Physica A 445 (2016) 283–294

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Quantifying long-range correlations with a multiscale ordinal pattern approach

Felipe Olivares [a,b,*], Luciano Zunino [a,c], Osvaldo A. Rosso [d,e]

[a] Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina

[b] Instituto de Física, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso (PUCV), Av. Brasil 2950, 234-0025, Valparaíso, Chile

[c] Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina

[d] Instituto de Física, Universidade Federal de Alagoas (UFAL), BR 104 Norte km 97, 57072-970, Maceió, Alagoas, Brazil

[e] Intituto Tecnológico de Buenos Aires (ITBA), C1106ACD, Av. Eduardo Madero 399, Ciudad Autónoma de Buenos Aires, Argentina

HIGHLIGHTS

- We use ordinal patterns probabilities for estimating Hurst exponents.
- Numerical and experimental results are included.
- Robustness to observational noise is confirmed.
- The spurious effect of low resolution data is carefully analysed.
- Results are contrasted with those obtained from DFA.

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form 4 November 2015
Available online 17 November 2015

Keywords:
Ordinal patterns probabilities
Fractional Brownian motion
Hurst exponent
Multiscale analysis

ABSTRACT

In this paper we use the ordinal patterns probabilities associated with fractional Brownian motions for estimating the Hurst exponent of artificially generated and experimentally measured data. Numerical analysis show a reliable estimation of this scaling parameter, even when data with low resolution are analysed. Robustness to observational noise is also obtained. Several experimental applications allow us to confirm the practical utility of the proposed approach. We contrast results obtained by implementing this multiscale symbolic tool with those obtained from the classical detrended fluctuation analysis.

1. Introduction

Long-range correlations in data sequences have been observed in different and heterogeneous scientific fields, like human gait [1], heart rate dynamics [2,3], ocean wave dynamics [4], financial time series [5,6], DNA code sequences [7,8], etc. This kind of natural phenomena has an empirical power-law spectra $1/f^{\alpha}$, where f stands for frequency and $\alpha \in (1, 3)$, with $\alpha = 2H + 1\,(0 < H < 1)$. The parameter H is called Hurst exponent [9], and quantifies the degree of long-range correlations in the time series. Given a record of equidistant measurements x_i with $i = 1, \ldots, N$ it is relevant to characterize the temporal correlations over different time scales s. On the one hand, short-range correlated data are described by an autocorrelation function that decays exponentially ($C(s) \propto e^{(s/s_x)}$ with s_x the decay time). On the other hand, if data are long-range correlated

* Corresponding author at: Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina.
E-mail address: olivaresfe@gmail.com (F. Olivares).

http://dx.doi.org/10.1016/j.physa.2015.11.015
0378-4371/

the autocorrelation function declines as a power-law ($C(s) \propto s^{-\gamma}$ with an exponent $0 < \gamma < 1$). The direct estimation of $C(s)$ is usually problematic and inappropriate. The Hurst exponent provides an indirect way to obtain the correlation exponent γ [10]. Therefore, it is of great interest the study of the processes that are parametrized by the Hurst exponent H and, of course, its adequate estimation for a suitable modelling of the associated phenomenon [11,12]. Many methods can be found in the literature for that purpose. Without being exhaustive we can mention: the method introduced by Hurst in his original paper using rescaled range statistical analysis (RS analysis) [9], the modified RS analysis proposed by Lo [13], detrended fluctuation analysis (DFA) [7], the wavelet maximum likelihood method [14], the periodogram method [15], the Whittle's estimator [16], and the linear regression of the FFT-estimated power spectrum after log transformation of frequency [11].

In the last decade, ordinal patterns analysis has constituted a significant advance in the treatment of time series originated from regular, chaotic and stochastic dynamics [17–19]. Particularly, a detailed theoretical analysis of the ordinal patterns relative frequencies for Gaussian and autoregressive moving average (ARMA) processes was developed by Bandt and Shiha [20]. We put especial attention to the link between the Hurst exponent and the ordinal pattern probabilities for fractional Brownian motion (fBm), which turns out to be a widely accepted theoretical framework to model fractal phenomena which have a power spectrum. In a later work, Sinn and Keller [21], analysed the statistical properties of estimators of ordinal patterns probabilities in discrete time Gaussian processes with stationary increments. They introduced the *Zero Crossing* estimator for the Hurst exponent in fBms. In addition, an empirical data application (Nile river data) was reported.

In this paper, we propose to apply the ordinal patterns analysis developed by Bandt and Shiha [20] to estimate the Hurst exponent for several real data sequences. In order to move forwards in their original proposal, and as a first step, we study numerical simulations of fBms in a noisy environment and with artificial low resolution. As a second step, in a real data scenario, we propose a multiscale symbolic approach by the estimation of the Hurst exponent for different time scales. When dealing with empirical data recorded at discrete frequency sampling (digitization), the range of temporal scales at which a reliable estimation of H can be obtained is unknown. If one is restricted to the finer temporal scales, H could be over or underestimated, as will be discussed in detail below. This multiscale analysis will provide an optimum range of temporal scales for the estimation of the Hurst exponent. Our results are compared with those obtained from the widely accepted DFA methodology [7].

The paper is organized as follows: in Section 2, we briefly describe the ordinal analysis in time series and the ordinal patterns probabilities for fractional Brownian motions. In Sections 3 and 4 numerical results and experimental applications, respectively, are presented and discussed. Finally, some concluding remarks are given in Section 5.

2. Theoretical background

Bandt and Pompe (BP) introduced a simple and robust symbolic method that takes notice of time causality in dealing with a given system dynamics [17]. This symbolization arises naturally from a given data sequence without any model assumptions. "Partitions" are devised by appropriately ranking the series' values rather than by apportioning amplitudes according to different levels. Even though this procedure entails losing some details of the original time series' amplitude information, by just referring to the series' intrinsic structure, a meaningful difficulty reduction is indeed achieved by BP recipe with regard to the description of complex systems. Its use has been shown to yield a clear improvement in the estimation quality of information theory-based quantifiers (please see Refs. [18,22–24] and references therein).

To build ordinal patterns from a given time series, first one has to choose two parameters, the embedding dimension $D \geq 2$ ($D \in \mathbb{N}$, the pattern length) and the embedding delay τ ($\tau \in \mathbb{N}$, the time separation between the values). After that, the time series is partitioned into subsets of length D of consecutive ($\tau = 1$) or non-consecutive ($\tau > 1$) values which allows for an accurate empirical reconstruction of the underlying phase-space. The elements of the embedded data are replaced by their ranks in the subset generating an ordinal pattern. Let us consider the following data sequence as an example: $x_i = \{2, 5, 11, 18, 9, 7, 4\}$. By setting $D = 3$ and $\tau = 1$ the first two triplets $(x_0, x_1, x_2) = (2, 5, 11)$ and $(x_0, x_1, x_2) = (5, 11, 18)$ are mapped to the pattern $\{012\}$ since the values are originally placed in ascending order. The third triplet $(x_0, x_1, x_2) = (11, 18, 9)$ gives $\{201\}$ since $x_2 < x_0 < x_1$. Finally the last two 3-dimensional vectors $(x_0, x_1, x_2) = (18, 9, 7)$ and $(x_0, x_1, x_2) = (9, 7, 4)$ are mapped to the pattern $\{210\}$ since $x_2 < x_1 < x_0$. Then, a probability distribution $P = \{p(\pi_i), i = 1, \ldots, D!\}$ can be obtained by counting the relative frequencies of the $D!$ possible permutations π_i. For the present example: $p(\pi_{012}) = p(\pi_{210}) = 2/5$, $p(\pi_{201}) = 1/5$, and $p(\pi_{021}) = p(\pi_{102}) = p(\pi_{120}) = 0$. In order to get a unique result BP suggest that if two consecutive values are equal, one must rank them according to their order of emergence. This is justified by assuming that the values x_i have a continuous distribution, so that equal values are unusual. Otherwise it is possible to break these equalities by adding small random perturbations [17].

