Detecting dynamical changes in time series by using the Jensen Shannon divergence

D. M. Mateos, L. E. Riveaud, and P. W. Lamberti

Citation: Chaos **27**, 083118 (2017); doi: 10.1063/1.4999613
View online: http://dx.doi.org/10.1063/1.4999613
View Table of Contents: http://aip.scitation.org/toc/cha/27/8
Published by the American Institute of Physics

Detecting dynamical changes in time series by using the Jensen Shannon divergence

D. M. Mateos,[1,a)] L. E. Riveaud,[2] and P. W. Lamberti[2]
[1]*Neuroscience and Mental Health Programme, Division of Neurology, Hospital for Sick Children, Institute of Medical Science and Department of Paediatrics, University of Toronto, Toronto, Ontario M5G 0A4, Canada*
[2]*Facultad de Matemática, Astronomía, Física y Computación (FaMAF), Universidad Nacional de Córdoba, CONICET, Córdoba, Argentina*

(Received 17 March 2017; accepted 9 August 2017; published online 23 August 2017)

Most of the time series in nature are a mixture of signals with deterministic and random dynamics. Thus the distinction between these two characteristics becomes important. Distinguishing between chaotic and aleatory signals is difficult because they have a common wide band power spectrum, a delta like autocorrelation function, and share other features as well. In general, signals are presented as continuous records and require to be discretized for being analyzed. In this work, we introduce different schemes for discretizing and for detecting dynamical changes in time series. One of the main motivations is to detect transitions between the chaotic and random regime. The tools here used here originate from the Information Theory. The schemes proposed are applied to simulated and real life signals, showing in all cases a high proficiency for detecting changes in the dynamics of the associated time series. *Published by AIP Publishing.*
[http://dx.doi.org/10.1063/1.4999613]

Detecting different dynamics present in a single signal is a topic of great importance from both theoretical and practical point of view. Therefore, for a long time, it has become a subject of relevance in many areas of science. The detection of changes in the dynamics of a system is not an easy task, especially between chaotic and random dynamics, because they share many characteristics.
In this paper, we propose to address these issues by using a tool derived from theInformation Theory, known as the Divergence of Jensen Shannon. This quantity is a measure of dissimilarity between probability distributions. In order to provide the pertinent probability distributions, we propose a method for discretizing the original signal, known as the alphabetic mapping.
The union of these two methods allows discretizing and analyzing any continuous signal. From this, we can evaluate the "distance" between two or more signals, or to detect the point at which the signal changes its dynamics.
To test our schemes we have used artificially generated signals taken from the bibliography (chaotic maps and noises), as real (electrocardiogram, mechanics signals) ones. The methods here introduced give a new way to distinguish different dynamical behavior, in an easy and fast way.

I. INTRODUCTION

In many areas of science, the dynamics of processes underlying the studied phenomena can be represented by means of finite time series, interpreted in some cases as realizations of stochastic processes (SPs) and in other cases as orbits of deterministic processes. Although these two types of processes are very different in nature, it is possible to show that under certain restrictions there is a relationship between them.[1] A theorem by H. Wold establishes that a stationary stochastic process can be decomposed as a deterministic part (DP), which can be described accurately by a linear combination of its own past, and another part represented as a moving average component of finite order.

A different situation occurs when we have a non stationary time series. In these situations it is not possible to separate the series in a deterministic and a stochastic part. Moreover in cases in which the DP is chaotic, it is possible to find a DP which produces a time series that could be very difficult to distinguish from a SP.[2,3] For these reasons, the issue of distinguishing between time series produced by deterministic chaos from those produced by a random dynamics has led to the development of different approaches.

This problem has already been treated with various techniques such as: complexity-entropy plane,[4–6] Lyapunov exponents[2,7,8] or applying neural networks,[9] among others. Chaotic signals always produce time series with a strong physical structure, unlike those originated in stochastic processes which have a little structure depending on their correlations factors.[10]

The basic idea behind our schemes is to use quantifiers that measure how close or far two time series are. The quantifier is the *Jensen–Shannon Divergence* (JSD), which was introduced as a measure of similarity between two probability distribution function (PDF).[11] Quantifiers originated in the Information Theory, such as the JSD require to associate to a given time series in a probability distribution (PD). Thus the determination of the most suitable PD is a fundamental point, because PD and the sample space are inextricably linked. Many methods to deal with this problem have been proposed. Among others methods of mapping between the

[a)]mateosdiego@gmail.com

time series and the PD, we can mention the binary symbolic dynamics,[12] Fourier analysis,[13] frequency counting,[14] wavelet transform[15] or comparing the order of neighboring relative values.[16] Here, we propose to explore a way of mapping continuous -state sequences into discrete-state sequences *via* the so called *alphabetic mapping*.[17]

In this work, we present two methods for analyzing time series. The first one involves a relative distance between two time series. It consists in evaluating the JSD to previously discretized signals, by using the alphabetic mapping. The quantity resulting from this evaluation will be called the *alphabetic Jensen Shannon divergence* (aJSD). The second method detects the changes in the dynamics of a time series through a sliding window which runs over the previously discretized signal. This window is divided into two parts. We can assign the corresponding PD to each part, which can be compared by using the JSD. In both methods, the changes in the underlying dynamics of the time series can be detected by analyzing the changes in the values of the JSD at the neighborhood of the time when the change occurs. Both procedures were tested by using well known chaotic and random signals taken from the literature and from biophysical and mechanical real records.

The structure of this manuscript is as follows: in Sec. II, we give a brief introduction of the JSD, and we explain in detail the alphabetic mapping. In Sec. III, we introduce the aJSD. In Sec. IV, we give a characterization of the chaotic maps and colored noises used in the implementation of the methods proposed here. In Sec. V, we display the results obtained by the application of those schemes. In Sec. VI, we illustrate how the aJSD can be used with real data extracted from physiological and mechanical records. Finally, we draw up some conclusions.

II. JENSEN–SHANNON DIVERGENCE FOR ALPHABETIC MAPPING

A. Jensen–Shannon divergence

The Jensen–Shannon divergence (JSD) is a measure of distinguishability between probability distributions introduced by Rao[18] and Lin[19] as a symmetrized version of the Kullback–Leibler divergence. This quantity has been used to analyse problems which arise in many areas of science.[20–24] Grosse and co-workers presented a recursive algorithm for the segmentation (i.e., the detection of non-stationarity) for symbolic sequences based on the JSD.[24] Our procedures are extensions of that segmentation procedure.

Let X be a discrete-states variable x_i, $i = 1, 2, \ldots, N$ and let P_1 and P_2 be two probability distributions for X, which we denote as $p_i^{(1)} = P_1(x_i)$ and $p_i^{(2)} = P_2(x_i)$, with $0 \leq p_i^{(k)} \leq 1$ and $\sum_{i=1}^{N} p_i^{(k)} = 1$ for all $i = 1, 2, \ldots, n$ and $k = 1, 2$. If π_1 and π_2 denote the weights of P_1 and P_2 respectively, with the restrictions $\pi_1 + \pi_2 = 1$ and $\pi_1, \pi_2 \geq 0$, the JSD is defined by

$$D_{JS}(P_1, P_2) = H(\pi_1 P_1 + \pi_2 P_2) - (\pi_1 H(P_1) + \pi_2 H(P_2)), \quad (1)$$

where

$$H(P) = -C \sum_{i=1}^{N} p_i \log(p_i) \quad (2)$$

is the Shannon entropy for the probability distribution P. We take $C = 1/\log(2)$ so that the entropy (and also the JSD) is measured in bits.

