

# Information theory in non-linear dynamics and causality: state-space partition or ordinal patterns?

Milan Paluš<sup>1, a)</sup>

*Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 00 Praha 8, Czech Republic*

(Dated: 26 June 2026)

Information theory provides powerful tools for the analysis of time series generated by interacting nonlinear systems. Coupling and causal relationships can be identified and characterized in terms of information transfer, while the complexity of the underlying dynamics can be quantified through rates of entropy production. The practical evaluation of information-theoretic measures, however, requires the estimation of multidimensional probability distributions from experimental time series, which is often challenging. In this work, we analyze and discuss the applicability of replacing state-space partitioning with ordinal patterns encoding continuous data into discrete symbolic representations.

PACS numbers: 05.45.Tp, 05.45.Xt

Keywords: nonlinear time series, Granger causality, conditional mutual information, transfer entropy, entropy rates, ordinal patterns, permutation entropy

Nonlinear, potentially chaotic systems evolve in time in a nontrivial manner, which imposes fundamental limits on the predictability of their future states. In real-world settings, such systems do not evolve in isolation but instead interact with one another, introducing additional complexity into their temporal dynamics. As a result, complex and chaotic systems generate uncertainty, commonly quantified as entropy, and their interactions can be interpreted as exchanges of information.

Scientific observation of such evolving phenomena is enabled by measurable quantities that can be recorded at consecutive time instants and stored as datasets known as time series. Time-series analysis therefore plays a pivotal role in understanding the evolution of interacting systems, particularly through information-theoretic tools that characterize dynamical complexity via entropy rates and quantify interactions by measuring exchanged or directionally transferred information.

The evaluation of information-theoretic measures, however, requires the estimation of multidimensional probability distributions, the reliability of which is limited by the amount of available data when methods based on state-space partitioning are employed. Ordinal patterns enable the encoding of continuous data into discrete symbolic sequences and are supposed to mitigate the curse of dimensionality; however, this approach is not suitable for all types of experimental data and/or underlying mechanisms.

## I. INTRODUCTION

Many phenomena in nature and society evolve over time and can be viewed as dynamical systems. Mathematical models provide a framework for characterizing their current state, understanding the mechanisms underlying their behavior, and anticipating their future evolution.

In experimental studies of complex dynamical systems, it is often impossible to fully characterize the system state using a single measurement or a set of instantaneous measurements of observable quantities. In many cases, however, the dynamics or evolution of system components can be monitored by recording the temporal evolution of one or more observable quantities and registering their values at successive points in time  $t_1, t_2, \dots, t_N$ . The collection of measurements  $\{s(t)\}$  is called a *time series*.

Time series analysis is a traditional field of statistics rooted in the theory of stochastic systems.<sup>1</sup> Although the classical framework is primarily linear, several extensions have been developed to account for specific types of nonlinearity.<sup>2,3</sup>

An independent approach to the analysis of complex, potentially nonlinear time series emerged from dynamical systems theory, particularly from the study of deterministic chaos. A bridge between chaos theory and time series analysis was established by the seminal paper “Geometry from a Time Series.”<sup>4</sup> In this work, Packard et al.<sup>4</sup> demonstrated how a subset of a mathematical state space – within which the variables of a chaotic system evolve – can be reconstructed using only a single measurable quantity. Recording an observable quantity of the system under study in the form of a time series can therefore provide a key means to characterize, model, and predict the behavior of the system.

For a regular deterministic system, knowledge of its present state allows its future evolution to be predicted, with accuracy limited only by measurement noise.

---

<sup>a)</sup> Electronic mail: mp@cs.cas.cz

Chaotic systems, however, despite their deterministic nature, exhibit inherently limited predictability. Small errors in measuring their state (the “initial conditions”) grow exponentially over time as the system evolves. Consequently, a chaotic system is said to lose information about its initial state.

In other words, a chaotic system can be said to generate information or, equivalently, to increase entropy by amplifying uncertainty in its prediction. These ideas establish a bridge between dynamical systems theory<sup>5</sup> and Shannon’s information theory.<sup>6</sup> The applicability of information-theoretic methods to nonlinear dynamics and the analysis of experimental data was demonstrated by Shaw<sup>7</sup>, Martien et al.<sup>8</sup> and by Fraser.<sup>9</sup>

A. N. Kolmogorov,<sup>10</sup> who introduced the theoretical concept of classifying dynamical systems by information rates, was strongly influenced by information theory and generalized the concept of the entropy of an information source into the notion of Kolmogorov-Sinai entropy.<sup>5</sup>

Natural and social systems do not evolve in isolation but interact with one another. In the field of chaotic dynamics, cooperative behavior of systems has been observed: despite their complex dynamics and limited predictability, chaotic systems can synchronize.<sup>11</sup> The synchronization of chaotic dynamical systems has become a field of intensive research.<sup>12,13</sup> Paluš et al.<sup>14</sup> described the synchronization of dynamical systems as adjustment of their information rates due to mutual information transfer on the route to synchronization.<sup>14</sup> In the case of unidirectionally coupled dynamical systems, the distinction between the driving and the driven systems was recognized as a special case of Granger causality which can be identified using the conditional mutual information.<sup>14</sup>

Entropies, entropy rates, and measures of information transfer are used to characterize the dynamics and interactions of dynamical processes. These information-theoretic quantities are defined as functionals of probability distribution functions (PDFs). Estimating PDFs from experimental data is challenging, particularly in systems with many variables. The simplest and most naive approach is to partition the state space – that is, the ranges of observed variables – into rectangular “boxes” and to count the number of data points within each box (the so-called “box-counting” or “quantization” methods).<sup>15</sup>

Bandt and Pompe<sup>16–18</sup> proposed an interesting alternative to quantization: instead of discretizing amplitudes, continuous-valued data are encoded into a finite set of discrete symbols known as ordinal patterns. By counting ordinal patterns, which capture the local ups and downs of a time series, they introduced the concept of permutation entropy as a measure of the complexity of the dynamical system underlying the data. Bandt<sup>19</sup> proved that permutation entropy converges to the Kolmogorov-Sinai entropy<sup>10</sup> for certain classes of dynamical systems. Owing to its robustness and computational efficiency, this approach has become widely used in time series analysis.<sup>20</sup>

In this paper, we compare a slightly more sophisticated box-counting approach – equiquantal binning – with ordinal pattern encoding, with the aim of identifying the conditions under which the latter provides greater efficiency.

Information-theoretic functionals are introduced in Sec. II A. Their use in characterization of dynamics and inference of causality is described in Sec. II B, and in Sec. II C, respectively. Related statistical testing, based on surrogate data, is presented in Sec. II D. Binning estimation of information-theoretic functionals is introduced in Sec. II E and their evaluation using ordinal patterns in Sec. II F. Numerical data used in this study and their generation is described in Sec. II G. Results are presented in Sec. III and discussed in Sec. IV, conclusions are given in Sec. V.