The probability distribution P is obtained once we choose the embedding dimension D and the embedding delay τ. The value of D determines the number of accessible states, given by $D!$. Moreover, it has been established that the length N of the time series must satisfy the condition $N \gg D!$ in order to achieve a reliable statistics in the estimation of P. With respect to the selection of the parameters, BP suggest in their cornerstone paper to work with $3 \leq D \leq 7$ and a time lag $\tau = 1$. Nevertheless, other values of τ might provide additional information. It has been showed that this parameter is strongly related, when it is relevant, to the intrinsic time scales of the system under analysis [19,25–28]. By changing the value of the embedding delay τ different time scales are being considered since τ physically corresponds to multiples of the sampling time of the signal under analysis.

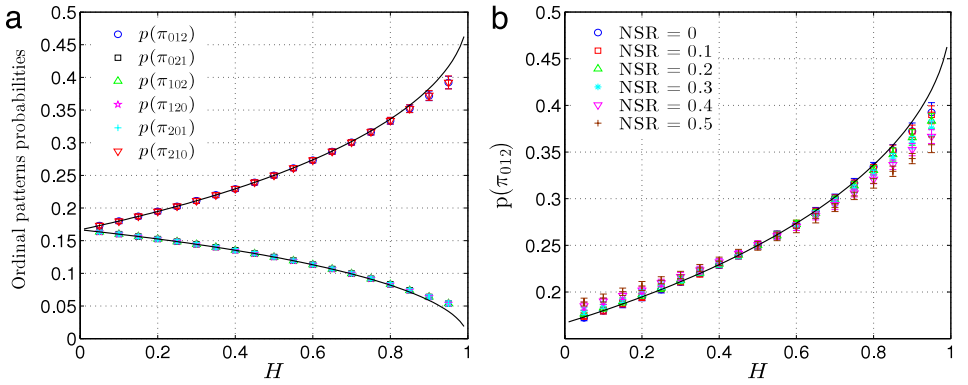

Fig. 1. (a) Probabilities of the six ordinal patterns (with $\tau = 1$) as a function of the Hurst exponent for fBm processes; (b) The probability $p(\pi_{012})$ for different noise-to-signal ratio (NSR) conditions. Numerical results are plotted. Mean and standard deviation from 1000 independent fBm simulations of length $N = 10,000$ data point each are depicted ($H \in \{0.05, 0.1, \ldots, 0.95\}$). Analytical curves (continuous line) are also included for comparison purposes.

Fractional Brownian motion (fBm) is a family of non-stationary stochastic processes which is Gaussian, self-similar, and endowed with stationary increments [fractional Gaussian noise (fGn)], widely used for modelling fractal phenomena that have empirical spectra of power-law type $1/f^{\alpha}$ and $\alpha = 2H + 1$ with $1 < \alpha < 3$ [29]. Its temporal correlations are characterized by the Hurst exponent $H \in (0, 1)$. These processes exhibit temporal memory for any Hurst parameter except for $H = 1/2$, which corresponds to classical Brownian motion (random walk). Thus, the Hurst parameter defines two distinct regions in the interval $(0, 1)$. When $H > 1/2$, consecutive increments tend to have the same sign so that these processes are persistent. For $H < 1/2$, on the other hand, consecutive increments are more likely to have opposite signs, and the underlying temporal dynamics are anti-persistent [30].

The ordinal patterns probabilities associated with different stochastic processes have been analysed by Bandt and Shiha (BS) [20]. They provide us with theoretical expressions for the relative frequencies $p(\pi_i)$ for fractional Brownian motion with arbitrary time delay τ, and $D = 3$ and $D = 4$. Particularly for $D = 3$, BS have shown that

$$p(\pi_{012})(\tau) = \frac{1}{\pi} \arcsin\left(2^{H-1}\right) \quad \forall\, \tau, \tag{1}$$

with H the Hurst exponent. Moreover, the ordinal patterns probabilities exhibit a particular symmetry: $p(\pi_{012}) = p(\pi_{210})$ and $p(\pi_{021}) = p(\pi_{102}) = p(\pi_{120}) = p(\pi_{201}) = (1/2 - p(\pi_{012}))/2$. Thus, for a given value of H one can compute analytically, all the theoretical components of the ordinal patterns probability distribution, and vice versa. On the other hand, for the case $D = 4$, the theoretical expressions are more complicated and for a given probability distribution the value of H can be obtained only numerically. For that reason, in the present study, the embedding dimension used was $D = 3$ and, for the sake of simplicity from hereafter it will not be explicitly stated. For a detailed study of these analytical results applied to permutations quantifiers in fBm and its stationary increments (fGn) please see Ref. [31].

3. Numerical analysis

In order to study the behaviour of the ordinal patterns probabilities as a function of the Hurst exponent, we compare the analytical values with numerical simulations for fBm processes. These results are depicted in Fig. 1(a). More specifically, mean and standard deviation of the estimated ordinal patterns probabilities (with $\tau = 1$) for 1000 independent fBm realizations of length $N = 10,000$ data points each, with Hurst exponent $H \in \{0.05, 0.1, \ldots, 0.95\}$, are shown. The simulated fBm time series were generated by implementing the method of Wood and Chan [32]. We find that the agreement between theoretical curves and numerical simulations is excellent, except for the simulated fBm with H close to 1. Taking into account that artificial series generators are obviously not exact, it is not surprising to find differences between numerical and theoretical results. Similar results were obtained using the function *wfbm* of MATLAB for simulating fBms, that follows the algorithm proposed by Abry and Sellan [33].

Since our purpose is to analyse temporal sequences of measurements (or observations), which always have a noise component, we have tested the effect that an additive observational noise has on the proposed approach. Taking into account the low sensitivity to noise of the symbolic BP methodology, robustness in a noisy environment is expected in the estimation of the probability distribution of the ordinal patterns. To perform this analysis, Gaussian white noises with different noise-to-signal ratio (NSR) conditions were added to the increments of the original simulations of the fBm processes. The NSR is defined as the standard deviation of the added noise over the standard deviation of the stationary increments of the original signal (process). Fig. 1(b) shows the influence of this additive noise on the estimation of the ordinal pattern probability $p(\pi_{012})$ as a function of H. Same quantitative results for the pattern probability $p(\pi_{210})$, and similar qualitative results for the rest of the probabilities are found, not shown for the sake of a better visualization. It can be concluded that reliable estimation

