DFA-based abacuses providing the Hurst exponent estimate for short-memory processes

Eric Grivel, Bastien Berthelot, Pierrick Legrand, Audrey Giremus

DFA-based abacuses providing the Hurst exponent estimate for short-memory processes

Eric Grivel[a], Bastien Berthelot[b,c], Pierrick Legrand[c], Audrey Giremus[a]

[a] *Bordeaux University - INP Bordeaux - IMS - UMR CNRS 5218, FRANCE*
[b] *THALES AVS France, Campus Merignac, FRANCE*
[c] *Bordeaux University - IMB UMR CNRS 5251 - ASTRAL team, INRIA, FRANCE*

Abstract

The detrended fluctuation analysis (DFA) and its higher-order variant make it possible to estimate the Hurst exponent and therefore to quantify the long-range dependence of a random process. These methods are popular and used in a wide range of applications where they have been proven to be discriminative to characterize or classify processes. Nevertheless, in practice, the signal may be short-memory. In addition, depending on the number of samples available, there is no guarantee that these methods provide the true value of the Hurst exponent, leading the user to draw erroneous conclusions on the long-range dependence of the signal under study. In this paper, using a matrix formulation and making no approximation, we first propose to analyze how the DFA and its higher-order variant behave with respect to the number of samples available. Illustrations dealing with short-memory data that can be modeled by a white noise, a moving-average process and a random process whose autocorrelation function exponentially decays are given. Finally, to avoid any wrong conclusions, we propose to derive abacuses linking the value provided by the DFA or its variant with the properties of the signal and the number of samples available.

Keywords: DFA, higher-order DFA, Hurst coefficient, sensitivity, abacus.

1. Introduction

In many applications, the estimation of the Hurst exponent [1] can be used to characterize or classify signals. This is of real interest in different cases: biomedical applications [2–9], econometrics [10, 11], meteorology [12, 13], or geophysics [14].

Various approaches have been proposed in the literature to estimate the Hurst exponent, from the rescaled range (R/S) analysis [15] and the variance-of-residuals method [16] [17] to the analysis of the power spectral density (PSD) of the time series [18–20], passing by semi-parametric methods [21, 22]. The reader may refer to [23] for an exhaustive state of the art and details about each method. Among all these approaches, the detrended fluctuation analysis (DFA) -initially introduced by Peng *et al.* in [24] after a first approach called fluctuation analysis (FA) [25]- is one of the most popular methods to estimate the long range dependence of a time series. It is mainly based on two principles: on the one hand, the extraction of the trend[1] of what is called the profile, which is the integrated centered signal, and on the other hand, the square of the fluctuation function which can be seen as the estimation of the variance of the residual, *i.e.* the detrended profile, decimated by a factor equal to N. By setting N at different values, the variances of the resulting sequences are related to one another through a function depending on the self-similarity of the signal, leading to an estimation of the Hurst exponent. The success of the DFA can probably be explained by the fact that the method does not require advanced skills in mathematics and signal processing. In addition, its computational cost is not high as it is based on standard resolutions of least-squares (LS) optimisation issues such as the determination of a regression line. A great deal of attention has been paid to this method, leading to different variants which mainly differ by the way the trend is estimated. The first family of methods is based on an

[1] A trend can be defined as a long-term change in a time series that does not appear to be periodic. In the standard DFA, the trend is assumed to be a piecewise linear function.

a priori model of the local trends. Thus, in the higher-order DFA, the local trends are no longer considered as straight lines, *i.e.* polynomials of order 1, but as polynomials of order larger than 1. Concerning the continuous DFA (CDFA) [26], a constraint of continuity is added between the consecutive local trends. We can also cite the regularized DFA [27], based on a regularized LS criterion, and the adaptive fractal analysis (AFA) [28], the purpose of which is to combine two piecewise linear trends shifted in time from each other by several samples and obtained with a DFA-inspired method in order to get a continuous global trend. The second family includes low-pass filtering based methods. In this case, the algorithms differ in the way to design the low-pass linear filter to get the trend. It can be based on a single filter causal or not, with a finite or infinite impulse response, but with a linear phase in order to preserve the signal phase relationships in the passband. An alternative is to design a structure of a null-phase filter with a non-linear-phase filter. In [29], the authors propose to analyze one structure based on two filters in parallel whose inputs are the signal and its reversed version in time to get the detrended signal. In this family, one of the most popular methods is the detrended moving average (DMA) [30–32]. More recently, extensions of the DMA method have been proposed in the literature. The first one is labelled as the higher-order DMA [33]. If one refers to mathematical concept, it amounts to applying the locally estimated scatterplot smoothing (LOESS)-based method. In other words, a part of signal is modeled by a d^{th}-degree polynomial but only one sample of this local trend is kept to model the global trend. A sliding-window approach is then used to deduce the global trend. Its derivation with the locally weighted scatterplot smoothing-based method has been also presented in [34]. The implementation of these approaches significantly increases the computational cost, but they make it possible to bridge the gap between the filtering-based method and the one based on *a priori* model of the local trends.

Developing variants of the DFA has not been necessarily the only type of contributions that have been made for the last years. In addition to comparative studies based on sets of time series generated using different generators of frac-

tional Gaussian noises and fractional Brownian motions (see [35] for instance), many other questions have been addressed by different authors. Among them and without being exhaustive: how to derive bivariate linear regression analysis [36]? How to deal with multifractal time series [37]? How to conceive fast versions [38, 39]? How to make connections between the estimation of the variance of the detrended profile and the correlation function of the signal [23, 40, 41]? How to analyze the single-frequency responses of the DFA and some of its variants such as the centered DMA and the higher-order ones [42–44]? How to deal with non-stationarities [45, 46]? How to derive probabilistic properties of the estimation of the variance of the detrended profile for wide-sense stationary (w.s.s.) processes or not, for the DFA and eventually its variants [40, 47–51]: in these papers, some signals have been more particularly considered such as a white noise, a first-order autoregressive process and a fractional Gaussian noise process. The cases of profiles modeled by a fractional Gaussian noise or a fractional brownian motion have been also addressed. Nevertheless, the mathematical formalism used by the authors are not necessarily the same.

When the practitioners use the DFA or its higher-order variant, they do not necessarily check *a priori* if the signal under study is a short or long-memory process[2]. The users can ask themselves different questions. Among them, how to trust the result provided by the DFA or its higher-order variant, all the more so as the number of samples may be small? In this paper, we focus our attention on the use of these methods when dealing with short-memory processes. Our contributions are threefold:

1/ using a matrix formulation for each step of the methods and making no approximation, we analyze the statistical mean of the square of the fluctuation function for a white noise for the DFA and its higher-order variant. It is true that similar analyzes were made in [40] or [41]. However, their methodology was

[2]Indeed, when the sum of all the terms of the normalized autocovariance function is bounded, the process is said to be short-memory. When this condition is no longer satisfied, the signal is a long-memory process.

different. If all the approaches provide the same results for the standard DFA, the expression proposed by Höll in [40] and the one we obtain for a second-order DFA tend to coincide when N is large but are all the more different as N becomes small. This is due to the fact that our approach is based on the standard definition of the correlation function of a random process whereas Höll's analysis is based on a definition of the correlation function using a temporal mean.

2/ we analyze the accuracy of the Hurst exponent estimate based on the DFA when a finite number of samples of a wide-sense-stationary (w.s.s.) short-memory process is available. We will see that the methods give the theoretical value of the Hurst exponent if the number M of samples available tends to infinity and the values of the decimation factor N are as large as possible. Otherwise, the estimate of the Hurst exponent may be drastically different. Illustrations are given for three types of signals such as a white noise, a moving-average process and a random process whose autocorrelation function exponentially decays. Since the way to address the problem is not the same, our work is complementary to the recent studies conducted by Höll *et al.* [40, 46], the empirical analysis for 1^{st} and 2^{nd}-order AR processes presented in [52] and the paper written by Maraun *et al.* [53] dealing with the evolution of the Hurst-coefficient estimate averaged on 1000 realizations provided by the standard DFA and the higher-order DFA with respect to the duration of a synthetic long-range correlated process or a linear combination of three 1^{st}-order autoregressive (AR) processes. Finally, our study is related to the work done in [54] where the author suggested correcting the values of the fluctuation function for small values of N, as he noticed that the fluctuation function was under-evaluated.

3/ given the mathematical developments presented in this paper, abacuses linking the number of samples with the model parameters (*e.g.* the MA parameters or the parameter describing the way the correlation decays) are provided. The methodology to obtain them is presented so that it can be easily extended to other types of processes.

The remainder of this paper, which is a follow-up of the theoretical comparative study we recently made on variants of the DFA [23], is organized as follows:

in Section 2, the notion of self-similarity is briefly recalled. In Section 3, the fluctuation analysis (FA) is presented whereas Section 4 deals with the main steps of the DFA and its higher-order variant. Presenting the approaches using a matrix form, as we did in [26] [23], has the advantage of characterizing the fluctuation function by means of a matrix $\Gamma_{1,d}$. Its structure and properties are then studied in Section 5. This part will be useful for the rest of the paper. The accuracy of the Hurst exponent estimate with the DFA when dealing with a finite number of samples of short-memory processes is then addressed in Section 6. Illustrations and abacuses are given in Section 7. Conclusions and perspectives end up the paper.

2. Brief reminder of the self-similarity of a signal

A time-continuous process $y(t)$ is said to be self-similar with parameter H if and only if:

$$y(Nt) \overset{d}{=} N^H y(t) \tag{1}$$

where $y(Nt)$ is time-scaled signal by a scale factor N, $\overset{d}{=}$ means an equivalency in terms of distribution and H is the self-similarity parameter or the Hurst exponent. One also speaks of the Hurst exponent when analyzing the regularity of a signal [1]. Thus, H is equal to 0.5, 0 and -0.5 for a Brownian noise, a pink noise and a white noise, respectively. In a general manner, the Hurst exponent takes the value 0.5 for the short-memory processes that are of interest in this study [55].

However, evaluating if two processes have similar probability density functions is not necessarily straightforward. One option involves the study of divergences such as Kullback-Leibler, Jeffreys or Rényi divergences. When dealing with Gaussian processes, an alternative is to compare the means and the standard deviations of both processes. If zero-mean processes are considered, one has to compare the standard deviation $F(N)$ of $y(Nt)$ with $F(1)$, the standard deviation of $y(t)$. They are linked by:

$$F(N) = N^H F(1) \tag{2}$$

Given a set of values $\{N_i\}_i$ whose cardinal is larger or equal than 2. the slope of the regression line of the set of values $\{\log(N_i), \log(F(N_i)\}_i$ can be estimated in the LS sense.

Finally, in most of real-world applications, the processes under study are bounded. They cannot exhibit relatively large values, no matter the size of the analysis window. As a consequence, bounded processes with different patterns may lead to an identical estimation of H. One solution consists in mapping these bounded processes into self-similar time-series by integrating them. Using these integrated processes may provide a significantly greater differentiation in terms of self-similarity. It is then an essential step in the study of the latter. In this case, the integration step adds $+1$ to the estimation of H.

Among the approaches proposed in the literature, a first solution was the fluctuation analysis (FA) which is presented in the next section, with additional comments. It should be noted that, for the sake of clarity, the subscript $\mathbf{0}$ will be used in some notations to refer to this method. This choice is motivated by the fact that FA was the initial method proposed by Peng.

3. Properties of the fluctuation analysis (FA) method

To estimate the Hurst exponent of a time series, the FA was proposed early in the 90ies by Peng [25]. It operates in three steps: the signal y is first integrated. This leads to a new sequence, denoted as $y_{int,\mathbf{0}}$.

$$y_{int,\mathbf{0}}(m) = \sum_{i=1}^{m} y(i) \tag{3}$$

Then, what can be called the fluctuation function $F_{\mathbf{0}}(N)$ is computed for different values of the lag N, as follows:

$$F_{\mathbf{0}}(N) = \sqrt{< (y_{int,\mathbf{0}}(i+N) - y_{int,\mathbf{0}}(i))^2 >} \tag{4}$$

where $< . >$ is the temporal mean.

Finally, as $F_{\mathbf{0}}(N)$ is proportional to N^{H+1}, $\log(F_{\mathbf{0}}(N))$ is represented as an affine function of $\log(N)$ in order to estimate H in the LS sense.

