Physics Letters A ••• (••••) •••–•••

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 xxxx
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.

© 2017 Elsevier B.V. All rights reserved.

UNCORRECTED PROOF

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/© 2017 Elsevier B.V. All rights reserved.

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,

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

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:

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

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.

Please cite this article in press as: L. Zunino et al., A simple and fast representation space for classifying complex time series, Phys. Lett. A (2017), http://dx.doi.org/10.1016/j.physleta.2017.01.047

UNCORRECTED PROOF

Download English Version:

https://daneshyari.com/en/article/5496849

Download Persian Version:

https://daneshyari.com/article/5496849

Daneshyari.com