Physics Letters A 381 (2017) 1021–1028

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

A simple and fast representation space for classifying complex time series

Luciano Zunino [a,b,*], Felipe Olivares [c], Aurelio F. Bariviera [d], Osvaldo A. Rosso [e,f,g]

[a] Centro de Investigaciones Ópticas (CONICET La Plata – CIC), C.C. 3, 1897 Gonnet, Argentina
[b] Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina
[c] Instituto de Física, Pontificia Universidad Católica de Valparaíso (PUCV), 23-40025 Valparaíso, Chile
[d] Department of Business, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain
[e] Instituto de Física, Universidade Federal de Alagoas (UFAL), BR 104 Norte km 97, 57072-970, Maceió, Alagoas, Brazil
[f] Instituto Tecnológico de Buenos Aires (ITBA) and CONICET, C1106ACD, Av. Eduardo Madero 399, Ciudad Autónoma de Buenos Aires, Argentina
[g] Complex Systems Group, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Av. Mons. Álvaro del Portillo 12.455, Las Condes, Santiago, Chile

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 17 January 2017
Accepted 26 January 2017
Available online 30 January 2017
Communicated by C.R. Doering

Keywords:
Time series analysis
Abbe value
Turning points
Financial data
Electroencephalogram data
Heart rate variability

ABSTRACT

In the context of time series analysis considerable effort has been directed towards the implementation of efficient discriminating statistical quantifiers. Very recently, a simple and fast representation space has been introduced, namely the number of turning points versus the Abbe value. It is able to separate time series from stationary and non-stationary processes with long-range dependences. In this work we show that this bidimensional approach is useful for distinguishing complex time series: different sets of financial and physiological data are efficiently discriminated. Additionally, a multiscale generalization that takes into account the multiple time scales often involved in complex systems has been also proposed. This multiscale analysis is essential to reach a higher discriminative power between physiological time series in health and disease.

1. Introduction

Typically, time series of measured variables are employed to analyze the dynamical behavior of complex systems. These temporal records need to be suitably characterized in order to reach a more reliable comprehension of the underlying nature of the phenomenon of interest. Obviously, this understanding is essential for modeling and forecasting purposes. In particular, numerous algorithms for quantifying the disorder and complexity of time series generated from nonlinear dynamical systems have been developed. Without being exhaustive, we can mention Lempel–Ziv complexity [1], correlation dimension [2,3], Lyapunov exponents [4,5], Kolmogorov [6], approximate [7], sample [8] and permutation [9] entropies, fractal [10] and multifractal [11] measures, and statistical complexity [12]. Moreover, combinations of these measures have been also proposed especially for discriminating and classifying dynamical systems. The usefulness of these multidimensional schemes has been confirmed for heterogeneous goals such as the distinction between noise and chaos [13,14], the characterization of a language corpus [15], the quantification of financial market efficiency [16,17], the automatic detection of epileptic seizure from electroencephalograms [18], the discrimination of songs in massive databases [19], and the classification of cardiac signals [10,20,21] and texture images [22], pointing out only a few of many applications. Despite the existing contributions, characterizing the underlying dynamics of complex system from time series is still a challenging problem of current research.

Tarnopolski has very recently introduced a representation space by plotting two statistical features associated with time series: the Abbe value and the number of turning points [23]. Numerical realizations of stationary and non-stationary long-range dependence stochastic processes are successfully discriminated in this plane. More precisely, fractional Brownian motion (fBm), fractional Gaussian noise (fGn), and differentiated fGn (dfGn) were found to form distinct branches in the proposed space. In this work, we go one step further by showing that this bidimensional scheme can be used as a discriminator of dynamics. Analysis of financial and physiological time series have been included for illustrating the

* Corresponding author at: Centro de Investigaciones Ópticas (CONICET La Plata – CIC), C.C. 3, 1897 Gonnet, Argentina.
E-mail addresses: lucianoz@ciop.unlp.edu.ar (L. Zunino), olivaresfe@gmail.com (F. Olivares), aurelio.fernandez@urv.cat (A.F. Bariviera), oarosso@gmail.com (O.A. Rosso).

http://dx.doi.org/10.1016/j.physleta.2017.01.047
0375-9601/

robustness of the technique when dealing with real time series. A multiscale generalization, inspired by the multiscale entropy algorithm proposed by Costa et al. [24], is introduced for unveiling hidden information over different levels of temporal resolution of the original signal. The higher discriminative power at particular time scales observed in the physiological applications confirms the advantages of implementing the proposed multiscale analysis. As it will be shown below, our results demonstrate that the Abbe value and the number of turning points are two distinctive features for identifying differences in complex systems dynamics. Consequently, the location in the representation space, that results from computing simultaneously both quantifiers, deserves special consideration for time series classification purposes.

The remainder of this paper is structured as follows. In Section 2, the Tarnopolski's diagram together with the proposed multiscale recipe and a couple of benchmark tests are discussed. The performance of the method as a diagnostic tool is analyzed in Section 3 through several real-world applications. Finally, in Section 4, the main results and conclusions of this work are summarized.

2. Methods

2.1. Tarnopolski plane

Giving a time series $\{x_i\}_{i=1}^{n}$, the Abbe value, denoted $\mathcal{A}$ in this paper, is defined as half of the ratio of the mean square successive difference to the variance,

$$\mathcal{A} = \frac{n}{2(n-1)} \frac{\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{1}$$

with $\bar{x}$ the mean of $\{x_i\}$ [25–27]. The Abbe statistic quantifies the smoothness of a time series: it is close to zero for time series displaying a high degree of smoothness while it tends to one for white noise [23]. According to our knowledge, very few works have implemented this measure for practical applications. Within these few exceptions, the Abbe value and another related measure, the excess Abbe value, have been successfully applied in stellar variability studies for identifying transients in large-scale surveys [28].

A turning point in a time series is observed when the middle value x_i of a sequence of three consecutive observations is lower or higher than the other two values, x_{i-1} and x_{i+1}, that surround it [29]. Equal values, *i.e.* $x_j = x_k$ for $j \neq k$, are neglected. This assumption is justified whenever $\{x_i\}_{i=1}^{n}$ has a continuous distribution. From an arbitrary time series the probability of finding a turning point, denoted by $\mathcal{T}$, can be empirically estimated by its relative frequency. In particular, $\mathcal{T}$ is asymptotically equal to 2/3 for random time series. It is important to stress here that estimating the turning points probability is equivalent to calculate the zero crossing rate (ZCR) of the differentiated time series. ZCR has been previously implemented for diverse applications, *e.g.* the detection of voiced and unvoiced sounds in speech signals [9] and the automatic diagnosis of tonic-clonic epileptic seizures [30]. The probability of finding a turning point is also linked with ordinal patterns. Indeed, estimating $\mathcal{T}$ is equivalent to calculate the relative frequency of four of the six possible motifs when embedding dimension $D = 3$ is considered (please see permutation indices 2, 3, 4 and 5 in Fig. 2a of Ref. [21]).