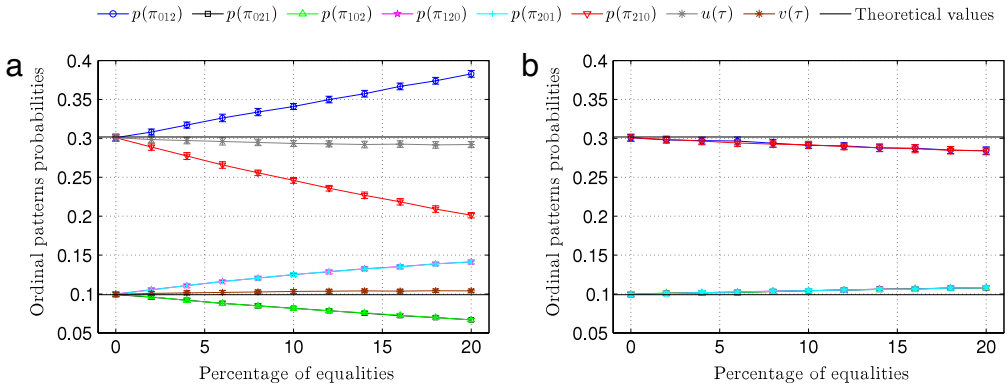

Fig. 2. Probabilities of the six ordinal patterns (with $\tau = 1$) as a function of the percentage of equalities in a fBm process with $H = 0.7$. Probabilities were estimated by (a) using the crude time series. (b) Adding a random perturbation. Average probabilities u and v are included in (a). Theoretical values of the probabilities for $H = 0.7$ are also included for comparison purposes (solid black lines). Note the superposition of the ordinal patterns probabilities ($p(\pi_{021}) = p(\pi_{102})$ and $p(\pi_{120}) = p(\pi_{201})$ in (a), $p(\pi_{012}) = p(\pi_{210})$ and $p(\pi_{021}) = p(\pi_{102}) = p(\pi_{120}) = p(\pi_{201})$ in (b)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the ordinal probabilities is performed up to a 20% of observational noise. It is worth noting that the influence of the additive noise is less significant for intermediate H values, since those processes are already close to a random walk ($H = 0.5$).

Other important issue in dealing with experimental sequences, is a possible low resolution in the data acquisition. When an observed time series is digitized with low resolution, a greater number of equal values in the sequence is generated. This could entail possible wrong interpretations when an ordinal analysis is used, since the pattern π_{012} will be overestimated, generating spurious temporal correlations [17,34]. In that frame, we have tested how the ordinal approach responds to simulated fBm processes with artificial low resolutions. To perform this analysis, a percentage of null values was generated in the simulations of the stationary increments (fGn) by substitution of the original values in arbitrary temporal positions, which leads to an amount of equalities in the process (fBm) after the integration of the time series. In Fig. 2(a), mean and standard deviation of the estimated ordinal pattern probabilities (with $\tau = 1$) for 100 independent fBm realizations with $H = 0.7$, of length $N = 10{,}000$ data points each, as a function of the percentage of equalities in the sequence are shown. Similar qualitative results are obtained for other values of H, not displayed here because of space saving reasons. It is observed that the presence of equalities leads to a spurious symmetry in the ordinal patterns, since their probabilities break the original clustering. On the one hand, $p(\pi_{012})$ and $p(\pi_{210})$ are no longer equal because they are over and underestimated, respectively. On the other hand, we observe two new clusters, *i.e.* $p(\pi_{021}) = p(\pi_{102})$ and $p(\pi_{120}) = p(\pi_{201})$ (please see Fig. 2(a)).

To solve this artificial symmetry, the BP proposal, *i.e.* adding a very small random perturbations, indeed recovers approximately the symmetry of the original process as can be seen in Fig. 2(b). Here, we propose a second option. To average over the probabilities patterns which have to be equal by theoretical definition, and numerically proved above, that is

$$u(\tau) = \frac{1}{2}(p(\pi_{012}) + p(\pi_{210})), \tag{2}$$

$$v(\tau) = \frac{1}{4}(p(\pi_{021}) + p(\pi_{102}) + p(\pi_{120}) + p(\pi_{201})). \tag{3}$$

The averaged probability $u(\tau)$ quantifies the degree of persistence in the series [20]. It can be seen in Fig. 2(a) (please see grey and brown stars) that by reducing the six probabilities to the two averaged probabilities, $u(\tau)$ and $v(\tau)$, it is possible to avoid the effect of a low resolution in a fBm process. Besides, these averaged probabilities are located closer to the theoretical values as the number of equalities grows, in comparison with the BP proposal. From this analysis, we conclude that an improved estimation of the Hurst exponent by using an ordinal analysis of a real measured data associated with a fBm behaviour can be achieved by the estimation of the averaged probability $u(\tau)$ in the process. Thus, we propose a Hurst exponent estimator, $\widehat{H}_p(\tau)$, using Eqs. (1) and (2), as follows

$$\widehat{H}_p(\tau) = \frac{\ln(\sin[\pi u(\tau)])}{\ln(2)} + 1. \tag{4}$$

This last expression permits us an estimation of the Hurst exponent for different temporal scales by varying τ, *i.e.* a multiscale estimation of H, which results necessarily in real data as will be discussed in detail in the next section. It should be remarked that for a "pure" fBm the estimator $\hat{H}_p$ is independent of the time delay τ because of the self-similarity property of this process. For further details, please see Appendix A.

Although a small embedding dimension is used, a multiscale approach has been proposed, therefore finite-size effects might be seen. In order to illustrate finite-size effects on the present approach, we studied the $\widehat{H}_p$ estimator for a Brownian

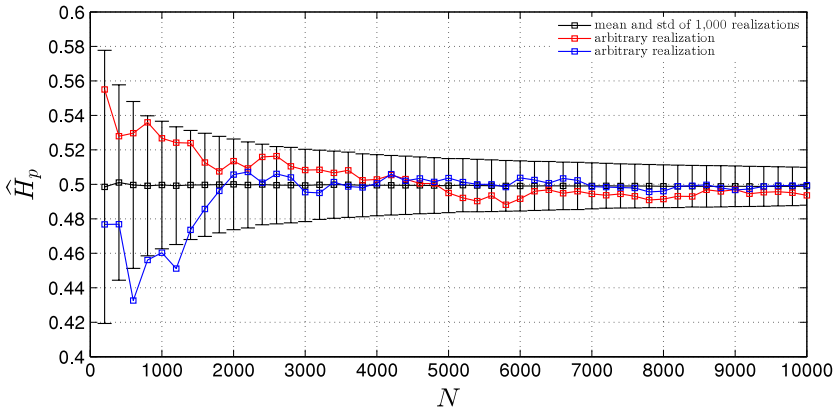

Fig. 3. Mean and standard deviation from 1000 independent random walks of the $\widehat{H}_p(\tau = 1)$ estimator values versus the time series length. Two arbitrary realizations are also included (red and blue squares). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

motion ($H = 0.5$) for different series length $N \in \{200;\ 400;\ 600;\ \ldots;\ 10,000\}$ data points. Mean and standard deviation of the $\widehat{H}_p(\tau = 1)$ estimator values for 1000 independent realizations for each length are depicted in Fig. 3. The particular behaviour for two arbitrary realizations are also depicted. Huge fluctuations are observed for small lengths. These fluctuations tend to die out to converge to the expected value of H for longer time series. The behaviour depicted is representative for other values of H. Thus, for finite time series' length an over or underestimated value of H could be obtained. For completeness sake, we have included a detailed study of the root mean square error associated with the estimator $\hat{H}_p(\tau = 1)$ for different Hurst exponent H and for different time series length N in Appendix B.

We have compared the computational efficiency of the present symbolic approach with the DFA methodology. The running times in the estimation of H from 10 independent realizations of a fBm with $H = 0.5$ were estimated. For the ordinal symbolic technique we have implemented an estimator $\langle \widehat{H}_p \rangle$ with $\tau = 1$ and also with $\tau \in [1, 50]$. A detrending polynomial of second order and one hundred time scales between $[10, N/4]$ were employed in the DFA implementation following the algorithm proposed by Ihlen in Ref. [35]. The linear fit for estimating H was performed in the whole range of scales. We have found a difference of two orders of magnitude in the running times for time series length $N \in [10^3, 10^6]$, being our symbolic technique, the faster. We have confirmed that this result is independent of the Hurst exponent value. For more details of this analysis please see Appendix C.

