Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach

L. Zunino,[1,2,*] M. C. Soriano,[3,†] and O. A. Rosso[4,5,‡]

[1]*Centro de Investigaciones Ópticas (CONICET La Plata–CIC), C.C. 3, 1897 Gonnet, Argentina*

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

[3]*Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC) CSIC-UIB, Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain*

[4]*LaCCAN/CPMAT–Instituto de Computação, Universidade Federal de Alagoas, BR 104 Norte km 97, 57072-970 Maceió, Alagoas, Brazil*

[5]*Laboratorio de Sistemas Complejos, Facultad de Ingeniería, Universidad de Buenos Aires. 1063 Av. Paseo Colón 840, Ciudad Autónoma de Buenos Aires, Argentina*

(Received 3 May 2012; revised manuscript received 16 August 2012; published 17 October 2012)

In this paper we introduce a multiscale symbolic information-theory approach for discriminating nonlinear deterministic and stochastic dynamics from time series associated with complex systems. More precisely, we show that the *multiscale complexity-entropy causality plane* is a useful representation space to identify the range of scales at which deterministic or noisy behaviors dominate the system's dynamics. Numerical simulations obtained from the well-known and widely used Mackey-Glass oscillator operating in a high-dimensional chaotic regime were used as test beds. The effect of an increased amount of observational white noise was carefully examined. The results obtained were contrasted with those derived from correlated stochastic processes and continuous stochastic limit cycles. Finally, several experimental and natural time series were analyzed in order to show the applicability of this scale-dependent symbolic approach in practical situations.

DOI: 10.1103/PhysRevE.86.046210 PACS number(s): 05.45.Tp, 89.70.Cf, 05.40.Ca

I. INTRODUCTION

It is clear that recorded signals from experimental measurements give us very useful information to reveal the deterministic or stochastic character of the system under analysis. However, the task to distinguish between regular, chaotic, and stochastic dynamics from complex time series can be a critical and subtle issue [1]. Particularly, deterministic chaotic time series arising from high-dimensional nonlinear systems share several properties with those generated by stochastic processes, e.g., a wide-band power spectrum and a long-term unpredictable behavior. Moreover, experimental chaotic signals are unavoidably contaminated by noise, making the discrimination task even more challenging [2]. It is also well known that the observed dynamics can be strongly dependent on the resolution scale used to sample the signal. The notion of chaotic or stochastic behavior of a system on a certain range of scales was introduced by Cencini *et al*. [3]. The time scales where these two behaviors dominate the system's dynamics are usually different since the stochastic forcing kicks the dynamics to larger scales and this effect is bounded to scales smaller than those showing chaotic motion [4]. The identification of these temporal scales is fundamental to develop suitable models for simulation and forecasting purposes. Consequently, it is essential to explicitly include the time scale notion in the measure devised for a more proper characterization. We can mention the multiscale entropy [5] and the scale-dependent Lyapunov exponents [4] as two alternatives introduced with this aim in mind. The importance of addressing the multiscale feature is straightforwardly concluded from different applications associated with these previously introduced measures [6].

In this paper we present a reliable and computationally fast multiscale symbolic information-theory approach to identify the time scales where stochastic and nonlinear deterministic components govern the system's dynamics. It can be easily applied to the time series obtained from some representative variable of the complex system under analysis. As will be shown below, short noisy scalar time series can be efficiently characterized on a wide range of scales, discriminating where the stochastic and chaotic dynamics are intrinsically present. Thus, this methodology can help shed new light on the characterization of noisy chaotic time series.

II. MULTISCALE COMPLEXITY-ENTROPY CAUSALITY PLANE

An information-theory quantifier can be defined as a measure that is able to characterize some property of the probability distribution associated with an observable or measurable quantity. Entropy, regarded as a measure of uncertainty, is the most paradigmatic example. For a discrete variable x, which can take a finite number M of possible values $x_i \in \{x_1, \ldots, x_M\}$ with corresponding probabilities $p_i \in P = \{p_1, \ldots, p_M\}$, Shannon's logarithmic information measure is defined by $S[P] = -\sum_{i=1}^{M} p_i \ln p_i$ [7]. This functional is equal to zero when we are able to predict with full certainty which of the possible outcomes i whose probabilities are given by p_i will actually take place. Our knowledge of the underlying process, described by the probability distribution, is maximal in this instance. In contrast, this knowledge is minimal, commonly, for a uniform distribution. It is well known, however, that the degree of structure present in a process is not quantified by randomness measures and,

*lucianoz@ciop.unlp.edu.ar
†miguel@ifisc.uib-csic.es
‡oarosso@fibertel.com.ar

consequently, measures of statistical or structural complexity are necessary for a better understanding of chaotic time series [8]. The opposite extremes of perfect order and maximal randomness (a periodic sequence and a fair coin toss, for example) are very simple to describe because they do not have any structure. The complexity should be zero in these cases. At a given distance from these extremes, a wide range of possible degrees of physical structure exists. The complexity measure allows one to quantify this array of behavior [9]. In this work we have considered the effective statistical complexity measure (SCM) introduced by Lamberti *et al*. [10] since it is able to detect essential details of the dynamics, and it also discriminates different degrees of periodicity, i.e., cycles of period 2,4,8,16,32, ... associated with the period-doubling bifurcation route to chaos are clearly distinguished. This statistical complexity measure is defined, following the seminal notion advanced by López-Ruiz *et al*. [11], through the product

$$\mathcal{C}_{JS}[P] = \mathcal{Q}_J[P, P_e]\,\mathcal{H}_S[P] \tag{1}$$

of the normalized Shannon entropy

$$\mathcal{H}_S[P] = S[P]/S_{\max} \tag{2}$$

with $S_{\max} = S[P_e] = \ln M$, $(0 \leqslant \mathcal{H}_S \leqslant 1)$ and $P_e = \{1/M, \ldots, 1/M\}$ the uniform distribution, and the disequilibrium $\mathcal{Q}_J$ defined in terms of the Jensen-Shannon divergence. That is, $\mathcal{Q}_J[P, P_e] = \mathcal{Q}_0 \mathcal{J}[P, P_e]$ with $\mathcal{J}[P, P_e] = \{S[(P + P_e)/2] - S[P]/2 - S[P_e]/2\}$ the above-mentioned Jensen-Shannon divergence and $\mathcal{Q}_0$ a normalization constant, equal to the inverse of the maximum possible value of $\mathcal{J}[P, P_e]$. This value is obtained when one of the components of P, say p_m, is equal to one and the remaining p_i are equal to zero. The Jensen-Shannon divergence, that quantifies the difference between two (or more) probability distributions, is especially useful to compare the symbolic composition between different sequences [12]. Note that the above introduced SCM depends on two different probability distributions, the one associated to the system under analysis, P, and the uniform distribution, P_e. Furthermore, it was shown that for a given value of $\mathcal{H}_S$, the range of possible $\mathcal{C}_{JS}$ values varies between a minimum $\mathcal{C}_{JS}^{\min}$ and a maximum $\mathcal{C}_{JS}^{\max}$, restricting the possible values of the SCM in a given complexity-entropy plane [13]. Thus, it is clear that important additional information related to the correlational structure between the components of the physical system is provided by evaluating the statistical complexity measure.