Let us deduce the mean of the square of the fluctuation function. To this end, the definition (4) of $F_{\mathbf{0}}(N)$ can be rewritten as follows:

$$F_{\mathbf{0}}^2(N) = \left\langle \left(\sum_{j=i+1}^{i+N} y(j) \right)^2 \right\rangle = \left\langle \sum_{j=i+1}^{i+N} y^2(j) + 2 \sum_{j=i+1}^{i+N-1} \sum_{k=j+1}^{i+N} y(j)y(k) \right\rangle \quad (5)$$
$$= \left\langle \sum_{j=i+1}^{i+N} y^2(j) + 2 \sum_{r=1}^{N-1} \sum_{j=i+1}^{i+N-r} y(j)y(j+r) \right\rangle$$

When dealing with a w.s.s. signal, the expectation of $F_{\mathbf{0}}^2(N)$ is expressed from the autocorrelation function $R_{y,y}(\tau)$ of the signal in the following manner:

$$E[F_{\mathbf{0}}^2(N)] = \sum_{r=1-N}^{N-1} \Big(Tr(\Gamma_{\mathbf{0}}, r) \Big) R_{y,y}(r) = \sum_{r=1-N}^{N-1} (N - |r|) R_{y,y}(r) \quad (6)$$

where $E[.]$ is the expectation, $\Gamma_{\mathbf{0}}$ is a square matrix of size N whose every element is equal to 1 and $Tr([.], r)$ denotes the r^{th} diagonal of the matrix $[.]$. Based on the Wiener-Khintchine theorem and by introducing the inverse Fourier transform (FT^{-1}) and $S_{yy}(f)$ the power spectral density (PSD) of y, $E[F_{\mathbf{0}}^2(N)]$ can be rewritten as follows:

$$E[F_{\mathbf{0}}^2(N)] = FT^{-1} \left(\left(\sum_{r=1-N}^{N-1} Tr(\Gamma_{\mathbf{0}}, r) e^{-j2\pi f_n r} \right) S_{yy}(f) \right) \Big|_{\tau=0}$$
$$= FT^{-1} \Big(\Psi_{\mathbf{0}}(f) S_{yy}(f) \Big) \Big|_{\tau=0} \quad (7)$$

where τ is the time variable of the inverse Fourier transform and f_n is the normalized frequency. In this case, $\Psi_{\mathbf{0}}(f) S_{yy}(f)$ corresponds to the PSD of the signal y filtered by a filter whose transfer function $H_{\mathbf{0}}(z) = \sum_n h_{\mathbf{0}}(n) z^{-n} = \sum_{n=0}^{N-1} z^{-n}$ satisfies:

$$\Psi_{\mathbf{0}}(f) = |H_{\mathbf{0}}(z)|^2_{z=exp(j\theta)} \quad (8)$$

with $\theta = 2\pi f / f_s$ the normalized angular frequency.

As a consequence, $E[F_{\mathbf{0}}^2(N)]$ is the power of the filter output. $F_{\mathbf{0}}(N)$ can be seen as an approximation of the standard deviation $F(N)$, under some implicit assumptions made by Peng that are introduced in the previous section.

Since the FA was sensitive to signals exhibiting fluctuations around trends varying over time, the detrended fluctuation analysis (DFA) was developed in [24].

Then, its variant of order d was proposed. They are presented in the next section. They consist in estimating the trend of the integrated centered process, and then analyzing the self-affinity of the process when the trend has been removed. In the standard version of the DFA, the trend is defined as a piecewise linear approximation of the integrated centered process whereas a polynomial of degree $d > 1$ is considered for a higher-order DFA. Finally, the subscript $_{\mathbf{1,}\boldsymbol{d}}$ may be used in some notations to make the difference with the ones used with the FA.

4. About the detendred fluctuation analysis (DFA) and its higher-order variant

In this section, the DFA and its higher-order variant are presented by using a matrix form, using the same formalism as the one proposed in [27] [23].

4.1. Notations

• $\mathbb{1}_{j\times k}$ and $\mathbf{0}_{j\times k}$ are matrices of size $j \times k$ filled with 1s and 0s, respectively.

• $diag([.], l)$ is a matrix whose elements are null except the l^{th} diagonal which is equal to $[.]$. Thus, $I_j = diag(\mathbb{1}_{1\times j}, 0)$ is the identity matrix of size j. $diag(\mathbb{1}_{1\times N-1}, 1)$ is the square matrix of size N whose 1^{st} sub-diagonal above the main one has its elements equal to 1.

• $J_j = I_j - \frac{1}{j}\mathbb{1}_{j\times j}$.

• T_l is a $N \times 1$ vector storing the values of the l^{th} local trend $t_l(n)$.

• Y and Y_{int} are two column vectors storing respectively the samples $\{y(n)\}_{n=1,...,M}$ and $\{y_{int}(n)\}_{n=1,...,M}$ that are related as follows:

$$Y_{int} = [y_{int}(1), y_{int}(2), ..., y_{int}(M)]^T = \mathrm{H}_M J_M Y \tag{9}$$

with $H_M = \sum_{r=0}^{M-1} diag(\mathbb{1}_{1\times M-r}, -r)$ a lower triangular matrix filled with 1s.

• $A(i : j, k : l)$ is the part of the matrix A corresponding to the elements belonging to the rows i to j and to the columns k to l .

• Given the following matrix of size (j, M):

$$C_{j,k} = [\mathbf{0}_{j\times k} \quad I_j \quad \mathbf{0}_{j\times(M-(j+k))}] \tag{10}$$

the first LN elements of the vector Y_{int} can be expressed as follows:

$$Y_{int}(1:LN) = [y_{int}(1), y_{int}(2), ..., y_{int}(LN)]^T = C_{LN,0}Y_{int} \underset{(9)}{=} C_{LN,0}\mathrm{H}_M J_M Y \tag{11}$$

4.2. Steps of the DFA and its higher-order variant with some comments

The DFA and its higher-order variant operate with following steps when M consecutive samples $\{y(m)\}_{m=1,...,M}$ are available.

- **Step 1:** The profile y_{int} is obtained by centering and integrating y:

$$y_{int}(m) = \sum_{i=1}^{m}(y(i) - \mu_y) \tag{12}$$

  where $\mu_y = \frac{1}{M}\sum_{m=1}^{M} y(m)$ denotes the mean of y.

- **Step 2:** The trend of the profile is estimated by splitting the profile into L non-overlapping segments of length N, denoted as $\{y_{int,l}(n)\}_{l=1,...,L}$ with $n \in [\![1; N]\!]$. This means that only the first LN samples of the sequences are considered. Using a vector form, the d^{th}-degree polynomial trend of the l^{th} segment can be expressed as:

$$\mathrm{T}_{loc,l,d} = A_{l,d}\theta_{l,d} \ \forall l \in [\![1; L]\!] \tag{13}$$

  where the parameter vector is $\theta_{l,d} = \left[a_{l,0}\ a_{l,1}\ ...\ a_{l,d}\right]^T$. When $d = 1$, the local trend is a linear function, $a_{l,1}$ being the local slope and $a_{l,0}$ the vertical intercept. In addition, $A_{l,d}$ is a $N \times d$ matrix whose d^{th} column is defined by the set of values $\{[(l-1)N + n]^d\}_{n=1,...,N}$.
  Given the parameter vector $\Theta_{\mathbf{1},d} = \left[\theta_{1,d} \dots \theta_{L,d}\right]^T$ of size $dL \times 1$, and the $(LN \times dL)$ matrix $A_{\mathbf{1},d}$ which is block diagonal defined from the set of matrices $\{A_{l,d}\}_{l=1,...,L}$, the parameters of the local trends can be deduced by minimizing $\left|\left|C_{LN,0}Y_{int} - A_{\mathbf{1},d}\Theta_{\mathbf{1},d}\right|\right|^2$. For $d > 0$, this leads to:

$$\hat{\Theta}_{\mathbf{1},d} = (A_{\mathbf{1},d}^T A_{\mathbf{1},d})^{-1} A_{\mathbf{1},d}^T C_{LN,0} Y_{int} \tag{14}$$

Therefore, combining (9) and (14), the global trend vector $T_{\mathbf{1},d}$ is given by:

$$T_{\mathbf{1},d} = A_{\mathbf{1},d}\hat{\Theta}_{\mathbf{1}} = A_{\mathbf{1},d}(A_{\mathbf{1},d}^T A_{\mathbf{1},d})^{-1} A_{\mathbf{1},d}^T C_{LN,0} \mathrm{H}_M J_M Y \tag{15}$$

Remark: When dealing with the zero-th order DFA, the trend vector is given by:

$$\begin{aligned} T_{\mathbf{1},0} &= \frac{1}{N} \begin{bmatrix} \mathbb{1}_{N\times N} & \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} \\ 0_N & \ddots & \ddots & \mathbf{0}_{N\times N} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{N\times N} & \dots & \mathbf{0}_{N\times N} & \mathbb{1}_{N\times N} \end{bmatrix} C_{LN,0} \mathrm{H}_M J_M Y \\ &= \frac{1}{N}(I_L \otimes I_N) \mathrm{H}_M J_M Y \end{aligned} \tag{16}$$

where $\otimes$ denotes the Kronecker product.

- **Step 3:**. Given (15), the residual vector $R_{\mathbf{1},d}$, *i.e.* the expression of the difference between the profile and its trend, can be expressed from the signal vector Y:

$$R_{\mathbf{1},d} = C_{LN,0} \mathrm{H}_M J_M \mathrm{Y} - T_{\mathbf{1},d} = B_{\mathbf{1},d} Y \tag{17}$$

where $d > 0$:

$$B_{\mathbf{1},d} = \big(I_{LN} - A_{\mathbf{1},d}(A_{\mathbf{1},d}^T A_{\mathbf{1},d})^{-1} A_{\mathbf{1},d}^T\big) C_{LN,0} \mathrm{H}_M J_M \tag{18}$$

Remark: When dealing with the zero-th order DFA, one has:

$$B_{\mathbf{1},0} = (I_L \otimes J_N) C_{LN,0} \mathrm{H}_M J_M \tag{19}$$

Taking into account the properties of self-similarity presented in Section 2, the next step is to estimate the standard deviation of the residual $y_{int}(Nn) - t(Nn)$. To this end, N subsequences of size L are obtained by decimating the residual by a factor N. The first one starts with the first sample of $R_{\mathbf{1},d}$, the second one with the second sample of $R_{\mathbf{1},d}$, and so on. Assuming a statistical mean equal to 0, the variances of the resulting

sub-sequences are estimated and their average is computed. By reorganizing the terms, this leads to the definition of the square of the fluctuation function:

$$F^2_{\mathbf{1},d}(N) = \frac{1}{LN}\sum_{n=1}^{LN}(y_{int}(n) - t(n))^2 \tag{20}$$

Once again, the fluctuation function $F_{\mathbf{1},d}(N)$ is an approximation of the standard deviation $F(N)$, under some assumptions made by Peng. From a signal processing point of view, the square of the fluctuation function also corresponds to the power of the residual. Then, introducing the matrix $\Gamma_{\mathbf{1},d} = \frac{1}{LN}B^T_{\mathbf{1},d}B_{\mathbf{1},d}$ of size $M \times M$ and the trace operator, $F^2_{\mathbf{1},d}(N)$ can be rewritten as follows:

$$F^2_{\mathbf{1},d}(N) = Tr\left(\Gamma_{\mathbf{1},d}YY^T\right) \tag{21}$$

- **Step 4:** Steps 2 and 3 are repeated for $l_{N_{1:max}}$ different values of N, *e.g.* N_1, N_2, ..., N_{max} usually selected in the interval $[3; \lfloor M/2 \rfloor]$ where $\lfloor\rfloor$ is the floor function.

- **Step 5:** As mentioned in Section 2, the final step is to search a straight line fitting the log-log representation of $F_{\mathbf{1},d}(N)$ with regards to N. Its slope α, which is no longer equal to H but to $= H + 1$ due to the integration, is estimated in the LS sense. Each pair $(\log(F_{\mathbf{1},d}(N)), \log(N))$ satisfies the following equation:

$$\log(F_{\mathbf{1},d}(N)) = b + \alpha\log(N) + \epsilon(N), \quad \text{for } N = N_1, ..., N_{max} \tag{22}$$

with b the vertical intercept and $\epsilon(N)$ the error variable.

