

Wavelet-based estimators of scaling behavior

B. Audit, E. Bacry, J.F. Muzy and A. Arneodo

Abstract

Various wavelet-based estimators of self-similarity or long-range dependence scaling exponent are studied extensively. These estimators mainly include the (bi)orthogonal wavelet estimators and the Wavelet Transform Modulus Maxima (WTMM) estimator. This study focuses both on short and long time-series. In the framework of Fractional Auto-Regressive Integrated Moving Average (FARIMA) processes, we advocate the use of *approximately* adapted wavelet estimators. For these “ideal” processes, the scaling behavior actually extends down to the smallest scale, i.e., the sampling period of the time series, if an adapted decomposition is used. But in practical situations, there generally exists a cut-off scale below which the scaling behavior no longer holds. We test the robustness of the set of wavelet-based estimators with respect to that cut-off scale as well as to the specific density of the underlying law of the process. In all situations the WTMM estimator is shown to be the best or among the best estimators in terms of the mean square error. We also compare the wavelet estimators with the Detrended Fluctuation Analysis (DFA) estimator which was recently proved to be among the best estimators which are not wavelet-based estimators. The WTMM estimator turns out to be a very competitive estimator which can be further generalized to characterize multiscaling behavior.

Keywords

Scaling Phenomena, Self-similarity, Parameter estimation, Wavelet analysis, Fractional Brownian motion, Fractional ARIMA.

I. INTRODUCTION

Data displaying scaling behavior are observed in various fields. The velocity field of fully developed turbulent flows [1], financial time-series [2], [3], the telecommunication traffic load in high speed networks [4], medical time-series [5], [6], geological shapes [7] or the random walks associated to DNA sequences [8], [9] are only six out of numerous examples. There is no precise mathematical definition of what the notion of scaling behavior refers to. It is often associated with very different features such as long range dependence, self-similarity, multifractality or local regularity. However, from a general point of view, it basically corresponds to the fact that a certain scale-dependent quantity (e.g., the increments of a time-series, its power spectrum, ...) behaves as a power-law of the scale. Thus, characterizing the scaling behavior amounts to estimating some power-law exponents.

Many methods for estimating these exponents have been developed. By far, the most precise estimations are given by parametric methods such as the maximum likelihood estimator (the Whittle estimator) [10], [11], [12], [13]. However, in most practical situations, exact correlation structure of the data is unknown, these estimators should be handled very carefully, and certainly cannot be applied “blindly” for estimating any type of scaling behavior.

Most of the nonparametric estimators of scaling behavior exponents consist in performing a linear regression fit of a scale-dependent quantity versus the scale in a logarithmic representation. From a mathematical point of view, depending upon the range of scales over which power-law scaling is satisfied, different interpretations are given for the so-measured exponent [14]. For instance, if the power-law scaling extends to arbitrarily large scales, this exponent can be seen as a long-range dependence exponent [16] whereas if it extends to arbitrarily small scales it can be seen as a regularity (Hölder) exponent [17]. If the scaling extends from arbitrarily small to arbitrarily large scales, the exponent is a self-similarity exponent [18]. However, in practice, there is always a smallest scale (e.g., the sampling, the dissipation scale or some ultra-violet cut-off) and a largest scale (e.g., the total size of the signal, the integral scale or some infra-red cut-off) one cannot go beyond. Consequently, long-range dependence, self-similarity or regularity can generally be measured only on a given bounded range of scales with any nonparametric estimator whatever the particular theoretical framework it derives from.

The most common nonparametric estimators are based on fitting a power-law on the n -order (absolute) moment of the data values themselves or of their variations as a function of some scale/lag parameter. This includes the aggregated first and second-order methods [17], [19], the Higuchi method [20], the Detrended Fluctuation Analysis (DFA) [13], [21] and various wavelet-based estimators [14], [22], [23], [24], [25], [26]. Other commonly used estimators are the R/S method [27] and the estimators based on the periodogram [28]. Let us note that confidence intervals have only been obtained for the aggregated first and second-order estimators [17] and particular wavelet estimators [16]. Taqqu, Teverovsky and Willinger [13] have compared numerically most of these estimators except the ones based on wavelets. They have used them for estimating the scaling exponent of fractional Gaussian noises (fGn) or of Fractional ARIMA (FARIMA) processes. They have found that the “best” nonparametric estimator in terms of minimizing the Mean Squared Error (MSE) is the DFA estimator. This estimator has been initially introduced by Peng *et al.* [21] for characterizing long-range dependence properties of DNA sequences and has been justified theoretically by Taqqu *et al.* [13]. Not only is its MSE low but it is a robust estimator since it is blind to eventual superimposed linear trends.

B. Audit and A. Arneodo are in the Centre de Recherche Paul Pascal, Avenue Schweitzer, 33600 Pessac, France.

B. Audit current address is Computational Genomics Group, EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, United Kingdom

E. Bacry is in the Centre de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau Cedex France.

J.F. Muzy is in the UMR 6134, CNRS, Université de Corse, Grossetti, 20250 Corte, France

Wavelet-based estimators have been used very successfully, in very different contexts for estimating scaling behavior [26], [30], [31], [32], [33]. Since wavelets can be made orthogonal to low-order polynomial behavior [29], these estimators are blind to eventual superimposed polynomial trends. They have been used very successfully, in very different contexts for estimating scaling behavior [26], [30], [31], [32], [33]. Abry *et al.* [14] have shown that these estimators are also very robust when the shape of the underlying law (Gaussian, Exponential, LogNormal, α -stable, ...) is changed. However, Abry *et al.* [14] have been testing mostly estimators based on (bi)orthogonal wavelet bases [29], [34]. Other estimators based on the continuous wavelet transform can also be used [35], [36]. The Wavelet Transform Modulus Maxima (WTMM) method [22], [23], [26], [35], [37], [38] allows us to build an estimator that is based on the local maxima of the continuous wavelet transform. This method was proved very efficient to compute the *singularity spectrum* of multifractal signals. It is a very robust estimator that has been successfully applied to the analysis of various experimental signals such as DNA sequences [9], [26], [32], [39], fully developed turbulence data [22], [33], [35], [38], [40], [41], financial time-series [42], diffusion limited aggregates [43], [44], [45] and high resolution satellite images of fractal clouds [32], [46]. However, the WTMM method has never been studied from a statistical point of view. One of the goals of this work is to compare it to other estimators in the simplest situation when the scaling properties of the studied process are characterized by a single exponent H . Comparison in terms of MSE with “standard” wavelet estimators [14] and DFA estimator [13], [21] will be made. We will focus our study on both short and long time-series and study the influence of an eventual “dissipation scale”, i.e., a small scale below which the power law scaling does not hold anymore. Our second purpose is to compare all these estimators in “practical” situations when, e.g., the size of the series is rather small, the statistics are non-Gaussian or the scaling does not hold on the full range of scales. Let us note that, to our knowledge, this paper is the first systematic study on the influence of the presence of some ultra-violet cut-off or of the fact that the time-series may be dramatically short, on the robustness of nonparametric scaling behavior estimators. As we will show in this paper, its robustness makes the WTMM estimator a very good candidate for analyzing any kind of real data without any prior assumption.

The paper is organized as follows. Section II presents the different models of scaling processes that will be used all along the paper. Section III focuses on the wavelet transform [29] with special emphasis on how it applies to the study of scaling processes. Both the continuous wavelet transform and the (bi)orthogonal wavelet bases are discussed. The different wavelet based estimators using (bi)orthogonal wavelet bases are introduced in Section IV. Section V is devoted to the WTMM estimator. The statistical properties of all these estimators are analyzed both from a theoretical point of view and from a numerical point of view. We study the influence of both the size of the signal and the range of scales that is used for the linear regression fit on each of these estimators. Section VI compare all these estimators when applied to estimate the exponent of a FARIMA(0, D , 0) process. In Section VII, we study the performances of the same estimators on two other processes: (i) the FARIMA(0, D , Q) and (ii) the two-valued FARIMA model (introduced in Section II) with the specific goal to test the robustness with respect (i) to the possible existence of an ultra-violet cut-off scale and (ii) to the shape of the underlying probability law. We finally conclude in Section VIII with some perspectives about the generalization of the WTMM estimator to characterize the multiscaling behavior of multifractal processes.

II. MODELS OF SCALING PROCESSES

A. The fractional Brownian motion

The most popular model that displays scaling behavior is certainly the fractional Brownian motion (fBm) [18] introduced by Mandelbrot and Van Ness in 1968 [51]. It is a family of Gaussian processes $\{B_H(t), t \geq 0\}$ indexed by a single parameter H ($0 < H < 1$) with mean 0, stationary increments and $B_H(0) = 0$. Its covariance is given by

$$E(B_H(t)B_H(s)) = \frac{\sigma^2}{2} (s^{2H} + t^{2H} - |s - t|^{2H}). \quad (1)$$

Let us note that for $H = 1/2$, $B_{1/2}(t)$ corresponds to the *classical* Brownian motion with independent increments. Moreover, it can be shown that the process $B_H(t)$ can be obtained by fractional integration of order $H + 1/2$ of a white noise [18].

Fractional Brownian motions display scaling behavior in the *wide sense*, i.e., all the features that we have mentioned in the introduction are present in the process. Indeed, one can prove [18] that the increments display long-range dependence (for $H \neq 1/2$, their autocorrelation function decays as l^{2H-2} as $l \rightarrow +\infty$), that it is a self-similar process in the sense that $B_H(\lambda t) = \lambda^H B_H(t)$ in law for all $\lambda \in \mathbb{R}^{+*}$, that it is monofractal (i.e., its singularity spectrum is reduced to a single point) and that the Hölder exponent at any positive time t of almost all the realizations of B_H is H . From a theoretical point of view, one can clearly state what the differences between all these features are. However, as we have already explained in the introduction, in practical situations, the scaling behavior appears only on a finite range of scales and consequently, these features cannot generally be distinguished.

B. The fractional ARIMA (FARIMA) process

To use fractional Brownian motions in numerical experiments, we need to sample it. It has been shown by Meyer, Sellan and Taqqu [50] that a Fractional Auto-Regressive Integrated Moving Average (FARIMA) time-series with Gaussian innovations [18], [47], [48] corresponds to such a sampling (using a particular filter). FARIMA have been chosen for this work because these discrete processes have a strong relationship to particular biorthogonal wavelet decomposition of the fBm (see Section III-A.3) which will enable us to perform simulations in the exact same situation as those assumed from our theoretical calculations of Section IV.

An ARIMA(P, Q) corresponds to the primitive of a regular ARMA(P, Q) time-series. In the case of FARIMA(P, D, Q), this integration is replaced by a fractional integration of order D , i.e., a convolution with the filter h_D whose z -transform reads:

$$\hat{h}_D(z) = (1 - z)^{-D}. \quad (2)$$

A sampled fBm $b_H(k)$ correspond a FARIMA(P, D, Q) using Gaussian innovations with $D = H + 1/2$ and without using any autoregressive ($P = 0$) nor moving average ($Q = 0$) terms [34], [49], [50]. To assure the corresponding convolution converges, it is necessary to impose that $b_H(0) = 0$. As a consequence a sampled fBm is obtained by simply integrating a FARIMA($0, D = H + 1/2, 0$) and corresponds to the primitive of a simple fractional integration of a Gaussian white noise [18]:

$$b_H(k) = \sum_{m=0}^k h_{H-1/2} * \epsilon(m), \quad (3)$$

where the $\epsilon(m)$'s are i.i.d. Gaussian variables of mean 0 and variance σ^2 . That is, $b_H(k)$ is a FARIMA($0, D = H + 1/2, 0$) forced to 0 for $k = 0$ ¹. In the following, most of the time, we will refer to $b_H(k)$ as a FARIMA time-series of exponent $D = H + 1/2$ using Gaussian innovations. The cases of non-Gaussian (log-normal or exponential) innovations will be also studied in Section IV.

As it was pointed out in the introduction, in practical situations, there exists generally a small scale cut-off (e.g., a dissipation scale) that defines a lower bound to the scaling range. In order to test the robustness of these estimators to the presence of such an ultra-violet cut-off, one can work with FARIMA($0, D, Q$) processes [26].

C. The FARIMA($0, D, Q$) process

This model corresponds to the same model as before but with an additional moving average term of order Q , i.e.,

$$sb_H(k) = w * b_H(k), \quad (4)$$

where w is a compact support smoothing filter of length Q . We arbitrarily choose w to be a single arch of a sinusoid of amplitude 1. In Section VII-A, we will use this process in order to test the robustness of the different estimators with respect to the range of scales used in the linear regression fit procedure.

D. The two-valued FARIMA($0, D, 0$) process

Abry *et al.* [14] have shown that orthogonal wavelet-based estimators are very robust when the underlying law is changed (e.g., Gaussian, Exponential, LogNormal,...). In order to study how, in that respect, these estimators compare to the DFA estimator [13], [21] and to nonorthogonal wavelet-based estimators (mainly the WTMM estimator [26], [35]), we introduce here a new model that displays long-range dependence and which is *very far* from being Gaussian since its increments can only take two values: -1 or 1.

This model is well adapted to describe the scaling behavior of DNA walks that are generated from some binary coding of DNA sequences [8], [9], [21], [26]. It simply consists in replacing the successive increments of a FARIMA($0, D, 0$) $b_H(k)$ by their signs [26]:

$$s_H(k) = \sum_{m=0}^k \text{sign}(b_H(m+1) - b_H(m)), \quad (5)$$

where the $\text{sign}(x)$ function is 1 for $x \geq 0$ and -1 otherwise.

One can easily show that if X and Y are two Gaussian variables of mean 0, variance 1 and covariance $\rho < 1$, then the variables $\text{sign}(X)$ and $\text{sign}(Y)$ have a covariance whose first order is of the order of ρ . Thus, since

$$E((b_H(k+1) - b_H(k))(b_H(k+1+l) - b_H(k+l))) \sim l^{2H-2}, \quad (6)$$

for $l \rightarrow +\infty$, we deduce that

$$E((s_H(k+1) - s_H(k))(s_H(k+1+l) - s_H(k+l))) \sim l^{2H-2}, \quad (7)$$

¹Note that for $H < 1$ the convolution $h_{H-1/2} * \epsilon(m) = \sum_{i=0}^{\infty} h(i)\epsilon(m-i)$ is convergent and can be computed to an arbitrary precision by a simple truncation of the discrete sum.

and thus the increments of $s_H(k)$ display the same second-order long-range dependence than the increments of $b_H(k)$. Consequently, although (i) the underlying law is no longer Gaussian and (ii) the range of scales on which scaling operates is truncated from below, the two-valued FARIMA(0, D , 0) process displays the same scaling behavior exponent as the regular FARIMA process. In Section VII-B, we will test the robustness of all the wavelet-based estimators and the DFA estimator for estimating the scaling exponent H of this newly defined process that provides a very attractive model to mimic DNA walks [9], [26].