4. Experimental applications

Here, experimental time series will be analysed and the usual course of action is assuming that they can be suitably modelled as a "pure" fBm. Thus, an estimation of H from $\hat{H}_p$ would be invariant with τ (please see Appendix A). However, in practical contexts, the expected invariance is realizable only in a limited range of temporal scales. Consequently, a multiscale estimation of the Hurst exponent is needed for searching this scale range at which a τ-invariance of $\hat{H}_p$ is observed. Otherwise, erroneous estimation of H may occur. It is also well-known that in some cases there exist crossover time scales separating regimes with different scaling exponents, *e.g.* long-range correlations on small scales and another type of correlations or uncorrelated dynamics on larger scales [10]. Therefore, a multiscale estimation for different values of τ is of vital importance when dealing with this kind of behaviours.

For comparison purposes we have estimated the Hurst exponents, for all the experimental examples included in this section, using the traditional DFA methodology [7,10], which has been widely proved to be robust in the analysis of experimental data. For its implementation in MATLAB we recommend Ref. [35]. Hurst exponent estimations with a detrending polynomial of second order are reported.

4.1. Human gait

As a first experimental application we have studied the fractal dynamics of human gait. In 1996 Hausdorff et al. analysed data obtained from ten healthy subjects who walked one hour at their usual pace. They found that Stride Interval Fluctuations (SIF) exhibited long-range correlations [1]. The data consist of 10 time series of SIF of lengths in a range between 2902 and 3397 data points (available under http://physionet.org/physiobank/database/umwdb/). Participants had no history of any neuromuscular, respiratory or cardiovascular disorders, and were taking no medications. Subjects walked continuously on level ground around an obstacle free path, and the stride interval was measured using ultra-thin, force sensitive switches taped inside one shoe. The recorded gait signal was digitized, and the time between foot strikes was automatically computed. For more details, please see Ref. [1].

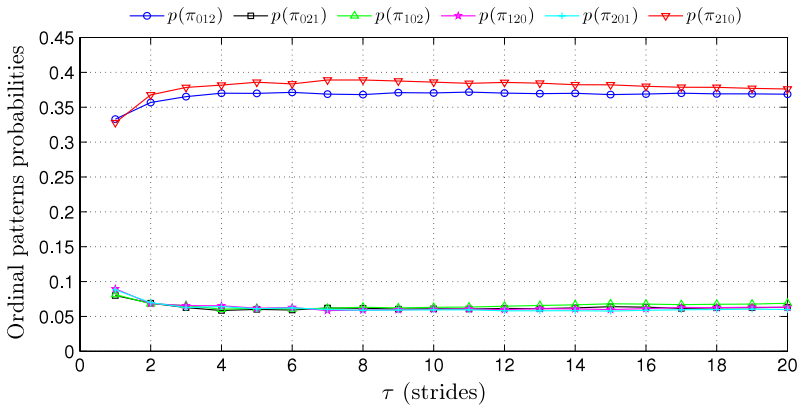

Fig. 4. Average of the ordinal patterns probabilities of the 10 subjects as a function of the embedding delay τ.

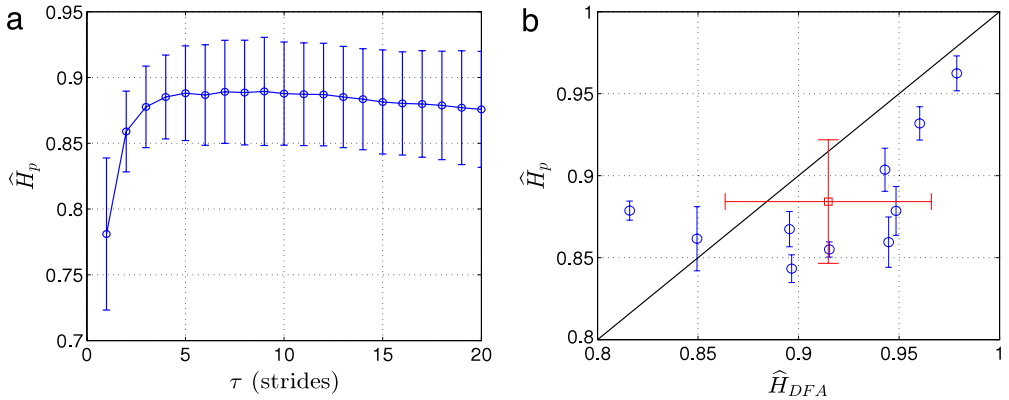

Fig. 5. (a) Mean and standard deviation of the multiscale $\widehat{H}_p$ estimator for the 10 subjects, as a function of τ. (b) Mean and standard deviation of the $\widehat{H}_p$ estimator for $\tau \in [4, 20]$ versus Hurst exponents estimated by DFA. Mean and standard deviation of the 10 values of H estimated with the two methodologies are also included (red cross). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As it was mentioned previously, a priori, the optimal range of time scales to estimate a Hurst exponent H from an experimental time series is unknown. Thus, we implement a multiscale strategy by analysing the behaviour of the ordinal patterns probabilities as a function of τ. In Fig. 4, the average of the ordinal patterns probabilities of the 10 subjects versus τ is depicted. We observe that the values of the probabilities tend to stabilize after $\tau = 4$. They exhibit the patterns' symmetry expected for a fBm process. The variations in the probabilities with τ will influence, of course, the estimation of the Hurst exponent, as can be seen in Fig. 5(a), where mean and standard deviation of the multiscale $\widehat{H}_p(\tau)$ estimator for the 10 subjects are shown. It can be seen that $\widehat{H}_p$ stabilizes for all the scales larger than $\tau = 4$. From these results we conclude that a Hurst exponent can be estimated by averaging over the time scales $\tau \in [4, 20]$. Mean and standard deviations of $\widehat{H}_p$ over the selected range of τ for each subject compared with the exponents estimated by DFA are depicted in Fig. 5(b). Our methodology quantifies a strong persistent behaviour characterized by $\langle \widehat{H}_p \rangle = 0.88 \pm 0.04$, meanwhile DFA estimates an exponent $\langle \widehat{H}_{DFA} \rangle = 0.91 \pm 0.05$. It is important to note that the mean $\langle\rangle$ is the average over the 10 subjects.

4.2. Heartbeat

As a second practical application, we have analysed normal physiological data [36]. More precisely, sequential time intervals between consecutive beats of long-term ECG recordings of subjects in normal sinus rhythm are considered. The original ECG recordings were digitized at 128 samples per second, and the beat annotations were obtained by automated analysis with manual review and correction. The data consist of 18 time series of length in a range between 75,106 and 115,911 data points, of subjects who did not have significant arrhythmias (they include 5 men, aged 26–45, and 13 women, aged 20–50). Data are available at http://physionet.org/physiobank/database/nsrdb.