In order to calculate the two information-theory-derived quantifiers mentioned previously, a probability distribution should be estimated from the time series of the system. The Bandt and Pompe permutation methodology was employed in our analysis due to its simplicity and effectiveness [14]. This symbolic technique, based on the ordinal relation between the amplitude of neighboring values, arises naturally from the time series, and allows one to avoid amplitude threshold dependencies that affect more conventional methods based on range partitioning. It is clear that, applying this prescription for symbolizing time series, some details of the original amplitude information and variability are lost. However, a meaningful reduction of the complex systems to their basic inherent structure is provided. Furthermore, the ordinal pattern distribution is invariant with respect to nonlinear monotonous transformations [14]. Thus, nonlinear drifts or scalings artificially introduced by a measurement device do not modify the quantifiers' estimations, a property highly desired for the analysis of experimental data. Technically speaking, the ordinal pattern probability distribution is obtained once we fix the embedding dimension D and the embedding delay time τ. The former parameter, D, refers to the number of symbols that form the ordinal pattern. Its choice depends on the length N of the time series in such a way that the condition $N \gg D!$ must be satisfied in order to obtain reliable statistics [15]. It is worth remarking that there are $D!$ possible permutations, and accessible states, for a D-dimensional vector. For practical purposes, Bandt and Pompe recommend $D = 3, \ldots, 7$ in their cornerstone paper. The embedding delay τ is the time separation between symbols, and it physically corresponds to multiples of the sampling time of the signal under analysis. Consequently, different time scales are considered by changing the embedding delays of the symbolic reconstruction [16]. (Please see Refs. [17,18] for further details about the Bandt and Pompe permutation methodology.) A related approach, based on computing the number of forbidden ordinal patterns present in the time series, has been successfully used to find evidence of determinism in noisy time series [19]. In this work the normalized Shannon entropy $\mathcal{H}_S$ [Eq. (2)], and the SCM $\mathcal{C}_{JS}$ [Eq. (1)], are evaluated using the permutation probability distribution. Defined in this way, these quantifiers are usually known as *permutation entropy* and *permutation statistical complexity* [20]. They characterize the diversity and correlational structure, respectively, of the orderings present in the complex time series. We have previously shown that these quantifiers are able to efficiently identify the delay phenomena from chaotic time series [17]: $\mathcal{H}_S$ is minimized and $\mathcal{C}_{JS}$ is maximized when the embedding delay τ of the symbolic reconstruction matches the intrinsic time delay τ_S of the system.

The *complexity-entropy causality plane* (CECP) is defined as the two-dimensional (2D) diagram obtained by plotting permutation statistical complexity (vertical axis) versus permutation entropy (horizontal axis) for a given system [21]. This representation space follows directly from the original complexity-entropy diagram approach introduced by Crutchfield and Young [22]. The term causality remembers the fact that temporal correlations between successive samples are taken into account through the Bandt and Pompe recipe used to estimate both information-theory quantifiers. This diagnostic tool was shown to be particularly efficient to distinguish between the deterministic chaotic and stochastic nature of a time series since the permutation quantifiers have distinctive behaviors for different types of motion. According to the findings obtained by Rosso *et al*. [21], chaotic maps have intermediate $\mathcal{H}_S$ values, while $\mathcal{C}_{JS}$ reaches larger values, very close to those of the limit. For regular processes, both quantifiers have small values, close to 0. Finally, totally uncorrelated stochastic processes are located in the planar location associated with $\mathcal{H}_S$ and $\mathcal{C}_{JS}$ near 1 and 0, respectively. It has also been found that $1/f^{\alpha}$ correlated stochastic processes with $1 < \alpha < 3$ are characterized by intermediate permutation entropy and intermediate statistical complexity values [20].

As pointed out by Cencini *et al*. [3] the underlying chaotic or stochastic nature of a system may depend on the resolution

of the data record, and, as a consequence, it is more suitable to define the concept of deterministic or stochastic behavior on a certain range of scales. Hence, a scale-dependent scheme is the most natural for dealing with complex multiscaled data [4]. Basically, we propose to generalize the estimation of both symbolic quantifiers, permutation entropy and statistical complexity, to different embedding delays. Therefore, the scale is explicitly included in these measures and their dependence on resolution can be thoroughly explored. The *multiscale complexity-entropy causality plane* refers to the parametric curve described by the permutation quantifiers estimated from a time series with the embedding delay as parameter and a chosen embedding dimension. According to the location in this representation space, the regular, chaotic or stochastic character for different scale ranges can be concluded. This approach offers a more versatile tool for discriminating dynamics at different temporal or spatial scales. In fact, the importance of selecting an appropriate embedding delay in the estimation of the permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) has been recently confirmed for other purposes, such as identifying periodicities in natural time series [16], estimating intrinsic time scales of delayed systems [17,18], quantifying the degree of unpredictability of the high-dimensional chaotic fluctuations of a semiconductor laser subject to optical feedback [23], and classifying cardiac biosignals [24].

III. NUMERICAL RESULTS

We have initially analyzed the behavior of the proposed scale-dependent approach in a numerically controlled situation. More precisely, we consider the well-known Mackey-Glass equation, a paradigmatic time-delay system [25], given by

$$\frac{dx}{dt} = -x + \frac{ax(t-\tau_S)}{1+x^c(t-\tau_S)} \tag{3}$$

with t being a dimensionless time, τ_S the time-delay feedback, a the feedback strength, and c the degree of nonlinearity. In particular, we choose the typical values $a = 2$, $c = 10$, and $\tau_S = 60$ for which the system operates in a high-dimensional chaotic regime. The time series were numerically integrated by using the Heun method (also called the modified Euler method) with an integration step $\Delta t = 0.01$ and sampling step $\delta t = 0.02$ time units/sample. We analyzed time series with $N = 10^5$ data points (the total integration time was 2×10^3 time units). Gaussian white noise was added to the original noise-free chaotic simulations in order to test the effect that an uncorrelated stochastic component has on the results. Different noise levels (NL), defined by the standard deviation of the noise divided by the standard deviation of the original signal, were considered. In Fig. 1 we plot the normalized permutation entropy $\mathcal{H}_S$ and the permutation SCM $\mathcal{C}_{JS}$ as a function of the