It can be shown that the JSD is positively defined, symmetric and it is zero if and only when $P_1 = P_2$.[11] Moreover JSD is always well defined, and it is the square of a metric[25] and it can be generalized to compare an arbitrary number of distributions, in the following way: let $P_1(x), \ldots, P_M(x)$ be a set of probability distributions and let $\pi_1, \ldots, \pi_M$ be a collection of non negative numbers such that $\sum_{j=1}^{M} \pi_j = 1$. Then the JSD among the probability distributions $P_j(x)$ with $j = 1, .., M$ is defined by

$$D_{JS}(P_1, \ldots, P_M) = H\left[\sum_{j=1}^{M} \pi_j P_j\right] - \sum_{j=1}^{M} \pi_j H[P_j]. \quad (3)$$

Finally it is worth mentioning that, although the JSD was introduced as a measure of discernability between discrete probability distributions, it is possible to extend its evaluation to continous probability distributions (see for example Ref. 26). In this work we only need to evaluate the JSD between discrete PD.

B. Alphabetic mapping

In nature, symbolic sequences can be found, such as DNA sequences or some symbolic sequences generated by logical circuits. But most of "real life" signals are represented as continuous records. Thus, for certain kinds of processing, it is necessary to discretize them. This is the case when JSD is applied (It is possible to calculate entropy for continuous state signals but the estimation of a differential entropy from the data is not an easy task.[27]). There exist several ways to discretize the original signal. We use a scheme introduced by Yang.[17]

For a given continuous series $\{x_t\}$, we can map this real-values series in a binary series easily, depending on the relative values between two neighboring points $[x_t; x_{t+1}]$ in the following way:

$$s_n = \begin{cases} 0, & \text{if } x_t \leq x_{t+1} \\ 1, & \text{if } x_t > x_{t+1}. \end{cases}$$

So we obtain a binary sequence $S = \{s_1, s_2, \ldots, s_{N-1}\}$, where $s_i = 0$ or 1.

Consider two integers $d > 2$ and $\tau \geq 1$ and let us define a trajectory in the d-dimensional space associated with $\{x_t\}$:

$$\mathbf{Y}_t \rightarrow [s_{t-(d-1)\tau}, \ldots, s_t] \qquad t \geq (d-1)\tau.$$

The vector $\mathbf{Y}_t^{(d,\tau)}$ is called "alphabetic word," where d is the "embedding dimension" (the number of bit taken to create the word) and τ is called the "delay." Taken's theorem gives conditions on d and τ, such that $\mathbf{Y}_t^{d,\tau}$ preserves the dynamical properties of the full dynamical system (e.g.,, reconstruction of strange attractors).[28,29]

By shifting one data point at time, the algorithm produces a collection of bit-word $\{\mathbf{Y}^{d,\tau}\}$ over the whole series. Therefore, it is plausible that the occurrence of this bit-word reflects the underlying dynamics of the original time series. Different types of dynamics produce different distributions of these $\mathbf{Y}^{d,\tau}$ series. We define $w_i^{d,\tau}$, as the symbol corresponding to the word $\mathbf{Y}_i^{d,\tau}$. From these, we construct a new series $W^{d,\tau} = \{w^{d,\tau}\}$ which quantify the original series. The number of the different symbols (alphabet length) depends on the number of bits taken; in this case is 2^d.

To give an example of the mapping, we can consider the series $\{3.5, 4.8, 3.6, 2, 1, 4.1, 3.7, 8.5, 10.4, 8.9\}$, which has a corresponding series $S = \{1, 0, 0, 1, 0, 1, 1, 0\}$. For the evaluation of the parameter $d = 4$ and $\tau = 1$, the first word to appear is $\mathbf{Y}_1^{4,1} = (1, 0, 0, 1)$, the second one is $\mathbf{Y}_2^{4,1} = (0, 0, 1, 0)$, and the other three are $\mathbf{Y}_3^{4,1} = (0, 1, 0, 1)$, $\mathbf{Y}_4^{4,1} = (1, 0, 1, 1)$, and $\mathbf{Y}_5^{4,1} = (0, 1, 1, 0)$.

It is necessary to frequently process signals of two or more dimensions such as bi-dimensional chaotic maps, polysomnography, EEG, etc. The components of such signals are mostly coupled, given that signal values depend not only on the previous values but also on the values reached by the other signals. Therefore by making a one-dimensional analysis, we can lose some valuable information. Based on the ideas discussed above for one-dimensional signals (1D), we have extended the same algorithm, with a slight modification to analyze 2-dimensional (2D) signals without losing information. For a given continuous 2D series $\mathbf{X}_t = (x(t), y(t))$, we assign a 1D string, by the relative values between the two component vector at each time t, $[x_t; y_t]$ in the following way:

$$S_t = \begin{cases} 0 & \text{si} \quad x_t \geq y_t \\ 1 & \text{si} \quad x_t < y_t \end{cases}.$$

For the 1D case, we use a time delay method because we consider $x(t)$ and $x(t+1)$ to binarize the signal. For the 2D signal, the binarization is produced by a spatial assignation as we compare the values of the two coordinate $(x(t), y(t))$ at the same time. However, when we define the embedding vector for the binary sequence the procedure is equal in both 1D and 2D signals.

It should be emphasized that it is possible to extend the procedure for N-dimencional signals. For example for N-dimensional signals, we can use the permutation vector approach, introduced by Bandt and Pompe[16] (see for example Ref. 30). In the N-dimensional case, the sequence S it is no longer a binary one, but it depends on the size of the signal, having an alphabet length $L = N!$ and the number of possible words W results in $N!^2$.

III. JSD COMBINED WITH THE ALPHABETIC MAPPING

Once the signals are discretized, we can approximate the PD by the frequency of occurrence of the symbols $w_i^{d,\tau}$. From this PD, we can develop the following analysis schemes:

- Distance between two signals: to define a distance between two signals we use the following procedure – given two different time series, for example, one originated from a chaotic dynamics and the other from a random dynamics, we map each one in a symbolic sequence by using the methods explained in Sec. II B. In this way, we get two sets of symbols $W^{(d,\tau)} = \{w^{(d,\tau)}\}$ and $\tilde{W}^{(d,\tau)} = \{\tilde{w}^{(d,\tau)}\}$. Then we can calculate the frequency of appearance of the symbols for both sequences $P^W = P(W^{d,\tau})$ and $P^{\tilde{W}} = P(\tilde{W}^{d,\tau})$. Finally, we evaluate the Jensen Shannon divergence between these two distributions $D_{JS}^{W,\tilde{W}} = D_{JS}(P^W|P^{\tilde{W}})$.
- Sliding Window: here we introduce a sliding window that moves over the symbolic sequence corresponding to the original signal. The window has a width $\Delta > 0$ and position k (referring to the position of the center of the window over the sequence). For each position k, we can divide the window in two sub windows, one to the left and the other to the right of the position k. For each windows, we evaluate the frequency of occurrence of symbolic patterns: to the right window $P(W_r^{d,\tau}) = P_r^W$ and to the left window $P(W_l^{d,\tau}) = P_l^w$. Finally we evaluate the associated JSD, $D_{JS}(k) = D_{JS}(P_r^W|P_l^W)$ as a function of the window's position k. The position where the maximum value of the JSD occurs, $D_{JS_{max}} = max[D_{JS}(k)]$, it is interpreted as the place where a significant change in the probability distribution patterns $W^{(d,\tau)}$, has occurred. This change can be associated to a variation in the statistical properties of the original signal. The only restriction in the election of window's width is that Δ must be greater than the number of possible patterns generated by the alphabetic mapping ($\Delta \gg 2^d$).