III. WAVELET TRANSFORM OF SCALING PROCESSES

A. The wavelet transform

A.1 The continuous wavelet transform

The continuous wavelet transform [29], [52], [53], [54] of a given function $f(t)$ is defined as the scalar product of f with a family of test functions $\psi_{b,a}(t)$, i.e.,

$$T_\psi[f](b, a) = \langle f, \psi_{b,a} \rangle. \quad (8)$$

Each test function $\psi_{b,a}$ is derived from a single function $\psi(t)$, referred to as the *analyzing wavelet*, by means of a translation and a dilation:

$$\psi_{b,a}(t) = \frac{1}{a} \psi\left(\frac{t-b}{a}\right), \quad (9)$$

where $b \in \mathbb{R}$ and $a \in \mathbb{R}^{+*}$. The analyzing wavelet $\psi(t)$ is chosen to be well localized in time around $t = 0$ and such that its spread in Fourier space is almost limited. Moreover ψ is required to satisfy the *admissibility condition* which in its weak form implies that ψ must be of zero mean. Actually, we will often choose ψ so that its first n_ψ moments are 0, i.e.,

$$\int t^k \psi(t) dt = 0, \quad \text{for } 0 \leq k < n_\psi. \quad (10)$$

A very common way to build admissible wavelets of arbitrary order n (i.e., with $n_\psi = n$) is to successively differentiate a smoothing function. Thus if we use the Gaussian function as the smoothing function, we get a very popular family of wavelets [29], [30], [53], [54]:

$$g_n(t) = g^{(n)}(t) = \frac{d^n}{dt^n} e^{-t^2/2}. \quad (11)$$

In order to compute the wavelet transform of a time-series $\{f(k)\}_{k \in \mathbb{Z}}$, one needs to discretize the scalar product expression (8). Moreover both parameters b and a must be discretized. Generally b is discretized using the discretization step of the time-series (here, we will arbitrarily choose the sampling rate equal to 1) and a is discretized on n_a scales chosen on a uniform logarithmic grid $\{a_j = \lambda^{j-1} a_{min}\}_{1 \leq j \leq n_a}$ (with $\lambda > 1$). Since, for a fixed scale a , the continuous wavelet transform can be seen as a simple convolution, the algorithm for computing the continuous wavelet transform on a time-series of size N is very fast and of the order of $n_a N \log_2 N$.

A.2 The orthogonal wavelet transform

In order to reduce the redundancy of the continuous wavelet transform, one can subsample it. Actually, one can show that if ψ is properly chosen, then the family $\{2^{j/2} \psi_{j,k}\}_{j,k \in \mathbb{Z}}$, with

$$\psi_{j,k}(t) = \psi_{b_{j,k}, a_j}(t) = 2^{-j} \psi(2^{-j} t - k), \quad \text{with } a_j = 2^j, b_{j,k} = 2^j k, \quad (12)$$

is an orthonormal basis of L^2 [29]. Let us note that the $2^{j/2}$ term in $\{2^{j/2} \psi_{j,k}\}_{j,k \in \mathbb{Z}}$ is just a normalization factor in order the basis to be normalized. The orthogonal wavelet coefficients can then be defined by

$$d_f(j, k) = \langle f, \psi_{j,k} \rangle. \quad (13)$$

A very popular class of orthogonal wavelets is the one introduced by Daubechies [56]. It is a class of compact support wavelets indexed by a single parameter which corresponds to the order of the wavelet. In the following, the wavelet d_n will refer to the Daubechies wavelet of order n .

In the same spirit as for the continuous wavelet transform, when dealing with discretized time-series, the scale parameter must be chosen greater than the sampling rate 1. Moreover it cannot go beyond the total length N of the time-series. Thus the available scales are restricted to $1 < a \leq N$, i.e., $J_{min} = 1 \leq j \leq J_{max} = \lfloor \log_2(N) \rfloor$. The algorithm which computes the orthogonal wavelet coefficients of a time-series of length N is very fast and its complexity is of the order of N [29].

Let us note that the rigorous definition of the so-obtained wavelet coefficients is done in the framework of a *Multiresolution Analysis* [29], [57]. We will not detail here what a Multiresolution Analysis is. One just needs to know that in this framework, the samples $f(k)$ are seen as samples of a continuous time function $f(t)$ using a particular sampling filter generally referred to as the *scaling function* $\phi(t)$ associated to the wavelet $\psi(t)$. In the case of a FARIMA(0, D , 0) (with $D = H + 1/2$), $b_H(k)$ is simply seen as the samples of a fBm $B_H(t)$ [50].

A.3 The biorthogonal wavelet bases

One can build biorthogonal wavelet bases, using 2 wavelet families instead of one [29]. One $\{2^{j/2}\psi_{j,k}^{(d)}\}_{j,k \in \mathbb{Z}}$ is used for the decomposition

$$d_f(j, k) = \langle f, \psi_{j,k}^{(d)} \rangle, \quad (14)$$

and the other one $\{2^{j/2}\psi_{j,k}^{(r)}\}_{j,k \in \mathbb{Z}}$ is used for the reconstruction $f(t) = \sum_{j,k \in \mathbb{Z}} 2^j d_f(j, k) \tilde{\psi}_{j,k}^{(r)}(t)$.

Commonly used biorthogonal wavelets are built by fractionally integrating and differentiating orthogonal wavelets. Indeed, if ψ corresponds to an orthogonal wavelet, we set $\psi^{(d)}(t) = \psi^{(H+1/2)}(t)$ (i.e., the fractional derivative of order $(H + 1/2)$ of ψ) and $\psi^{(r)}(t) = \psi^{-(H+1/2)}(t)$ (i.e., the fractional primitive of order $(H + 1/2)$ of ψ). These biorthogonal bases play a particular role when analyzing an fBm. Indeed, since the fBm $B_H(t)$ of parameter H is basically defined as the fractional primitive of order $H + 1/2$ of a Gaussian white noise $W(t)$, one can easily show that:

$$d_{B_H(t)}(j, k) = \langle B_H, \psi_{j,k}^{(H+1/2)} \rangle = 2^{jH} \langle W(t), \psi_{j,k} \rangle. \quad (15)$$

Eq. (15) shows that the wavelet coefficients $d_{B_H(t)}(j, k)$ of the fBm $B_H(t)$ are exactly scaling and, since $\{2^{j/2}\psi_{j,k}\}$ is an orthonormal basis of L^2 , are decorrelated Gaussian variables. For this two reasons, we will say that this biorthogonal wavelet basis is *adapted* to $B_H(t)$.

Since one can build adapted biorthogonal basis from any orthogonal basis, in this paper, we will use the biorthogonal bases derived from the Daubechies orthogonal bases. We will note $\text{dnh}H$ the biorthogonal wavelets derived from the Daubechies wavelet dn and adapted to the value H . Thus for instance d9h7 will refer to the biorthogonal wavelets derived from the orthogonal wavelet d9 and adapted to the value $H = 0.7$. Let us point out that using d9h5 amounts to use d9 on the derivative of the analyzed signal. One recovers the fact that d9 (as any orthogonal basis) is adapted to white noise.

From a numerical point of view, it is important to note that in order to use the pyramidal algorithm to compute the (bi)orthogonal decomposition of a continuous-time signal, one needs to apply it to a particular sampling (depending on the wavelet used) of this signal [29]. In the case of fBm and adapted biorthogonal decomposition, this sampling corresponds to the FARIMA(0,D,0) processes. That is adapted biorthogonal wavelets are only adapted to this particular sampling of the fBm. The implication of this assumption will be discussed in Section VI.

B. Wavelet analysis of scaling processes

The wavelet transform is a particularly well adapted tool for studying scaling processes. Indeed, one can prove [14], [15], that the wavelet transforms of scaling processes with stationary increments share some statistical properties. Mainly, one can prove that if the stochastic process $f(t)$ is H self-similar or if it has some long-range dependence properties characterized by the exponent H then both the continuous and the (bi)orthogonal wavelet transform of $f(t)$ satisfy the 3 following properties:

- **(P1)** the wavelet coefficients at a fixed scale form a stationary process, i.e., both $\{T_\psi[f](b, a)\}_{b \in \mathbb{R}}$ (fixed a) and $\{d_{j,k}\}_{k \in \mathbb{Z}}$ (fixed j) are stationary.
- **(P2)** The wavelet coefficients have short-range correlations. Moreover, the greater the number n_ψ of vanishing moments of ψ (Eq. (10)) the shorter the correlations. In the case of the continuous wavelet transform, it translates into

$$E(T_\psi[f](b, a)T_\psi[f](b', a')) \sim |b - b'|^{2H-2n_\psi}, \quad (16)$$

for $|b - b'| \rightarrow +\infty$ and $a, a' \gg 1$. In the case of (bi)orthogonal wavelet transform, it translates into

$$E(d_f(j, k)d_f(j', k')) \sim |2^j k - 2^{j'} k'|^{2H-2n_\psi}, \quad (17)$$

for $|2^j k - 2^{j'} k'| \rightarrow +\infty$ and $j, j' \gg 0$. From a numerical point of view, it has been shown [14] that as long as n_ψ is large enough the wavelet coefficients $\{d_{j,k}\}_{k \in \mathbb{Z}}$ can be considered as decorrelated.

- **(P3)** the q -order moment $M_q(a)$ of the wavelet coefficients at scale a reproduces the scaling property of the process. More precisely, one can prove that (in the case of continuous wavelet transform)

$$M_q(a) = E(|T_\psi[f](b, a)|^q) \sim a^{qH}, \quad \text{for } a \gg 1. \quad (18)$$

or (in the case of (bi)orthogonal wavelet transform)

$$M_q(a = 2^j) = E(|d_f(j, k)|^q) \sim a^{qH}, \quad \text{for } j \gg 0. \quad (19)$$

Thus, according to the last property, the parameter H can be estimated by simply performing a linear regression fit of the logarithm of any q moment of the wavelet coefficients versus the logarithm of the scale, as long as the fit is performed on scales much greater than 1. The fact that the coefficients are stationary and “almost” decorrelated will allow us to estimate the q -order moment of a wavelet coefficient at scale a by just averaging wavelet coefficients at scale a of a given realization of the process. This is the purpose of the next section.

IV. WAVELET-BASED ESTIMATION OF THE SCALING EXPONENT USING (BI)ORTHOGONAL WAVELET BASED ESTIMATORS

A. Basis of the estimation

As we have already pointed out, the q -order moment $M_q(a)$ of the wavelet coefficients at scale a of a scaling process scales like $M_q(a) \sim a^{qH}$. This is indeed the case for the continuous wavelet coefficients (Eq. (18)) and the (bi)orthogonal coefficients (Eq. (19)). Thus, a very common way to estimate H is to build an estimator of $\ln(M_q(a))/q$ and then perform a linear fit of this estimator versus the logarithm of the scale [14], [26], [35], [38]. Following the lines of [14], in order to enable analytical computations, we will consider that n_ψ is large enough so that (using properties (P1), (P2) and (P3) of Section III-B), one can reasonably (from a numerical point of view) make the assumptions that:

- **(H1)** $\{d_{j,k}\}_{k \in \mathbb{Z}}$ (fixed j) and $\{T_\psi[f](b, a)\}_{b \in \mathbb{R}}$ (fixed a) are stationary,
- **(H2)** $\{d(j, k)\}_{j,k}$ are independant random variables and
- **(H3)** $M_q(a) = C_q a^{qH}$ for a certain range of scale $2^{J_0} \leq a \leq 2^J$.

These assumptions will be made all along Section IV.

Let us note that Abry *et al.* have already studied this estimator in the case of (bi)orthogonal bases [14], [15]. However, these authors did not tackle several problems that we will address here. Mainly they have only studied the case of very long time-series ($N > 10000$ samples). In that case, they have proved that the estimators can be considered as unbiased estimators and that the optimal moment for estimating H is the variance ($q = 2$). As we will see, this is no longer true neither for short time-series due to nonnegligible bias nor for non-Gaussian distributions (such as the WTMM distribution). Moreover, all along this paper, we will study the influence of the largest and smallest scale involved in the linear regression fit procedure on the MSE of the estimator.

Before moving on, let us point out that one could use an alternative wavelet-based estimator that the one suggested here [26], [35], [38]. Indeed, it would consist in building an estimator of $h_q(a) = \partial \ln M_q(a) / \partial q$ and then performing a linear fit of this quantity versus the logarithm of the scale. This estimator has been studied extensively in Ref. [26] in the (bi)orthogonal and WTMM cases. It has been shown that the $h_q(a)$ -based estimator is far more stable than the standard $M_q(a)$ estimator when dealing with multifractal time-series (i.e., involving several scaling exponents depending on q). However, since in this paper we only focus on monofractal time-series (i.e., involving a single scaling exponent) we will only consider $M_q(a)$ -based estimators.

B. An estimation of $\ln(M_q(a))/q$.

B.1 Analytical computations.

In this Section, we assume the hypothesis (H1), (H2) and (H3) introduced in Section IV-A. Thus, a natural estimator of $\ln(M_q(a))/q$ is simply given by

$$\mathcal{L}_{N,q} = \frac{1}{q} \ln \left(\sum_{k=1}^N |d(j, k)|^q / N \right). \quad (20)$$

Since the $d(j, k)$ are i.i.d. variables, using the Corollary (1) in Appendix A, one gets an expansion (for N large enough) of the bias:

$$B(\mathcal{L}_{N,q}) = -\frac{\Lambda_2(q)}{2qN} + \frac{1}{N^2} \left(\frac{\Lambda_3(q)}{3q} - \frac{3\Lambda_2(q)}{4q} \right) + O\left(\frac{1}{N^3}\right), \quad (21)$$

and of the variance

$$Var(\mathcal{L}_{N,q}) = \frac{\Lambda_2(q)}{q^2 N} + \frac{1}{q^2 N^2} \left(\frac{5}{2} \Lambda_2^2(q) - \Lambda_3(q) \right) + O\left(\frac{1}{N^3}\right), \quad (22)$$

where

$$\Lambda_2(q) = \frac{Var(|d(j, k)|^q)}{E^2(|d(j, k)|^q)}, \quad (23)$$

and

$$\Lambda_3(q) = \frac{M_3(|d(j, k)|^q)}{E^3(|d(j, k)|^q)}, \quad (24)$$

where $M_3(|d(j, k)|^q)$ is the centered moment of order 3 of $|d(j, k)|^q$

$$\rangle. \quad (25)$$

From Eqs (21) and (22), one gets the expression of the MSE

$$\begin{aligned} e_{N,q}^2 &= B(\mathcal{L}_{N,q})^2 + Var(\mathcal{L}_{N,q}) \\ &= \frac{\Lambda_2(q)}{q^2 N} + \frac{1}{N^2} \left(\frac{11\Lambda_2^2(q)}{4q^2} - \frac{\Lambda_3(q)}{q^2} \right) + O\left(\frac{1}{N^3}\right). \end{aligned} \quad (26)$$

Thus, as already shown by Abry *et al.* [14], this estimator is asymptotically unbiased and consistent. Moreover for large N , $B(\mathcal{L}_{N,q})^2 \ll \text{Var}(\mathcal{L}_{N,q}) \sim 1/N$. Thus, in this limit, the MSE is dominated by the variance and consequently is of the order of $1/N$:

$$e_{N,q}^2 = \frac{\Lambda_2(q)}{q^2 N} + O\left(\frac{1}{N^2}\right). \quad (27)$$

B.2 Numerical simulations.