## II. METHODS

### A. Information-theoretic functionals

Consider two discrete random variables  $X$  and  $Y$  with sets of values  $\Xi$  and  $\Upsilon$ , respectively. The probability distribution function (PDF)  $p_X(x)$  for the variable  $X$ , for simplicity denoted as  $p(x)$ , is  $p(x) = \Pr\{X = x\}$ ,  $x \in \Xi$ . The probability distribution function  $p(y)$  for the variable  $Y$  is defined in the full analogy; and the joint PDF  $p(x, y)$  is  $\Pr\{(X, Y) = (x, y)\}$ ,  $x \in \Xi$ ,  $y \in \Upsilon$ . Uncertainty in a random variable, say  $X$ , is characterized by its entropy

$$H(X) = - \sum_{x \in \Xi} p(x) \log p(x). \quad (1)$$

The joint entropy  $H(X, Y)$  of  $X$  and  $Y$  is

$$H(X, Y) = - \sum_{x \in \Xi} \sum_{y \in \Upsilon} p(x, y) \log p(x, y). \quad (2)$$

The two variables  $X$  and  $Y$  are independent if and only if  $p(x, y) = p(x)p(y)$ , i.e.

$$\log \frac{p(x, y)}{p(x)p(y)} = 0.$$

The average digression from independence, i.e., the averaged value of  $\log \frac{p(x, y)}{p(x)p(y)}$  is known as *mutual information*

$$I(X; Y) = \sum_{x \in \Xi} \sum_{y \in \Upsilon} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (3)$$

The mutual information can be expressed using the entropies (1), (2) as

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \quad (4)$$

Thus the mutual information  $I(X; Y)$  quantifies the decrease of uncertainty in  $H(X, Y)$  due to the dependence

between  $X$  and  $Y$ , i.e., it measures the average amount of common information, contained in the variables  $X$  and  $Y$ . The mutual information is a measure of general statistical dependence for which the following statements hold:

- $I(X; Y) \geq 0$ ,
- $I(X; Y) = 0$  iff  $X$  and  $Y$  are independent.

In some cases, the common information in the variables  $X$  and  $Y$  can be induced by influence of another variable,  $Z$ , on both  $X$  and  $Y$ . Then, *conditional mutual information (CMI)* of the random variables  $X$  and  $Y$  given  $Z$ , defined as

$$I(X, Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z), \quad (5)$$

measures the average amount of common information in the variables  $X$  and  $Y$  after removing the influence of the variable  $Z$ . Obviously, for  $Z$  independent of  $X$  and  $Y$ ,  $I(X, Y|Z) = I(X, Y)$  holds.

For bivariate normal distributions,

$$I(X, Y) = \frac{1}{2} \log(1 - \rho^2(X, Y)), \quad (6)$$

where  $\rho$  is the correlation coefficient between  $X$  and  $Y$ .

## B. Information in dynamics

Let us consider  $n$  discrete random variables  $X_1, \dots, X_n$  with values  $(x_1, \dots, x_n) \in \Xi_1 \times \dots \times \Xi_n$ . The PDF for an individual  $X_i$  is  $p(x_i) = \Pr\{X_i = x_i\}$ ,  $x_i \in \Xi_i$ , the joint PDF for the  $n$  variables  $X_1, \dots, X_n$  is  $p(x_1, \dots, x_n) = \Pr\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\}$ . The joint entropy of the  $n$  variables  $X_1, \dots, X_n$  with the joint PDF  $p(x_1, \dots, x_n)$  is

$$H(X_1, \dots, X_n) = - \sum_{x_1 \in \Xi_1} \dots \sum_{x_n \in \Xi_n} p(x_1, \dots, x_n) \log p(x_1, \dots, x_n). \quad (7)$$

A stochastic process  $\{X_i\}$  is an indexed sequence of random variables  $X_1, \dots, X_n$ , characterized by the joint PDF  $p(x_1, \dots, x_n)$ . Uncertainty in a variable  $X_i$  is characterized by its entropy  $H(X_i)$ . The rate at which a stochastic process “produces” uncertainty is measured by its entropy rate

$$h = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, \dots, X_n). \quad (8)$$

The entropy rate (8) is a useful characterization of the dynamics of a system or a process underlying the time series  $\{x(t)\}$ . In information theory the entropy rate (8) is considered as a measure of production of information of an information source.<sup>21</sup>

If a stochastic process  $\{X_i\}$  is a zero-mean, stationary, Gaussian process with power spectral density  $\Phi(\omega)$ , where  $\omega$  is a normalized frequency, then its entropy rate  $h_G$ , apart from a constant term, is<sup>22,23</sup>

$$h_G = \frac{1}{2\pi} \int_0^{2\pi} \log \Phi(\omega) d\omega. \quad (9)$$

Alternatively, a time series  $\{x(t)\}$  can be considered as a projection of a trajectory of a dynamical system, evolving in a measurable state space. A. N. Kolmogorov introduced the theoretical concept of classification of dynamical systems by information rates. The Kolmogorov-Sinai entropy (KSE thereafter) or metric entropy<sup>5</sup> is a topological invariant, suitable for the classification of dynamical systems or their states, and is related to the sum of the system’s positive Lyapunov exponents<sup>24</sup>. The concept of entropy rates is common to theories based on philosophically opposite assumptions (randomness vs. determinism) and is ideally applicable for the characterization of complex processes, where possibly deterministic rules are always accompanied by random influences.

## C. Information in causality

In practical applications one deals with time series  $\{x(t)\}$  and  $\{y(t)\}$  which can be considered as realizations of stationary, ergodic stochastic processes  $\{X(t)\}$  and  $\{Y(t)\}$ . Alternatively, the time series  $\{x(t)\}$  and  $\{y(t)\}$  can be understood as one-dimensional projections of trajectories of dynamical systems  $\dot{\mathbf{X}} = f_X(\mathbf{X}, \mathbf{Y})$  and  $\dot{\mathbf{Y}} = f_Y(\mathbf{Y}, \mathbf{X})$ , where  $\mathbf{X}$  and  $\mathbf{Y}$  are vectors of dimensions  $d_1$  and  $d_2$ , respectively.

Paluš *et al.*<sup>14</sup> studied synchronization of chaotic dynamical systems using tools from information theory. The route to synchronization is considered as a process of adjustment of information rates and the information transferred from system (process)  $\{Y(t)\}$  to system (process)  $\{X(t)\}$  is measured using the conditional mutual information  $I(\mathbf{Y}; \mathbf{X}_\tau | \mathbf{X})$ , where  $\mathbf{X} = \mathbf{X}(t)$  and  $\mathbf{X}_\tau = \mathbf{X}(t + \tau)$ . Analogously, the information transferred from system  $\{X\}$  to system  $\{Y\}$  is measured by  $I(\mathbf{X}; \mathbf{Y}_\tau | \mathbf{Y})$ . In the case of unidirectionally coupled systems, Paluš *et al.*<sup>14</sup> interpreted the process in which the master system drives the slave system as a special case of causal influence in the sense of Granger causality. The conditional mutual information was proposed as an information-theoretic formulation and a nonlinear generalization of the Granger causality.

Using the concept of Markov processes, Schreiber<sup>25</sup> introduced a functional of conditional probability distributions called transfer entropy. Paluš and Vejmelka<sup>26</sup> showed that the transfer entropy is equivalent to CMI  $I(\mathbf{X}; \mathbf{Y}_\tau | \mathbf{Y})$ . Barnett *et al.*<sup>27</sup> demonstrated analytically that transfer entropy (i.e., CMI  $I(\mathbf{X}; \mathbf{Y}_\tau | \mathbf{Y})$ ) is equivalent to Granger causality for Gaussian processes. Therefore, causal influence in the Granger sense is often interpreted as information transfer.