Two sequences with the same statistical properties, should lead to identical probability distributions and therefore, the divergence between them should take a very small value, close to zero but non zero. This fact is due to the statistical fluctuations. The estimators for probability distributions corresponding to sequences must be constructed and then the fluctuations of this construction will yield to JSD values greater than zero. To address this problem, Grosse *et al.*[24] introduced a quantity called "significance" which allows to see if the values reached by the JSD are greater than the statistical fluctuations. This amount depends on the sequence length and the size of the alphabet of symbols used in the representation of the sequence. An expression for the significance value has been presented in Ref. 24. A limitation of that expression is that it is valid only for an alphabet with no more than five symbols. Therefore, we must modify the criteria introduced by Grosse *et al.*, to identify values of the JSD that are genuinely above statistical fluctuations. To do that, we proposed to calculate the aJSD on a set of $Ns = 10^5$ ensembles generated for each signal using the same parameters but with different initial conditions. After quantifying the signals, we have different Ns sequences $(W_1^{d,\tau}, \ldots, W_{Ns}^{d,\tau})$.

The first step is to calculate the aJSD between all sequences belonging to the same group of segments, that we call "auto-aJSD"

$$D_{JS}^W = D_{JS}(P^{W_i}|P^{W_j}), \quad i, j = 1, \ldots, Ns, \quad \textit{for } i \neq j.$$

Then we evaluate the average value $\mu^W = \langle D_{JS}^W \rangle$ over all the sets of sequences with its respective standard deviation

($\sigma^W = \langle (D_{JS}(P^{W_i}|P^{W_j}) - \mu^W)^2 \rangle^{1/2}$). The next step is the same, but using the two group of signals to be compared. For all the aJSD values resulting from all the signals, we take the average $\mu^{W,\tilde{W}} = \langle D_{JS}(P^{W_i}|P^{\tilde{W}_i}) \rangle$, with its respective standard deviation ($\sigma^{W,\tilde{W}} = \langle (D_{JS}(P^{W_i}|P^{\tilde{W}_i}) - \mu^{W,\tilde{W}})^2 \rangle^{1/2}$).

Finally, two sequences are different (in the statistical sense) if the inequality:

$$\mu^{W,\tilde{W}} - \sigma^{W,\tilde{W}} \geq \max\left[\mu^W + \sigma^W, \mu^{\tilde{W}} + \sigma^{\tilde{W}}\right]$$

is satisfied. If the values of the aJSD do not pass this criteria we say that the two signals are not statistically distinguishable one from each other.

IV. CHARACTERIZATION OF CHAOTIC MAPS AND COLORED NOISES

To test our scheme based on the distance between two signals, we use sequences extracted from the bibliography. We use 18 chaotic maps and 5 colored noises in what follows will be described briefly.

A. Chaotic maps

We consider 18 chaotic maps which were taken from Ref. 31. They can be grouped as follows:

(1) 1D chaotic maps: also called non–inverted maps. They are dynamical sequences for which the image has more than one pre-image and in each interaction a loss of information occurs, generating in this way a chaotic system.[31]

- The lineal congruential generator.[32]
- The gaussian map.[33]
- The logistic map.[34]
- The Pinchers map.[35]
- The Ricker's population model.[36]
- The sine circle.[37]
- The sine map.[38]
- The Spencer map.[39]
- The tent map.[40]
- The Hénon map.[41]
- The Lonzi map.
- The delayed logistic map.[42]
- The Tinkerbell map.[43]
- The dissipative standard map.[44]
- The Arnold's cat map.[45]
- The chaotic web map.[46]
- The Chirikov standard map.[47]
- The Gingerbreadman map.[48]
- The Hénon area-preserving quadratic map.[41]

B. Colored noises

The noise power spectrum often varies with frequency as $1/f^{\alpha}$ (some times called *Hurst noise*). *White noise* corresponds to $\alpha = 0$. To generate all noises we used the algorithm described in Refs. 4 and 49.

V. DISTINGUISHING BETWEEN CHAOTIC AND RANDOM SEQUENCES

In this section, we present the results obtained from evaluating the aJSD between chaotic maps and colored noises. We apply the two methods described in Sec. II. First, we create an aJSD distance matrix between chaotic and noise signals and between chaotic maps. Next, we merge two different signals (i.e., one chaotic and one noisy) and through the aJSD sliding window method we show that it is possible to detect changes in the types of dynamics.

A. Distance matrix between sequences

For each type of process explained in Secs. IV A and IV B we generate $N_s = 10^6$ time series of $L_s = 10^6$ data points with identical parameters[31] and a random initialization. We compute a distance matrix between chaotic and colored noises, using the significance criterion for the aJSD already explained. For the discretization of the signals, we used the parameters $\tau = 1$ and $8 \leq d \leq 12$. Figure 1 displays the matrix for the corresponding parameters $d = 8$ and $\tau = 1$. For different embedding dimensions, we obtained similar results.

In the case of the aJSD distances matrix corresponding to chaos-noise (Fig. 1), we observe that most of the chaotic maps are distinguishable from the different types of colorated noises. The numbers in the boxes represent the values of the aJSD. The lower these values, more similar are the sequences.

Only for the particular case of the lineal congruential generator map (LCG) and white noise (WN), the aJSD value does not pass the significance criterion. This fact means that the LCG map is an example of a random number generator passing the Miller–Rabin test.[50] Therefore, the distribution of words $\{W^{(8,1)}\}$ corresponding to LCG and WN is very similar.

The same evaluation has been done between chaotic maps. Figure 2 displays the corresponding distance matrix. As it can be seen, all the aJSD values are above the significance criterion, showing that our method is adequate to distinguish between different types of chaos. A more detailed analysis shows a strong relationship between the values of aJSD and the phase diagram of the chaotic maps. For example, the logistic map and the tent map have similar phase diagrams (all phase diagrams belong to the chaotic time series used in this work are shown in the Sprott book (Appendix[31]), and the value of the corresponding distance is almost null. The same behaviour can be found between the Pincher's map and the Spencer map.

The same analysis was performed for the 2D chaotic series. In this case, we use the two-dimensional assignment described in Sec. II B, where the parameters $8 \leq d \leq 12$ and $\tau = 1$ were chosen. Figure 3 shows the corresponding aJSD distance matrix for the parameters $d = 8$ and $\tau = 1$. As observed for one-dimensional chaotic maps, all distances passed the significance criteria, with all maps being distinguishable one from each other.

