Symbolic recurrence plots to analyze dynamical systems

M. Victoria Caballero-Pintado, 1 Mariano Matilla-García, 2,a) and Manuel Ruiz Marín 3

1Departamento de Métodos Cuantitativos para la Economía y la Empresa, Universidad de Murcia, Campus de Espinardo, Murcia 30100, Spain
2Departamento de Economía Aplicada y Estadística, UNED, Paseo Senda del Rey, 11, 28040 Madrid, Spain
3Departamento de Métodos Cuantitativos e Informáticos, Universidad Politécnica de Cartagena, Cartagena 30201, Spain

(Received 23 February 2018; accepted 15 May 2018; published online 12 June 2018)

This paper, based on the concept of symbolic correlation integral, introduces a set of symbolic recurrence plots and associated invariant measures, which are independent of the distance parameter, serving as a tool for quantifying the dynamic structure. These new measures allow the study of transient behavior, coexistence of attractors, bifurcations, and structural change. The final user does not have to choose the free distance parameter. An empirical application to electrocardiography data illustrates the use of symbolic recurrence measures. Published by AIP Publishing. https://doi.org/10.1063/1.5026743

I. INTRODUCTION

Recurrence plots (RPs), correlation integrals, and symbolic pattern analysis are standard concepts (and tools) and the palette of any researcher in nonlinear science. We investigate how symbolic recurrence plots (SRPs) and their derived measures can be easily used as productive tools to analyze the dynamic structure of dynamical systems. Symbolic recurrences are introduced under the theoretical support of the concept of symbolic correlation integral (SCI). A symbolic recurrence plot has several advantages: (i) it can be used for stationary and nonstationary data sets; (ii) it is robust to noise and invariant under certain data transformations; and (iii) it can be used as a tool for diagnostic analysis given its statistical support.

Recurrence plots (RPs), correlation integrals, and symbolic pattern analysis are standard concepts (and tools) and the palette of any researcher in nonlinear science. We investigate how symbolic recurrence plots (SRPs) and their derived measures can be easily used as productive tools to analyze the dynamic structure of dynamical systems. Symbolic recurrences are introduced under the theoretical support of the concept of symbolic correlation integral (SCI). A symbolic recurrence plot has several advantages: (i) it can be used for stationary and nonstationary data sets; (ii) it is robust to noise and invariant under certain data transformations; and (iii) it can be used as a tool for diagnostic analysis given its statistical support.

A. Symbolic Recurrence Plots

Given a real-valued time series \( \{x_t\}_{t \in I} \) and an embedding dimension \( m \), an RP is a graphical representation of a \( n \times n \) matrix, namely \( (R_{ij}) \), representing neighboring \( m \)-histories \( \bar{x}_i = (x_t, x_{t+1}, \ldots, x_{t+m-1}) \) in an \( m \)-dimensional phase space where

\[
R_{ij}(\varepsilon) = \begin{cases} 
1 & \text{if } \|\bar{x}_i - \bar{x}_j\| \leq \varepsilon, \\
0 & \text{otherwise,} 
\end{cases} 
\]

for \( i, j = 1 \ldots n \), where \( n \) is the number of considered states \( \bar{x}_i \); \( \varepsilon \) is a threshold distance (proximity parameter), and \( \| \cdot \| \)

\( a \)Author to whom correspondence should be addressed: mmatilla@cece.uned.es

\( 1 \) is a norm. Hence, (1) is a pairwise test of the closeness of points on a phase space trajectory: points that fall in the neighborhood of size \( \varepsilon \) are recurrence points.

This idea is based on the mathematical concept of a correlation integral

\[
C(m, \varepsilon) = \frac{2}{n(n-1)} \sum_{i,j} R_{ij},
\]

introduced by Grassberger and Procaccia (1983). Roughly speaking, the correlation integral measures the probability that two \( m \)-histories are within a ball of radius \( \varepsilon \), which coincides with the probability of finding a recurrence point. The RP is therefore clearly related to the correlation integral. The probability \( C(m, \varepsilon) \) is currently a concept and also an analytic technique that together with RP analysis is widely used in science. In particular, RPs (and the correlation integral) have been used to develop several measures (see Webber and Zbilut, 1994 and Marwan et al. 2002) to quantify structures that occur in these RPs. These measures are referred as to the quantification of a recurrence plot (RQA) as they provide quantification tools for analyzing complex and time-varying dynamical systems, and also for the analysis of nonlinear systems in many different research fields including biology, neuroscience, psychology, physiology, engineering, physics, geosciences, linguistics, finance, and economics (Marwan et al. 2007 and Webber, Jr. and Marwan, 2015).

Recently, Caballero et al. (2017) have proposed the concept of a symbolic correlation integral (SCI), which is the theoretical support of what we present and refer to in this paper as a symbolic recurrence plot (SRP). Although it was initially proposed by Marwan et al. (2007) it has not yet been developed fully. In particular, in this paper, we propose several measures aiming at constructing a tool for symbolic recurrence quantification analysis. It is shown how these measures link patterns in an SRP with specific underlying dynamic structures and how (partially because of its theoretical underpinning on SCI) those measures can be interpreted.
The well-known correlation integral and RP analysis are both used throughout a wide range of scientific domains. By construction, recurrence-based methods depend on a proximity parameter. This parameter, \( \varepsilon \), should be chosen ex ante, with limited guidance from theory. If \( \varepsilon \) is chosen too large, almost every point is a neighbor of every other point. On the other hand, if \( \varepsilon \) is taken too small, we have no recurrence points and we cannot explain anything about the dynamic system. Many scholars have devoted their time to obtaining criteria for selecting \( \varepsilon \) (Koebbe and Mayer-Kress, 1992; Zbilut and Webber, 1992; Zbilut et al., 2002; and Marwan, 2003). There is no consensus on the topic. Many techniques recommend choosing \( \varepsilon \) such that \( \varepsilon > 5\sigma \), where \( \sigma \) is the standard deviation of measures of the observed process (Thiel et al., 2002). This choice is especially crucial in the case of experimental or observational data where there is only a univariate time series available.