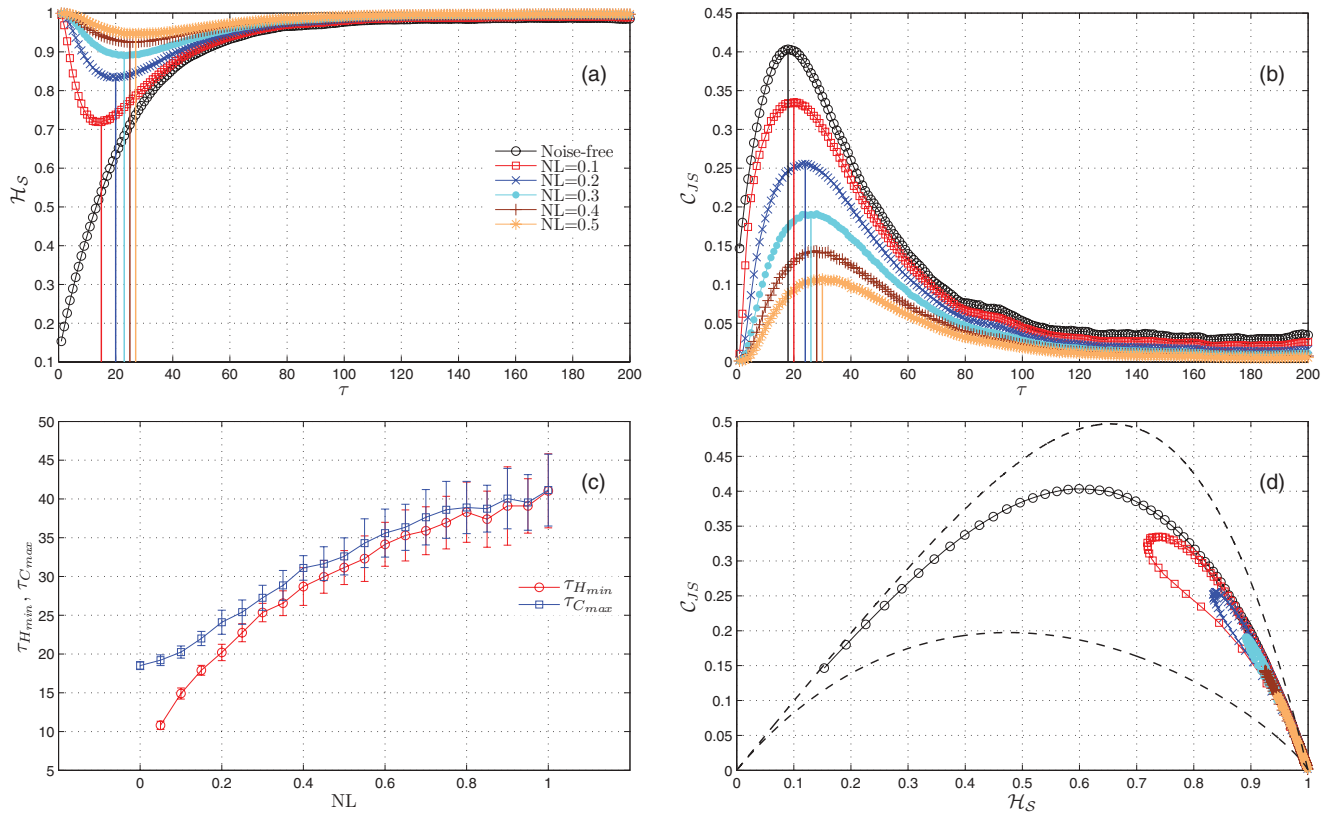

FIG. 1. (Color online) (a) Permutation entropy $\mathcal{H}_S$ and (b) permutation statistical complexity $\mathcal{C}_{JS}$ as a function of the embedding delay τ with embedding dimension $D = 6$ for the Mackey-Glass oscillator operating in a chaotic regime. The effect of different levels of observational noise can be contrasted with the results obtained for the original noise-free chaotic dynamics. Mean and standard deviation for $\tau_{H_{\min}}$ and $\tau_{C_{\max}}$ values corresponding to 20 independent numerical realizations of length $N = 10^5$ data points are depicted in (c) for different noise levels (NL $= 0.05, 0.1, \ldots, 0.95, 1.0$). Curves described by the permutation quantifiers in the CECP are illustrated in (d). The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

embedding delay τ for embedding dimension $D = 6$. Similar results were also obtained for other embedding dimensions ($D = 4$, $D = 5$, and $D = 7$). Our main intention is to compare the results obtained in the noise instance with respect to those obtained for the pure chaotic dynamics. In the absence of observational noise, Figs. 1(a) and 1(b) show, respectively, that $\mathcal{H}_S$ has a monotonically increasing behavior for increasing embedding delays, while $\mathcal{C}_{JS}$ reaches a maximum at an intermediate embedding delay $\tau_{C_{\max}}$. The value associated with this temporal scale, which appears to be slightly dependent on the noise environment [please see vertical lines in Fig. 1(b)], represents the minimally required sampling rate to capture all the information related to the nonlinear correlations of the underlying chaotic dynamics [18,26]. In the presence of noise, the entropy quantifier reaches a minimum value at an embedding delay smaller than $\tau_{C_{\max}}$. This embedding delay, henceforth denoted as $\tau_{H_{\min}}$, can be interpreted as the transitional time scale between the stochastic and the chaotic dynamics. Note that $\tau_{H_{\min}}$ increases with the noise level, as is evident following vertical lines added to Fig. 1(a). What is more, $\tau_{H_{\min}}$ tends to $\tau_{C_{\max}}$ for highly noisy data [Fig. 1(c)]. Since this behavior is directly related to the amount of observational noise added, it might be possible to quantify the noise level with this procedure. A more detailed investigation of this issue is beyond the scope of the present paper and will be performed elsewhere.

From Fig. 1(d) it can be concluded that, under the presence of uncorrelated observational noise, the permutation quantifiers describe a clear clockwise loop in the CECP as the embedding delay increases. This result can be interpreted taking into consideration that (i) for small embedding delays, the uncorrelated stochastic dynamics is detected and, consequently, the quantifiers are located in the bottom right corner with $\mathcal{H}_S$ and $\mathcal{C}_{JS}$ near 1 and 0, respectively; (ii) for intermediate embedding delays, the nonlinear structures related to the intrinsic chaotic dynamics are optimally discriminated with $\mathcal{C}_{JS}$ reaching a maximum value for $\tau_{C_{\max}}$; and (iii) for larger embedding delays the original noisy chaotic dynamics is undersampled, any information about the nonlinear determinism is progressively lost, and the data appear to be stochastic rather than chaotic. Contrarily, in the pure chaotic case, the positions described by these symbolic quantifiers move gradually from the left-hand corner to the right-hand corner of the CECP. The difference is principally due to the fact that, now, the underlying deterministic behavior is oversampled for very low values of the embedding delay, and an artificial regular behavior is spuriously concluded for these small scales. Colinearities are introduced when a high sampling rate is used to capture a low frequency dynamics, leading to a decrease in the entropy value [27]. Summarizing, the curve described by the permutation quantifiers as a function of the embedding delay allows us to characterize the relative importance of the information redundancy, determinism, and stochasticity present in the underlying complex dynamics. We have also confirmed a quantitatively similar behavior for the Lorenz and Rössler systems by numerical simulations in the chaotical regime. Our results clearly demonstrate that the choice of an inappropriate sampling time can hinder the evidence on nonlinear determinism from a chaotic dynamics.

With the aim to test the approach in a purely stochastic scenario, we have studied numerical simulations of correlated stochastic processes. Particularly, we have considered the fractional Brownian motion (fBm) process because this is a paradigmatic model for $1/f^{\alpha}$ processes. Moreover, the ubiquity of time series with long-range correlation in many areas of science and engineering is widely accepted [28]. Twenty independent realizations of length $N = 10^5$ data points for fBm with Hurst exponents $H = 0.3$ (antipersistent correlation), $H = 0.5$ (ordinary Brownian motion), and $H = 0.7$ (persistent correlation) were simulated, each series starting at a different initial condition, by employing the method of Wood and Chan, which is both exact and fast [29]. Small amounts of Gaussian white noise (NL $= 0.01, 0.05$, and 0.1) were added to investigate the effect of an observational additive noise. From Fig. 2 it can be concluded that the curves described by the permutation quantifiers in the CECP for these stochastic instances are different than those obtained for the chaotic ones. In the noise-free case and for the different Hurst exponent, it is found that both permutation quantifiers have constant values independently of the embedding delay τ. This constant behavior is expected since fBm are self-similar processes, i.e., invariant in distribution under suitable scaling of time. So, the relative frequencies of the ordinal patterns do not depend on the value of the embedding delay, and quantifiers derived from the permutation probability distribution, such as permutation entropy and permutation statistical complexity, are independent of the time scale considered [30]. When observational noise is added, permutation quantifier estimations are especially affected for smaller time scales: $\mathcal{H}_S \to 1$ and $\mathcal{C}_{JS} \to 0$ when $\tau \to 1$ independently of the Hurst exponent value. The effect is much more pronounced for higher noise levels and for higher H. It should be stressed that a very small amount of noise is enough to destroy the self-similarity feature of the original fBm.