In order to write (22) in a matrix form, let us introduce the column vector $\underline{F}_{\log,\mathbf{1},d}$ of length $l_{N_{1:max}}$ storing the values of $\log(F_{\mathbf{1},d}(N))$ and the matrix $A_{\log}$ of size $(l_{N_{1:max}} \times 2)$, whose first column is a vector of 1s and whose second column is equal to the values of $\log(N)$. In addition, $\underline{\eta}$ is the column vector storing the parameters b and α and $\underline{\epsilon}$ is the column vector of length $l_{N_{1:max}}$ storing the different errors $\{\epsilon(N)\}$. Thus, one has:

$$\underline{F}_{\log,\mathbf{1},d} = A_{\log}\underline{\eta} + \underline{\epsilon} \tag{23}$$

As a consequence, the LS estimation of $\underline{\eta}$ can be written as:

$$\underline{\hat{\eta}} = (A_{\log}^T A_{\log})^{-1} A_{\log}^T \underline{F}_{\log,\mathbf{1},d} \tag{24}$$

Then, the estimation of α can be deduced since $\alpha = [0\ 1]\underline{\eta}$.

By expressing the inverse of the 2×2 matrix $A_{\log}^T A_{\log}$ in (24), developing and simplifying, the estimation of the slope α can be expressed as follows[3]:

$$\hat{\alpha}_{\mathbf{1},d}(N_1, ..., N_{max}) = \frac{\overline{\log(N)\log(F_{\mathbf{1}}(N))} - \overline{\log(N)} \times \overline{\log(F_{\mathbf{1}}(N)}}{\overline{\log^2(N)} - \overline{\log(N)}^2} \tag{25}$$

where $\overline{\log(N)}$ is the mean of $\log(N)$ computed for the different values of N that have been chosen, *i.e.* $N_1, N_2, ..., N_{max}$. For instance, consecutive values of N could be considered. It is true that this estimation of the slope depends on the values of N that have been selected. The larger M is, the larger N_{max} can be. This is the reason why we have chosen to point out this dependence in the above equation by writing $\hat{\alpha}_{\mathbf{1}}(N_1, ..., N_{max})$. This will be useful in the remainder of the paper. Moreover, the result obtained in (25) is consistent with the result in statistics where $\log(F_{\mathbf{1},d}(N))$ and $\log(N)$ are considered as two random variables and $\log(F_{\mathbf{1},d}(N))$ is assumed to be an affine function of $\log(N)$. In this case, α satisfies:

$$\begin{aligned} \alpha(N_1, ..., N_{max}) &= \frac{cov(\log(F_{\mathbf{1},d}(N)), \log(N))}{var(\log(N))} \\ &= \frac{E[\log(F_{\mathbf{1},d}(N))\log(N)] - E[\log(F_{\mathbf{1},d}(N))]E[\log(N)]}{E[\log(N)^2] - E^2[\log(N)]} \end{aligned} \tag{26}$$

where *cov* and *var* respectively denote the cross-covariance and the variance.

As one could expect, the estimation of α by using the DFA is a non linear function of the signal due to the presence of the logarithm function in the vector $\underline{F}_{log,\mathbf{1},d}$.

Depending on the applications, the number M of samples that are available may vary. It can take small or large values, *e.g.* from a few dozens to a few

[3]The same expression could be obtained with the FA. This would amount replacing the subscript $_{\mathbf{1},d}$ by $_{\mathbf{0}}$ in the equation.

thousands. In the next sections, we propose to analyze the properties of the matrix $\mathbf{\Gamma_{1,}}_d$ in order to derive statistical properties of the square of the fluctuation functions. This will be the opportunity to deduce some properties on the slope α with respect to $N_{max} \leq \lfloor {}^{M}/_{2} \rfloor$.

5. Structure and properties of the matrix $\mathbf{\Gamma_{1,}}_{\boldsymbol{d}}$

5.1. Preamble

In (17)-(18), for $d > 0$, the matrix $I_{LN} - A_{\mathbf{1},d}(A_{\mathbf{1},d}^T A_{\mathbf{1},d})^{-1} A_{\mathbf{1},d}^T$ is block diagonal. Each block is defined by $I_N - A_{l,d}(A_{l,d}^T A_{l,d})^{-1} A_{l,d}^T$ with $l = 1, ..., L$, but the latter are all equal. Indeed, $A_{l,d}(A_{l,d}^T A_{l,d})^{-1} A_{l,d}^T$ corresponds the orthogonal projector onto the space spanned by the columns of $A_{l,d}$. However, the 1^{st} column of $A_{l,d}$ is equal to the 1^{st} column of $A_{1,d}$. The 2^{nd} column of $A_{l,d}$ is a linear combination of the 1^{st} and 2^{nd} columns of $A_{1,d}$. Then, when $d > 1$, as the third column of $A_{l,d}$ is a linear combination of the first two columns of $A_{l,d}$, it is a linear combination of the first two columns of $A_{1,d}$. More generally, for $c \leq d$, one can easily show that the c^{th} column of $A_{l,d}$ is a linear combination of the first $(c-1)$ columns of $A_{l,d}$. Consequently, it is a linear combination of the first two columns of $A_{1,d}$. So, for $l = 1, ..., L$, one has:

$$\mathbf{A_{1,d}} = I_N - A_{l,d}(A_{l,d}^T A_{l,d})^{-1} A_{l,d}^T = I_N - A_{1,d}(A_{1,d}^T A_{1,d})^{-1} A_{1,d}^T \tag{27}$$

Remark: When dealing with the zero-th order DFA, one has:

$$\mathbf{A_{1,0}} = J_N \tag{28}$$

The above matrix $\mathbf{A_{1,d}}$ with $d > 0$ has four main properties that will be used in the following:

1. $\mathbf{A_{1,d}} = \mathbf{A_{1,d}}^T$.
2. The space spanned by the columns of $\mathbb{1}_{N \times N}$ corresponds to the space spanned by the first column of $\mathbf{A_{1,d}}$. Therefore, $\mathbf{A_{1,d}}\mathbb{1}_{N \times N}$ is the null matrix.

3. The space spanned by the columns of $\mathrm{H}_N \mathbb{1}_{N\times N}$ corresponds to the space spanned by the two columns of $A_{1,d}$. Therefore, $\mathbf{A_{1,d}}\mathrm{H}_N \mathbb{1}_{N\times N}$ is also the null matrix.
4. By combining the first properties, it can be shown that $\mathbb{1}_{N\times N}\mathbf{A_{1,d}}$ and $\mathbb{1}_{N\times N}\mathrm{H}_N^T\mathbf{A_{1,d}}$ are null matrices.

5.2. Structure of the matrix $\Gamma_{\mathbf{1},d}$

First of all, let us look at the zero-th order DFA, one has:

$$\begin{aligned}\Gamma_{\mathbf{1},0} &= \frac{1}{LN}\mathrm{H}_M^T C_{LN,0}^T (I_L \otimes J_N)^T (I_L \otimes J_N) C_{LN,0}\mathrm{H}_M \\ &= \frac{1}{LN}\mathrm{H}_M^T C_{LN,0}^T (I_L \otimes J_N^T J_N) C_{LN,0}\mathrm{H}_M \\ &= \frac{1}{LN}\mathrm{H}_M^T C_{LN,0}^T (I_L \otimes J_N) C_{LN,0}\mathrm{H}_M\end{aligned} \tag{29}$$

In what follows, our purpose is to give a detailed expression of $\Gamma_{\mathbf{1},d}$, by highlighting its structure. To this end, using the properties introduced in the Section 5.1 and what is written above, $\Gamma_{\mathbf{1},d}$ can expressed as follows for $d \geq 0$:

$$\begin{aligned}\Gamma_{\mathbf{1},d} &= \frac{\mathbf{1}}{LN}\mathrm{H}_M^T C_{LN,0}^T (I_L \otimes \mathbf{A_{1,d}})^T (I_L \otimes \mathbf{A_{1,d}}) C_{LN,0}\mathrm{H}_M \\ &= \frac{1}{LN}\mathrm{H}_M^T C_{LN,0}^T (I_L \otimes \mathbf{A_{1,d}}) C_{LN,0}\mathrm{H}_M\end{aligned} \tag{30}$$

In (30), $\Gamma_{\mathbf{1},d}$ depends on $C_{LN,0}\mathrm{H}_M$. This matrix is of size $LN \times M$ whose many parts are null matrices:

$$C_{LN,0}\mathrm{H}_M = \begin{bmatrix} H_N & 0_N & 0_N & \dots & 0_{N\times(M-LN)} \\ 1_N & H_N & 0_N & \dots & 0_{N\times(M-LN)} \\ \vdots & & \ddots & \ddots & \vdots \\ 1_N & \dots & 1_N & H_N & 0_{N\times(M-LN)} \end{bmatrix} \tag{31}$$

Due to the properties of $\mathbf{A_{1,d}}$ given in the Section 5.1, one can deduce:

$$(I_L \otimes \mathbf{A_{1,d}}) C_{LN,0}\mathrm{H}_M = \begin{bmatrix} \mathbf{A_{1,d}}H_N & 0_N & 0_N & \dots & 0_{N\times(M-LN)} \\ 0_N & \mathbf{A_{1,d}}H_N & 0_N & \dots & 0_{N\times(M-LN)} \\ \vdots & & \ddots & \ddots & \vdots \\ 0_N & \dots & 0_N & \mathbf{A_{1,d}}H_N & 0_{N\times(M-LN)} \end{bmatrix} \tag{32}$$

Finally, given (30), it can be shown that $\Gamma_{\mathbf{1},d}$ is a block-diagonal matrix. By introducing the square matrix $\gamma_{\mathbf{1},d}$ of size N defined by:

$$\gamma_{\mathbf{1},d} = \frac{1}{N} H_N^T {\mathbf{A_{1,d}}}^T \mathbf{A_{1,d}} H_N \tag{33}$$

the first L blocks of $\Gamma_{\mathbf{1},d}$ are defined by $\frac{1}{L}\gamma_{\mathbf{1},d}$ while the last of size $(M - LN) \times (M - LN)$ is a matrix of zeros.

$$\Gamma_{\mathbf{1},d} = \begin{bmatrix} \frac{1}{L}\gamma_{\mathbf{1},d} & 0_N & 0_N & \dots & 0_{N\times(M-LN)} \\ 0_N & \frac{1}{L}\gamma_{\mathbf{1},d} & 0_N & \dots & 0_{N\times(M-LN)} \\ \vdots & & \ddots & \ddots & \vdots \\ 0_N & \dots & 0_N & \frac{1}{L}\gamma_{\mathbf{1},d} & 0_{N\times(M-LN)} \\ 0_{(M-LN)\times N} & \dots & \dots & 0_{(M-LN)\times N} & 0_{(M-LN)\times(M-LN)} \end{bmatrix} \tag{34}$$

5.3. Properties of the matrix $\Gamma_{\mathbf{1},d}$, for $d \geq 0$

In Section 5.2, and more particularly in (33), we saw that $\Gamma_{\mathbf{1},d}$ could be expressed from $\gamma_{\mathbf{1},d}$. The latter depends on $\mathbf{A_{1,d}}H_N$. Our first goal in this subsection is to give an analytic expression of each element of the matrix $\mathbf{A_{1,d}}H_N$ for $d = 0, ..., 3$. This intermediary result will be useful to express $Tr(\Gamma_{\mathbf{1},d}, r)$.

Case $d = 0$, zero-th order DFA: In this particular case, one has:

$$\mathbf{A_{1,0}}H_N = J_N H_N \tag{35}$$

The element of the $N \times N$ matrix $\mathbf{A_{1,0}} = J_N$ located at the r^{th} row and the c^{th} column is given by:

$$\mathbf{A_{1,0}}(r, c, N) = \delta_{c,r} - \frac{1}{N} \tag{36}$$

where $\delta_{c,r} = 1$ if $c = r$ and 0 elsewhere. In the notation above, we have added the dependency on the size N explicitly as this will be useful in the following. As a consequence, one has for $c > 1$:

$$\mathbf{A_{1,0}}H_N(r, c) = \sum_{c_1=c}^{N} \mathbf{A_{1,0}}(r, c_1) \tag{37}$$

Case $d = 1$, standard DFA: As the first column of H_N is a column of 1s, the first column of $\mathbf{A_{1,1}}H_N$ is a column of zeros due to the properties mentioned in subsection 5.1.

