Transfer entropy rate through Lempel-Ziv complexity

Juan F. Restrepo,[1,2,*] Diego M. Mateos,[2,3,4] and Gastón Schlotthauer[1,2]

[1]Laboratorio de Señales y Dinámicas no Lineales, Instituto de Investigación y Desarrollo en Bioingeniería y Bioinformática, CONICET–UNER, Entre Ríos, Argentina

[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 CABA, Argentina

[3]Instituto de Matemática Aplicada del Litoral, CONICET–UNL, Paraje El Pozo 3000, Santa Fe, Argentina

[4]Facultad de Ciencia y Tecnología, Universidad Autónoma de Entre Ríos, Entre Ríos, Argentina

(Received 17 October 2019; accepted 21 April 2020; published 15 May 2020)

The transfer entropy and the transfer entropy rate are closely related concepts that measure information exchange between two dynamical systems. These measures allow us to study linear and nonlinear causality relations and can be estimated through the use of different methodologies. However, some of them assume a data model and/or are computationally expensive. This article depicts a methodology to estimate the transfer entropy rate between two systems through the Lempel-Ziv complexity. This methodology offers a set of advantages: It estimates the transfer entropy rate from two single discrete series of measures, it is not computationally expensive, and it does not assume any data model. The simulation results over three different unidirectional coupled dynamical systems suggest that this methodology can be used to assess the direction and strength of the information flow between systems. Moreover, it provides good estimations for short-length time series.

DOI: 10.1103/PhysRevE.101.052117

I. INTRODUCTION

Transfer entropy (TE) and the transfer entropy rate (TER) are closely related concepts that measure information transport. The former was proposed by Schreiber [1] and independently by Paluš [2]. The latter was described by Amblard *et al.* [3,4]. These measures are able to quantify the strength and direction of the coupling between simultaneously observed systems [5]. Moreover, they have provoked a general interest because they can be used to study complex interaction phenomena found in several disciplines [6].

Lempel-Ziv complexity (LZC) is a classical measure that relates the concepts of complexity (in the Kolmogorov-Chaitin sense) and entropy rate [7,8]. For ergodic dynamical processes, the amount of new information gained per unit of time (entropy rate) can be estimated by measuring the capacity of this source to generate new patterns (LZC). Because of the simplicity of the LZC method, the entropy rate can be estimated from a single discrete sequence of measurements with a low computational cost [9]. This idea has been successfully applied to measure the entropy rate of neurological signals [10]. However, the LZC has not been used to estimate either the TE or the TER.

Recently, several algorithms devoted to estimate the transfer entropy have been reported. In Refs. [11,12] a k-nearest-neighbor (kNN) methodology is used to estimate joint Shannon entropies in different reconstructed spaces, which provides good estimations of transfer entropy (TE) even though it is computational expensive. Staniek and Lehnertz [13] proposed a faster algorithm founded on the concept of permutation entropy [14]. Nevertheless, because it is built on the relative frequencies of permutation patterns, the algorithm is highly dependent of the data length and poses a problem when there is a small amount of data taken from a high-dimensional system. As an alternative, we propose to exploit the advantages of the LZC methodology to calculate the TER between two ergodic dynamical systems.

In this article, we aim to relate the well-established concepts of Lempel-Ziv complexity and transfer entropy rate. To be precise, we propose a novel methodology that exploits the algorithmic advantages of the former to estimate the latter. This approach reconstructs the joint dynamics of the systems using a delay-embedding procedure and computes the TER through the LZC of sequences build in that reconstructed space. This measure is able to quantify the direction of the information flow between two systems from short-length time series, it is not computationally expensive and it does not assume any data model.

The remainder of this paper is organized as follows. Section II begins with a brief review of the concepts of TE, TER, and LZC. In Sec. III, we described the proposed methodology to estimate the transfer entropy rate through the Lempel-Ziv complexity. In Sec. IV, we present and analyze the results of the simulations carried out to evaluate the performance of our approach. Finally, in Secs. V and VI the discussion and conclusions are described.

II. METHODOLOGY

In this section, we will briefly review some theoretical concepts related to the LZC, TE, and TER. Moreover, we will introduce the notation used along the document.

*Corresponding author: jrestrepo@ingenieria.uner.edu.ar; a preprint version of this article can be accessed at: http://export.arxiv.org/pdf/1903.07720.

Since our intent is to investigate a possible causality connection between two dynamical systems, we need to analyze the signals that they produce. We will assume the existence of ergodic probability measures that describe the density of trajectories in phase space and which can be treated as probability densities. This allows us to analyze the systems' dynamics through the construction of random processes from their signals.

Consider a system X that produces a time series $x_t = x_1 \cdots x_T$. We can compose samples of an m-dimensional time-embedded process $\{X^{(m)}\} = \{X_1, \dots, X_m\}$ by sampling x_t with a frequency of $1/\tau$ [15,16]:

$$\boldsymbol{x}_n^{(m)} = [x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau}],$$

where $n = 1, \dots, T-(m-1)\tau$. The process $\{X^{(m)}\}$ is characterized by the joint probability distribution:

$$p\big(\boldsymbol{x}_n^{(m)}\big) = P\big\{(X_1, \dots, X_m) = \boldsymbol{x}_n^{(m)}\big\}.$$

We can define the m-order entropy rate as [16]:

$$h(X^{(m)}) = H(X^{(m+1)}) - H(X^{(m)}),$$

where $H(X^{(m)})$ is the entropy of the joint distribution $p(\boldsymbol{x}_n^{(m)})$:

$$\begin{aligned} H(X^{(m)}) &= H(X_1, \dots, X_m), \\ &= -\sum_{x_1} \cdots \sum_{x_m} p\big(\boldsymbol{x}_n^{(m)}\big) \ln p\big(\boldsymbol{x}_n^{(m)}\big). \end{aligned}$$

The m-order entropy rate measures the variation of the total information in the time-embedded process when the embedding dimension m is increased by 1. Based on this definition, we can calculate the entropy rate of the system X as follows [7,17]:

$$h(X) = \lim_{m\to\infty} h(X^{(m)}), \tag{1}$$

$$= \lim_{m\to\infty} \frac{H(X^{(m)})}{m}. \tag{2}$$

Equations (1) and (2) relate two different interpretations of the entropy rate. The first equation shows that $h(X)$ is a measure of our uncertainty about the present state of the system under the assumption that its entire past has been observed. The second one states that the entropy rate is the average information gained by observing the system. In this respect, systems with a higher entropy rate generate information at a superior scale which makes their dynamics more complex and difficult to predict.