For long time-series, it follows from Eq. (27) that the optimum value $q = q^*$ which minimizes $e_{N,q}^2$ does not depend on N and cancels the derivative of $\Lambda_2(q)/q^2$. This value strongly depends on the probability distribution law of the wavelet coefficients [26]. Straightforward calculations show that when wavelet coefficients are Gaussian ² $q^* = 2$ and $\Lambda_2(q^*)/q^{*2} = 1/2$, whereas for log-normal coefficients ³ $q^* = 0$ and $\Lambda_2(q^*)/q^{*2} = 1$, while for exponential distribution $q^* = 1$ and $\Lambda_2(q^*)/q^{*2} = 1$. (Section VI-B.2 shows that $q^* \simeq 2.5$ when using WTMM). In figure 1b we compare $\Lambda_2(q)/q^2$ for these three shapes of wavelet coefficient distribution. In order to be able to compare these curves to each other, we have plotted $\frac{1}{q^2}\Lambda_2(q)/\frac{1}{q^{*2}}\Lambda_2(q^*)$ and the curves have been translated horizontally so that the minimum is reached at the same abscissa (0). It illustrates that choosing the optimum $q = q^*$ for a given law instead of $q = 2$ as suggested by the Gaussian case, leads to significant improvement of the MSE. For the exponential law, using $q = 2$ instead of $q = q^* = 1$ leads to a 25% increase of the variance. For a log-normal distribution of the form $\exp(\mathcal{N}(m, \sigma))$ with $\sigma = 1$, using $q = 2$ instead of $q = q^* = 0$ leads to a 10 fold increase of the variance.

However, for short time-series (i.e., small N), Eq. (27) is no longer sufficient to account for the actual MSE and Eq. (26) must be used. In the following we quantify this effect in the Gaussian case only. On figure 1a, we have superposed for $N = 32$, the variance (---) and the MSE (---) computed with the first two terms of the expansions (22) and (26), with the variance and MSE (—) computed with their common first term only. One can see that the second-order terms are not small. The actual MSE is greater (4.6%) than the MSE computed with the first-order terms. Moreover the MSE is clearly greater than the variance which indicates that the bias is not negligible. As a result, the optimal value for q is slightly smaller than 2. To validate these results, numerical experiments have been performed replacing the $d(j, k)$ in Eq. (20) by a Gaussian noise to ensure that hypothesis (H1) and (H2) hold. For $N = 32$ and $\sigma = 10\,000$, the numerical results presented in figure 1a for the variance (\triangle) and the MSE (\blacktriangle) match perfectly the theoretical curves obtained with the first two terms of the expansions (22) and (26). In figure 2, for $N \in [8, 2048]$ and $q = 2$, we compare the theoretical estimations of the MSE using the first or the first two terms of Eq. (26) to numerical experiment with $\sigma = 1$ or $10\,000$. Again, it appears clearly that for small N ($N \leq 64$), the first-order term in (26) is not enough for describing the MSE. One needs to take into account the second-order term. Figure 1 and 2 demonstrate that down to very small sizes ($N < 8$) it is sufficient to use these two terms to get an accurate estimation of the variance and the MSE (in the same manner, it can be shown that the bias is well estimated using the first term only in Eq. (21)).

As we have already said, Abry *et al.* [24] have only computed the first-order term in Eq. (26) and they have concluded that, in the Gaussian case, for N “large enough”, $q = 2$ is optimal, the estimator is unbiased and the Cramér-Rao bound is reached. Let us note that in [16] Abry *et al.* obtained exact expressions (for any N) for the bias and the variance in the case $q = 2$. Since asymptotically (for large N) the Cramér-Rao bound is reached, it implicitly proves that $q = 2$ is asymptotically optimal. As we have seen, for small signals, this lead to pick up a nonoptimal q and to underestimate the MSE. Picking up a nonoptimal q is not dramatic since the MSE does not change much when q is varied around 2. However, for short time-series, whatever value of q is used, it is very important to compute the MSE not forgetting the second-order term since it is almost of the same order as the first one.

C. Estimation of H

A linear regression of the estimator \mathcal{L} as a function of $\ln a$ leads to the estimation of H . We will use the standard linear regression estimator which minimizes the MSE. As already explained, in the case of the (bi)orthogonal wavelet bases, the scale parameter a is sampled as $a_j = 2^j$ with $J_{min} = 1 \leq j \leq J_{max} = \lfloor \log_2 N \rfloor$. Moreover the number of coefficients at scale a_j is $N_j = N2^{-j}$ where N is the length of the initial time-series. We will perform the regression over the range of scales corresponding to $j \in [J_m, J_M]$ where $J_{min} = 1 \leq J_m < J_M \leq J_{max} = \lfloor \log_2 N \rfloor$.

We are looking for $\hat{H}_{J_m, J_M, q}$ and \hat{C} which minimize

$$(\hat{H}_{J_m, J_M, q}, \hat{C}) = \underset{(h, c)}{\operatorname{argmin}} \sum_{j=J_m}^{J_M} p_j E((\mathcal{L}_{N_j, q} - h \ln a_j - c)^2), \quad (28)$$

²Note that Eq. (27) thus becomes $e_{N,q^*}^2 \simeq \frac{1}{2N}$, which corresponds to the Cramér-Rao lower bound for the estimation of the logarithm of the standard deviation of a centered Gaussian distribution [24], [26].

³ $q = 0$ actually means replacing Eq. (20) by the mean of the logarithm of the wavelet coefficients.

where p_j is the weight given to the j th point. It is natural to choose the p_j so that to give more weight to the points which correspond to the smallest variance, i.e., $p_j \sim \text{Var}(\mathcal{L}_{N_j,q})^{-1}$. This leads to the following estimator:

$$\hat{H}_{J_m, J_M, q} = \frac{\sum_j p_j \sum_j p_j \ln(a_j) \mathcal{L}_{N_j, q} - \sum_j p_j \ln a_j \sum_j p_j \mathcal{L}_{N_j, q}}{\sum_j p_j \sum_j p_j \ln^2 a_j - \left(\sum_j p_j \ln a_j \right)^2}. \quad (29)$$

From Eq.(21), we know that the bias of $\mathcal{L}_{N_j, q}$ is of the form:

$$B_{N_j, q} = -\frac{\Delta_1(q)}{N_j} + O\left(\frac{1}{N_j^2}\right), \quad (30)$$

and from Eq. (22) that the variance can be written as

$$\text{Var}(\mathcal{L}_{N_j, q}) = \frac{\Xi_1(q)}{N_j} + \frac{\Xi_2(q)}{N_j^2} + O\left(\frac{1}{N_j^3}\right). \quad (31)$$

For the sake of simplicity, in order not to introduce q -dependent terms in the weights p_j , we decide to follow Abry *et al.* [24] by choosing $p_j = N_j$ (though, of course, it does not correspond to the optimal choice). Then, using the decorrelation hypothesis of the wavelet coefficients at different scales, long but straightforward computations lead to the following expressions for the bias and the MSE of $\hat{H}_{J_m, J_M, q}$ [26]:

$$B(\hat{H}_{J_m, J_M, q}) = -\frac{\Delta_1(q) D_1(J_M - J_m + 1)}{2N 2^{-(J_m-1)} \ln 2} + O\left(\frac{1}{N^2}\right), \quad (32)$$

$$\begin{aligned} e^2(\hat{H}_{J_m, J_M, q}) &= \frac{\Xi_1(q) X_1(J_M - J_m + 1)}{2^{-(J_m-1)} N \ln^2 2} + \frac{\Xi_2(q) X_2(J_M - J_m + 1) + \Delta_1(q)^2 D_1^2(J_M - J_m + 1)/4}{2^{-2(J_m-1)} N^2 \ln^2 2} \\ &\quad + O\left(\frac{1}{N^3}\right), \end{aligned} \quad (33)$$

where the J_m, J_M dependent terms can be found in [26]. Thus, the so-obtained estimators $\hat{H}_{J_m, J_M, q}$ of H are asymptotically unbiased and consistent (let us note that neither the bias nor the MSE depends on H). We must pick up the optimal estimator, i.e., the optimal J_m, J_M and q which minimize the MSE. One can show very easily that if we take into account in Eq. (33) only the first-order term, then the optimal value for q is q^* , i.e., the value minimising the first term of $e_{N,q}^2$, and the optimal values for J_m and J_M are $J_m^* = J_{min} = 1$ and $J_M^* = J_{max} = \lceil \log_2 N \rceil$.

Now, if the second-order term in Eq. (33) is not negligible then the optimal values of the fitting parameters are modified. In the Gaussian case, the optimal value for J_m remains 1 (in the following, we will write $H_{J_M, q}$ instead of $H_{J_m, J_M, q}$ when $J_m = 1$) but, as illustrated in figures 3b and 3c, the optimal value for q (resp. J_M) is no longer 2 (resp. J_{max}). Actually, as N goes to infinity, though the optimal q goes to $q^* = 2$, the optimal J_M is always strictly smaller than J_{max} . Thus, the best estimator is not obtained using the largest range of scales. However, as shown in figure 3a, the MSE using the optimal values q^* and J^* is just a few percent smaller than the MSE using $q = 2$ and $J_M = J_{max}$. Indeed, for N ranging in [32, 1024], the gain that is made when using optimal values ranges between 6% to 8%. Therefore, for the sake of simplicity and without hardly any loss, we advocate the use of the nonoptimal parameter values $\bar{q} = 2$ and $\bar{J}_M = J_{max}$. Note that the estimation of the MSE using only the first-order term in Eq. (33), as done in Abry *et al.* [24], can lead to drastic errors. Indeed, for N ranging in [32, 1024], the first order term alone underestimates the error by 19% to 39%. As a consequence, the complete expression (33) has to be used to obtain an accurate estimation of the MSE.

V. WAVELET-BASED ESTIMATION OF THE SCALING EXPONENT USING THE WTMM ESTIMATOR

A. Definition

The WTMM estimator has been introduced in 1989 by Arneodo *et al.* [22], [23], [35], [37], [38]. Since, it has proved successful in very various domains ranging from the study of DNA sequences [9], [26], [32], [39], of turbulent flows [22], [33], [35], [38], [40], [41], financial time-series [42], cloud images [32], [46]. It is based on the local maxima of the modulus of the continuous wavelet transform, i.e., on the local maxima of the function $b \rightarrow |T_\psi[f](b, a)|$, where a is a fixed scale. One generally chooses to sample the scales using a uniform logarithmic scale $a_j = \lambda^{j-1} a_{min}$ (with $\lambda > 1$) and the b parameter is generally sampled using the sampling rate of the analyzed time-series. Let $x_{j,k}$ be the k th local maximum which appears at scale a_j . We are going to decrease the redundancy of the continuous wavelet transform by just keeping

the positions and the values of the wavelet transform at the local maxima. Let us notice that, in practical situations, one can recover the original time-series using only the information contained in the WTMM [29].

With some very basic assumptions, one can prove that the density of the local maxima at scale a_j is proportional to a_j^{-1} (see Eq. (43)). Thus if $\lambda = 2$ (i.e., $a_j = 2^{j-1}a_{min}$), the density at the different scales mimic the density of the dyadic grid of a (bi)orthogonal wavelet transform. Even though two different local maxima are not decorrelated, it seems natural to use the same estimators as we did for the (bi)orthogonal bases. Thus the WTMM estimator consists first in estimating at each scale a_j the $\ln(M_q(a_j))/q$ using the $\mathcal{L}_{N_j,q}$ estimator, i.e.,

$$\mathcal{L}_{N_j,q} = \frac{1}{q} \ln \left(\sum_{k=1}^{N_j} |T_\psi[f](x_{j,k}, a_j)|^q / N_j \right), \quad (34)$$

and secondly in performing the linear regression defined in Eq. (29) ⁴.

B. Analytical computation

The estimation of $\ln(M_q(a_j))/q$ is made using the estimator $\mathcal{L}_{N_j,q}$ defined by Eq. (34). Since it involves local maxima, this estimator is much more complex (from an analytical point of view) than the estimators based on (bi)orthogonal wavelet transform. Consequently, we will make the assumption that the scaling process $f(t)$ is a Gaussian process with stationary increments. $R(x, y)$ will denote the covariance of the process.

B.1 The law of $|T_\psi[f](x_{j,k}, a_j)|$

Let $x_{j,k}$ be the location of a modulus maximum of the wavelet transform at a fixed scale a_j (see below for a precise definition). The goal of this section is to derive the analytical expression of $p_M(z)$ the unconditional probability density function of continuous wavelet coefficient modulus at modulus maxima location, i.e., the law of $|T_\psi(x_{j,k}, a_j)|$. For that purpose, let us define $\lambda_{n,m}$ as:

$$\lambda_{n,m} = \frac{1}{2} \int \int \psi^{(n)}(x) \bar{\psi}^{(m)}(y) R(x, y) dx dy, \quad (35)$$

where $\psi^{(k)}$ refers to the k -th derivative of $\psi(x/a_j)$. Let us remark that $\lambda_{n,m}$ is nothing but the covariance of the stationnary random processes $\partial^m T(x, a_j)/\partial x^m$ and $\partial^n T(x, a_j)/\partial x^n$. It is easy to show that the previous integral is zero for odd $m + n$. Since $f(t)$ is Gaussian, $T_\psi(b, a_j)$ and its derivatives is a multivariate Gaussian process. It follows that the joint law $\rho(x_0, x_1, x_2)$ of $x_0 = T_\psi(b, a_j)$, $x_1 = \partial T_\psi(b, a_j)/\partial b$ and $x_2 = \partial^2 T_\psi(b, a_j)/\partial b^2$ can be expressed in terms of $\lambda_{n,m}$:

$$\rho(x_0, x_1, x_2) = \frac{e^{-\frac{x_1^2}{2\lambda_{1,1}} - (\frac{x_0^2}{2\lambda_{0,0}} + \frac{x_2^2}{2\lambda_{2,2}} - 2\frac{x_0 x_2 r}{\sqrt{\lambda_{0,0}\lambda_{2,2}}})/(2(1-r^2))}}{2\pi^{3/2} \sqrt{\lambda_{0,0}\lambda_{1,1}\lambda_{2,2}(1-r^2)}} \quad (36)$$

where r is the correlation coefficient $-1 < r = \frac{\lambda_{0,2}}{\sqrt{\lambda_{0,0}\lambda_{2,2}}} < 0$.