To define the other columns of $\mathbf{A_{1,1}}H_N$, the reasoning is the following. First, it can be shown that:

$$A_{1,1}^T A_{1,1} = \begin{bmatrix} N & \sum_{n=1}^N n \\ \sum_{n=1}^N n & \sum_{n=1}^N n^2 \end{bmatrix} = \begin{bmatrix} N & \frac{N(N+1)}{2} \\ \frac{N(N+1)}{2} & \frac{N(N+1)(2N+1)}{6} \end{bmatrix} \tag{38}$$

Therefore, one has:

$$(A_{1,1}^T A_{1,1})^{-1} = \frac{12}{N^2(N^2-1)} \begin{bmatrix} \frac{N(N+1)(2N+1)}{6} & -\frac{N(N+1)}{2} \\ -\frac{N(N+1)}{2} & N \end{bmatrix} \tag{39}$$

After premultiplying $(A_{1,1}^T A_{1,1})^{-1}$ by $A_{1,1}$ and postmultiplying by $A_{1,1}^T$, the element of the $N \times N$ matrix $\mathbf{A_{1,1}} = I_N - A_{1,1}(A_{1,1}^T A_{1,1})^{-1} A_{1,1}^T$ located at the r^{th} row and the c^{th} column is given by:

$$\mathbf{A_{1,1}}(r,c,N) = \delta_{c,r} - \frac{12}{N(N^2-1)} \Big[\frac{N+1}{2} \Big(\frac{2N+1}{3} - (r+c) \Big) + rc \Big] \tag{40}$$

In the notation above, we have added the dependency on the size N explicitly as this will be useful in the following.

As a consequence, one has for $c > 1$:

$$\mathbf{A_{1,1}}H_N(r,c) = \sum_{c_1=c}^{N} \mathbf{A_{1,1}}(r,c_1) \tag{41}$$

Case $d = 2$, 2^{nd}order DFA: As done in the previous case, let us first look at the expression of $A_{1,2}^T A_{1,2}$:

$$A_{1,2}^T A_{1,2} = \begin{bmatrix} N & \sum_{n=1}^N n & \sum_{n=1}^N n^2 \\ \sum_{n=1}^N n & \sum_{n=1}^N n^2 & \sum_{n=1}^N n^3 \\ \sum_{n=1}^N n^2 & \sum_{n=1}^N n^3 & \sum_{n=1}^N n^4 \end{bmatrix} \tag{42}$$

It is a 3×3 Hankel matrix whose antidiagonals are defined by $\sum_{n=1}^N n^l$, with $l = 0, ..., 4$. After development and simplification, this leads to:

$$(A_{1,2}^T A_{1,2})^{-1} = \frac{3}{N(N-1)(N-2)} \times \tag{43}$$
$$\begin{bmatrix} 3\,N^2 + 3\,N + 2 & -6(2\,N+1) & 10 \\ -6(2\,N+1) & 4\,\frac{(8\,N+11)(2\,N+1)}{(N+2)(N+1)} & -\frac{60}{N+2} \\ 10 & -\frac{60}{N+2} & \frac{60}{(N+2)(N+1)} \end{bmatrix}$$

The element of $I_N - A_{1,2}(A_{1,2}^T A_{1,2})^{-1}A_{1,2}^T$ located at the r^{th} row and the c^{th} column is given by:

$$\begin{aligned}\mathbf{A_{1,2}}(r,c,N) = \delta_{c,r} - \frac{3}{N(N-1)(N-2)}\Big[&3N^2+3N+2-6r(2N+1)+10r^2 \\ &+ c(-6(2N+1)+4\frac{r(8N+11)(2N+1)}{(N+2)(N+1)} - 60\frac{r^2}{N+2}) \\ &+ c^2(10 - \frac{60r}{N+2} + \frac{60r^2}{(N+2)(N+1)})\Big]\end{aligned} \tag{44}$$

Case $d = 3$, 3^{rd}-order DFA: First, it can be shown that:

$$A_{1,3}^T A_{1,3} = \begin{bmatrix} N & \sum_{n=1}^N n & \sum_{n=1}^N n^2 & \sum_{n=1}^N n^3 \\ \sum_{n=1}^N n & \sum_{n=1}^N n^2 & \sum_{n=1}^N n^3 & \sum_{n=1}^N n^4 \\ \sum_{n=1}^N n^2 & \sum_{n=1}^N n^3 & \sum_{n=1}^N n^4 & \sum_{n=1}^N n^5 \\ \sum_{n=1}^N n^3 & \sum_{n=1}^N n^4 & \sum_{n=1}^N n^5 & \sum_{n=1}^N n^6 \end{bmatrix} \tag{45}$$

It is a 4×4 Hankel matrix with $A_{1,3}^T A_{1,3}(2,1) = \frac{N(N+1)}{2}$, $A_{1,3}^T A_{1,3}(3,1) = \frac{N^2(N+1)^2}{4}$, $A_{1,3}^T A_{1,3}(4,1) = \frac{N^2(N+1)^2}{4}$, $A_{1,3}^T A_{1,3}(4,2) = \frac{N(N+1)(2N+1)(3N^2+3N-1)}{30}$, $A_{1,3}^T A_{1,3}(4,3) = \frac{N^2(N+1)^2(2N^2+2N-1)}{12}$ and $A_{1,3}^T A_{1,3}(4,4) = \frac{N(N+1)(2N+1)(3N^4+6N^3-3N+1)}{42}$.

Following the same reasoning, one has:

$$(A_{1,3}^T A_{1,3})^{-1} = \tag{46}$$

$$\begin{bmatrix} 8\frac{2N^3+3N^2+7N+3}{NB_{1,3}} & -20\frac{6N^2+6N+5}{NB_{1,3}} & 120\frac{2N+1}{NB_{1,3}} & -\frac{140}{NB_{1,3}} \\ -20\frac{6N^2+6N+5}{NB_{1,3}} & 200\frac{6N^4+27N^3+42N^2+30N+11}{NC_{1,3}} & -300\frac{9N^2+21N+10}{ND_{1,3}} & 280\frac{6N^2+15N+11}{NC_{1,3}} \\ 120\frac{2N+1}{NB_{1,3}} & -300\frac{9N^2+21N+10}{ND_{1,3}} & 360\frac{18N^2+35N+13}{NC_{1,3}} & -\frac{4200}{ND_{1,3}} \\ -\frac{140}{NB_{1,3}} & 280\frac{6N^2+15N+11}{NC_{1,3}} & -\frac{4200}{ND_{1,3}} & \frac{2800}{NC_{1,3}} \end{bmatrix}$$

with

$$B_{1,3} = N^3 - 6N^2 + 11N - 6 = (N-1)(N-2)(N-3) \tag{47}$$

$$C_{1,3} = N^6 - 14N^4 + 49N^2 - 36 \tag{48}$$

$$\begin{aligned} D_{1,3} &= N^5 - N^4 - 13N^3 + 13N^2 + 36N - 36 \\ &= (N-1)(N-2)(N-3)(N+2)(N+3) \end{aligned} \tag{49}$$

The element of $I_N - A_{1,3}(A_{1,3}^T A_{1,3})^{-1}A_{1,3}^T$ located at the r^{th} row and the c^{th} column then satisfies:

$$\mathbf{A_{1,3}}(r,c) = \delta_{c,r}$$
$$- \Big[8\frac{2N^3+3N^2+7N+3}{NB_{1,3}} - 20\frac{r(6N^2+6N+5)}{NB_{1,3}} + 120\frac{r^2(2N+1)}{NB_{1,3}} - 140\frac{r^3}{NB_{1,3}}$$
$$+ \frac{c}{N}\big(-20\frac{6N^2+6N+5}{B_{1,3}} + 200\frac{r(6N^4+27N^3+42N^2+30N+11)}{C_{1,3}}$$
$$- 300\frac{r^2(9N^2+21N+10)}{D_{1,3}} + 280\frac{r^3(6N^2+15N+11)}{C_{1,3}}\big)$$
$$+ \frac{c^2}{N}\big(120\frac{2N+1}{B_{1,3}} - 300\frac{r(9N^2+21N+10)}{D_{1,3}} + 360\frac{r^2(18N^2+35N+13)}{C_{1,3}} - 4200\frac{r^3}{D_{1,3}}\big)$$
$$+ \frac{c^3}{N}\big(-\frac{140}{B_{1,3}} + 280\frac{r(6N^2+15N+11)}{C_{1,3}} - 4200\frac{r^2}{D_{1,3}} + 2800\frac{r^3}{C_{1,3}}\big)\Big] \quad (50)$$

Given the above mathematical development and the properties of $\mathbf{A_{1,d}}H_N$, let us comment on the block $\frac{1}{L}\gamma_{\mathbf{1},d}$ which is repeated L times on the main diagonal of the matrix of $\Gamma_{\mathbf{1},d}$. Due to the structure of $\mathbf{A_{1,d}}H_N$, the elements of its first row and first column are all equal to 0. Then, each element can be easily deduced by taking advantage of (36) when $d = 0$, (40) and (41) when $d = 1$, (44) when $d = 2$ and (50) when $d = 3$. Therefore, when computing the trace of $\Gamma_{\mathbf{1},d}$, it corresponds to $\frac{1}{N}Tr(H_N^T {\mathbf{A_{1,d}}}^T \mathbf{A_{1,d}} H_N)$. It does not depend on L. More generally, for $r = 0, ..., N-1$, one has:

$$Tr(\Gamma_{\mathbf{1},d}, r) = \frac{1}{N}Tr(H_N^T {\mathbf{A_{1,d}}}^T \mathbf{A_{1,d}} H_N, r) = Tr(\gamma_{\mathbf{1},d}, r) \quad (51)$$

$Tr(\Gamma_{\mathbf{1},d}, r)$ does not depend on L. In addition, when $r > N-1$, the trace is null.

5.4. Asymptotic properties of the matrix $\Gamma_{\mathbf{1},d}$ when N increases

When N increases, one has:

$$\lim_{N\to+\infty} \mathbf{A_{1,d}}(r,c,N) = \delta_{c,r} \quad (52)$$

or equivalently:

$$\lim_{N\to+\infty} \mathbf{A_{1,d}} = I_N \quad (53)$$

It should be noted that the Q-convergence can be studied. Thus, for $d = 1$ one can show that $\mathbf{A_{1,1}}(r,c,N)$ logarithmically converges to $\delta_{c,r}$. Indeed, as

$$\lim_{N\to+\infty} \frac{\mathbf{A_{1,1}}(r,c,N+1) - \delta_{c,r}}{\mathbf{A_{1,1}}(r,c,N) - \delta_{c,r}} = 1 \quad (54)$$

the convergence is sublinear. In addition, one has:

$$\lim_{N\to+\infty} \frac{\mathbf{A_{1,1}}(r,c,N+2) - \mathbf{A_{1,1}}(r,c,N+1)}{\mathbf{A_{1,1}}(r,c,N+1) - \mathbf{A_{1,1}}(r,c,N)} = 1 \tag{55}$$

When $d > 0$, the expression of the trace becomes:

$$\lim_{N\to+\infty} \frac{1}{N} Tr(H_N^T \mathbf{A_{1,d}}^T \mathbf{A_{1,d}} H_N, r) = \lim_{N\to+\infty} \frac{1}{N} Tr(H_N^T H_N, r) \tag{56}$$

As $H_N^T H_N = \begin{bmatrix} N & N-1 & \dots & 2 & 1 \\ N-1 & N-1 & \dots & 2 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 2 & 2 & \dots & 2 & 1 \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}$, and if N takes large values, it comes:

$$\begin{aligned} Tr(\Gamma_{\mathbf{1},d}, r) &\approx \frac{1}{N} Tr(H_N^T H_N, r) \approx \frac{1}{2} \frac{(N-|r|)(N-|r|+1)}{N} \\ &\approx \frac{1}{N} Tr(\Gamma_{\mathbf{0}}, r)(1 - \frac{|r|-1}{2}) \approx \frac{1}{2} Tr(\Gamma_{\mathbf{0}}, r) w(r) \end{aligned} \tag{57}$$

with $w(r) = (1 - \frac{|r|-1}{N})$ for $r = 1-N, ..., N-1$. When N tends to infinity, the influence of the weights disappears. This result will be useful in the next section.

6. Accuracy of the Hurst exponent estimate with the DFA when dealing a finite number of samples of short-memory processes

After showing that the statistical mean of the square of the fluctuation function $E[F^2_{\mathbf{1},d}(N)]$ converges to $E[F^2_{\mathbf{0}}(N)]$ when N tends to infinity up to a multiplicative factor, we will see that the estimation of the slope tends to the theoretical value of the Hurst exponent when larges values of N are considered and so for different short-memory processes such as a white noise, a moving-average process and a random process whose correlation function exponentially decays. However, for small values of M, the estimation of the slope may be drastically different and depend on the number M of samples available.