A. Lempel-Ziv complexity

The concepts of entropy rate and Lempel-Ziv complexity are closely related, because systems with higher entropy rate tend to generate more complex sequences (time series). In that context, the entropy rate of an ergodic system can be estimated by measuring its capacity to generate new patterns [9]. Assessing the entropy rate of a system using the Lempel-Ziv algorithm carries a set of practical advantages: It can be estimated from a single discrete series of measures, the algorithm is fast and it does not assume any model for the data.

Suppose a stationary, discrete stochastic process $\{X_t\}$ that generates a sequence x_t of length T, where for a fixed t, the random variable X_t can take values from an alphabet Ω_x of α symbols. To estimate the complexity of this process, we will use the Lempel and Ziv scheme proposed in 1976 [18]. In this approach, a sequence x_t is parsed into a number $\mathcal{C}_{x_t}$ of words by considering any subsequence that has not yet been encountered as a new word. For example, the sequence 10011011100101010001011 is parsed in seven words: $1 \cdot 0 \cdot 01 \cdot 101 \cdot 1100 \cdot 1010 \cdot 001011$. As a result, the entropy rate can be computed as [8]:

$$h(X) = \lim_{T\to\infty} \frac{\mathcal{C}_{x_t}\big[\ln\alpha + \ln\mathcal{C}_{x_t}\big]}{T}. \tag{3}$$

By extending the alphabet size, this approach can be easily generalized to multivariate discrete processes [19]. Consider an m-dimensional stationary process $\{X^{(m)}\}$ that generates the sequences $x_{t,i} = x_{1,i}, \dots, x_{T,i}$ with $i = 1, \dots, m$, each one of them from an alphabet of α symbols. Let $z_t = z_1, \dots, z_T$ to be a new sequence defined over an extended alphabet of size α^m [19]:

$$z_t = \sum_{i=1}^{m} \alpha^{i-1} x_{t,i}$$

and then the joint Lempel-Ziv complexity $\mathcal{C}_{x_{t,i}} = \mathcal{C}_{z_t}$ and the m-order entropy rate can be calculated as [8,19]:

$$\begin{aligned} h[X^{(m)}] &= h(Z), \\ &= \lim_{T\to\infty} \frac{\mathcal{C}_{z_t}\big[\ln(\alpha^m) + \ln\mathcal{C}_{z_t}\big]}{T}. \end{aligned}$$

The same approach can be extended for continuous processes using a quantization scheme. The most popular is the binarization of the time series using its median value. There exists other approaches like the one proposed in Ref. [20] that uses permutation patterns to produce a symbolic sequence from the data. In this article we use an α quantile scheme, where the symbolic sequence $\{x_t\}$ is built by assigning to each sample of the time series a number from 0 to $\alpha - 1$ depending on which α quantile it belongs. For example, if $\alpha = 2$, then the time series will be binarized according to its median value. This ensures that all symbols in the alphabet have the same outcome probability which maximizes the Shannon entropy of the source.

B. Transfer entropy and transfer entropy rate

The transfer entropy is able to assess the amount of information transferred from process $Y = \{Y_t\}$ (driver-source) to process $X = \{X_t\}$ (driven-target). It is defined as [1,5,6,21]:

$$\begin{aligned} \mathbf{T}_{Y\to X}^{(m)} = {} & H\big(X_t^{(m)}, Y_t^{(m)}\big) - H\big(X_t, X_t^{(m)}, Y_t^{(m)}\big) \\ & + H\big(X_t, X_t^{(m)}\big) - H\big(X_t^{(m)}\big). \end{aligned} \tag{4}$$

where $X_t^{(m)} = (X_{t-m\tau}, \dots, X_{t-\tau})$ and $Y_t^{(m)} = (Y_{t-m\tau}, \dots, Y_{t-\tau})$. The parameter m is commonly called the history length or embedding dimension and τ is the lag or embedding lag [6]. $\mathbf{T}_{Y\to X}^{(m)}$ quantifies the amount of information contained in m-past states of process Y ($Y_t^{(m)}$) about the current state of the process X (X_t) that is not already explained by m-past states of process X ($X_t^{(m)}$). This measure is asymmetric, $\mathbf{T}_{Y\to X}^{(m)} \neq \mathbf{T}_{X\to Y}^{(m)}$, and increases along with the coupling level, allowing us to determine the direction and strength of the information flow [5]. Despite it is the most widely used measure, its estimation is sensitive to faulty observations of the states of the driving system, which can lead to spurious causality detection [22].

Amblard and Michel suggest that the TE can be considered as an information flow rate under stationarity conditions [3]. This idea leads to the following definition of transfer entropy rate [3,4,23,24]:

$$\begin{aligned}\mathbf{t}^{(m)}_{Y\to X} &\equiv h(X) - h(X|Y),\\ &= h\big(X_t, X_t^{(m)}\big) - h\big(Y_t^{(m)}, X_t^{(m)}, X_t\big),\end{aligned} \tag{5}$$

where $h(X)$ is the entropy rate of X and $h(X|Y)$ is the conditional entropy rate [23]:

$$\begin{aligned}h(X|Y) &\equiv \lim_{m\to\infty} H\big(X_t \big| X_t^{(m)}, Y_t^{(m)}\big),\\ &= \lim_{m\to\infty} \frac{H\big(Y_t^{(m)}, X_t^{(m)}, X_t\big)}{m},\\ &= h\big(Y_t^{(m)}, X_t^{(m)}, X_t\big).\end{aligned}$$

The TER lies between zero and the entropy rate of the target X, being equal to zero if X and Y are independent [23].

If the processes X and Y had no relationship, then $\mathbf{t}^{(m)}_{Y\to X}$ should be equal to zero. However, in practical applications, the estimation of $\mathbf{t}^{(m)}_{Y\to X}$ could present a bias due to the finite length of the data. Bossomaier *et al.* proposed to correct this bias by empirically finding the distribution of the surrogate measurement $\hat{\mathbf{t}}^{(m)}_{\hat{Y}\to X}$. The surrogate data must be generated in such a way that the temporal correlation between the source and the target processes is destroyed but the statistical properties and the temporal structure of both processes are preserved [6,24]. Note that only the second term in Eq. (5) depends on the source, so the surrogate transfer entropy from Y to X is defined as:

$$\hat{\mathbf{t}}^{(m)}_{\hat{Y}\to X} = -\big\langle h_k\big(\hat{Y}_t^{(m)}, X_t^{(m)}, X_t\big)\big\rangle_K, \tag{6}$$

where $\hat{Y}_t^{(m)}$ is obtained by redrawing with replacement samples from $Y_t^{(m)}$, and $\langle\cdot\rangle_K$ is the mean value over the $k = 1, 2, \ldots, K$ surrogate realizations.