Tarnopolski introduced a model representation space by plotting the fraction of turning points of a time series versus its associated Abbe value. This $\mathcal{T}$ vs $\mathcal{A}$ diagram is able to discriminate fBm, fGn and dfGn (please see Fig. 5 of Ref. [23]). Moreover, an invertible relationship is found between $\mathcal{A}$ and the Hurst exponent H. This functional form has been then used for estimating the Hurst exponent of several real world data. Briefly, the Hurst exponent H is a scaling exponent that measures the long-range dependence in time series. Further details about H can be found in Refs. [31,32]. For illustrating the ability of the Tarnopolski plane to characterize long-range dependence in time series, we have analyzed the location of generic $1/f^{\alpha}$ noises in this bidimensional scheme. In Fig. 1 a), we depict the position of colored noises with α ranging from -1 to 3 in steps of size 0.1. Average and standard deviation (SD) (displayed as error bars) of estimated $\mathcal{A}$ and $\mathcal{T}$ values for one hundred independent realizations of length $n = 2^{14}$ for each α exponent have been plotted. The Fourier Filtering Method (FFM) has been implemented in Matlab for generating these long-range power-law correlated time series. In the FFM, the Fourier components of an uncorrelated sequence of Gaussian-distributed random numbers are filtered with a suitable power-law filter in order to introduce correlations among the variables. We address the reader to Refs. [33,34] for more details about this algorithm. Some examples of these artificial long-range correlated time series are shown in Fig. 1 b). It can be concluded that colored noises with α between -1 and 1 are more noisy and better discriminated by the Abbe value. Whereas, when the power-law exponent is between 1 and 3, the fraction of turning points is more appropriate for distinguishing between them. We have also confirmed that a very similar evolution in the Tarnopolski plane is followed by longer $1/f^{\alpha}$ artificial time series ($n = 100{,}000$). As expected, in this case, shorter SD error bars are obtained.

2.2. Multiscale analysis

It is widely recognized that time series arising from some representative variable of nonlinear complex systems have a multiscale nature, *i.e.* the observed dynamics is often strongly dependent on the resolution scale used to sample the signal. For illustrating this multiscale phenomenon, we consider the analysis of time series derived from nonlinear dynamics in a numerically controlled situation. More precisely, we estimate $\mathcal{A}$ and $\mathcal{T}$ from realizations of the x-variable of the Lorenz system:

$$\dot{x} = \sigma(y-x), \quad \dot{y} = x(\rho - z) - y, \quad \dot{z} = xy - \beta z\,. \tag{2}$$

Following the example included in Ref. [23], time series of length $n = 2^{14}$ data points were generated with initial conditions $(x_0, y_0, z_0) = (1, 5, 10)$, and standard parameters $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$ for which the system exhibits chaotic behavior. The time series were numerically integrated by using the Matlab's *ode45* function, that implements fourth and fifth order Runge–Kutta numerical integration algorithms, with an integration step equal to 0.001. Sampling periods δ_t ranging from 0.001 to 1 with a step equal to 0.001 are considered. We analyzed time series with $n = 2^{14}$ data points for each δ_t. The first 10^5 iterations were discarded to avoid possible transients. The evolution of the location in the Tarnopolski plane of these one thousand numerical realizations of length $n = 2^{14}$ with different temporal resolutions is depicted in Fig. 2. It is worth remarking here that a very similar behavior is obtained by analyzing longer numerical realizations ($n = 100{,}000$). On the one hand, for low values of δ_t, an artificial regular behavior is spuriously observed due to oversampling and both quantifiers are near zero. This oversampling generates redundancy in the information contained in the signals. On the other hand, for high values of the sampling period, the signal appears to be stochastic and fully uncorrelated. Essentially, relevant information about the nonlinear temporal correlations is lost due to undersampling, and the value of quantifiers are close to that expected for a white noise, *i.e.* $\mathcal{A} \approx 1$ and $\mathcal{T} \approx 2/3$. Through this toy example it is easily concluded that the estimated value for $\mathcal{A}$ and $\mathcal{T}$, and consequently the location in the bidimensional scheme, is strongly dependent on the temporal resolution. These findings imply the need to explicitly include the time scale notion in the implemented measure to reach a more proper characterization.