Finally, we have tested the behavior of the introduced symbolic approach in the case of stochastic oscillations, where other methods, such as the *noise titration technique* [31], are shown to fail by giving a false positive detection of chaos [32]. Following the analysis performed by Gao *et al.* [4,32], we have studied the stochastically driven van der Pol's oscillator given by

$$dx/dt = y + D_1\eta_1(t), \quad dy/dt = -(x^2 - 1)y - x + D_2\eta_2(t), \tag{4}$$

where $\langle \eta_i(t) \rangle = 0$, $\langle \eta_i(t)\eta_i(t') \rangle = \delta_{ij}\delta(t - t')$, $i, j = 1,2$, and the parameters D_i, $i = 1,2$ characterize the strength of noise. Time series of length $N = 10^5$ data points, integration step $\Delta t = 0.001$, and sampling step $\delta t = 0.02$ were numerically generated by using the Euler-Maruyama method. The x coordinate of the system was chosen for the study. Results obtained for both permutation quantifiers as a function of the embedding delay τ with embedding dimension $D = 6$ are shown in Fig. 3. Curves described by the estimated quantifiers in the CECP are displayed in Fig. 4. Specifically, we have compared the results obtained in the noise-free case [Fig. 3(a)] with those derived for strengths of noise $D_1 = D_2 = 0.01$ [Fig. 3(b)], $D_1 = D_2 = 0.03$ [Fig. 3(c)], and $D_1 = D_2 = 0.05$ [Fig. 3(d)]. On the one hand, from Fig. 3(a), it can be concluded that both quantifiers have small estimated values, close to 0, when the embedding delay τ matches the period and multiples of the period of the van der Pol's oscillator. On the other

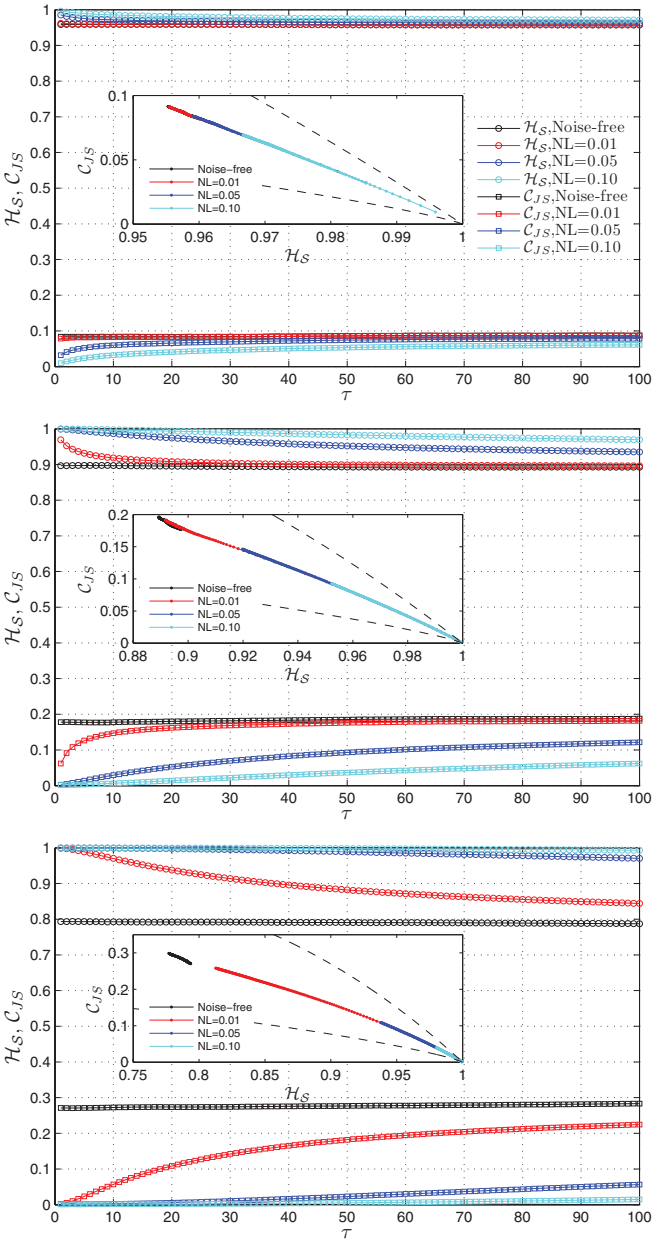

FIG. 2. (Color online) Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delay τ with embedding dimension $D = 6$ for fBm with Hurst exponent $H = 0.3$ (upper plot), $H = 0.5$ (central plot), and $H = 0.7$ (lower plot). Mean values corresponding to 20 independent numerical realizations of length $N = 10^5$ data points are depicted for the noise-free case and for different noise levels (NL = 0.01, NL = 0.05 and NL = 0.10). Curves described by the permutation quantifiers in the CECP are displayed in the insets. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

hand, in the stochastic instances, Figs. 3(b)–3(d), $\mathcal{H}_S$ jumps abruptly to values close to 0.9 for these particular embedding delay values. Therefore, the presence of dynamical noise can be clearly detected by the permutation entropy quantifier when the embedding delay matches the oscillation period and its multiples. As a consequence, curves described by the permutation quantifiers in the CECP, depicted in Fig. 4,

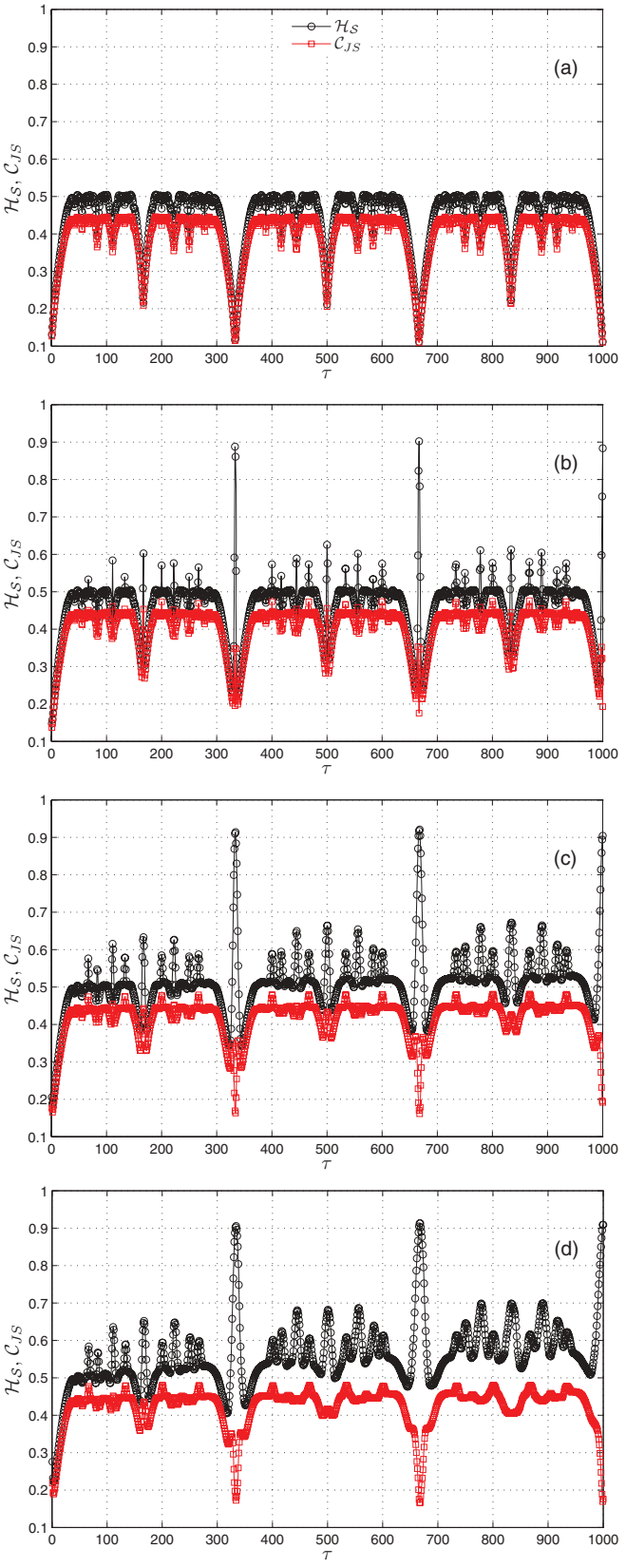

FIG. 3. (Color online) Permutation quantifiers as a function of the embedding delay τ with embedding dimension $D = 6$ for the van der Pol's oscillator. (a) Noise-free case, (b) $D_1 = D_2 = 0.01$, (c) $D_1 = D_2 = 0.03$, and (d) $D_1 = D_2 = 0.05$.

