Quantifying the diversity of multiple time series with an ordinal symbolic approach

Luciano Zunino [1,2,*] and Miguel C. Soriano [3,†]

[1]*Centro de Investigaciones Ópticas (CONICET La Plata - CIC - UNLP), C.C. 3, 1897 Gonnet, La Plata, Argentina*
[2]*Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina*
[3]*Instituto de Física Interdisciplinar y Sistemas Complejos CSIC-UIB, Campus Universitat de les Illes Balears,*
E-07122 Palma de Mallorca, Spain

The main motivation of this paper is to introduce the *ordinal diversity*, a symbolic tool able to quantify the degree of diversity of multiple time series. Analytical, numerical, and experimental analyses illustrate the utility of this measure to quantify how diverse, from an ordinal perspective, a set of many time series is. We have shown that ordinal diversity is able to characterize dynamical richness and dynamical transitions in stochastic processes and deterministic systems, including chaotic regimes. This ordinal tool also serves to identify optimal operating conditions in the machine learning approach of reservoir computing. These results allow us to envision potential applications for the handling and characterization of large amounts of data, paving the way for addressing some of the most pressing issues facing the current big data paradigm.

I. INTRODUCTION

It is a fact that the advent of new technologies has brought with it an unprecedented increase in the amount of available data to manage and analyze. Astronomy, climate science, physics and genomics are some of the scientific fields that currently need to cope with terabyte-sized data sets [1]. Very often, scientists from these and other disciplines cannot yet be fully successful in getting the best out of this modern deluge of data principally due to the complexity of the analysis task they are facing [2]. Consequently, the development of computationally efficient methods to handle this huge volume of information in a reliable and robust way is vastly sought for researchers from the time series analysis and data mining communities. Unveiling relevant and hidden information from large data sets is a challenge that deserves special consideration and any progress on this matter is welcome.

Particularly, in recent years, great effort has been made to measure the degree of similarity between time series [3–5]. Approaches able to quantify the similarity or dissimilarity of time series are especially suited to address time series clustering and classification problems with higher accuracy [4,6]. Most of these measures are designed to estimate the difference between two time series. However, a generalization of this concept for multiple time series is necessary for many applications for which the overall similarity or, otherwise, diversity among a finite number of time series provides a very useful piece of information. For example, assessing the degree of diversity of a test dataset is a relevant issue before its use for testing and evaluating the performance of a new method. This is due to the fact that only datasets with a high level of diversity enable a rigorous validation of any time series algorithm [7]. In the context of machine learning techniques

for time series processing, reservoir computing stands out as a powerful method for predicting and generating arbitrary dynamical systems [8]. For example, it has recently been shown that reservoir computers can be used to create a large number of different chaotic signals with desirable properties from a single chaotic system. Making these generated signals as diverse as possible is required to enhance its applicability [9]. Consequently, suitable strategies for quantifying and maximizing diversity are needed. Finally, minimizing the redundancy among different variables in multivariate time series analysis is another significant task that could be potentially addressed by implementing a suitable diversity measure. This approach might be useful to complement standard dimensionality reduction methods [10].

In this paper, we attempt to shed some light on this problem by introducing the *ordinal diversity*. This symbolic measure is defined as the generalized Jensen-Shannon divergence of the ordinal pattern probability distributions associated with the different time series of the dataset under analysis. On the one hand, the generalized Jensen-Shannon divergence is a distributional distance measure with important advantages over other approaches, such as flexibility for all distributional data types and intuitive theoretical interpretation [11,12]. On the other hand, the ordinal mapping of time series satisfies relevant properties that make it especially suitable for handling a high volume of complex data, namely, simplicity, low computational cost, wide applicability, and less susceptibility to outliers and artifacts [13,14]. Combining both concepts, the ordinal diversity straightforwardly inherits all these features and emerges as a potentially valuable tool for evaluating how diverse multiple time series are in practical contexts. We illustrate this through several analytical, numerical, and experimental analyses.

The remainder of the paper is structured as follows. In Sec. II, the ordinal diversity is introduced. Some analytical and numerical findings are detailed in Sec. III. Next, in Sec. IV, an experimental application is included to illustrate

*Corresponding author: lucianoz@ciop.unlp.edu.ar
†miguel@ifisc.uib-csic.es

the practical utility of the proposed measure in a real context. The main conclusions reached along this work are highlighted in Sec. V. For the sake of completeness, an Appendix is added to describe a complementary tool used to characterize the synchronization degree in the experimental application.

II. ORDINAL DIVERSITY: BACKGROUND AND DEFINITION

In the following, we start by revisiting the ordinal symbolic representation together with its main advantages. Then, the generalized Jensen-Shannon divergence is briefly introduced. Finally, the proposed measure for diversity is defined.

A. Ordinal symbolization

Since the publication, more than 20 years ago, of the foundational paper by Bandt and Pompe [15], the use of ordinal or permutation patterns for the analysis of time series has been continuously increasing [16–18]. The success of this coarse-grained representation can be principally attributed to the fact that the temporal structure of the underlying data is straightforwardly taken into account in the symbols [19], and to its wide applicability; i.e., datasets generated by unknown dynamic processes with unknown levels of noise can be robustly analyzed by implementing this approach [20]. Furthermore, this approach offers a natural and simple way to find an appropriate symbolic representation from a time series, with a weak condition of stationarity (order stationarity) that is often fulfilled in practice, at least approximately [14,21].

The ordinal framework essentially consists of mapping equidistant successive D values of a time series $\{x_t, x_{t+\tau}, \ldots, x_{t+(D-1)\tau}\}$ to one of the $D!$ possible permutations of the same size that describe the order relation between these elements. A symbol π_i with $i = 1, 2, \ldots, D!$ is then associated with each permutation and, consequently, the original time series $\{x_t\}_{t=1}^{t=L}$ of length L is mapped to an ordinal sequence $\{s_t\}_{t=1}^{t=L^*}$ of length $L^* = L - (D-1)\tau$ with $s_t \in S_D = \{\pi_1, \pi_2, \ldots, \pi_{D!}\}$ (the set of permutations of length D that defines the new alphabet). D is usually known as the order or embedding dimension and the time separation between the elements in the subsequence is the lag or embedding delay τ ($\tau = 1$ when considering consecutive points). Ordinal time series analysis is commonly based on the discrete probability distribution of the symbols,

$$P_\pi = \{p(\pi_i), i = 1, \ldots, D!\}, \quad (1)$$

with $p(\pi_i)$ the probability of each ordinal pattern estimated as the relative number of occurrences of them in the time series. The presence of any kind of temporal correlation in the data manifests itself in a nonuniform distribution of the ordinal patterns while equiprobability is obtained for a completely random sequence, i.e., for white noise. In order to robustly estimate P_π, the condition $L \gg D!$ should be satisfied, limiting the maximum value that can be chosen for the order D. Regarding the parameter of the embedding delay, it is commonly suggested to use $\tau = 1$. However, estimations with lagged data points, $\tau \geqslant 2$, are particularly useful for a better understanding of continuous and/or scale-dependent systems [14,22–24].