We have applied a multiscale approach by analysing the behaviour of the probabilities of the patterns as a function of τ. Results are shown in Fig. 6(a) for low time scales ($\tau \leq 14$). We observe that there exists, for $\tau = 1$, a spurious symmetry due to a possible low resolution in the data because of the digitization of the original ECG recordings. This result illustrates

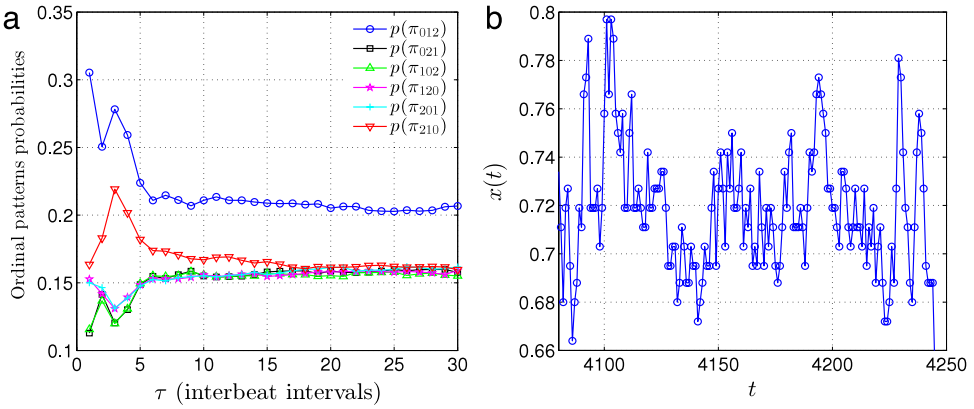

Fig. 6. (a) Probabilities of the six ordinal patterns of the time intervals between consecutive beats of an arbitrary subject as a function of the embedding delay τ. (b) A portion of sequential time intervals between consecutive beats of an arbitrary subject.

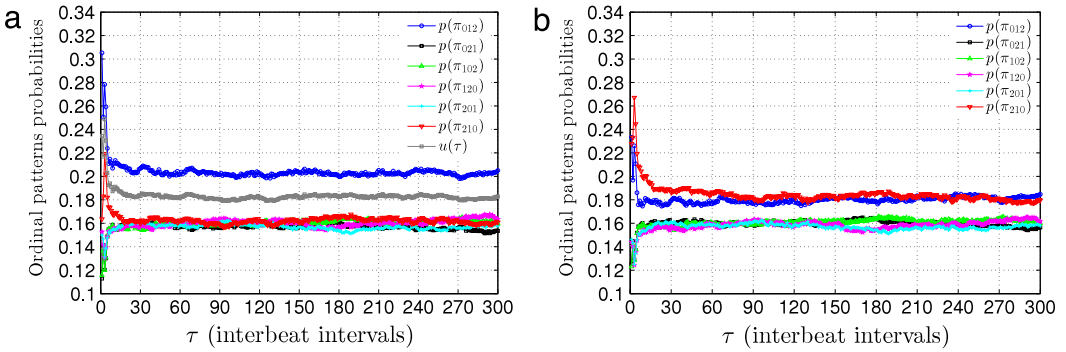

Fig. 7. Probabilities of the six ordinal patterns of the time intervals between consecutive beats of an arbitrary subject (a) using the crude signal and the averaged probability $u(\tau)$, (b) using the signal with a small random perturbations.

the expected spurious symmetry in the ordinal patterns probabilities seen in the simulated data (please see Fig. 2(a)). The artificial symmetry disappears as τ grows, but the pattern {012} tends to keep overestimated even for larger values of τ, due to the presence of a considerable number of equalities in the data, as can be seen in Fig. 6(b), where a portion of measured data is depicted. In Fig. 7(a) the behaviour of the pattern's probabilities estimated from the crude data for larger values of τ is shown. It can be seen that after $\tau \sim 30$, stability is reached, but the pattern {012} is still overestimated. To break the equalities, we can work either, on the averaged probability $u(\tau)$ (please see Fig. 7(a) grey squares), or following the proposal of BP (please see Fig. 7(b)). In both cases we obtain the expected symmetry in the patterns of a fBm. These results associated with a particular subject are representative for the whole dataset.

From this analysis we conclude that we are able to estimate a reliable Hurst exponent for these experimental data from the averaged probability $u(\tau)$. Mean and standard deviation of the multiscale $\widehat{H}_p$ estimator are shown in Fig. 8(a). After an oscillation, stability of the $\widehat{H}_p$ for $\tau \geq 100$ is observed. We chose the interval $[100, 150]$ to estimate a value for H. Mean and standard deviation of the $\widehat{H}_p$ versus the $\widehat{H}_{DFA}$ are shown in Fig. 8(b). An anti-persistent dynamics is found, characterized by $\langle \widehat{H}_p \rangle = 0.09 \pm 0.03$. On the other hand, DFA methodology estimates $\langle \widehat{H}_{DFA} \rangle = 0.05 \pm 0.03$.

4.3. Significant wave height

The dynamic of the ocean waves is a complex system mainly driven by a combination or interaction of wind and swell waves. The significant wave height (SWH) is generally used to represent this interaction [4]. SWH is approximately equal to the average of the highest one-third of the waves, as measured from the trough to the crest of the waves. Specifically, it is calculated as SWH $= 4 * \sqrt{m_0}$ where m_0 is the variance of the wave displacement time series acquired during the wave acquisition period. We analysed a SWH data previously investigated by Ozger in Ref. [4], available at http://www.ndbc.noaa.gov/wavecalc.shtml. Ozger found that there are two distinct scaling regions corresponding to finer and coarser scales, characterized by Brownian motion and persistent fluctuations, respectively [4]. The data originates from 25 buoys, located off the west coast of the US. The resolution is hourly. The range of data length is between 16,319 and 61,368 data points.

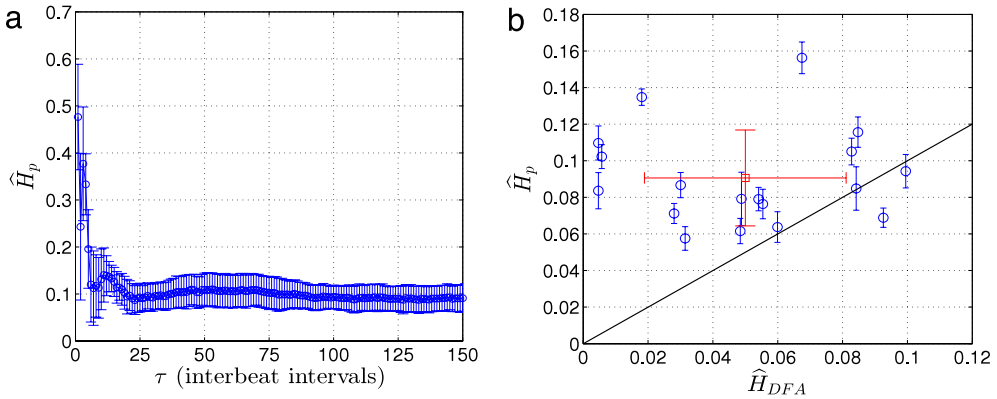

Fig. 8. (a) Mean and standard deviation of the multiscale $\widehat{H}_p$ estimator for the 18 subjects as a function of τ. (b) Mean and standard deviation of the $\widehat{H}_p$ estimator for $\tau \in [100, 150]$ versus Hurst exponents estimated by DFA. Mean and standard deviation of the 18 values of H estimated with the two methodologies are also depicted (red cross). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

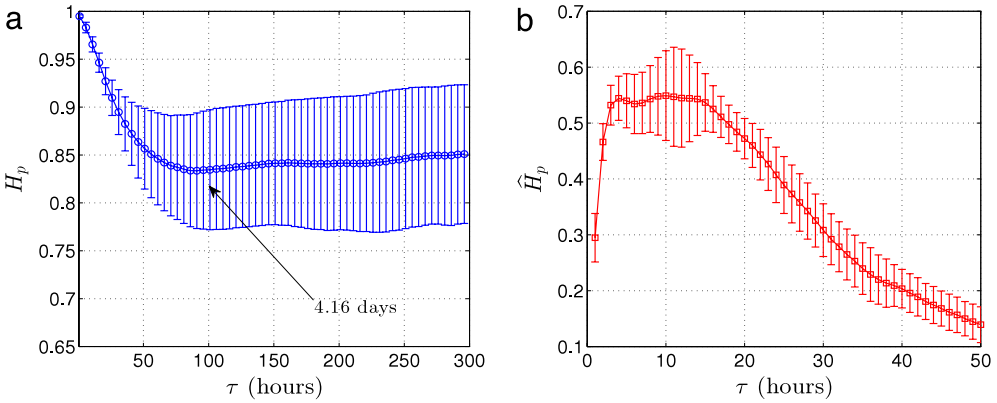

Fig. 9. Mean and standard deviation of the multiscale $\widehat{H}_p$ estimator for the 25 buoys as a function of τ; (a) From the integrated SWH-fluctuations, (b) From the original SWH-data.