There exists a strong correspondence between the values of the aJSD and the similarity (or dissimilarity) of the topology of the phase diagrams of the maps. The aJSD's values

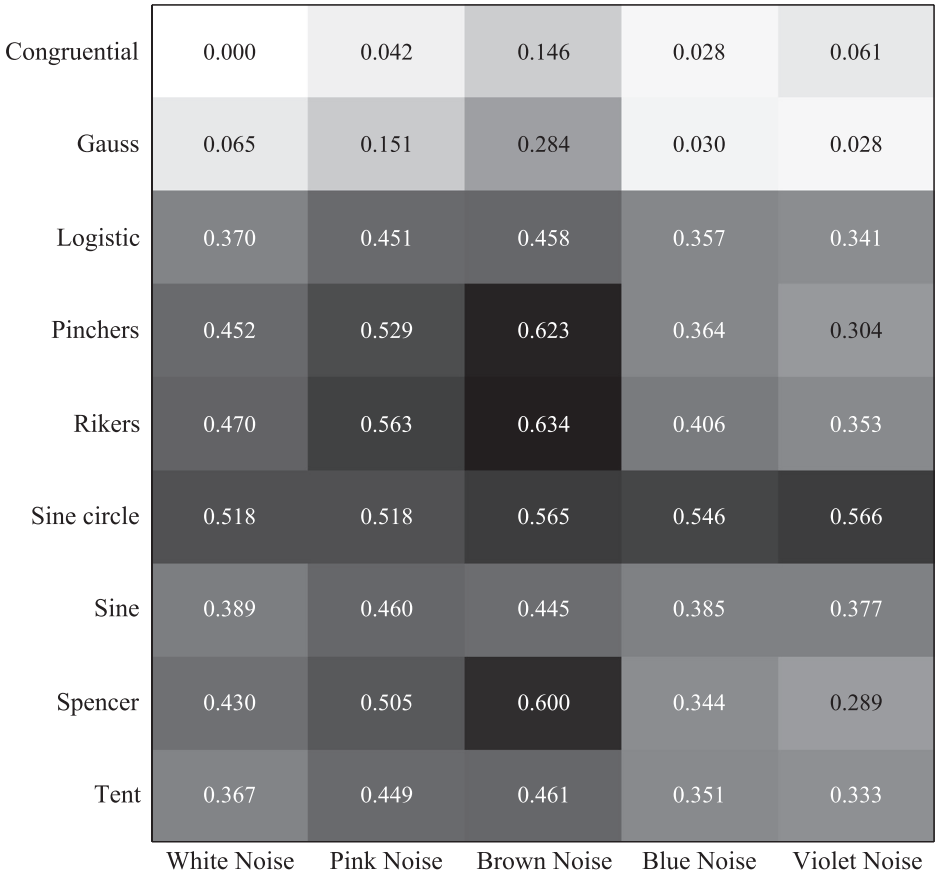

FIG. 1. aJSD distance matrix between 1D chaotic sequences and colored noise signals. The values chosen for the parameters d and tau are: $d = 8$ and $\tau = 1$. It can be observed that there is a good discrimination between chaotic sequences and noise. aJSD values vary according to the correlation noise and the dynamics of the chaotic series. The value 0.00 means that the aJSD has not exceeded the significance criterion settled out in Sec. III so that the two sequences are indistinguishable from each other. Similar results were found for parameters $8 \leq d \leq 12$ and $\tau = 1$, showing that the method is robust with respect to the election parameter d.

decrease as the topology of the phase diagram tends to be similar, as can be observed in the case of the Hénon map and the Lonzi map. Conversely, when two maps have a different topological structure in their phase space, such as the Hénon map and the Chirikov map, the values of the aJSD increase. Different embedding dimensions change the absolute value of the aJSD but the relative value between the elements of the distance matrix remains unaltered.

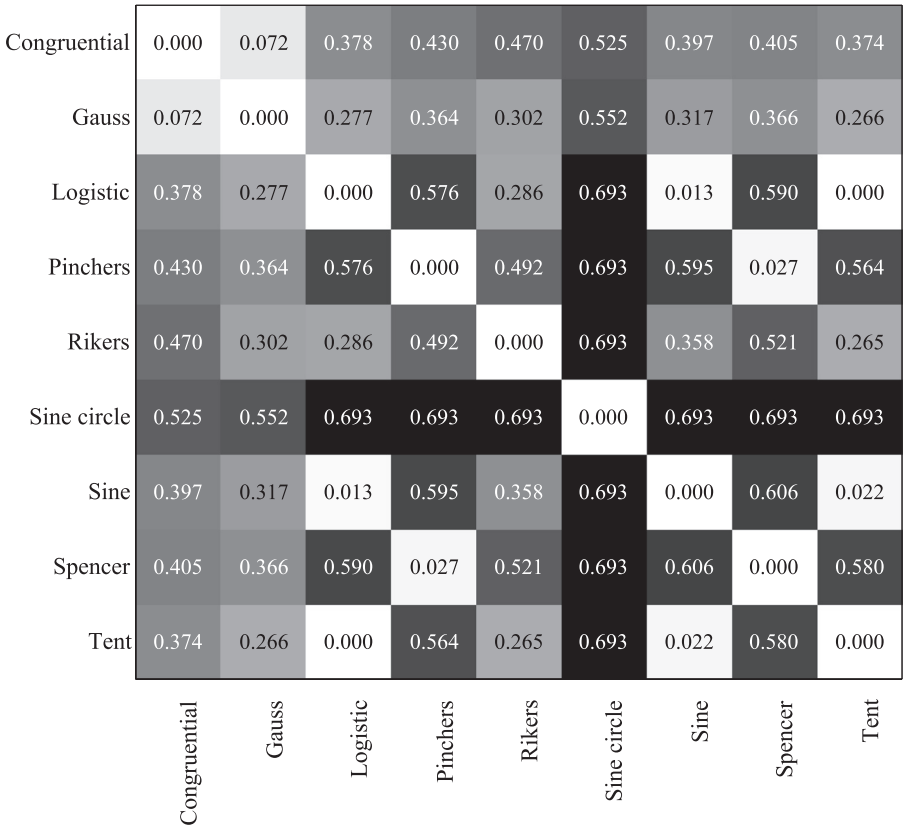

FIG. 2. aJSD distance matrix between 1D chaotic sequences. The values chosen for the parameters d and τ are: $d = 8$ and $\tau = 1$. A good discrimination can be observed between the different chaotic sequences. The absolute values of aJSD vary according to the similarity between the phase diagrams of the chaotic sequences, giving information of the similarity about the dynamics of such sequences. Similar results were found for parameters $8 \leq d \leq 12$ and $\tau = 1$.

	Henon	Lonzi	Delayed	Tinkerbell	Dissipative	Arnolds cat	Gingerbread	Chirikov	Henon area
Henon	0.000	0.019	0.691	0.645	0.502	0.554	0.693	0.674	0.693
Lonzi	0.019	0.000	0.692	0.649	0.503	0.555	0.693	0.680	0.693
Delayed	0.691	0.692	0.000	0.265	0.671	0.550	0.594	0.453	0.633
Tinkerbell	0.645	0.649	0.265	0.000	0.611	0.478	0.649	0.524	0.663
Dissipative	0.502	0.503	0.671	0.611	0.000	0.161	0.593	0.656	0.603
Arnolds cat	0.554	0.555	0.550	0.478	0.161	0.000	0.551	0.456	0.574
Gingerbread	0.693	0.693	0.594	0.649	0.593	0.551	0.000	0.693	0.052
Chirikov	0.674	0.680	0.453	0.524	0.656	0.456	0.693	0.000	0.693
Henon area	0.693	0.693	0.633	0.663	0.603	0.574	0.052	0.693	0.000

FIG. 3. aJSD distance matrix between 2D chaotic sequences. The values chosen for the parameters d and τ are: $d=8$ and $\tau=1$. The matrix shows a good discrimination between the different chaotic sequences and aJSD values. As in the case of 1D chaotic maps, which are more similar for the phase space, the value of aJSD is lower. Similar results were found for parameters $8 \leq d \leq 12$ and $\tau=1$.