A word of caution should be raised when time series from high-dimensional dynamics are symbolized by using the ordinal mapping. On the one hand, as it has been carefully analyzed in Ref. [25], the order D has to exceed a lower bound $D_{\min}$ to successfully resolve the underlying temporal structures for data from high-dimensional systems. Actually, an almost uniform distribution of ordinal patterns is obtained if $D < D_{\min}$, making the distinction of these highly complex dynamics from purely random noise unfeasible. But, on the other hand, it is also true that the data length L is often limited in real situations and, consequently, the order D cannot be chosen arbitrarily large taking into account the aforementioned condition $L \gg D!$. As a result, the order D does not have to exceed an upper bound $D_{\max}$. Thus, any chosen pattern length in the range $[D_{\min}, D_{\max}]$ would guarantee a proper estimation of the ordinal pattern probability distribution. The major challenge faced when dealing with time series from high-dimensional dynamics is that $D_{\min}$ can be larger than $D_{\max}$, making it practically impossible to unfold the underlying complex structures.

It is worth noticing that amplitude threshold dependencies that negatively affect more conventional symbolization recipes based on range partitioning [26] are naturally circumvented with the Bandt and Pompe recipe, and, also, that the distribution of ordinal patterns is invariant with respect to strictly monotonically increasing transformations of the original data [27]. Finally, since the ordinal mapping requires sorted values, the presence of equal values in the subsequences deserves to be carefully considered, especially if the number of them is not negligible. This could be the case of, e.g., signals with low resolution. The estimation of P_π can be significantly biased if ties are frequent and the addition of a small random perturbation helps to mitigate this spurious effect [28].

B. Generalized Jensen-Shannon divergence

The Jensen-Shannon divergence between the distributions $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$ with weights ω_P and ω_Q ($\omega_P \geqslant 0$, $\omega_Q \geqslant 0$, and $\omega_P + \omega_Q = 1$), respectively, is defined [29] as

$$D_{JS}^{(\omega_P, \omega_Q)}(P, Q) = S(\omega_P P + \omega_Q Q) - \omega_P S(P) - \omega_Q S(Q), \quad (2)$$

with S the Shannon entropy function, i.e., $S(P) = -\sum_{i=1}^{n} p_i \ln p_i$ (as usual, we assume the convention that $0 \ln 0 = 0$). This divergence is a measure of discernibility between two probability distributions that has proven to be useful for heterogeneous applications such as the segmentation of nonstationary symbolic sequences (e.g., DNA sequences) into stationary subsequences [11,30], the definition of a statistical measure of complexity [31], and the statistical analysis of language [3], to name just a few of them.

But what is more important for our present concerns is the fact that this divergence measure can be easily generalized to quantify the overall difference among any finite number of probability distributions. The extension to M multiple distributions, usually known as generalized Jensen-Shannon

divergence (GJSD) [29], is given by

$$D_{JS}^{(\omega_1,\omega_2,\dots,\omega_M)}(P_1, P_2, \dots, P_M) = S\left(\sum_{i=1}^{M} \omega_i P_i\right) - \sum_{i=1}^{M} \omega_i S(P_i),$$

$$(3)$$

with P_i the probability distributions and ω_i positive weights such that $\sum_{i=1}^{M} \omega_i = 1$. The GJSD is bounded in the range $[0, -\sum_{i=1}^{M} \omega_i \ln \omega_i]$. The minimum is achieved if and only if all probability distributions under comparison are identical, while the maximum value is obtained whenever their supports are disjoints, i.e., when their nonzero probabilities are entirely nonoverlapping [12]. In particular, the maximum possible value for the GJSD is equal to $\ln M$ when the same weight is assigned to all the time series under analysis ($\omega_i = 1/M, \forall i$).

C. Ordinal diversity

Leveraging on the two previous concepts, the ordinal diversity (OD) is defined as the GJSD [Eq. (3)] calculated over the ordinal pattern probability distributions [Eq. (1)] of the multiple time series that are being contrasted. This measure quantifies the degree of diversity among the ordinal coarse-grained representations of the multiple signals under consideration. Intuitively, when the estimated ordinal distributions are all similar, the OD will be close to 0. In contrast, OD will be larger (significantly different from zero) if the associated distributions of symbols obtained via the ordinal mapping are relatively dissimilar from each other.

A related ordinal metric, the permutation Jensen-Shannon distance (PJSD), has been recently introduced as a versatile and robust tool for quantifying the degree of similarity between two arbitrary time series [5]. Briefly, it is defined as the square root of the Jensen-Shannon divergence between the ordinal probability distributions P_π and Q_π of the two signals under analysis with equal weights, i.e., $[D_{JS}^{(1/2,1/2)}(P_\pi, Q_\pi)]^{1/2}$. The exponent $1/2$ is included since it has been shown that $[D_{JS}^{(1/2,1/2)}(P, Q)]^{1/2}$ is a true metric for arbitrary probability distributions P and Q [32]. For further mathematical and technical details about the PJSD, we refer the interested reader to Ref. [5]. OD is essentially a generalization of the PJSD for more than two time series, which additionally enables one to assign appropriate weights to each ordinal probability distribution.

III. ANALYTICAL AND NUMERICAL ANALYSIS

Next, we resort to the analysis of well-known stochastic and chaotic dynamical systems for assessing the usefulness of the proposed tool within controlled scenarios. It is important to clarify here that equal weights were assigned to the multiple time series in all analyses included in this work.