6.1. Statistical mean of the square of the fluctuation function

Let us recall the statistical mean of the square of the fluctuation function we derived in [50]. For this purpose, (21) was written in various ways: either by operating along the subdiagonals of $\Gamma_{\mathbf{1},d}$ or by operating along each row and by taking advantage of the symmetry of $\Gamma_{\mathbf{1},d}$:

$$\begin{aligned} F^2_{\mathbf{1},d}(N) &= \sum_{k=1}^{LN} \Gamma_{\mathbf{1},d}(k,k) y^2(k) \\ &+ \sum_{r=1}^{LN-1} \sum_{k=1}^{LN-r} [\Gamma_{\mathbf{1},d}(k,k+r) + \Gamma_{\mathbf{1},d}(k+r,k)] y(k) y(k+r) \\ &= \sum_{k=1}^{LN} \Gamma_{\mathbf{1},d}(k,k) y^2(k) + 2 \sum_{l=1}^{LN-1} \sum_{m=l+1}^{LN} \Gamma_{\mathbf{1},d}(l,m) y(l) y(m) \end{aligned} \tag{58}$$

By assuming that y is w.s.s. and taking the statistical mean of (58), one has:

$$E[F^2_{\mathbf{1},d}(N)] = \sum_{r=-LN+1}^{LN-1} Tr(\Gamma_{\mathbf{1},d}, r) R_{y,y}(r) \tag{59}$$

where $R_{y,y}(r)$ is the autocorrelation function of the process y.[4] As $\Gamma_{\mathbf{1},d}$ is a block-diagonal matrix with blocks of size $N \times N$, $Tr(\Gamma_{\mathbf{1},d}, r) = 0$ for $r \geq N$. Therefore, the sum reduces to the sum of $2N-1$ terms:

$$\begin{aligned} E[F^2_{\mathbf{1},d}(N)] &= \sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{1},d}, r) R_{y,y}(r) \\ &\underset{(51)}{=} \sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{1},d}, r) R_{y,y}(r) \end{aligned} \tag{60}$$

As $\gamma_{\mathbf{1},d}$ only depends on N, $E[F^2_{\mathbf{1},d}(N)]$ only depends on N. In other words, it does not depend on L and M.

[4] Let us recall that the correlation function is here defined by $R_{yy}(k_1, k_2) = E[y(k_1)y(k_2)]$. When y is wide-sense stationary, the following three conditions are satisfied: $E[y(k_1)]$ is a constant, $R_{yy}(k_1, k_2) = R_{yy}(\tau)$ where $\tau = k_1 - k_2$ and $R_{yy}(0)$ is bounded. Finally, when y is also ergodic, the statistical mean can be substituted by the temporal mean over an infinite duration of one realization of the random process, leading to $R_{yy}(\tau) = \lim_{K\to\infty} \frac{1}{2K+1} \sum_{k=-K}^{K} y(k)y(k-\tau)$. When only a few samples are available, the correlation of the function is estimated. This comment will be useful in the rest of the paper.

6.2. Impact on the estimation of the slope

In the following, we propose to study the behaviour of the estimated slope when the square of the fluctuation function is replaced by its statistical mean. It should be noted that our purpose is not to analyze the statistical mean of the slope, but to obtain meaningful information about the influence of the number of samples and consequently of N_{max}.

Given (25), let us look at the following quantity:

$$\hat{\alpha}_{\mathbf{1}}(N_1, ..., N_{max}) = \frac{1}{2} \frac{\overline{\log(N)\log(E[F^2_{\mathbf{1},d}(N)])} - \overline{\log(N)} \times \overline{\log(E[F^2_{\mathbf{1},d}(N)])}}{\overline{\log^2(N)} - \overline{\log(N)}^2} \tag{61}$$

or equivalently:

$$\hat{\alpha}_{\mathbf{1}}(N_1, ..., N_{max}) = \frac{1}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} \times \tag{62}$$

$$\overline{\log(N)\log\Big(\sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{1},d}, r) R_{y,y}(r)\Big)} - \overline{\log(N)} \times \overline{\log\Big(\sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{1},d}, r) R_{y,y}(r)\Big)}$$

At this stage, we suggest combining all the results above in order to analyze the accuracy of the estimation of the Hurst coefficient depending on the number of samples available. Different types of processes are analyzed: a white noise, a moving-average process, a random process with an exponentially decaying correlation.

6.2.1. Case 1: white noise

When the process is a w.s.s. zero-mean white noise with variance σ_m^2, its correlation function satisfies:

$$R_{y,y}(\tau) = \sigma_m^2 \delta(\tau) \tag{63}$$

where $\delta(\tau) = 1$ if $\tau = 0$ and 0 otherwise. In this case, (60) reduces to:

$$E[F^2_{\mathbf{1},d}(N)] = \sigma_m^2 Tr(\Gamma_{\mathbf{1},d}, 0) = \sigma_m^2 Tr(\gamma_{\mathbf{1},d}, 0) \tag{64}$$

When $d = 1$, using (33), (40) and (41), calculating the trace $Tr(\gamma_{\mathbf{1},1}, 0)$ amounts to summing the square of each element of $\mathbf{A}_{\mathbf{1},\mathbf{1}} H_N(r, c)$ and dividing the result

by N. Using numerical simulations, it confirms the result obtained by [40, 51] based on alternative formalisms:

$$E[F^2_{\mathbf{1},1}(N)] = \sigma_m^2 \frac{N^2-4}{15N} \tag{65}$$

The order of magnitude of the difference between our method and their analytical expression is 10^{-13}.

When $d = 2$, Höll [40] proposed the following expression:

$$E[F^2_{\mathbf{1},2}(N)] = \sigma_m^2 \frac{3N^2-18}{70N} \tag{66}$$

The expectation of the square of the fluctuation function obtained by Höll and by us for $\sigma_m^2 = 1$ and the normalized difference, expressed in percentage, are respectively given in Fig. 1 and 2. Given these numerical simulations, it appears that the expression (66) is not exactly the true one, *i.e.* (64), when N is small (< 40). It should be noted that the definition of the correlation function of the random process we consider is the standard one based on the expectation whereas Höll takes into account a definition based on a temporal mean, which implicitly means that the random process is assumed to be wide-sense-stationary and ergodic. This hence explains the difference for small value of N represented in Fig. 1 and Fig. 2. We already pointed out some differences with Höll's work when N is small in [23].

Let us now go back to the expression (62) of the slope. It becomes:

$$\hat{\alpha}_{\mathbf{1}}(N_1, ..., N_{max}) = \frac{1}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} \times \frac{}{\overline{\log(N)\log\Big(\sigma_m^2 Tr(\gamma_{\mathbf{1},d},0)\Big)} - \overline{\log(N)} \times \overline{\log\Big(\sigma_m^2 Tr(\gamma_{\mathbf{1},d},0)\Big)}} \tag{67}$$

However,

$$\log\Big(\sigma_m^2 \gamma_{\mathbf{1},d},0)\Big) = \log(\sigma_m^2) + \log\Big(Tr(\gamma_{\mathbf{1},d},0)\Big) \tag{68}$$

Therefore, as $\overline{\log(\sigma_m^2)} = \log(\sigma_m^2)$, the terms σ_m^2 vanish in (68) and one has:

$$\hat{\alpha}_{\mathbf{1}}(N_1, ..., N_{max}) = \frac{1}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} \times \frac{}{\overline{\log(N)\log\Big(Tr(\gamma_{\mathbf{1},d},0)\Big)} - \overline{\log(N)} \times \overline{\log\Big(\gamma_{\mathbf{1},d},0)\Big)}} \tag{69}$$

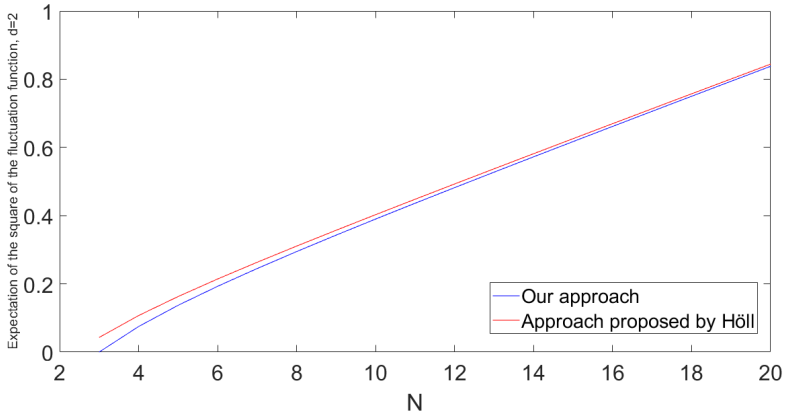

Figure 1: Difference between expressions of $E[F^2_{\mathbf{1},2}(N)]$: comparison between Höll's approach and ours for $d = 2$ when dealing with a white noise

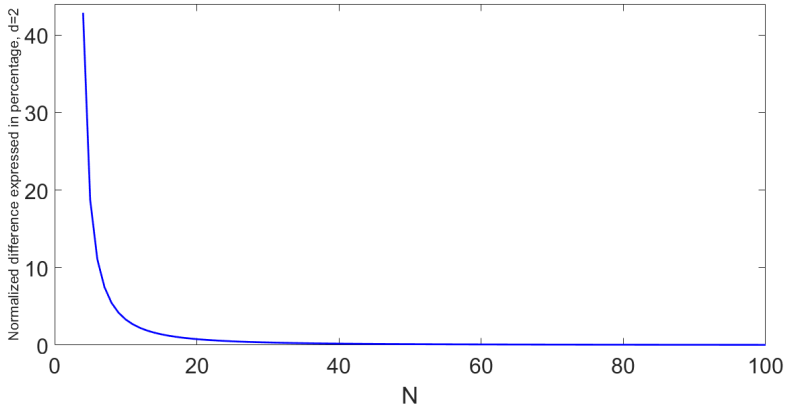

Figure 2: Zoom on the normalized difference between expressions of $E[F^2_{\mathbf{1},2}(N)]$ (in %): comparison between Höll's approach and ours for $d = 2$ when dealing with a white noise

It is also coherent with the fact that the slope cannot depend on the power of the signal under study.

Moreover, given the properties of the trace of $\Gamma_{\mathbf{1},d}$ and $\Gamma_{\mathbf{0}}$ we mentioned in (51), the slope $\hat{\alpha}_{\mathbf{1}}$ should converge to $\hat{\alpha}_{\mathbf{0}}$ obtained with the FA provided large values of N are considered. When the latter is used, one has:

$$E[F_{\mathbf{0}}^2(N)] = N\sigma_m^2 \tag{70}$$

The expression (62) of the slope applied to the FA with the statistical mean of the square of the fluctuation function is given by:

$$\begin{aligned}\hat{\alpha}_{\mathbf{0}}(N_1,...,N_{max}) &= \frac{\overline{\log(N)\log\left(N\sigma_m^2\right)} - \overline{\log(N)} \times \overline{\log\left(\sigma_m^2 N\right)}}{2\left(\overline{\log^2(N))} - \overline{\log(N)}^2\right)} \\ &= \frac{\overline{\log^2(N)} - \overline{\log(N)}^2}{2\left(\overline{\log^2(N))} - \overline{\log(N)}^2\right)} = \frac{1}{2}\end{aligned} \tag{71}$$

Therefore, when the variates $N_1, ..., N_{max}$ take large values:

$$\lim_{N_1,...,N_{max}\to+\infty} \hat{\alpha}_{\mathbf{1}}(N_1,...,N_{max}) \approx \lim_{N_1,...,N_{max}\to+\infty} \hat{\alpha}_{\mathbf{0}}(N_1,...,N_{max}) = \frac{1}{2} \tag{72}$$

As a conclusion, when large values of $N_1, ..., N_{max}$ are selected and if the statistical mean of the square of the fluctuation function is considered in the DFA, the estimation of the slope tends to the true one, *i.e.* $\alpha = H + 1 = -0.5 + 1 = 0.5$ in the case of a white noise.

6.2.2. Case 2: moving-average process

Let us consider the w.s.s. q^{th}-order MA process. Its k^{th} sample, denoted as $y(k)$, is defined as follows:

$$y(k) = \sum_{j=0}^{q} b_j u(k-j) \tag{73}$$

where $u(k)$ is the k^{th} sample of the driving process, assumed to be white, Gaussian, zero-mean with variance σ_m^2. $b_0 = 1$ and $\{b_j\}_{j=0,...,q}$ denote the MA parameters. This can be seen as a filtering of the driving process whose transfer function is given by $H(z) = \sum_{j=0}^{q} b_j z^{-j}$. The latter is defined by its zeros which may be inside or outside the unit-circle in the z-plane.