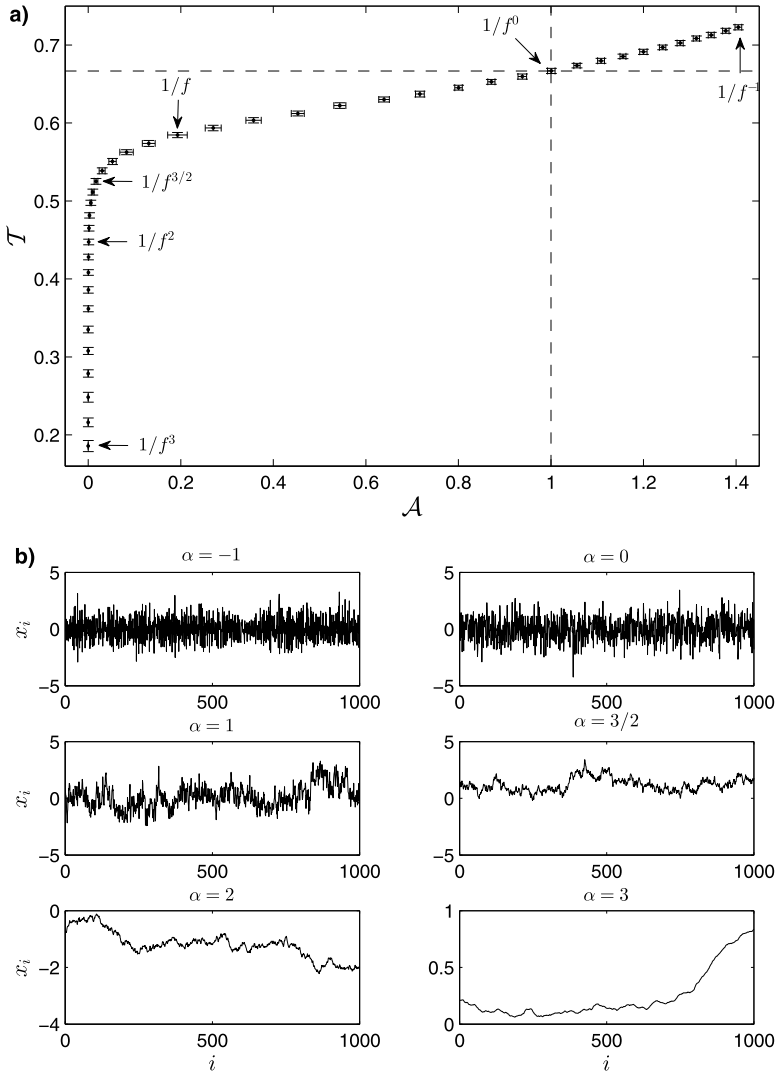

Fig. 1. a) Location in the Tarnopolski representation space of generic $1/f^{\alpha}$ noises with $\alpha \in \{-1, -0.9, ..., 3\}$. One hundred numerical realizations of length $n = 2^{14}$ data points for each α exponent were generated with FFM. Mean and standard deviation (displayed as error bars) of estimated values for both quantifiers, $\mathcal{A}$ and $\mathcal{T}$, over these one hundred simulations are shown. Positions of quantifiers move gradually from the right upper corner to the left bottom corner as α increases. Horizontal and vertical dashed lines indicate the theoretical value of quantifiers for white noise stochastic processes. b) Some examples of the numerically generated long-range power-law linearly correlated time series. Only 10^3 data points are plotted for a better visualization.

Multiscale entropy [24] and scale-dependent Lyapunov exponents [35] are two generalized quantifiers introduced precisely with the aim of characterizing the different signal behaviors on a wide range of scales simultaneously. Bearing in mind the inherent scale-dependent nature of the physiological systems we will analyze in the next section, a multiscale tool is proposed by constructing coarse-grained time series at multiple temporal scales and estimating the statistical quantifiers $\mathcal{A}$ and $\mathcal{T}$ for each of these transformed time series. Coarse-grained sequences are reconstructed by following the same procedure introduced in Ref. [24]. That is, the original record of length n is divided into non-overlapping segments of length τ, and the mean value is calculated for each segment generating smoothing sequences $\{y_j^{\tau}\}$ of length $\lfloor n/\tau \rfloor$,

$$y_j^{\tau} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \lfloor n/\tau \rfloor \tag{3}$$

with $\lfloor n/\tau \rfloor$ the largest integer not greater than n/τ. The analysis of the location of both quantifiers, $\mathcal{A}$ and $\mathcal{T}$, in the Tarnopolski plane as a function of the temporal scale factor τ allows detection of intrinsic complex structures across multiple temporal resolutions. Furthermore, the identification of optimal time scales for the classification of complex systems could be achieved by implementing this approach. For checking the usefulness of this multiscale recipe, we have analyzed an oversampled long numerical realization of the x-variable of the Lorenz system with the same parameters and initial conditions detailed previously. The chosen sampling period was $\delta_t = 0.001$ and the time series length $n = 2^{24}$. A multiscale analysis with temporal scale factors $1 \leq \tau \leq 1000$ has been performed. Evolution of the location of quantifiers for these temporal scales is depicted in Fig. 3 (blue curve). The evolution obtained for the original analysis for different sampling period developed previously is also plotted (black curve) for the sake of comparison. Overall, the results are qualitatively similar. The observed quantitative differences are attributed to the non-overlapping moving average filter (Eq. (3)) proposed for constructing the coarse-grained time series.

3. Real-world applications

Time series generated in a wide range of fields ranging from physiology to economy result from very complex dynamics and/or from coupled dynamics of many dimensional systems. Besides,

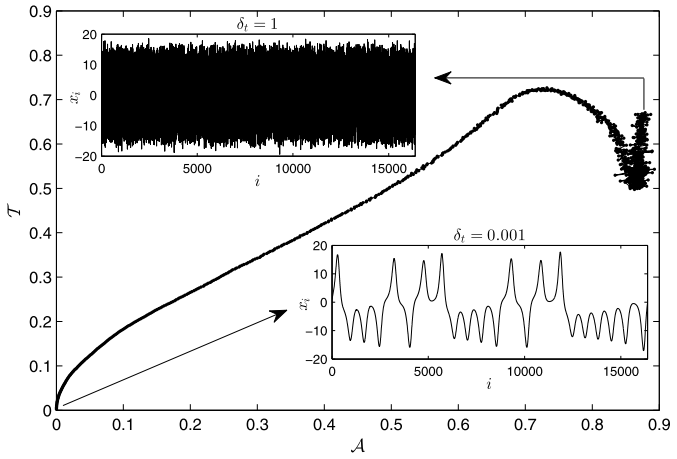

Fig. 2. Location in the Tarnopolski plane of one thousand numerical realizations of the x-variable of the Lorenz system ($\sigma = 10$, $\rho = 28$, $\beta = 8/3$, and $(x_0, y_0, z_0) = (1, 5, 10)$) with different sampling periods ($\delta_t \in \{0.001, 0.002, ..., 1\}$). Time series of length $n = 2^{14}$ data points were generated with initial conditions $(x_0, y_0, z_0) = (1, 5, 10)$. Positions of quantifiers move gradually from the left bottom corner to the right upper corner as δ_t increases. Simulations of the x-variable for the two extreme sampling periods, *i.e.* $\delta_t = 0.001$ and $\delta_t = 1$, are shown in the lower and upper insets, respectively. A similar evolution has been confirmed for longer numerical realizations ($n = 100{,}000$).

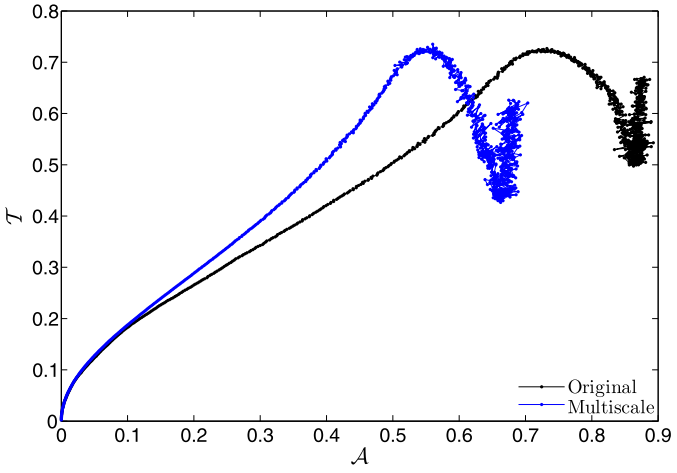

Fig. 3. Evolution of location in the Tarnopolski plane for an oversampled ($\delta_t = 0.001$) and long ($n = 2^{24}$) numerical realization of the x-variable of the Lorenz system ($\sigma = 10$, $\rho = 28$, $\beta = 8/3$, and $(x_0, y_0, z_0) = (1, 5, 10)$) by implementing a multiscale analysis with temporal scale factor τ ranging from 1 to 1000 (blue curve). Positions of quantifiers move gradually from the left bottom corner to the right upper corner as τ increases. The original evolution obtained for different sampling periods is also displayed (black curve) for easy comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these real signals are often contaminated by noise and other artifacts. Thus, the extraction of meaningful information from them is usually a challenging task. Taking this into account, next we will test the efficiency of the Tarnopolski representation space in real situations.