A. Fractional Brownian motion

Fractional Brownian motion (fBm) is a generalization of Brownian motion that allows for long-range dependence and self-similarity of the random increments [33]. The so-called Hurst exponent H, which is a real number in (0, 1), characterizes the raggedness of the motion. If $H = 1/2$, fBm reduces
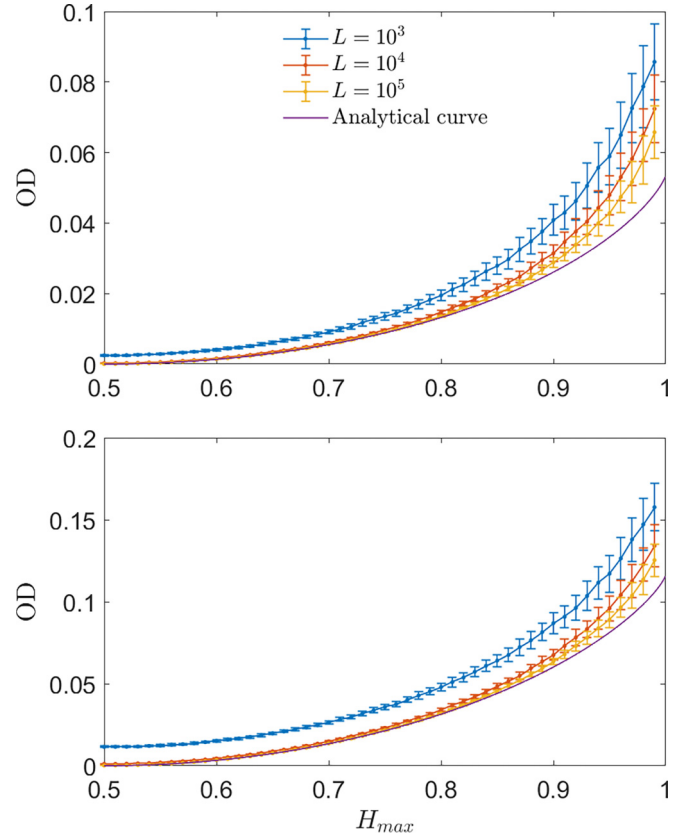

FIG. 1. OD analytically calculated for sets of $M = 100$ fBms with Hurst exponent values linearly equispaced in the range $[1 - H_{\max}, H_{\max}]$. Results for $D = 3$ (top plot) and $D = 4$ (bottom plot) are displayed. Mean and standard deviation (as error bar) obtained from OD estimations, with the same orders and lag $\tau = 1$, for 100 independent ensembles of the different sets of $M = 100$ numerical simulations with different data lengths L are also included for comparison purposes. It is observed that the OD increases monotonically as a function of $H_{\max}$ since time series with a wider range of long-range dependencies are included in the set. Besides, numerical curves get closer to the analytical counterpart as L increases.

to Brownian motion; if $H > 1/2$, the increments are positively correlated; if $H < 1/2$, the increments are negatively correlated. fBm has been successfully used to model fractal phenomena appearing in very heterogeneous fields such as hydrology, telecommunication, turbulence, image processing, and finance [34], and the practical characterization of this family of fractal stochastic processes remains as a topic of interest among researchers [35,36].

We start our analysis of the OD with the example of fBm as we can compare the results of the numerical estimates with the ground truth. Analytical probabilities of the ordinal patterns with orders $D = 3$ and $D = 4$ for fBm have previously been determined by Bandt and Shiha [37]. Thanks to the self-similarity property satisfied by this family of stochastic processes, the ordinal pattern probabilities are independent of the lag τ. There is, however, no closed analytical formula for orders $D \geqslant 5$. Making use of the analytical results for $D = 3$ and $D = 4$, the OD for sets of fBms with different H is plotted in Fig. 1. To generate these analytical curves, $M = 100$ fBms with H linearly equispaced in the range $[1 - H_{\max}, H_{\max}]$
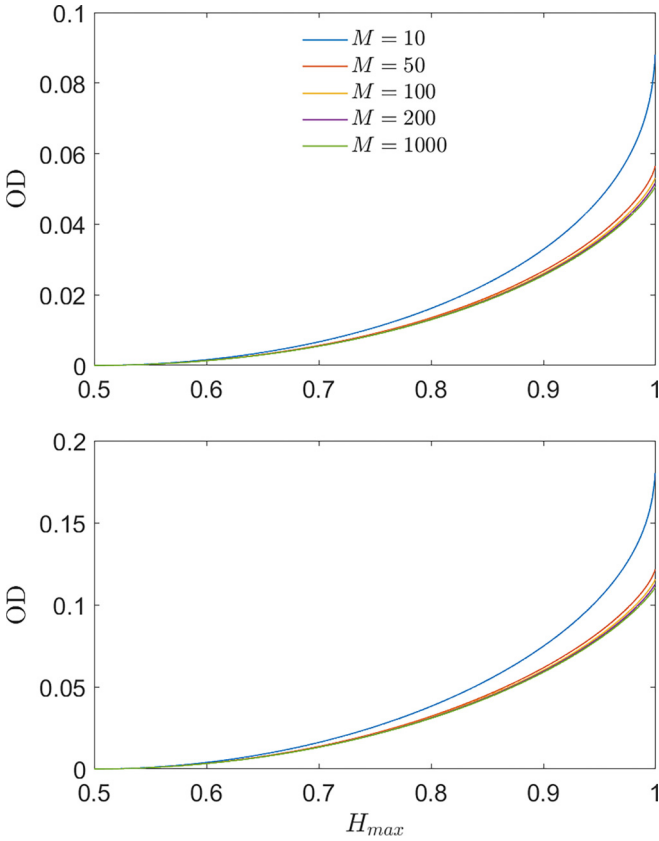
FIG. 2. OD analytically calculated for sets of M different fBms with Hurst exponent values linearly equispaced in the range $[1 - H_{max}, H_{max}]$. Results for $D = 3$ (top plot) and $D = 4$ (bottom plot) are displayed. Analytical curves collapse when the number M of different fBms in the set is larger than or equal to 100.

have been considered, where H_{max} is varied between 0.5 and 0.9995 with step 0.0005. In Fig. 1, OD estimations from numerical realizations have also been included. Mean and standard deviations (as error bars) for 100 independent ensembles of the different sets of $M = 100$ fBm realizations with $H_{max} \in \{0.5, 0.51, \ldots, 0.99\}$ and different data lengths L are displayed for comparison purposes. The simulated fBm time series were generated by implementing the method of Wood and Chan [38]. We show that the OD increases for increasing H_{max}. The numerical results become closer to the analytical curves as the length of the simulations increases.

We have also analyzed how the analytical curves change with the number M of different fBms in the set with H linearly equispaced in the range $[1 - H_{max}, H_{max}]$. As was done above, fBms with $H_{max} \in \{0.5, 0.5005, \ldots, 0.9995\}$ have been considered. Figure 2 shows results obtained for $D = 3$ (top) and $D = 4$ (bottom) for $M \in \{10, 50, 100, 200, 1000\}$. The analytical curves are almost indistinguishable for $M \geqslant 100$, which is the value that has been used in Fig. 1 and will be used in the remainder of the manuscript.