A modulus maximum at $x_{j,k}$ of the wavelet transform is defined by the following two conditions: (i) Zero crossing of $\partial T/\partial b$ and (ii) $T\partial^2 T/\partial b^2 < 0$. Let $\epsilon \ll 1$ be a finite resolution. Condition (i) can be expressed as $(T_\psi(b, a_j) - T_\psi(b - \epsilon, a_j))(T_\psi(b + \epsilon, a_j) - T_\psi(b, a_j)) < 0$. To the first order in ϵ this condition becomes:

$$|\partial T/\partial b| < \frac{\epsilon}{2} |\partial^2 T/\partial b^2|. \quad (37)$$

The density function $p_M(z)$ (recall that z is a dummy variable representing wavelet coefficient modulus at a modulus maximum location), at a finite resolution ϵ , is thus equal to:

$$p_M(z) = \frac{C_0}{\epsilon} \int_{-\infty}^0 dx_2 \int_0^{-\epsilon x_2/2} dx_1 \rho(z, x_1, x_2) = -\frac{C_0}{\epsilon} \int_{-\infty}^0 dx_2 \text{Erf} \left(\frac{\epsilon x_2}{\sqrt{\lambda_{1,1}}} \right) \tilde{\rho}(z, x_2), \quad (38)$$

where $\tilde{\rho}(x_0, x_2) = \int \rho(x_0, x_1, x_2) dx_1$, $\text{Erf}(x)$ is the error function $\text{Erf}(x) = \frac{1}{\sqrt{(2\pi)}} \int_0^x e^{-x^2/2} dx$ and C_0 a normalization constant.

When $\epsilon \rightarrow 0$, the function $\text{Erf}(\epsilon x) \sim \epsilon x$ to the first order, so that we finally obtain:

$$p_M(z) = -C_1 \int_{-\infty}^0 dx_2 x_2 \tilde{\rho}(z, x_2) \quad (39)$$

$$= C_1 e^{-\frac{z^2}{2\lambda_{0,0}}} \left[\pi \sqrt{2} r z \text{Erf} \left(\frac{r \sqrt{(2)} z}{2 \sqrt{\lambda_{0,0}(1-r^2)}} \right) - \pi \sqrt{2} r z + 2 \sqrt{\pi \lambda_{0,0}(1-r^2)} e^{-\frac{r^2 z^2}{2\lambda_{0,0}(1-r^2)}} \right] \quad (40)$$

⁴Let us note that there exists a *scale adapted* version [35], [38] that leads to a much more robust estimation of negative order moments ($q < 0$),

where C_1 is a normalization constant that can be computed using the normalization equation $\int_0^\infty p_M(z)dz = 1$.

B.2 Estimation of H

In order to get an estimator of H , one can proceed along the same line as for orthogonal coefficients. The bias and variance of this estimator can be computed from the bias and variance of expression (34). From appendix A, we thus have to compute the order q moments of WTMM: $M(q) = \int z^q p_M(z)dz$. Straightforward computations leads to the following analytical expression for $M(q)$:

$$M(q) = K \left[\Gamma(q/2 + 1/2) 2^{q/2} \left((1-r^2)^{q/2+3/2} + r^2 H \left([3/2 + q/2, 1/2], 3/2, \frac{r^2}{r^2-1} \right) (1+q) \right) - \sqrt{\pi} r 2^{q/2} \Gamma(q/2 + 1) \sqrt{1-r^2} \right] \quad (41)$$

where r is the correlation coefficient defined previously, $K = \frac{\lambda_{0,0}^{q/2}}{(1-r)\sqrt{\pi(1-r^2)}}$, and $H([w, x], y, z)$ is the generalized hypergeometric function:

$$H([w, x], y, z) = \prod_{k=0}^{\infty} \frac{\Gamma(w+k)\Gamma(x+k)\Gamma(y)}{\Gamma(x)\Gamma(w)\Gamma(y+k)} \frac{z^k}{k!}. \quad (42)$$

Strictly speaking, in order to apply theorem 1, the number of WTMM at scale a_j should be non random. Eventhough N_j fluctuates from realisation to realisation in Eq. (34), numerical experiments (see next section) suggest that it can be replaced by its mean value $E(N_j)$. This value (hereafter denoted as N_j) can be exactly computed for a Gaussian process [29]:

$$N_j = \frac{N}{\pi} \left(\sqrt{\frac{-\lambda_{0,2}}{\lambda_{0,0}}} + \frac{1}{2} \sqrt{\frac{-\lambda_{0,4}}{\lambda_{0,2}}} \right) \quad (43)$$

where $\lambda_{n,m}$ is defined in Eq. (35). Since $\lambda_{n,m} \sim a_j^{-(n+m)}$, we have the classical result [30] $N_j \sim a_j^{-1}$.

VI. ESTIMATION OF THE SCALING EXPONENT OF A FARIMA(0, D , 0)

In this section, we study the problem of the estimation of H given some realizations of a FARIMA(0, D , 0) (with $D = H + 1/2$). Both the (bi)orthogonal estimators and the WTMM estimator will be considered. The final comparison will also include the DFA estimator [13], [21].

A. Estimation using the (bi)orthogonal wavelet based estimators

A.1 Numerical simulations

The biorthogonal basis adapted to the FARIMA(0, D , 0) would be, of course, particularly well adapted to estimate H . In this basis, all the wavelet coefficients are independant within and between scales. More importantly, they strickly follow the scaling hypothesis across all scales (15). This means that, when estimating the parameter H using a log-log fit of the variance of the coefficients, the fit can be performed down to the smallest scale $j = 1$. This is, of course, neither the case for the continuous wavelet nor the orthogonal wavelet estimators. Thus, the use of adapted biorthogonal wavelet basis will lead to much better estimation. Actually one can prove that, in the limit of infinite length series, this estimator achieves the Cramér-Rao lower bound. However, one must keep in mind that the adapted basis can only be built knowing a priori the parameter H ! Thus, the so-defined estimator is a theoretical estimator that cannot be used in practical situations. In the following, we will use this estimator as a *reference* estimator to compare the other estimators with.

We have used the (bi)orthogonal wavelet based estimators to estimate the exponent $H = 0.7$ of a computer generated FARIMA time-series. Following Section IV-C, we use $q = \bar{q} = 2$, $J_m = J_{min} = 1$ and $J_M = \bar{J} = J_{max}$ and we manage not to be bothered by any border effect by using slightly longer time-series. In figure 4, we display both the MSE and the variance of the estimator as functions of $\log_2 N$. Actually, three different estimators have been used corresponding to three different sets of curves. First the orthogonal wavelet-based estimator using d9 (\triangle), the adapted estimator using d9h7 (\bullet) and finally an “approximately adapted” estimator using d9h5 (\circ).

Let us first note that for the 3 estimators, the variance is very close to the theoretical variance. That indicates that, even when using d9, the wavelet coefficients are almost decorrelated. Thus, the main source of error does not come from a “bad” decorrelation ($n_\psi = 9$ seems to be large enough). Moreover, the MSE remarkably fits the theoretical curve and is very small in the case the wavelet is “perfectly” adapted (i.e., using d9h7). It is still very small, but not as good as for d9h7, when using the “approximately adapted” d9h5 wavelet. But the MSE is clearly much larger when using d9. Thus, the main source of error is the bias. Let us explain why it is so.

The time-series one performs the (bi)orthogonal wavelet transform on, can be seen as the sampling of a time-continuous function using, as the sampling filter, the scaling function ϕ associated to the (bi)orthogonal wavelet ψ . We know, in the case of FARIMA's, that if one uses an adapted wavelet ψ_a (corresponding to an adapted scaling function ϕ_a), then

the continuous time function is nothing but a fBm [50]. Then, it is very easy to prove [26] that, analyzing a FARIMA time-series using an arbitrary wavelet ψ associated to a scaling function ϕ , amounts to sample the corresponding fBm using a sample filter h whose Fourier transform is $\hat{h} = \hat{\phi}/\hat{\phi}_a$. In our case, we have $\phi_a = \phi_{d9h7}$ and $\phi = \phi_{d9}$ or $\phi = \phi_{d9h5}$. Actually, one can show that $\hat{\phi}_{d9}$ and $\hat{\phi}_{d9h5}$ are very similar to $\hat{\phi}_{d9h7}$ at small frequencies but that they are different at large frequencies. Indeed this difference is much more significant for $\hat{\phi}_{d9}$ than for $\hat{\phi}_{d9h5}$. It implies that the wavelet coefficients at small scales do not follow a perfect scaling (as for the adapted basis d9h7). Let us note that it is all the more important that there are much more wavelet coefficients at small scales than at large scales. This explains why the MSE is much larger when using d9 than when using d9h5 or d9h7.

As stated above, we used $J_m = J_{min} = 1$. Remark that for d9, the bias at the smallest scales is such that a smaller MSE can be achieved by using $J_m > 1$. But, even though we could in practice perform this optimization on J_m , d9 would remain the worst choice. (Observe that, as a first approximation, the loss of the first ($j = 1$) scale of a (bi)orthogonal decomposition amounts to replacing N by $N/2$ in Eq. (33), i.e., multiplying the MSE by 2.)

A.2 Wich (bi)orthogonal wavelet to use for FARIMA(0,D,0)?

As we have just seen, the adapted (bi)orthogonal wavelet estimator performs a very small MSE mainly due to the small scale coefficients. The “less” adapted the wavelet is, the more biased are the small scale wavelet coefficients and the worse is the estimation. Of course, in order to estimate the exponent H of a FARIMA, one does not know a priori which is the “ideal” adapted biorthogonal wavelet estimator to use (e.g., d9h7). However, it is always better to use a wavelet which has more chance to be “approximately” adapted (e.g., d9h5) than a wavelet that one knows it will be always very far from being adapted such as d9.

The idea is that, in order to get the best estimation of H , one has to make the time-series as “close” as possible to a white noise. Thus, for instance, when analyzing a fBm, using d9 directly on the fBm is not as good as using d9 on the (discrete) derivative of the fBm, i.e., using d9h5 directly on the fBm. Since the “ideal” estimator (i.e., the exactly adapted wavelet) cannot be built before measuring H , one could first perform a “rough” estimate of the exponent H using any estimator (for instance the wavelet-based estimator using d9 or d9h5) and then perform a “finer” estimation using an “approximately” adapted (biorthogonal) wavelet.

These results apply to FARIMA(0,D,0) processes, in section VII we will see how they transposes to a more general framework.

B. Estimation using the WTMM

B.1 The WTMM absolute moments in the case of a FARIMA(0, D, 0)

In the particular case of a FARIMA(0, D, 0), to compute the expressions (21) and (22) and consequently the bias (Eq. (32)) and the variance (Eq. (31)) of the estimator, we just need to compute the expression for $\lambda_{n,m}$ and r in Eqs (41) and (43). If the wavelet is $\psi(t) = g_p(t)$ as defined in Eq. (11), one gets

$$\lambda_{n,m} = \sigma^2 (-1)^{\frac{n-m}{2}} \Gamma\left(\frac{m+n}{2} + p - H\right). \quad (44)$$

Using this expression, we have been able, for each wavelet g_p and exponent H , to compute the Cramér-Rao lower bound and the expression for the bias and the variance of the estimator $\mathcal{L}_{N_j,q}$. The value q^* that minimizes the variance is found to be close to 2.3 for all wavelets and the Cramér-Rao bound is smaller than 1/2 (e.g. 0.308 for $H = 0.7$ and $\psi = g_2$).

B.2 Numerical estimation of $\ln(M_q(a))/q$

We have applied the WTMM estimator for computing the $\ln(M_q(a_j))/q$ of a FARIMA of size 1024 and of exponent $H = 0.7$. We made 60000 realizations of this discrete process. The sampling rate is supposed to be 1 and we have systematically used the analyzing wavelets g_1, g_2, g_3 and g_4 defined by Eq. (11). Moreover, the minimum scale is fixed to $a_{min} = 2.15$ (the reference scale 1 is chosen so that the scale a corresponds to a Gaussian function whose standard deviation corresponds to a samples). In figure 5, we show $N_j \text{Var}(\mathcal{L}_{N_j,q})$ as a function of q . Let us note that in that case N_j is the average number of maxima found at scale $a_j = 2^{j-1} a_{min}$. Since the more oscillations a wavelet has the more maxima are found, N_j increases with the order of the wavelet. We find $N_j = 127$ for g_1 , 200 for g_2 , 251 for g_3 and 293 for g_4 in perfect agreement with theoretical expression (43). Since the more vanishing moments, the better the wavelet coefficients are decorrelated (16), we would expect g_4 to correspond to the best estimator. This is clearly not the case. This is due to the fact that if there are more vanishing moments then, although the decay of the autocorrelation function is faster, the wavelets have a larger support so that “close” wavelet coefficients are actually more correlated. Indeed, as illustrated in figure 6, the trade off between the number of vanishing moments and of the size of the support of the wavelet leads to g_1 as the optimal choice for time-series shorter than 500 samples [26]. Let us point out that the Cramér-Rao bound (i.e., the minimum variance of the estimator) for $\ln M_q(a_j)/q$ using N

Gaussian independent variables is $1/2N$. As shown in Section IV-B.1, from a numerical point of view, this bound is reached for $q = 2$ when using the N_j wavelet coefficients corresponding to the adapted wavelet decomposition, i.e., $Var(\mathcal{L}_{N_j, q^*=2}) \simeq 1/2N_j$. As illustrated in figure 5, not only we find that for the WTMM estimator, the optimal q value is rather $q = q^* \simeq 2.5$, but the minimum value of $N_j Var(\mathcal{L}_{N_j, q})$ is clearly below $1/2$. As mentioned above, using the analytical expressions (41) and (44), we can show that the value that optimizes $N_j Var(\mathcal{L}_{N_j, q})$ is close to 2.3 for all wavelet orders p and that the Cramér-Rao bound is below $1/2$. This can be interpreted as the fact that each local maximum is “more” adapted to the signal and thus contains “more” information than any single (bi)orthogonal wavelet coefficient. Moreover, some additional numerical computations prove that summing over all the continuous wavelet transform coefficients instead of just the local maxima do not improve the estimation. Thus, not only the local maxima “carry” more information than any other coefficient, but the set of all the local maxima “carry” as much information as the set of all the continuous wavelet transform coefficients. The discrepancy between the curves in Fig. 5 and the analytical computations actually results from the correlations between successive maxima pointed out above. We have checked that if one selects only independent maxima in the computation (e.g. by retaining only one over $2p$ maxima in the computation of $Var(\mathcal{L}_{N_j, q})$), one recovers the exact theoretical curve with a minimum value at $q^* = 2.3$. Moreover, if one extends theorem 1 to correlated variables, the influence of the correlation between successive maxima can be evaluated precisely. This rather technical problem will be studied in a forthcoming work.

B.3 Estimation of H : numerical simulations

As already explained, we use the estimator $\hat{H}_{J_m, J_M, q}$ defined in Eq. (29) in which $\mathcal{L}_{N_j, q}$ is defined by Eq. (34). As for the adapted biorthogonal wavelet estimator, we have tried to optimize on the three parameters J_m , J_M and q as a function of N in order to minimize the MSE. Again, we have found that we could simply fix $J_m = J_{min}$, $J_M = \bar{J} = J_{max}$ and q to be the value which minimizes the variance in figure 5, i.e., $q = \bar{q} = 2.5$, without losing much in terms of MSE. Thus in all numerical estimations, we will use these three values.