3.1. Efficiency of developed, emerging and frontier stock markets

It is well-known that a stock market is considered efficient whenever its prices follow a random walk. That is, the increments of the prices should be independent and obey a Gaussian distribution. However, it is also widely accepted that this is only an idealized first approximation, and deviations from this white noise model, violating either the independence or Gaussian assumptions, have been found in many empirical studies since the revolutionary paper of Benoit Mandelbrot [36]. In particular, correlated markets open the door to arbitrage opportunities because past prices can help to predict future prices. Indeed, it has been shown that emerging markets have greater correlations than developed ones, suggesting more predictability [37,38]. Moreover, the Hurst exponent has been widely proposed to quantify the stock market efficiency [39–43].

Given the link found between the Hurst exponent and the two statistical quantifiers that define the Tarnopolski plane, we propose to use this representation space to distinguish the stage of stock market development. With this aim in mind, we analyze the price returns of forty-eight stock market indices for different countries. Codes and names of these indices, collected from the Datastream platform (http://financial.thomsonreuters.com/en/products/tools-applications/trading-investment-tools/datastream-macroeconomic-analysis.html), are detailed in Table 1. Daily prices beginning on January 3, 2000 and ending on May 27, 2016 are considered (4,280 observations). Following the classification provided by the Morgan Stanley Capital Index (MSCI) methodology (https://www.msci.com), there are twenty developed, seventeen emerging and eleven frontier stock markets.

Locations in the $\mathcal{T}$ vs $\mathcal{A}$ diagram of the daily price returns of the forty-eight stock market indices are plotted in Fig. 4. Mean and standard deviation (displayed as error bars) of one thousand numerical independent realizations of fGn with Hurst exponents $H \in \{0.05, 0.10, ..., 0.95\}$ and the same length of the returns ($n = 4{,}279$) are also shown. These simulations were obtained by consecutive differences of fBm generated via the MATLAB function *wfbm*. On the one hand, it is observed that positions of indices associated with developed countries (blue circles) are, in average, closer to the ideal efficiency point, *i.e.* $\mathcal{A} \approx 1$ and $\mathcal{T} \approx 2/3$, that corresponds to white noise. Particularly, price returns of the American stock market index (Standard & Poor's 500) show an antipersistent behavior and, according to its location in the Tarnopolski plane, could be modeled as a fGn with $H = 0.45$. On the other hand, price return indices of stock market from emerging countries (green triangles) have lower estimated values for both quantifiers confirming the presence of persistency in their dynamics. There are two exceptions, namely Thailand ($\mathcal{A} \approx 0.987$, $\mathcal{T} \approx 0.661$) and Turkey ($\mathcal{A} \approx 1.006$, $\mathcal{T} \approx 0.662$), located within the region of maximum efficiency. Finally, price return indices of frontier countries (red squares) can be considered as the most inefficient since they are more distant with respect to the white noise location. The presence of strong long-range correlations in their dynamics is the main reason of this inefficient location. Argentina ($\mathcal{A} \approx 0.945$, $\mathcal{T} \approx 0.649$) and Croatia ($\mathcal{A} \approx 0.911$, $\mathcal{T} \approx 0.653$) are the two frontier countries whose price return indices appear to be better behaved (please see the two red squares nearest to the ideal efficiency location). It is also worth noting that the positions of all indices are below the curve described by the family of fGn. We conjecture that this behavior could be attributed to non-Gaussian distributions. It is then concluded that the fGn stochastic processes do not seem to be suitable for modeling the price returns of stock market indices of many emerging and frontier countries. This is especially clear in the case of Oman index ($\mathcal{A} \approx 0.819$, $\mathcal{T} \approx 0.555$). We confirm, through this financial application, that the Tarnopolski bidimensional scheme is a powerful tool for discriminating market dynamics since it is able to distinguish simultaneously different degrees of correlations and deviations from Gaussianity. The daily original resolution, that is $\tau = 1$, has been used in this analysis since improvements in the classification are not observed for larger temporal scales $2 \leq \tau \leq 10$. Indeed, for the larger scale factors ($\tau \geq 5$), locations of the different countries overlap in a small region making the distinction between the three groups almost impossible. Qualitatively similar findings have been confirmed for the same database but in a shorter data span (since January 1, 2010 to

Table 1
Codes and names of the stock market indices downloaded from Datastream. Developed, emerging and frontier stock markets (classified following the MSCI methodology) are denoted as D, E and F, respectively.