B. Logistic map

The logistic map is a recurrence relation of the form

$$x_{n+1} = rx_n(1 - x_n), \tag{4}$$
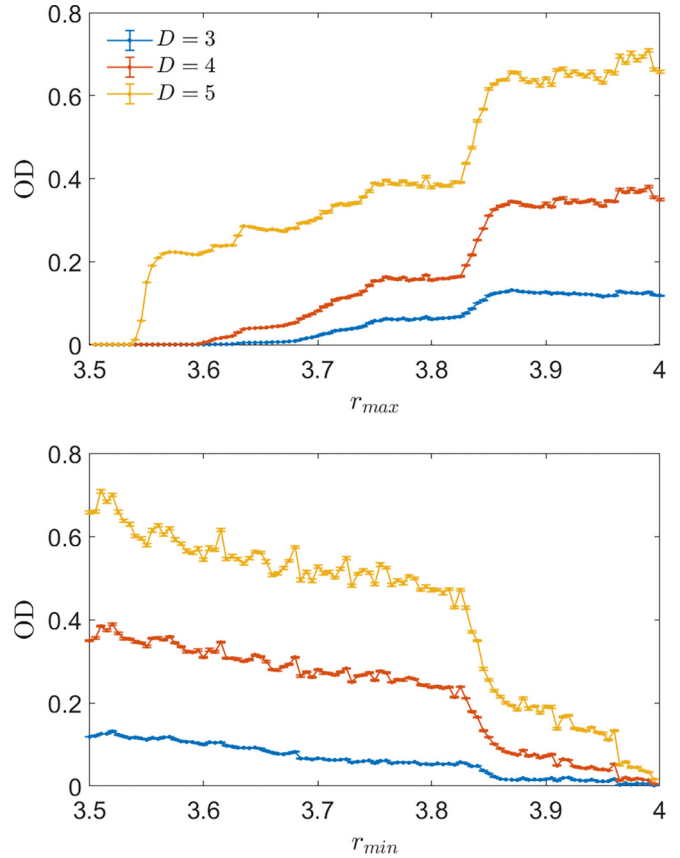

FIG. 3. OD estimated for sets of $M = 100$ realizations of the logistic map with values of the parameter r linearly equispaced between 3.5 and r_{max} (top plot) and between r_{min} and 4 (bottom plot). Mean and standard deviation (as error bar) obtained from OD estimations with orders $D \in \{3, 4, 5\}$ and lag $\tau = 1$ for 100 independent ensembles of the different sets of $M = 100$ numerical simulations are displayed. Realizations of length $L = 10^3$ data points (after discarding the first 10^4 generations to avoid the effect of transient behavior on the results) were generated for this numerical analysis. Results obtained for other data lengths are analogous. The OD is able to capture the dynamical richness and dynamical transitions of the logistic map, reaching higher values when time series from periodic and chaotic regimes coexist in the set.

where x_n is a number between 0 and 1 and r is a positive constant. The logistic map is often used as a simple model of population dynamics and chaotic behavior [39]. Depending on the value of r, the logistic map can exhibit different types of behavior, such as convergence to a fixed point when r is between 0 and 3, or oscillations between two or more values when r is between 3 and approximately 3.57. The logistic map undergoes a period-doubling route to chaos, leading to increasingly complex and unpredictable behavior. Chaotic behavior appears when r is between 3.57 and 4, with narrow windows of periodicity for particular values of r (e.g., a stable period-3 cycle emerges around $r \sim 3.83$). The logistic map is a paradigmatic example of how simple nonlinear equations can generate complex phenomena that are not easily predictable or controllable.

Figure 3 shows the OD numerically estimated for sets of $M = 100$ time series originating from Eq. (4). On the one

hand, Fig. 3 (top) presents the results obtained via the OD when the time series are generated with linearly equispaced r values in the range [3.5, r_{max}]. We find that the OD captures the dynamical changes present in the time series as r_{max} is increased. In particular, the OD starts to increase when the time series set under analysis includes oscillatory signals of different periods and continues to grow as chaotic time series are included in the set. The OD measure also captures the transition to periodic windows around $r \sim 3.83$ as this increases the dynamical richness in the time series set. On the other hand, Fig. 3 (bottom) focuses on the estimation of the OD when the time series are generated in the range [r_{min}, 4]. As r_{min} is decreased, the OD grows as soon as time series with different dynamics are included in the set under analysis. An abrupt jump in the OD is again found when the periodicity regime around $r \sim 3.83$ is present in the time series set. Overall, we find the largest OD for the time series generated by the logistic map when chaotic and oscillatory dynamics with different periods are combined.

C. Coupled Hénon maps

The Hénon map is a two-dimensional discrete-time dynamical system that can exhibit chaotic behavior [40]. Here, we consider two unidirectionally coupled identical Hénon maps with equations

$$x_1(t+1) = 1.4 - x_1^2(t) + 0.3x_2(t), \qquad (5)$$

$$x_2(t+1) = x_1(t), \qquad (6)$$

for the driving system X, and

$$y_1(t+1) = 1.4 - [Cx_1(t)y_1(t)$$
$$+ (1-C)y_1^2(t)] + 0.3y_2(t), \qquad (7)$$

$$y_2(t+1) = y_1(t), \qquad (8)$$

for the response system Y. This drive-response configuration, proposed in Ref. [41], has been widely used as a testbed for evaluating the performance of causality measures [42–45]. For the parameter values chosen in Eqs. (5)–(8), the dynamics of the systems exhibits chaos. The response system becomes identically synchronized to the drive as the coupling parameter C approaches unity.