Let us recall that in our scheme the aJSD measures the distance between the PDF associated with the set of words $\{W^{(d,\tau)}\}$ and $\{\tilde{W}^{(d,\tau)}\}$. As a consequence of the Taken's theorem,[29] these sets of words are in correspondence with certain aspects of the phase space of each original signal. Series which have a similar dynamics, have similar phase space and therefore PDF giving low values of the aJSD.

B. Detection changes in the time series

Here, we use the proposed sliding window scheme for detecting changes in a signal. For this purpose, we use two different signals x_1 and x_2 of equal length $L_{x_1}=L_{x_2}=5\times10^4$ symbols, which are merged in a single sequence, where signal x_1 is a chaotic one and x_2 is a random one, or two different chaotic sequences. Examples of two combined normalized sequences are plotted in Figs. 4(a) and 4(b).

In both figures, we plot the results of applying the segmentation procedure for different combinations of signals. The aJSD achieves its maximum value exactly at the merging point of the two sequences, which is marked with a dotted vertical line. Similar results were observed in the case of sequences generated by chaotic processes. In these cases, the aJSD value reaches several orders of magnitude higher than those corresponding to a single stationary sequence.

It is of interest to see the robustness of the aJSD divergence for the detection of dynamic changes, under different noise contents of the signals. With this objective, we use two chaotic maps (tent map and Pincher map) with different levels of noise. $NSR=0\%$, 1%, 2%, 5%, and 10% noise level. Figure 5 shows the behavior of the aJSD for different noise contents in two chaotic maps. The parameters of the discretization are $d=6$ and $\tau=1$; the width of the windows is $L=10000$. From these results we can conclude that the methods distinguish between the two dynamics independent of the level noise, but when the noise level increases the aJSD values decrease.

VI. APPLICATION OF THE AJSD TO REAL DATA

Now, we test our proposed schemes for a group of real world signals. The first one is a set of electrocardiograms (ECG) signals on which we calculated the distance matrix of the aJSD between groups of patients with different cardiac pathologies. In the second application, we use our methods to detect the misalignment of the axis of an electric motor.

A. Distinction between groups of patients with heart diseases

In this study case, the signals correspond to the beat to beat intervals (BBI) registered from 15 patients. The patients were grouped into 3 sets of 5 patients each. The first group consists of healthy persons with a normal sinus rhythm (NSR); the second set contains patients suffering from congestive heart failure (CHF), and the third one is composed of patients suffering from atrial fibrillation (AF) data available in Ref. 51. An example of each ECG signal is shown in Fig. 6(a).

Each series represents a record of 24 hours (approximately 100,000 intervals). The analysed records do not have any previous filter. Figure 6 plots the distance matrix for the aJSD among the three sets, for the parameters ($d=8$ and $\tau=1$). The aJSD can discriminate between the members of

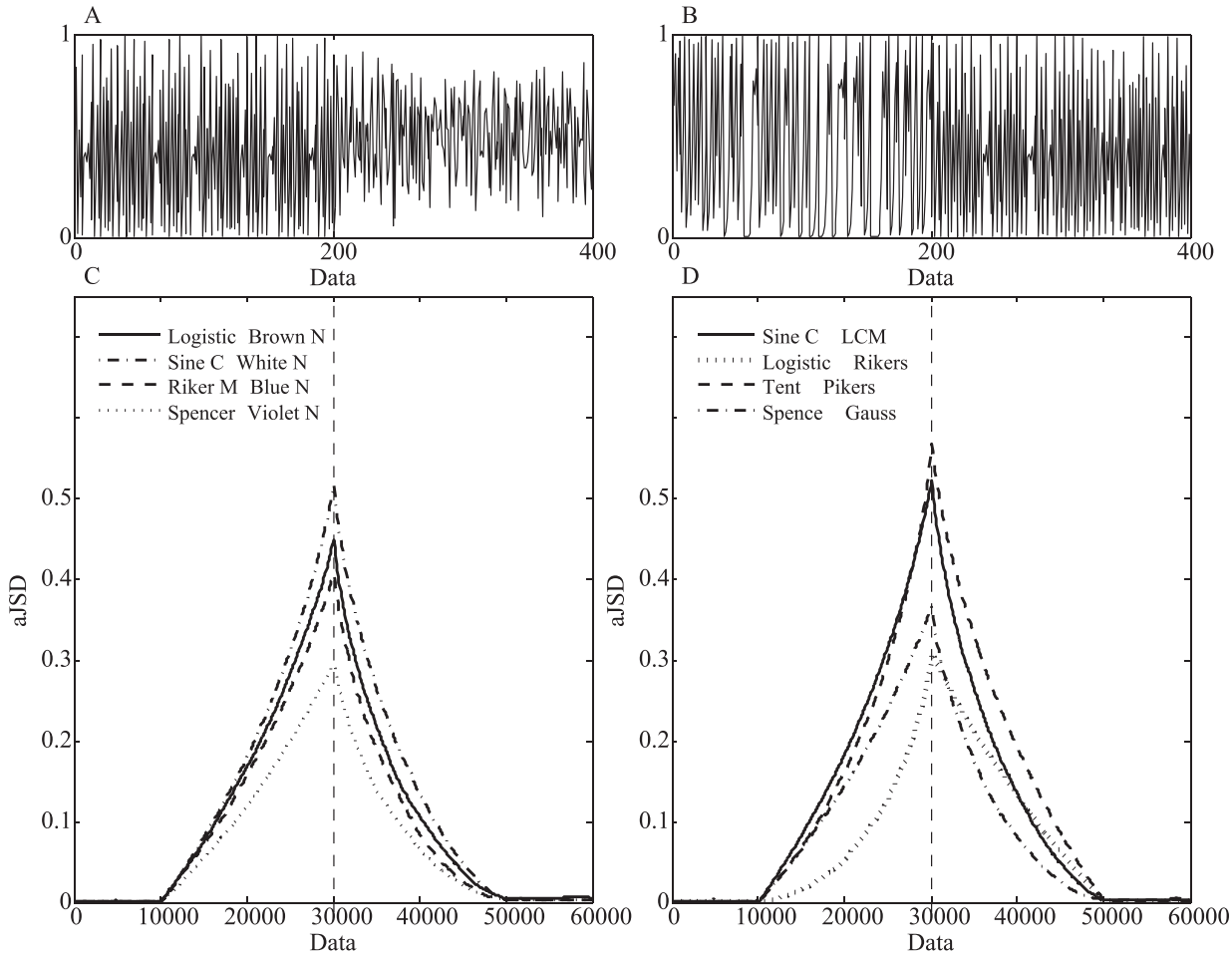

FIG. 4. (a) Example of a combined chaos-noise signal (Rikers maps—Blue noise). The signals were merged at the meddle position (point = 200). (b) idem A but using a chaos-chaos combined signals (Sine Circle–Linear congruential Map). (c) aJSD calculation using the running window method, applied to four composed chaotic and random sequences. The values chosen for the parameters d, τ, and Δ are: $d = 8$, $\tau = 1$, and $\Delta = 40\,000$. (d) Idem C applied to four composed different chaotic signals. In both plots, the maximum aJSD value is reached at the point where the two different sequences were merged (dotted vertical line).