Under these circumstances, alleviating practitioners from choosing a particular or a set of \( \varepsilon \), will avoid potential conflict about distinct outputs for the same measure. As occurs with SCI, SRP is \( \varepsilon \)-independent and therefore all the measures that will be shown in this paper benefit from this advantage. This allows us to think of SRP and its derived measures as invariants of the measured system, and therefore it can be used a tool for distinguishing among time series.

Symbolic analysis is based on the information theory and on the theory of dynamical systems. This methodological approach has had a strong resurgence from the studies of (Bandt and Pompe, 2002). Since then the number of studies that use symbolic dynamics has grown significantly and in several disciplines [see Amigó (2010)]. The properties of symbolic encodings are crucial to the theory of communication Shannon (2001). The mathematical discipline of symbolic dynamics, which studies the behavior of dynamical systems, started in 1898 with the pioneering studies of Hadamard, which developed a symbolic description of sequences of geodesic flows, and was later extended by Morse (1921), who coined the name “symbolic dynamics.” Collet and Eckmann (2009) showed that a complete description of the behavior of a dynamical system can be captured in terms of certain appropriate symbols. The basic idea of symbolic dynamics is very simple: divide the phase space into a finite number of regions and label each region by a letter (a symbol) from a certain alphabet (a set of symbols); instead of following a trajectory point by point, one only keeps recording the alternation of symbols. As occurs with coarse-grained methods, usually used to provide some description of the data generating process, symbolic analysis focus on some essential features of the generating dynamics, which are frequently of interest to the researcher, for example, periodicity, dependence, structure, etc. In the particular case of this paper, the basic block of analysis is the ordinal pattern, which only demands a totally order set to be fully applicable.

In this paper, we show the versatility and power of the symbolic recurrence quantification as a tool to analyze the dynamic structure of dynamical systems that avoids the use of the nuisance proximity parameter \( \varepsilon \). More concretely, we consider a special symbolic dynamic of the system where the time series is encoded by using ordinal patterns. A similar technique was used by Groth (2005) who used cross SRP to visualize dependencies between two-time series and derived a measure of the coupling strength. Another kind of SRP was proposed by Donner et al. (2008). As will be evident in this paper, a code of colors plays a relevant role in the characterization of the recurrence. It is shown that SRP when applied to real data, allows segregation between groups of observations. Along with these added values, using SRP has several other potential advantages: (i) it can be used for stationary and non-stationary data sets, which facilitates studying long time series; (ii) it is robust to noise and invariant under certain data transformations; and (iii) it can be used as a tool for diagnostic analysis given its statistical support.

The paper is organized as follows: in Sec. II, we fix the notation and give the main definition of this paper: the SRP. In Sec. III, we show different SRPs for several time series from dynamical processes. In these plots, we can distinguish patterns that are linked to different dynamical behaviors. Section IV deals with measures defined on the studied plot. In Secs. V and VII, we apply this methodology to deterministic, stochastic, and real dynamical systems. Section VI is devoted to the comparison of classical versus symbolic RQA and finally, we provide a summary of the main ideas in Sec. VIII.

II. DEFINITIONS AND NOTATION

Let \( \{\vec{x}_t\}_{t=1}^n \) be an \( m \)-dimensional time series of length \( n \). For \( m \geq 2 \), we denote by \( S_m \) the symmetric group of order \( m! \), that is, the group formed by all the permutations of length \( m \). We will call an element \( \pi = (i_1, i_2, \ldots, i_m) \in S_m \) a symbol. We say, that an element \( \vec{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m \) is of \( \pi \)-type if and only if \( \pi = (i_1, i_2, \ldots, i_m) \) is the unique symbol in the symmetric group \( S_m \) satisfying the two following conditions:

\[
\begin{align*}
(a) & \quad x_{i_1} \leq x_{i_2} \leq \ldots \leq x_{i_m}, \quad \text{and} \\
(b) & \quad i_{s-1} < i_s \text{ if } x_{i_{s-1}} = x_{i_s}.
\end{align*}
\]

Condition (b) guarantees the uniqueness of the symbol \( \pi_i \). This is justified if the values of \( x_i \) have a continuous distribution so that equal values are very uncommon, with a theoretical probability of occurrence of zero.

Next, we define the symbolization map

\[
s : \{\vec{x}_t\}_{t=1}^n \rightarrow S_m,
\]

defined by \( s(\vec{x}_t) = \pi \) if and only if \( \vec{x}_t \) is of \( \pi \)-type. Note that the symbolization map \( s \) transforms the vectorial time series into a sequence of symbols (permutations).

Recall that for this \( m \) there exist \( \frac{m(m-1)}{2} \) hyperplanes providing a partition

\[
P = \{A_{\pi_1}, A_{\pi_2}, \ldots, A_{\pi_{m!}}\},
\]

of \( \mathbb{R}^m \) of size \( m! \). This partition satisfies that for each \( j = 1, 2, \ldots, m! \), there is a unique symbol \( \pi_j \in S_m \) such that all elements in the interior of \( A_{\pi_j} \) are of the same \( \pi_j \)-type. We will say, that two states \( \vec{x}_t \) and \( \vec{x}_s \) are symbolic recurrence
states to the symbol $\pi$ if and only if $s(\bar{x}_i) = s(\bar{x}_j) = \pi$, that is, they belong to the same element $A_\pi$ of the partition $P$.

Similar to classical RP (Eckmann et al., 1987), these symbolic recurrence states can be represented in a matrix $\text{SR} \in M_n(\{0, 1\})$, where the entries are given by

$$\text{SR}_{ij} = \begin{cases} 1 & \text{if } s(\bar{x}_i) = s(\bar{x}_j), \\ 0 & \text{otherwise}, \end{cases}$$

with $i, j = 1, 2, \ldots, n$.

The SRP visualizes symbolic recurrence states and it is obtained by plotting the symbolic recurrence matrix. Both axes of the SRP are time axes and each point $(t, s)$ is marked with a black dot if $\bar{x}_i$ and $\bar{x}_j$ are recurrence states to symbol $\pi$ for some symbol $\pi \in S_m$ (that is, they belong to the same set $A_\pi$ of the symbolic partition of $\mathbb{R}^m$). We call these dots in the SRP symbolic recurrence points. By definition, the SRP always has a black main diagonal line and is symmetric with respect to this diagonal.