The numerical tests have been made on FARIMA’s of different sizes corresponding to the same exponent $H = 0.7$ (let us note that the same results would be obtained with a different exponent [26]). For each length N , we have performed 5000 realizations of the same FARIMA process in order to compute the MSE of the estimator $\hat{H}_{J_{max}, q=2.5}$. Moreover we have chosen to discretize the scale parameter of the wavelet transform as $a_j = \lambda^{j-1} a_{min}$ with $\lambda = 2^{0.1}$ and we have repeated the computations for the 4 analyzing wavelets g_1, g_2, g_3 and g_4 (Eq. (11)).

As shown in figure 6, for time-series of size greater than $N^* \simeq 500$, the more vanishing moments, the smaller the MSE. Actually, one can show that the main source of error comes from the variance and that the bias is very small [26]. As we already explained, increasing the number of vanishing moments increases the decay of the autocorrelation function, i.e., decreases the correlation between wavelet coefficients which are far from each others. It thus leads to a smaller variance of the estimator. However, for small time-series (smaller than N^*), there is not enough coefficients to be able to take advantage of the faster decay of the autocorrelation function. Moreover, increasing the number of vanishing moments increases the size of the support of the analyzing wavelet. That results in increasing the correlation between close coefficients, i.e., between all the coefficients. This leads to an increase of the variance and consequently of the MSE.

As a conclusion, in order to minimize the MSE, one should choose the WTMM g_1 estimator for time-series shorter than 500 samples, whereas for time-series longer than 500 samples, g_4 seems to be a good choice.

C. Comparisons of the different estimators (including DFA)

We are now ready to give a comparison between the different MSE’s obtained using the different estimators we have described in the last Section. In Table I, we compare the estimation of the scaling exponent of FARIMA’s of size 64 up to 2048 (using 5000 realizations) with $H = 0.7$. The different estimators are: the adapted estimator (d9h7), the “approximately” adapted estimator (d9h5), the orthogonal estimators (d2 and d9) and the WTMM estimators using the best estimator among g_1, g_2, g_3 and g_4 . In all these computations, we used mirror border effects and J_m and J_M set to 1 and J_{max} . We have also tested the DFA method introduced by Peng *et al.* [8], [21]. Let us recall that Taqqu *et al.* [13] have shown that, in terms of the MSE, it is the best nonparametric estimator of the scaling exponent of a FARIMA that is not wavelet-based.

These results raise the following comments:

- (1) The adapted estimator (d9h7) leads to a very small MSE. As already explained, this estimator achieves the Cramér-Rao bound in the limit $N \rightarrow +\infty$. For finite N , due to the first-order term of the bias (the term in $1/N$ in Eq. (21)) and to the second-order term of the variance (the term in $1/N^2$ in Eq. (22)), the MSE stays above this limit. Let us recall that this estimator cannot be used in practical situations since, in order to build the adapted basis, one needs to know a priori the value of H .
- (2) For short time-series, an “approximately” adapted estimator (such as d9h5) is almost as good as the exactly adapted estimator.

N	64	128	256	512	1024	2048	8192
Cramér-Rao	0.108	0.0713	0.0483	0.0332	0.0231	0.0162	0.0080
d9h7	0.167	0.0982	0.0630	0.0415	0.0283	0.0189	0.0087
d9h5	0.169	0.114	0.0790	0.0567	0.0422	0.0327	0.0235
d2	0.237	0.178	0.143	0.122	0.108	0.0986	0.0910
d9	0.324	0.229	0.163	0.132	0.115	0.106	0.0971
WTMM	0.278 g_1	0.194 g_1	0.136 g_1	0.0845 g_4	0.0529 g_4	0.0348 g_4	0.0164 g_4
DFA	0.129	0.0979	0.0787	0.0649	0.0535	0.0454	0.0339

TABLE I

MSE OF THE ESTIMATION OF THE SCALING EXPONENT $H = 0.7$ OF A FARIMA BASED ON THE STUDY OF 5000 REALIZATIONS.

(3) The classical orthogonal estimators (d9 and d2) are clearly not as good. The main problem comes from the bias due to the bad scaling at small scales.

(4) The WTMM estimator leads to larger MSE for short time series ($N \leq 128$) as compared to the orthogonal estimators but to much smaller MSE for long time-series (for time-series of size $N = 8192$, it also performs better than the d9h5 estimator). Let us note that, for these estimators, for each N , we have used the analyzing wavelet which leads to the smallest MSE. Basically, as already explained in section IV-C for the orthogonal estimators, when using the WTMM estimator for short (resp. long) time-series, one should use a low-order (resp. high-order) wavelet in order to maximize the decorrelation of the wavelet coefficients.

(5) For time-series of length smaller than 128, the DFA estimator beats all the other estimators, including the adapted wavelet d9h7 estimator! As soon as the time-series is longer than 256 (resp. 512), the bias of the d9h7 (resp. d9h5) estimator is small enough and leads to a smaller MSE than the DFA estimator.

Thus for very short time-series, the most efficient estimator of the scaling exponent H of a FARIMA appears to be the DFA estimator. For time series of size 8192, the WTMM estimator using the wavelet g_4 should be used. For time-series of length ranging from 512 to 8192 samples, the biorthogonal estimator associated to d9h5 is the most performant. Let us note that it simply amounts to apply the orthogonal d9 estimator on the increments of the FARIMA. Actually, one could even improve the estimation by first estimating H using d9h5 leading to a preliminary estimate H_1 and then by using a “better adapted” estimator d9h H_1 . Let us note that a different value of H would lead to the same conclusions [26].

As we have already anticipated in the introduction and in Section II, in order to test the robustness of these estimators, we have used each of them for estimating the scaling exponent of two other models which possess some characteristic features frequently found in experimental data.

VII. STUDYING THE ROBUSTNESS OF THE ESTIMATORS

As previously discussed, real time-series generally display a smallest scale below which the scaling behavior breaks down whatever the (Fourier, Wavelet, ...) basis one uses. It could be a dissipation scale (in the case of turbulent flow signals [1]) or any other physical smoothing phenomenon which, for instance, keeps from modeling the time-series values themselves by an exact FARIMA(0, D , 0). From a numerical point of view, it means that the linear fits should not be done from the smallest scale available a_{min} but from a larger but still small scale a_m . From our knowledge, all the estimators that have been introduced in the literature have been facing the problem of detecting this scale a_m , but no automatic method for estimating a_m has been found. People generally just “look” at the scaling behavior over the whole available range of scales, and just “decide” down to which scale a_m the scaling extends to and then perform the linear fit from this scale.

In this Section, we test the robustness of the different wavelet-based estimators with respect to the existence of such a scale a_m and to the fact that one might poorly estimate it. For that purpose, we will test the estimators on the two previously introduced models: the FARIMA(0, D , Q) model (Section II-C) and the two-valued FARIMA(0, D , 0) model (Section II-D).

A. Estimation of the scaling exponent of the FARIMA(0, D , Q) process

We have used all the estimators presented in the last 2 sections in order to estimate the scaling exponent H of a FARIMA(0, D , Q). As explained in Section II-C, the FARIMA(0, D , 0) has been smoothed with a smoothing filter which is an arch of a sinusoid of amplitude 1. We have chosen to keep its size Q constant all along the analysis with value $Q = 32$ samples. The numerical simulations have been made on 5000 realizations of a FARIMA(0, D , Q) with $H = D + 1/2 = 0.7$ (a different value of H would lead to the same conclusions [26]). Figure 7 displays the MSE versus the length of the FARIMA when using the WTMM estimators corresponding to the analyzing wavelets g_1 , g_2 , g_3 and g_4 (Eq. (11)) or the (bi)orthogonal estimator corresponding to d1, d2, d2h5 and d9h7. Let us note that in these numerical experiments, the scale a_m has been chosen in order to minimize the MSE. This, of course, cannot be done in a practical

N/Q	32	64	128	256	512	1024	2048
d9h7	0.219	0.139	0.0888	0.0645	0.0548	0.0520	0.0517
d9h5	0.229	0.146	0.0915	0.0640	0.0515	0.0477	0.0474
d2	0.230	0.155	0.0976	0.0646	0.0454	0.0350	0.0304
d9	0.286	0.174	0.103	0.0657	0.0476	0.0398	0.0385
WTMM (g_1)	0.176	0.131	0.0955	0.0694	0.0540	0.0396	0.0294
DFA	0.161	0.119	0.0923	0.0741	0.0620	0.0507	0.0440

TABLE II

MSE OF THE ESTIMATION (5000 REALIZATIONS) OF THE SCALING EXPONENT $H = 0.7$ OF A FARIMA(0, D , Q) WITH $Q = 32$ SAMPLES. THE SCALE a_m IS ESTIMATED BY JUST “LOOKING” AT THE CORRESPONDING LOG-LOG PLOTS. WE FOUND $a_m = 2^6$ FOR THE (BI)ORTHOGONAL WAVELET-BASED ESTIMATORS, $a_m = 2^4 a_{min}$ (WITH $a_{min} = 2.15$) FOR THE WTMM ESTIMATOR AND $a_m = 2^5 a_{min}$ (WITH $a_{min} = 4$) FOR THE DFA ESTIMATOR.

situation, but this exercise will allow us to understand which estimator is “potentially” the best estimator. We will test the robustness with respect to the estimation of a_m later on.

As seen in figure 7(a), the less vanishing moments the analyzing wavelet has the smaller the MSE is. As already explained in the previous Section, for small N , this is due to the fact that when the number of vanishing moments increases, even though the decay of the autocorrelation function increases, the correlation between two coefficients (that cannot be far one from each other since N is small) increases. However, the fact that the MSE increases with the number of vanishing moments, whatever N is, comes from a different phenomenon, namely some cross-over between the scaling behavior observed at scales smaller than a_m and at scales greater than a_m respectively. Indeed, it can be shown that, since below a_m the time-series is very regular, the wavelet transform behaves like [23], [29], [60], [61]:

$$|T_\psi(b, a)| \sim a^{n_\psi}, \quad \forall b \in \mathbb{R} \text{ and } \forall a < a_m. \quad (45)$$

Since for large scales, the wavelet transform is likely to display the scaling behavior a^H , these two power-law behavior are all the more different than n_ψ is large and consequently, the estimation of H is more precise for smaller n_ψ .

The same phenomenon happens when using (bi)orthogonal wavelet-based estimators. Indeed, as illustrated in figure 7(b), for any length N , the estimation is better when using the smallest order wavelet d1. Even the adapted wavelet d9h7 leads to a much greater MSE.

Thus, in terms of the MSE, under the hypothesis that one has access to a_m , the “potentially” best estimators to use are the WTMM g_1 estimator or the orthogonal d1 estimator.

Testing the robustness with respect to the estimation of a_m

In many practical situations, a_m is estimated by just “looking” at the $\ln \sigma(a)$ vs $\ln(a)$ plot. When the analyzing wavelet has only one vanishing moment, the power-law behaviors below a_m ($\propto a$) and above a_m ($\propto a^H$) look very similar on a noisy background. Therefore, the estimation of a_m may lead to a rather important error (i.e., more than an octave). Figure 8 illustrates the robustness of the estimation of H with respect to the choice of the cut-off scale for wavelets with $n_\psi = 1$. It clearly shows that the WTMM g_1 estimator is much more robust than the orthogonal d1 estimator.

Thus, we end up with an apparently paradoxical conclusion. The d1 wavelet has a much better frequency selectivity than the wavelet g_1 in the sense that the transition from the $\propto a$ behavior to the $\propto a^H$ comportment is sharper with d1 than with g_1 . Thus d1 leads to a better optimal estimation of H as seen on figure 7. However, the fact that the transition is sharper with d1 also induces a stronger instability of this estimator with respect of the estimation of a_m (Fig. 8). Consequently, one should rather use the WTMM g_1 estimator

Comparing the MSE

In the Table II, we compare the estimation of the scaling exponent of a FARIMA(0, D , Q) ($H = D + 1/2 = 0.7$ and $Q = 32$) of size 1024 to 65536, when estimating a_m by just “looking” at the log-log plots as one usually does for experimental data. In order to compare these results with the ones presented in Table I (basically corresponding to the case $Q = 1$), one has to compare the MSE obtained over the same range of H power law scaling. For that purpose, instead of considering the number of samples N of the realizations, we refer to the corresponding size of the H scaling range, i.e., N/Q .

As shown by these results

(1) For very short time-series ($N/Q \leq 64$), the DFA method leads to the best MSE and the WTMM g_1 estimator is the best wavelet-based estimator. Moreover the MSE obtained with this latter estimator is very close to the optimal one using the DFA estimator.

N	64	256	512	1024	4096	16384
Cramér-Rao	0.108	0.0483	0.0332	0.0231	0.0114	0.00565
d9h7	0.202	0.101	0.0828	0.0706	0.0597	0.0492
d9	0.377	0.176	0.130	0.105	0.0677	0.0457
WTMM (g_1)	0.299	0.143	0.105	0.0805	0.0477	0.0293
DFA	0.180	0.125	0.108	0.0947	0.0734	0.0590

TABLE III

MSE OF THE ESTIMATION OF THE SCALING EXPONENT $H = 0.7$ OF A TWO-VALUED FARIMA. THE SCALE a_m IS OPTIMIZED IN THE CASE OF D9H7 AND D9 WHEREAS NO OPTIMIZATION IS PERFORMED NEITHER FOR THE WTMM ESTIMATOR NOR FOR THE DFA ESTIMATOR.

(2) For larger size ($N/Q \geq 512$), the DFA method is no longer competitive as compared to WTMM or any other wavelet-based estimator. Moreover the larger the series is, the better are the low-order wavelet-based estimators such as d2 or WTMM with g_1 . This is due to the fact that a_m is generally underestimated and thus the smaller n_ψ , the softer the cross-over to the wavelet-dependent power-law behavior of the wavelet coefficients (Eq. (45)).

(3) Only for a very limited range ($128 \leq N/Q \leq 256$), the adapted wavelet d9h7 performs a smaller MSE than WTMM or d2. Let us recall that d9h7 is an “ideal estimator” that cannot be used in practical situations. Moreover, the MSE obtained using the “approximately” adapted d9h5 estimator is close to the one obtained using d2 or WTMM.

As a conclusion, we advocate the use of the WTMM g_1 estimator which, though not always optimal, has the advantage to lead to very competitive MSE for any size N .

B. Estimation of the scaling exponent of the two-valued FARIMA(0, D, 0) process

The same tests as in the previous Section has been performed on 5000 realizations of two-valued FARIMA time-series with $H = 0.7$ (Section II-D). The results of these numerical experiments are illustrated in figure 9. Preliminary results have shown that the WTMM and DFA estimators are robust in respect to the estimation of a_m [26]. Thus for these estimators we have chosen a_m to be the smallest scale available in the corresponding continuous wavelet transform ($a_m = a_{min}$). However, the (bi)orthogonal estimators are not robust at all. Thus, for these estimators, as for smoothed FARIMA’s in the previous Section, we have optimized a_m in order to minimize the MSE (which of course cannot be done in practical situations). Even though no optimization is used for the WTMM estimators, they appear to be, for N greater than 512, better estimators than the DFA estimator or than any adapted or not adapted (bi)orthogonal estimators. For very small N , it is better to use the DFA estimator.