If the measurement of information about the future  $X_\tau$  of the process  $\{X\}$ , shifted  $\tau$  time units forward, contained in the process  $\{Y\}$ , is used for testing the existence of a causal link from  $\{Y\}$  to  $\{X\}$ , denoted as  $Y \rightarrow X$ , Paluš & Vejmelka<sup>26</sup> show that the vectors  $\mathbf{X}$  and  $\mathbf{Y}_\tau$  can be replaced with one-dimensional components  $x$  and  $y_\tau$ , and the CMI in the time series representation reads as

$$I(y(t); x(t+\tau)|x(t), x(t-\eta_1), \dots, x(t-(d_1-1)\eta_1)). \quad (10)$$

The condition in CMI (10) must contain complete information about the state of the system  $X$ . According to Takens theorem<sup>28</sup> the state of a  $d_1$ -dimensional dynamical system (a point in the state space) is mapped by the set of time-lagged coordinates  $x(t), x(t-\eta_1), \dots, x(t-(d_1-1)\eta_1)$ , where  $\eta_1$  is the backward time-lag used in the embedding of system  $X$ . This time-lag can be set according to the embedding construction recipe based on the first minimum of the mutual information.<sup>29</sup>

The causal link  $X \rightarrow Y$  is tested in analogy with (10):

$$I(x(t); y(t+\tau)|y(t), y(t-\eta_2), \dots, y(t-(d_2-1)\eta_2)). \quad (11)$$

Many interesting processes in nature exhibit oscillatory behavior. From a mathematical point of view, such systems can be modeled as weakly coupled oscillators, whose interactions can be inferred by analyzing the dynamics of their instantaneous phases.<sup>12</sup> The instantaneous phase of a signal  $s(t)$  can be determined by using the analytic signal concept of Gabor.<sup>30</sup> The analytic signal  $\psi(t)$  is a complex function of time defined as

$$\psi(t) = s(t) + j\hat{s}(t) = A(t)e^{j\phi(t)}. \quad (12)$$

Usually, the imaginary part  $\hat{s}(t)$  of the analytic signal  $\psi(t)$  can be obtained by using the Hilbert transform of  $s(t)$

$$\hat{s}(t) = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau. \quad (13)$$

(P.V. means that the integral is taken in the sense of the Cauchy principal value.)  $A(t)$  is the instantaneous amplitude and the instantaneous phase  $\phi(t)$  of the signal  $s(t)$  is

$$\phi(t) = \arctan \frac{\hat{s}(t)}{s(t)}. \quad (14)$$

Paluš & Stefanovska<sup>31</sup> have shown that the conditional mutual information can be applied also in inference of coupling of oscillatory systems from their instantaneous phases, confined in interval  $[0, 2\pi)$  or  $[-\pi, \pi)$  (so-called wrapped phases). Thus one can use time series  $\{x(t)\}$  and  $\{y(t)\}$  and compute their instantaneous phases  $\phi_1(t)$  and  $\phi_2(t)$ , respectively, according to Eqs. (13) and (14) and then evaluate the conditional mutual information  $I(\phi_1(t); \phi_2(t+\tau)|\phi_2(t))$  and  $I(\phi_2(t); \phi_1(t+\tau)|\phi_1(t))$ . Paluš & Stefanovska<sup>31</sup> demonstrated that better distinction of the direction of causality can be obtained using phase increments

$$\Delta_\tau \phi_{1,2}(t) = \phi_{1,2}(t+\tau) - \phi_{1,2}(t), \quad (15)$$

in conditional mutual information defined as  $I(\phi_1(t); \Delta_\tau \phi_2(t)|\phi_2(t))$  and  $I(\phi_2(t); \Delta_\tau \phi_1(t)|\phi_1(t))$ , in a shorter notation  $I(\phi_1; \Delta_\tau \phi_2|\phi_2)$  and  $I(\phi_2; \Delta_\tau \phi_1|\phi_1)$ , respectively.

#### D. Surrogate data for statistical testing

Finite-sample estimates of mutual information are always nonzero. To assess the presence of causal relationships in the analyzed data, it is therefore appropriate to compare the conditional mutual information values computed from the observed data with ranges of CMI values obtained from uncoupled processes that share the same statistical properties. Within a surrogate data<sup>32,33</sup> testing framework, the original data are manipulated using a randomization procedure that preserves the original frequency spectra or variance across all relevant time scales. In this study, we employ circularly time-shifted surrogate data, which have been shown to be effective for causal inference.<sup>34</sup> For the analyzed time series  $\{x\}$  of the length  $N$ , we generate independent realizations of time-shifted surrogates as follows: For each realization an integer variable  $k < N$  is randomly chosen. Then, by moving the first  $k$  values of  $x(1), x(2), \dots, x(k)$  to the end of the time series, we generate the circular time-shifted surrogate series  $x^{surr}$  as

$$x^{surr} = \{x(k+1), x(k+2), \dots, x(N), x(1), x(2), \dots, x(k)\}. \quad (16)$$

In order to avoid the surrogates very close to the original series,  $k$  is constrained as follows:  $\min(k, N-k) > 500$ .

#### E. Estimating entropies and information

In practice, however, the PDF's used in the definitions of the information-theoretic functionals are not known and we only have a set of measurements, e.g.,  $\{x(1), x(2), \dots, x(N)\}$  for the variable  $X$  and  $\{y(1), y(2), \dots, y(N)\}$  for the variable  $Y$ . Then PDF  $p(x)$  is estimated as probabilities  $p_x(i)$ ,  $i = 1, \dots, m$  of the  $m$  subsets into which the interval  $[\min(x), \max(x)]$  is divided. The number  $n_i$  of samples  $x(t)$  in each interval ("box")  $i$  is counted and  $p_x(i)$  is computed as  $n_i/N$ . PDF  $p(y)$  is estimated in the same way, while the joint PDF  $p(x, y)$  is obtained by counting the number  $n_{ij}$  of pairs  $x, y$  so that the value of  $x$  lies in the subset  $i$  of the interval  $[\min(x), \max(x)]$  and the value of  $y$  lies in the subset  $j$  of the interval  $[\min(y), \max(y)]$  and  $p_{xy}(i, j)$  is computed as  $n_{ij}/N$ . The univariate probability distributions  $p(x) \approx p_x(i)$  and  $p(y) \approx p_y(i)$  are called marginal distributions of the bivariate distribution  $p(x, y) \approx p_{xy}(i, j)$ .

The simplest form of box counting is equidistant binning, obtained by dividing the range of each variable, for example,  $[\min(x), \max(x)]$ , into  $m$  subintervals of equal length. In this study, however, we evaluate conditional mutual information and entropies using a box-counting

algorithm based on marginal equiquantization.<sup>15,35,36</sup> In this approach, the partition is generated adaptively in one dimension (for each variable) so that the marginal bins are approximately equiprobable, that is,  $p(i) \approx 1/m$  for each  $i = 1, \dots, m$ . The only free parameter of this method is the number  $m$  of marginal bins.

Paluš<sup>36</sup> proposed that, when computing mutual information of  $n$  variables, the number of marginal bins should not exceed the  $(n+1)$ -st root of the number of data samples, that is,  $m \leq \sqrt[n+1]{N}$ . Cellucci et al.<sup>37</sup>, who employed the same binning procedure, determined the number of bins using the minimum description length criterion. In this paper, we adopt a simple pragmatic choice<sup>26</sup> and set  $m = 8$  in all computations, unless explicitly stated otherwise.

The equiquantization method effectively transforms each variable to have a uniform marginal distribution, thereby maximizing the individual (marginal) entropies. Mutual information estimates obtained in this way are invariant under any monotonic (possibly nonlinear) transformation of the data.<sup>35</sup>

There are also more sophisticated algorithms for the estimation of entropy and mutual information available,<sup>15</sup> however, we will not consider them here.