In order to assess the directionality of the information flow, we need to analyze the global TER estimator:

$$\mathcal{T} = \mathbf{t}^{(m)}_{Y\to X} - \mathbf{t}^{(m)}_{X\to Y} - \big(\hat{\mathbf{t}}^{(m)}_{\hat{Y}\to X} - \hat{\mathbf{t}}^{(m)}_{\hat{X}\to Y}\big). \tag{7}$$

A positive value of $\mathcal{T}$ suggests that the information flow goes from system Y to system X, whereas a negative value suggests the contrary. Finally, if $\mathcal{T} = 0$, then there is no information flow between systems.

III. TRANSFER ENTROPY RATE BASED ON LEMPEL-ZIV COMPLEXITY

In this section, we formalize our approach to estimate the transfer entropy rate using the Lempel-Ziv complexity. The intention is to estimate the two joint entropy rates on the right-hand side of Eq. (5) by means of their associated joint Lempel-Ziv complexities. To this end, we propose a methodology based on the construction of delayed embedding vectors from quantized time series. To facilitate the description of the method (see Algorithm 1), we will assume binarized time series, although this methodology can be extended to higher quantization levels.

Algorithm 1. LeZTER.

MATLAB code:
https://bitbucket.org/jrinckoar/tentropyrate-lzc/src/master/

1: Obtain the symbolic sequences x_t and y_t by assigning to each sample of the time series a number from 0 to $\alpha - 1$ depending on which α quantile it belongs.
2: Set x_t as target-driven series and y_t as source-driver series.

3: Set the matrix of embedding vectors according to the given value of m and τ [Eq. (8)]:

$$\mathcal{V} = \big[\boldsymbol{y}_n^{(m)}, \boldsymbol{x}_n^{(m)}, x_n\big],$$

and obtain the sequence z_n (see Fig. 1).
4: Calculate the LZC of z_n and the entropy rate $h(Y_t^{(m)}, X_t^{(m)}, X_t)$ using Eq. (9).
5: Calculate the entropy rate $h(X_t^{(m)}, X_t)$. Obtain the corresponding z_n sequence considering the last $m+1$ columns of $\mathcal{V}$.
6: Calculate the transfer entropy rate $\mathbf{t}^{(m)}_{Y\to X}$ using Eq. (5).
7: Set a number K of surrogate data sets. For $k = 1, 2, \ldots, K$ build the matrices:

$$\mathcal{V}_k = \big[\hat{\boldsymbol{y}}_n^{(m)}, \boldsymbol{x}_n^{(m)}, x_n\big]_k.$$

Calculate $h_k(\hat{Y}_t^{(m)}, X_t^{(m)}, X_t)$ using Eq. (9) and $\hat{\mathbf{t}}^{(m)}_{\hat{Y}\to X}$ using Eq. (6).

8: Set y_t as target series, x_t as source series and repeat steps 3–7 to calculate $\mathbf{t}^{(m)}_{X\to Y}$ and $\hat{\mathbf{t}}^{(m)}_{\hat{X}\to Y}$.
9: Obtain the global estimation of transfer entropy rate $\mathcal{T}$ using Eq. (7).

Consider two binarized time series ($\alpha = 2$) from a coupled system: $x_t = x_1 \cdots x_T$ as the target and $y_t = y_1 \cdots y_T$ as the source. Set the parameter m (embedding dimension) and τ (embedding lag) and create a collection of embedding vectors $\{\boldsymbol{v}_n\}$ as samples of a multidimensional process $\boldsymbol{V} = (Y_t^{(m)}, X_t^{(m)}, X_t)$ (see Fig. 1):

$$\{\boldsymbol{v}_n\} = \big\{\big(\boldsymbol{y}_n^{(m)}, \boldsymbol{x}_n^{(m)}, x_n\big)\big\},$$

where:

$$\begin{aligned}&n = 1, 2, \ldots, N, \quad \text{with} \quad N = T - m\tau,\\ &\boldsymbol{y}_n^{(m)} = [y_n, y_{n+\tau}, \ldots, y_{n+(m-1)\tau}],\\ &\boldsymbol{x}_n^{(m)} = [x_n, x_{n+\tau}, \ldots, x_{n+(m-1)\tau}],\\ &x_n = x_{n+m\tau}.\end{aligned} \tag{8}$$

By construction $\{\boldsymbol{v}_n\}$ is a collection of $(2m+1)$-uples, so we can define the sequence $z_n = \sum_{i=1}^{2m+1} 2^{i-1} v_{n,i}$, over an extended alphabet of size 2^{2m+1}. Then the joint entropy rate $h(Y_t^{(m)}, X_t^{(m)}, X_t)$ can be calculated as [19]:

$$\begin{aligned}h\big(Y_t^{(m)}, X_t^{(m)}, X_t\big) &= h(\boldsymbol{V}),\\ &= h(Z),\\ &= \lim_{N\to\infty} \frac{\mathcal{C}_{z_n}\big[\ln(2^{2m+1}) + \ln \mathcal{C}_{z_n}\big]}{N},\end{aligned} \tag{9}$$

