Applied Mathematics and Computation 362 (2019) 124554

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Learning and distinguishing time series dynamics via ordinal patterns transition graphs

João B. Borges[a,b,*], Heitor S. Ramos[b], Raquel A.F. Mini[c], Osvaldo A. Rosso[d,e], Alejandro C. Frery[f], Antonio A.F. Loureiro[b]

[a] Department of Computing and Technology, Universidade Federal do Rio Grande do Norte, Caicó, RN 59300-000, Brazil
[b] Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG 30123-970, Brazil
[c] Department of Computer Science, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, MG, 30535-610, Brazil
[d] Institute of Physics, Universidade Federal de Alagoas, Maceió, AL 57072-900, Brazil
[e] Institute of Translational Medicine and Biomedical Engineering, University Institute of the Italian Hospital, Hospital Italiano de Buenos Aires & Conicet, Ciudad Autónoma de Buenos Aires, Argentina
[f] Laboratory of Scientific Computation and Numerical Analysis, Universidade Federal de Alagoas, Maceió, AL 57072-970, Brazil

ARTICLE INFO

Keywords:
Time series dynamics
Chaos
Randomness
Time series characterization
Time series classification
Bandt-Pompe transformation

ABSTRACT

Strategies based on the extraction of measures from ordinal patterns transformation, such as probability distributions and transition graphs, have reached relevant advancements in distinguishing different time series dynamics. However, the reliability of such measures depends on the appropriate selection of parameters and the need for large time series. In this paper we present a method for the characterization of distinct time series behaviors based on the probability of self-transitions, a measure extracted from their transformation onto ordinal patterns transition graphs. We validate our method by investigating the main characteristics of periodic, random, and chaotic time series. By the application of learning strategies, we precisely classify different randomness levels in time series, reaching 100% in accuracy, and advances in performing the hard task of distinguishing random noises from chaotic time series, correctly distinguishing 96.61% of the cases. Furthermore, we show that this strategy is well suitable to be used by many applications, even for short time series, and does not depend on the selection of parameters.

1. Introduction

An essential initial step for any learning strategy is to properly understand the data. When dealing with time ordered data, an important aspect to be concerned is its dynamics. The knowledge on how a system evolves over time may lead to the development of better solutions, by understanding the underlying data generating process [1,2]. However, time series dynamics has been the focus of many studies, which propose to solve theoretic problems in areas such as physics and statistics. Moreover, there are few studies of using dynamical aspects of time series for knowledge discovery, which is the basis for several data mining strategies (e.g., clustering, classification, anomaly detection).

The characterization of time series by its dynamics has achieved important advances in recent investigations, mainly guided by their transformation onto ordinal patterns (OP) [3–7]. Following the cornerstone contributions of Bandt and

* Corresponding author at: Department of Computing and Technology, Universidade Federal do Rio Grande do Norte, Caicó, RN 59300-000, Brazil.
E-mail addresses: joaoborges@dct.ufrn.br (J.B. Borges), ramosh@dcc.ufmg.br (H.S. Ramos), raquelmini@pucminas.br (R.A.F. Mini), oarosso@gmail.com (O.A. Rosso), acfrery@laccan.ufal.br (A.C. Frery), loureiro@dcc.ufmg.br (A.A.F. Loureiro).

https://doi.org/10.1016/j.amc.2019.06.068
0096-3003/

Pompe [8], these proposals are based on the transformation of a time series into a set of symbolic patterns. They are obtained from the ordinal relation between successive data samples, which is based on two parameters: the embedding dimension D and the embedding delay τ. This transformation captures the ordinal structure of the series, thus the ordinal name, which is related to the dynamics of their temporal correlation [3,9,10].

The feasibility of such characterization has already been examined and proven [3,10]. The main strategy involves the construction of probability distributions from the frequency of patterns, the extraction of information theory quantifiers (normalized permutation entropy and statistical complexity), and the construction of a plane with these measures. It was observed that time series with different dynamics refer to different regions of the plane, which allow to distinguish different time series dynamics, such as deterministic, noisy and chaos. Thus, this method is able to differentiate most of these dynamics, but there are some challenging regions with no clear separability between those time series within them, leading to misclassifications. For instance, there are some chaotic maps that could be easily confounded as random noises if one considers only their placements in the plane.

Another challenge when dealing with strategies based on OP transformations is related to their dependence on the proper choice of parameters D and τ. For instance, to assure the reliability of measures extracted from this transformation, the length n of time series must be long enough so the sampling in the $D!$ space of patterns is representative, i.e., $n \gg D!$ [3]. On the other hand, there is still no clear definition regarding the choice of the embedding delay τ. Also it is known that the most suitable τ is domain specific, and inappropriate values may obscure important characteristics of the phenomenon under analysis [11]. A common choice is $\tau = 1$, but some studies follow different strategies, such as the first zero of the autocorrelation of the time series [12].

Recent approaches to obtain some knowledge about the time series dynamics consider the creation of graphs based on the observed ordinal patterns of a given time series [12–15]. These graphs are constructed after the transformation of a time series onto the set of OP, taking into account the transitions between consecutive patterns. Each $D!$ possible OP is a vertex in the graph and a directed edge connects two OPs in the graph if they appear sequentially in the time series. Each edge represents the transition between patterns, thus the name "ordinal patterns transition graphs", denoted here as G_π.

The study of time series via their transformation into graphs is a very successful strategy [16–24]. Some notable examples are the visibility graph (VG) [21,23,24] and horizontal visibility graph (HVG) [22]. For these approaches, each point in the data series is a vertex in the graph, and two vertices are connected by an edge if they satisfy the visibility criterion, i.e., it is possible to trace a line between two data points without intersecting intermediate points or, more strictly, a horizontal line for the HVG. However, the visibility approaches may not scale depending on the time series length, and, thus, the G_π is the best of both worlds.

The analysis of these graphs are often performed on their structure, by accounting both graph measures and information quantifiers, which may require to create graphs with a large enough number of vertices. For the case of G_π, the number of vertices depends on D. Thus, the reliability of such metrics may suffer from the same inherited problems of the underlying OP transformations. For instance, the range $6 \leq D \leq 10$ may invalidate the strategy for small time series [12].

In this study, we present a novel method for the characterization of time series based on the probability of self-transitions (p_{st}), a measure extracted from the graph G_π. The main contributions of this work are the following: (i) we show that the p_{st} measure is directly related to the temporal correlation of time series, which is a valuable indicative of their underlying dynamics and adequate to their proper characterization; (ii) we advance the state of the art by providing a precise distinction among different time series dynamics within those challenging regions in the plane, as discussed above, even for short time series. In our experiments, we use time series with length $n = 5 \times 10^3$, being necessary small values of D, in contrast to previous research results that considered lengths from $2^{15} (\approx 3.2 \times 10^4)$ to 10^7, i.e., one to four orders of magnitude of difference; and (iii) we reduce the dependence of the ordinal patterns transformation on the selection of the parameters. We assume small values of D and, in order to be independent on the choice of τ, we propose evaluating the evolution of their behavior for different time scales. Furthermore, this strategy also aims to contribute in motivating the community towards a broaden applicability of the ordinal patterns transformation, being possible to be used for several different domains, such as Medical Sciences, Engineering, Mathematics, Computing and Physics, where time series representing the dynamics of a system must be characterized and better understood.

The rest of this paper is organized as follows. Section 2 describes the background necessary for transforming a given time series onto ordinal patterns and to compute metrics that can be used for understanding its dynamics. We also discuss some relevant challenges and limitations of the Bandt-Pompe method, which serves as motivation for our problem definition. Section 3 details the method for constructing the ordinal patterns transition graph, presenting its main properties and describing the intuition behind the probability of self-transition (p_{st}). Section 4 presents our results for learning and distinguishing different time series dynamics. We validate our method by investigating the main characteristics of periodic and random time series, and by applying our strategy to perform the hard task of distinguishing between random noises and chaotic time series. Finally, Section 5 discusses our final remarks for this work and presents some future directions.