Figure 4 presents the PJSD and the OD for the coupled Hénon maps, where we have considered 100 response systems with different values of the coupling. Each of these 100 realizations of the response system y_1 has a coupling strength C, which is randomly selected from a Gaussian distribution with mean $\mu = C_{mean}$ and standard deviation $\sigma = 0.05$ (qualitatively similar results are obtained for other values of σ). In Fig. 4 (top), we present the PJSD between the time series of the drive x_1 and each of the 100 different responses y_1. The PJSD measure increases for increasing values of the coupling in the range of C_{mean} [0, 0.5]. For $C_{mean} > 0.5$, the PJSD starts to decrease and eventually we find that the PJSD approaches 0 for larger coupling strengths. The low values of the PJSD for large couplings suggest that x_1 and y_1 are then identically synchronized. Figure 4 (bottom) shows the OD measure for the $M = 100$ independent responses y_1 and for different values of C_{mean}. We identify that the OD is maximal for low values of
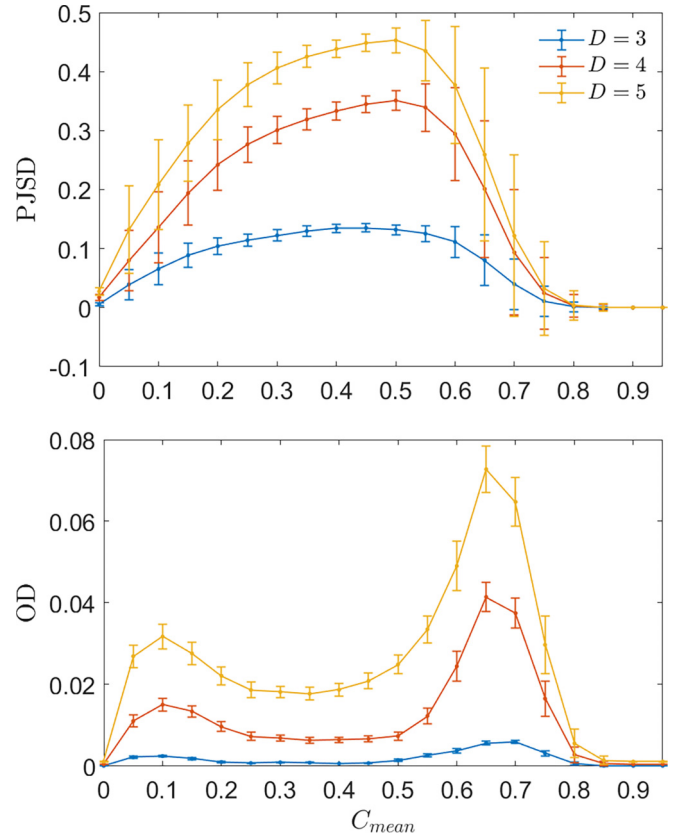

FIG. 4. Results obtained for the analysis of two unidirectionally coupled identical Hénon maps with coupling strength C randomly selected from a Gaussian distribution with mean $\mu = C_{mean}$ and standard deviation $\sigma = 0.05$. One hundred independent pairs of time series x_1 and y_1 were generated for each mean coupling strength C_{mean}. Top: PJSD between pairs of time series x_1 and y_1. Bottom: OD estimated for the set of $M = 100$ independent responses y_1. Mean and standard deviation (as error bar) obtained from estimations with orders $D \in \{3, 4, 5\}$ and lag $\tau = 1$ for 100 independent ensembles of the set of $M = 100$ numerical responses y_1 are displayed. Both ordinal quantifiers are plotted as a function of the increasing mean coupling strength C_{mean}. Realizations of length $L = 10^4$ data points (after discarding the first 10^4 generations to avoid the effect of transient behavior on the results) were generated for this numerical analysis. A noticeable maximum of the OD is reached at $C_{mean} \sim 0.65$, in the middle of the transition to identical synchronization.

the coupling strength $C_{mean} \sim 0.1$ and for intermediate values of $C_{mean} \sim 0.65$. It is particularly interesting to observe that the responses become more diverse in the transition to identical synchronization at intermediate values of C_{mean}. In order to better understand the different behaviors displayed by the two ordinal quantifiers on the coupling strength, it should be noted that PJSD quantifies the degree of ordinal similarity between drive x_1 and responses y_1, while the OD measures the overall difference among the ordinal coarse-grained representations of the 100 independent responses y_1. Thus, at $C_{mean} \sim 0.65$, the 100 responses reach the maximum ordinal diversity among them even when each of them is quite ordinally similar to the drive signal. In other words, for this coupling strength, the 100 responses y_1 are comparatively heterogeneous but, simultaneously, similar to the drive x_1 from an ordinal perspective.
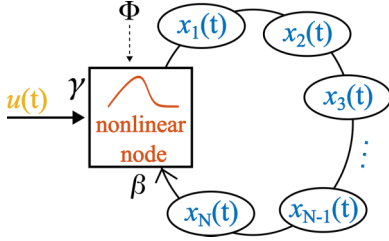
FIG. 5. Illustration of the experimental application. An input stream $u(t)$ drives the reservoir with input strength γ. The reservoir consists of a nonlinear node subject to delayed feedback, with β denoting the feedback strength. The operating point of the Mackey-Glass nonlinear node is controlled by the external bias Φ. The output of the nonlinear node is measured at different time shifts, $x_i(t)$, for the $N = 100$ virtual nodes.

IV. AN ILLUSTRATIVE EXPERIMENTAL APPLICATION

In this section, we focus on evaluating experimental data that originates from a reservoir computing implementation. Reservoir computing (RC) is a machine learning paradigm that exploits dynamical systems for time series forecasting and classification [46,47]. The computational core of RC is the so-called reservoir, which is typically an input-driven non-linear dynamical system. While the reservoir was originally considered to be a recurrent neural network, it is also possible to implement a reservoir with different physical substrates [48]. Main ingredients for operational reservoirs include the accessibility to a high-dimensional input-output mapping, the ability to perform nonlinear transformations, and the existence of a memory that fades over time [49]. In practice, these ingredients need to be tested with large amounts of data. Specific measures to properly quantify them are scarce and typically computationally demanding [50]. We propose here the OD as an efficient tool to identify favorable operating regimes for RC.

The data under analysis originate from a physical implementation that follows the paradigm of delay-coupled reservoir computing (delay RC) [51]. Instead of a network of discrete elements, delay RC employs only a single nonlinear node that is coupled to itself via a temporal feedback loop. A set of N *virtual* nodes is then created by considering the response of the nonlinear node at equidistant time slots, here $N = 100$. A scheme of the physical system is presented in Fig. 5, where an input stream drives the reservoir and N output signals are measured at different temporal locations. More details about the experimental implementation, which is a Mackey-Glass-based reservoir computer, can be found in Ref. [52].