In addition to knowing whether $\bar{x}_i$ and $\bar{x}_j$ are recurrence states, it is also informative to know to which set of the symbolic partition they belong. To distinguish among different symbols of symbolic recurrence states, each symbol $\pi$ will be represented by a number $c_\pi$ representing a color in the SRP. Then, we construct the matrix $\overline{\text{SR}}$ with entries

$$\overline{\text{SR}}_{ij} = \begin{cases} c_\pi & \text{if } s(\bar{x}_i) = s(\bar{x}_j) = \pi \\ 0 & \text{otherwise}. \end{cases}$$

In this case, $\overline{\text{SR}}$ is represented using the colored SRP. This plot has (at most) $m!$ different colors where each point $(t, s)$ is left empty if $\overline{\text{SR}}_{ij} = 0$ and it is marked with a colored dot of color represented by $c_\pi$ always $\bar{x}_i$ and $\bar{x}_j$ are recurrence states to $\pi$.

We are interested in the analysis of real-valued time series. To this end, we consider that the time series is embedded in an $m$-dimensional space as follows:

$$\bar{x}_i = (x_i, x_{i+1}, \ldots, x_{i+(m-1)}) .$$

The positive integer $m$ is usually known as the embedding dimension.

We illustrate the symbolization procedure on the time series shown in Fig. 1 and Fig. 2, whose seven values are

$$\{x_1 = 3, \ x_2 = 9, \ x_3 = 7, \ x_4 = 6, \ x_5 = 5, \ x_6 = 10, \ x_7 = 4\} .$$

Given the embedding dimension $m = 3$, we have six symbols comprising the symmetric group, that is,

$$S_3 = \{(0, 1, 2), (0, 2, 1), (1, 0, 2), (1, 2, 0), (2, 0, 1), (2, 1, 0)\} .$$

Each of the 3-histories generated from the time series given by (2) can be uniquely mapped into a symbol in $S_3$. For example, for $t = 2$, $\bar{x}_2 = (9, 7, 6)$, we have that $x_{r+2} = 6 < x_{r+1} = 7 < x_{r+0} = 9$, which implies that $\bar{x}_2$ is of $(2, 1, 0)$-type.

Some symbols may have a higher occurrence than others in a given time series, encapsulating the recurrence of ordinal patterns in the time series.

III. USING SRP TO ANALYZE DYNAMIC STRUCTURE

In this section, we show the usefulness and power of SRP in analyzing and describing different dynamical aspects of a dynamical system. To this end, we use real-valued time series from continuous and discrete deterministic dynamical

![FIG. 1. Illustration of the time series $\{x_i\}$ of seven values given in (2).](image1)

![FIG. 2. Symbolization of the time series of seven values given by (2) for $m = 3$. From left to right, we find the five 3-histories that can be constructed and its associated symbol.](image2)
systems and two stochastic processes. The deterministic dynamical systems are as follows:

(a) Hénon system (discrete), where \( a, b \in \mathbb{R} \):
\[
(x_{t+1}, y_{t+1}) = (1 - ax_t^2 + y_t, bx_t).
\] (3)

(b) Lorenz system (continuous), where \( a, r, b \in \mathbb{R} \):
\[
(x, y, z) = (\sigma(y-x), x(r-z) - y, xy - bz).
\] (4)

The stochastic processes are as follows.

(c) White noise:
\[
\varepsilon_t \sim \text{i.i.d. } N(0, 1).
\]

(d) Autoregressive process of order one AR(1):
\[
x_t = 0.8x_{t-1} + \varepsilon_t,
\]
with \( \varepsilon_t \sim \text{i.i.d. } N(0, 1) \).

We have chosen the \( x \)-component of an orbit in each one of the deterministic cases as the time series. The length of the time series has been chosen as \( T = 1000 \) and an embedding dimension \( m = 4 \). Figures 3, 4, 5, and 9 show the SRP of the embedded time series that we have considered. We can see that the color is important. When the color is present, we can distinguish between the repetition of pattern of symbols and the time interval that the dynamical system spends in every element of symbolic partition of \( \mathbb{R}^4 \).

Figure 3(b) shows the SRP of the time series (the \( x \)-component) of a purely periodic orbit of a Hénon system. This SRP consists of diagonal lines separated by the temporal distance \( \tau = 7 \), where \( \tau \) is the period. This is due to the fact that, after one period, the position of the system in phase space is exactly the same and, therefore, we find identical recurrence. In addition, the color tells us how the embedded time series visits the different regions of symbolic partition of \( \mathbb{R}^2 \).

Moreover, in relation to the Henon system, Fig. 4(b) shows the SRP of an embedded time series from the first component of a chaotic orbit. Note that, in this case, symbolic recurrence states form vertical lines and diagonal lines with colored dots in the same order. Again, this plot shows that the color plays an important role in symbolic recurrence analysis.

In Figs. 5 and 6, we find the SRP of a periodic and a chaotic time series from the Lorenz system, respectively. As will commonly be the case in continuous dynamical systems, mainly two different symbols are represented in the SRP: those that provide increasing and decreasing ordinal patterns of the considered \( m \)-histories.

Figure 5(b) shows the SRP obtained from the \( x \)-component of a periodic orbit of the Lorenz system (with parameters \( r = 150 \), \( \sigma = 10 \), and \( b = 8/3 \)). As can be seen, the blue square represents the set of symbolic recurrence points corresponding to the increasing ordinal pattern, whereas the red square corresponds to the decreasing ordinal pattern. In addition, the size of the blue and red squares indicates for how long the time series is increasing and decreasing, respectively, as can be checked by viewing the plot of the time series in Fig. 7.