2. Background and problem definition

The transformation of time series data onto a set of ordinal patterns, which is the basis for the transition graph, is obtained by the ordinal relation between successive data samples. It is based on two parameters: the embedding dimension

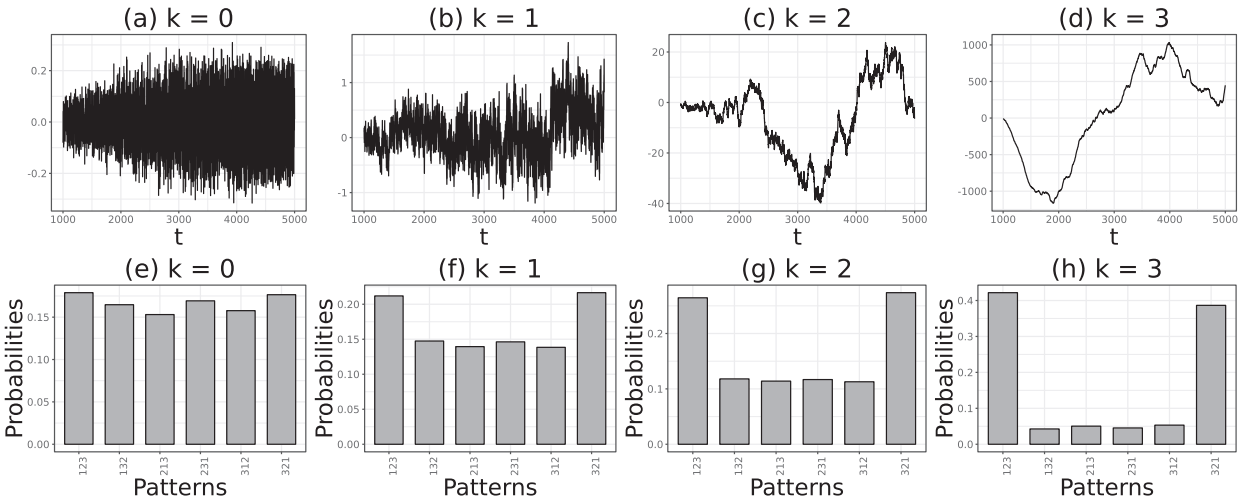

Fig. 1. Illustration of ordinal patterns probability distributions constructed from random time series with different correlation levels. Where (a) is a white noise ($k = 0$), (b) pink noise ($k = 1$), (c) red/brown noise ($k = 2$), and (d) a black noise ($k = 3$). Their corespondent ordinal patterns probability distributions are presented in (e), (f), (g), and (h), respectively. Each time series was generated with 5,000 samples in length, and the OP transformations were computed with $D = 3$ and $\tau = 1$.

D and the embedding delay τ, which maps the samples onto the space of $D!$ possible patterns, by computing sequential or τ-spaced sliding windows of size D. These transformations are detailed below.

2.1. Ordinal patterns transformation

For a time series $\mathbf{x} = \{x_1, \ldots, x_n\}$ of length n and an embedding dimension $D \in \mathbb{N}$, the Bandt-Pompe method [8] requires the generation of sliding windows $\mathbf{w}_t \subseteq \mathbf{x}$ of size D. For each instant $t = 1, \ldots, n-(D-1)$, $\mathbf{w}_t = \{x_t, x_{t+1}, \ldots, x_{t+(D-2)}, x_{t+(D-1)}\}$. The ordinal relation [10] for each instant t consists of the permutation $\pi = (r_0, r_1, \ldots, r_{D-1})$ of $(0, 1, \ldots, D-1)$, such that $x_{t-r_{D-1}} \leq x_{t-r_{D-2}} \leq \cdots \leq x_{t-r_1} \leq x_{t-r_0}$. In other words, π represents the necessary permutation for the elements of $\mathbf{w}_t$ to be sorted in ascending order.

A proposal that advances this method considers the insertion of an embedding delay $\tau \in \mathbb{N}$ [6], such that the elements within each sliding window are spaced by intervals of size τ, corresponding to a sample of the time series by regular spaced intervals [11]. Thus, the sliding windows for each instant t are defined as $\mathbf{w}_t = \{x_t, x_{t+\tau}, \ldots, x_{t+(D-2)\tau}, x_{t+(D-1)\tau}\}$.

The method originally proposed by Bandt and Pompe [8] is equivalent to the case where $\tau = 1$, with sliding windows sampled by consecutive data points. After the transformation, the time series is converted into a set of ordinal patterns $\Pi = \{\pi_i : i = 1, \ldots, m\}$ where $m = n - (D-1)\tau$ and each π_i represents a permutation from the set of $D!$ possible permutations.

2.2. Ordinal patterns probability distribution

Given the set Π of ordinal patterns, the ordinal patterns probability distribution p_π consists of assigning a probability distribution to the permutations observed in the time series. Thus, for each possible permutation $\pi_t \in \Pi$, with $t \in \{1, \ldots, D!\}$, let $|s_{\pi_t}| \in \{0, \ldots, m\}$ be the number of observed patterns of type π_t, and the ordinal patterns probability distribution $p_\pi = \{p(\pi_t)\} : \forall t \in \{1, \ldots, D!\}$ be defined as

$$p(\pi_t) = \frac{|s_{\pi_t}|}{n - (D-1)\tau}, \tag{1}$$

satisfying the conditions $p(\pi_t) \geq 0$ and $\sum_{\pi_t} p(\pi_t) = 1$.

Fig. 1 gives examples of ordinal patterns probability distributions constructed from different time series. Fig. 1a–d illustrate random time series with different correlation levels between their points. Those time series were synthetically generated according to their power spectra f^{-k} [23], where (a) is a white noise ($k = 0$), (b) pink noise ($k = 1$), (c) red/brown noise ($k = 2$), and (d) a black noise ($k = 3$). Their respective ordinal patterns probability distributions where constructed with $D = 3$ and $\tau = 1$, and are presented in Fig. 1e–h. For this example, the frequency of patterns 123 and 321 are higher as the correlation levels increases in the series, indicating the more regular behavior [10]. A suitable strategy to account for different time series dynamics is possible by computing information theory quantifiers from p_π, as presented in next sections.

2.3. Metrics from the ordinal patterns distribution

Given the ordinal patterns distribution of a time series, some metrics could be extracted for its analysis. Such metrics can highlight important aspects from the time series, which are not readily available from the raw data and can be used to better distinguishing them.

2.3.1. Permutation entropy

Following the initial purpose of the Bandt-Pompe method [8], the main extracted metrics are information quantifiers calculated from the distribution p_π, for all $D!$ permutations π of order D. Notice that p_π depends on the permutations of consecutive data points.

In this work, we propose a variation of the classical Shannon entropy, called *permutation entropy*, and defined as

$$H[p_\pi] = -\sum p(\pi)\log p(\pi), \tag{2}$$

where $0 \le H[p_\pi] \le \log D!$. The permutation entropy is equivalent to the Shannon entropy and is a measure of uncertainty associated to the process described by p_π [6]. Lower values of $H[p_\pi]$ represent an increasing or decreasing sequence of values in the permutation distribution, indicating that the original time series is deterministic. On the other side, high values of $H[p_\pi]$ indicate a completely random system [8].

2.3.2. Normalized permutation entropy

The maximum value for $H[p_\pi]$ occurs when all $D!$ possible permutations have the same probability to occur, which is the case for the uniform distribution p_u of permutations. Thus, $H_{\max} = H[p_u] = \log D!$, where $p_u = \{1/D!, \ldots, 1/D!\}$ [11]. The normalized Shannon entropy can defined [10], based on the permutation entropy case, as

$$H_S[p_\pi] = \frac{H[p_\pi]}{H_{\max}}, \tag{3}$$

where $0 \le H_S[p_\pi] \le 1$.

2.3.3. Statistical complexity

Another statistical measure that can be computed from the permutation distribution is the statistical complexity. This measure [25] is another point of view concerning the knowledge of some underlying process, based on the Jensen-Shannon divergence *JS* between the associated probability distribution p_π and the uniform distribution p_u, i.e., the trivial case for the minimum knowledge from the process. The statistical complexity is given by

$$C_{JS}[p_\pi] = Q_{JS}[p_\pi, p_u]H_S[p_\pi], \tag{4}$$

where $p_\pi = \{p(\pi)\}$ is the ordinal patterns probability distribution, p_u is the uniform distribution, and $H_S[p_\pi]$ is the normalized Shannon entropy, as described above.

The disequilibrium $Q_{JS}[p_\pi, p_u]$ is given by

$$Q_{JS}[p_\pi, p_u] = Q_0 JS[p_\pi, p_u] = Q_0\left\{S\left[\frac{p_\pi + p_u}{2}\right] - \frac{S[p_\pi] + S[p_u]}{2}\right\}, \tag{5}$$

where S is the Shannon entropy measure and Q_0, given by

$$Q_0 = -2\left\{\left(\frac{D!+1}{D!}\right)\ln(D!+1) - 2\ln(2D!) + \ln(D!)\right\}^{-1}, \tag{6}$$

i.e., it is a normalization constant equals to the inverse of the maximum value of $JS[p_\pi, p_u]$, so $0 \le Q_{JS} \le 1$ [6,26].

2.4. Problem definition: limitations of Bandt-Pompe's method

The effectiveness of using ordinal patterns for the characterization of time series has already been largely discussed by many recent studies [3–7,10]. These are strong results that have driven extensive discussions in the literature, but they are confined to solve theoretic problems in areas such as physics and statistics. When trying to apply those strategies to a broad domain, there are some issues that still hinder the achievement of similar success. In the following, we conjecture that specific aspects of Bandt-Pompe's method and the current methods used to extract time series behaviors are responsible for adding limitations to a wide applicability.