However, there may be statistical links between a MA process with zeros inside the unit-circle in the z-plane and a MA process with zeros outside the unit-circle in the z-plane. Thus, let us take the example of a 1^{st}-order MA process defined by the following transfer function:

$$H_{z_l}(z) = (1 - z_l z^{-1}) = |z_l| H_{bla,z_l}(z) \left(1 - \frac{1}{z_l^*} z^{-1}\right) \tag{74}$$

where z_l^* denotes the conjugate of z_l, $H_{bla,z_l}(z) = \frac{1}{|z_l|} \frac{1 - z_l z^{-1}}{1 - \frac{1}{z_l^*} z^{-1}}$ is a Blaschke product, representative of a transfer function of an all-pass filter.

The above equation (74) amounts to saying that the two filters with transfer functions $H_{z_l}(z)$ and $\left(1 - \frac{1}{z_l^*} z^{-1}\right)$ have the same frequency responses up to a multiplicative factor equal to $|z_l|$. In other words, the MA process defined by the driving process with variance σ_m^2 and the MA parameters $(1, -z_l)$ and the MA process defined by the driving process with variance $K_l \sigma_m^2 = |z_l|^2 \sigma_m^2$ and by the MA parameters $(1, -\frac{1}{z_l^*})$ have the same power spectral densities and consequently the same correlation function. In addition, the MA processes with transfer functions $H_{z_l}(z)$ and $\left(1 - \frac{1}{z_l^*} z^{-1}\right)$ have the same normalized correlation function[5]. This result can be generalized to any q^{th}-order MA process.

Following the same reasoning as in the previous section, one can show that the expression of the slope does not depend on σ_m^2. Therefore, if 1^{st}-order MA processes are analyzed by using the DFA, the value obtained with the DFA for the set of MA parameters $(1, b_1)$ is the same as $(1, \frac{1}{b_1})$. One can reduce the simulation protocol to MA processes corresponding to minimum-phase filters, *i.e.* with zeros inside the unit-circle in the z-plane.

When the FA is used and in the case of a 1^{st}-order process, one has:

$$E[F_{\mathbf{0}}^2(N)] = \Big(N(1 + b_1^2) + 2(N-1)b_1\Big)\sigma_m^2 = \Big(N(1 + b_1)^2 - 2b_1\Big)\sigma_m^2 \tag{76}$$

[5]Note that the correlation function is defined as follows:

$$\begin{cases} R_{y,y}(|\tau|) = \sigma_m^2 \sum\limits_{j=|\tau|}^{q} b_j b_{j-|\tau|} \text{ for } 0 \leq |\tau| \leq q \\ R_{y,y}(|\tau|) = 0 \text{ otherwise} \end{cases} \tag{75}$$

When $1+2b_1+b_1^2 = (1+b_1)^2 \neq 0$ or equivalently $b_1 \neq -1$, $E[F_{\mathbf{0}}^2(N)]$ depends on N. When N increases, $N(1+b_1)^2 >> 2b_1$. Taking into account the expression of the slope using the FA with the statistical mean of the square of the fluctuation function, one can easily deduce that for large values of $\{N_1, \ldots, N_{max}\}$:

$$\begin{aligned}\hat{\alpha}_{\mathbf{0}}(N_1, ..., N_{max}) &= \frac{\overline{\log(N)\log(N(1+b_1)^2)} - \overline{\log(N)} \times \overline{\log\left(N(1+b_1)^2\right)}}{2\left(\overline{\log^2(N))} - \overline{\log(N)}^2\right)} \\ &= \frac{\overline{\log^2(N)} - \overline{\log(N)}^2}{2\left(\overline{\log^2(N))} - \overline{\log(N)}^2\right)} = \frac{1}{2}\end{aligned} \tag{77}$$

Then, the estimation of the Hurst coefficient tends to $-\frac{1}{2}$.
When $b_1 = -1$, $E[F_{\mathbf{0}}^2(N)] = 2\sigma_m^2$, the slope is equal to:

$$\hat{\alpha}_{\mathbf{0}}(N_1, ..., N_{max}) = \frac{\overline{\log(N)\log(2\sigma_m^2)} - \overline{\log(N)} \times \overline{\log\left(2\sigma_m^2\right)}}{2\left(\overline{\log^2(N))} - \overline{\log(N)}^2\right)} = 0 \tag{78}$$

Therefore, the estimation of the Hurst exponent tends to be equal to -1.
Let us now analyze a 2^{nd}-order MA process. In this case, when the FA is used, one has:

$$\begin{aligned}E[F_{\mathbf{0}}^2(N)] &= \Big(N(1+b_1^2+b_2^2) + 2(N-1)(b_1+b_1b_2) + 2(N-2)b_2\Big)\sigma_m^2 \\ &= \Big(N(1+b_1^2+b_2^2+2b_1+2b_2+2b_1b_2) - 2(b_1+2b_2+b_1b_2)\Big)\sigma_m^2\end{aligned} \tag{79}$$

In the above expression, $(1+b_1^2+b_2^2+2b_1+2b_2+2b_1b_2) = (1+b_1+b_2)^2$. Therefore, $E[F_{\mathbf{0}}^2(N)]$ does not depend on N when $1+b_1+b_2 = 0$. In this case, $\hat{\alpha}_{\mathbf{0}} = 0$. In the other case, following the same reasoning as above, one can easily show that $\hat{\alpha}_{\mathbf{0}} = \frac{1}{2}$. One can generalize the above result for a q^{th}-order MA process. $\hat{\alpha} = \frac{1}{2}$ except when $\sum_{j=0}^{q} b_j = H(z)|_{z=1} = 0$ for which $\hat{\alpha} = 0$. This would mean that that the FA using the statistical mean of the square of the fluctuation function tends to the true result except when $\sum_{j=0}^{q} b_j = H(z)|_{z=1} = 0$, *i.e.* when the PSD of the MA process is characterized by a total rejection of the null frequency. Consequently, when $N_1, ..., N_{max}$ take large values, the DFA based on the statistical mean of the square of the fluctuation function would lead to the same results.

6.2.3. Case 3: random process with correlation function decaying exponentially

The ARMA process of order (p, q) is defined defined by[6]:

$$y(k) = -\sum_{i=1}^{p} a_i y(k-i) + \sum_{j=0}^{q} b_j u(k-j) \tag{80}$$

The correlation function for $\tau \geq q+1$ satisfies:

$$R_{y,y}(\tau) = -\sum_{i=1}^{p} a_i R_{y,y}(\tau - i) \tag{81}$$

By setting $a_0 = 1$, this is a recurrence relation defined by the following characteristic equation:

$$\sum_{i=0}^{p} a_i x^{p-i} = 0 \tag{82}$$

It has its roots equal to $\{p_i\}_{i=1,\ldots,p}$. The latter correspond to the poles of the transfer function $H(z) = \frac{1}{\sum_{i=0}^{p} a_i z^{-i}}$. If the roots are of order of multiplicity equal to 1, the correlation function can be expressed as follows:

$$R_{y,y}(\tau) = \sum_{i=1}^{p} c_i p_i^{|\tau|} \tag{83}$$

If the root p_p has its order of multiplicity larger than 1 and equal to m, one has:

$$R_{y,y}(\tau) = \sum_{i=1}^{p-1} c_i p_i^{|\tau|} + c_{p,0} p_p^{|\tau|} + c_{p,1}|\tau| p_p^{|\tau|} + \ldots + + c_{p,m-1}|\tau|^{m-1} p_p^{|\tau|} \tag{84}$$

As a consequence, the correlation function of an ARMA process decays exponentially.

In the following, let us consider another type of w.s.s. process with short memory whose correlation function decays exponentially, with a decay constant β:

$$R_{y,y}(\tau) = \sigma_m^2 e^{-\beta|\tau|T} = \sigma_m^2 \rho^{|\tau|} \tag{85}$$

where τ is the difference between two consecutive instants, T is the sampling frequency, $\rho = e^{-\beta T}$, and σ_m^2 is the power of the process. The latter characterizes the acceleration of a Singer motion model.

[6]The AR processes is defined by $y(k) = -\sum_{i=1}^{p} a_i y(k-i) + u(k)$. It is an ARMA process of order $(p, q = 0)$.

Therefore, (60) becomes:

$$E[F^2_{\mathbf{1},d}(N)] = \sigma^2_m \sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{1},d}, r)\rho^{|r|} \tag{86}$$

Remark: When the FA is used, one has:

$$E[F^2_{\mathbf{0}}(N)] = \sigma^2_m \sum_{r=-N+1}^{N-1} Tr(\Gamma_{\mathbf{0}}, r)\rho^{|r|} = \sigma^2_m \sum_{r=-N+1}^{N-1} (N-|r|)\rho^{|r|} \tag{87}$$

Let us express it differently.

$$\begin{aligned} E[F^2_{\mathbf{0}}(N)] &= \sigma^2_m \Big(N + 2\sum_{r=1}^{N-1}(N-r)\rho^r\Big) \\ &= \sigma^2_m \Big(N + 2N\sum_{r=1}^{N-1}\rho^r - 2\rho\sum_{r=1}^{N-1} r\rho^{r-1}\Big) \\ &= \sigma^2_m \Big(N + 2N\rho\frac{1-\rho^{N-1}}{1-\rho} - 2\rho\frac{d}{d\rho}\Big(\frac{\rho-\rho^N}{1-\rho}\Big)\Big) \\ &= \sigma^2_m \Big(N + 2N\rho\frac{1-\rho^{N-1}}{1-\rho} - 2\rho\frac{1-N\rho^{N-1}+(N-1)\rho^N}{(1-\rho)^2}\Big) \end{aligned} \tag{88}$$

When taking the limit of the above expression when N increases and takes a large value, one has:

$$\begin{aligned} &\lim_{N\to+\infty} E[F^2_{\mathbf{0}}(N)] \\ &= \lim_{N\to+\infty} \sigma^2_m \Big(N + 2N\rho\frac{1-\rho^{N-1}}{1-\rho} - 2\rho\frac{1-N\rho^{N-1}+(N-1)\rho^N}{(1-\rho)^2}\Big) \\ &= \lim_{N\to+\infty} \sigma^2_m \Big(N(1+\frac{2\rho}{1-\rho}) - \frac{2\rho}{(1-\rho)^2}\Big) \end{aligned} \tag{89}$$

Therefore, when $N_1, ..., N_{max}$ are large and the statistical mean of the square of the fluctuation function is considered, the slope obtained with the FA satisfies:

$$\lim_{N_1,...,N_{max}\to+\infty} \hat{\alpha}_{\mathbf{0}}(N_1, ..., N_{max}) = \frac{1}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} \times \frac{\overline{\log(N)\log\Big(N(1+\frac{2\rho}{1-\rho}) - \frac{2\rho}{(1-\rho)^2}\Big)} - \overline{\log(N)} \times \overline{\log\Big(N(1+\frac{2\rho}{1-\rho}) - \frac{2\rho}{(1-\rho)^2}\Big)}}{} \tag{90}$$

This leads to:

$$\lim_{N_1,...,N_{max}\to+\infty} \hat{\alpha}_{\mathbf{0}}(N_1,...,N_{max}) = \frac{1}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} \times \tag{91}$$

$$\overline{\log(N)\log\Big(N(1+\frac{2\rho}{1-\rho})\Big)} - \overline{\log(N)} \times \overline{\log\Big(N(1+\frac{2\rho}{1-\rho})\Big)}$$

$$= \frac{\overline{\log^2(N)} - \overline{\log N}^2}{2\Big(\overline{\log^2(N))} - \overline{\log(N)}^2\Big)} = \frac{1}{2}$$

Consequently, the DFA used with large values of $N_1, ..., N_{max}$ and the statistical mean of the square of the square of the fluctuation function leads to the same result.

In what follows, various illustrations are given.

7. Illustrations

In this section, we suggest analyzing how the estimation $\hat{\alpha}$ evolves when M increases. It is true that our analysis is based on the statistical mean of the square of the fluctuation function whereas the DFA uses the square of the fluctuation function only. Thus, the performance of the DFA in its standard version are necessarily worse. Note that the practitioner could consider a larger set of samples, create several realizations of the process by using a sliding window, compute the squares of the resulting fluctuation functions and average them. We would then come closer to the analysis conditions that we introduced in this paper.

Given what we presented in the above sections, the following comments can be made:

- Whatever M, $E[F^2_{\mathbf{1},d}(N)]$ has always the same value.
- Assuming that M is even, the diagram representing $\log(E[F_{\mathbf{1},d}(N)])$ as a function of $\log(N)$ for a signal of length $M+1$ is the same as the one obtained for the signal of length M.
- The diagram representing $\log(E[F_{\mathbf{1},d}(N)])$ as a function of $\log(N)$ for a signal of length $M+2$ is composed of the $M/2$ points of the curve obtained

from the signal of length M and an additional point whose coordinates are $\log(E[F_{\mathbf{1}}((M+1)/2)]), \log((M+1)/2)$ to be computed.