We analyze the experimental data of $N = 100$ virtual nodes for an input strength $\gamma = 60$ and a feedback strength $\beta = 0.8$ over 20 different values of the bias Φ in the range [12, 725]. The dynamics of the experimental system, in the absence of input and for these parameter values, is typically chaotic for $\Phi < 237$, exhibits a window of periodic dynamics for intermediate values of Φ, and a single fixed point (stable dynamics) for $\Phi > 650$. In the following, we compute and analyze the PJSD and the OD for these data with the ultimate goal to relate the values of the quantifiers to the properties
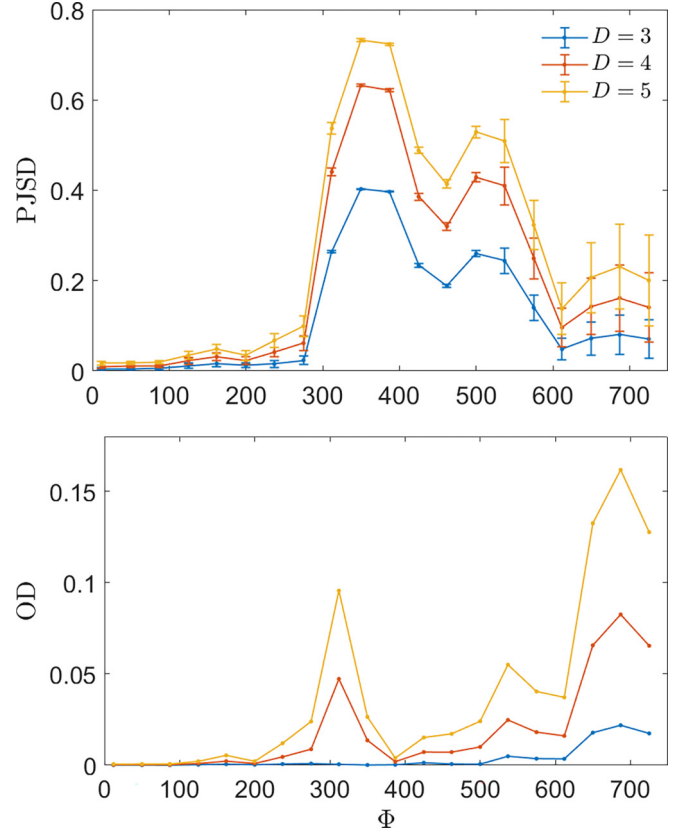

FIG. 6. Results obtained for the analysis of the experimental Mackey-Glass-based reservoir computer with input scaling $\gamma = 60$ and feedback scaling $\beta = 0.8$ over 20 different values of the bias Φ (x axis) in the range [12, 725] (that spans the domain of the Mackey-Glass nonlinearity). Top: PJSD between each one of the outputs associated with the 100 virtual nodes and the input stream that feeds the reservoir. Mean and standard deviation (as error bar) obtained from these 100 estimations are depicted. Bottom: OD estimated for the set of $M = 100$ outputs. Estimations with orders $D \in \{3, 4, 5\}$ and lag $\tau = 1$ for time series of length 3×10^5 data points are displayed. The parameter region at around $\Phi \sim 687$ satisfies having large OD with low PJSD, i.e., large values of diversity among the output signals, but preserving their similarity with the input stream that feeds the reservoir.

desired for performing reservoir computers. For characterization purposes, the input stream follows a random uniform distribution.

Figure 6 (top) shows the average value of the PJSD between each one of the measured 100 virtual output nodes and the input stream that feeds the reservoir. Three distinct regions can be observed in Fig. 6 (top). For $\Phi \leqslant 275$, the input and output ordinal pattern distributions are equivalent as characterized by a low PJSD. For $312 \leqslant \Phi \leqslant 575$, the large PJSD values indicate that the output distributions are significantly different than the input one. For $\Phi \geqslant 612$, the input and output distributions are similar but not identical as illustrated by the intermediate values of PJSD. In this range, the relatively large error bars hint at the fact that the ordinal pattern output distributions exhibit some diversity.

Figure 6 (bottom) shows the OD estimated for the set of 100 virtual node outputs. The OD exhibits two distinct local
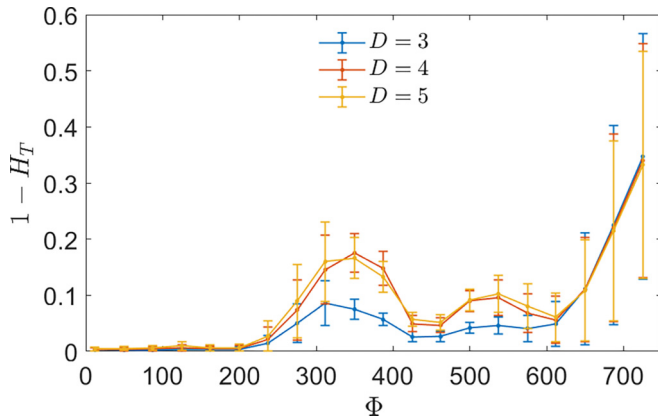
FIG. 7. Transcript synchronization $1 - H_T$ as a function of the bias Φ between the input stream that feeds the reservoir and each one of the 100 virtual output nodes. Mean and standard deviation (as error bar) obtained from the 100 estimations with orders $D \in \{3, 4, 5\}$ and lag $\tau = 1$ are depicted. A drive-response relationship is confirmed between the input and the output signals for intermediate and large values of the bias Φ.

maxima around $\Phi \sim 312$ and $\Phi \sim 687$, respectively, which are close to the transitions between dynamical regimes in the absence of external input. A third local maximum can be seen at $\Phi \sim 537$. The maximum at $\Phi \sim 687$ exhibits the largest OD for all embedding dimensions.

For reservoir computing purposes, a large output diversity that preserves the input driving is favorable. The parameter region around $\Phi \sim 687$ best fulfills the desired properties of large OD with low PJSD. A final test needs to be performed in order to guarantee that the observed large output diversity is a signature of a high-dimensional input-output mapping. We resort to the transcript entropy, a measure of desynchronization [53], as an indicator of the existence of a mapping between the input and the output signals. Please see the Appendix for more details about the transcript entropy analysis. As shown in Fig. 7, the input and the output time series are related in similar parameter regions where the OD peaks.

One of the main reasons for the popularity of reservoir computing for time series processing is the existence of a simple training scheme, i.e., usually a linear regression suffices [46]. For this reservoir computing training, only the signals at the output of the reservoir are needed. However, the final performance of the reservoir for a given computational task depends nontrivially on the so-called hyperparameters of the system, such as, in our case, the feedback strength or the input scaling. Brute-force evaluation of the optimal hyperparameters for each task is time consuming and contradicts the simplicity of the reservoir computing approach. For this reason, the evaluation of task-independent metrics can help to reduce the hyperparameter search for each task. Our results with the ordinal quantifiers are in good agreement with those obtained with the more involved information processing capacity measure [52] for the identification of favorable parameter regions for computing in the same system evaluated in our work. In turn, the relationship between information processing capacity and task-specific performance has been derived in [54]. Using the same data analyzed here, we have

verified that good performance in a nonlinear autoregressive moving average task [55] coincides with the parameter range where the OD is greatest.