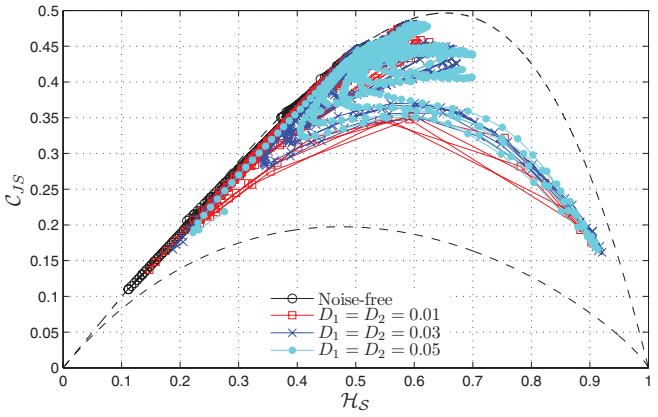

FIG. 4. (Color online) Curves described by the permutation quantifiers in the CECP for the van der Pol's oscillator ($D = 6$). The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

move from the regular behavior associated with the original oscillatory dynamics to the stochastic and unpredictable one detected for these more noise-sensitive temporal scales.

IV. EXPERIMENTAL RESULTS

Several experimental and natural records were tested in order to illustrate the performance of our method in real contexts. In particular, we have analyzed the behavior described by the permutation quantifiers in the multiscale complexity-entropy causality plane for time series originating from a chaotic laser, three different geophysical processes (river flow dynamics, Madden-Julien oscillation, and North Atlantic oscillation), the price evolution of two different commodities (crude oil and gold), and the human postural dynamics.

A. Chaotic laser data

We have analyzed the chaotic intensity pulsations recorded from a single-mode far-infrared NH_3 laser by employing a LeCroy oscilloscope. This experimental time series was used for the Santa Fe Time Series Competition (Series A from Ref. [33]). Further details of the recording procedure of this data set can be found in Ref. [34]. The longer data set ($N = 10^4$ data points) was considered. Figure 5 shows the behaviors of the permutation quantifiers as a function of the embedding delay for different embedding dimension ($D = 4, \ldots, 7$) together with the curves described by these quantifiers in the CECP. The chaotic fingerprint is observed for the higher embedding dimensions ($D = 6$ and $D = 7$), with a maximum for the permutation statistical complexity and an increasing behavior for the permutation entropy. Indeed, the chaotic dynamics is discriminated without any noise influence. This is reasonable if we take into consideration that the signal-to-noise ratio was about 300, slightly under the half bit uncertainty of the analog to digital conversion. The analysis of this experimental data confirms that higher values of D allow the use of larger minimally required sampling times, retaining all the information about the chaotic dynamics of

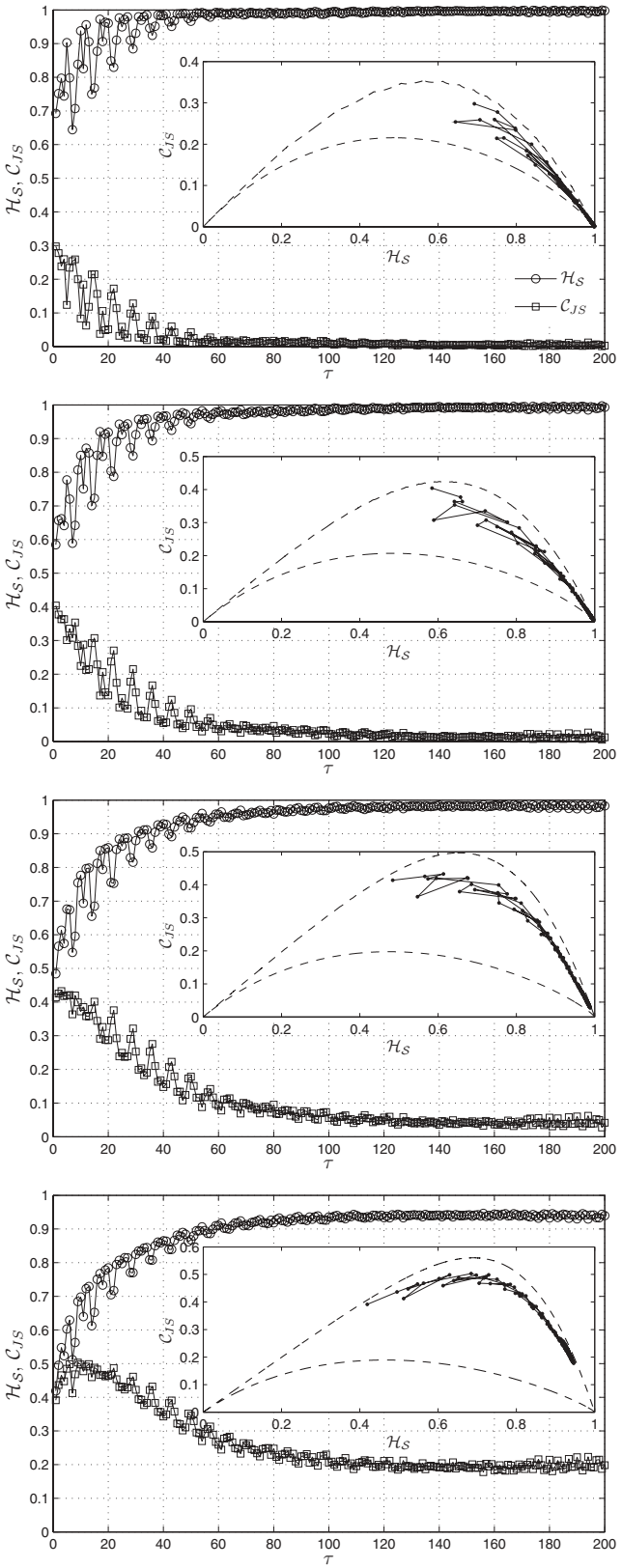

FIG. 5. Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with different embedding dimensions ($D = 4, \ldots, 7$ from top to bottom) for the chaotic laser data. Curves described by the symbolic quantifiers in the CECP are shown in the insets. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

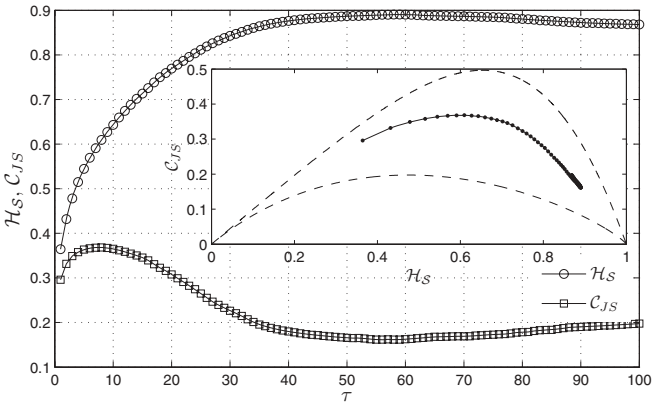

FIG. 6. Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with embedding dimensions $D = 6$ for the Grand River flow dynamics. Qualitative similar results were also obtained for $D = 4$, 5, and 7. Curve described by the symbolic quantifiers in the CECP is shown in the inset. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