## F. Ordinal patterns

Bandt and Pompe<sup>16–18</sup> proposed to encode a continuous-valued time series  $x(t)$ ,  $t = 1, \dots, N$  into a symbolic sequence using ordinal patterns of order  $n$ .  $n$ -tuples  $\{x(t), x(t+\eta), \dots, x(t+(n-1)\eta)\}$  for  $t = 1, \dots, N - (n-1)\eta$  are constructed and turned into symbols according to the ordering of their components. For instance, for  $n = 3$ , if for the triplet  $\{x(t), x(t+\eta), x(t+2\eta)\}$  the ordering  $\{x(t) < x(t+\eta) < x(t+2\eta)\}$  holds, it can be classified as “123” and marked as symbol 1. For  $n = 3$  six different orderings, or permutations can be obtained. In general,  $n!$  permutations or symbols is available to encode time series. Then, the numbers of different symbols are counted and the probability for each symbol is given as  $p_i^n[x] = n_i / (N - (n-1)\eta)$  where  $n_i$  is the number of occurrence of symbol  $i$ ,  $i = 1, \dots, n!$ . Finally, we define the permutation entropy as

$$h_p(x) = - \sum_{i=1}^{n!} p_i^n[x] \log p_i^n[x]. \quad (17)$$

Permutation mutual information can also be defined. Since we consider time lag  $\tau$ , we use a more precise notation. The symbol probability  $p_i^n[x(t)]$  is defined as above, while the symbol probabilities  $p_i^n[y(t+\tau)]$  for the lagged time series  $y(t+\tau)$  are obtained by applying the same ordinal-pattern encoding to the time-shifted series,  $\{y(t+\tau), y(t+\tau+\eta_2), \dots, y(t+\tau+(n-1)\eta_2)\}$ . In the definition (3) of mutual information,  $p(x)$  is replaced by  $p_i^n[x(t)]$ ,  $p(y)$  by  $p_j^n[y(t+\tau)]$ , and  $p(x, y)$  by the joint

probability  $p_{i,j}^n[x(t), y(t+\tau)]$ . Permutation conditional mutual information can be obtained in full analogy. Care must be taken to avoid overlap between ordinal symbols representing the future of the effect (driven) time series and those used in the conditioning variables. This can be achieved either by introducing backward time lags  $\eta_k$  or by replacing the forward lag  $\tau$  with  $\tau + (n-1)\eta_k$ .

In all of the above formulas, either the natural logarithm or the binary logarithm can be used, in which case the resulting measures are expressed in nats or bits, respectively.

## G. Data

As an example of a route to synchronization based on unidirectional information transfer, we consider the unidirectionally coupled Rössler systems, studied in detail, for example, by Paluš & Vejmelka.<sup>26</sup> The autonomous system  $X$  is defined as

$$\begin{aligned} \dot{x}_1(t) &= -\omega_1 x_2(t) - x_3(t) \\ \dot{x}_2(t) &= \omega_1 x_1(t) + 0.15 x_2(t) \\ \dot{x}_3(t) &= 0.2 + x_3(t)[x_1(t) - 10] \end{aligned} \quad (18)$$

and the driven system  $Y$  as

$$\begin{aligned} \dot{y}_1(t) &= -\omega_2 y_2(t) - y_3(t) + \epsilon[x_1(t) - y_1(t)] \\ \dot{y}_2(t) &= \omega_2 y_1(t) + 0.15 y_2(t) \\ \dot{y}_3(t) &= 0.2 + y_3(t)[y_1(t) - 10] \end{aligned} \quad (19)$$

where  $\omega_1 = 1.015$ ,  $\omega_2 = 0.985$ . The first component of  $Y$  contains the diffusive coupling term  $\epsilon[x_1(t) - y_1(t)]$ , where  $\epsilon$  is the coupling strength.

The unidirectionally coupled Rössler systems were integrated using the adaptive Bulirsch-Stoer method<sup>38</sup>. It uses an adaptive integration step, however, the final sampling time 0.314 was prescribed, which leads to approximately 20 samples per period. After the initialization, 5000 samples were discarded and the following 65536 samples were recorded and used for the CMI computations, using variable time series length from 8092 to 65536 samples. The first system was initialized using coordinates from the Rössler attractor (11.120979, 17.496796, 51.023544), for the second these values were multiplied by a random number between 0.5 and 1.5 [e.g.  $y_1(0) = 11.120979 * (\text{ran}(\text{iseed}) + 0.5)$ ]. The experiments with the Bulirsch-Stoer method were repeated for different sets of random initial conditions. Due to discarding of the transient data, all runs gave equivalent results, i.e., the presented results do not depend on initial conditions.

The parameters of the systems were chosen as in the previous study.<sup>26</sup>

We also consider probably the simplest demonstration of “ $X$  Granger causes  $Y$ ” – a bivariate, order one autore-

gressive model (AR1)

$$x(t) = a_1 x(t-1) + \xi_1(t) \quad (20)$$

$$y(t) = b_1 y(t-1) + c_1 x(t-1) + \xi_2(t) \quad (21)$$

where  $\xi_{1,2}$  are independent, independently distributed (IID) normal random deviates with zero mean and variance  $\sigma^2 = 0.1$ ,  $a_1 = 0.7$ ,  $b_1 = 0.5$  and  $c_1 = 0.2$ . The parameters of the AR model (20,21) were chosen such that the model provides long runs of nontrivial stationary time series. Equivalent results can be obtained also for different sets of parameters. The first ten thousand iterations were discarded and then 65536 samples recorded.

Studying the entropy rates, we use single dynamical systems, first the discrete logistic map:<sup>39</sup>

$$x_{n+1} = ax_n(1 - x_n); \quad (22)$$

and the continuous Lorenz system:<sup>40</sup>

$$\begin{aligned} \dot{x}(t) &= \sigma[y(t) - x(t)] \\ \dot{y}(t) &= rx(t) - y(t) - x(t)z(t) \\ \dot{z}(t) &= x(t)y(t) - bz(t) \end{aligned} \quad (23)$$

with  $\sigma = 16$ ,  $b = 4$ .

Iterating the logistic map, the first ten thousand iterations were discarded and then 65536 samples recorded. The Lorenz system was integrated again using the adaptive Bulirsch-Stoer method,<sup>38</sup> just the sampling time was set to 0.01. The entropy rates were computed for different system states, obtained by varying the parameter  $a$  of the logistic map, and  $r$  of the Lorenz system.

### III. RESULTS

#### A. Components of the unidirectionally coupled Rössler systems

The route to synchronization for the unidirectionally coupled Rössler systems (18,19) is illustrated in Fig. 1, presenting Lyapunov exponents (LE) and causality/information transfer measures as functions of the coupling strength  $\varepsilon$ .