V. CONCLUSIONS

We have proposed the ordinal diversity as a measure to estimate the heterogeneity in a set of time series. With the presented analytical, numerical, and experimental results, we have identified a number of useful properties that can improve the toolkits for time series analysis. The quantification of the fractional Brownian motion shows first that the OD measure increases monotonically as the range of Hurst exponents in the set under analysis is extended, while the numerical results approach the analytical values of the OD. Second, the example of the logistic map allows us to visualize that the OD characterizes the dynamical richness in a set of time series as well as serves to identify dynamical transitions. Third, the OD can identify the region of transition to synchronization in the case of heterogeneous coupled Hénon maps. Finally, we have validated the OD as a way to characterize the machine learning paradigm of reservoir computing, where diversity can be viewed as a resource for computation. For large amounts of data, the OD can also be used as a tool to identify the level of redundancy. These examples illustrate the potential applications of reducing the diversity of time series to a single number. Of course, the OD can be complemented with other measures, as details of the statistical properties of the time series are lost in the reduction to a single value.

There are two other outstanding features of the OD that deserve to be fully explored in future research. The first one is the possibility to assign different weights to the time series. This could be useful, for example, when time series with different lengths and/or different amplitude resolutions are compared. The second one is the fact that a multiscale analysis can be easily carried out by setting different τ values in the symbolization mapping. This multiscale implementation might be especially valuable in those analyses in which signals with different sampling frequencies are considered.

Thanks to its wide applicability, versatility, and robustness, we are optimistic about the utility of the proposed tool for the analysis of complex signals in heterogeneous scientific fields. In particular, we conjecture that it could be of high interest in neuroscience to help overcome some current issues related to the spatiotemporal dynamics of the human brain during physiological and pathological conditions [56].

MATLAB code to estimate ordinal diversity is available in a GitHub repository [57].

APPENDIX: TRANSCRIPT ENTROPY ANALYSIS

The symbolic transcription scheme has been proposed by Monetti *et al.* [58] to study synchronization between two time series. Through basic permutation operations, the ordinal patterns from the source and the target time series are mapped into each other. More precisely, the transfer symbol or transcript $T = (T_1, T_2, \ldots, T_D)$ is the unique ordinal pattern that transforms the source ordinal symbol $X = (X_1, X_2, \ldots, X_D)$ into the target ordinal motif $Y = (Y_1, Y_2, \ldots, Y_D)$ ($T, X, Y \in S_D$), such that [59]

$$T \circ X = (X_{T_1}, X_{T_2}, \ldots, X_{T_D}) = Y. \tag{A1}$$

A transcript series can then be calculated by repeating this procedure for each pair of simultaneous ordinal patterns from the source and the target time series. Let us illustrate this procedure with the help of a toy example. Given the source time series {4.1, 1.2, 6.3, 5.3, 10.8, 7.3} and the target time series {2.2, 8.3, 9.4, 3.7, 1.4, 2.6}, their associated ordinal sequences with $D = 4$ and $\tau = 1$ are

$$\{(2, 1, 4, 3), (1, 3, 2, 4), (2, 1, 4, 3)\} \tag{A2}$$

and

$$\{(1, 4, 2, 3), (4, 3, 1, 2), (3, 4, 2, 1)\}, \tag{A3}$$

respectively. The transcript series is given by the sequence of transfer symbols,

$$\{(2, 3, 1, 4), (4, 2, 1, 3), (4, 3, 1, 2)\}, \tag{A4}$$

that allows us to generate (A3) from (A2) by applying (A1).

The normalized Shannon entropy $H_T = S(T_\pi)/S_{\max}$, with $S_{\max} = \ln D!$, of the ordinal patterns probability distribution T_π associated with the transcript series quantifies the desynchronization between the two original, i.e., source and target, time series from an ordinal perspective. Consequently, the so-called transcript synchronization $1 - H_T$ is a measure of synchronization [53]. Intuitively, a high transcript synchronization implies that a low number of transcripts (small amount of information) is needed to deduce the ordinal sequence of one series, given the ordinal sequence of the other one. On the contrary, a low transcript synchronization is obtained when many different transcripts are required to infer the target ordinal sequence from the source one.

[1] C. A. Mattmann, A vision for data science, Nature (London) **493**, 473 (2013).

[2] D. M. Blei and P. Smyth, Science and data science, Proc. Natl. Acad. Sci. USA **114**, 8689 (2017).

[3] M. Gerlach, F. Font-Clos, and E. G. Altmann, Similarity of symbol frequency distributions with heavy tails, Phys. Rev. X **6**, 021009 (2016).

[4] A. Lahreche and B. Boucheham, A fast and accurate similarity measure for long time series classification based on local extrema and dynamic time warping, Expert Syst. Appl. **168**, 114374 (2021).

[5] L. Zunino, F. Olivares, H. V. Ribeiro, and O. A. Rosso, Permutation Jensen-Shannon distance: A versatile and fast symbolic tool for complex time-series analysis, Phys. Rev. E **105**, 045310 (2022).

[6] A. Jastrzebska, G. Nápoles, Y. Salgueiro, and K. Vanhoof, Evaluating time series similarity using concept-based models, Knowledge-Based Syst. **238**, 107811 (2022).

[7] F. Bahrpeyma, M. Roantree, P. Cappellari, M. Scriney, and A. McCarren, A methodology for validating diversity in synthetic time series generation, MethodsX **8**, 101459 (2021).

[8] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, Phys. Rev. Lett. **120**, 024102 (2018).

[9] T. L. Carroll, Creating new chaotic signals with reservoir computers, Chaos, Solitons & Fractals **164**, 112688 (2022).

[10] W. Jia, M. Sun, J. Lian, and S. Hou, Feature dimensionality reduction: A review, Complex Intell. Syst. **8**, 2663 (2022).

[11] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley, Analysis of symbolic sequences using the Jensen-Shannon divergence, Phys. Rev. E **65**, 041905 (2002).

[12] A. Kirkley, Information theoretic network approach to socioeconomic correlations, Phys. Rev. Res. **2**, 043212 (2020).

[13] H. Y. D. Sigaki, M. Perc, and H. V. Ribeiro, History of art paintings through the lens of entropy and complexity, Proc. Natl. Acad. Sci. USA **115**, E8585 (2018).

[14] C. Bandt, Small order patterns in big time series: A practical guide, Entropy **21**, 613 (2019).

[15] C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, Phys. Rev. Lett. **88**, 174102 (2002).