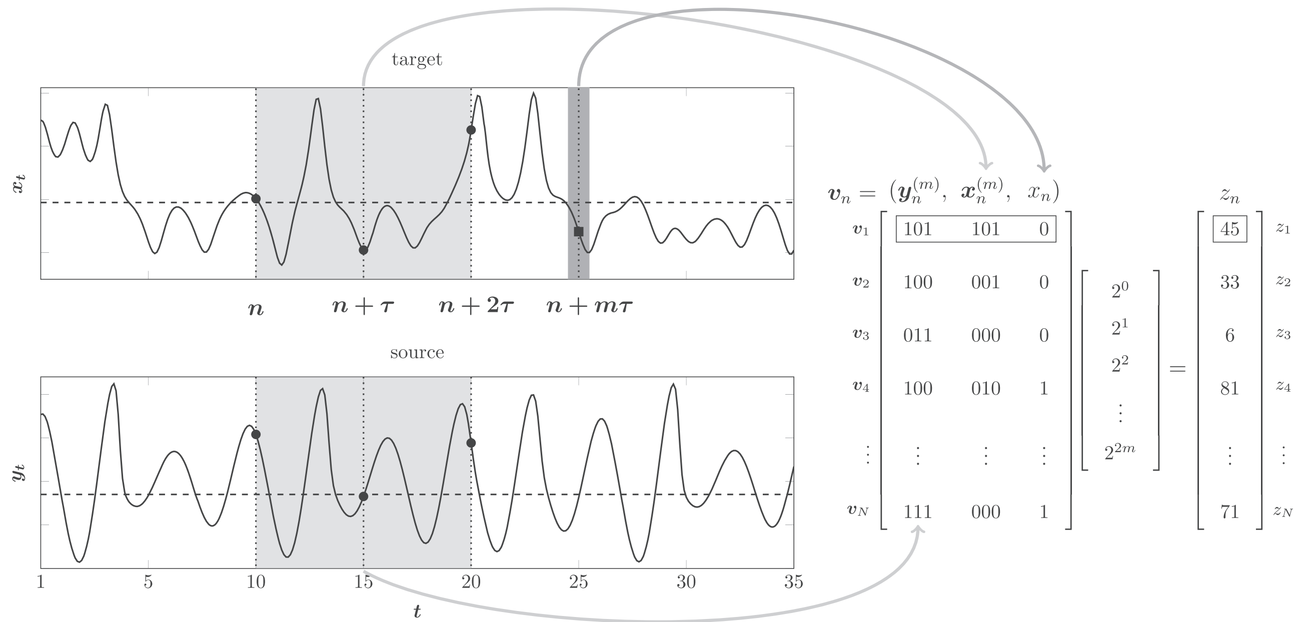

FIG. 1. Diagram to obtain the sequence z_n to calculate the entropy rate $h(Y_t^{(m)}, X_t^{(m)}, X_t)$. The matrix $\mathcal{V}$ is obtained by embedding ($m = 3$ and $\tau = 5$) the binarized version of x_t and y_t. The median values of both time series are shown as horizontal dashed lines.

where $\mathcal{C}_{z_n}$ is the LZC of the sequence z_n. The same procedure can be followed to estimate the first term of Eq. (5), but considering the process $\boldsymbol{V} = (X_t^{(m)}, X_t)$, the collection of embedding vectors $\{\boldsymbol{v}_n\} = \{(\boldsymbol{x}_n^{(m)}, x_n)\}$, and the sequence $z_n = \sum_{i=1}^{m+1} 2^{i-1} v_{n,i}$.

The surrogate measurement $h_k(\hat{Y}_t^{(m)}, X_t^{(m)}, X_t)$ can be obtained by taking the collection of embedding vectors:

$$\{\hat{\boldsymbol{v}}_n\}_k = \left\{\left(\hat{\boldsymbol{y}}_n^{(m)}, \boldsymbol{x}_n^{(m)}, x_n\right)\right\}_k,$$

where the samples $\{\hat{\boldsymbol{v}}_n\}_k$ are obtained by shuffling (or redrawing with replacement) $\boldsymbol{y}_n^{(m)}$ amongst the set of $(\boldsymbol{y}_n^{(m)}, \boldsymbol{x}_n^{(m)}, x_n)$ tuples.

The procedure to estimate the global transfer entropy rate is summarized in the following algorithm:

IV. RESULTS

We have conducted three simulations using different unidirectional coupled systems: the Henon-Henon, the Lorenz driven by Rössler, and the Lorenz-Lorenz. They were chosen because of their relevance and frequent use in the literature.

The coupled Henon-Henon system is described by the following equation [2,25]:

$$\begin{aligned}
y_1[n+1] &= 1.4 - y_1^2[n] + by_2[n]\\
y_2[n+1] &= y_1[n]\\
x_1[n+1] &= 1.4 - (\epsilon y_1[n] + (1-\epsilon)x_1[n])x_1[n] + bx_2[n]\\
x_2[n+1] &= x_1[n]
\end{aligned}$$

where $b = 0.3$. For the simulation, the coupling parameter ϵ was varied from zero to 1 in steps of 0.1. For each ϵ, 200 realizations were computed using random initial conditions. The transfer entropy rate was calculated with $m = \{2, 3, 4, 5, 6, 7\}$ and $\tau = \{1, 3, 5, 7, 10\}$. This procedure was repeated for data lengths $N = \{500, 3000, 5000\}$.

The results are shown in Fig. 2. Each plot presents the global TER ($\mathcal{T}$), calculated with $m = 4$, $\alpha = 4$, and $\tau = 1$, as a function of the coupling parameter ϵ. It can be observed in Fig. 2(a) ($N = 500$) that the median value of estimator $\mathcal{T}$ is close to zero for $\epsilon = 0$. This is an expected result since there is no information flow between the two systems. Besides, the median value of $\mathcal{T}$ increases along with the coupling parameter until $\epsilon = 0.6$. The positivity of $\mathcal{T}$ points out the correct direction of coupling and its increasing magnitude indicates a rising strength of the coupling. In contrast, for $\epsilon \geqslant 0.7$ the median value of $\mathcal{T}$ is zero. For these values of the coupling parameter, the Henon-Henon system is synchronized in such a way that both systems are statistically indistinguishable. In this kind of situation, $\mathcal{T}$ is unable indicate any information flow. This behavior has been already observed on other transfer entropy estimators [2,21,25]. Regarding the variance and bias of the estimations, it can be seen in Figs. 2(b) and 2(c) ($N = 3000$ and $N = 5000$, respectively) that it decreases as long as the data length is increased. This behavior is confirmed by the results shown in Fig. 3, where we show the estimation of $\mathcal{T}$ versus data length. Note that the values of $\mathcal{T}$ approaches asymptotically toward ≈0.48 as N increases, which means that the bias, relative to our measure, is reduced when the data length is enlarged.