In contrast to the periodic orbit of the Lorenz system, in Fig. 6, we see that this SRP is more irregular (many different sizes of the symbolic recurrence squares). Nevertheless, as
Fig. 6 shows, the SRP exhibits some structure that represents the part of the orbit that is being visited. This structure is separated by six vertical bands. The first and third bands correspond to symbolic recurrence points when the orbit lies in the left wing of the attractor (in blue color). The second band corresponds to symbolic recurrence points when the orbit starts to leave the left wing, visits the right wing, and returns to the left wing three times, and then remains in the left wing [that is why the main diagonal of the square $[256, 525] \times [256, 525]$ is formed by six colored rectangles corresponding to the increasing (blue) and decreasing (red) symbols]. Similar to the second band, the fourth band corresponds with symbolic recurrence points when the orbit exits the left wing and visits the right wing twice [that is why the main diagonal of the square $[701,900] \times [701,900]$ is formed by five colored rectangles corresponding to the increasing (blue) and decreasing (red) symbols] and the third time the orbit exits the left wing and remains in the right wing whose recurrences correspond with the fifth band. This dynamic behavior is consistent with the plot of the time series given in Fig. 8.

### IV. MEASURES ON A SYMBOLIC RECURRENCE PLOT

The recurrence pattern in an SRP is linked with a specific dynamic behavior of the time series under consideration. To go beyond the graphical interpretation of SRPs, some descriptive measures have been proposed.

We start with two global recurrence measures: the symbolic recurrence rate (SRR) and the recurrence rate to a given symbol.

1. The SRR measures the probability that a symbolic recurrence point occurs. It is computed by

\[
SRR = \frac{1}{n^2} \sum_{i,j=1}^{n} SR_{ij}.
\]

2. The recurrence rate of a symbol $\pi$ measures the probability that a symbolic recurrence point to the symbol $\pi \in S_m$ occurs. It is computed by

\[
SRR(\pi) = \frac{\sum_{i,j=1}^{n} SR_{ij}(\pi)}{\sum_{i,j=1}^{n} SR_{ij}},
\]

where

\[
SR_{ij}(\pi) = \begin{cases} 
1 & \text{if } s(\bar{x}_i) = s(\bar{x}_j) = \pi, \\
0 & \text{otherwise}.
\end{cases}
\]
Therefore, SRR(π) measures the amount of color c_π in the SRP corresponding to symbol π.

As an example of the information given by these measures in Tables I and II, we have computed SRR and SRR(π) for all π ∈ S_4, respectively, for the examples shown Sec. II. We assume a fixed embedding dimension m = 4 and the length of the time series as T = 1000.

As expected, a larger the value of the SSR is obtained for the deterministic time series (see Table I). For the two stochastic processes, the largest value is obtained for the more structured time series, namely the AR(1) process. Therefore, a large symbolic recurrence rate is associated with a higher dynamic structure in the time series, whereas a low recurrence rate is associated with a stochastic system converging to \( \frac{1}{m} (\text{in } 0.0416) \) in the i.i.d. case (Caballero et al., 2017).

Table II shows the percentage of symbolic recurrence points for the 24 symbols of S_4 for the considered time series [SRR(π) × 100%]. As can be seen, a different behavior exists between the deterministic and stochastic processes. In the i.i.d. case, all symbols have essentially the same probability of recurrence. For AR(1), all symbols have a positive probability of symbolic recurrence although in this case the symbols corresponding to the increasing and decreasing order have the largest probabilities. In contrast to the stochastic processes, for the deterministic processes, either chaotic or periodic, we find that not all symbols have a positive probability of recurrence. This fact seems to be a characteristic that differentiates the stochastic and deterministic time series.

In an SRP, the symbolic recurrence points can be grouped into vertical, horizontal, and diagonal lines. The remaining points will be called isolated. All of them provide information about the dynamics of the time series. Now we define measures for all these symbolic recurrence structures.

- A **vertical line** of length q starting at \((i, j)\) occurs when \(SR_{i+k} = 1\) for \(k = 0, 1, \ldots, q - 1\). We denote by \(L^v(q)\) the number of vertical lines of length q in an SRP. A **horizontal line** of length q starting at \((i, j)\) occurs when \(SR_{i+k} = 1\) for \(k = 0, 1, \ldots, q - 1\). Hence, in a diagonal line two consecutive symbolic recurrence points are of different color. We denote the number of diagonal lines of length q in a symbolic recurrence plot by \(L^d(q)\).

- An isolated symbolic recurrence point at location \((i, j)\) occurs when \(SR_{ij} = 1\) and \(SR_{ij} \cdot SR_{eq} = 0\) where \(p = i - 1, i, i + 1, q = j - 1, j, j + 1\), and both \(p \neq i\) and \(q \neq j\). We denote it by \(I_{ij}\).

Note that, owing to the symmetry in the SRP horizontal and vertical lines are related. More concretely, if we find a horizontal line of length q starting at location \((i, j)\), then we will also find a vertical line of the same length q starting at location \((j, i)\). Therefore, in what follows, we will only compute the vertical lines. Other interesting measures that provide information on the dynamics of the time series derived from the previous definitions are as follows.

- The percentage of symbolic recurrence points that form diagonal lines of lengths between \(q_0\) and \(q_1\) over all symbolic recurrence points

\[
DR_{q_0}^{q_1} = \frac{\sum_{q=q_0}^{q_1} qL^d(q)}{n^2\text{SRR}} \times 100\%.
\]

- The percentage of symbolic recurrence points that form vertical lines of lengths between \(q_0\) and \(q_1\) over all symbolic recurrence points

\[
VR_{q_0}^{q_1} = \frac{\sum_{q=q_0}^{q_1} qL^v(q)}{n^2\text{SRR}} \times 100\%.
\]
The average length of vertical lines of lengths between \( q_0 \) and \( q_1 \) is given by

\[
\text{AVR}_{q_0}^{q_1} = \frac{\sum_{q=q_0}^{q_1} q L^v(q)}{\sum_{q=q_0}^{q_1} L^v(q)}.
\]

The average length of diagonal lines of lengths between \( q_0 \) and \( q_1 \) is given by

\[
\text{ADR}_{q_0}^{q_1} = \frac{\sum_{q=q_0}^{q_1} q L^d(q)}{\sum_{q=q_0}^{q_1} L^d(q)}.
\]