Three-dimensional chaotic systems, such as each of the studied Rössler systems, are characterized by three Lyapunov exponents:<sup>39</sup> one positive exponent indicating sensitivity to initial conditions, one zero exponent associated with motion along the trajectory, and one negative exponent corresponding to contraction in state space. The positive LE (black line in Fig. 1a) and the zero LE (blue line) of the driving system  $X$  remain constant for all values of  $\varepsilon$ , while both positive and zero LE for the driven system  $Y$  (solid red and dashed green lines, respectively) decrease, although nonmonotonously, with increasing  $\varepsilon$ . When the positive LE of the driven system (the solid red

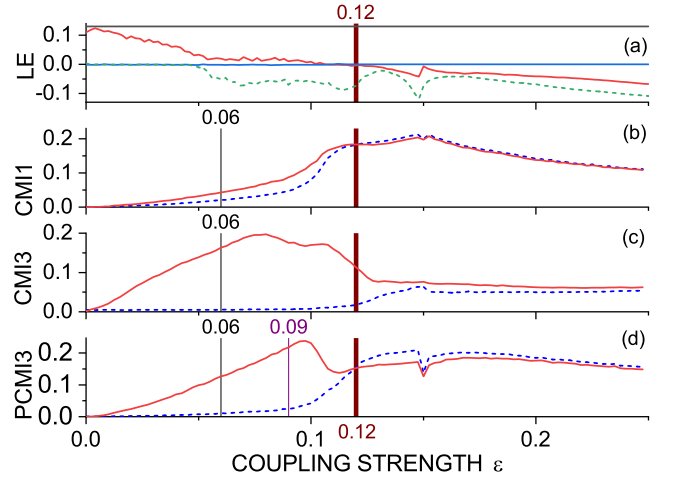


FIG. 1. (a) Two largest Lyapunov exponents of the driver  $X$  (black and blue constant lines) and the response system  $Y$  (decreasing curves, originally positive LE is marked by solid red line, originally zero LE by dashed green line). (b) Averaged conditional mutual information with 1-dimensional condition for the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). (c) Averaged conditional mutual information with 3-dimensional condition for the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). (d) Permutation conditional mutual information of order 3 for the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). All results for the unidirectionally coupled Rössler systems are plotted as functions of the coupling parameter  $\varepsilon$ . CMI and PCMI curves were obtained as averages for  $\tau = 1, \dots, 50$  samples.

line in Fig. 1a) becomes negative, the systems synchronize.

In the synchronized state, the trajectories of both systems become topologically equivalent, making it impossible to determine the direction of coupling; see the discussion and references in Coufal *et al.*<sup>41</sup> Consequently, for systems capable of synchronization, the direction of coupling can be inferred from experimental data only when the systems are coupled but not yet synchronized.<sup>26,42,43</sup> In the studied case (Fig. 1) the synchronization threshold appears at  $\varepsilon = 0.12$ .

Figure 1b presents CMI  $I(x(t); y(t+\tau)|y(t))$  for the causal direction  $X \rightarrow Y$  (solid red line) and  $I(y(t); x(t+\tau)|x(t))$  for the causal direction  $Y \rightarrow X$  (dashed blue line) with the one-dimensional condition (CMI1 thereafter). CMI1 increases with increasing  $\varepsilon$  in both directions, which could be incorrectly interpreted as a bidirectional coupling.

Using CMI with three-dimensional condition (CMI3 thereafter),

$$I(x(t); y(t+\tau)|y(t), y(t-\eta_2), y(t-2\eta_2)) \quad (24)$$

for the causal direction  $X \rightarrow Y$  (solid red line), and

$$I(y(t); x(t+\tau)|x(t), x(t-\eta_1), x(t-2\eta_1)) \quad (25)$$

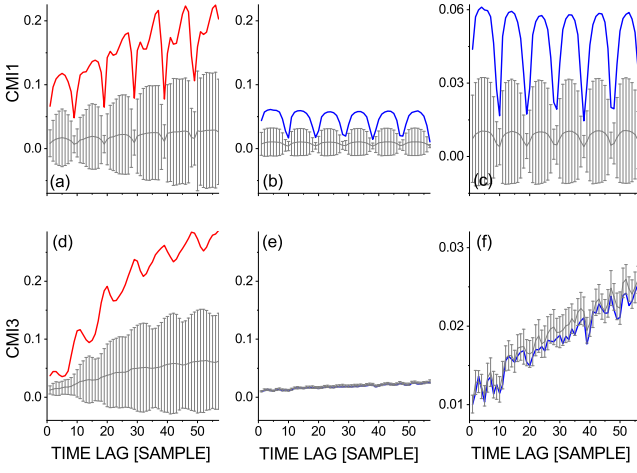


FIG. 2. (a) Conditional mutual information  $I(x(t); y(t + \tau) | y(t))$  in the direction  $X \rightarrow Y$  with the 1-dimensional condition (red curve, a) and  $I(y(t); x(t + \tau) | x(t))$  in the opposite direction (blue, b) on the common scale with (a); (c) on the individual scale); CMI  $I(x(t); y(t + \tau) | y(t), y(t - \eta_2), y(t - 2\eta_2))$  in the direction  $X \rightarrow Y$  with the 3-dimensional condition (red curve, d) and  $I(y(t); x(t + \tau) | x(t), x(t - \eta_1), x(t - 2\eta_1))$  in the opposite direction (blue, e) on the common scale with (d); (f) on the individual scale); plotted as functions of time lag  $\tau$  for the coupling strength  $\varepsilon = 0.06$  and time series length 32768 samples. Gray curves and whiskers represent the mean and mean  $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations.

for the causal direction  $Y \rightarrow X$  (dashed blue line in Fig. 1c), we obtain the correct result: CMI3 detects positive causal information flow in the direction  $X \rightarrow Y$  (solid red line), while in the direction  $Y \rightarrow X$  (dashed blue line) it remains near the zero level up to the synchronization threshold. This result conforms with the above requirement that the condition in CMI (10,11) must contain complete information about the state of the effect system which, in our case, is three-dimensional. The embedding construction lags were obtained from the first minimum of the mutual information<sup>29</sup>  $\eta_1 = \eta_2 = 5$  samples.

Figure 1d presents the results using the permutation conditional mutual information of order three (PCMI3 thereafter). PCMI3 seems to give an equivalent distinction of the causal direction as CMI3, just it ends earlier before the synchronization threshold. Apparently the ordinal patterns of order three successfully encode the states of three-dimensional dynamical systems, however, the pattern components  $\{x(t), x(t + \eta), x(t + 2\eta)\}$  should be properly embedded coordinates. Therefore the choice of  $\eta$  matters, and the same criterion as for the CMI3 conditions gives  $\eta = 5$  samples.

We now turn the above visual considerations into a quantitative analysis using statistical tests based on surrogate data. Choosing  $\varepsilon = 0.06$  we plot CMI in each direction as functions of the forward lag  $\tau$  and compare the results from the original Rössler data with the CMI

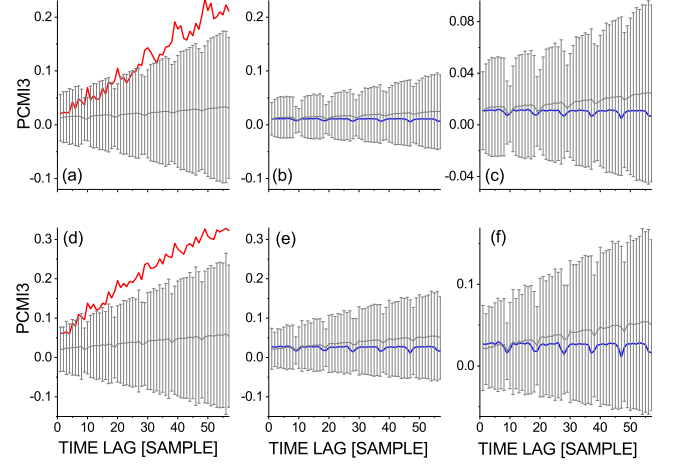


FIG. 3. (a) Permutation conditional mutual information of the order 3 in the direction  $X \rightarrow Y$  (red curve, a) and PCMI in the opposite direction (blue, b) on the common scale with (a); (c) on the individual scale); plotted as functions of time lag  $\tau$  for the coupling strength  $\varepsilon = 0.06$  and time series length 32768 samples. Gray curves and whiskers represent the mean and mean  $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations. (d, e, f): The same as (a, b, c), respectively, but for  $\varepsilon = 0.09$ .

ranges obtained from 30 realizations of circularly shifted surrogate data. The latter are presented as means  $\pm 2$ SD (SD=standard deviation) using gray lines and whiskers. The results for the direction  $Y \rightarrow X$  (blue line) without coupling are plotted twice, once on the common scale with the results for the causal direction  $X \rightarrow Y$  (red line) and then on their individual scales to enhance visual separation from the surrogate ranges.