Country	Datastream code	Index	MSCI classification
1. Netherlands	AMSTEOE	AEX INDEX (AEX)	D
2. Jordan	AMMANFM	AMMAN SE FINANCIAL MARKET	F
3. Argentina	ARGMERV	ARGENTINA MERVAL	F
4. Greece	GRAGENL	ATHEX COMPOSITE	E
5. Austria	ATXINDX	ATX – AUSTRIAN TRADED INDEX	D
6. Thailand	BNGKSET	BANGKOK S.E.T.	E
7. Belgium	BGBEL20	BEL 20	D
8. Turkey	TRKISTB	BIST NATIONAL 100	E
9. Hungary	BUXINDX	BUDAPEST (BUX)	E
10. Chile	IGPAGEN	CHILE SANTIAGO SE GENERAL (IGPA)	E
11. Sri Lanka	SRALLSH	COLOMBO SE ALL SHARE	F
12. Croatia	CTCROBE	CROATIA CROBEX	F
13. Germany	DAXINDX	DAX 30 PERFORMANCE	D
14. Egypt	EGHFINC	EGYPT HERMES FINANCIAL	E
15. France	FRCAC40	FRANCE CAC 40	D
16. United Kingdom	FTSE100	FTSE 100	D
17. Malaysia	FBMKLCI	FTSE BURSA MALAYSIA KLCI	E
18. Italy	FTSEMIB	FTSE MIB INDEX	D
19. South Africa	JSEOVER	FTSE/JSE ALL SHARE	E
20. Hong Kong	HNGKNGI	HANG SENG	D
21. Hong Kong	HKHCHAF	HANG SENG CHINA AFFILIATED CORP	D
22. Hong Kong	HKHCHIE	HANG SENG CHINA ENTERPRISES	D
23. Spain	IBEX35I	IBEX 35	D
24. Indonesia	JAKCOMP	IDX COMPOSITE	E
25. Ireland	ISEQUIT	IRELAND SE OVERALL (ISEQ)	D
26. Israel	ISTA100	ISRAEL TA 100	D
27. Pakistan	PKSE100	KARACHI SE 100	F
28. Kenya	NSEINDX	KENYA NAIROBI SE (NSE20)	F
29. Korea	KORCOMP	KOREA SE COMPOSITE (KOSPI)	E
30. Lebanon	LBBLOMI	LEBANON BLOM	F
31. Germany	MDAXIDX	MDAX FRANKFURT	D
32. Mexico	MXIPC35	MEXICO IPC (BOLSA)	E
33. Oman	OMANMSM	AN MUSCAT SECURITIES MKT.	F
34. Denmark	COSEASH	OMX COPENHAGEN (OMXC)	D
35. Finland	HEXINDX	OMX HELSINKI (OMXH)	D
36. Sweden	SWSEALI	OMX STOCKHOLM (OMXS)	D
37. Estonia	ESTALSE	OMX TALLINN (OMXT)	F
38. Philippine	PSECOMP	PHILIPPINE SE I(PSEi)	E
39. Czech Republic	CZPXIDX	PRAGUE SE PX	E
40. Romania	RMBETRL	ROMANIA BET (L)	F
41. Russia	RSMICEX	RUSSIAN MICEX INDEX	E
42. USA	S&PCOMP	S&P 500 COMPOSITE	D
43. China	CHSASHR	SHANGHAI SE A SHARE	E
44. China	CHZBSHR	SHENZHEN SE B SHARE	E
45. Switzerland	SWISSMI	SWISS MARKET (SMI)	D
46. Taiwan	TAIWGHT	TAIWAN SE WEIGHED TAIEX	E
47. Japan	TOKYOSE	TOPIX	D
48. Tunisia	TUTUNIN	TUNISIA TUNINDEX	F

May 27, 2016). Comparison of the performance of the Tarnopolski representation space with other implemented tools for characterizing financial data efficiency, such as entropy-related methodologies [44–47] and the Efficiency Index [48–50], is beyond the purpose of the present work and will be studied in a future research.

3.2. Electroencephalograms from healthy and epileptic patients

Discrimination of brain electrical activity from different regions and from different physiological and pathological brain states is obviously a very relevant issue since this information could be potentially useful for medical diagnosis. Motivated by this fact, we analyze five different sets of electroencephalogram (EEG) time series for different groups and recording regions: surface (scalp) EEG recordings from five healthy volunteers in an awake state with eyes open (Set A) and closed (Set B), intracranial EEG recordings from five epilepsy patients during the seizure free interval from outside (Set C) and from within (Set D) the seizure generating area, and intracranial EEG recordings of epileptic seizures (Set E). These artifact-free records are available under www.meb.unibonn.de/epileptologie/science/physik/eegdata.html. The sampling rate of the data was 173.61 Hz. One hundred single channel EEG segments of 23.6 s of duration (4,097 data points) for each one of the five sets of data have been considered. These segments were selected from continuous multichannel EEG recordings after visual inspection for artifacts (*e.g.* muscle activity or eye movements). Additionally, only segments that satisfy a weak stationarity criterion were chosen. Further details about the recording technique of these EEG data can be found in the original paper by Andrzejak et al. [51].

Taking into account that the optimal time scale for discriminating these five sets of EEG data is *a priori* not known, a multiscale analysis has been implemented. That is, we have analyzed the location in the Tarnopolski diagram for coarse-grained time series with temporal scale factors $1 \leq \tau \leq 10$. Mean and standard deviation (displayed as error bars) of the one hundred $\mathcal{A}$ and $\mathcal{T}$ values computed for the five EEG sets have been plotted for the different resolution temporal scales. Fig. 5 shows the results obtained for four different temporal scales ($\tau \in \{1, 4, 7, 10\}$). It is visually

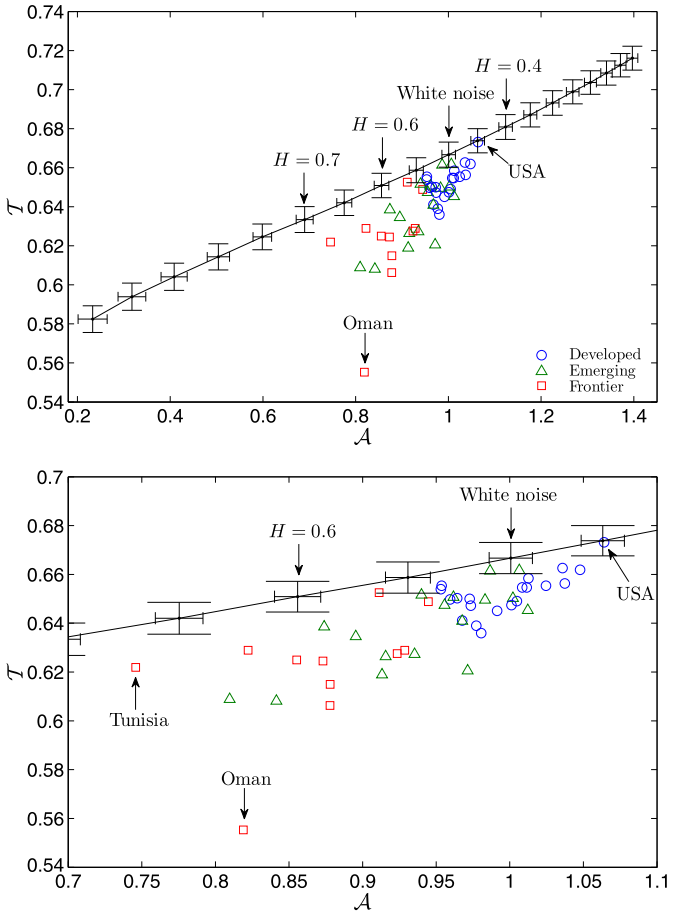

Fig. 4. Top: Position of the price returns of the forty-eight stock market indices in the Tarnopolski plane. Developed, emerging and frontier countries are identified by blue circles, green triangles and red squares, respectively. Location of fGn with Hurst exponents $H \in \{0.05, 0.10, ..., 0.95\}$ and the same length of the price returns ($n = 4{,}279$) has been also included. Being more precise, mean and standard deviation (displayed as error bars) of one thousand independent realizations are depicted in black color. A continuous black curve joining the mean values for fGn with different H is included for visual reference. Estimated values for both quantifiers decrease as the Hurst exponent of fGn increases. Bottom: Enlargement for a better view of locations associated with stock market indices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