The percentage of symbolic recurrence points that are isolated over all symbolic recurrence points

\[
\text{IR} = \frac{\sum_{i,j=1}^{n} I_{ij}}{n^2 S_{\text{RPP}}} \times 100\%.
\]

In Table III, we list the values of the above-defined measures together with the maximum length of vertical and diagonal lines for the six time series described in Sec. III with line lengths between \( q_0 = 2 \) and \( q_1 = n = 997 \). The measures computed in Table III reveal strong differences between deterministic and stochastic processes, and within the deterministic processes, between discrete and continuous processes (see Figs. 10, 11, 12, 13, and 14). More concretely, the percentage of isolated symbolic recurrence points (IR) is the measure that better distinguishes between deterministic and stochastic processes, taking this measure as the highest value in the stochastic processes. When accounting only for the deterministic processes, the percentage of recurrence of points forming vertical lines (respectively, diagonal lines) has the highest value in continuous (respectively, discrete) systems. Finally, when comparing stochastic processes, the i.i.d. processes have a higher percentage of isolated symbolic recurrence points (IR) and a lower percentage of symbolic recurrence points forming vertical lines (VR) than a structured stochastic process [AR(1)].

The results in Table III are summarized in Figs. 10, 11, 12, 13, and 14. More concretely, Fig. 10 illustrates the graphs corresponding to SRP (a), recurrence points forming vertical lines (b), recurrence points forming diagonal lines (c), and isolated points (d) of the chaotic Henon system (3) for \( a = 1.4, b = 0.3 \) and initial condition \( x = 0.1, y = 0.15 \) with \( m = 4 \). The graphs of the periodic Henon orbit are not shown because the SRP and the plot of diagonal lines are the same and there are no vertical nor isolated points [see Fig. 3(b)]. Figures 11(a) and 11(b) show the SRP and the plot of diagonal lines for a periodic Lorenz orbit, respectively, and Figs. 12(a) and 12(b) show the SRP and the plot of diagonal lines for the chaotic orbit, respectively. As there are no isolated symbolic recurrence points in the periodic or chaotic Lorenz orbits, we do not show the corresponding plot. Moreover, plots of vertical lines are not shown because they are undistinguishable from the SRP ones.

<table>
<thead>
<tr>
<th>Measures</th>
<th>VR (%)</th>
<th>IR (%)</th>
<th>DR (%)</th>
<th>AVR</th>
<th>ADR</th>
<th>max V</th>
<th>max D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lorenz chaos</td>
<td>99.87</td>
<td>0</td>
<td>0.127</td>
<td>27.55</td>
<td>2</td>
<td>69</td>
<td>2</td>
</tr>
<tr>
<td>Lorenz periodic</td>
<td>99.68</td>
<td>0</td>
<td>0.31</td>
<td>22.34</td>
<td>2</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Hénon chaos</td>
<td>3.063</td>
<td>10.97</td>
<td>85.96</td>
<td>2.22</td>
<td>3.6</td>
<td>4</td>
<td>129</td>
</tr>
<tr>
<td>Hénon periodic</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>498.25</td>
<td>0</td>
<td>997</td>
</tr>
<tr>
<td>AR(1)</td>
<td>26.66</td>
<td>29.01</td>
<td>44.33</td>
<td>2.7</td>
<td>2.46</td>
<td>8</td>
<td>66</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>4.19</td>
<td>44.92</td>
<td>50.88</td>
<td>2.36</td>
<td>2.50</td>
<td>3</td>
<td>311</td>
</tr>
</tbody>
</table>

In Table III, we list the values of the above-defined measures together with the maximum length of vertical and diagonal lines for the six time series described in Sec. III with line lengths between \( q_0 = 2 \) and \( q_1 = n = 997 \). The measures computed in Table III reveal strong differences between deterministic and stochastic processes, and within the deterministic processes, between discrete and continuous processes (see Figs. 10, 11, 12, 13, and 14). More concretely, the percentage of isolated symbolic recurrence points (IR) is the measure that better distinguishes between deterministic and stochastic processes, taking this measure as the highest value in the stochastic processes. When accounting only for the deterministic processes, the percentage of recurrence of points forming vertical lines (respectively, diagonal lines) has the highest value in continuous (respectively, discrete) systems. Finally, when comparing stochastic processes, the i.i.d. processes have a higher percentage of isolated symbolic recurrence points (IR) and a lower percentage of symbolic recurrence points forming vertical lines (VR) than a structured stochastic process [AR(1)].

The results in Table III are summarized in Figs. 10, 11, 12, 13, and 14. More concretely, Fig. 10 illustrates the graphs corresponding to SRP (a), recurrence points forming vertical lines (b), recurrence points forming diagonal lines (c), and isolated points (d) of the chaotic Henon system (3) for \( a = 1.4, b = 0.3 \) and initial condition \( x = 0.1, y = 0.15 \) with \( m = 4 \). The graphs of the periodic Henon orbit are not shown because the SRP and the plot of diagonal lines are the same and there are no vertical nor isolated points [see Fig. 3(b)]. Figures 11(a) and 11(b) show the SRP and the plot of diagonal lines for a periodic Lorenz orbit, respectively, and Figs. 12(a) and 12(b) show the SRP and the plot of diagonal lines for the chaotic orbit, respectively. As there are no isolated symbolic recurrence points in the periodic or chaotic Lorenz orbits, we do not show the corresponding plot. Moreover, plots of vertical lines are not shown because they are undistinguishable from the SRP ones.