[16] M. Zanin and F. Olivares, Ordinal patterns-based methodologies for distinguishing chaos from noise in discrete time series, Commun. Phys. **4**, 190 (2021).

[17] I. Leyva, J. H. Martínez, C. Masoller, O. A. Rosso, and M. Zanin, 20 years of ordinal patterns: Perspectives and challenges, Europhys. Lett. **138**, 31001 (2022).

[18] J. M. Amigó and O. A. Rosso, Ordinal methods: Concepts, applications, new developments, and challenges—In memory of Karsten Keller (1961–2022), Chaos **33**, 080401 (2023).

[19] J. M. Amigó, K. Keller, and V. A. Unakafova, Ordinal symbolic analysis and its application to biomedical recordings, Phil. Trans. R. Soc. A **373**, 20140091 (2015).

[20] E. Bradley and H. Kantz, Nonlinear time-series analysis revisited, Chaos **25**, 097610 (2015).

[21] C. Bandt, Statistics and contrasts of order patterns in univariate time series, Chaos: Interdiscipl. J. Nonlinear Sci. **33**, 033124 (2023).

[22] L. Zunino, M. C. Soriano, I. Fischer, O. A. Rosso, and C. R. Mirasso, Permutation-information-theory approach to unveil

delay dynamics from time-series analysis, Phys. Rev. E **82**, 046212 (2010).

[23] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics, Comput. Biol. Med. **42**, 319 (2012).

[24] L. Zunino, M. C. Soriano, and O. A. Rosso, Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach, Phys. Rev. E **86**, 046210 (2012).

[25] I. Kottlarz and U. Parlitz, Ordinal pattern-based complexity analysis of high-dimensional chaotic time series, Chaos: Interdiscipl. J. Nonlinear Sci. **33**, 053105 (2023).

[26] C. S. Daw, C. E. A. Finney, and E. R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. **74**, 915 (2003).

[27] K. Keller, M. Sinn, and J. Emonds, Time series from the ordinal viewpoint, Stochast. Dynam. **07**, 247 (2007).

[28] L. Zunino, F. Olivares, F. Scholkmann, and O. A. Rosso, Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions, Phys. Lett. A **381**, 1883 (2017).

[29] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory **37**, 145 (1991).

[30] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, Compositional segmentation and long-range fractal correlations in DNA sequences, Phys. Rev. E **53**, 5181 (1996).

[31] P. Lamberti, M. Martin, A. Plastino, and O. Rosso, Intensive entropic nontriviality measure, Physica A **334**, 119 (2004).

[32] D. Endres and J. Schindelin, A new metric for probability distributions, IEEE Trans. Inf. Theory **49**, 1858 (2003).

[33] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. **10**, 422 (1968).

[34] L. Decreusefond and A. S. Üstünel, Fractional Brownian motion: Theory and applications, ESAIM: Proc. **5**, 75 (1998).

[35] A. A. B. Pessa and H. V. Ribeiro, Characterizing stochastic time series with ordinal networks, Phys. Rev. E **100**, 042304 (2019).

[36] H. Verdier, F. Laurent, A. Cassé, C. L. Vestergaard, and J.-B. Masson, Variational inference of fractional Brownian motion with linear computational complexity, Phys. Rev. E **106**, 055311 (2022).

[37] C. Bandt and F. Shiha, Order patterns in time series, J. Time Ser. Anal. **28**, 646 (2007).

[38] J.-F. Coeurjolly, Simulation and identification of the fractional Brownian motion: A bibliographical and comparative study, J. Stat. Software **5**, 1 (2000).

[39] R. M. May, Simple mathematical models with very complicated dynamics, Nature (London) **261**, 459 (1976).

[40] M. Henon, A two-dimensional mapping with a strange attractor, Commun. Math. Phys. **50**, 69 (1976).

[41] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble, Phys. Rev. E **54**, 6708 (1996).

[42] R. Quian Quiroga, J. Arnhold, and P. Grassberger, Learning driver-response relationships from synchronization patterns, Phys. Rev. E **61**, 5142 (2000).

[43] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Synchronization as adjustment of information rates: Detection from bivariate time series, Phys. Rev. E **63**, 046211 (2001).

[44] T. Kreuz, F. Mormann, R. G. Andrzejak, A. Kraskov, K. Lehnertz, and P. Grassberger, Measuring synchronization in coupled model systems: A comparison of different approaches, Physica D **225**, 29 (2007).

[45] A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, Comparison of six methods for the detection of causality in a bivariate time series, Phys. Rev. E **97**, 042207 (2018).

[46] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. **3**, 127 (2009).

[47] K. Nakajima and I. Fischer, *Reservoir Computing: Theory, Physical Implementations, and Applications* (Springer, Singapore, 2021).

[48] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, Neural Networks **115**, 100 (2019).

[49] M. C. Soriano, D. Brunner, M. Escalona-Morán, C. R. Mirasso, and I. Fischer, Minimal approach to neuro-inspired information processing, Front. Comput. Neurosci. **9**, 68 (2015).

[50] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, Information processing capacity of dynamical systems, Sci. Rep. **2**, 514 (2012).

[51] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, Nat. Commun. **2**, 468 (2011).

[52] B. Vettelschoss, A. Röhm, and M. C. Soriano, Information processing capacity of a single-node reservoir computer: An experimental evaluation, IEEE Trans. Neural Networks Learn. Syst. **33**, 2714 (2022).

[53] M. López Pérez and R. Mansilla Corona, Ordinal synchronization and typical states in high-frequency digital markets, Physica A **598**, 127331 (2022).

[54] T. Hülser, F. Köster, K. Lüdge, and L. Jaurigue, Deriving task specific performance from the information processing capacity of a reservoir computer, Nanophotonics **12**, 937 (2023).

[55] H. Jaeger, Adaptive nonlinear system identification with echo state networks, in *Advances in Neural Information Processing Systems*, edited by S. Becker, S. Thrun, and K. Obermayer (MIT Press, 2002), Vol. 15, pp. 609–616.

[56] K. Lehnertz, Ordinal methods for a characterization of evolving functional brain networks, Chaos **33**, 022101 (2023).

[57] L. Zunino (31 October 2023), github.com/lucianoz-codes/ordinal-diversity.

[58] R. Monetti, W. Bunk, T. Aschenbrenner, and F. Jamitzky, Characterizing synchronization in time series using information measures extracted from symbolic representations, Phys. Rev. E **79**, 046207 (2009).

[59] J. M. Amigó, R. Monetti, T. Aschenbrenner, and W. Bunk, Transcripts: An algebraic approach to coupled time series, Chaos **22**, 013105 (2012).