These results are summarized in the Table III where the MSE is given versus N for the different estimators. The robustness of the WTMM estimators with respect to a_m and their quality of estimation make the WTMM g_1 estimator clearly the best to use as long as $N \geq 512$.

VIII. CONCLUSIONS

Most of the (nonparametric) techniques for estimating the scaling exponent of time-series that display scaling behavior consist essentially in the measurement of the slope in a log-log plot. Abry *et al.* [14], [16], [25] have advocated the use of orthogonal wavelet-based estimators. Indeed, these estimators have several advantages. They are very fast to compute, they are very competitive estimators in terms of MSE, they are blind to eventual superimposed smooth behavior (such as trends) and they are very robust when changing the shape of the underlying probability law [14] (Gaussian, exponential, Levy,...). Abry *et al.* performed their study only for long time-series (> 10000 samples) and they have shown that, in that case, the estimators can be considered as unbiased [15], [24]. As we have seen along this paper, this is no longer true for shorter time-series. Indeed, the bias can be the main source of error, and it is very important to take it into account in order to avoid underestimating the MSE of the estimator. In that framework, in the case of FARIMA time-series which are very commonly used for modelling scaling behavior, we have shown that the more “adapted” the (bi)orthogonal wavelet, the better the estimation (i.e., the smaller the MSE). In particular, we advocate the use of “approximately” adapted biorthogonal wavelets that can be obtained through an iterative method.

As a very efficient alternative estimator (as far as practical applications are concerned), we have introduced a wavelet-based estimator which is derived from the WTMM method [23], [26], [35], [38] for computing the multifractal spectrum of a time-series. This estimator is based on the local maxima of the modulus of the continuous wavelet transform. It is no longer based on (bi)orthogonal wavelet analysis and consequently is slightly slower to compute than the orthogonal wavelet-based estimators. However, we have shown that, in practical situations, for which generally the scaling behavior does not extend to the smallest available scale (i.e, the sampling scale), the WTMM estimator is more robust than any other wavelet-based estimators. Moreover, we have shown that, in the case of time-series of size $N \geq 512$, it leads to smaller MSE than the DFA method which has been shown by Taqqu *et al.* [13], to be among the “best” estimators which

are not wavelet-based estimators. Thus, in this case, we strongly advocate the use of the WTMM estimator. As far as analytical considerations are concerned, we have computed the WTMM probability distribution and absolute moments. This enable us to compute optimal q value for MSE and to estimate the Cramér-Rao lower bound . However, because successive maxima are correlated, we are not exactly in the conditions assumed in Appendix A so further analytical work is necessary to solve completely the problem.

Most of the work done on the estimation of the scaling exponent concern monofractal models which can be seen as derived versions of fBm's. This means that the regularity of the process is the same everywhere and can be quantified by a unique exponent H . This is a strong assumption that is often not satisfied in practical situations. In the case the process is multifractal, one has to be very careful since the scaling exponent of the q -order moments of the scale dependent quantity under consideration depends on q [30], [35], [38]. Then, clearly, the standard estimators mentioned in the introduction using, for instance, first-order absolute moment or second-order moment, will not lead to the same estimation at all. Thus it is of great importance, before applying any of these estimators, to run some prior tests for estimating the multifractality of the process. Basically it consists in measuring the different scaling exponents $H(q)$ corresponding to the different q -order absolute moments. For positive q 's, this can be done basically with any of the estimators mentioned in the introduction. However, for negative values of q , these estimators are unstable and consequently only half (corresponding to the strongest singularities) of the multifractal spectrum can be estimated [35], [37], [38]. In previous work, the WTMM method has been shown to be very stable, even for negative q 's and proved to give access to the whole multifractal spectrum [30], [35], [37]. Let us note that the negative q values could be of interest even in the framework of self-similar processes. Indeed, some commonly used estimators of the scaling exponent of Lévy walks are based on the way the probability distribution function of the variations of the walk at a given scale behaves around 0 when the scale is changed [3]. It can be shown that it is actually related to looking at the negative order moment of the variations. Thus, negative absolute moment-based estimators could be very competitive in various situations.

In this paper, we have assumed that the process is monofractal. We are currently performing the same theoretical work on multifractal processes. We are elaborating numerical tests for the existence of a scaling behavior (which should be made prior to any estimation) as well as for discriminating between multifractality and monofractality. Moreover, we are currently studying the robustness and the precision of various estimators of the so-called multifractal spectra including the $D(h)$ singularity spectrum.

IX. ACKNOWLEDGMENTS

All the computations involved in this paper have been made using a free GNU licensed software LastWave [62].

APPENDIX A

Theorem 1: Let $\{X_i\}_{1 \leq i \leq N}$ be i.i.d. positive random variables (with $E(X) > 0$) and let $S_N = \sum_{i=1}^N x_i/N$. Let us assume that the density of X is bounded and that there exists $\alpha > 0$ such that $E(e^{\alpha X}) < +\infty$. Let $g : \mathbb{R}^{+*} \rightarrow \mathbb{R}$. Let us assume that g is C^m ($m \geq 2$, m even), that g_m is bounded on $[E(X) - \epsilon, +\infty]$ (for a given $\epsilon > 0$) and that g belongs to $L^2([0, E(X) - \epsilon])$. Then, in the limit $N \rightarrow +\infty$

$$E(g(S_N)) = \sum_{k=0}^{m-2} \frac{1}{k!} g^{(k)}(E(X)) M_k(S_N) + O\left(\frac{1}{N^{m/2}}\right), \quad (46)$$

where $M_k(S_N)$ stands for the moment of order k of S_N .

Proof

Let $T_m(t)$ be the Taylor expansion of g at $t = E(X)$ and let F_N be the density function of S_N . Thus,

$$\begin{aligned} E(g(S_N) - T_{m-1}(S_N)) &= \int (g(t) - T_{m-1}(t)) F_N(t) dt, \\ &\leq \int_0^{E(X)-\epsilon} |g - T_{m-1}| F_N dt + \int_{E(X)-\epsilon}^{+\infty} |g - T_{m-1}| F_N dt. \end{aligned} \quad (47)$$

On the one hand, there exists a constant K , such that for all $t \in [E(X) - \epsilon, +\infty[$,

$$|g(t) - T_{m-1}(t)| \leq K(t - E(X))^m; \quad (48)$$

then

$$\int_{E(X)-\epsilon}^{+\infty} |g(t) - T_{m-1}(t)| F_N(t) dt \leq K M_m. \quad (49)$$

Moreover, one can show [26] that $M_k(S_N) = O(N^{-k/2})$ if k is even and $M_k(S_N) = O(N^{-(k+1)/2})$ otherwise. Thus

$$\int_{E(X)-\epsilon}^{+\infty} |g(t) - T_{m-1}(t)| F_N(t) dt = O(N^{-m/2}). \quad (50)$$

On the other hand, since $g \in L^2([0, E(X) - \epsilon])$ and F_N is bounded, using the Schwartz inequality, one gets,

$$\left(\int_0^{E(X)-\epsilon} |g - T_{m-1}| F_N dt \right)^2 \leq \int_0^{E(X)-\epsilon} |g - T_{m-1}|^2 F_N dt \int_0^{E(X)-\epsilon} F_N dt. \quad (51)$$

Using the law of large numbers, one gets

$$\int_0^{E(X)-\epsilon} |g - T_{m-1}| F_N dt \leq C \left(\int_0^{E(X)-\epsilon} F_N dt \right)^{1/2}. \quad (52)$$

Since there exists $\alpha > 0$ such that $E(e^{\alpha X}) < +\infty$, one can use the large deviation theorem to prove that this last term is decreasing exponentially in N . Thus, from Eqs (47) and (50), one gets

$$E(g(S_N) - T_{m-1}(S_N)) = O(N^{-m/2}). \quad (53)$$

Moreover,

$$E(T_{m-1}(S_N)) = \sum_{k=0}^{m-1} \frac{1}{k!} g^{(k)}(E(X)) M_k(S_N). \quad (54)$$

Since m is even, the last term is $O(N^{-m/2})$ and, from Eqs (53) and (54), one gets Eq. (46).

Corollary 1: By applying this theorem to $m = 6$, one gets

$$\begin{aligned} E(g(S_N)) &= g + \frac{g_2}{2N} \text{Var}(X) \\ &\quad + \frac{1}{N^2} \left(\frac{g_4}{8} \text{Var}^2(X) + \frac{g_3}{6} M_3(X) \right) + O\left(\frac{1}{N^3}\right), \end{aligned} \quad (55)$$

where $g^{(k)}$ stands for $g^{(k)}(E(X))$. Moreover, if $g^2 \in L^2([0, E(X) - \epsilon])$, by applying Eq. (46) to g^2 (with $m = 6$), one gets

$$\begin{aligned} \text{Var}(g(S_N)) &= \frac{(g_1)^2}{N} \text{Var}(X) \\ &\quad + \frac{1}{N^2} \left(\left(\frac{(g_2)^2}{2} + g_1 g_3 \right) \text{Var}^2(X) + g_1 g_2 M_3(X) \right) + O\left(\frac{1}{N^3}\right). \end{aligned}$$

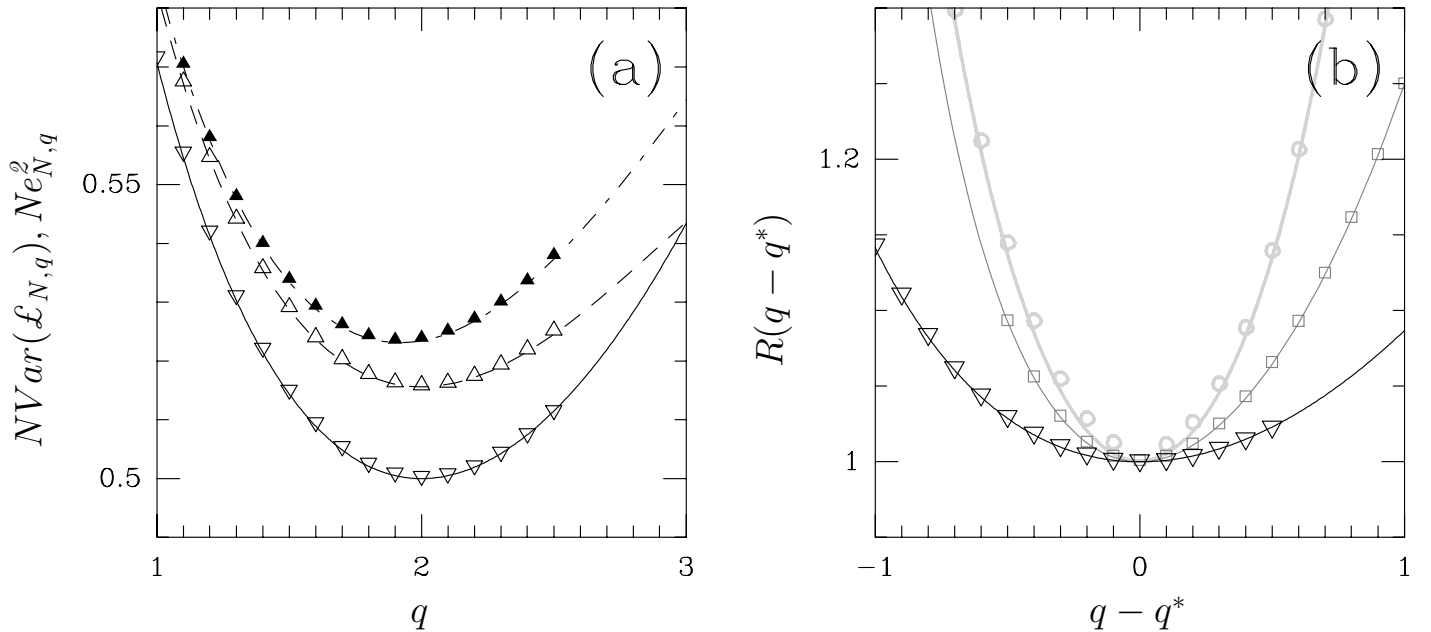


Fig. 1. Variance and MSE of the estimator \mathcal{L} . (a) Gaussian case: $NVar(\mathcal{L}_{N,q})$ as a function of q : (—) first term and (---) first two terms of Eq. (22) (for $N = 32$). The symbols (Δ) (resp. (∇)) correspond to numerical estimations of $NVar(\mathcal{L}_{N,q})$ for $N = 32$ and $\sigma = 10\,000$ (resp. $N = 2\,048$ and $\sigma = 1$). $Ne_{N,q}^2$ as a function of q : (---) first two terms of Eq. (26) for $N = 32$. The symbols (\blacktriangle) correspond to numerical estimations of $Ne_{N,q}^2$ for $N = 32$ and $\sigma = 10\,000$. The first-order terms in Eq. (22) (resp. Eq. (26)) are shown to underestimate the variance (resp. MSE) of \mathcal{L} for small values of N . (b) $R(q - q^*) = \frac{\Lambda_2(q - q^*) / (q - q^*)^2}{\min_q (\Lambda_2(q - q^*) / (q - q^*)^2)}$ as a function of $q - q^*$ for different probability distribution laws. From bottom to top : Gaussian ($q^* = 2$), exponential ($q^* = 1$) and log-normal ($\exp(\mathcal{N}(m, \sigma))$ with $\sigma = 1$) ($q^* = 0$) law. As illustrated by this figure, choosing the optimum $q = q^*$ for a given law leads to significant improvement of the MSE compared to the MSE obtained with a nonoptimal value $q \neq q^*$. Symbols correspond to numerical estimation using 50 000 signals of size 2 048 made of uncorrelated gaussian (see (a), ∇), exponential (with unit expectation, \square) and log-normal ($\exp(\mathcal{N}(0, 1))$, \circ) variables.

REFERENCES

- [1] U. Frisch, *Turbulence*, Cambridge Univ. Press, Cambridge, 1995.
- [2] J.-P. Bouchaud and M. Potters, *Théorie des Risques Financiers*, Eyrolles, Aléa-Saclay, 1997.
- [3] R.N. Mantegna and H.E. Stanley, *An introduction to Econophysics*, Cambridge Univ. Press, Cambridge, 2000.
- [4] Parks and W. Willinger, Eds., *Self Similarity in Network Traffic*, chapter "Wavelets for the analysis, estimation and synthesis of scaling data", Wiley, New York, 1998.
- [5] B.J. West, *Fractal Physiology and Chaos in Medicine*, World Scientific, Singapore, 1990.
- [6] B.J. West and W. Deering, *Fractal Physiology for Physicists*, Phys. Rep. 254:1, 1994.
- [7] G. G. Wilkinson, J. Kanellopoulos, and J. Megier, Eds., *Fractals in Geoscience and Remote Sensing*, Image Understanding Research Series, vol.1, ECSC-EC-EAEC. Bruxelles, Luxembourg, 1995.
- [8] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, "Long-range correlations in nucleotide sequences," *Nature*, vol. 356, pp. 168–170, 1992.
- [9] A. Arneodo, Y. d'Aubenton-Carafa, E. Bacry, P.V. Graves, J.F. Muzy, and C. Thermes, "Wavelet-based fractal analysis of DNA sequences," *Physica D*, vol. 96, pp. 291–320, 1996.
- [10] R. Fox and M.S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, vol. 14, pp. 517–532, 1986.
- [11] I. Giraitis and D. Sukgailis, "A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to the asymptotic normality of Whittle estimates," *Prob. Theory Relat. Fields*, vol. 86, pp. 87–104, 1990.
- [12] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.
- [13] M.S. Taqqu, V. Teverovsky, and W. Willinger, "Estimator for long-range dependence: an empirical study," *Fractals*, vol. 3, pp. 785–798, 1995.
- [14] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, *Self Similarity in Network Traffic*, chapter "Wavelets for the analysis, estimation and synthesis of scaling data", Wiley, New York, 1998.
- [15] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, *Self Similarity and long-range dependence through the wavelet lens*, to appear in *Long range dependence: theory and applications*, eds. Doukhan, Oppenheim, Taqqu, 2000.
- [16] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. Info. Theory*, vol. 45, pp. 878–897, 1999.
- [17] J.-M. Poggi and M.-C. Viano, "An estimate of the fractal index using multiscale aggregates," *J. Time Ser. Anal.*, vol. 19, pp. 221–233, 1998.
- [18] G. Samorodnisky and M.S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York, 1994.
- [19] L.F. Burgala and L.W. Klein, "Fractal structure of the interplanetary magnetic field," *J. Geophys. Res.*, vol. 91, no. A1, pp. 347–350, 1986.
- [20] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol. 31, pp. 277–283, 1988.