<table>
<thead>
<tr>
<th>Measures</th>
<th>VR (%)</th>
<th>IR (%)</th>
<th>DR (%)</th>
<th>AVR</th>
<th>ADR</th>
<th>max V</th>
<th>max D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lorenz chaos</td>
<td>99.87</td>
<td>0</td>
<td>0.127</td>
<td>27.55</td>
<td>2</td>
<td>69</td>
<td>2</td>
</tr>
<tr>
<td>Lorenz periodic</td>
<td>99.68</td>
<td>0</td>
<td>0.31</td>
<td>22.34</td>
<td>2</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Hénon chaos</td>
<td>3.063</td>
<td>10.97</td>
<td>85.96</td>
<td>2.22</td>
<td>3.6</td>
<td>4</td>
<td>129</td>
</tr>
<tr>
<td>Hénon periodic</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>498.25</td>
<td>0</td>
<td>997</td>
</tr>
<tr>
<td>AR(1)</td>
<td>26.66</td>
<td>29.01</td>
<td>44.33</td>
<td>2.7</td>
<td>2.46</td>
<td>8</td>
<td>66</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>4.19</td>
<td>44.92</td>
<td>50.88</td>
<td>2.36</td>
<td>2.50</td>
<td>3</td>
<td>311</td>
</tr>
</tbody>
</table>
Figures 13 and 14 illustrate the SRP (a) and the symbolic recurrence points forming vertical lines (b), symbolic recurrence points forming diagonal lines (c), and isolated points (d) for the i.i.d. and AR(1) processes, respectively.

V. EXAMPLES OF APPLICATIONS OF SRP

In this section, we show some examples of applications of the tools described in Secs. II–IV for the study of dynamical systems. We focus on the transient behavior, coexistence of attractors, bifurcation diagram, and structural change.

A. Transient behavior

In this section, we first explore the transient behavior to an asymptotic periodic orbit for the following dynamical system that has been proposed by Liu et al. (2004)

\[
\begin{align*}
\dot{x} &= 10(y - x) \\
\dot{y} &= 40x - xz + 5.81 \\
\dot{z} &= -2.5z + 4x^2.
\end{align*}
\]

We have simulated the orbit with the initial condition \((x_0, y_0, z_0) = (4, 4, 28)\) of length 9500. As shown in Fig. 15(a), the simulated orbit has a long transient behavior before it reaches its periodic asymptotic orbit [see Fig. 15(b)].

Next, we show how SRR, VR, and DR help us to understand the moment in which the orbit leaves the transient behavior to enter the limit periodic behavior. To this end, we have calculated SRR, VR, and DR with a moving windows procedure. More concretely, for the computations we have considered the first component of the simulated orbit with the initial condition \((4, 4, 28)\), an embedding dimension \(m = 4\), window of length \(w = 500\), and the following window is calculated by disregarding the first \(nw = 10\) observations of the previous one. The length of the time series is chosen to be 9000. Figure 16 illustrates the evolution of SRR, VR, and DR for each one of the 901 windows. As can be seen, the three measures find that the first window such that the whole window belongs to the periodic asymptotic behavior is the number 520.

B. Parameter instability in a dynamic model

In this section, we show how the set of symbolic recurrence measures SRR, VR, DR, and IR can identify the period in which a parameter defining the data-generating process...
changes. To this end, we have used two time series of length 4000, one coming from a stochastic model and one coming from a deterministic model, where the change in the parameter was imposed at $t = 2000$ in both cases.

More concretely, the stochastic time series comes from a simple AR(1) where the first 2000 values are generated with an autoregressive parameter of 0.8 and the last 2000 observations were generated with an autoregressive parameter of 0.3. We have used the same moving windows procedure as in Sec. VA, with an embedding dimension of $m = 4$, windows of length $w = 500$, and the following window is calculated by disregarding the first 10 observations of the previous one.

Figure 17 illustrates how SRR and VR detect the structural change in the time series when passing from the AR(1) with autoregressive parameter 0.8 to the AR(1) with autoregressive parameter 0.3. More concretely, the first window containing points of the two processes is the window number 151 where a decreasing trend of SRR and VR starts and it finishes in the last window containing points of both autoregressive models, namely window number 199. From window 200 to 350, the data come from an AR(1) with autoregressive parameter 0.3 and therefore the SRR (respectively, VR) is lower than in the first 150 windows where the data come from a more structured process, the AR(1) with autoregressive parameter 0.8.

The deterministic time series comes from the well-known logistic dynamical system $x_t = ax_t(1 - x_{t-1})$, where the first 2000 values are generated with $a = 3.827$, that is, a chaotic orbit, and the last 2000 values are generated with $a = 3.829$, that is, a periodic cycle. The same moving windows procedure applies in this case.

As in the stochastic case, Fig. 18 illustrates how SRR and IR distinguish between the windows containing data points from both logistic models. We find a similar behavior as in the stochastic case. An increasing (respectively, decreasing) trend of SRR (respectively, IR) is found from windows 151 to 199 where the data contain points of the two

![Figure 15](image1.png)  
FIG. 15. Transient behaviour (a) and asymptotic periodic behaviour of the orbit of Liu’s system (5) with initial condition (4; 4; 28) and $m = 4$.  

![Figure 16](image2.png)  
FIG. 16. SRR (a), VR (b), and DR (c) computed with a moving window procedure ($w = 500$, $nw = 10$) of the orbit of Liu’s system (5) with initial condition (4, 4, 28) with $m = 4$.  

![Figure 17](image3.png)  
FIG. 17. SRR (a) and VR (b) computed with a moving window procedure ($w = 500$, $nw = 10$), for a time series of length 4000 where the first 2000 observations come from an AR(1) with autoregressive parameter 0.3 and the last 2000 come from an AR(1) with autoregressive parameter 0.8, using $m = 4$.  

![Figure 18](image4.png)  
FIG. 18. SRR (a) and IR (b) computed with a moving window procedure ($w = 500$, $nw = 10$), for a time series of length 4000 where the first 2000 observations come from the logistic model with parameter $a = 3.827$ and the last 2000 for $a = 3.829$ with $m = 4$.  

logistic orbits (the chaotic and the periodic orbits). All data contained in windows from 200 to 350 come from the highly structured orbit (the periodic orbit) and that is why the SRR (respectively, IR) is higher (respectively, lower) than in windows before window number 150 where the time series comes from a chaotic orbit.

C. Bifurcation diagram

Trulla et al. (1996) and Iwanski and Bradley (1998) showed that RQA can detect changes in the dynamic behavior of the logistic dynamical system and Lorenz system, respectively. Now, we show how symbolic recurrence measures can also be used to construct symbolic bifurcations diagrams in the same way as bifurcation diagrams.

In Fig. 19(a), we illustrate the classical bifurcation diagram and the symbolic bifurcations diagrams [using SRR in (b) and IR in (c)] for the logistic model when the parameter $a$ varies from 3 to 4.