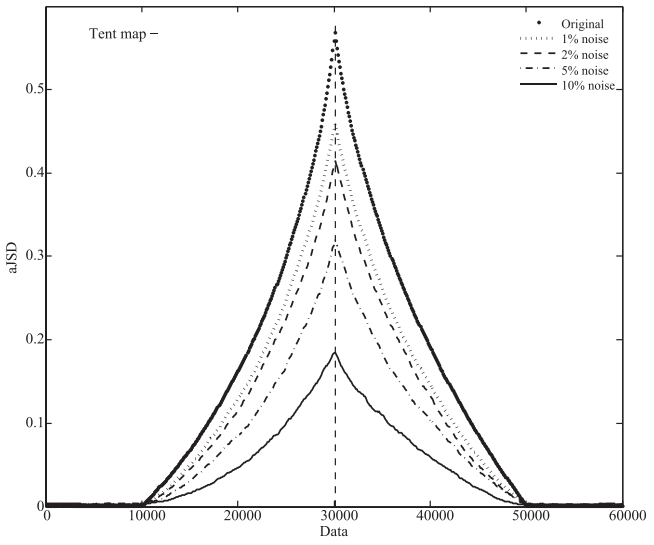

FIG. 5. The aJSD running windows method applied to the combination of two chaotic maps (Tent map and Riker's map) with different noise levels: $NSR : 0\%, 1\%, 2\%, 5\%, 10\%$. As the noise level increases, the aDJS values decrease. However the merging point of two signals is clearly detectable for all noise levels. The parameters used are: windows length $L = 10\,000$, $d = 6$, and $\tau = 1$.

the control group and the pathological groups, with the distance between the members of the pathological groups and the members of the control group being greater. The frequency time between heartbeats for the patients belonging to the control group, remains virtually unchanged, while for patients with pathologies the interbeat time is altered. The difference between heartbeats is captured through alphabetical mapping, giving different pattern distributions of $W^{d,\tau}$ for each group.

This example shows that our scheme could be of interest from a clinical point of view. It could be an adequate tool for controlling the evolution of patients under some kind of medical treatment.

B. Misalignment detection of an electric motor

The second example corresponds to a record taken from the axis movement of an electric motor. The first signal corresponds to the vibration measure using a capacitive accelerometer, mounted along the axis on the motor bearing [Fig. 7(a)]. The second one is a signal taken from an optical incremental rotary encoder drum, with about 145 pulses per revolution, as shown in Fig. 7(b). The third is a signal generated by a piezoelectric accelerometer mounted in the same place

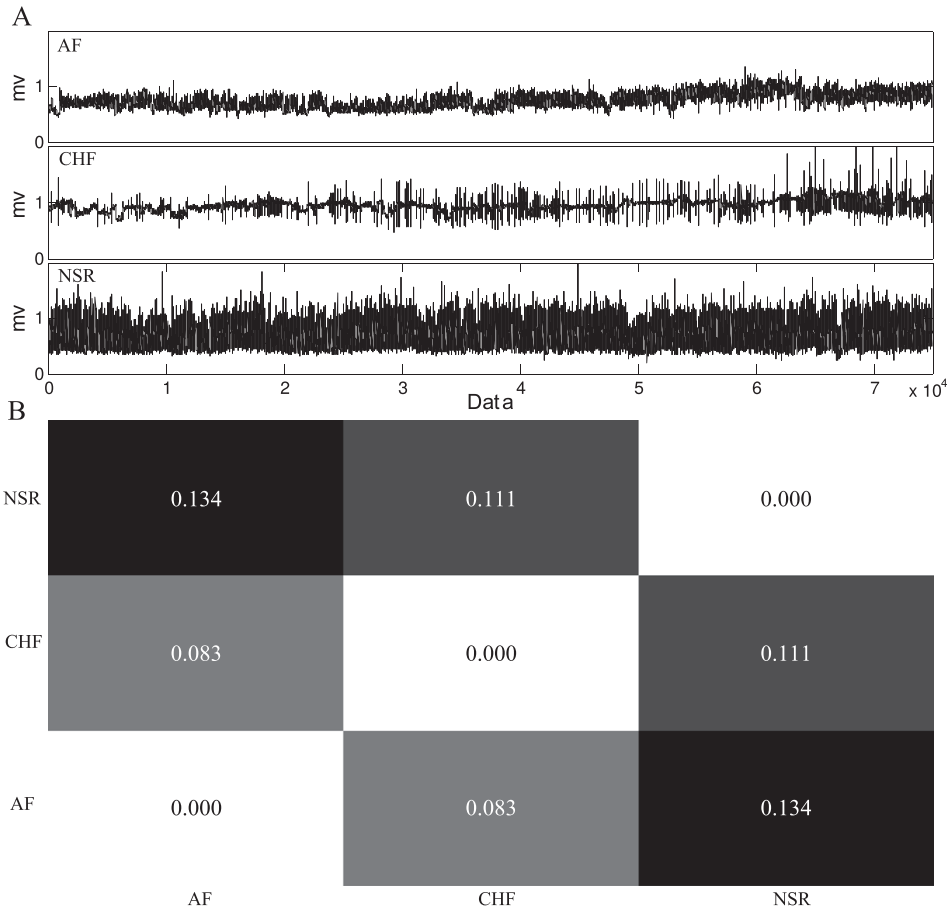

FIG. 6. (a) ECG signals belonging to the different study groups: the normal sinus rhythm (NSR), the congestive heart failure (CHF), and the atrial fibrillation (AF). (b) aJSD distance matrix for patients belonging to each group. The aJSD can discriminate between these three groups. The parameters used in this matrix are $d = 8$ and $\tau = 1$. The same results were obtained for higher embedding dimensions ($8 \leq d \leq 12$).

like the capacitive signal, which is plotted in Fig. 7(c). The last signal is the rate of the engine load and is depicted in Fig. 7(d) (It is related to the frequency of rotation of the shaft by means of a "Keyphaser."). The data were obtained with a sampling frequency of 25×10^3 Hz without any preprocessing. For more technical details on the recording setup, see Ref. 52.

Each signal has a length of $N = N_a + N_m = 14 \times 10^4$, in which the first $N_a = 7 \times 10^4$ measurement values correspond to the axis in the aligned position and $N_m = 7 \times 10^4$ in the misaligned position. For all the signals we used the following parameters: $d = 8$, $\tau = 1$, and $\Delta = 4 \times 10^4$. The results are shown in Fig. 7. It should be noted that for all signals, the aJSD maximum value is reached exactly when the shaft alignment state changes. That point is identified by the dotted vertical line. In the case of the signals from the vibration measurement by the piezoelectric accelerometers, the value of the maximum aJSD is smaller and more volatile than the other two methods. This fact can be associated with the efficacy of the measurement method. In all cases, the different signals are clearly detectable by the methods proposed here.

VII. DISCUSSION

In this paper, we have introduced the alphabetic JSD as a measure of distance between time series. Using this measure we have developed two methods to distinguish between different types of signals. Our schemes can be applied in two dimensional signals.