The statistical tests confirm the suspicion suggested by Fig. 1b: CMI1 detects bidirectional coupling, since its values exceed the surrogate threshold mean  $\pm 2$ SD for large intervals of time lag  $\tau$  (Fig. 2a,b,c). Using CMI3 (Fig. 2d,e,f) the causality in the causal direction  $X \rightarrow Y$  (red line in Fig. 2d) is statistically significant for all considered time lags  $\tau$ , while there is no significant causality in the direction  $Y \rightarrow X$  without coupling (blue line in Fig. 2e,f).

PCMI3 for  $\varepsilon = 0.06$  (Fig. 3a,b,c) confirms the results of CMI3 (Fig. 2d,e,f), however, PCMI3 is statistically significant only for larger time lags  $\tau$  (Fig. 3a). Upon increasing the coupling strength to  $\varepsilon = 0.09$  (Fig. 3d,e,f), the  $\tau$ -range of significant PCMI3 extends also to smaller  $\tau$  values. These observations raise some doubts about the sensitivity of PCMI3 compared with CMI3. Nevertheless, the specificity of both measures remains high, and both confirm the absence of significant causality in the uncoupled direction  $Y \rightarrow X$ .

We continue the investigation of the sensitivity of both CMI3 and PCMI3 in detecting causality in the causal direction  $X \rightarrow Y$ . While the above results were obtained using 32768 samples of the time series  $\{x\}$  and  $\{y\}$ , now we compute CMI3 and PCMI3 using 8092 samples. CMI3

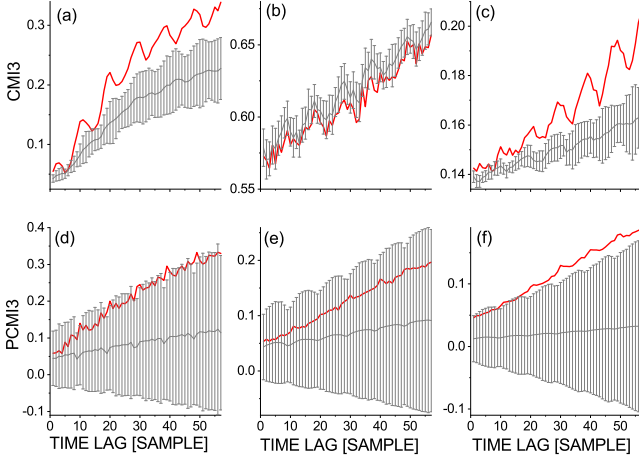


FIG. 4. (a) CMI3  $I(x(t); y(t+\tau)|y(t), y(t-\eta_2), y(t-2\eta_2))$  in the direction  $X \rightarrow Y$  with the 3-dimensional condition (red curve) plotted as a function of time lag  $\tau$  for the coupling strength  $\varepsilon = 0.06$  and time series length 8092 samples; (b) the same as (a) but for data with 30% of additive Gaussian noise; (c) CMI  $I(x(t); y(t+\tau)|y(t), y(t-\eta_2), y(t-2\eta_2))$  for time series length 65536 samples and 30% of additive Gaussian noise. Gray curves and whiskers represent the mean and mean $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations. (d, e, f): The same as (a, b, c), respectively, but for PCMI of order 3 and  $\varepsilon = 0.09$ .

still correctly detect the causal coupling (Fig. 4a), however, PCMI3 fails (Fig. 4d). Next, time series are contaminated with 30% of additive Gaussian noise, using independent noise realizations for each time series. Neither CMI3 (Fig. 4b), nor PCMI3 (Fig. 4e) is able to detect the present causality from 8092 samples of noisy time series. Increasing the time series length to 65536 samples, the detection ability of both the measures is restored, however, CMI3 (Fig. 4c) provides more robust detection than PCMI3 (Fig. 4f).

Although the behavior of both estimators across different noise levels merits a systematic study, we restrict our analysis to 30% noise as a representative example of a non-negligible perturbation. Such noise can effectively mask causal interactions in short datasets, leading to a failure of causality detection. However, the causal information remains present in the data and becomes detectable again when longer time series are available, providing sufficient statistical power for reliable inference.

## B. Phases of the unidirectionally coupled Rössler systems

Since the Rössler time series  $\{x(t)\}$  and  $\{y(t)\}$  have oscillatory character, we compute their instantaneous phases  $\phi_1(t)$  and  $\phi_2(t)$ , respectively, according to Eqs. (13) and (14) and then evaluate the conditional mutual information  $I(\phi_1(t); \Delta_\tau \phi_2(t) | \phi_2(t))$  and  $I(\phi_2(t); \Delta_\tau \phi_1(t) | \phi_1(t))$  using the phase increments (15). They are presented, as functions of the coupling param-

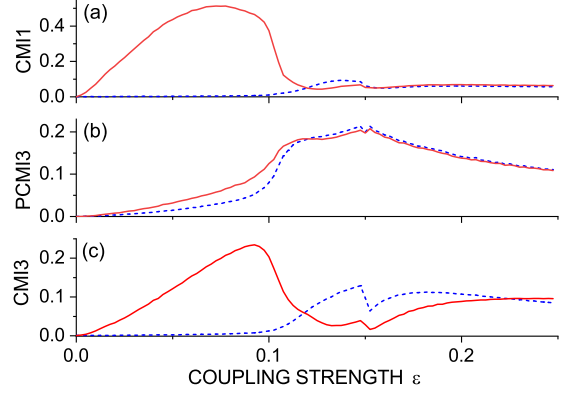


FIG. 5. (a) Averaged conditional mutual information for the instantaneous phases with 1-dimensional condition for the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). (b) Permutation conditional mutual information of order 3 for the instantaneous phases in the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). (c) Averaged conditional mutual information for the instantaneous phases with 3-dimensional condition for the  $X \rightarrow Y$  direction (solid red line) and for the opposite direction (dashed blue line). All results for the unidirectionally coupled Rössler systems are plotted as functions of the coupling parameter  $\varepsilon$ . CMI and PCMI curves were obtained as averages for  $\tau = 1, \dots, 50$  samples.

eter  $\varepsilon$ , in Fig. 5a. We can see that the phases provide clear distinction of the causal direction already with the one-dimensional condition.

Then we apply PCMI3 on the instantaneous phases  $\phi_1(t)$  and  $\phi_2(t)$  (Fig. 5b). Using the phases, PCMI3 apparently loses its ability to reliably infer the correct direction of coupling. In order to have a fair comparison, we apply also CMI3 on the phases: In formulations (24) and (25) we just replace  $x(t)$  by  $\phi_1(t)$  and  $y(t)$  by  $\phi_2(t)$  (Fig. 5c). Increasing the dimensionality of the conditions in the phase CMI does not jeopardize the correct inference of the causal direction.

The visual consideration of Fig. 5 is again complemented with the quantitative testing. Phase CMI1  $I(\phi_1; \Delta_\tau \phi_2 | \phi_2)$  for  $\varepsilon = 0.06$  as a function of time lag  $\tau$  is presented in Fig. 6a by the red line, compared with the surrogate range mean $\pm 2$ SD (gray line and whiskers). CMI  $I(\phi_2; \Delta_\tau \phi_1 | \phi_1)$  in the opposite direction is again plotted on the common scale (Fig. 6b) with the causal direction in Fig. 6a, and on the individual scale in Fig. 6c. The phase CMI1 correctly identifies the statistically significant information transfer only in the coupled direction  $X \rightarrow Y$ .