For the second simulation we have chosen the Lorenz driven by Rössler system (Rössler-Lorenz) [13,21,25]:

$$\begin{aligned}
\dot{y}_1 &= -6(y_2 + y_3)\\
\dot{y}_2 &= 6(y_1 + 0.2y_2)\\
\dot{y}_3 &= 6[0.2 + y_3(y_1 - 5.7)]\\
\dot{x}_1 &= 10(-x_1 + x_2)\\
\dot{x}_2 &= 28x_1 - x_2 - x_1x_3 + \epsilon y_2^{\beta}\\
\dot{x}_3 &= x_1x_2 - \tfrac{8}{3}x_3
\end{aligned}$$

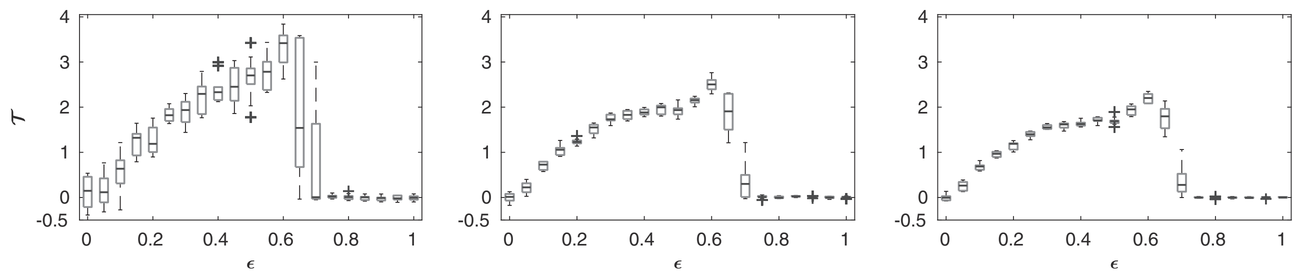

FIG. 2. Henon-Henon coupled system. Boxplot of global transfer entropy rate as a function of the coupling parameter ϵ. $\mathcal{T}$ was calculated with $m = 4$, $\alpha = 4$, and $\tau = 1$ for different data lengths: (a) $N = 500$, (b) $N = 3000$, and (c) $N = 5000$.

where $\beta = 2$ and $\epsilon \in \{0, 0.25, \ldots, 5\}$. We have computed 200 realizations for the values of the coupling parameter, each of them starting from a different initial condition. The numerical integration was performed using the ode45 function of Matlab with step size $\Delta t = 0.0217$ [21]. For each realization, the first 10 000 data points were discarded. Then $\mathcal{T}$ was calculated for the combination of all parameters: $m = \{2, 3, 4, 5, 6, 7\}$ and $\tau = \{1, 3, 5, 7, 10\}$. We applied the above procedure varying the data length $N = \{3000, 5000, 10\,000\}$.

In Fig. 4, the behavior of $\mathcal{T}$ as a function of the coupling parameter for $m = 6$, $\alpha = 6$, and $\tau = 10$ is shown. For this coupled system, the synchronization threshold is $\epsilon \approx 2.1$ [25,26]. In Fig. 4(a) ($N = 3000$) it can be observed that the median value of $\mathcal{T}$ is always positive, even for $\epsilon = 0$. This means that the $\mathcal{T}$ estimator identifies false coupling for $\epsilon = 0$. This phenomenon has also been observed in the symbolic transfer entropy [13]. However, the $\mathcal{T}$ estimator detects the correct coupling direction. Figures 4(b) and 4(c) display a similar behavior, but they reveal that the variance of $\mathcal{T}$ decreases as long as the data length is increased.

There already exist two similar methodologies to our approach that estimate the transfer entropy rate. The first one is the symbolic transfer entropy [13], which is founded in the permutation entropy [14]. The second one is based on the kNN estimation method proposed by Kraskov *et al.* [11,12]. In order to compare our methodology with the ones mentioned above, we have implemented both algorithms and calculated the global TER and the computation time for different parameters values. The simulation was performed in a cluster of 10 nodes, each of which has two Intel Xeon E5-2670 v3 2.5-GHz processors of 12 cores. The chosen system is the coupled Lorenz-Lorenz [13,25]:

$$\begin{aligned}
\dot{y}_1 &= 10(-y_1 + y_2)\\
\dot{y}_2 &= \rho_1 y_1 - y_2 - y_1 y_3\\
\dot{y}_3 &= y_1 y_2 - \tfrac{8}{3} y_3\\
\dot{x}_1 &= 10(-x_1 + x_2) + \epsilon(y_1 - x_1)\\
\dot{x}_2 &= \rho_2 x_1 - x_2 - x_1 x_3\\
\dot{x}_3 &= x_1 x_2 - \tfrac{8}{3} x_3
\end{aligned}$$

where $\rho_1 = 28.5$, $\rho_2 = 27.5$, $\epsilon \in \{0, \ldots, 15\}$, and $\Delta t = 0.03$ [13].

In Fig. 5 the estimator $\mathcal{T}$ is shown as a function of ϵ for the Lorenz-Lorenz coupled system and the three different methods: the LZC-based method (first column), the symbolic transfer entropy (second column), and the kNN-based approach (third column). The data in the plots have been normalized to the range $[-1, 1]$ for comparative purposes.

Regarding LZC-based approach, in Fig. 5(d) ($N = 1000$) it can be observed that for uncoupled systems ($\epsilon = 0$) the median value of $\mathcal{T} \approx 0$. As the coupling parameter increases, $\mathcal{T}$ is positive and grows until the synchronization threshold is reached ($\epsilon \approx 8$ [25]). From this point, the value of $\mathcal{T}$ goes toward zero despite the systems are coupled. Moreover, in Fig. 5(g) (N = 10 000) it can be observed that the variance of $\mathcal{T}$ decreases as the number of data points is increased, making easier to distinguish the direction of the information flow.

It is important to mention that our methodology produces similar results to the symbolic transfer entropy and the kNN approach, as it can be observed in Figs. 5(d), 5(e) and 5(f) for $N = 1000$ and Figs. 5(g), 5(h), and 5(i) for $N = 10\,000$. Nevertheless, for $N = 200$ neither the symbolic transfer entropy nor the kNN method can point out the direction of the information flow [Figs. 5(b) and 5(c)]. In contrast, the Lempel-Ziv complexity approach shows the correct path of the information flow despite the variance of the estimation Fig. 5(a).

Figures 6(a), 6(b) and 6(c) show a boxplot of the execution time of a single realization ($N = 10\,000$), for the three methods, as a function of the embedding dimension. Observe that the computational cost of each method increases with m in different ways. The increasing is linear for our approach, since the m parameter is linked to the alphabet size in the Lempel-Ziv algorithm. Comparing with the kNN approach, both methods present the same linear dependence

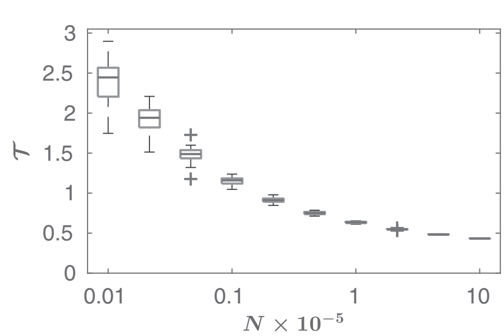

FIG. 3. Henon-Henon coupled system. Boxplot of global transfer entropy rate as a function of the data length N for $m = 4$, $\alpha = 4$, $\tau = 1$, and $\epsilon = 0.6$.