Using this measure, we have developed two methods to distinguish between different types of signals. Particularly, we used them to distinguish random from chaotic signals, and to compare different types of chaotic signals between them. Through the aJSD distance matrix, we were able to show that chaotic signals are clearly distinguishable from random signals with a diverse power spectrum. It is important to note that chaotic signals with different phase space structures are distinguishable from one another.

The results were presented in two ways. The first one by evaluating the distance matrix corresponding to different pairs of signals, i.e., chaotic-noisy and chaotic-chaotic signals; and the second one by plotting the values of the aJSD in terms of the position of a moving window along the signal.

We also developed a procedure to detect dynamical changes in a signal by using a sliding window that moves on the resulting symbolic sequence after mapping the original series by an alphabetical assignment, being the centre of the original window moved one place per step. Then the aJSD is evaluated for the two corresponding sub-sequences belonging to each half of the window. The maximum value of the divergence corresponds to the point where a change occurs in the probability distribution of the discretized signal. We tested the method using two different signals merged together. The aJSD allows detecting the point where the two different signals were coupled in order to constitute signals with chaos-noise and chaos-chaos parts. Finally, we use this scheme to evaluate the alignment condition of the axis of an electric motor.

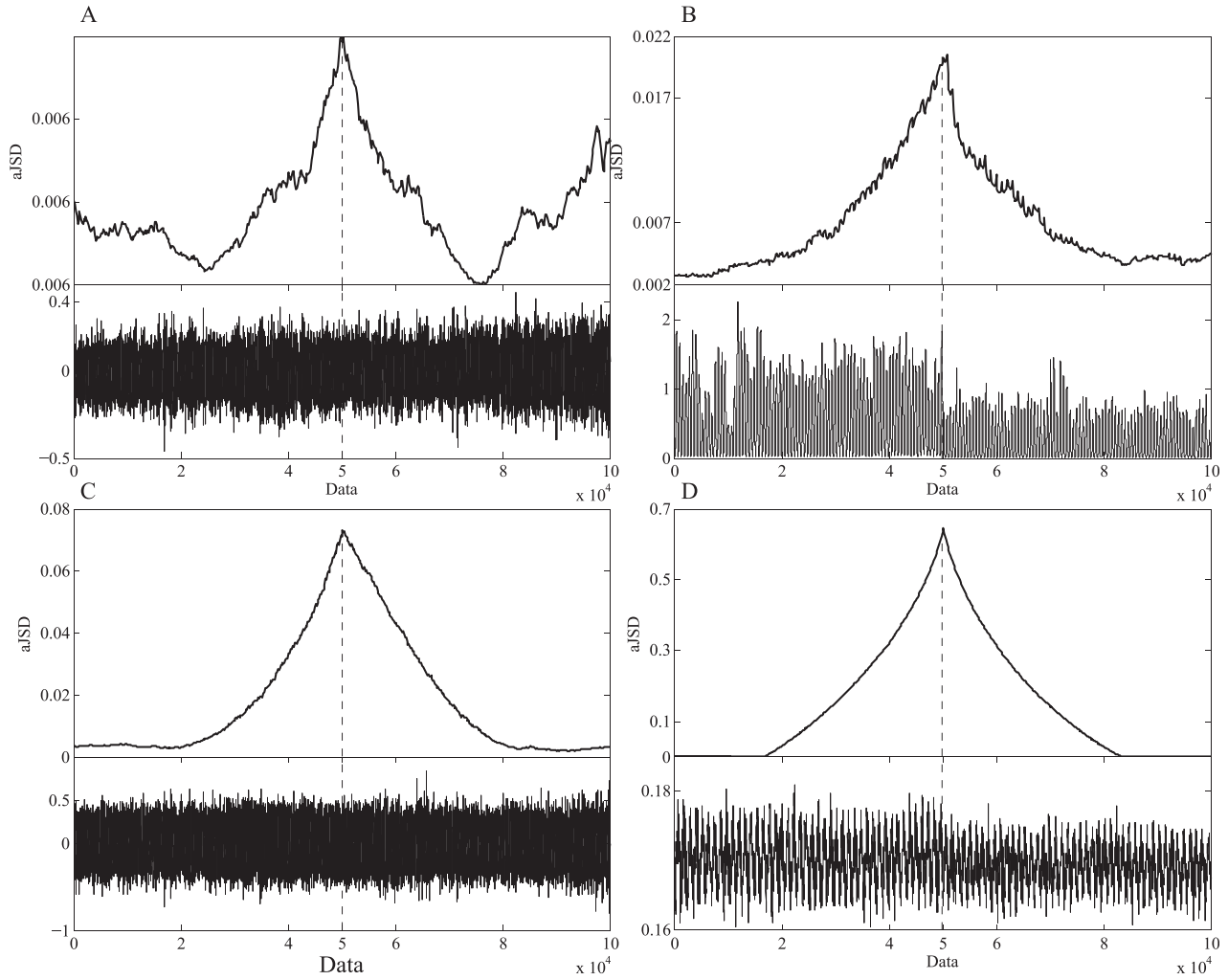

FIG. 7. aJSD analysis using a sliding windows, corresponding to the four different signals taken on the axis of an electric motor. (a) Signal of a capacitive accelerometer mounted on the z axis of the motor bearing. (b) An optical signal incremental rotary encoder drum with 145 pulses per revolution. (c) Signal generated by a piezoelectric accelerometer mounted on the same place as A. (d) Signal engine load rate. All signals were analysed choosing the parameters $d = 8$, $\tau = 1$, $\Delta = 40\,000$.

The methods were tested in real world signals, showing to be robust, and suitable for the detection of different regimes, both in electrophysiological records and in the alignment of the axis of an electric motor. These two applications show that the proposed schemes could be useful in the context of clinical treatment and other practical situations. Concerning the inner noise present in real signals, we deal with this problem using the significance criterion explained in Sec. III. This criterion is capable of discerning between differences of probabilities of each dynamics and differences that are produced by the noise or statistical fluctuations. Finally, it is worth mentioning that our schemes could be implemented for a real time analysis of the signal under study.

Is important to note that one of the weak points of the proposed methods is that they do not allow detecting changes in the signal amplitude. This occurs due to the discretization procedures required for the evaluation of the JSD along the signals. For example, the signal $X_1 = \{12321312\}$ and the signal $X_2 = \{1020102010201020\}$ lead to the same binarization $S_1 = S_2 \equiv 1100101$. So the methods are not suitable to study changes in the amplitude of the signals. On the other hand, if the signals are very similar, a huge embedding size and therefore very long signals are required, in order that the proposed methods are suitable for detecting dynamical changes in the signals. However, one of the main advantages of our schemes is the relatively low computational cost they have. In that sense we think that they might be adequate to the development of real time schemes of analysis.

ACKNOWLEDGMENTS

The authors thank Dr. Pierre Granjon from the GIPSA-lab, Grenoble, France, for providing the recording electric motor data. We thank the CONICET and the Secyt-UNC for the financial assistance.

[1]H. Wold, *A Study in the Analysis of Stationary Time Series* (Almqvist & Wiksell, 1938).
[2]M. Cencini, M. Falcioni, E. Olbrich, H. Kantz, and A. Vulpiani, "Chaos or noise: Difficulties of a distinction," Phys. Rev. E **62**(1), 427 (2000).
[3]H. Sakai and H. Tokumaru, "Autocorrelations of a certain chaos," IEEE Trans. Acoust. Speech Signal Process. **28**(5), 588–590 (1980).
[4]O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, "Distinguishing noise from chaos," Phys. Rev. Lett. **99**(15), 154102 (2007).