2.4.1. Strong dependence on parameters definition and time series length

Strategies based on OP transformations depend on the proper choice of parameters D and τ. For instance, to assure the reliability of measures, the length n of the time series must be long enough so the sampling in the $D!$ space of patterns is representative, i.e., $n \gg D!$ [3,11]. For practical purposes, D is recommended to be within the interval $[3 \ldots 7]$ [8]. If this condition is not satisfied, it may lead to analysis of statistics that are not consistent to the real time series behavior and the occurrence of missing patterns [27]. Recall that the number of possible patterns grows as factorial of D. Thus, if the time

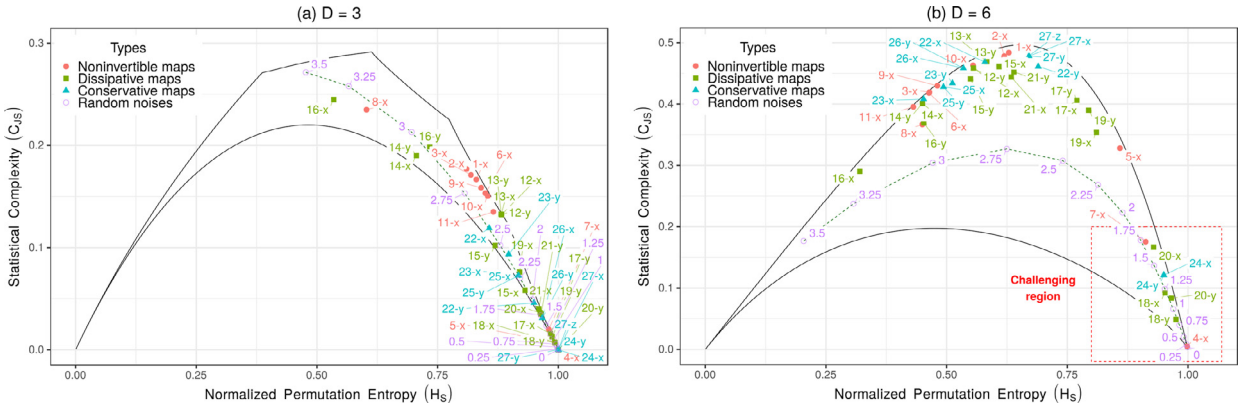

Fig. 2. Reproduction of the CCEP [3] of 27 chaotic maps and 15 random noises. Chaotic maps are illustrated by their notation "$n - d$", where n is the map number and d its dimension. Random noises are illustrated as sequences of open circles with dashed lines, with respective values k, defined by their power spectra f^{-k}. The measures were computed from series with length 10^5 each, after discarding the first 5000 samples, with $\tau = 1$, and (a) $D = 3$ and (b) $D = 6$. For random noises, each measure was obtained as the average of 10 runs with different seeds. Continuous lines represent the minimum and maximum limits for statistical complexities [26].

series are not long enough, the absence of some patterns may occur, but just due to its limited length n, not because these patterns do not occur within the time series behavior.

This effect also occurs on the OP-based graph transformations. For large values of D, the constructed graph may have a large number of vertices, up to $D!$. Thus, if the time series length is also not long enough, some of these vertices may be wrongly absent. However, for the current strategies based on OP transformations, the best results are achieved when large values of D are used. For instance, the range $6 \leq D \leq 10$ has been suggested [12] for the extraction of representative behaviors from the constructed graphs. This fact practically invalidates the application of those strategies for small time series, which represent a severe limitation for many real-world scenarios.

Another challenge when dealing with OP is the choice of the embedding delay τ. The most suitable τ is domain specific, and inappropriate values may obscure important characteristics of the phenomenon under analysis [11]. A common choice found in many studies is $\tau = 1$. Although some of them follow different strategies, such as using the first zero of the autocorrelation of the time series [12], there is still no closed strategy for choosing its value.

One of our main contributions with this work is to advance the applicability of Bandt-Pompe's method, being possible to be used even for short time series, in contrast to previous strategies. Moreover, the proposed method must present minimum dependency on the selection of parameters. It must consider small values of D and be independent of the choice of τ. We accomplish this last requirement by comparing the behavior of time series when evaluated as a function of the embedding delay τ, which is better described later on.

2.4.2. Clear separability for time series dynamics within CCEP

An intriguing aspect of time series arising from chaotic systems is that they share several properties with stochastic processes, making their distinction a hard task [1,3,7,10,11,23,28,29]. The causality complexity-entropy plane (CCEP) has been successfully applied [3,10] to effectively distinguish noise from chaos, by placing each type of series at different locations in the plane. This plane is formed by the normalized permutation entropy (H_S) and the statistical complexity (C_{JS}) measures, computed from the OP probability distribution and represented by the x-axis and y-axis of the plane, respectively. Fig. 2a and b show the CCEP for $D = 3$ and $D = 6$, respectively, in which we reproduce the scenario for the distinction of 27 chaotic maps and 15 colored random noises.

Altogether, we analyzed 59 time series: 15 random noises and 44 chaotic maps, considering their dimensions. The chaotic maps are arranged into three different groups (noninvertible, dissipative and conservative maps), and each chaotic map may have one or more dimensions. Table 1 presents the list of maps, with dimensions in curly brackets. The random noises were created to have different levels of correlation between points, defined by their power spectra f^{-k}. Each noisy time series was generated with $n = 5000$ samples and the OP transformation was computed as the average of 10 runs with different seeds.

Fig. 2a shows a version of the CCEP for $D = 3$, which is not sufficient for distinguishing the time series by positions in the plane. Instead, larger values of D (such as $D = 6$ in Fig. 2b) are able to contrast different placements for them. However, as the method requires larger values of D, it also requires longer time series. The time series length to compute this plane is $n = 10^5$.

Considering the case for $D = 6$, the colored noises have a well-defined placement in the plane, with intermediate values of C_{JS} and, for the entropy, there is an inverse relation between H_S and the correlation degree k. The random time series, with lower k, have larger entropy, while the more correlated noises, with higher k, present lower H_S. The general placement for the chaotic maps is close to the upper limits of C_{JS}, being straightforward to be distinguished from noises. However,

Table 1
Descriptions of the 27 chaotic maps used in the characterization [3]. They are arranged into three different groups, and each chaotic map may have one or more dimensions. Each time series arising for these chaotic maps dimensions are defined in text by the syntax "n-d", where $n = \{1, \ldots, 27\}$ represents the number of the chaotic map and $d = \{x, y, z\}$ its respective dimension. More details regarding the generation of these maps and their initial conditions can be found in the supplementary materials of [23].

#	Description	Dim.	Grp.	#	Description	Dim.	Grp.
1	Logistic map	{x}	Noninvertible	15	Tinkerbell map	{x,y}	Dissipative
2	Sine map	{x}		16	Burgers' map	{x,y}	
3	Tent map	{x}		17	Holmes cubic map	{x,y}	
4	Linear congruential generator	{x}		18	Dissipative standard map	{x,y}	
5	Cubic map	{x}		19	Ikeda map	{x,y}	
6	Ricker's population model	{x}		20	Sinai map	{x,y}	
7	Gauss map	{x}		21	Discrete predator-prey map	{x,y}	
8	Cusp map	{x}		22	Chirikov standard map	{x,y}	Conservative
9	Pinchers map	{x}		23	Hnon area-preserving quad. map	{x,y}	
10	Spence map	{x}		24	Arnold's cat map	{x,y}	
11	Sine-circle map	{x}		25	Gingerbreadman map	{x,y}	
12	Hnon map	{x,y}		26	Chaotic web map	{x,y}	
13	Lozi map	{x,y}		27	Lorenz 3-d chaotic map	{x,y,z}	
14	Delayed logistic map	{x,y}			**Total of time series**	**44**	

there are a few chaotic maps whose placement lies at the bottom right of the plane, with large H_S and low C_{JS}, thus being very similar to random noises with small correlation degree k, hindering their distinction.

This challenging region is highlighted by a dashed-line rectangle in the bottom-right of Fig. 2b. It comprises the colored random noises with $k = \{0, \ldots, 1.75\}$, and chaotic maps: linear congruential generator (4-x), Gauss map (7-x), dissipative standard map (18-x, 18-y), sinai map (20-x, 20-y), and Arnold's cat map (24-x, 24-y). Although the CCEP method is able to mostly distinguish noisy from chaotic time series, these chaotic maps could be easily confused as noises. There is no clear separability between those time series within the challenging region, so a distance-based algorithm could not be applied to correctly make the distinction among them. Thus, another contribution aimed with this work is to present a method for distinguishing different time series dynamics, even within this region.

3. Ordinal patterns transition graphs

Let us recall the methodology proposed by Bandt and Pompe, which consists in transforming time series based on the ordinal relation between successive data samples [6,8,11]. It is a transformation from the time domain into a new representation composed of a sequence of symbolic ordinal patterns Π. Although the distribution of these OP reveals important properties of the underlying dynamics, it has some characteristics which limit their applicability to broad scenarios.