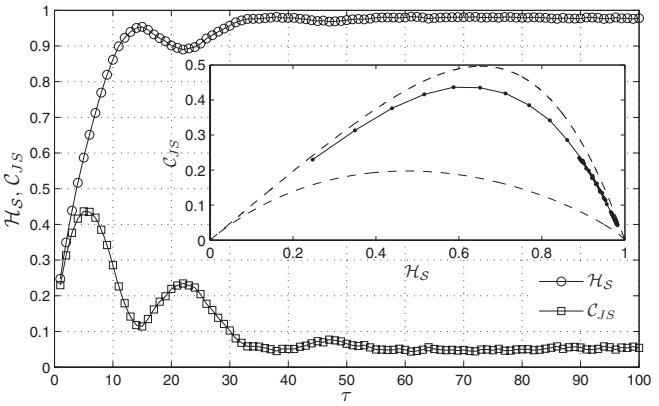

FIG. 7. Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with embedding dimensions $D = 6$ for the Madden-Julian oscillation. Qualitative similar results were also obtained for $D = 4$, 5, and 7. Curve described by the symbolic quantifiers in the CECP is shown in the inset. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

the system under analysis. This fact was recently proved numerically [18].

B. River flow dynamics

Over the last decade controversial results have been obtained about the hypothetical chaotic nature of river flow dynamics [35]. We have analyzed the streamflow data corresponding to the Grand River at Lansing [Michigan; US Geological Survey (USGS) ID 04113000] trying to provide new insights regarding this issue. A high frequency record with a sampling time equal to 15 min over a period of more than two years (from May 2007 to September 2009, $N = 78\,578$ data points) was retrieved from the USGS Instantaneous Data Archive website [36]. The results obtained are shown in Fig. 6. The curve described by the permutation quantifiers in the CECP suggests that the Grand River flow dynamics might be chaotic and, at least for the temporal resolution of the data, noise effects are not observed. $\mathcal{C}_{JS}$ has a well-defined maximum value around $\tau = 8$, and $\mathcal{H}_S$ is increasing over this domain. According to this evidence, the apparent chaotic dynamics associated with this river can be optimally captured with a sampling time equal to 2 h.

C. Madden-Julian oscillation

As a second geophysical application, we have analyzed the Madden-Julian oscillation (MJO) [37]. This is the dominant component of the intraseasonal variability in the tropical atmosphere. Developing suitable models for the MJO is relevant because a wide range of tropical weather and climate are affected by this pattern [38]. Daily MJO time series from 1979 to the middle of 2002 ($N = 8572$ data points) was analyzed via our multiscale symbolic technique looking for evidence of deterministic chaos. The original data are available at the historical climate data archive of the Joint Institute for the Study of the Atmosphere and Ocean website [39]. Figure 7 shows the permutation quantifiers as a function of the embedding delay for this time series. As can be seen in the inset of Fig. 7, the curve described by the symbolic quantifiers in the CECP suggests a noiseless chaotic dynamics for the MJO. The permutation statistical complexity reaches a maximum value for $\tau \sim 5$. Thus, a sampling period equal to five days appears to be the optimal one for uncovering information related with the intrinsic nonlinear correlated dynamics. It is worth noting that another time scale is clearly discriminated, around $\tau = 22$, where both quantifiers reach extreme values simultaneously. This time scale is consistent with a characteristic period of between 40 and 50 days easily estimated through the power spectrum. In the case of periodic functions certain ordinal patterns do not appear, or have very small probabilities, for embedding delays at the half of the period [16]. For this reason, $\mathcal{H}_S$ has a minimum and $\mathcal{C}_{JS}$ has a maximum for these particular embedding delay values [18].

D. North Atlantic oscillation

The North Atlantic oscillation (NAO) is an atmospheric spatiotemporal phenomenon observed over the North Atlantic Ocean. Its fluctuations are quantified through the NAO index, calculated as the difference between the normalized sea level pressure at two action centers, the southernmost one located at the Azores High and the northernmost at the Icelandic Low [40]. Taking into account that the NAO fluctuations have a significant influence on the winter weather over Western and Central Europe, there is a great interest in predicting its behavior. We have studied the monthly mean NAO index, from January 1950 to May 2012 ($N = 749$ data points). The time series was extracted from the Climate Prediction Center website [41]. The results obtained for this third geophysical application are displayed in Fig. 8. Both permutation quantifiers reach nearly constant values independently of the embedding delay τ. According to this finding and by comparing with the results previously shown for fBm (Fig. 2), a chaotic dynamics should be rejected in favor of a purely stochastic dynamics. This is consistent with the stochastic model suggested by other studies on similar data [42]. Additionally, it is shown that

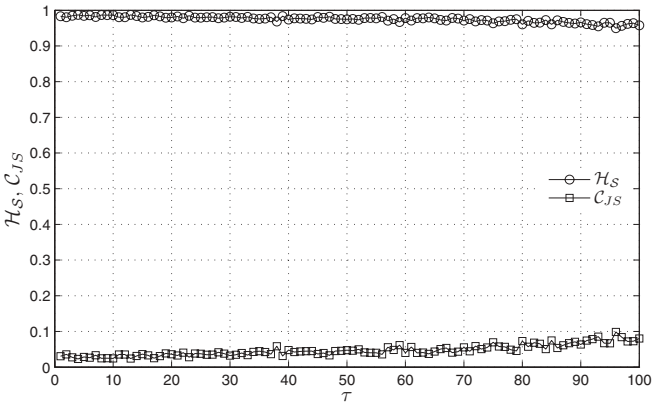

FIG. 8. Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with embedding dimension $D = 5$ for the monthly mean North Atlantic oscillation index. Estimated values for both quantifiers are practically constant whatever the time scale used for the analysis. Qualitative similar results were also obtained for $D = 4$.

very short time series can be efficiently characterized with the proposed scheme.

E. Crude oil and gold price dynamics

It is clear that crude oil and gold are two very important and representative commodities. The former is the major energy source for the present economic activity [43] and the latter one is extremely popular for investment purposes by governments, households, and institutional and private equity investors because it provides economic safety against inflation and deflation [44]. Both commodities tend to rise in periods of financial instability. With the purpose of unveiling some information about the nature of their dynamics, we have analyzed their historical price evolution. The daily closing spot price of the West Texas Intermediate from January 2nd, 1986 to July 10th, 2012 ($N = 6691$ oil price observations) obtained from the US Energy Information Administration website [45] (quoted in US dollars per barrel), and daily gold price from January 2nd, 1973 to July 17th, 2012 ($N = 9841$ gold price observations) extracted from the USAGOLD website [46] (quoted in US dollars per ounce) were analyzed by employing our diagnostic tool. The estimated values for the permutation quantifiers as a function of the embedding delay can be found in Fig. 9. $\mathcal{H}_S$ decreases and $\mathcal{C}_{JS}$ increases for an increasing embedding delay. This effect is more noticeable for the daily gold price evolution. Based on this evidence, stochastic dynamics appear to govern the underlying processes that generated the price fluctuations of both commodities. The lack of a significant chaotic component in the crude oil spot market is in accordance with the results obtained by a very recently published work [47]. Taking into consideration that these two commodity markets are highly correlated [48], it is completely reasonable to find that stochastic rather than deterministic rules are also present in the gold price evolution.