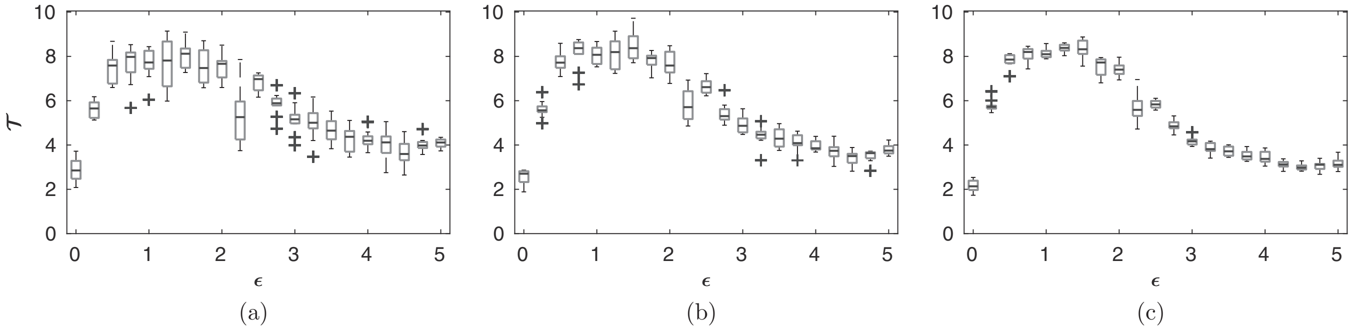

FIG. 4. Rössler-Lorenz coupled system. Boxplot of global transfer entropy rate as a function of the coupling parameter ϵ. $\mathcal{T}$ was calculated with $m = 6$, $\alpha = 6$, and $\tau = 10$ for different data lengths: (a) $N = 3000$, (b) $N = 5000$, and (c) $N = 10\,000$.

in the computation time, the scales are very different, being the computation time of the kNN method much longer than ours.

Regarding the symbolic transfer entropy, the results shown in Fig. 6(b) suggest that it is the fastest method for $m < 6$. For $m = 6$, its computation time is similar to the one of our algorithm. Moreover, for the symbolic transfer entropy, the embedding dimension is directly related with the size of the alphabet (factorial of m) used to quantize the time series. For $m = 7$, the size of the alphabet is 5040 (number of possible permutation patterns), this means that to estimate the joint probability $p(X_t, X_t^{(m)}, Y_t^{(m)})$ we need a three-dimensional matrix of size 5040^3 that requires ≈238.5 GB of memory (using int16 precision) and this mount of memory can not be handled