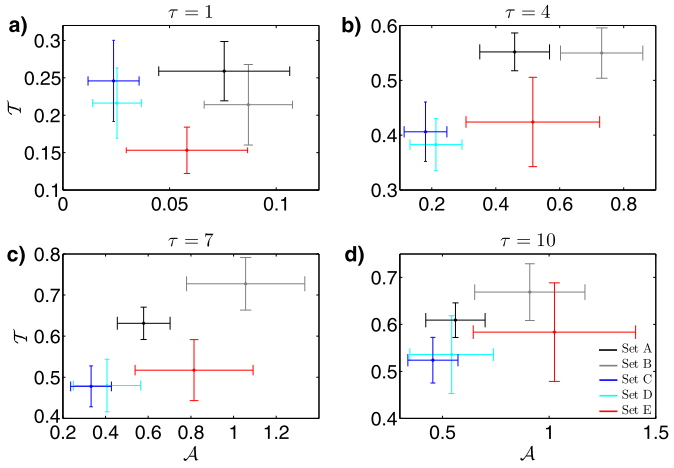

Fig. 5. Location of the five sets of EEG data in the Tarnopolski plane for different temporal scale factors: a) $\tau = 1$, b) $\tau = 4$, c) $\tau = 7$, and d) $\tau = 10$. Mean and standard deviation (displayed as error bars) of both quantifiers, $\mathcal{A}$ and $\mathcal{T}$, for the one hundred single channel EEG segments associated with each set are plotted. Groups are differentiated by color: Set A in black, Set B in gray, Set C in blue, Set D in cyan, and Set E in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

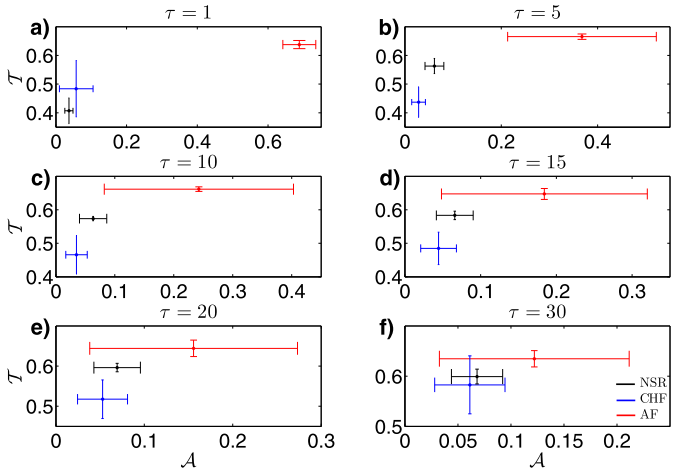

Fig. 6. Location of the three groups of HRV data in the Tarnopolski plane for different temporal scale factors: a) $\tau = 1$, b) $\tau = 5$, c) $\tau = 10$, d) $\tau = 15$, e) $\tau = 20$, and f) $\tau = 30$. Mean and standard deviation (displayed as error bars) of both quantifiers, $\mathcal{A}$ and $\mathcal{T}$, for the five BBI time series associated with each set are plotted. Groups are differentiated by color: NSR in black, CHF in blue, and AF in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

confirmed that an improved classification is reached for intermediate resolution scales. In such a case not only healthy (Sets A and B) and pathological groups (Sets C, D and E) are discriminated, but also differences between interictal (Sets C and D) and ictal activities (Set E) are achieved. Regrettably, interictal epileptiform activities from the epileptogenic zone (Set D) and those found at recording sites distant from the epileptogenic zone (Set C) cannot be separated. Last but not least, for these intermediate temporal scales, the two healthy groups are clearly distinguished. This differentiation between the conditions of eyes closed and eyes open has not been reached by previous implemented algorithms much more complicated, both computationally and conceptually [18,51–53]. We consider this a very interesting finding since it is well-known that alpha waves in a frequency range of 8–13 Hz are predominant in relaxed healthy subjects with eyes closed while broader frequency characteristics are obtained for open eyes [51].

3.3. Physiological and pathological beat-to-beat intervals

In this application, heart rate variability (HRV) of healthy and pathological subjects have been analyzed. HRV time series are derived from electrocardiogram signals by measuring consecutive beat-to-beat intervals (BBI). As cardiac diseases have effects on BBI dynamics, a reliable classification between physiological and pathological BBI time series could be useful for developing a new diagnostic tool [21]. Three different groups have been considered in our analysis. More precisely, the location in the Tarnopolski representation space for a total of fifteen BBI time series of which five were obtained from healthy persons in normal sinus rhythm (NSR), five from congestive heart failure (CHF) patients, and five from subjects suffering from atrial fibrillation (AF), have been tested. These records, freely available in PhysioNet (www.physionet.org/challenge/chaos/), are about 24 h long (roughly 100,000 intervals). Filtered records, *i.e.* with outliers removed, were used. We have again implemented a multiscale analysis by estimating both quantifiers, $\mathcal{A}$ and $\mathcal{T}$, for coarse-grained time series with $1 \leq \tau \leq 30$. In Fig. 6, we have plotted mean and standard deviation (displayed as error bars) of these quantifiers for the three HRV groups (NSR, CHF and AF)

for $\tau \in \{1, 5, 10, 15, 20, 30\}$. By comparing the location of these groups in the Tarnopolski plane for the different scale factor, an improved classification is visually concluded for intermediate temporal scales. Being more precise, the best separation of the given datasets of the three groups is achieved for $\tau = 5$. Higher values for both quantifiers are obtained for the AF group. That is, dynamics in HRV of AF subjects are more irregular and noisy than those related to NSR and CHF subjects, which is in agreement with previous findings [24,54–56]. Moreover, it should be also stressed, that for intermediate scale factors ($\tau \in \{5, 10, 15, 20\}$) cardiovascular dynamics from CHF patients have, in average, lower estimated values for $\mathcal{A}$ and $\mathcal{T}$ than those obtained from healthy subjects. This fact is reflecting a more regular dynamics for the CHF pathological state, which is consistent with results derived earlier through fractal analysis [10,54,55].

We demonstrate that the Tarnopolski diagram is able to efficiently separate the three different groups at particular time scales, evidencing once again the usefulness of the proposed multiscale analysis. We should emphasize here that this is a very simple qualitative analysis, and studies with larger datasets are required to confirm the reliability of the proposed multiscale bidimensional scheme as cardiac biomarker. Particularly, it is worth noting here that locations of the three HRV groups do not seem to follow the transition from more regular to more noisy dynamics as τ increases. This is a curious finding even more if we take into account that results obtained from EEG time series (please see Fig. 5) do follow the expected behavior. The reasons behind such unexpected behavior are not clear for us and will be investigated through further research by analyzing larger beat-to-beat intervals datasets.