As a first step, we analysed the behaviour of the multiscale $\widehat{H}_p$ estimator for the 25 integrated time series (processes). In Fig. 9(a), mean and standard deviation of the $\widehat{H}_p$ are shown. We report $\tau \in \{1, 5, 10, 15, \ldots, 200\}$ for better visualization. We observe that for $\tau = 1, \widehat{H}_p$ saturates to its maximum possible value (unity) and after $\tau \geq 100$, it stabilizes to a constant value. In real time, $\tau = 100$ corresponds to 4.16 days, which coincides with the crossover time reported by Ozger [4]. The maximum value of $\widehat{H}_p(\tau = 1)$ may be due to the fact that for small time scales the dynamic of the data sequences is affected by the integration of the series. Therefore, as a second step, we study the multiscale $\widehat{H}_p$ estimator for the 25 series without integration. Fig. 9(b) shows mean and standard deviation of the $\widehat{H}_p$ versus τ for small scales. Now, the $\widehat{H}_p$ estimator reaches a *plateau* for $\tau \in [3, 15]$. From this analysis, we conclude that for small scales the data sequences measured can be associated with a stochastic process, and for long scales they behave as stochastic fluctuations. Then, the integration of the data is needed to estimate a Hurst exponent for long scales. We have estimated H for the 25 data sequences with and without integration as the average of $\widehat{H}_p$ for $\tau \in [100, 300]$ and $\tau \in [3, 15]$, respectively. In Fig. 10 the estimated Hurst exponents compared with those obtained by using DFA methodology are depicted. We found a good agreement between both methodologies.

From this multiscale analysis we conclude that for small time scales, a slightly persistent stochastic process with $\langle \widehat{H}_p \rangle = 0.54 \pm 0.06$ is observed from the SWH-data without integration. This dynamic is originated by the sea waves, specifically they are surface waves as a result from the wind blowing over the surface, with limited predictability. On the other hand, for long time scales, we estimated $\langle \widehat{H}_p \rangle = 0.84 \pm 0.07$ for the integrated SWH-fluctuations, corresponding to a very persistent stochastic behaviour that describes the dynamics of swell waves, which consists of wind-generated waves that are not significantly affected by the local wind at that time and are travelled into the measurement area.

4.4. Financial time series

As a last practical example, we study the Dow–Jones Industrial Average (DJIA) closing values from 1900 to 1951. This time series was previously widely investigated using different methodologies reporting the presence of an uncorrelated

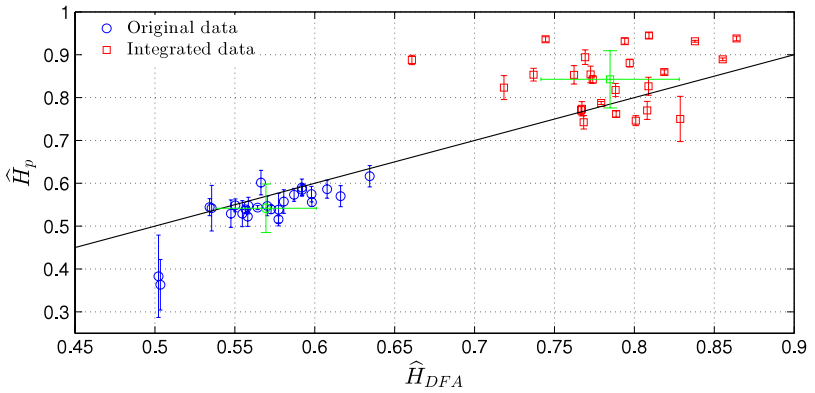

Fig. 10. Mean and standard deviation of the $\widehat{H}_p$ estimator versus $\widehat{H}_{DFA}$ for the SWH-data without integration ($\tau \in [3, 15]$, blue circles) and for the integrated SWH-fluctuations ($\tau \in [100, 300]$, red squares). Mean and standard deviation of the 25 values estimated with the two methodologies for both cases are also depicted (green crosses). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

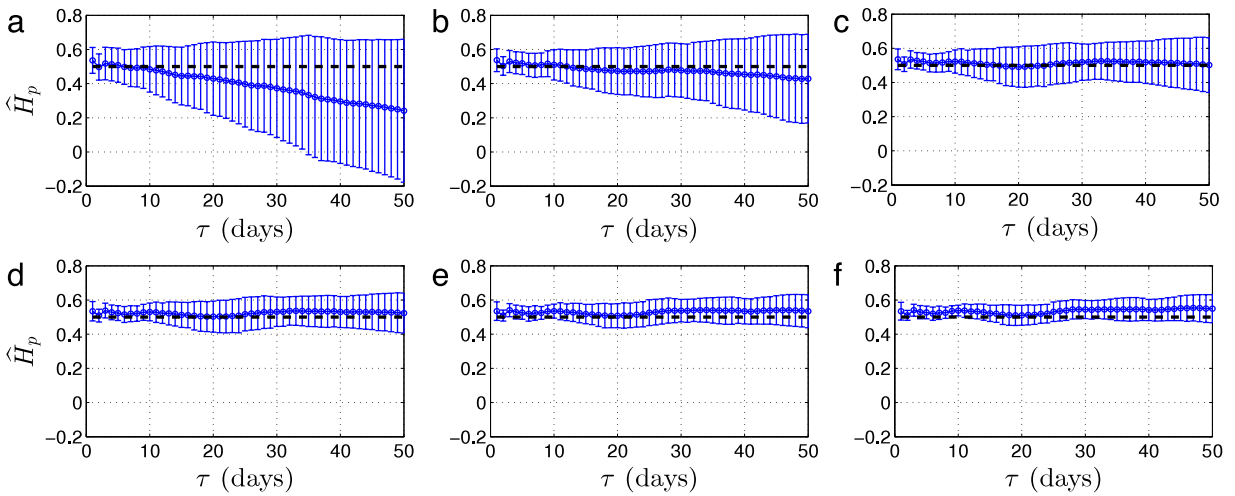

Fig. 11. Mean and standard deviation from the dynamical $\widehat{H}_p$ estimator as a function of τ for windows of size (a) 1 year, (b) 2, (c) 3, (d) 4, (e) 5, and (f) 6 years (each year is approximately 274 data points). The black dashed line indicates $H = 0.5$.

dynamics ($H \simeq 0.5$ in Ref. [11] and $H = 0.51$ in Ref. [12]). Thus it is a good example to test the performance of the proposed ordinal approach. The closing values were recorded approximately 311 times a year and are available at http://lib.stat.cmu.edu/datasets/djdc0093. After removing several spikes (possibly due to transcription errors [11]) the length of the series is 15,309 data points and a year is reduced to approximately 274 data points. Particularly, following the work of Pilgram et al. [11], we are interested in a dynamical analysis of the data by sliding time windows. Besides, this analysis will illustrate finite-size effects on the $\widehat{H}_p$ estimator.