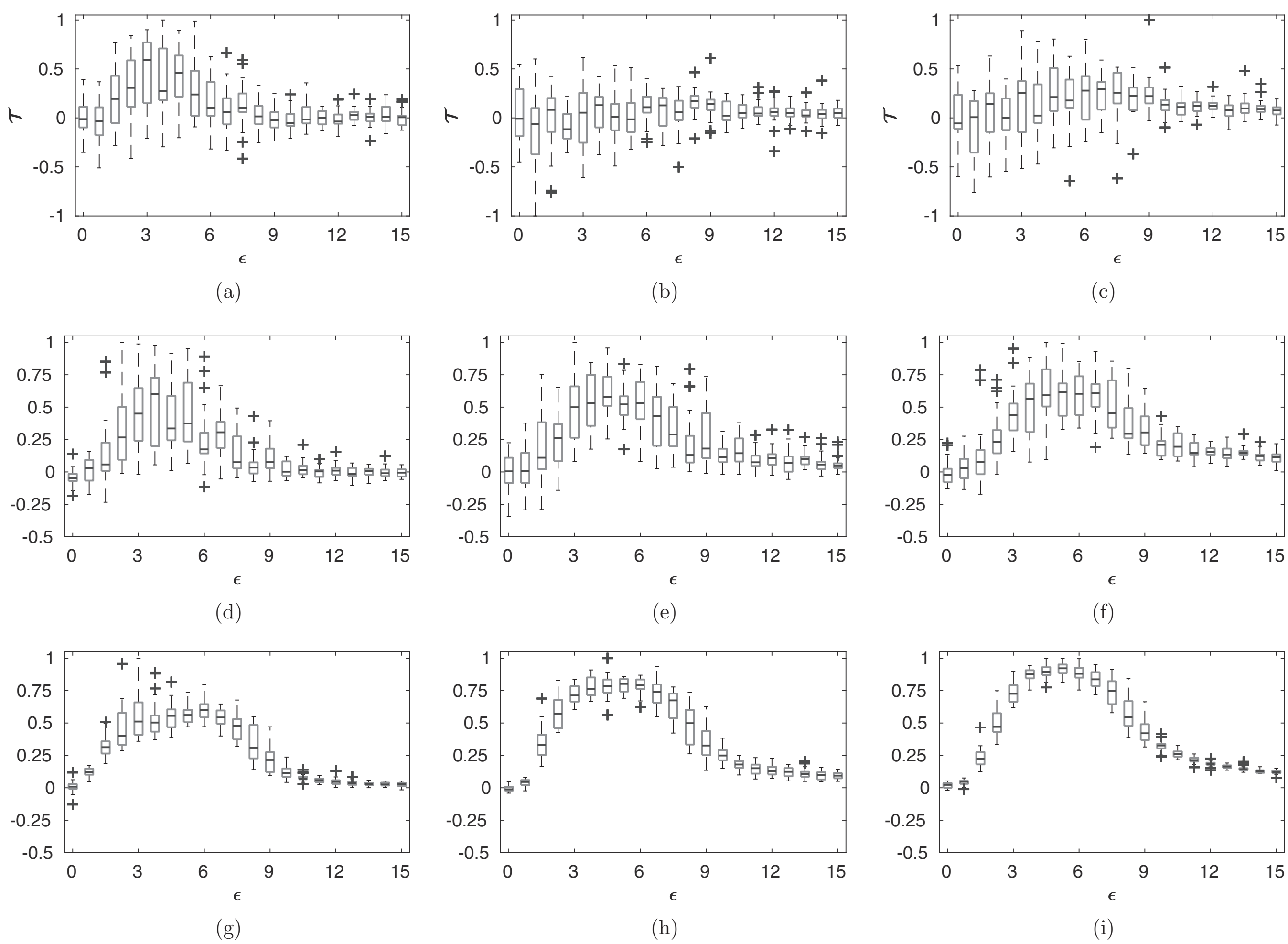

FIG. 5. Comparison of transfer entropy rate estimation with three different methods. Lempel-Ziv complexity-based method (a) $N = 200$, (d) $N = 1000$, and (g) $N = 10\,000$. Symbolic transfer entropy (b) $N = 200$, (e) $N = 1000$, and (f) $N = 10\,000$. The kNN method (c) $N = 200$, (f) $N = 1000$, and (i) $N = 10\,000$. Boxplot of the $\mathcal{T}$ as a function of the coupling parameter, calculated for the coupled Lorenz system with $m = 6$, $\tau = 5$, and $\alpha = 4$.

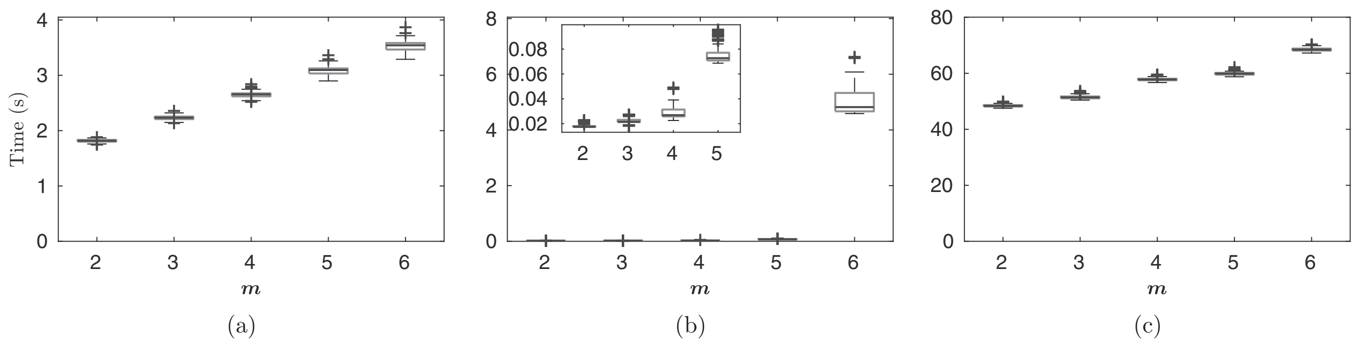

FIG. 6. Computation time as a function of the embedding dimension (m) for different transfer entropy rate estimation methods: (a) Lempel-Ziv complexity-based method, (b) symbolic transfer entropy and (c) kNN method. The simulation was made using the coupled Lorenz system with parameters $N = 10000$, $\tau = 5$, $\alpha = 4$, $\epsilon = 3$, and $K = 30$.

by a personal computer. This suggests that our methodology presents an additional advantage over the symbolic transfer entropy when the analysis of high-dimensional systems is needed.

V. DISCUSSION

One of the most widely used estimator of transfer entropy is the one based on the Kraskov-Stögbauer-Grassberger (KSG) estimator [12], which was implemented in the open source toolbox TRENTOOL [11]. This kNN methodology is derived by expressing volumes of neighborhoods of data points in terms of distances from each data point to other data points in the neighborhood. Despite it provides an unbiased estimation of transfer entropy, this approach carries two disadvantages. First, it depends on a good reconstruction of the geometry of the dynamic's manifold [27]. When the sample size is limited the local volume elements might not be descriptive of the geometry of the underlying probability measure, resulting in a biased estimation [27]. Second it is computationally expensive compared with the symbolic transfer entropy or our Lempel-Ziv-based approach. The symbolic transfer entropy algorithm is based on the estimation of relative frequencies of permutation patterns [13]. The main drawback of all methods based on counting appearance of patterns is that for short-length series the frequency of appearance of those patters may not be representative of the underlying probability of being observed. For short-length sequences just a few patterns are observed and the reconstructed pdf will be biased to a uniform distribution [28]. This situation is worsened when the alphabet length is large and for that reason the parameter m is often chosen to be small. However, a small m value leads to a poor quantization of the series and also to a poor embedding. Other observed issue is that this methodology requires a big amount of memory when the embedding parameter is large ($m > 6$), deteriorating its computational efficiency. In Ref. [29] Lesne *et al.* assert that the definition of Shannon entropy rate requires the knowledge of the invariant measure of the dynamics. Even under the assumption of ergodicity, the reconstruction of the probability distribution of the source from the observation of a typical single sequence is not warranted. On the contrary, Lempel-Ziv theorems, for stationary and ergodic sources, ensure two things. First, the Lempel-Ziv complexity coincides with the Shannon entropy rate up to a constant factor. Second, almost all symbolic sequences have the same compressibility features, this means that the estimation of the Shannon entropy rate can be performed with any typical sequence and the result coincides (in a statistical sense) with the average [30]. Moreover, in the same article [29] they compare four entropy rate estimators (two based on counting relative frequencies of patterns and the other two based on the Lempel-Ziv complexity) for short-length sequences. The results suggest that the Lempel-Ziv approach based on the 76 algorithm [18] gives better estimations than the other ones. As it was shown in Fig. 5, all the approaches give similar results as long as $N > 1000$. For smaller data lengths we have observed that the Lempel-Ziv complexity-based approach gives better results. This strengthens the hypothesis that our methodology can be used as an information transfer measure. Moreover, it presents an advantage over the other two methods when short-length series are available [compare Figs. 5(a), 5(b) and 5(c)].

The results obtained for the greatest embedding dimension ($m = 6$) and the longest data length ($N = 10\,000$) show that our approach is able to estimate transfer entropy rate as well as the other two methods but in less than 4 s. This lead us to conclude that the proposed methodology can detect the direction and strength of the information flow between two coupled ergodic systems in a time-effective period for online applications.