F. Human postural dynamics

The human postural control system, i.e., the human task of maintaining an upright equilibrium, is especially suitable to

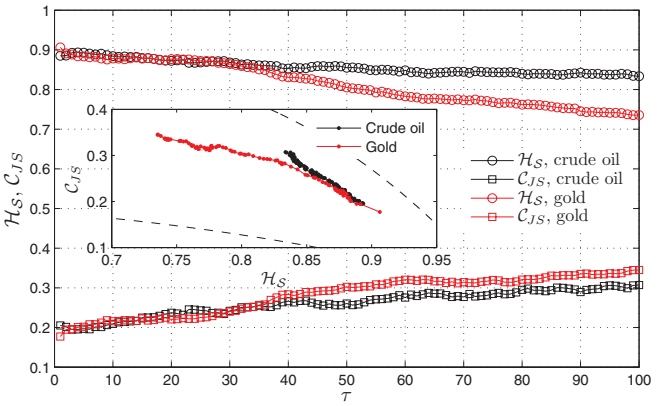

FIG. 9. (Color online) Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with embedding dimension $D = 6$ for the daily price data of the crude oil and gold. In both cases $\mathcal{H}_S$ decreases and $\mathcal{C}_{JS}$ increases for increasing time scales. Qualitative similar results were also obtained for $D = 4$ and $D = 5$. Curves described by the symbolic quantifiers in the CECP are depicted in the inset. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

test our multiscale symbolic approach. Some authors have suggested a correlated stochastic process for modeling the complex behavior exhibited by postural sway in humans [49], while others have claimed the existence of a chaotic dynamics [50]. Moreover, noise is inherently present in all physiological signals due to the recording equipment. Postural dynamics is commonly studied by analyzing the center of pressure (CoP) fluctuations recorded over time while a subject is asked to stand on a force platform. CoP is the point of application of the resultant of vertical forces acting on the basis of support [51]. Particularly, CoP data are never free of noise [27]. We have analyzed the time series of the CoP fluctuations during quiet standing for a healthy young volunteer. Ten independent trials of the displacement in the antero-posterior direction recorded during 30 s at 60 Hz ($N = 1800$ data points) were analyzed.

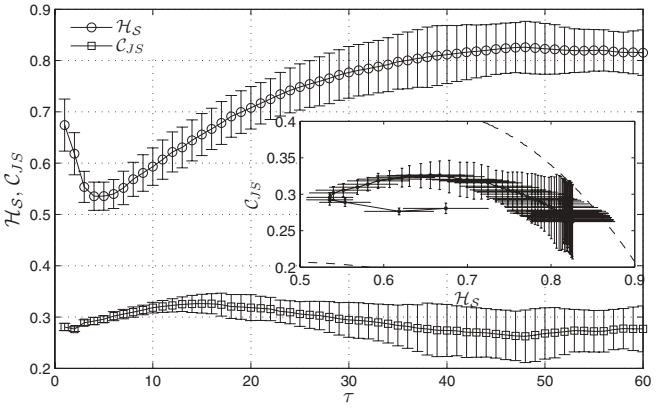

FIG. 10. Permutation quantifiers ($\mathcal{H}_S$ and $\mathcal{C}_{JS}$) as a function of the embedding delays with embedding dimensions $D = 5$ for postural sway measurements. Mean and standard deviation of the permutation quantifiers for ten independent trials associated to the same volunteer are depicted. Curve described by the symbolic quantifiers in the CECP is shown in the inset. The dashed lines represent the maximum and minimum complexity values for a fixed value of the entropy.

Details about the experiment can be found in Ref. [52]. The original data are available in the PhysioNet database [53]. More precisely, records associated to the control trials for volunteer MT1419 were considered. Figure 10 shows the results obtained. The fingerprint related to a noisy chaotic dynamics is observed with a clear clockwise loop in the CECP. Similar results were confirmed for other subjects. Then, a noisy chaotic dynamics seems to be confirmed for the center of pressure fluctuations. Through this example we also confirm that the multiscale CECP can robustly characterize short and noisy data sets. Finally, it is worth remarking that this multiscale symbolic approach has been recently successfully applied for distinguishing hyperchaotic dynamics and noise in experimental time series recorded from a delayed optical system [54].

V. CONCLUSIONS

In this paper we have shown that complex multiscaled time series can be reliably characterized by estimating the permutation entropy and permutation statistical complexity as a function of the embedding delay. The scale is explicitly incorporated in this approach by changing the embedding delay. The location of these permutation quantifiers in the multiscale CECP allows us to infer useful information about the underlying dynamics of the complex time series. More precisely, we are able to identify the range of scales where deterministic or noisy behaviors dominate the system's dynamics. The obtained numerical and experimental results confirm that this multiscale symbolic information-theory approach provides a conceptually simple and computationally efficient tool for characterizing complex time series on a wide range of scales.

According to the evidence gathered from the several tests performed, the presence of a maximum for the permutation statistical complexity at a particular time scale together with an increasing behavior of the permutation entropy around this domain is a necessary condition for the presence of an intrinsic chaotic nature. Nevertheless, it is not possible to confirm that this fingerprint constitutes a sufficient proof for chaos. The existence of stochastic processes for which the permutation quantifiers describe a similar behavior in the CECP cannot be discarded. Further studies should be performed in order to try to elucidate this issue. We would like to emphasize, however, that the symbolic tool proposed in this work is particularly useful for characterizing the complex interplay between the deterministic and stochastic components that are usually found in time series derived from natural and man-made complex phenomena.

In spite of the fact that an inherent limitation of the proposed approach is that oversampled signals are required, the present availability of real-time acquisition reaching sampling rates of up to 120 GSamples/s makes our method feasible for more and more systems. Actually, the coexistence of regular, nonlinear, and stochastic dynamics is a major problem directly associated with such larger temporal resolutions. Taking into account that the multiscale CECP inherits all the good properties of the Bandt and Pompe symbolization recipe, i.e., simplicity, robustness, and very low computational cost, we believe that this approach can be a practical alternative for unveiling the intricate nature of complex multiscaled time series.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their useful comments and suggestions that greatly helped to improve an earlier version of this paper. L.Z. and O.A.R. were supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. O.A.R. acknowledges support as a CNPq fellow, Brazil.

[1] A. R. Osborne and A. Provenzale, Physica D **35**, 357 (1989); A. Provenzale, A. R. Osborne, and R. Soj, *ibid.* **47**, 361 (1991); J.-P. Eckmann and D. Ruelle, *ibid.* **56**, 185 (1992); A. Provenzale, L. A. Smith, R. Vio, and G. Murante, *ibid.* **58**, 31 (1992); M. Dämmig and F. Mitschke, Phys. Lett. A **178**, 385 (1993); L. Bergamasco and M. Serio, Fractals **4**, 73 (1996); U. S. Freitas, C. Letellier, and L. A. Aguirre, Phys. Rev. E **79**, 035201(R) (2009).

[2] A. Ben-Mizrachi, I. Procaccia, and P. Grassberger, Phys. Rev. A **29**, 975 (1984); P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos **3**, 127 (1993); T. Schreiber, Phys. Rev. E **48**, R13 (1993); A. Serletis, A. Shahmoradi, and D. Serletis, Chaos, Solitons Fractals **32**, 883 (2007); M. Lei and G. Meng, *ibid.* **36**, 512 (2008).

[3] M. Cencini, M. Falcioni, E. Olbrich, H. Kantz, and A. Vulpiani, Phys. Rev. E **62**, 427 (2000).

[4] J. B. Gao, J. Hu, W. W. Tung, and Y. H. Cao, Phys. Rev. E **74**, 066204 (2006).

[5] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).