Taking into account that a dynamical analysis could entail some finite-size effects (please see Fig. 3), we propose to do different dynamic multiscale estimations of the Hurst exponent by sliding windows. Specifically, windows of size $L \in \{1, 2, \ldots, 6\}$ years, with a shift equal to 1 data point were used. In Fig. 11, mean and standard deviation from the dynamical $\widehat{H}_p$ estimator for the different window sizes as a function of τ are shown. Clearly, the finite-size effects disappear as the window size grows by reducing the dispersion and the underestimation of the Hurst exponent. A more stable behaviour of $\hat{H}_p$ with τ is found for the larger window sizes (three to six years). Therefore, as illustrated here, we confirm that the sequence of DJIA closing values can be associated with a random walk process when using sliding windows of size greater than 3 years (822 data points approximately).

5. Conclusions

In this work, we present an extended study of the estimation of the Hurst exponent of real data sequences from a multiscale ordinal patterns probability analysis. The proposed approach is a generalization of a theoretical result, originally introduced by Bandt and Shiha [20], for multiple time scales. We perform numerical studies to illustrate robustness in a noisy environment, and the effects of low resolution in the data acquisition and finite-size series length.

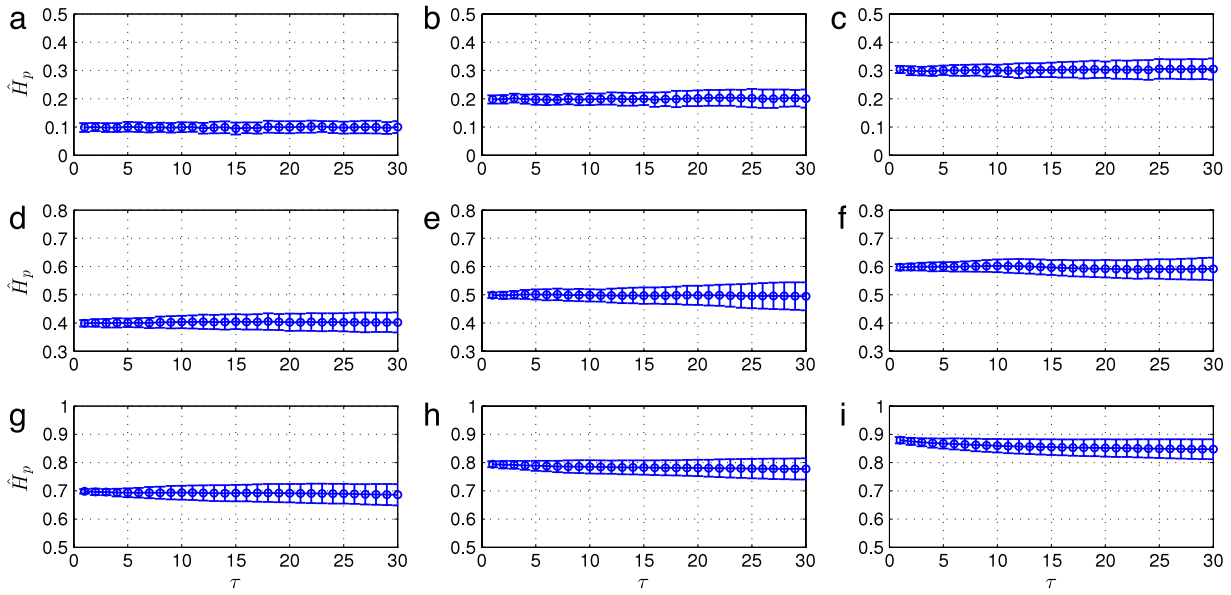

Fig. 12. Mean and standard deviation of the multiscale $\hat{H}_p$ estimator for 100 independent realizations of a fBm with length $N = 10,000$ as a function of τ, for (a) $H = 0.1$, (b) $H = 0.2$, (c) $H = 0.3$, (d) $H = 0.4$, (e) $H = 0.5$, (f) $H = 0.6$, (g) $H = 0.7$, (h) $H = 0.8$, (i) $H = 0.9$.

Empirical evidences of the last two situations mentioned above were illustrated by the heart rate dynamics and financial time series analysis, respectively. We exemplify the performance of the method in a time series with a crossover between two different stochastic behaviours by studying the dynamics of the significant wave height. In the analysis of real data, there exists an optimum range of time scales at which stability of the estimated Hurst exponents is observed, *i.e.* $\hat{H}_p$ is invariant with τ. Finally, all the results reported were contrasted with those obtained with the widely accepted DFA methodology, confirming the reliability of the proposed multiscale ordinal pattern approach.

Acknowledgements

The authors thank reviewers for their constructive comments and suggestions. This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. Felipe Olivares also thanks support from Pontificia Universidad Católica de Valparaíso. Luciano Zunino acknowledges partial support from Universidad Nacional de La Plata, Argentina (Incentive Project 11/I170).

Appendix A

Self-similar stochastic processes are invariant in distribution under suitable scaling of time. Therefore, the estimator $\hat{H}_p$ for a perfect, theoretical fBm is independent of the parameter τ. Fig. 12 shows the mean and standard deviation of $\hat{H}_p$ for 100 independent realizations of a fBm with length $N = 10,000$ as a function of τ for different values of H. Clearly, the estimation of the Hurst exponent results invariant over τ. In addition we have studied how this invariance changes with the time series length. Fig. 13 depicts the mean and standard deviation of $\hat{H}_p$ for $\tau \in [1, 30]$ in one arbitrary realization for each $H \in \{0.1, 0.2, \ldots, 0.9\}$. As expected, the dispersion in the estimation decreases as the time series length grows.

Appendix B

We have estimated the root mean square error (RMSE) associated with the proposed estimator $\hat{H}_p(\tau = 1)$ for 100 independent realizations of fBms with $H \in \{0.1, \ldots, 0.9\}$ and time series lengths $N \in \{10^3, 10^4, 10^5, 10^6\}$. Results obtained are shown in Fig. 14. The pronounced bias observed for $H = 0.9$ is partially attributed to limitations of the simulation method for generating highly persistent fBms.

Appendix C

Fig. 15 shows the running times in the estimation of the Hurst exponent H from 10 independent realizations of a fBm with $H = 0.5$ using the $\hat{H}_p$ estimator (for $\tau = 1$ and $\tau \in [1, 50]$) and DFA methodology. These results are representative for all the values of H.

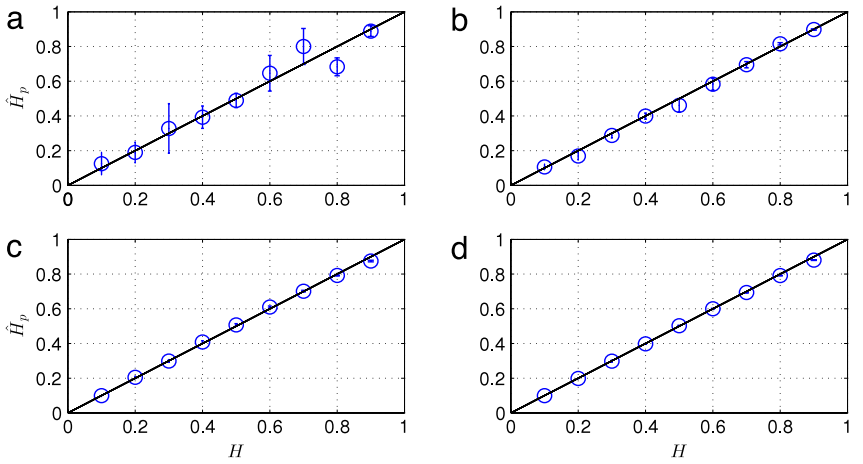

Fig. 13. Mean and standard deviation of the multiscale estimator $\hat{H}_p$ for $\tau \in [1, 30]$ as a function of H for different time series lengths (a) $N = 10^3$, (b) $N = 10^4$, (c) $N = 10^5$, (d) $N = 10^6$.