- The behaviour of the DFA tends to the one of the FA when M increases.

7.1. Case 1: white noise

In this section, three strategies are considered for the selection of $N_1, \ldots, N_{max}$. For the first one, $N_1, \ldots, N_{max}$ take consecutive values, whose lower bound is equal to $\lfloor \frac{M}{12} \rfloor$ and upper bound is $\lfloor \frac{M}{2} \rfloor$. The second one consists in taking consecutive values from $N = 20$ to $\lfloor \frac{M}{2} \rfloor$, whereas the last strategy is to consider all the values between 3 to $\lfloor \frac{M}{2} \rfloor$. The first two strategies avoid using too small values of N to be in accordance with the theory presented in the previous section and to take into account the comments made by Kantelhardt [54] where the fluctuation function is under-estimated and has to be corrected. The last approach has the advantage of being applicable when only very few samples are available. Selecting $d = 0$ and $d = 1$ provides the best results when the number M of samples is small. The abacuses based on the DFA and the higher-order DFA with $d = \{0, 1\}$ and $d = \{2, 3\}$ that provide $\hat{\alpha}_{\mathbf{1}} - 1$ are given respectively in Fig. 3 and Fig. 4.

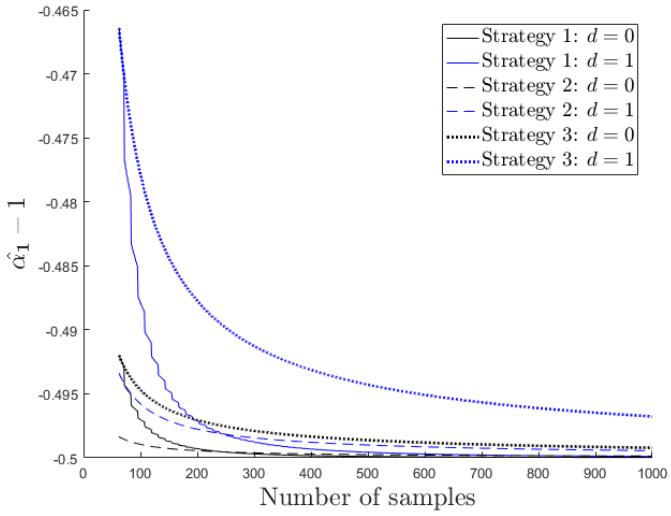

Figure 3: Abacuses for a white noise, giving $\hat{\alpha}_{\mathbf{1}} - 1$ for different values of M, based on three strategies and and $d = \{0, 1\}$

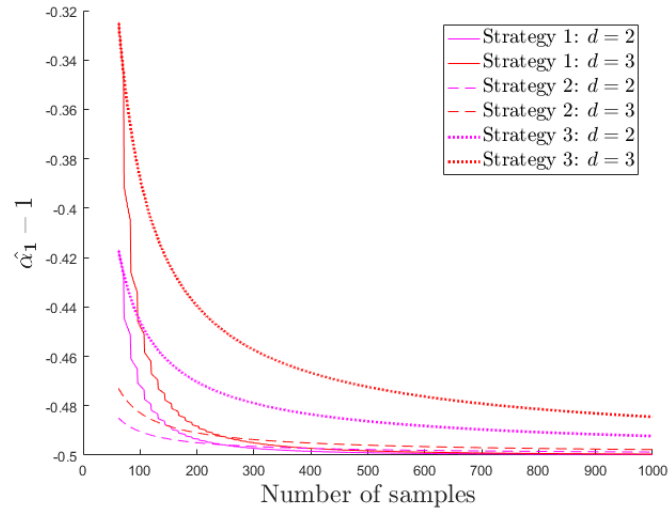

Figure 4: Abacuses for a white noise, giving $\hat{\alpha}_{\mathbf{1}} - 1$ for different values of M, based on three strategies and $d = \{2, 3\}$

One can see that when the number of samples is small, the value is not equal to

-0.5 whatever the strategy that is considered. The first strategy converges to the true value faster than the second method as it does not take into account small values for $\{N_i\}_{i=1,2,...}$. Indeed, the lower bound for N_1 increases when the first strategy is considered whereas it is fixed in the second strategy. Moreover, the abacus obtained with the third strategy is necessarily above the abacus obtained with the second one because small values of $\{N_i\}_{i=1,2,...}$ are considered in the third one, leading to underestimated values of the fluctuation function and consequently larger estimates of the slopes. When comparing the first and the second strategy, as the lower bound for N_1 are respectively equal to $\lfloor \frac{M}{12} \rfloor$ and 20, the abacus based on the first strategy is necessarily above the one obtained with the second for M smaller than $20 \times 12 = 240$. Finally, strategy 3 leads to the wider range of values for the estimation of the regularity, depending on the samples available.

In the next section, as few samples are usually available in various applications, we propose to provide abacuses based on the third strategy[7]. In the next cases, the abacuses are presented for $d = 1$, but the abacuses for other values of d can be easily obtained.

7.2. Case 2: moving-average process

In this sub-section, a 1^{st}-order MA process is analyzed. First of all, let us confirm that the results obtained for the set of MA parameters $(1, b_1)$ and $(1, \frac{1}{b_1})$ are the same. In Fig. 5 and 6 for the DFA and the FA respectively, the MA parameter b_1 varies from -3 to 3 and M can take different values equal to 50, 75, 100, 250, 500, 750, 1500, 2000, 5000, 10000 and 19000. This confirms what we show in the theoretical part in Section 6.

Let us now analyze how the estimation of the Hurst exponent evolves when the number M increases. The MA parameters varies from -1 to 1 with a step size equal to 0.2. This makes it possible to analyze a large panel of spectral

[7] As a matter of fact, the practitioner can design his own abacus following the approach we propose based on the values N_1, N_2, ...,N_{max} he chooses.

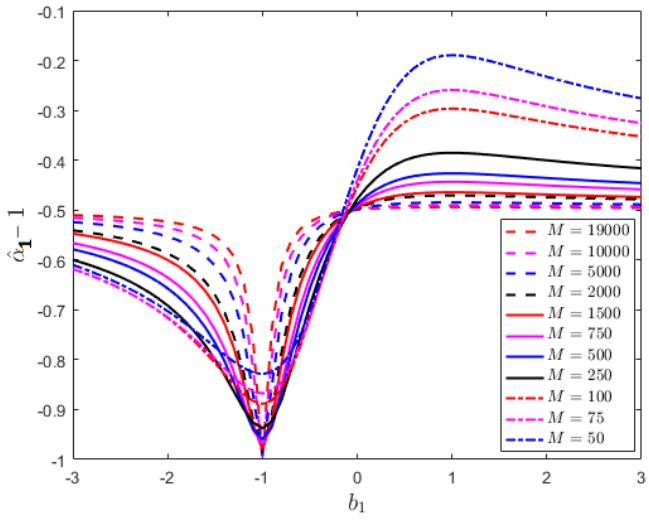

Figure 5: Abacus giving $\hat{\alpha}_{\mathbf{1}} - 1$ vs MA parameter b_1, for different values of M in the case of the DFA

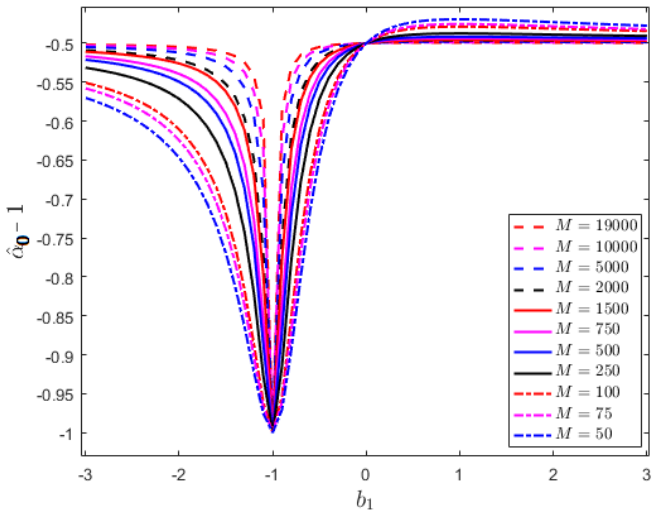

Figure 6: Abacus giving $\hat{\alpha}_{\mathbf{0}} - 1$ vs MA parameter b_1, for different values of M in the case of the FA

properties.

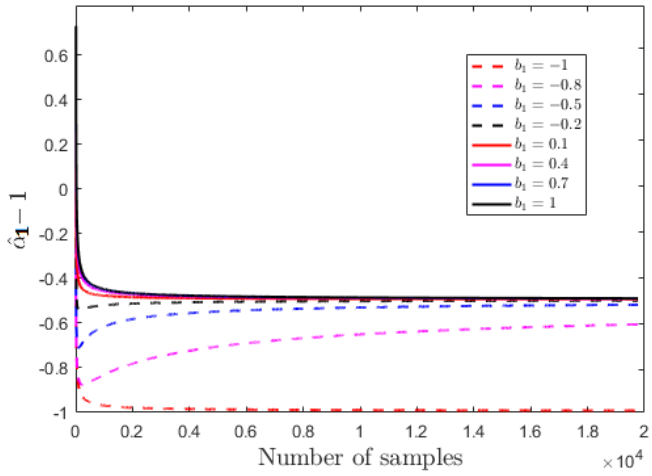

Figure 7: Abacus based on the DFA giving $\hat{\alpha}_{\mathbf{1}} - 1$ for different values of M and different MA parameters

When M becomes larger and larger, the value tends to be the same and equal to -0.5. However, when M is small, the results that are obtained differ and depend on the values of the MA parameter.

Remark: The abacus for the FA approach is given Fig. 8.

7.3. Case 3: random process with correlation decaying exponentially

Here is the abacus that provides $\hat{\alpha}_{\mathbf{1}} - 1$ for values of β equal to 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1 and 5. $T = 1$s and $\sigma_m^2 = 1$ (even if the latter parameter does impact on the result).

The larger β, the faster it converges to -0.5.

8. Conclusions and perspectives

In this paper, we first confirm that the DFA does not provide the true estimation of the Hurst exponent when the number of samples is not large for short-memory processes and the sizes N of the segments are not set to large

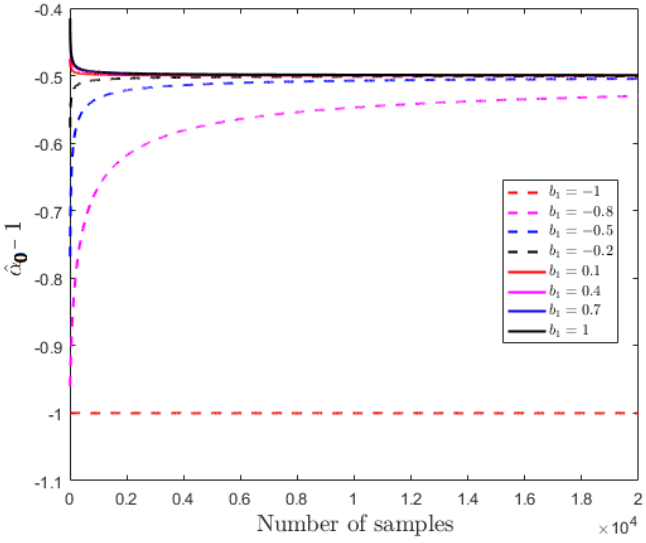

Figure 8: Abacus based on FA giving $\hat{\alpha}_{\mathbf{0}} - 1$ for different values of M and different MA parameters

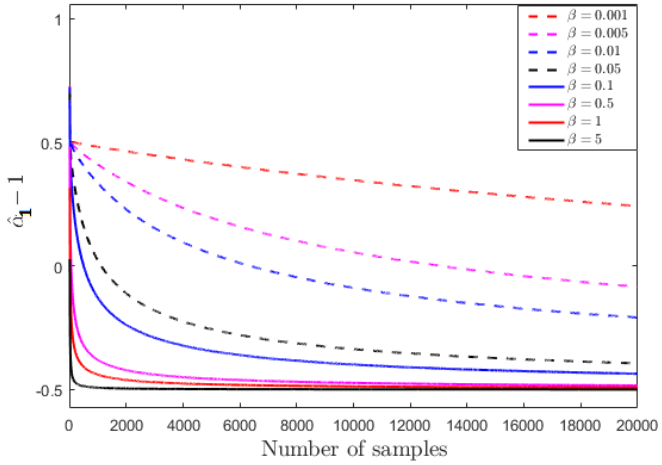

Figure 9: Abacus giving $\hat{\alpha}_{\mathbf{1}} - 1$ for different values of β and M

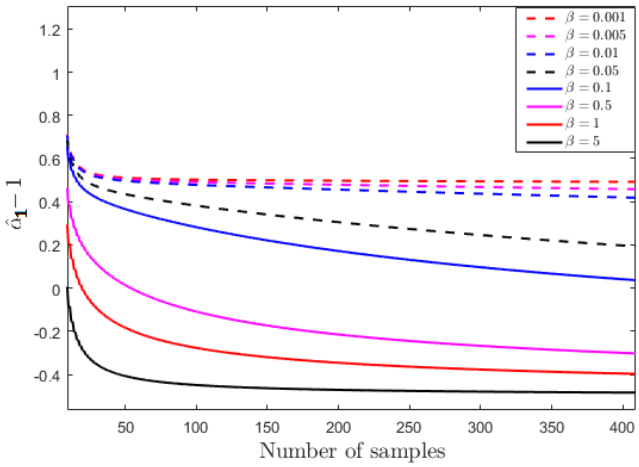

Figure 10: Zoom on the abacus giving $\hat{\alpha}_{\mathbf{1}} - 1$ for different values of β and M

values. Then, we present a method to derive abacuses linking the value provided by the DFA to the properties of the w.s.s. short-memory random process and the number of samples available. This approach is based on the expression of the statistical mean of the square of the fluctuation function, without any approximation and by using a matrix formulation. No assumption of ergodicity is made.