The results are different when PCMI3 is applied on the phases (Fig. 6d-f). Although the values in the uncoupled direction  $Y \rightarrow X$  (Fig. 6e) are smaller than those in the coupled direction  $X \rightarrow Y$  (Fig. 6d), in both directions they are statistically significant. The phase PCMI is unable to reliably infer the correct coupling direction

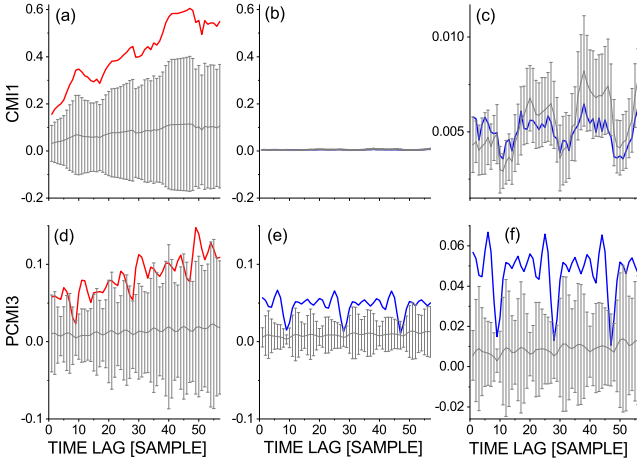


FIG. 6. Conditional mutual information for the instantaneous phases  $I(\phi_1; \Delta_\tau \phi_2 | \phi_2)$  in the direction  $X \rightarrow Y$  with the 1-dimensional condition (red curve, a) and  $I(\phi_2; \Delta_\tau \phi_1 | \phi_1)$  in the opposite direction (blue, (b) on the common scale with (a); (c) on the individual scale); permutation conditional mutual information of order 3 for the instantaneous phases in the direction  $X \rightarrow Y$  (red curve, d) and PCMI3 in the opposite direction (blue, (e) on the common scale with (d); (f) on the individual scale); plotted as functions of time lag  $\tau$  for the coupling strength  $\varepsilon = 0.06$  (a,b,c) and  $\varepsilon = 0.09$  (d,e,f), for phase time series of 32768 samples. Gray curves and whiskers represent the mean and mean  $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations.

of unidirectionally coupled systems.

### C. A simple stochastic system

In addition to analysis of causality in nonlinear dynamical system, we present results from an example of simple linear Granger causality: a bivariate, order one autoregressive model (20,21) in which  $x(t)$  depends on  $x(t-1)$  while  $y(t)$  depends on both  $x(t-1)$  and  $y(t-1)$ . CMI1  $I(x(t); y(t+\tau)|y(t))$  in the direction  $X \rightarrow Y$  (Fig. 7a) highly exceeds the surrogate range mean  $\pm 2$ SD for lags 1–6 samples, indicating “short-living” but statistically significant causal influence  $X \rightarrow Y$ . CMI1 in the opposite direction  $Y \rightarrow X$  (Fig. 7b, plotted on the common scale with Fig. 7a, and on its individual scale in Fig. 7c) is confined within the surrogate range mean  $\pm 2$ SD, implying no causality in the direction  $Y \rightarrow X$ . The one-dimensional condition is sufficient for this simple example, however, we also present the results using CMI3 (Fig. 7d,e,f) demonstrating that an extended condition dimension does not change the correct results, providing enough data is available for the estimation of higher dimensional PDFs. 65536 samples were used.

The results differ using PCMI3, exceeding the surrogate range mean  $\pm 2$ SD for small lags in both causal directions (Fig. 7g,h,i). Thus PCMI3 indicates bidirectional causality in simple unidirectionally coupled AR1 systems.

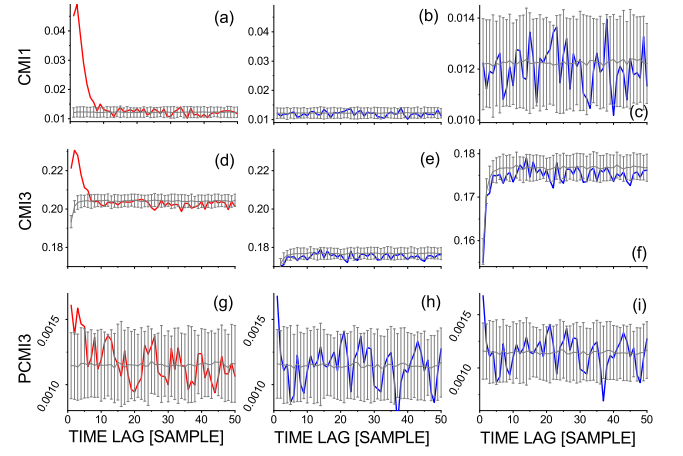


FIG. 7. Conditional mutual information  $I(x(t); y(t+\tau)|y(t))$  in the direction  $X \rightarrow Y$  with the 1-dimensional condition (red curve, a) and  $I(y(t); x(t+\tau)|x(t))$  in the opposite direction (blue, (b) on the common scale with (a); (c) on the individual scale); CMI  $I(x(t); y(t+\tau)|y(t), y(t-\eta_2), y(t-2\eta_2))$  in the direction  $X \rightarrow Y$  with the 3-dimensional condition (red curve, d) and  $I(y(t); x(t+\tau)|x(t), x(t-\eta_1), x(t-2\eta_1))$  in the opposite direction (blue, (e) on the common scale with (d); (f) on the individual scale). Permutation conditional mutual information of the order 3 in the direction  $X \rightarrow Y$  (red curve, g) and PCMI in the opposite direction (blue, (h) on the common scale with (g); (i) on the individual scale); for the coupled AR1 model plotted as functions of time lag  $\tau$  for the time series length 65536 samples. Gray curves and whiskers represent the mean and mean  $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations.

### D. Entropy rates of dynamical systems and Gaussian processes

While the previous section focused on evaluating information transfer between systems, we now turn to the measurement of information produced by systems during their evolution. The appropriate measure is the entropy rate (8), which for dynamical systems is formulated as the Kolmogorov–Sinai entropy (KSE).<sup>10</sup> We consider two example systems, each characterized by a single positive Lyapunov exponent (LE), which, according to Pesin’s theorem,<sup>24</sup> is equal to KSE. Bandt<sup>19</sup> proved that permutation entropy converges to KSE for a certain class of dynamical systems. For Gaussian processes, the entropy rate can be expressed analytically by Eq. (9), which we will refer to as the Gaussian entropy rate (GS ER). Paluš<sup>23</sup> demonstrated that GS ER provides the same classification of states of nonlinear chaotic systems as that given by the KSE/LE. Therefore, for different states of chaotic systems we compare LE, GS ER and permutation entropy of order 3 (PE3).

Different states of the logistic and Lorenz systems were obtained by varying parameter  $a$  in (22), and parameter  $r$  in (23), respectively. All three measures for the logistic (Fig. 8) and Lorenz (Fig. 9) systems provide consistent results: Increasing the parameter  $a$  (or  $r$ ), the “chaotic-

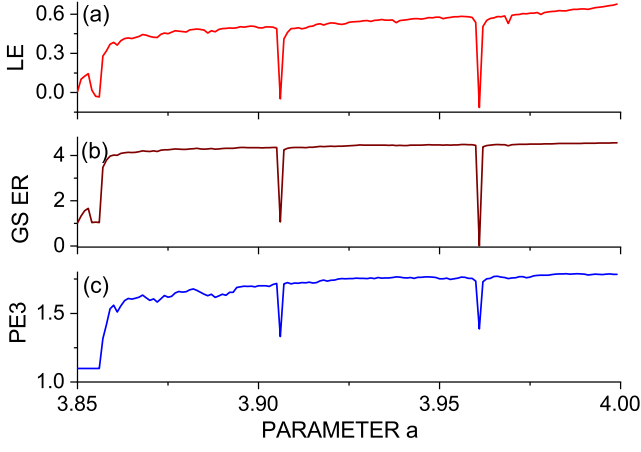


FIG. 8. (a) The Lyapunov exponent; (b) entropy rate for Gaussian processes (GS ER); and (c) permutation entropy of order 3 (PE3) for the logistic system plotted as functions of the parameter  $a$ .