[5]L. Zunino, M. C. Soriano, and O. A. Rosso, "Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach," Phys. Rev. E **86**(4), 046210 (2012).
[6]F. Olivares, A. Plastino, and O. A. Rosso, "Contrasting chaos with noise via local versus global information quantifiers," Phys. Lett. A **376**(19), 1577–1583 (2012).
[7]W. A. Brock, "Distinguishing random and deterministic systems: Abridged version," J. Econ. Theory **40**(1), 168–195 (1986).
[8]J. B. Gao, J. Hu, W. W. Tung, and Y. H. Cao, "Distinguishing chaos from noise by scale-dependent Lyapunov exponent," Phys. Rev. E **74**(6), 066204 (2006).
[9]J. B. Elsner, "Predicting time series using a neural network as a method of distinguishing chaos from noise," J. Phys. A: Math. General **25**(4), 843 (1992).
[10]Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," Proc. R. Soc. London A: Math. Phys. Eng. Sci. **460**, 1597–1611 (2004).
[11]T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
[12]K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak, "Construction of symbolic dynamics from experimental time series," Phys. Rev. Lett. **82**(6), 1144 (1999).
[13]G. E. Powell and I. C. Percival, "A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems," J. Phys. A **12**(11), 2053 (1979).
[14]O. A. Rosso, H. Craig, and P. Moscato, "Shakespeare and other english renaissance authors as characterized by information theory complexity quantifiers," Phys. A: Stat. Mech. Appl. **388**(6), 916–926 (2009).
[15]O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," J. Neurosci. Methods **105**(1), 65–75 (2001).
[16]C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," Phys. Rev. Lett. **88**, 174102 (2002).
[17]A. C. Yang, S. Hseu, H. Yien, A. L. Goldberger, and C. Peng, "Linguistic analysis of the human heartbeat using frequency and rank order statistics," Phys. Rev. Lett. **90**(10), 108103 (2003).
[18]J. Burbera and C. R. Rao, "Entropy, differential metric, distance and divergence measures in probability spaces: a unified approach," J. Multivariate Statist. **12**, 537–596 (1982).
[19]J. Lin, "Divergence measures based on the shannon entropy," Inf. Theory, IEEE Trans. **37**(1), 145–151 (1991).
[20]S. López-Rosa, J. Antolín, J. C. Angulo, and R. O. Esquivel, "Divergence analysis of atomic ionization processes and isoelectronic series," Phys. Rev. A **80**(1), 012505 (2009).
[21]J. C. Angulo, J. Antolín, S. López-Rosa, and R. O. Esquivel, "Jensen–Shannon divergence in conjugate spaces: The entropy excess of atomic systems and sets with respect to their constituents," Phys. A: Stat. Mech. Appl. **389**(4), 899–907 (2010).
[22]J. Antolín, J. C. Angulo, and S. López-Rosa, "Fisher and jensen–shannon divergences: Quantitative comparisons among distributions: Application to position and momentum atomic densities," J. Chem. Phys. **130**(7), 074110 (2009).
[23]A. L. Martín, S. López-Rosa, J. C. Angulo, and J. Antolín, "Jensen-Shannon and Kullback–Leibler divergences as quantifiers of relativistic effects in neutral atoms," Chem. Phys. Lett. **635**, 75–79 (2015).
[24]I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley, "Analysis of symbolic sequences using the Jensen-Shannon divergence," Phys. Rev. E **65**(4), 041905 (2002).
[25]D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," IEEE Trans. Inf. Theory **49**(7), 1858–1860 (2003).
[26]H.-V. Nguyen and J. Vreeken, "Non-parametric Jensen-Shannon divergence," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer, 2015), pp. 173–189.
[27]J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, "Nonparametric entropy estimation: An overview," Int. J. Math. Stat. Sci. **6**(1), 17–39 (1997).
[28]J. C. Robinson, *Dimensions, Embeddings, and Attractors* (Cambridge University Press, 2011), Vol. 186.
[29]F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, 1981), pp. 366–381.
[30]S. Zozor, D. Mateos, and P. W. Lamberti, "Mixing bandt-pompe and lempel-ziv approaches: Another way to analyze the complexity of continuous-state sequences," Eur. Phys. J. B **87**(5), 107 (2014).
[31]J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Oxford, 2003), Vol. 69.
[32]D. E. Knuth, The Art of Computer Programming, Vol. 3, 2nd ed. (Adison-Wesley, 1998).
[33]W. H. Steeb and M. A. Van Wy, *Chaos and Fractals: Algorithms and Computations* (Wissenschaftsverlag, 1992).
[34]R. M. May, "Simple mathematical models with very complicated dynamics," Nature **261**(5560), 459–467 (1976).
[35]A. Potapov and M. K. Ali, "Robust chaos in neural networks," Phys. Lett. A **277**(6), 310–322 (2000).
[36]W. E. Ricker, "Stock and recruitment," J. Fish. Board Can. **11**(5), 559–623 (1954).
[37]V. I. Arnol'd, "Small denominators. I. Mapping the circle onto itself," Izv. Ross. Akad. Nauk. Ser. Mat. **25**(1), 21–86 (1961).
[38]S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology and Chemistry* (Perseus Publishing, 2001).
[39]R. Shaw, Strange Attractors, "Chaotic Behavior and Information Flow," Z. Naturforsch. **38a**, 80–112 (1981).
[40]M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos* (Academic Press, 2004), Vol. 60.
[41]M. Hénon, "A two-dimensional mapping with a strange attractor," Commun. Math. Phys. **50**(1), 69–77 (1976).
[42]D. G. Aronson, M. A. Chory, G. R. Hall, and R. P. McGehee, "Bifurcations from an invariant circle for two-parameter families of maps of the plane: A computer-assisted study," Commun. Math. Phys. **83**(3), 303–354 (1982).
[43]H. E. Nusse, *Dynamics: Numerical Explorations* (Springer, 1998), Vol. 101.
[44]G. Schmidt and B. W. Wang, "Dissipative standard map," Phys. Rev. A **32**(5), 2994 (1985).
[45]V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, 1968), Vol. 9.
[46]A. A. Chernikov, R. Z. Sagdeev, and G. M. Zaslavskii, "Chaos-how regular can it be?," Phys. Today **41**(11), 27–35 (1988).
[47]B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," Phys. Rep. **52**(5), 263–379 (1979).
[48]R. L. Devaney, "A piecewise linear model for the zones of instability of an area-preserving map," Phys. D: Nonlinear Phenom. **10**(3), 387–393 (1984).
[49]H. A. Larrondo, http://www.mathworks.com/matlabcentral/fileexchange/35381 for program: noisef k.m (2012).
[50]M. O. Rabin, "Probabilistic algorithm for testing primality," J. Number Theory **12**(1), 128–138 (1980).
[51]Physionet, "Is the normal heart rate chaotic?," https://www.physionet.org/challenge/chaos/ for physionet database (2008).
[52]M. Blödt, P. Granjon, B. Raison, and J. Regnier, "Mechanical fault detection in induction motor drives through stator current monitoring-theory and application examples," Fault Detect. W. Zhang (Ed), 451–488 (2010).