4. Conclusions

The performance of a bidimensional scheme, obtained by plotting the fraction of turning points versus the Abbe value associated with a time series, has been tested for classification purposes. Several practical applications allow us to assess its reliability. A multiscale algorithm is also introduced for capturing relevant aspects of the complex dynamics at different temporal resolutions. We demonstrate that this new multiscale description is required to accurately distinguish between physiological time series in health and disease. We guess that the outstanding classification results shown by the Tarnopolski plane in the practical applications have their root in the excellent capability for separating colored noises (please see the high discriminative power of colored noises and fGn stochastic processes in Fig. 1 and Fig. 4, respectively). Obviously, theoretical arguments to support this heuristic observation are necessary and will be the main aim of a future study. We consider that results obtained are quite encouraging and justify further analysis in other research fields for testing the potentiality of this multiscale bidimensional approach as a discriminative tool.

Acknowledgements

We thank three anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions, which greatly helped to improve the manuscript. LZ and OAR acknowledge financial support from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. FO thanks support from Pontificia Universidad Católica de Valparaíso.

References

[1] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory 22 (1) (1976) 75–81, http://dx.doi.org/10.1109/TIT.1976.1055501.
[2] P. Grassberger, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (5) (1983) 346–349, http://dx.doi.org/10.1103/PhysRevLett.50.346.
[3] P. Grassberger, Do climatic attractors exist?, Nature 323 (1986) 609–612, http://dx.doi.org/10.1038/323609a0.
[4] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Physica D 16 (3) (1985) 285–317, http://dx.doi.org/10.1016/0167-2789(85)90011-9.
[5] M.T. Rosenstein, J.J. Collins, C.J. De Luca, A practical method for calculating largest Lyapunov exponents from small data sets, Physica D 65 (1–2) (1993) 117–134, http://dx.doi.org/10.1016/0167-2789(93)90009-P.
[6] P. Grassberger, I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signal, Phys. Rev. A 28 (4) (1983) 2591(R), http://dx.doi.org/10.1103/PhysRevA.28.2591.
[7] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. USA 88 (6) (1991) 2297–2301, http://dx.doi.org/10.1073/pnas.88.6.2297.
[8] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol., Heart Circ. Physiol. 278 (6) (2000) H2039–H2049.
[9] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17) (2002) 174102, http://dx.doi.org/10.1103/PhysRevLett.88.174102.
[10] C.-K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos 5 (1) (1995) 82–87, http://dx.doi.org/10.1063/1.166141.
[11] P.C. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, Multifractality in human heartbeat dynamics, Nature 399 (1999) 461–465, http://dx.doi.org/10.1038/20924.
[12] R. López-Ruiz, H.L. Mancini, X. Calbet, A statistical measure of complexity, Phys. Lett. A 209 (5–6) (1995) 321–326, http://dx.doi.org/10.1016/0375-9601(95)00867-5.
[13] O.A. Rosso, H.A. Larrondo, M.T. Martin, A. Plastino, M.A. Fuentes, Distinguishing noise from chaos, Phys. Rev. Lett. 99 (15) (2007) 154102, http://dx.doi.org/10.1103/PhysRevLett.99.154102.
[14] F. Olivares, A. Plastino, O.A. Rosso, Contrasting chaos with noise via local versus global information quantifiers, Phys. Lett. A 376 (19) (2012) 1577–1583, http://dx.doi.org/10.1016/j.physleta.2012.03.039.
[15] O.A. Rosso, H. Craig, P. Moscato, Shakespeare and other English Renaissance authors as characterized by Information Theory complexity quantifiers, Physica A 388 (6) (2009) 916–926, http://dx.doi.org/10.1016/j.physa.2008.11.018.
[16] L. Zunino, M. Zanin, B.M. Tabak, D.G. Pérez, O.A. Rosso, Complexity-entropy causality plane: a useful approach to quantify the stock market inefficiency, Physica A 389 (9) (2010) 1891–1901, http://dx.doi.org/10.1016/j.physa.2010.01.007.
[17] L. Zunino, A.F. Bariviera, M.B. Guercio, L.B. Martinez, O.A. Rosso, On the efficiency of sovereign bond markets, Physica 391 (18) (2012) 4342–4349, http://dx.doi.org/10.1016/j.physa.2012.04.009.
[18] J.B. Gao, J. Hu, W.W. Tung, Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering, PLoS ONE 6 (9) (2011) e24331, http://dx.doi.org/10.1371/journal.pone.0024331.
[19] H.V. Ribeiro, L. Zunino, R.S. Mendes, E.K. Lenzi, Complexity-entropy causality plane: a useful approach for distinguishing songs, Physica A 391 (7) (2012) 2421–2428, http://dx.doi.org/10.1016/j.physa.2011.12.009.
[20] N. Wessel, M. Riedl, J. Kurths, Is the normal heart rate “chaotic” due to respiration?, Chaos 19 (2) (2009) 028508, http://dx.doi.org/10.1063/1.3133128.
[21] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, N. Wessel, Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics, Comput. Biol. Med. 42 (3) (2012) 319–327, http://dx.doi.org/10.1016/j.compbiomed.2011.03.017.
[22] L. Zunino, H.V. Ribeiro, Discriminating image textures with the multiscale two-dimensional complexity-entropy causality plane, Chaos Solitons Fractals 91 (2016) 679–688, http://dx.doi.org/10.1016/j.chaos.2016.09.005.
[23] M. Tarnopolski, On the relationship between the Hurst exponent, the ratio of the mean square successive difference to the variance, and the number of turning points, Physica A 461 (2016) 662–673, http://dx.doi.org/10.1016/j.physa.2016.06.004.
[24] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (6) (2002) 068102, http://dx.doi.org/10.1103/PhysRevLett.89.068102.
[25] J. von Neumann, The mean square successive difference, Ann. Math. Stat. 12 (2) (1941) 153–162, http://dx.doi.org/10.1214/aoms/1177731746.
[26] J. von Neumann, Distribution of the ratio of the mean square successive difference to the variance, Ann. Math. Stat. 12 (4) (1941) 367–395, http://dx.doi.org/10.1214/aoms/1177731677.
[27] M.G. Kendall, Studies in the history of probability and statistics. XXVI: the work of Ernst Abbe, Biometrika 58 (2) (1971) 369–373, http://dx.doi.org/10.2307/2334525.
[28] N. Mowlavi, Searching transients in large-scale surveys – a method based on the Abbe value, Astron. Astrophys. 568 (2014) A78, http://dx.doi.org/10.1051/0004-6361/201322648.
[29] M.G. Kendall, A. Stuart, The Advanced Theory of Statistics, vol. 3, 2nd edition, Griffin, London, 1966.