[6] M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, Physica A **330**, 53 (2003); M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. E **71**, 021906 (2005); J. Escudero, D. Abásolo, R. Hornero, P. Espino, and M. López, Physiol. Meas. **27**, 1091 (2006); M. Costa, I. Ghiran, C.-K. Peng, A. Nicholson-Weller, and A. L. Goldberger, Phys. Rev. E **78**, 020901(R) (2008); J. Hu, J. B. Gao, and W. W. Tung, Chaos **19**, 028506 (2009); J. Hu, J. B. Gao, W. W. Tung, and Y. H. Cao, Ann. Biomed. Eng. **38**, 854 (2010); T. Takahashi, R. Y. Cho, T. Mizuno, M. Kikuchi, T. Murata, K. Takahashi, and Y. Wada, Neuroimage **51**, 173 (2010); J. B. Gao, J. Hu, W. W. Tung, and Y. Zheng, Quant. Finance (2011), doi: 10.1080/14697688.2011.580774.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, New York, 2006).

[8] D. P. Feldman and J. P. Crutchfield, Phys. Lett. A **238**, 244 (1998).

[9] D. P. Feldman, C. S. McTague, and J. P. Crutchfield, Chaos **18**, 043106 (2008).

[10] P. W. Lamberti, M. T. Martín, A. Plastino, and O. A. Rosso, Physica A **334**, 119 (2004).

[11] R. López-Ruiz, H. L. Mancini, and X. Calbet, Phys. Lett. A **209**, 321 (1995).

[12] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley, Phys. Rev. E **65**, 041905 (2002).
[13] M. T. Martín, A. Plastino, and O. A. Rosso, Physica A **369**, 439 (2006).
[14] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[15] M. Staniek and K. Lehnertz, Int. J. Bifurcation Chaos **17**, 3729 (2007).
[16] C. Bandt, Ecol. Modell. **182**, 229 (2005).
[17] L. Zunino, M. C. Soriano, I. Fischer, O. A. Rosso, and C. R. Mirasso, Phys. Rev. E **82**, 046212 (2010).
[18] M. C. Soriano, L. Zunino, O. A. Rosso, I. Fischer, and C. R. Mirasso, IEEE J. Quantum Electron. **47**, 252 (2011).
[19] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Europhys. Lett. **79**, 50001 (2007); M. Zanin, Chaos **18**, 013119 (2008); G. Ouyang, X. Li, C. Dang, and D. A. Richards, Phys. Rev. E **79**, 041146 (2009); J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Int. J. Bifurcation Chaos **20**, 2915 (2010); J. Tiana-Alsina, J. M. Buldú, M. C. Torrent, and J. García-Ojalvo, Philos. Trans. R. Soc. London, Ser. A **368**, 367 (2010).
[20] O. A. Rosso, L. Zunino, D. G. Pérez, A. Figliola, H. A. Larrondo, M. Garavaglia, M. T. Martín, and A. Plastino, Phys. Rev. E **76**, 061114 (2007).
[21] O. A. Rosso, H. A. Larrondo, M. T. Martín, A. Plastino, and M. A. Fuentes, Phys. Rev. Lett. **99**, 154102 (2007).
[22] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[23] L. Zunino, O. A. Rosso, and M. C. Soriano, IEEE J. Sel. Top. Quantum Electron. **17**, 1250 (2011).
[24] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, Comput. Biol. Med. **42**, 319327 (2012).
[25] M. C. Mackey and L. Glass, Science **197**, 287 (1977).
[26] L. De Micco, J. G. Fernández, H. A. Larrondo, A. Plastino, and O. A. Rosso, Physica A **391**, 2564 (2012).
[27] C. K. Rhea, T. A. Silver, S. L. Hong, J. H. Ryu, B. E. Studenka, C. M. L. Hughes, and J. M. Haddad, PLoS ONE **6**, e17696 (2011).
[28] J. B. Gao, J. Hu, W. W. Tung, Y. H. Cao, N. Sarshar, and V. P. Roychowdhury, Phys. Rev. E **73**, 016117 (2006).
[29] J.-F. Coeurjolly, J. Stat. Software **5**, 1 (2000).
[30] C. Bandt and F. Shiha, J. Time Ser. Anal. **28**, 646 (2007); M. Sinn and K. Keller, Comput. Stat. Data Anal. **55**, 1781 (2011).
[31] C.-S. Poon and M. Barahona, Proc. Natl. Acad. Sci. USA **98**, 7107 (2001).
[32] J. B. Gao, J. Hu, X. Mao, and W. W. Tung, Chaos, Solitons Fractals **45**, 213 (2012).
[33] See http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html.
[34] U. Hübner, N. B. Abraham, and C. O. Weiss, Phys. Rev. A **40**, 6354 (1989).
[35] B. Sivakumar, R. Berndtsson, J. Olsson, and K. Jinno, Hydrol. Sci. J. **46**, 131 (2001); D. Schertzer, I. Tchiguirinskaia, S. Lovejoy, P. Hubert, H. Bendjoudi, and M. Larchevêque, *ibid.* **47**, 139 (2002); B. Sivakumar, Chaos, Solitons Fractals **19**, 441 (2004); J. D. Salas, H. S. Kim, R. Eykholt, P. Burlando, and T. R. Green, Nonlinear Processes Geophys. **12**, 557 (2005).
[36] See http://ida.water.usgs.gov.
[37] R. A. Madden and P. R. Julian, Mon. Weather Rev. **122**, 814 (1994).
[38] W. W. Tung, J. B. Gao, J. Hu, and L. Yang, Phys. Rev. E **83**, 046210 (2011).
[39] See http://jisao.washington.edu/data/mjo/.
[40] I. Fernández, C. N. Hernández, and J. M. Pacheco, Physica A **323**, 705 (2003).
[41] See http://www.cpc.ncep.noaa.gov/.
[42] D. B. Stephenson, V. Pavan, and R. Bojariu, Int. J. Climatol. **20**, 1 (2000); C. Collette and M. Ausloos, Int. J. Mod. Phys. C **15**, 1353 (2004); P. G. Lind, A. Mora, M. Haase, and J. A. C. Gallas, Int. J. Bifurcation Chaos **17**, 3461 (2007); M. D. Martínez, X. Lana, A. Burgueño, and C. Serra, Nonlinear Processes Geophys. **17**, 93 (2010); I. Fernández, J. M. Pacheco, and M. P. Quintana, Physica A **389**, 5801 (2010).
[43] E. Martina, E. Rodriguez, R. Escarela-Perez, and J. Alvarez-Ramirez, Energy Econ. **33**, 936 (2011).
[44] M. Bolgorian and Z. Gharli, Acta Phys. Pol. B **42**, 159 (2011); D. Ghosh, S. Dutta, and S. Samanta, *ibid.* **43**, 1261 (2012).
[45] See http://www.eia.gov/.
[46] See http://www.usagold.com/.
[47] J. T. Barkoulas, A. Chakraborty, and A. Ouandlous, Energy Econ. **34**, 584 (2012).
[48] Y.-J. Zhang and Y.-M. Wei, Resour. Pol. **35**, 168 (2010).
[49] J. J. Collins and C. J. De Luca, Phys. Rev. Lett. **73**, 764 (1994).
[50] N. Yamada, Hum. Mov. Sci. **14**, 711 (1995).
[51] S. Ramdani, B. Seigle, J. Lagarde, F. Bouchara, and P. L. Bernard, Med. Eng. Phys. **31**, 1023 (2009).
[52] A. A. Priplata, J. B. Niemi, J. D. Harry, L. A. Lipsitz, and J. J. Collins, Lancet **362**, 1123 (2003).
[53] See http://www.physionet.org/physiobank/database/nesfdb/.
[54] M. C. Soriano, L. Zunino, L. Larger, I. Fischer, and C. R. Mirasso, Opt. Lett. **36**, 2212 (2011).