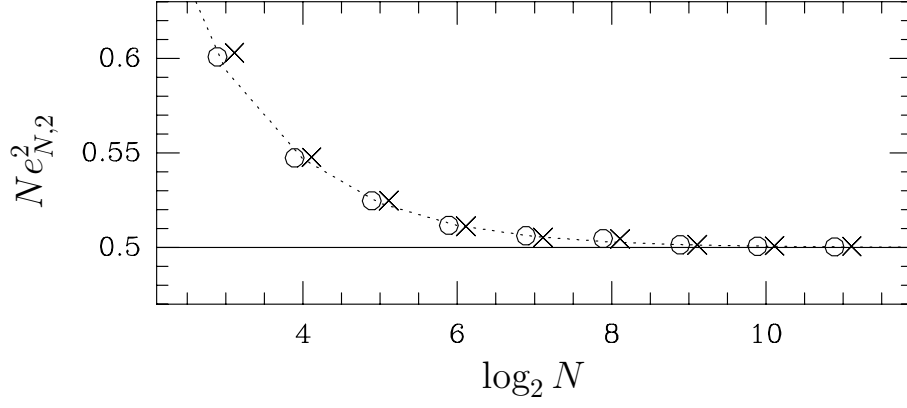


Fig. 2. MSE of the estimator \mathcal{L} as a function of N for $q = 2$. $Ne_{N,q=2}^2$ as a function of $\log_2 N$: (—) first term and (···) first two terms of Eq. (26). The symbols (o) (resp. (x)) correspond to numerical estimations of $Ne_{N,q}^2$ with $\sigma = 1$ (resp. $\sigma = 10\,000$).

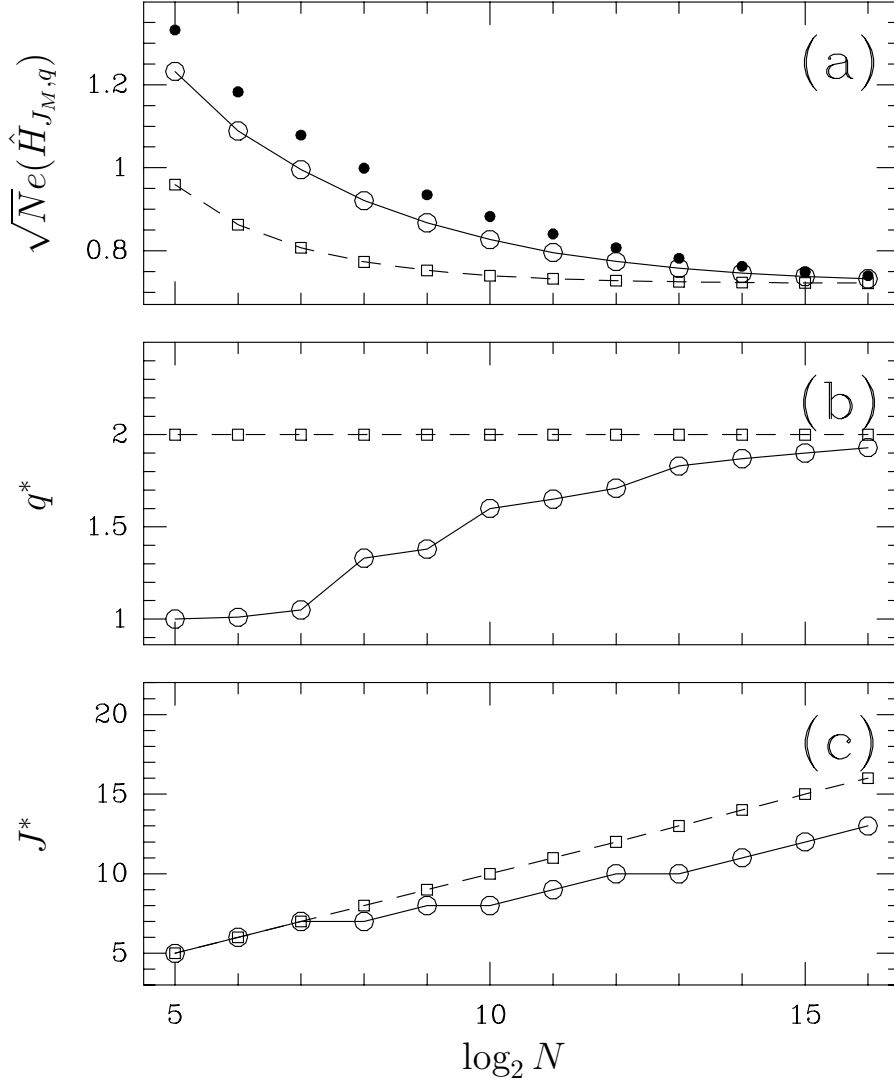


Fig. 3. (a) $\sqrt{N}e(\hat{H}_{J_M,q})$ as a function of $\log_2 N$. (b) and (c) optimal values q^* and J^* which minimize the MSE of the estimator $\hat{H}_{J_M=J^*,q=q^*}$. In all the figures the symbols have the following meaning: (o) correspond to computing the MSE using the expression (33), (□) to the MSE using only the first-order term in Eq. (33), and (•) to the MSE for $q = \bar{q} = 2$ and $J_m = 1$ and $J_M = \bar{J} = J_{max}$.

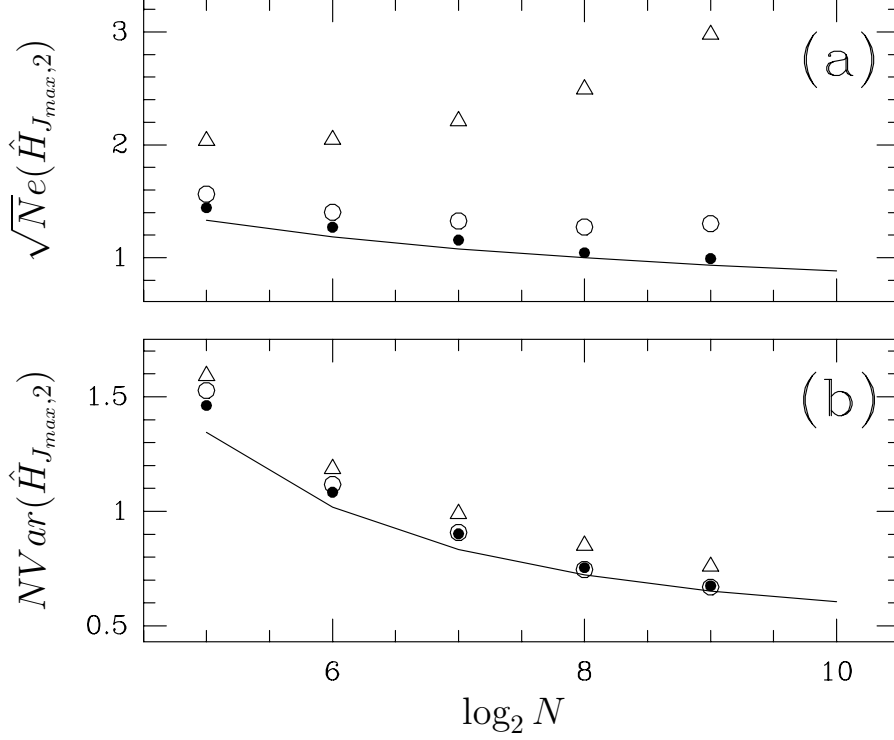


Fig. 4. Estimation of the scaling exponent $H = D + 1/2 = 0.7$ of FARIMA(0, D , 0) using the orthogonal wavelet d9 (\triangle), the “approximately” adapted wavelet d9h5 (\circ) and the adapted wavelet d9h7 (\bullet). The solid line corresponds to the theoretical predictions for the d9h7 estimator. (a) $\sqrt{N}e(\hat{H}_{J_{max},2})$ versus $\log_2 N$. (b) $NVar(\hat{H}_{J_{max},2})$ versus $\log_2 N$.

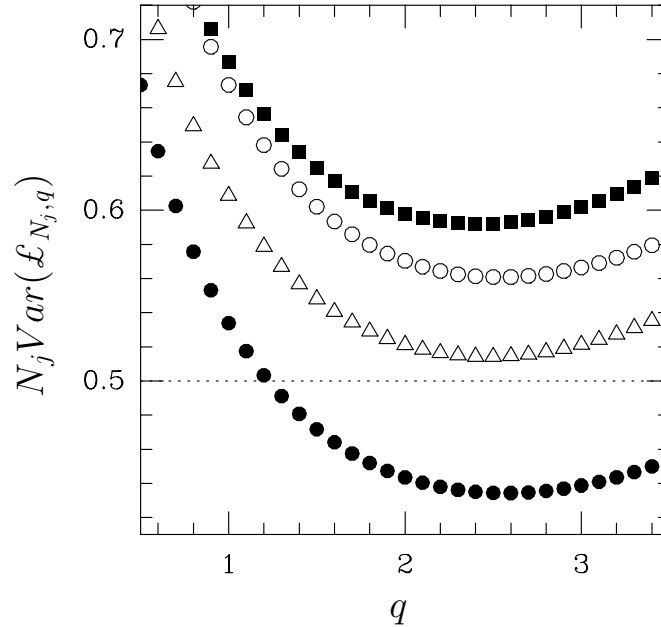


Fig. 5. $N_j Var(\mathcal{L}_{N_j,q})$ versus q using the WTMM estimator defined in Eq. (34). The computations have been performed with 60 000 realizations of length $N = 1024$ of FARIMA(0, D , 0) with exponent $H = D + 1/2 = 0.7$. The value of the scale is $a = 2.15$ and corresponds to $j = 1$ ($a_j = a_{min} 2^{j-1} = 2.15 \cdot 2^{j-1}$). The various symbols mean different estimations performed with the analyzing wavelets g_1 (\circ), g_2 (\bullet), g_3 (\triangle) and g_4 (\blacksquare) as defined in Eq. (11). The dashed line (---) indicates the Cramér-Rao bound.

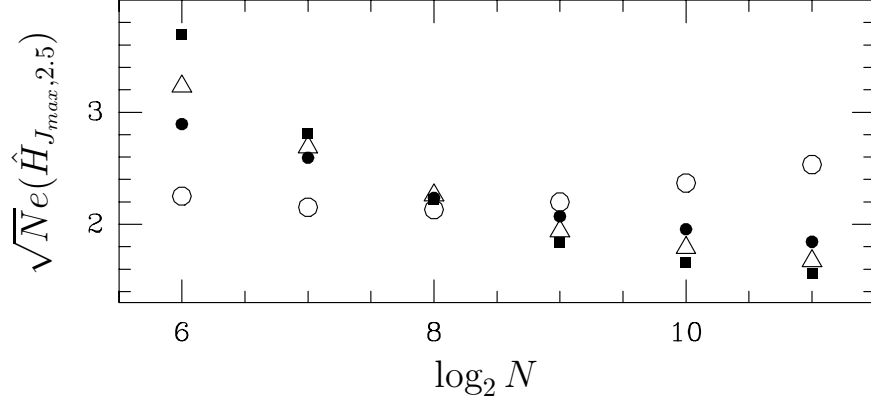


Fig. 6. Estimation of the exponent $H = D + 1/2$ for FARIMA(0, D , 0) time-series (5 000 realizations) using the WTMM estimator (Eq. (34)). $\sqrt{N}e(\hat{H}_{J_{max}, 2.5})$ versus $\log_2 N$. The various symbols have the same meaning as in figure 5.

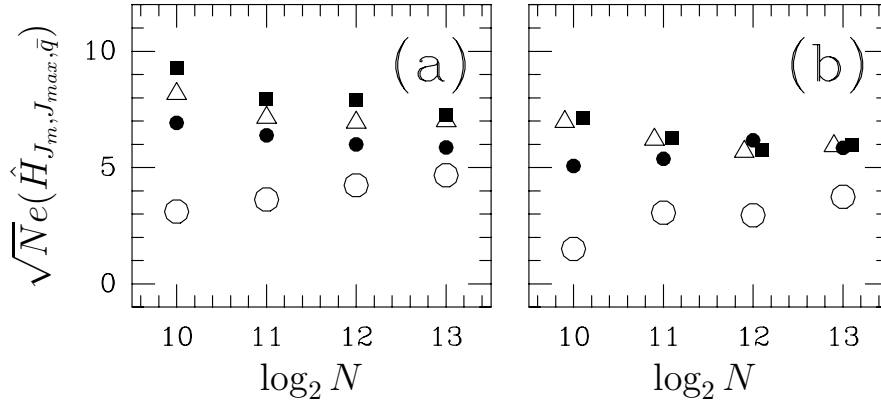


Fig. 7. Estimation of the scaling exponent $H = D + 1/2$ for a FARIMA(0, D , Q) time-series (5 000 realizations). The size of the smoothing filter is $Q = 32$ samples. (a) $\sqrt{N}e(\hat{H}_{J_m, J_{max}, \bar{q}})$ versus $\log_2 N$ using the WTMM estimator with (o) g_1 , (•) g_2 , (Δ) g_3 and (■) g_4 (Eq. (11)). (b) $\sqrt{N}e(\hat{H}_{J_m, J_{max}, 2})$ versus $\log_2 N$ using the (bi)orthogonal estimators with (o) d1, (•) d2, (Δ) d2h5 and (■) d9h7. In (a) and (b), q is chosen to be equal to the “simplified” values $\bar{q} = 2.5$ and 2 respectively and the value of J_m that minimizes the MSE has been selected (see text).

