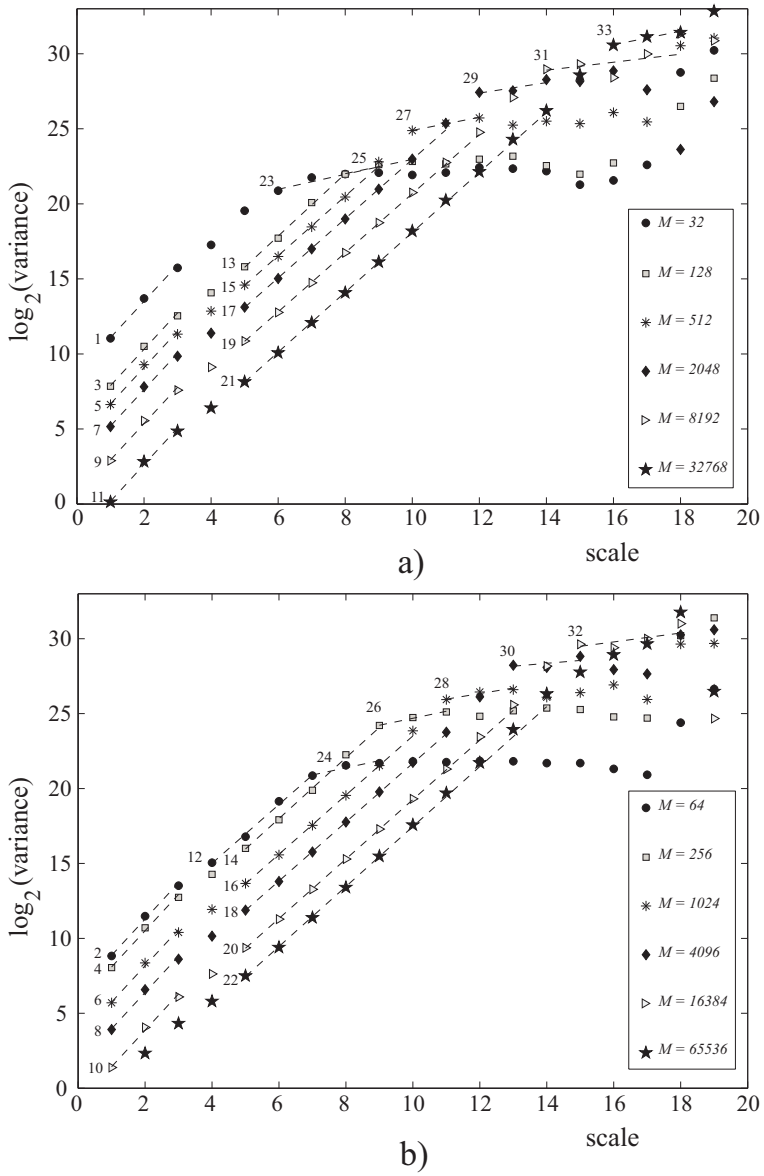

Fig. 8. DWT on scrambled fBm ($\log_2\left(\mathbb{E}\left[d_x^2(j)\right]\right)$ vs scale j) for $H = 0.7$ and for: a) $M \in \{2^5, 2^7, \ldots, 2^{15}\}$ and b) $M \in \{2^6, 2^8, \ldots, 2^{14}\}$.

shown that for different block sizes, and in spite of the shuffling process, the H value of the original fBm is still present not only at small-time scales, *i.e.* smaller than the block size, but also at long-time scales. This kind of process with mixed statistics seems to exchange its self-similarity, present at small-time scales for a stationarity at long-time scales and this, by holding the same value of H. This approach highlights the persistence level on the scale range for different block size values and reveals the reminiscent character of the fBm process behavior appearing from small-time scales to long-time scales. As a result, large value of H can be measured even without any long-range correlation and thus does not imply necessarily the existence of persistency at large time scales. This would put back into question the picture of a radial transport caused by avalanches at all scales. In order to confirm these first conclusions, the same work on other values of H is needed.

References

1. R. Jennane, R. Harba and G. Jacquet, *Traitement du Signal* **18**, 419 (2001).
2. A. Scipioni, P. Rischette, G. Bonhomme and P. Devynck, *Phys. Plasmas* **15**, p. 112303(november 2008), doi:10.1063/1.3006075.
3. P. Abry, P. Flandrin, M. Taqqu and D. Veitch, in *Theory and applications of long-range dependence*, (Birkhuser, 2003), ch. Self-similarity and long-range dependence through the wavelet lens, pp. 527–556.
4. D. Veitch and P. Abry, *IEEE Trans. on Info. Theory* **45**, 878 (1999).
5. B. B. Mandelbrot and J. W. van Ness, *SIAM Rev.* **10**, 422 (1968).
6. P. Flandrin, *IEEE Trans. on Info. Theory* **35**, 197 (1989).
7. F. Sellan, *CRAS* **321**, 351 (1995).
8. P. Abry and F. Sellan, *Applied and Computational Harmonic Analysis* **3**, 377 (1996).