[30] I. Conradsen, S. Beniczky, K. Hoppe, P. Wolf, H.B.D. Sorensen, Automated algorithm for generalized tonic-clonic epileptic seizure onset detection based on sEMG zero-crossing rate, IEEE Trans. Biomed. Eng. 59 (2) (2012) 579–585, http://dx.doi.org/10.1109/TBME.2011.2178094.

[31] H.E. Hurst, The long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng. 116 (1951) 770–799.

[32] B.B. Mandelbrot, J.W. van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. 10 (4) (1968) 422–437, http://dx.doi.org/10.1137/1010093.

[33] H.A. Makse, S. Havlin, M. Schwartz, H.E. Stanley, Method for generating long-range correlations for large systems, Phys. Rev. E 53 (5) (1996) 5445–5449, http://dx.doi.org/10.1103/PhysRevE.53.5445.

[34] M. Gómez-Extremera, P. Carpena, P.C. Ivanov, P.A. Bernaola-Galván, Magnitude and sign of long-range correlated time series: decomposition and surrogate signal generation, Phys. Rev. E 93 (4) (2016) 042201, http://dx.doi.org/10.1103/PhysRevE.93.042201.

[35] J.B. Gao, J. Hu, W.W. Tung, Y.H. Cao, Distinguishing chaos from noise by scale-dependent Lyapunov exponent, Phys. Rev. E 74 (6) (2006) 066204, http://dx.doi.org/10.1103/PhysRevE.74.066204.

[36] B. Mandelbrot, The variation of certain speculative prices, J. Bus. 36 (4) (1963) 394–419.

[37] M. Bęben, A. Orłowski, Correlation in financial time series: established versus emerging markets, Eur. Phys. J. B 20 (4) (2001) 527–530, http://dx.doi.org/10.1007/s100510170233.

[38] T. Di Matteo, T. Aste, M.M. Dacorogna, Scaling behaviors in differently developed markets, Physica A 324 (1–2) (2003) 183–188, http://dx.doi.org/10.1016/S0378-4371(02)01996-9.

[39] D.O. Cajueiro, B.M. Tabak, The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient, Physica A 336 (3–4) (2004) 521–537, http://dx.doi.org/10.1016/j.physa.2003.12.031.

[40] D.O. Cajueiro, B.M. Tabak, Ranking efficiency for emerging markets, Chaos Solitons Fractals 22 (2) (2004) 349–352, http://dx.doi.org/10.1016/j.chaos.2004.02.005.

[41] D.O. Cajueiro, B.M. Tabak, Ranking efficiency for emerging markets II, Chaos Solitons Fractals 23 (2) (2005) 671–675, http://dx.doi.org/10.1016/j.chaos.2004.05.009.

[42] L. Zunino, B.M. Tabak, D.G. Pérez, M. Garavaglia, O.A. Rosso, Inefficiency in Latin-American market indices, Eur. Phys. J. B 60 (1) (2007) 111–121, http://dx.doi.org/10.1140/epjb/e2007-00316-y.

[43] C. Eom, S. Choi, G. Oh, W.-S. Jung, Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets, Physica A 387 (18) (2008) 4630–4636, http://dx.doi.org/10.1016/j.physa.2008.03.035.

[44] E. Martina, E. Rodriguez, R. Escarela-Perez, J. Alvarez-Ramirez, Multiscale entropy analysis of crude oil price dynamics, Energy Econ. 33 (5) (2011) 936–947, http://dx.doi.org/10.1016/j.eneco.2011.03.012.

[45] J. Alvarez-Ramirez, E. Rodriguez, J. Alvarez, A multiscale entropy approach for market efficiency, Int. Rev. Financ. Anal. 21 (2012) 64–69, http://dx.doi.org/10.1016/j.irfa.2011.12.001.

[46] A. Ortiz-Cruz, E. Rodriguez, C. Ibarra-Valdez, J. Alvarez-Ramirez, Efficiency of crude oil markets: evidences from informational entropy analysis, Energy Policy 41 (2012) 365–373, http://dx.doi.org/10.1016/j.enpol.2011.10.057.

[47] L. Zunino, A.F. Bariviera, M.B. Guercio, L.B. Martinez, O.A. Rosso, Monitoring the informational efficiency of European corporate bond markets with dynamical permutation min-entropy, Physica A 456 (2016) 1–9, http://dx.doi.org/10.1016/j.physa.2016.03.007.

[48] L. Kristoufek, M. Vosvrda, Measuring capital market efficiency: global and local correlations structure, Physica A 392 (1) (2013) 184–193, http://dx.doi.org/10.1016/j.physa.2012.08.003.

[49] L. Kristoufek, M. Vosvrda, Measuring capital market efficiency: long-term memory, fractal dimension and approximate entropy, Eur. Phys. J. B 87 (7) (2014) 162, http://dx.doi.org/10.1140/epjb/e2014-50113-6.

[50] L. Kristoufek, M. Vosvrda, Commodity futures and market efficiency, Energy Econ. 42 (2014) 50–57, http://dx.doi.org/10.1016/j.eneco.2013.12.001.

[51] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001) 061907, http://dx.doi.org/10.1103/PhysRevE.64.061907.

[52] H. Adeli, S. Ghosh-Dastidar, N. Dadmehr, A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy, IEEE Trans. Biomed. Eng. 54 (2) (2007) 205–211, http://dx.doi.org/10.1109/TBME.2006.886855.

[53] L. Zunino, F. Olivares, O.A. Rosso, Permutation min-entropy: an improved quantifier for unveiling subtle temporal correlations, Europhys. Lett. 109 (1) (2015) 10005, http://dx.doi.org/10.1209/0295-5075/109/10005.

[54] A.L. Goldberger, L.A.N. Amaral, J.M. Hausdorff, P.C. Ivanov, C.-K. Peng, H.E. Stanley, Fractal dynamics in physiology: alterations with disease and aging, Proc. Natl. Acad. Sci. USA 99 (Suppl. 1) (2002) 2466–2472, http://dx.doi.org/10.1073/pnas.012579499.

[55] J. Alvarez-Ramirez, E. Rodriguez, J.C. Echeverría, Delays in the human heartbeat dynamics, Chaos 19 (2) (2009) 028502, http://dx.doi.org/10.1063/1.3152005.

[56] J. Hu, J. Gao, W.W. Tung, Characterizing heart rate variability by scale-dependent Lyapunov exponent, Chaos 19 (2) (2009) 028506, http://dx.doi.org/10.1063/1.3152007.