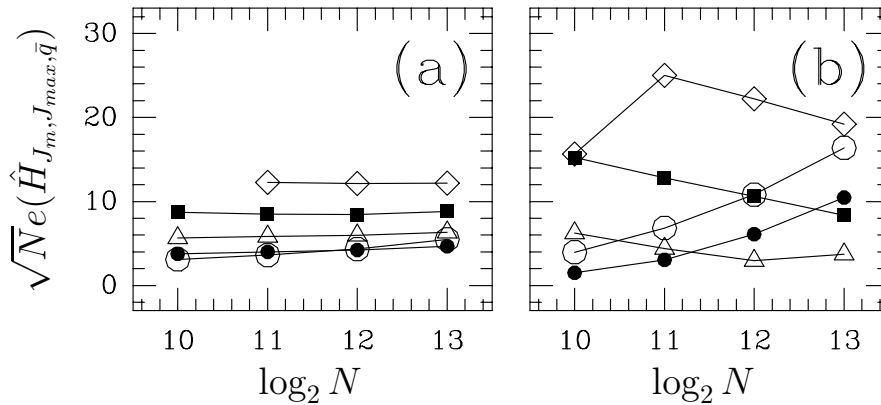


Fig. 8. Influence of the estimation of a_m on the MSE when estimating the scaling exponent $H = D + 1/2$ of a FARIMA(0, D , Q) of exponent $H = 0.7$ with a smoothing filter of size $Q = 32$. The cut-off scale a_m is chosen so that $a_m = a_{min}2^{J_m}$ where a_{min} is the smallest available scale. The symbols correspond to different values of J_m : (o) 2, (•) 3, (Δ) 4, (■) 5 and (◇) 6. (a) $\sqrt{N}e(\hat{H}_{J_m, J_{max}, 2.5})$ versus $\log_2 N$ when using the WTMM g_1 estimator ($a_{min} = 2.15$). (b) $\sqrt{N}e(\hat{H}_{J_m, J_{max}, 2})$ versus $\log_2 N$ when using the orthogonal d1 estimator ($a_{min} = 1$).

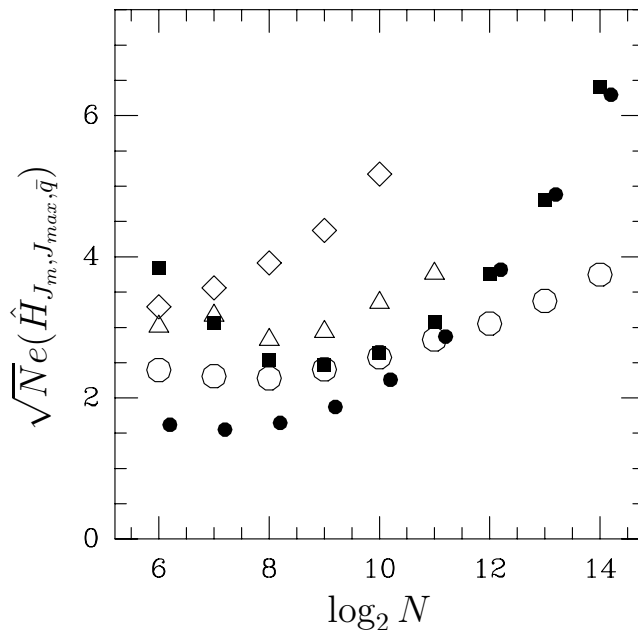
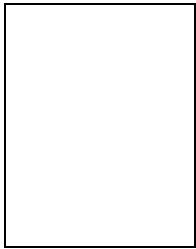


Fig. 9. Estimation of the scaling exponent of a two-valued FARIMA(0, D , 0) time-series (5000 realizations). $\sqrt{N}e(\hat{H}_{J_m, J_{max}, \bar{q}})$ versus $\log_2 N$ using the WTMM estimators (with $\bar{q} = 2.5$ and $a_m = a_{min}$) with (○) g_1 and (■) g_4 (Eq. (11)) or the (bi)orthogonal estimators ($\bar{q} = 2$ and optimized J_m) with (◇) d_1 , (△) d_9 and (●) d_{9h7} .

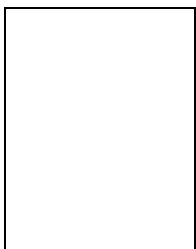
- [21] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A.L. Goldberger, "Mosaic organization of DNA nucleotides," *Phys. Rev. E*, vol. 49, pp. 1685–1689, 1994.
- [22] J.-F. Muzy, E. Bacry, and A. Arneodo, "Wavelets and multifractal formalism for singular signals: Application to turbulence data," *Phys. Rev. Lett.*, vol. 67, pp. 3515–3518, 1991.
- [23] E. Bacry, J.-F. Muzy, and A. Arneodo, "Singularity spectrum of fractal signals from wavelet analysis: exact results," *J. Stat. Phys.*, vol. 70, pp. 635–674, 1993.
- [24] P. Abry, P. Gonçalves, and P. Flandrin, *Lecture Notes in Statistics*, 105, chapter "Wavelets, spectrum analysis and $1/f$ processes", pp. 15–30, Springer Verlag, Berlin, 1995.
- [25] P. Abry, D. Veitch, and P. Flandrin, "Long-range dependence: revisiting aggregation with wavelets," *J. Time Ser. Anal.*, vol. 19, no. 3, pp. 253–266, 1998.
- [26] B. Audit, *Analyse statistique des séquences d'ADN par l'intermédiaire de la transformée en ondelettes*, Ph.D. thesis, Université de Paris VI Pierre et Marie Curie, 1999. <http://www.ebi.ac.uk/audit/PhD/>
- [27] B.B. Mandelbrot and M.S. Taqqu, "Robust R/S analysis of long-run serial correlation," *Bulletin of the International Statistical Institute*, vol. 48, pp. 69–104, 1979.
- [28] J. Geweke and S. Porter-Hudak, "The estimation and application of long-memory time series models," *J. Time Ser. Anal.*, vol. 4, pp. 221–238, 1983.
- [29] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1997.
- [30] A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J.-F. Muzy, *Ondelettes Multifractales et Turbulences: de l'ADN aux croissances cristallines*, Diderot Editeur, Arts et Sciences, Paris, 1995.
- [31] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic," *IEEE Trans. Info. Theory*, vol. 44, pp. 2–15, 1998.
- [32] A. Arneodo, B. Audit, N. Decoster, J.-F. Muzy, and C. Vaillant, *Facets of Universality in Complex Systems: Climate, Biodynamics and Stock Markets*, chapter "A wavelet-based multifractal formalism: application to DNA sequences, satellite images of the cloud structure and stock market data", Springer, Berlin, 2000, to appear.
- [33] A. Arneodo, S. Manneville, J.-F. Muzy, and S.G. Roux, "Revealing a log-normal cascading proces in turbulent velocity statistics with wavelet analysis," *Phil. Trans. R. Soc. Lond.*, vol. A 357, pp. 2415–2438, 1999.
- [34] P. Abry and F. Sellan, "The wavelet-based synthesis for fractional brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation," *Appl. Comp. Harm. Anal.*, vol. 3, pp. 377–383, 1996.
- [35] J.-F. Muzy, E. Bacry, and A. Arneodo, "The multifractal formalism revisited with wavelets," *Int. J. Bifurc. Chaos*, vol. 4, pp. 245–302, 1994.
- [36] A. Hansen, I. Simonsen and O.M. Nes, "Determination of the Hurst exponent by use of wavelet transforms," *Phys. Rev. E*, vol. 58, pp. 2779–2787, 1998.
- [37] J.-F. Muzy, E. Bacry, and A. Arneodo, "Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method," *Phys. Rev. E*, vol. 47, pp. 875–884, 1993.
- [38] A. Arneodo, E. Bacry, and J.-F. Muzy, "The thermodynamics of fractals revisited with wavelets," *Physica A*, vol. 213, pp. 232–275, 1995.
- [39] A. Arneodo, E. Bacry, P.V. Graves, and J.F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis," *Phys. Rev. Lett.*, vol. 74, pp. 3293–3296, 1995.
- [40] A. Arneodo, J.-F. Muzy, and S.G. Roux, "Experimental analysis of self-similarity and random cascade process: application to fully developed turbulence data," *J. Phys. II France*, vol. 7, pp. 363–370, 1997.
- [41] A. Arneodo, S. Manneville and J.-F. Muzy, "Towards log-normal statistics in high Reynolds number turbulence," *Eur. Phys. J. B*, vol. 1, pp. 129–140, 1998.
- [42] A. Arneodo, J.-F. Muzy, and D. Sornette, "Causal cascade in the stock market from the "infrared" to the "ultraviolet", *Eur. Phys. J. B*, , vol. 2, pp. 277–282, 1998.

- [43] A. Arneodo, F. Argoul, E. Bacry, J.-F. Muzy, and M. Tabard, "Golden mean arithmetic in the fractal branching of diffusion limited aggregates," *Phys. Rev. Lett.*, vol. 68, pp. 3456–3459, 1992.
- [44] A. Arneodo, F. Argoul, J.-F. Muzy, and M. Tabard, "Structural five-fold symmetry in the fractal morphology of diffusion-limited aggregates," *Physica A*, vol. 188, pp. 217–242, 1992.
- [45] A. Kuhn, F. Argoul, J.-F. Muzy, and A. Arneodo, "Structural analysis of electroless deposits in the diffusion-limited aggregates," *Phys. Rev. Lett.*, vol. 73, pp. 2998–3001, 1994.
- [46] A. Arneodo, N. Decoster, and S.G. Roux, "Intermittency, log-normal statistics and multifractal cascade process in high-resolution satellite images of cloud structure," *Phys. Rev. Lett.*, vol. 83, pp. 1255–1258, 1999.
- [47] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland, CA, 1976.
- [48] J.R.M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.
- [49] F. Sellan, "Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes," *C. R. Acad. Sci. Paris Sér. I*, vol. 321, pp. 351–358, 1995.
- [50] Y. Meyer, F. Sellan, and M.S. Taqqu, "Wavelets, generalized white noise and fractional integration: The synthesis of fractional brownian motion," *J. Fourier Anal. and App.*, vol. 5, pp. 466–494, 1999.
- [51] B.B. Mandelbrot and J.W. van Ness, "Fractional brownian motions, fractal noises and applications," *S.I.A.M. Rev.*, vol. 10, pp. 422–437, 1968.
- [52] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *S.I.A.M. J. of Math. Anal.*, vol. 15, pp. 723–736, 1984.
- [53] C. K. Chui, *An Introduction to Wavelets*, Academic Press, Boston, 1992.
- [54] B. Torresani, *Analyse Continue par Ondelettes*, Editions de Physique, Les Ulis, 1998.
- [55] A.H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Info. Theory*, vol. 38, pp. 904–909, 1992.
- [56] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 49, pp. 909–996, 1988.
- [57] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 11, pp. 674–693, 1989.
- [58] T. Lundhal, W.J. Ohley, S.M. Kay, and R. Siffert, "Fractional brownian motion: maximum likelihood estimator and its application to image texture," *IEEE Trans. Medical Imaging*, vol. MI-5, pp. 152–161, 1986.
- [59] G.W. Wornell and A.V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Sig. Proc.*, vol. 40, pp. 611–623, 1992.
- [60] S. Jaffard, "Pointwise smoothness, two-microlocalization and wavelet coefficients," *Publicacions Matemàtiques*, vol. 35, pp. 155–168, 1991.
- [61] M. Holschneider, *Wavelets: An Analysis Tool*, Oxford Univ. Press, Oxford, 1996.
- [62] LastWave software freely available at <http://wave.cmap.polytechnique.fr/soft/LastWave>.




Alain Arneodo Alain Arneodo got his thesis (Thèse d'Etat) in Elementary Particle Physics at the University of Nice (France) in 1978. Then he switched his scientific interest to dynamical system theory which led him to move to the Centre de Recherche Paul Pascal (CNRS) in Bordeaux (France), to collaborate with the experimental group that was working at that time on chemical chaos. Dr. Arneodo scientific contribution encompasses many fields of modern physics including statistical mechanics, dynamical system theory, chemical chaos, pattern formation in reaction-diffusion systems, fully-developed turbulence, the mathematics of fractals and multifractals, fractal growth phenomena, signal and image processing, wavelet transform analysis and its applications in physics, geophysics, chemistry, biology and finance. Dr. Arneodo has taught several courses on these topics as visiting professor at different universities including Nice, Marseille, Bordeaux, Lyon, Paris (France), Louvain-la-Neuve (Belgique), Wien (Austria), Austin (Texas), Budapest (Hungary), Wuhan (China), Montreal (Canada) and Natal (Brasil). Dr. Arneodo is a Director of Research at the CNRS (Centre National de la Recherche Scientifique) and has published extensively in the physics literature, including


over 200 peer-reviewed papers. He has trained 17 Doctors of Science. He serves on the Editorial Boards of the following journals: European Physical Journal B, International Journal of Bifurcation and Chaos, International Journal of Fractals, The Journal of Difference Equations and Applications, Applied and Computational Harmonic Analysis: Wavelets, Signal Processing and Applications. Dr. Arneodo is a fellow of the Société Française de Physique.



Benjamin Audit Benjamin Audit received the M.Eng. degree from the École Supérieure d'Électricité (Supélec), Paris, in 1995 and the M.Sc. and Ph.D. degrees from the Université Pierre et Marie Curie, Paris, in 1995 and 1999, respectively. For his PhD. BA was working in the Centre de Recherche Paul Pascal, Bordeaux, France, on the Statistical Analysis of DNA Sequences using the Wavelet Transform. He is now a postdoctoral fellow in the Computational Genomics Group, the European Bioinformatics Institute, Cambridge, UK. His research interests include structure and dynamics of chromatin, complete genome analysis and comparative genomics.



Emmanuel Bacry Emmanuel Bacry graduated from Ecole Normale Supérieure (Ulm, Paris, France) in 1990. He received the Ph.D. degree in applied mathematics from the university of Paris VII (Paris, France) in 1992 and obtained the “habilitation à diriger des recherches” from the same university in 1996. Since 1992, he is a researcher at the *Centre Nationale de Recherche Scientifique* (CNRS). After spending 4 years in the applied mathematics department of Jussieu (Paris VII), he moved, in 1996, to the *Centre de Mathématiques Appliquées* (CMAP) at Ecole Polytechnique, Paris, France. During the same year he became a part-time assistant professor at Ecole Polytechnique. His research interests include signal processing, wavelet transform, fractal and multifractal theory with applications to finance very various domains such as sound processing and finance.



Jean-François Muzy Jean-François Muzy was born in Bastia, France in 1966. He graduated his “Diplôme d’Etudes Approfondies”, in theoretical physics in 1989. He received the PhD degree in Physics from the University of Nice in 1993 and entered the CNRS the same year. He received his “Habilitation à diriger des Recherches” from the University of Bordeaux I in 2001. He worked at the Centre de Recherche Paul Pascal, CNRS UPR 8641 from 1989 to 2000. In 1998 he joined the Centre de Recherche Mathématique, Montreal University for a one year postdoctoral position. Since 2001, he works at the Laboratoire des Systèmes Physiques de l’Environnement, UMR 6134 CNRS University of Corsica. His main research interests concern the wavelet analysis of fractals and various applications ranging from fully developed turbulence to econophysics.