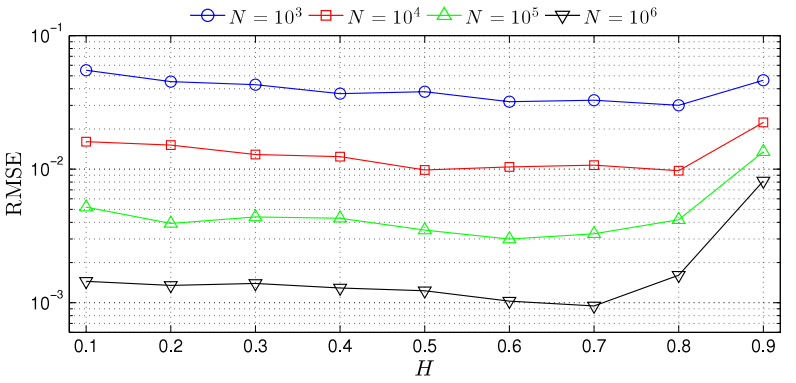

Fig. 14. Root mean square error of the estimator $\hat{H}_p$ with $\tau = 1$ for 100 independent realizations of fBm with different Hurst exponents and different time series length.

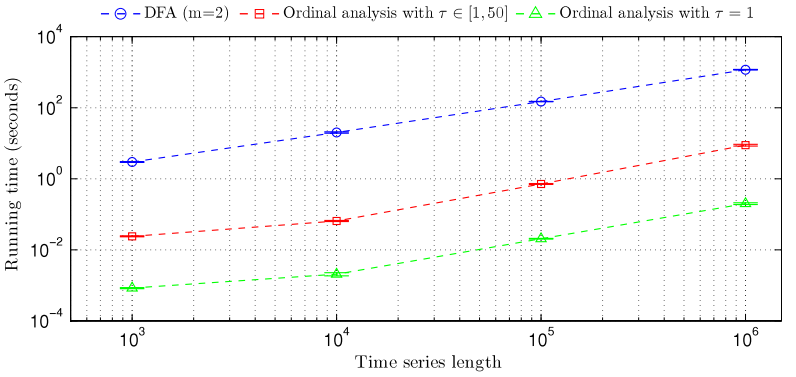

Fig. 15. Running times in the estimation of the Hurst exponent H from DFA methodology (blue circles) and the estimator $\hat{H}_p$, for $\tau = 1$ (green triangles) and $\tau \in [1, 50]$ (red squares), as a function of the time series length. Mean and standard deviation from 10 independent realizations of a fBm with $H = 0.5$ are reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

[1] J.M. Hausdorff, P.L. Purdon, C.-K. Peng, Z. Ladin, J.Y. Wei, A.L. Goldberger, J. Appl. Physiol. 80 (1996) 1448–1457.
[2] C.-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Phys. Rev. Lett. 70 (1993) 1343–1346.
[3] C.-K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Chaos 5 (1995) 82–87.
[4] M. Ozger, Physica A 390 (2011) 981–989.
[5] Y.H. Liu, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, Physica A 245 (1997) 437–440.

[6] P. Cizeau, Y.H. Liu, M. Meyer, C.-K. Peng, H.E. Stanley, Physica A 245 (1997) 441–445.
[7] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685–1689.
[8] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Matenga, M. Simons, H.E. Stanley, Physica A 221 (1995) 180–192.
[9] H.E. Hurst, Trans. Amer. Soc. Civ. Eng. 116 (1951) 770–808.
[10] J.W. Kantelhardt, E. Koscielny-Bunde, H.A. Rego, S. Havlin, A. Bunde, Physica A 295 (2001) 441–454.
[11] B. Pilgram, D.T. Kaplan, Physica D 114 (1998) 108–122.
[12] N. Makarava, S. Benmehdi, M. Holschneider, Phys. Rev. E 84 (2011) 021109.
[13] A.W. Lo, Econometrica 59 (1991) 1279–1313.
[14] E.J. McCoy, A.T. Walden, J. Comput. Graph. Statist. 5 (1996) 26–56.
[15] J. Geweke, S. Porter-Hudak, J. Time Ser. Anal. 4 (1983) 221–238.
[16] P. Whittle, J. R. Stat. Soc. Ser. B 15 (1953) 125–139.
[17] C. Bandt, B. Pompe, Phys. Rev. Lett. 88 (2002) 174102.
[18] O.A. Rosso, H.A. Larrondo, M.T. Martín, A. Plastino, M.A. Fuentes, Phys. Rev. Lett. 99 (2007) 154102.
[19] L. Zunino, M.C. Soriano, I. Fischer, O.A. Rosso, C.R. Mirasso, Phys. Rev. E 82 (2010) 046212.
[20] C. Bandt, F. Shiha, J. Time Ser. Anal. 28 (2007) 646–665.
[21] M. Sinn, K. Keller, Comput. Statist. Data Anal. 55 (2011) 1781–1790.
[22] O.A. Rosso, L.C. Carpi, P.M. Saco, M. Gómez Ravetti, A. Plastino, H.A. Larrondo, Physica A 391 (2012) 42–55.
[23] O.A. Rosso, L.C. Carpi, P.M. Saco, M. Gómez Ravetti, H.A. Larrondo, A. Plastino, Eur. Phys. J. B 85 (2012) 419.
[24] F. Olivares, A. Plastino, O.A. Rosso, Phys. Lett. A 376 (2012) 1577–1583.
[25] M.C. Soriano, L. Zunino, L. Larger, I. Fischer, C.R. Mirasso, Opt. Lett. 36 (2011) 2212–2214.
[26] M.C. Soriano, L. Zunino, O.A. Rosso, I. Fischer, C.R. Mirasso, IEEE J. Quantum Electron. 47 (2011) 252–261.
[27] L. Zunino, M.C. Soriano, O.A. Rosso, Phys. Rev. E 86 (2012) 046210.
[28] L. Zunino, F. Olivares, O.A. Rosso, Europhys. Lett. EPL 109 (2015) 10005.
[29] B.B. Mandelbrot, J.W. Van Ness, SIAM Rev. 10 (1968) 422–437.
[30] J. Feder, Fractals, Plenum, 1988.
[31] L. Zunino, D.G. Pérez, M.T. Martín, M. Garavaglia, A. Plastino, O.A. Rosso, Phys. Lett. A 372 (2008) 4768–4774.
[32] J.F. Coeurjolly, J. Stat. Softw. 5 (2000) 1.
[33] P. Abry, F. Sellan, Appl. Comput. Harmon. Anal. 3 (1996) 377–383.
[34] C. Bian, C. Qin, Q.D.Y. Ma, Q. Shen, Phys. Rev. E 85 (2012) 021906.
[35] E.A.F. Ihlen, Front. Physiol. 3 (2012) 141.
[36] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Circulation 101 (2000) 215–220.