As can be seen, symbolic bifurcation diagrams also detect the dynamic changes of the logistic model when the parameter $a$ varies. Note that the periodic windows and the chaotic windows are consistently detected and coincide with those detected by the classical diagram. For instance, when the parameter $a$ belongs to a periodic (respectively, chaotic) window, the SRR takes regular constant (respectively, irregular) values whereas the IR tends to vanish (respectively, increases).

D. Coexistence of attractors

For a dynamical system described by a set of autonomous ordinary differential equations, an attractor can be a point, a periodic cycle, or even a strange attractor. Sprott et al. (2013) provided a dynamical system such that for a certain initial condition, in addition to the point attractor and chaotic attractor, the system also has a coexisting stable limit cycle. This dynamical system is given by

\[
\begin{align*}
\dot{x} &= yz + 0.01 \\
\dot{y} &= x^2 - y \\
\dot{z} &= 1 - 4x. 
\end{align*}
\]

This system has the stable equilibrium at $P = (1/4, 1/16, -0.16)$. Moreover this stable equilibrium can coexist with a periodic orbit or a strange attractor depending on the initial condition. For example, if the initial condition is $(0.3, 0, 0)$, the system goes to the stable equilibrium $P$, but if the initial condition is $(0.5, 0, 0)$, this system has a strange attractor (see Fig. 20). Thus, to show the usefulness of the symbolic recurrence measures in the analysis of coexistence of attractors, for each initial condition $(x_0, 0, 0)$ with $x_0$ from 0.15 to 1 at steps of 0.00085, we have simulated an orbit and we consider the last 1000 observations to ensure that the orbit is not in the transient state. Then, for each of these orbits, we have calculated SRR with an embedding dimension $m = 4$. This procedure is illustrated in Fig. 21, which shows that approximately at $x_0 = 0.49$, the limit behavior changes from one attractor to a different one.

On the other hand, for the initial conditions $(1, 1, 0.7)$, the Sprot dynamical system (6) has a chaotic attractor, whereas for the initial condition $(1, 1, 0.5)$, the system has an asymptotic periodic behavior (see Fig. 22). Using a similar
procedure as above but changing the initial condition \((1, 1, z_0)\) with \(z_0\) from 0.3 to 0.7 at steps of 0.0004, Fig. 23 shows the values of SRR for embedding dimension \(m = 4\). The change from the chaotic attractor to the periodic limit attractor and vice versa is detected approximately.

VI. SYMBOLIC VERSUS CLASSICAL RQA

The dynamic behavior of the logistic function has been widely studied in the specialized literature allowing nowadays to know in terms of the logistic parameter \(a\) where the transition boundaries are located. Indeed in Trulla et al. (1996), this analysis is carried out with the classical RQA. Therefore, in this section, we have chosen the logistic map to compare classical RQA with symbolic RQA. Moreover, we highlight the importance of the selection of the distance parameter \(\varepsilon\) showing how small variations in this parameter may obscure the understanding of the dynamic of the time series.

Following Trulla et al. (1996), we have simulated a transient time series consisting of 120 001 points from the first order logistic difference equation by consistently increasing parameter \(a\) in steps of 0.00001 on each iteration. For this data generating process, Trulla et al. (1996) give two recurrence plots where one can see the transition boundaries at different values of the parameter \(a\) of the logistic map (see Fig. 23, p. 257). We have replicated this figure for two different values of \(\varepsilon\), the value used in original paper \(\varepsilon = 0.5\sigma\) and a second value of \(\varepsilon = \sigma\) (where \(\sigma\) is the standard deviation of
the data). This comparison allows for appreciating how sensitive the RP analysis is to the choice of this parameter.

Together with the classical RP, we show the SRP on the same data set. In Figs. 24 and 25, we can see that for both, the SRP and RP, the transition boundary is delimited by the large increase in recurrences starting at the middle of the plot ($a = 3.82907$). Interestingly, the zoom for SPR not only confirms the transition boundary, but also allows us to emphasize the importance of colors. Notice that different colors are very informative since each colored dot indicates in which part of the symbolic partition of $\mathbb{R}^3$ the embedded time series is located.

Figures 26 and 27 illustrate the transition boundaries for the classical RP and the SRP at the periodic window (between $a = 3.90254$ and $a = 3.91053$) at the center of the plot. Notice that the symbolic and the classical recurrence plots detect these changes. Nevertheless, Fig. 26 (right) shows that when the noisy parameter $\epsilon$ is not selected properly, the detection of the transition boundaries becomes less evident.

We have also replicated (see Fig. 28, left panel) Fig. 4 of Trulla et al. (1996). This figure represents the percentage of recurrence computed as in Trulla et al. (1996) (i.e., repeatedly performed with transients embedded in 3-space on 800-point epochs). The left panel of Fig. 28 illustrates the strong alterations at the bifurcation points when $\epsilon = 0.5\sigma$. Notice however that when the distance cutoff parameter is slightly changed to $\epsilon = \sigma$, some information on the presence of periodic windows is clearly lost (see Fig. 28, right panel). Our equivalent symbolic quantity is the percentage of symbolic recurrences (see the right panel of Fig. 29). According to this figure, bifurcations points are already detected, but now without the necessity of choosing the cutoff parameter.

Finally, we have considered how to increase the precision of the tool. The right panel of Fig. 29 hints that if the symbolic recurrence tool user wants to make more precise the detection of periodic windows, she has only to increase the embedding dimension.

VII. APPLICATION TO REAL DATA

In this section, we use the SRR to a symbol $\pi$, namely, SRR($\pi$), defined in Sec. IV, to study the dynamics of heart rate variability (HRV). The electrocardiography (ECG) record has been obtained from the Physiobank Database (Goldberger et al., 2000) where many of the databases currently in the PhysioBank Archives were developed at MIT and at Boston’s Beth Israel Hospital.