As mentioned, our methodology is based on the construction of embedding spaces from time series. In this direction, our algorithm has two parameters: the embedding dimension (m) and the embedding lag (τ). Like other embedding-based algorithms, we have found that the best results are achieved when a good reconstruction of the state space is guaranteed [31]. In other words, when m is bigger than the minimum embedding dimension of the system and τ is large enough so that the various coordinates of the embedding vectors contain as much new information as possible, without being entirely independent. For this reason a good choice of embedding dimension is $m = m_x + m_y + 1$, where m_x and m_y are estimations of the minimum embedding dimension of X and Y, respectively.[1] In addition, we propose to use an embedding

[1]The minimum embedding dimension for the Henon-Henon system is $m_x + m_y = 4$ and for the Lorenz-Lorenz and Rössler-Lorenz is $m_x + m_y = 6$.

lag value $\tau = \max(\tau_x, \tau_y)$, where τ_x and τ_y are the lags that minimize the mutual information function between x_t and $x_{t+\tau}$, and between y_t and $y_{t+\tau}$, respectively. Finally, some authors suggest that the data length, the dictionary size and the embedding dimension should be related as $N \geqslant \alpha^{m+1}$ [2,32].

VI. CONCLUSIONS

In this article we have presented a methodology to calculate the transfer entropy rate between two systems-based on the Lempel-Ziv's complexity. We were able to propose a computationally fast methodology to estimate the information flow between two systems because of the properties of the Lempel-Ziv algorithm. This methodology has been assessed using three unidirectional coupled systems: the Henon-Henon system, the Rössler-Lorenz system, and the Lorenz-Lorenz system. The results reveal two advantages compared to the standard approaches: First, it estimates better the transfer entropy rate for short data lengths. Second, its computation is faster than the other methods for high dimensional embeddings ($m > 5$) and large data lengths. In future studies, we will address the implementation of our methodology using different embedding parameters for the source and the target as well as a nonuniform embedding scheme.

ACKNOWLEDGMENTS

This work was supported by the National Scientific and Technical Research Council (CONICET) of Argentina, No. PIO-14620140100014CO (UNER-CONICET), and the National University of Entre Ríos (UNER), Grants No. PID-6171 (UNER).

[1] T. Schreiber, Measuring Information Transfer, Phys. Rev. Lett. **85**, 461 (2000).
[2] M. Paluš, V. Komárek, Z. Hrncír, and K. Sterbová, Synchronization as adjustment of information rates: Detection from bivariate time series, Phys. Rev. E **63**, 046211 (2001).
[3] P.-O. Amblard and O. J. J. Michel, Relating Granger causality to directed information theory for networks of stochastic processes, arXiv:0911.2873.
[4] P. O. Amblard and O. J. Michel, On directed information theory and Granger causality graphs, J. Comput. Neurosci. **30**, 7 (2011).
[5] A. Kaiser and T. Schreiber, Information transfer in continuous processes, Physica D **166**, 43 (2002).
[6] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Transfer Entropy* (Springer, Berlin, 2016).
[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2012).
[8] J.-L. Blanc, L. Pezard, and A. Lesne, Delay independence of mutual-information rate of two symbolic sequences, Phys. Rev. E **84**, 036214 (2011).
[9] E. Estevez-Rams, R. Lora Serrano, B. Aragón Fernández, and I. Brito Reyes, On the non-randomness of maximum Lempel Ziv complexity sequences of finite size, Chaos **23**, 023118 (2013).
[10] J. M. Amigó, J. Szczepański, E. Wajnryb, and M. V. Sanchez-Vives, Estimating the entropy rate of spike trains via lempel-ziv complexity, Neural Comput. **16**, 717 (2004).
[11] M. Lindner, V. Priesemann, R. Vicente, and M. Wibral, TRENTOOL: A Matlab open source toolbox to analyze information flow in time series data with transfer entropy, BMC Neurosci. **12**, 119 (2011).
[12] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, Phys. Rev. E **69**, 066138 (2004).
[13] M. Staniek and K. Lehnertz, Symbolic Transfer Entropy, Phys. Rev. Lett. **100**, 158101 (2008).
[14] C. Bandt and B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series, Phys. Rev. Lett. **88**, 174102 (2002).
[15] H. G. Schuster, *Deterministic Chaos: An Introduction*, 4th ed. (WileyVCH Verlag GmbH & Co. KGaA, 2005), pp. 1–287.
[16] C. Granero-Belinchon, S. G. Roux, P. Abry, M. Doret, and N. B. Garnier, Information theory to probe intrapartum fetal heart rate dynamics, Entropy **19**, 640 (2017).
[17] A. Papoulis and S.-U. Pillai, *Probabilities, Random Variables, and Stochastic Processes* (Tata McGraw-Hill Education, New York, 1991).
[18] A. Lempel and J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory **22**, 75 (1976).
[19] S. Zozor, P. Ravier, and O. Buttelli, On Lempel-Ziv complexity for multidimensional data analysis, Physica A **345**, 285 (2005).
[20] Y. Bai, Z. Liang, X. Li, L. J. Voss, and J. W. Sleigh, Permutation Lempel-Ziv complexity measure of electroencephalogram in GABAergic anaesthetics, Physiol. Meas. **36**, 2483 (2015).
[21] M. Paluš and M. Vejmelka, Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections, Phys. Rev. E **75**, 056211 (2007).
[22] D. A. Smirnov, Spurious causalities with transfer entropy, Phys. Rev. E **87**, 042917 (2013).
[23] T. Haruna and K. Nakajima, Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic Markov processes, Eur. Phys. J. B **86**, 230 (2013).
[24] R. Marschinski and H. Kantz, Analysing the information flow between financial time series, Eur. Phys. J. B **30**, 275 (2002).
[25] A. Krakovská, J. Jakubík, H. Budáčová, and M. Holecyová, Causality studied in reconstructed state space. Examples of uni-directionally connected chaotic systems, arXiv:1511.00505 (2015).
[26] R. Quian Quiroga, J. Arnhold, and P. Grassberger, Learning driver-response relationships from synchronization patterns, Phys. Rev. E **61**, 5142 (2000).
[27] W. M. Lord, J. Sun, and E. M. Bollt, Geometric k-nearest neighbor estimation of entropy and mutual information, Chaos **28**, 033114 (2018).
[28] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep. **441**, 1 (2007).

[29] A. Lesne, J. L. Blanc, and L. Pezard, Entropy estimation of very short symbolic sequences, Phys. Rev. E **79**, 046208 (2009).

[30] J. Ziv and A. Lempel, Compression of lndiwdual Sequences, IEEE Trans. Inf. Theory **24**, 530 (1978).

[31] M. Small, *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance* (World Scientific, Singapore, 2005).

[32] T. Craciunescu, A. Murari, and M. Gelfusa, Causality detection methods applied to the investigation of malaria epidemics, Entropy **21**, 784 (2019).