Our approach consists in using the set OP to create the ordinal patterns transition graphs (G_π), by taking into account the sequence of consecutive patterns. These graphs are the basis for the computation of the probability of self-transitions (p_{st}), our characterization criterion for different aspects of time series dynamics, which is defined as the probability of the occurrence of a sequence of equal patterns within the set OP. In the following, we detail these transformations and measures.

3.1. Definition and graph construction

Given the sequence of ordinal patterns Π, the ordinal patterns transition graph G_π represents the relations between consecutive ordinal patterns, and is defined as a directed weighted graph $G_\pi = (V, E)$ with $V = \{v_{\pi_i} : i = 1, \ldots, D!\}$, where each vertex v_{π_i} corresponds to one of the $D!$ possible permutations for an embedding dimension D, and a set of directed weighted edges $E = \{(v_{\pi_i}, v_{\pi_j}) : v_{\pi_i}, v_{\pi_j} \in V\}$. There is an edge $(v_{\pi_i}, v_{\pi_j}) \in E$, between the vertices v_{π_i} and v_{π_j}, if there is a sequence of ordinal patterns $\Pi_t = \pi_i$ and $\Pi_{t+1} = \pi_j$, $1 \le t \le m-1$, indicating a transition between these permutations.

The weight function $w : E \to \mathbb{R}$ of an edge represents the probability of the existence of a specific transition in Π. Thus, the weight of a given edge (v_{π_i}, v_{π_j}) is given by

$$w(v_{\pi_i}, v_{\pi_j}) = \frac{|\Pi_{\pi_i,\pi_j}|}{m-1}, \tag{7}$$

where $|\Pi_{\pi_i,\pi_j}| \in \{0, \ldots, m-1\}$ is the number of transitions between permutations π_i and π_j, and $\sum_{v_{\pi_i}, v_{\pi_j} \in V} w(v_{\pi_i}, v_{\pi_j}) = 1$.

3.2. Properties of the ordinal patterns transition graphs

The ordinal patterns representation was proposed focusing on its simplicity and fast calculation. It was intended to enable the extraction of information measures from time series that could have similar results of other well-known methods, such as the Lyapunov exponent for chaotic systems [30], but with a lower computational cost. Thus, once transition graphs

are constructed from the OP, some properties are directly inherited from that transformation. However, there are specific properties that could be noted when comparing it with other classical graph representations for time series.

3.2.1. Simplicity and fast calculation

The process of constructing transition graphs is very simple and fast. It only depends on the number $m = n - (D-1)\tau$ of OP, accounting the number of transitions in $m-1$ steps. The transformation from a time series into the set of OP, in turn, depends on the length n of the time series and the embedding dimension D. The time complexity to perform this transformation is bounded by $O(nD^2)$, assuming the permutations are obtained by sorting each sliding window by a simple sorting algorithm, such as selection sort, in $O(D^2)$ and the embedding delay $\tau = 1$ is the worst case. However, for practical purposes, D is recommended to be small [8], $3 \leq D \leq 7$, so the sorting will take at most 7 elements, and the time complexity largely depends on n.

3.2.2. Robustness

Since transition graphs are created from OP, the robustness in creating this set is also inherited. OP transformation is robust to the presence of observational and dynamic noise, and also invariant with respect to nonlinear monotonous transformations [6,10]. Although losing the amplitude of the original time series, it is still suitable for the analysis of experimental data, avoiding amplitude threshold dependencies that affect other methods based on range partitions [11].

3.2.3. Scalability

The main difference between G_π and VG approaches [21,22] is the scalability of graphs. In VG or HVG, each time series sample is mapped onto a vertex in the graph. Although this transformation can be satisfactorily performed in terms of time complexity, it requires a large space, which can be critical for long time series. Instead, the number of vertices for a transition graph is given by the embedding dimension D, independently of the time series length, and it is limited by $D!$, where $3 \leq D \leq 7$ [8].

3.3. Forbidden and missing transition patterns

Deterministic time series have an important characteristic: some patterns may be forbidden and will not occur in the set OP after data transformation, no matter the time series length [31]. Instead of forbidden patterns, time series with stochastic components, if long enough, will exhibit all possible patterns. Unobserved patterns in random time series are not forbidden, but missing, as the result of the finite sample size [27]. The same phenomenon can also be observed with respect to transition graphs. Similarly, there are forbidden and missing transition patterns, where some transitions between consecutive patterns do not occur. Unobserved patterns, whichever their nature, will not appear as a vertex in the constructed graph. Also, if a transition between two patterns does not occur, then this edge will not be present in the transition graph.

A special case of forbidden transition patterns that always occurs when $\tau = 1$, no matter the time series dynamics and value of D, is given in Theorem 3.1.

Theorem 3.1. *Given a time series x with length n, if the embedding delay $\tau = 1$ is used for computing ordinal patterns, its constructed ordinal patterns transition graph G_π will always present forbidden transitions, independently of n, of the chosen embedding dimension D, and of the dynamics of its underlying phenomena.*

Proof. Let $\mathbf{x} = \{x_1, \ldots, x_n\}$ be a time series that will be transformed into a set OP with $D = 3$ and $\tau = 1$. For a given time instant t, the sliding window $\mathbf{w}_t = \{x_t, x_{t+1}, x_{t+2}\}$ is the subset of $\mathbf{x}$ that will be used for computing the pattern at t. Let us assume, without loss of generality, that the pattern at t is $\pi_t = 123$, representing a ascending sequence of time series points, i.e., $x_t \leq x_{t+1} \leq x_{t+2}$. The next sliding window $\mathbf{w}_{t+1}$ will be composed of the sample points $\{x_{t+1}, x_{t+2}, x_{t+3}\}$, as the last two sample points was already fixed. Since we already know that $x_{t+1} \leq x_{t+2}$, then, the possible values for x_{t+3} are $x_{t+1} \leq x_{t+2} \leq x_{t+3}$, $x_{t+1} \leq x_{t+3} \leq x_{t+2}$, and $x_{t+3} \leq x_{t+1} \leq x_{t+2}$. Thus, the possible OP for these options at $t+1$ are $\pi_{t+1} = 123$, $\pi_{t+1} = 132$, and $\pi_{t+1} = 312$, respectively. By generalization, for each of the $D!$ patterns, the next possible transition patterns are limited to only D other patterns, since $D! - D$ patterns are forbidden due to the fixed previous points. □

Corollary 3.2. *As a consequence of Theorem 3.1, the maximum number of edges, including loops, for a transition graph when $\tau = 1$ is bounded by*

$$|E(G_\pi)|_{\tau=1} \leq D! \cdot D, \tag{8}$$

where $E(G_\pi)$ is the set of edges in G_π.

This limitation is given only for the case of $\tau = 1$, but it occurs for any value of D. For $\tau > 1$, the number of transitions depends on the time series behavior, and the maximum number of edges is given by a complete graph with loops, i.e.,

$$|E(G_\pi)|_{\tau>1} \leq D!^2. \tag{9}$$

If a transition graph has less edges, we may consider a scenario with more forbidden transitions or with missing transition patterns, but this last case occurs when the time series is not long enough so all the possible patterns can be observed.

3.4. Proposed metric: probability of Self-transitions

The way edges are set is an important aspect for further graph analysis. The weights can be normalized [12,13] such that all transitions from a given vertex sum up to 1, following a Markov chain representation. In this work, we follow a similar normalization [15], where the sum of all weights is 1. The main difference is that we accept the presence of self-transitions (loops) in the graph.

In fact, we show that self-transitions, when evaluated as a function of the embedding delay τ, represent a valuable indicative of the main characteristics of the time series.

The probability of self-transitions is defined as

$$p_{\text{st}} = p(\pi_i, \pi_i) = \sum_{i \in \{1,\dots,D!\}} w(v_{\pi_i}, v_{\pi_i}). \tag{10}$$

When considering the adjacency matrix $A_\pi = \{a_{\pi_i,\pi_j} : \pi_i, \pi_j \in \Pi\}$ of G_π, we can check that p_{st} is the trace of A_π. Thus, it can also be denoted as $p_{\text{st}} = \sum_{\pi_i \in \Pi} a_{\pi_i,\pi_i}$, representing the sum of its diagonal.

4. Results

A key aspect when dealing with graph representations of time series is what characteristics from the series are inherited by the graph, and how they can be extracted [21]. In this section, we apply learning algorithms for the characterization and distinction of time series, based on the measures extracted from the ordinal patterns transition graphs. We show how different dynamics can be expressed by analyzing inherited characteristics from periodic, random and chaotic synthetically generated time series.

4.1. Periodic time series

An important issue in a graph constructed from periodic time series is how to identify their period within the graph properties. For some graph representations, which is the case for the VG and HVG transformations, periodic time series are mapped onto regular graphs and the degree distribution of the vertices has a peak suggesting the period [21]. For the OP transformation, the number of vertices are limited to $D!$, avoiding a precise analysis of the degree distribution. Thus, an indicative of the time series period can be given from the transition graph when considering the p_{st} as a function of the embedding delay τ.