HRV represents the time interval between two consecutive heartbeats in seconds, namely the RR interval. We have selected the 40 time series from the Fantasia Database, 20 young (21–34 years old) and 20 elderly (68–85 years old) rigorously screened healthy subjects. Each subgroup of subjects includes equal numbers of men and women. All subjects remained in a resting state in sinus rhythm while watching the movie Fantasia (Disney, 1940) to help maintain wakefulness.
On the other hand, we have used 26 time series from the Congestive Heart Failure (CHF) database and 40 time series from Arrhythmia (MIT-BIH), subjects aged from 34 to 79, men and women, gender unknown. All considered time series correspond approximately to a half an hour of ECG.

We have calculated the SRR of the symbols increasing $\pi_i = (0, 1, 2, 3)$ and decreasing $\pi_d = (0, 1, 2, 3)$ measured for each one of the 106 (40 healthy people, 26 CHF people and 40 people with arrhythmia) time series under consideration. Table IV reports Welch's t-test and confidence
TABLE IV. Confident intervals and p-value of Welch’s t-test at 5% confidence level for SRR(\(\pi\)) and SRR(\(\pi_d\)) between healthy people and people with congestive heart failure (H-CHF), and between healthy people and people with arrhythmia (H-A).

<table>
<thead>
<tr>
<th>Measure</th>
<th>p-value</th>
<th>Confidence interval 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRR((\pi_d))</td>
<td>1.194 \times 10^{-6}</td>
<td>[0.0039, 0.0084]</td>
</tr>
<tr>
<td>SRR((\pi_d))</td>
<td>0.4251</td>
<td>[-0.0015, 0.0034]</td>
</tr>
<tr>
<td>SRR((\pi_i))</td>
<td>0.0095</td>
<td>[-0.0159, -0.0023]</td>
</tr>
<tr>
<td>SRR((\pi_i))H-A</td>
<td>0.0030</td>
<td>[0.0027, 0.0128]</td>
</tr>
</tbody>
</table>

Intervals for the mean differences of SRR(\(\pi\)) between unwell and healthy people, with \(\pi\) the two symbols described above.

The results show that with these two heart diseases, the mean value of SRR(\(\pi_i\)) differs between unwell and healthy people at a nominal level of 5%.

Given the good result provided by the symbolic recurrences to \(\pi_i\), and in order to complete the analysis, in Fig. 30, we show the value of SRR(\(\pi_i\)) computed on 500-point epochs and shifted by 10 point for the next computation, with embedding dimension \(m = 4\). As expected SRR(\(\pi_i\)), clearly separates the healthy person from the unwell one in both heart diseases (congestive heart failure and arrhythmia) showing that SRR(\(\pi\)) could be a useful diagnosis tool.

VIII. CONCLUSIONS

Based on the concept of Symbolic Correlation Integral, we have defined a set of SRP measures that do not depend on the distance parameter and that serve as a tool for quantifying the dynamic structure. The potential of these measures has been demonstrated for studying transient behavior, the coexistence of attractors, bifurcations, and structural change in several well-known dynamical systems. An important advantage of this approach is that, for a fixed embedding dimension, symbolic measures are invariant.

We want to remark that we have constructed similar symbolic recurrence measures to the classical RQA measures that might be used in the analysis of the dynamic of a complex time series. We did not focus only in the well-known logistic function but also in other (continuous, discrete and even stochastic) complex models. The interpretation of the symbolic recurrence measures is different than the classical ones. This is mainly due to the fact that the classical RQA uses the geometry in the analysis by introducing the norm and the noisy threshold, \(\varepsilon\) while in the symbolic context this analysis relies only on the partition of \(\mathbb{R}^m\) provided by the \(m!\) symbols. Moreover, we have pointed out that the color of the dot identifies the specific set of the symbolic partition and how the orbit moves. Therefore, a deeper analysis of the colors in the symbolic recurrence plot may help the researcher to understand how the orbit behaves. Even more, given the embedding dimension, \(m\), the symbolic recurrence plot is unique and it is not required to look for a correct \(\varepsilon\) as in the classical RQA, which is in general a difficult task when the researcher does not know where the complexity of the observed time series is located. Our comparative study with other classical measures leads us to conclude that symbolic-based measures, like SRP and others, detect the same transition boundaries as the classical RP, provided that the user of the RP chooses a proper \(\varepsilon\).

From an applied point of view, the final user of these tools is alleviated from choosing the free distance parameter (that is required in other tools) and therefore avoids the cost of obtaining two different conclusions for the same data set. We have illustrated the use of symbolic recurrence measures by applying them to ECG data. The results show that symbolic recurrence measures are useful for discriminating between healthy and unwell patients and, therefore, could be a very useful diagnostic tool.

Finally, it is worth mentioning that, to the best of our knowledge, this is the first paper studying and evaluating the power performance of symbolic RQA and SRP to analyze the dynamic behavior of different complex time series coming from discrete, continuous and stochastic data generating processes. The work of Groth (2005) on symbolic recurrence analysis is limited to the analysis of dependencies between two time series by applying cross symbolic recurrence plots. Marwan et al. (2007) and Schinkel et al. (2007) apply...
symbolic RQA to the analysis of ECG data. More recently in Donner et al. (2008), the authors use symbolic recurrence plots for a quantitative investigation of the dynamics of discrete-valued inventory levels of cooperating firms in a manufacturing network. In these papers, there is a lack of a description of the behavior of: symbolic recurrences (per symbol), vertical lines, diagonal lines, isolated points, and the different colors representing the symbols (and thus the region of the symbolic partition of \( \mathbb{R}^m \) that is being visited). In this regard, another contribution of this paper is to systematically show that SR-measures are powerful and useful tools to detect transient, bifurcations, periodicities, and in general, to analyze the complex time series behavior. Moreover, given its theoretical support in terms of correlation integral, statistical hypothesis tests could be devised in the near future.

ACKNOWLEDGMENTS

This article is the result of the activity performed under Grant ECO2015-65637-P from the Ministerio de Economía y Competitividad de España. This article is the result of the activity performed under the program Groups of Excellence of the Region of Murcia, the Fundacion Séneca, Science and the Technology Agency of the region of Murcia Project 19884/GERM/15. All remaining errors are our responsibility.


Webber, Jr., C. L. and Marwan, N., Recurrence Quantification Analysis (Springer, 2015).