In the future, we plan to analyze the case of long-memory processes such as fractionally integrated white noises and autoregressive fractionally integrated moving-average processes. We also plan to derive statistical properties of the estimation of the Hurst exponent by using the matrix formalism we have used.

References

[1] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," SIAM Review, vol. 10(4), pp. 422–437, 1968.

[2] S. Sanyal, A. Banerjee, R. Pratihar, A. Kumar Maity, S. Dey, V. Agrawal, R. Sengupta, and D. Ghosh, "Detrended fluctuation and power spectral

analysis of alpha and delta EEG brain rhythms to study music elicited emotion," International Conference on Signal Processing, Computing and Control, pp. 206–210, 2015.

[3] A. A. Pranata, G. W. Adhane, and D. S. Kim, "Detrended fluctuation analysis on ECG device for home environment," Consumer Communications and Networking Conference, pp. 4233–4236, 2017.

[4] A. G. Ravelo-Garcia, U. Casanova-Blancas, S. Martin-González, E. Hernández-Pérez, I. Guerra-Moreno, P. Quintana-Morales, N. Wessel, and J. L. Navarro-Mesa, "An approach to the enhancement of sleep apnea detection by means of detrended fluctuation analysis of RR intervals," Computing in Cardiology, pp. 905–908, 2014.

[5] R. U. Acharya, C. M. Lim, and P. Joseph, "Heart rate variability analysis using correlation dimension and detrended fluctuation analysis," ITBM-RBM, vol. 23, no. 6, pp. 333–339, 2002.

[6] R. Kumagai and M. Uchida, "Detrended fluctuation analysis of repetitive handwriting," in 2017 International Conference on Noise and Fluctuations (ICNF), 2017, pp. 1–4.

[7] W. Mumtaz, A.-S. Malik, S. S. A. Ali, M. A. M. R. Yasin, and A. Hafeezullah, "Detrended fluctuation analysis for major depressive disorder," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 08 2015, pp. 4162–4165.

[8] A. Adda and H. Benoudnine, "Detrended fluctuation analysis of EEG recordings for epileptic seizure detection," in 2016 International Conference on Bio-engineering for Smart Technologies (BioSMART), 2016, pp. 1–4.

[9] A. Kitlas Golińska, "Detrended fluctuation analysis (DFA) in biomedical signal processing: selected examples," Studies in Logic, Grammar and Rhetoric, vol. 29, pp. 107–115, 01 2012.

[10] A. Tiwari, C. Albulescu, and S.-M. Yoon, "A multifractal detrended fluctuation analysis of financial market efficiency: comparison using dow jones sector etf indices," Physica A: Statistical Mechanics and its Applications, vol. 483, pp. 182–192, 05 2017.

[11] A. Serletis, O.-Y. Uritskaya, and V.-M. Uritsky, "Detrended fluctuation analysis of the US stock market," International Journal of Bifurcation and Chaos, vol. 18, no. 02, pp. 599–603, 2008.

[12] P. Talkner and R. Weber, "Power spectrum and detrended fluctuation analysis: Application to daily temperatures," Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics, vol. 62, pp. 150–60, 08 2000.

[13] A. Király and I. M. Jánosi, "Detrended fluctuation analysis of daily temperature records: geographic dependence over australia," Meteorology and Atmospheric Physics, vol. 88, no. 3-4, pp. 119–128, 2005.

[14] W.-W. Tung, J. Gao, J. Hu, and L. Yang, "Detecting chaos in heavy-noise environments," Physical Review E, vol. 83, pp. 046210, Apr 2011.

[15] H. E. Hurst, "Long-term storage capacity of reservoirs," Transactions of the American Society of Civil Engineers, vol. 116, pp. 770–799, 1951.

[16] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long range dependence: an empirical study," Fractals, vol. 3, (4), pp. 785–788, 1995.

[17] M. S. Taqqu and V. Teverovsky, "On estimating the intensity of long range dependence in finite and infinite variance time series," A practical guide to heavy tails: statistical techniques and applications, pp. 177–217, 1996.

[18] G. Rilling, P. Flandrin, and P. Gonçalves, "Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation," International Conference on Acoustics, Speech and Signal Processing, pp. 489–492, 2005.

[19] R. Sun, "Fractional order signal processing: techniques and applications," Thesis of Master of science in electrical Engineering, Utah State University, 2007.

[20] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Self-similarity and long-range dependence through the wavelet lens," Theory and Applications of Long-range Dependence, pp. 527–556, 2003.

[21] J.-M. Bardet, "Testing for the presence of self-similarity of Gaussian time series having stationary increments," Journal of Time Series Analysis, vol. 21, pp. 497–515, 2000.

[22] E. Moulines, F. Roueff, and M. S. Taqqu, "Central limit theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context," Fractals, vol. 15, pp. 301–313, 2007.

[23] B. Berthelot, E. Grivel, P. Legrand, J.-M. André, and P. Mazoyer, "Alternative ways to compare the detendred fluctuation analysis and its variants. application to visual tunneling detection," Digital Signal Processing, vol. 108, pp. 102865, 2020.

[24] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," Physical Review E, vol. 49, (2), pp. 1685–1689, 1994.

[25] 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.

[26] B. Berthelot, E. Grivel, P. Legrand, J.-M. André, P. Mazoyer, and T. Ferreira, "Filtering-based analysis comparing the DFA with the CDFA for wide sense stationary processes," in EUSIPCO 2019, 2019.

[27] B. Berthelot, E. Grivel, P. Legrand, J.-M. André, and P. Mazoyer, "Regularized DFA to study the gaze position of an airline pilot," in EUSIPCO 2020, 2020.

[28] M. A. Riley, S. Bonnette, N. Kuznetsov, S. Wallot, and J. Gao, "A tutorial introduction to adaptive fractal analysis," Frontiers in Physiology, vol. 3, pp. 371, 2012.

[29] Rodriguez E., Echeverrıa J.-C., and Alvarez-Ramirez J., "Detrending fluctuation analysis based on high-pass filtering," Physica A, vol. 375, pp. 699–708, 2007.

[30] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro, "Second-order moving average and scaling of stochastic time series," The European Physical Journal B, vol. 27, 2, pp. 197–200, 2002.

[31] L. Xu, P. Ch. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, "Quantifying signals with power-law correlations: a comparative study of detrended fluctuation analysis and detrended moving average techniques," Physical Review E, vol. 71, 5, pp. 051101, 2005.

[32] A. Carbone, "Detrending moving average algorithm: a brief review," IEEE International Conference Science and Technology for Humanity (TIC-STH), 2009.

[33] S. Arianos, A. Carbone, and C. Türk, "Self-similarity of higher-order moving averages," Physical Review E, vol. 84(4 Pt 2), pp. 046113, 2011.

[34] B. Berthelot, E. Grivel, and P. Legrand, "New Variants of DFA based on LOESS and LOWESS methods: generalization of the detrended moving average," in ICASSP 2021, 2021.

[35] Y.-H. Shao, G.-F. Gu, Z.-Q. Jiang, W.-X. Zhou, and D. Sornette, "Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series," Scientific Reports, vol. 2, pp. 835, 2012.

[36] Q. Fan and F. Wang, "Detrending-moving-average-based bivariate regression estimator," Physical Review E, vol. 102, 2020.

[37] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," Physica A: Statistical Mechanics and its Applications, vol. 316, (1-4), pp. 87–114, 2002.

[38] Y. Tsujimoto, Y. Miki, S. Shimatani, and K. Kiyono, "Fast algorithm for scaling analysis with higher-order detrending moving average method," Physical Review E, vol. 93 (5), pp. 053304, 2016.

[39] Y. Tsujimoto, Y. Miki, E. Watanabe, J. Hayano, Y. Yamamoto, T. Nomura, and K. Kiyono, "Fast algorithm of long-range cross-correlation analysis using Savitzky-Golay detrending filter and its application to biosignal analysis," International Conference on Noise and Fluctuations, pp. 1–4, 2017.

[40] M. Höll and H. Kantz, "The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal," The European Physical Journal B, vol. 88, pp. 327, 2015.

[41] K. Kiyono, "Theory and applications of detrending-operation-based fractal-scaling analysis," International Conference on Noise and Fluctuations (ICNF), 2017.

[42] K. Kiyono, "Establishing a direct connection between detrended fluctuation analysis and fourier analysis," Physical Review E, vol. 92, pp. 042925, 2015.

[43] K. Kiyono and Tsujimoto Y, "Time and frequency domain characteristics of detrending-operation-based scaling analysis: Exact dfa and dma frequency responses," Physical review E, vol. 94, pp. 012111, 2016.

[44] A. Carbone and K. Kiyono, "Detrending moving average algorithm: Frequency response and scaling performances," Physical Review E, vol. 93, 2016.

[45] B. Berthelot, E. Grivel, P. Legrand, M. Donias, J.-M. André, P. Mazoyer, and T. Ferreira, "2D Fourier transform based analysis comparing the DFA with the DMA," in EUSIPCO 2019, 2019.

[46] M. Höll, K. Kiyono, and H. Kantz, "Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average," Physical Review E, vol. 99, pp. 033305, 2019.

[47] N. Crato, R. Linhares, and S. R. Costa Lopes, "Statistical properties of detrended fluctuation analysis," Journal of Statistical Computation and Simulation, vol. 80(6), pp. 625–641, 2010.

[48] M. Holl and H. Kantz, "The fluctuation function of the detrended fluctuation analysis – investigation on the ar(1) process," The European Physical Journal B, vol. 88, pp. 126, 2015.

[49] M. Höll, H. Kantz, and Y. Zhou, "Detrended fluctuation analysis and the difference between external drifts and intrinsic diffusionlike nonstationarity," Physical Review E, vol. 94, pp. 042201, 2016.

[50] B. Berthelot, P. Mazoyer, S. Egea, J.-M. André, E. Grivel, and P. Legrand, "Self-affinity of an aircraft pilot's gaze direction as a marker of visual tunneling. sae aerotech," SAE Aerotech, 2019.

[51] G. Sikora, M. Höll, J. Gajda, H. Kantz, A. Chechkin, and A. Wylomanska, "Probabilistic properties of detrended fluctuation analysis for Gaussian processes," Physical review E, vol. 94, pp. 032114, 2020.

[52] V. V. Morariu, L. Buimaga-Iarinca, C. Vamoş, and S. M. Soltuz, "Detrended fluctuation analysis of autoregressive processes," submitted to Fluctuation and Noise Letters, p. arXiv:0707.1437, 2007.

[53] D. Maraun, H. W. Rust, and J. Timmer, "Tempting long-memory – on the interpretation of DFA results," Nonlinear Processes in Geophysics, p. 495–503, 2004.

[54] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis," Physica A: Statistical Mechanics and its Applications, vol. 295, (3-4), pp. 441–454, 2001.

[55] F. Lillo and J. Doyne Farmer, "The long memory of the efficient market," Studies in Nonlinear Dynamics and Econometrics, vol. 8, pp. 1–33, 2004.