As a first analysis, let us consider the example of a periodic time series of a Visibility Graph (VG) [21], in which its first 20 samples are illustrated in Fig. 3a. It is a series of period $T = 4$, composed of n/T repetitions of the same pattern (0.87, 0.49, 0.36, 0.83). After the OP transformation with $D = 3$ and $\tau = 1$, due to its determinism, we have forbidden patterns that will not be observed, no matter the length of the time series [27,31]. The G_π for $D = 3$ and $\tau = 1$ is illustrated in Fig. 3c. The transitions between the four observed patterns have the same probability of occurrence, and there are no self-transitions, which give a $p_{st} = 0$.

However, different values of τ lead to different transition graphs. For instance, in Fig. 3d, for $\tau = 2$, we have only two patterns but four possible transitions. In this case, $p_{st} = 0.5$, considering the sum of loops. For $\tau = 3$, in Fig. 3e the graph is similar to the case $\tau = 1$, but patterns 213 and 312 were replaced by 132 and 231, respectively. Fig. 3f shows the case for $\tau = 4$, when the embedding delay is equal to the period of the series. Here we have a particular behavior: only pattern 123 is observed and, thus, all possible transitions are loops and $p_{st} = 1$. This occurs because when $\tau = T$ all sliding windows extract equal values from the time series, and the sequence is considered already ordered. Repeated behaviors can be observed in Fig. 3c–f for different values of τ.

Since forbidden patterns do not occur in all periodic time series, it is not expected that only one pattern will be observed like this example. Thus, we have to generalize this observation to be able to suggest the period of a given periodic time series based on the p_{st}. Thus, we conjecture that p_{st} assumes maximal values when τ is equal to the period T or multiples of it, i.e., $\tau = nT, n = \{1, 2, 3, \dots\}$. This maximal values can be observed as peaks in the values of p_{st}, as shown in Fig. 3b, which occurs regardless of D.

To examine this statement, let us consider the synthetic periodic time series, constructed from a sinusoidal function with period $T = 12.5$, in which its first 50 samples are shown in Fig. 4a. Without loss of generality, let us consider as example the cases for an embedding dimension $D = 3$, which enables a more clear visualization and explanation given the reduced number of $D!$ patterns. The ordinal patterns distribution of this time series is illustrated in Fig. 4b for $D = 3$ and $\tau = 1$, which reveals nothing regarding the time series period but shows that all possible patterns are observed. In fact, for any value of τ, all patterns will be observed for a large enough time series, and the main differences that may be observed are the number of edges and the probabilities of the transitions, as illustrated in Fig. 4d–g for different values of τ. According to these figures, we can see a repetition in the behavior of the graphs for different values of τ.

For the case of the sinusoidal series, a simple strategy to obtain the period consists in averaging the intervals between the values of τ where the peaks are higher than a particular threshold. As shown in Fig. 4c, we have $T = (12 + 13 + 12 + 13 + 12 + 13)/6 = 12.5$ as peaks for the p_{st}, when considering a peak the values higher than a threshold of 0.76. This result is in agreement with [11], where the normalized Shannon entropy and statistical complexity measures are evaluated as

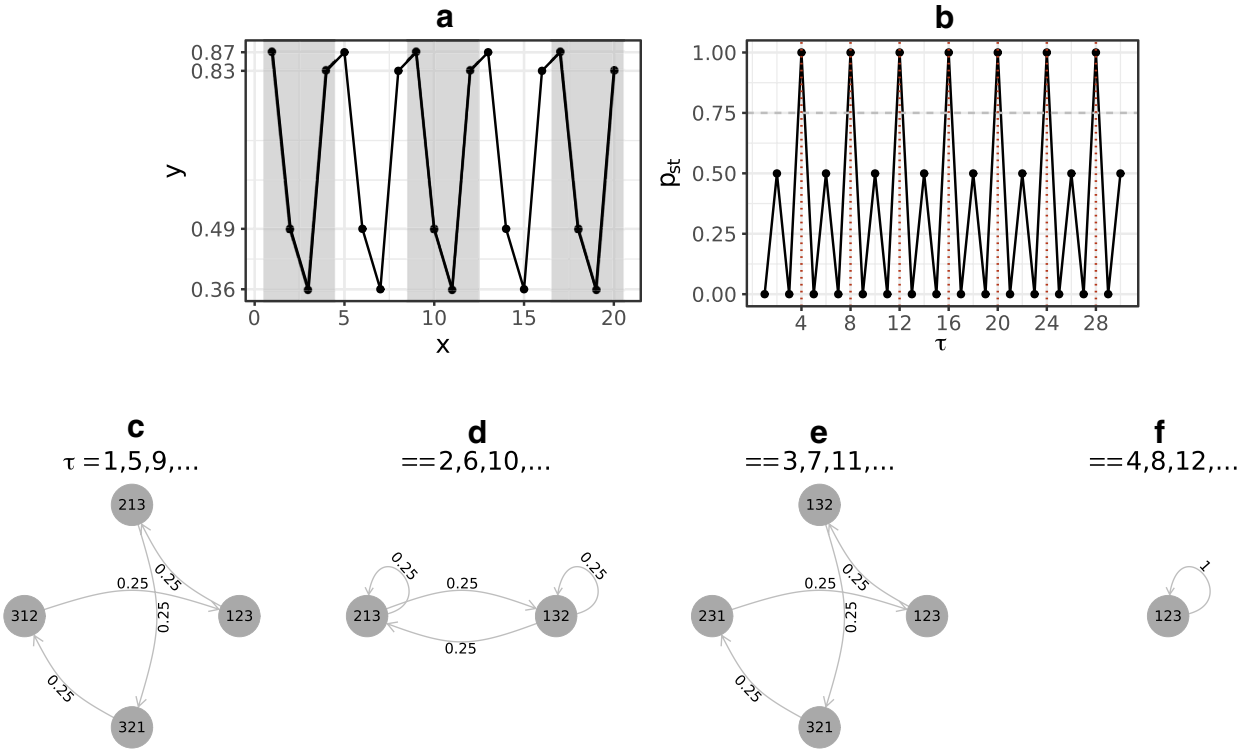

Fig. 3. Example of (a) the first 20 samples of the periodic time series from VG [21] and (b) its p_{st} with $D = 3$ as function of $\tau \in \{1, \ldots, 30\}$. It is also shown the G_π for the time series, with $D = 3$, and (d) $\tau = \{1, 5, 9, \ldots\}$, (e) $\tau = \{2, 6, 10, \ldots\}$, (f) $\tau = \{3, 7, 11, \ldots\}$, (g) $\tau = \{4, 8, 12, \ldots\}$.

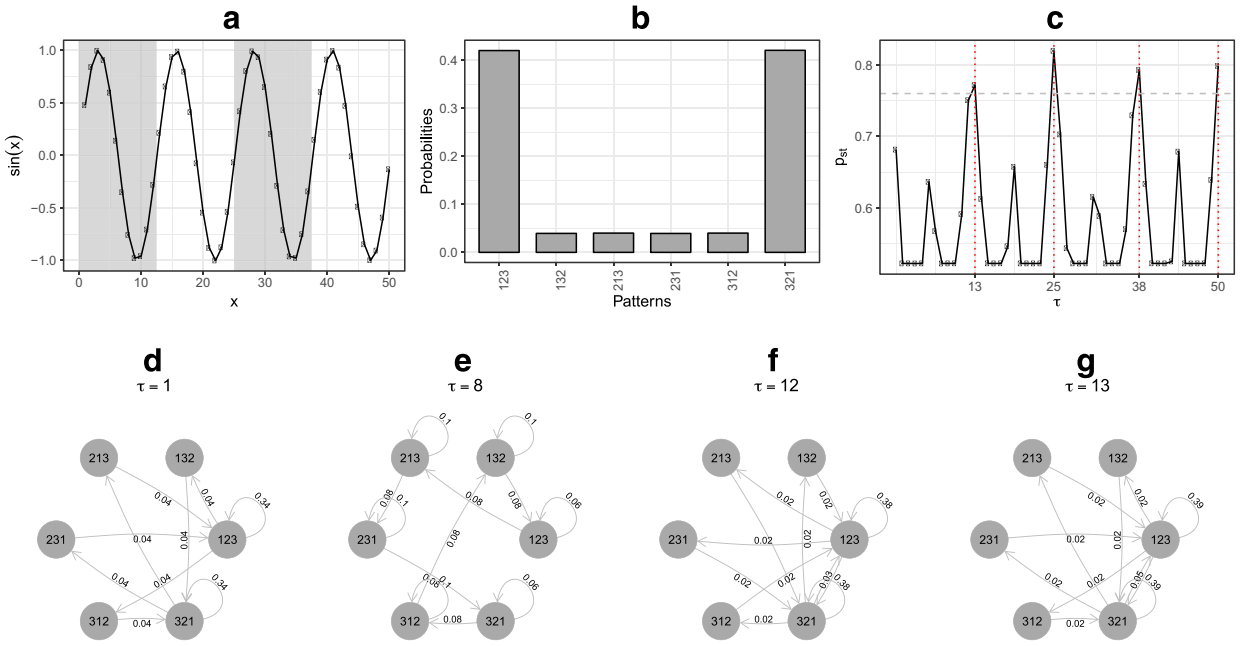

Fig. 4. Example of (a) the first 50 samples of a sinusoidal time series with period $T = 12.5$, (b) its OP probability distribution after the OP transformation with $D = 3$ and $\tau = 1$, and (c) its p_{st} as a function of $\tau \in \{1, 2, \ldots, 100\}$. Some G_π for this time series are illustrated, for $D = 3$ and (d) $\tau = 1$, (e) $\tau = 8$, (f) $\tau = 12$, and (g) $\tau = 13$.

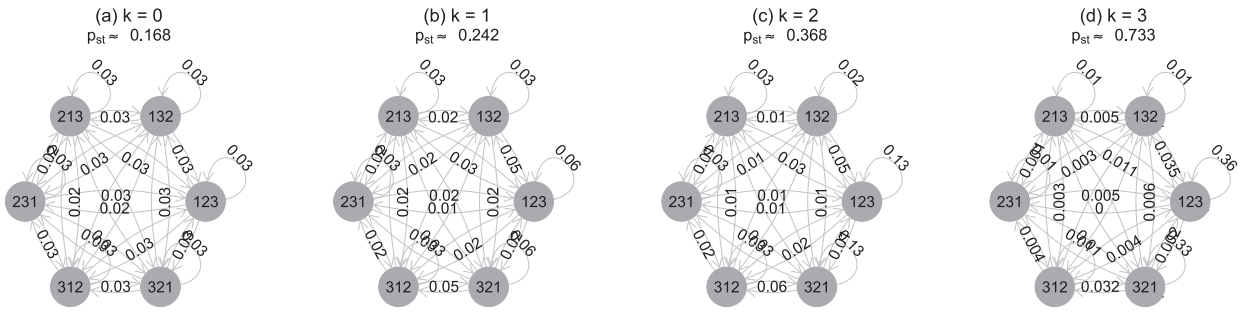

Fig. 5. Analysis of G_π from the random time series with different correlation levels presented in Fig. 1. Where (a) is a white noise ($k = 0$), (b) pink noise ($k = 1$), (c) red/brown noise ($k = 2$), and (d) a black noise ($k = 3$). Their corespondent ordinal patterns transition graphs are presented in (a) for the white noise ($k = 0$), (b) for the pink noise ($k = 1$), (c) for red/brown noise ($k = 2$), and (d) for the black noise ($k = 3$). Each G_π was constructed from the OP transformations computed as the average of 10 runs with different seeds for $D = 3$ and $\tau = 2$. The first 1000 points where omitted for the time series for better illustrate their randomness after the generation, and some minor transition probabilities (edges weights) were removed from the graphs to clear their visualization.

functions of τ for periodic time series. For those series, values are close to zero when τ matches the period or multiples of it.

4.2. Random time series

Studies in literature have investigated methods for identifying and characterizing randomness in time series [1,10,23,27,29]. Following the visibility graph approaches, random time series are mapped onto random graphs, with an exponential degree distribution [21,22]. While this can be used for a randomness test, in our case, besides this randomness identification, we also want to express the randomness level present at each time series.

In our strategy using the transition graph G_π, we assume a long enough random time series, so missing patterns are not expected to occur [31], and also no missing transition patterns when $\tau > 1$, as described in Theorem 3.1. Also, graphs will have a fixed number of $D!$ vertices, and the number of edges will reach their bound $D!^2$, defined in Eq. (9). In this case, each vertex will have the same degree, so the degree distribution does not reveal anything regarding the time series structure. Thus, we propose using the p_{st} as an indicative to the presence of some randomness, and to suggest at which levels the noise affects it.

To evaluate different randomness levels, we used the synthetically generated time series from Fig. 1, where Fig. 1a–d illustrate, respectively, a time series with a white ($k = 0$), pink ($k = 1$), red/brown ($k = 2$), and black ($k = 3$) noises. The transition graphs presented in Fig. 5a–d and were generated with from the OP transformations computed as the average of 10 runs with different seeds for $D = 3$ and $\tau = 2$.

Fig. 5a shows the OP transition graph created from the white noise time series. When $D = 3$ and $\tau = 2$, all the 36 possible transitions occurred, as expected by Eq. (9), with the same probability $p_{\text{st}} \approx 1/36 \approx 0.03, \forall i, j \in \{1, \ldots, D!\}$. This reinforces the notion that for a completely random time series (and it occurs independently of its marginal distribution) all the patterns [27] and all the transitions between consecutive patterns are expected to occur with the same probability, for long enough time series. Also, as the correlation between observations increases, i.e., $k > 0$, the randomness decreases and is expected that this should be observed in the graphs. In fact, when evaluated with examples of pink, red, and black noises, illustrated in Fig. 5b–d, p_{st} increases with the noise correlation. The values assumed for these noises are, approximately, 0.168, 0.242, 0.368 and 0.733, respectively.

Following our strategy to better understand these randomness levels, in Fig. 6 we present an analysis of the behavior of p_{st} for random time series when evaluated as functions of τ. We considered colored noises ranging from $k = 0$ to $k = 3.5$, by steps of $\Delta_k = 0.25$, illustrated as different data points for each k. We constructed the OP transformations for $D = \{3, 4, 5, 6\}$ and evaluated them as functions of $\tau = \{2, \ldots, 50\}$. In our analysis, we discarded the case for $\tau = 1$ to avoid the influence of forbidden transition patterns, as previously described in Theorem 3.1.

For the particular case of white noise ($k = 0$), there is a special situation where increasing τ does not affect p_{st}, giving approximately $1/D!$ for all combinations of D and τ. However, for $k > 0$ there is a growing tendency in p_{st} as τ increases. It is possible to see a direct relation between p_{st} and its determinism. The more deterministic the series, given by k, the higher the p_{st}. This increasing determinism as a function of τ was also pointed out in [11], and we could also verify its influence here. Another point related to this growing tendency of p_{st} is its behavior between different values of D. As Fig. 6 shows, there is a direct relation between D and an easy distinction of k. For instance, for $D = 3$, we can clearly see a better distinction of different correlation levels for lower values of k. On the contrary, for $D = 6$, lower values of k became hard to distinguish and higher values of k became clearer.

Following our strategy of distinguishing the randomness levels in time series, based on curves of Fig. 6, one may consider its characterization by analyzing the growing behavior of p_{st}. For this analysis, we quantify the correlation degree of random

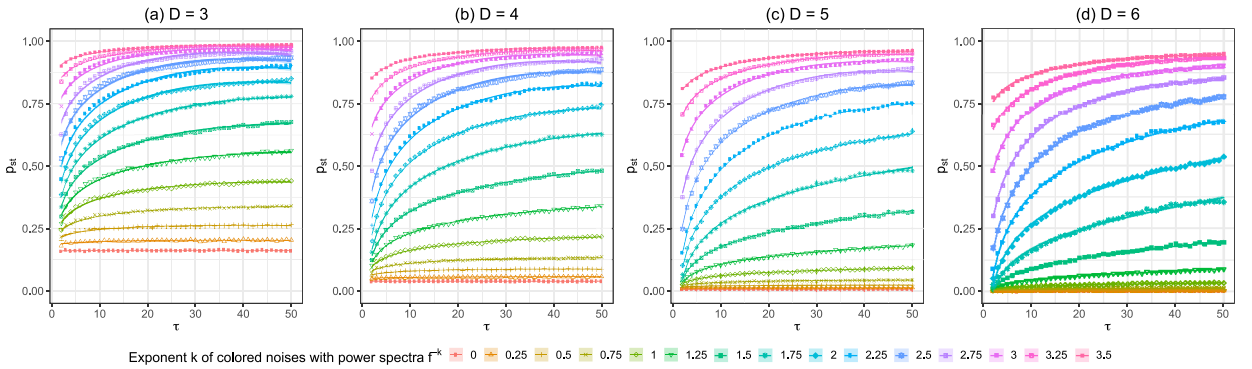

Fig. 6. Analysis of p_{st} from G_π of different colored noises with power spectra f^{-k}, ranging from $k = 0$ to $k = 3.5$, by steps of $\Delta_k = 0.25$, as a function of $\tau = \{2, \ldots, 50\}$, for $D = \{3, 4, 5, 6\}$. Fitted values for regression model are illustrated for each curve as solid lines, with upper and lower limits of 0.95 confidence intervals as shaded areas.

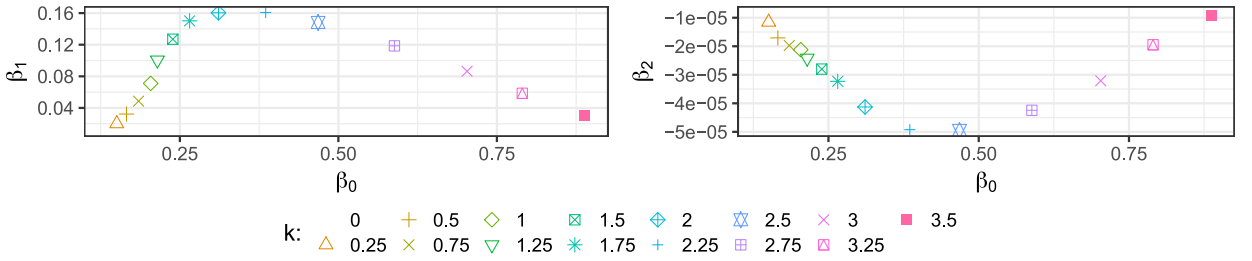

Fig. 7. Analysis of the parameters β_0, β_1, and β_2 from the regression model presented in Eq. (11), fitted for the colored random noises of Fig. 6a, considering the case for $D = 3$.

time series by fitting a regression model for each of these noise curves. For each $D = \{3, 4, 5, 6\}$, we considered the following linear model

$$p_{st} = \beta_0 + \beta_1 \log \tau + \beta_2 \tau^2 + \varepsilon, \tag{11}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is a random error following a normal distribution with zero mean and unknown variance σ^2. This model assumes the presence of an intercept, a logarithmic growing for the curves as τ increases and a quadratic component to better adjust the model. Other models may be used, but this one is satisfactory for our purpose. The fitted curves for each noise are plotted as solid lines in Fig. 6, with shaded 95% confidence regions. The unbiased mean squared error (MSE) of residuals in the regression model fitted for all random noises ranges from 3.58×10^{-8} to 1.7×10^{-4}.

Fig. 7 shows the estimated parameters of the model applied to the p_{st} curves of Fig. 6a, with $D = 3$. Parameters β_1 and β_2 are presented as functions of the intercept β_0. As β_0 increases, β_1 and β_2 present similar behaviors, but in opposite directions. For both cases, each noise level k is placed at different regions, indicating they can be used for distinction.

Thus, to evaluate the model, we performed the classification of time series with different randomness degrees, with the intention to effectively distinguishing among them. The dataset used for this task is composed of 150 time series, with 10 samples for each of the 15 different colored noises with power spectra f^{-k}, ranging from $k = 0$ to $k = 3.5$, by steps of $\Delta_k = 0.25$. The dataset was split in the proportion of 70% and 30% of the time series for the training and test sets, respectively, by keeping the same number of time series for each k. For each time series, we computed their p_{st} values for each $D = \{3, 4, 5, 6\}$, as functions of $\tau = \{2, \ldots, 50\}$. After fitting the regression model presented in Eq. (11), we only extract the parameters β_0 and β_1 for the fitted model. The reason for this choice is simply by the similar behavior between β_1 and β_2, as shown in Fig. 7.

For the classification, we used the Support Vector Machine [32], evaluated with radial (SVMR), linear (SVML), polynomial (SVMP), and sigmoid (SVMS) kernels. Table 2 presents the classification accuracy results achieved with this strategy. We can see its success in correctly distinguishing these different random levels. For SVM with both radial and linear kernels, we achieved 100% accuracy when $D = 3$ and $D = 4$. This highlights the effectiveness and efficiency of the method, since the lower the values of D, the lower the resource consumption.

4.3. Chaotic time series

As discussed before, a hard task when dealing with time series from both stochastic processes and chaotic systems is to distinguish them. This is an intriguing aspect since these systems share several properties, even being completely different

Table 2
Classification accuracy (%) results achieved with our strategy for distinguishing time series with different randomness levels. We present the results of different classification algorithms for $D = \{3, 4, 5, 6\}$.

D	SVMR	SVML	SVMP	SVMS
3	**100**	**100**	93.33	51.11
4	**100**	**100**	97.78	48.89
5	97.78	97.78	95.56	57.78
6	95.56	95.56	88.89	68.89

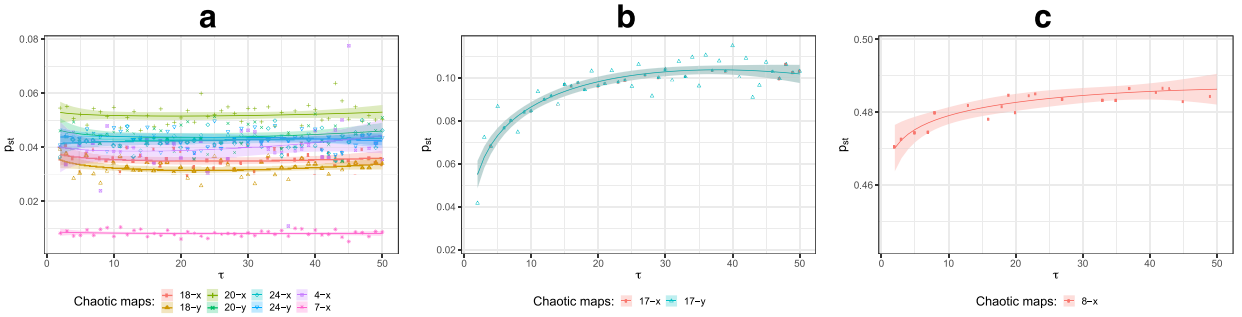

Fig. 8. Analysis of (a) challenging chaotic maps from Fig. 2b. These are the linear congruential generator (4-x), Gauss map (7-x), dissipative standard map (18-x, 18-y), sinai map (20-x, 20-y), and Arnold's cat map(24-x, 24-y). The (b) Holmes cubic map (17-x, 17-y) and (c) Cusp map (8-x) are also presented. Curves are computed for $\tau = \{2, \ldots, 50\}$ and $D = 4$. The fitted values for the regression model (Eq. (11)) are shown as solid lines, with upper and lower limits of 0.95 confidence intervals as shaded areas.

Table 3
Clustering results for distinguishing 59 time series (random process and chaotic systems). The accuracy values and errors for this distinction are also presented.

D	Accuracy	Errors
3	93.22	0, 8-x, 17-x, 17-y
4	94.91	0, 17-x, 17-y
5	**96.61**	0, 0.25
6	94.91	0, 0.25, 8-x

in concept [10,11,29]. Many studies were proposed to tackle this issue, such as the CCEP method [3,10], which is able to mostly distinguish noisy from chaotic time series, but some chaotic maps and random noises could still be misclassified. As presented in Section 2, there is no clear separation between those time series within the challenging region in Fig. 2, so a distance-based algorithm could not be applied to correctly make the distinction among them.

Before presenting the proposed method for distinguishing time series, let us first revisit the previously discussed p_{st} as functions of τ and understand its behavior when considering the chaotic maps. Fig. 8 analyzes the challenging chaotic maps (composed of 8 chaotic time series) and the Cusp map. In that figure we can also see the fitted line and 0.95 confidence interval region for the regression model presented in Eq. (11). We see that p_{st} for some of these time series, highlighted in Fig. 8a, does not necessarily follow a stable behavior as compared to the fitted results for the random time series in Fig. 6. For instance, the linear congruential generator (4-x), which presents the more challenging placement in CCEP, at the far bottom right of the plane, does not grow with τ. This leads to a poor fit of the regression model, which can be used as a considerable difference in behavior from the random series. On the other hand, there are chaotic maps where the regression model fits quite well, such as the Holmes cubic map (17-x, 17-y) and the Cusp map (8-x), presented in Fig. 8b and c, respectively. If compared to Fig. 6, we note that both curves and fitted regression model are similar, making their distinction more difficult.

Thus, in order to distinguish chaotic maps from random time series, based on their behavior as a function of τ, we have to consider both the parameters from the regression model of Eq. (11) and an analysis of the fitted model. The set of used features are the parameters β_0 and β_1, as used for the random time series case, together with the error for the fitted model represented by its coefficient of determination R^2. Thus, we performed an experiment that aims to evaluate our strategy in distinguishing each of the 59 time series presented in the analysis of Fig. 2 as "Random noise" or "Chaotic map". For each ith time series, we transformed it onto a feature vector $f_i = \{\beta_0^i, \beta_1^i, R^{2,i}\}$, and used the k-means [33] clustering algorithm for partitioning the time series into these two clusters. We compute these features for $D = \{3, 4, 5, 6\}$ and $\tau = \{2, \ldots, 50\}$ over time series with 5000 observations in length, which is sufficient for these values of D.

Table 3 presents the clustering results, with the accuracy and errors for this distinction. The best result is achieved for $D = 5$, with just 2 errors out of 59 cases, which gives an accuracy result of 96.61%. The errors in this case are, exactly, the

random noises with correlation degree $k = 0$ (white noise) and $k = 0.25$. In fact, as mentioned in Section 4.2, these lower levels of k are the ones with practically constant growing tendency as τ increases, mainly for higher values of D, which give poor fitting for the regression model, thus, being identified as a chaotic map. However, for this case, all the remaining time series were correctly identified. For other values of D, we achieved lower accuracies, 93.22% for $D = 3$, and 94.91% for $D = 4$ and $D = 5$, where some chaotic maps were misidentified as random noises. These are exactly the cases illustrated in Fig. 8b and c, the Holmes cubic maps (17-x, 17-y) and Cusp map (8-x), respectively, whose behaviors are similar to noises.

These results show the efficacy of the proposed method for distinguishing different time series dynamics, which is a hard task with several proposals found in literature. Also, they highlight the simplicity of the method, which could be easily employed in conjunction with simple learning algorithms, giving reasonable results with low computational complexity.

5. Conclusions and future directions

The approach based on the p_{st} from G_π, presented in this work, provides a valuable tool for the characterization of time series. We show that p_{st}, when evaluated as a function of the embedding delay τ, can be successfully applied as an indicative of different time series dynamics. For periodic time series, when τ matches the period (or multiples) of the time series, the measure reaches maximal levels, suggesting a strategy to discover it. This confirms the results described in previous studies about the observed behavior of information theory quantifiers when also evaluated as function of τ.

We show for random time series with different degrees of correlation and colored noises that p_{st} is directly related to their correlation. It follows a monotonically increasing function of τ for all colored noises except for the white noise ($k = 0$). For this particular case, the measure remains constant for all values of τ, implying in a higher randomness level. However, we could precisely distinguish those different randomness levels by using a supervised learning (classification) algorithm, reaching 100% in accuracy.

To distinguish random noises from chaotic maps, we show that the behavior of p_{st}, when evaluated as function of τ, is different for these types of time series. Thus, we map this behavior by fitting a regression model for each curve and show that, while the proposed model almost perfectly fits the random series, it does not properly represent chaotic maps. In fact, by assessing the model parameters and residual errors when fitting, the difference in their behaviors were used to supply an unsupervised learning (clustering) algorithm that correctly identified 96.61% of the underlying dynamics.

In general, our method is a simpler alternative for the characterization of time series, even short ones (we used here 5000 data points). However, there are still a few time series in which their dynamics are misclassified. For those cases, an alternative solution for future studies might be the integration between our proposed method with already established techniques, such as other more complex learning algorithms. When applicable, one strategy could solve the drawbacks of another, yielding a more robust and general solution. Furthermore, the method can also be evaluated when applied to realistic time series, to verify the results when facing new challenges present in real world scenarios.

Acknowledgments

The authors would like to thank the research agencies São Paulo Research Foundation (FAPESP) via grant #15/24536-2, Research Foundation of the State of Alagoas (FAPEAL), Brazilian National Council for Research and Development (CNPq) via grants #447229/2014-1 and #405364/2018-0, and Coordination for the Improvement of Higher Education Personnel (CAPES).

References

[1] O.A. Rosso, F. Olivares, A. Plastino, Noise versus chaos in a causal Fisher-Shannon plane, Pap. Phys. 7 (2015) 070006.
[2] J. Gama, I. Žliobait, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM Comput. Surv. 46 (4) (2014) 1–37.
[3] O.A. Rosso, F. Olivares, L. Zunino, L. De Micco, A.L.L. Aquino, A. Plastino, H.A. Larrondo, Characterization of chaotic maps using the permutation Bandt-Pompe probability distribution, Eur. Phys. J. B 86 (4) (2013) 116.
[4] A.L.L. Aquino, T.S.G. Cavalcante, E.S. Almeida, A.C. Frery, O.A. Rosso, Characterization of vehicle behavior with information theory, Eur. Phys. J. B 88 (10) (2015) 257.
[5] O.A. Rosso, R. Ospina, A.C. Frery, Classification and verification of handwritten signatures with time causal information theory quantifiers, PLOS One 11 (12) (2016) e0166868.
[6] A.L. Aquino, H.S. Ramos, A.C. Frery, L.P. Viana, T.S. Cavalcante, O.A. Rosso, Characterization of electric load with information theory quantifiers, Phys. A Stat. Mech. App. 465 (2017) 277–284.
[7] H.V. Ribeiro, M. Jauregui, L. Zunino, E.K. Lenzi, Characterizing time series via complexity-entropy curves, Phys. Rev. E 95 (6) (2017) 062106.
[8] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17) (2002) 174102.
[9] H. Larrondo, M. Martín, C. González, A. Plastino, O. Rosso, Random number generators and causality, Phys. Lett. A 352 (2006) 421–425.
[10] O.A. Rosso, H.A. Larrondo, M.T. Martin, A. Plastino, M.A. Fuentes, Distinguishing noise from chaos, Phys. Rev. Lett. 99 (15) (2007) 154102.
[11] L. Zunino, M.C. Soriano, O.A. Rosso, Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach, Phys. Rev. E 86 (4) (2012) 046210.
[12] M. McCullough, M. Small, T. Stemler, H.H.-C. Iu, Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems, Chaos 25 (5) (2015) 053101.
[13] T. Sorrentino, C. Quintero-Quiroz, A. Aragoneses, M.C. Torrent, C. Masoller, Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback, Opt. Express 23 (5) (2015) 5571.
[14] C.W. Kulp, J.M. Chobot, H.R. Freitas, G.D. Sprechini, Using ordinal partition transition networks to analyze ECG data, Chaos Interdiscipl. J. Nonlinear Sci. 26 (7) (2016) 073114.
[15] J. Zhang, J. Zhou, M. Tang, H. Guo, M. Small, Y. Zou, Constructing ordinal partition transition networks from multivariate time series, Scient. Rep. 7 (1) (2017) 7795.
[16] J. Zhang, M. Small, Complex network from pseudoperiodic time series: topology versus dynamics, Phys. Rev. Lett. 96 (23) (2006) 238701.

[17] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, M. Small, Characterizing pseudoperiodic time series through the complex network approach, Phys. D Nonlinear Phenomena 237 (22) (2008) 2856–2865.
[18] Z.-K. K. Gao, N.-D. D. Jin, A directed weighted complex network for characterizing chaotic dynamics from time series, Nonlinear Anal. Real World Appl. 13 (2) (2012) 947–952.
[19] J. Tang, Y. Wang, F. Liu, Characterizing traffic time series based on complex network theory, Phys. A Stat. Mech. Appl. 392 (18) (2013) 4192–4201.
[20] Z.-K. Gao, M. Small, J. Kurths, Complex network analysis of time series, EPL (Europhys. Lett.) 116 (5) (2016) 50001.
[21] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuno, From time series to complex networks: the visibility graph, Proc. Natl. Acad. Sci. 105 (13) (2008) 4972–4975.
[22] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: exact results for random time series, Phys. Rev. E 80 (4) (2009) 046103.
[23] M.G. Ravetti, L.C. Carpi, B.A. Gonçalves, A.C. Frery, O.A. Rosso, Distinguishing noise from chaos: objective versus subjective criteria using horizontal visibility graph, PLoS One 9 (9) (2014) e108004.
[24] B.A. Gonçalves, L. Carpi, O.A. Rosso, M.G. Ravetti, Time series characterization via horizontal visibility graph and Information Theory, Phys. A Stat. Mech. Appl. 464 (2016) 93–102.
[25] P. Lamberti, M. Martin, A. Plastino, O. Rosso, Intensive Entropic non-triviality measure, Phys. A Stat. Mech. Appl. 334 (1-2) (2004) 119–131.
[26] O.A. Rosso, L. Zunino, D.G. Pérez, A. Figliola, H.A. Larrondo, M. Garavaglia, M.T. Martín, A. Plastino, Extracting features of Gaussian self-similar stochastic processes via the Bandt-Pompe approach, Phys. Rev. E 76 (6) (2007) 061114.
[27] O.A. Rosso, L.C. Carpi, P.M. Saco, M. Gómez Ravetti, A. Plastino, H.A. Larrondo, Causality and the entropy complexity plane: Robustness and missing ordinal patterns, Phys. A Stat. Mech. Appl. 391 (1-2) (2012) 42–55.
[28] C.W. Kulp, L. Zunino, Discriminating chaotic and stochastic dynamics through the permutation spectrum test, Chaos Interdisci. J. Nonlinear Sci. 24 (3) (2014) 033116.
[29] B. Ye, J. Chen, C. Ju, H. Li, X. Wang, Distinguishing chaotic time series from noise: a random matrix approach, Commun. Nonlinear Sci. Numer. Simul. 44 (2017) 284–291.
[30] A. Wolf, J.B. Swift, H.L. Swinney, J.A. Vastano, Determining Lyapunov exponents from a time series, Phys. D Nonlinear Phenom. 16 (3) (1985) 285–317.
[31] J.M. Amigó, L. Kocarev, J. Szczepanski, Order patterns and chaos, Phys. Lett. A 355 (1) (2006) 27–31.
[32] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[33] J. Macqueen, Some methods for classification and analysis of multivariate observations, Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. 1 (233) (1967) 281–297.