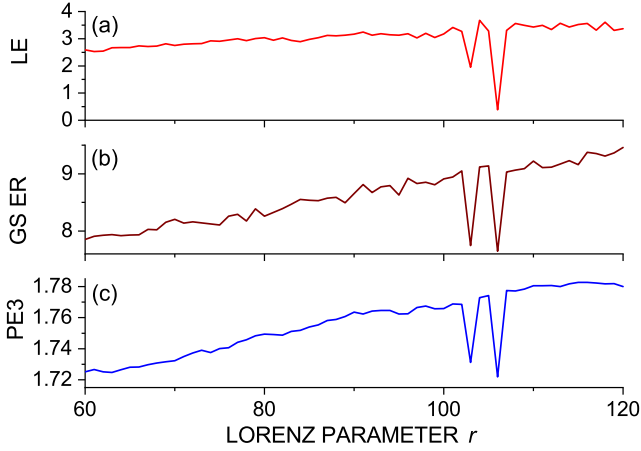


FIG. 9. (a) The largest Lyapunov exponent; (b) entropy rate for Gaussian processes (GS ER); and (c) permutation entropy of order 3 (PE3) for the Lorenz system plotted as functions of the parameter  $r$ .

ity” or entropy rate of the systems increases, however, the increase is interrupted by bifurcations into periodic or weakly chaotic states. Now, consider the simple stochastic system – the single AR1 process (20). By varying the parameter  $a_1$  (20), we change the relative contribution of the deterministic part of the AR1 process, thereby modulating its memory and, consequently, its entropy rate. This effect is confirmed by GS ER shown in Fig. 10a. Permutation entropy PE3 (Fig. 10c) also reflects the decrease of the entropy rate with increasing  $a_1$ .

Next, we consider measurement noise in addition to the dynamical noise  $\xi_1(t)$  in (20) by adding 30% additive Gaussian noise to the original AR1 series. This allows us to test the robustness of entropy rate measures to observational noise and to assess how well they capture the underlying dynamics under realistic measurement condi-

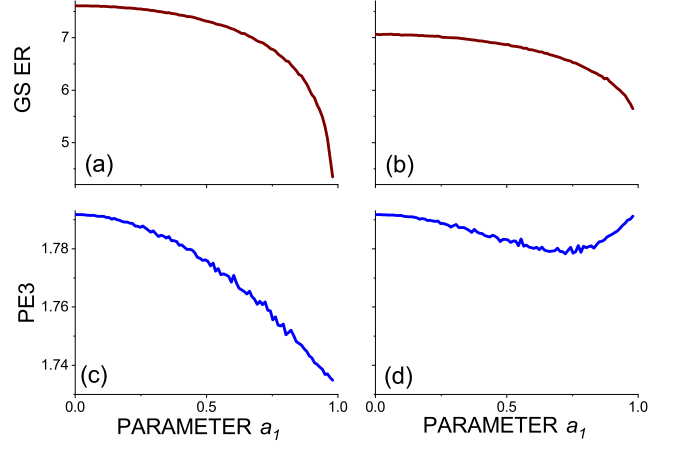


FIG. 10. (a, b) Entropy rate for Gaussian processes (GS ER); (c, d) permutation entropy of order 3 (PE3) for the AR1 process (a, c); and for the AR1 process with 30% of additive Gaussian noise (b, d); plotted as functions of the parameter  $a_1$ .

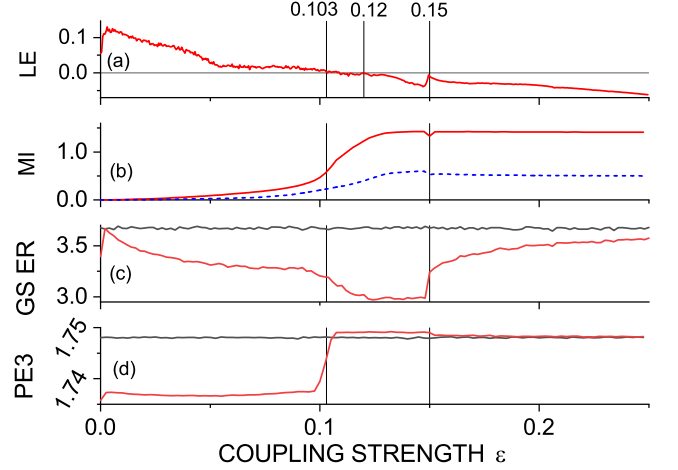


FIG. 11. (a) The largest Lyapunov exponent of the response system  $Y$  (decreasing solid red line); (b) averaged mutual information of components  $x(t)$  and  $y(t)$  (dashed blue line) and phases  $\phi_1(t)$  and  $\phi_2(t)$  (solid red line); (c) entropy rate for Gaussian processes (GS ER); and (d) permutation entropy of order 3 (PE3) for the driver  $X$  (approximately constant black line) and the response system  $Y$  (red line) for the unidirectionally coupled Rössler systems plotted as functions of the coupling parameter  $\varepsilon$ .

tions. The additive IID noise decreases the maximum values of entropy rate and increases its low values, as measured by GS ER (Fig. 10b), i.e. it reduces the range of GS ER, however, the decreasing dependence of the entropy rate on the increasing parameter  $a_1$  is preserved. Permutation entropy PE3 (Fig. 10d) is able to follow the decreasing entropy rate only up to  $a_1 \approx 0.5$ , then it starts to increase and does not follow the variation of the exact entropy rate.

Above, in Sec. III A, we studied information trans-

fer, or causality, along the route to synchronization for the unidirectionally coupled Rössler systems (18,19). Information transfer between the systems represents one aspect of the route to synchronization, while the other involves the adjustment of their information or entropy rates.<sup>14</sup> We therefore return to the unidirectionally coupled Rössler systems (18,19) and evaluate their LE, level of synchronization and entropy rates GS ER and PE3 as functions of the coupling parameter  $\varepsilon$ . The detailed evolution of the largest Lyapunov exponent of the response system  $Y$  is presented in Fig. 11a (decreasing solid red line). As identified above,  $\varepsilon = 0.12$  corresponds to the synchronization threshold; however, the transition to synchronization extends over a broader interval of the coupling parameter  $\varepsilon$ . LE first reaches zero at approximately  $\varepsilon = 0.103$ , then fluctuates around zero, and becomes persistently negative only for  $\varepsilon > 0.12$ . Even then, its decrease is followed by a sharp increase to values close to zero before  $\varepsilon = 0.15$ ; only for  $\varepsilon > 0.15$  does the Lyapunov exponent decrease monotonically in the negative domain. This behaviour is reflected in the measures of synchronization – the mutual information  $I[x(t); y(t+\tau)]$  (dashed blue line in Fig. 11b) and the MI of the phases  $I[\phi_1(t); \phi_2(t+\tau)]$  (solid red line in Fig. 11b), averaged over  $\tau = -20, \dots, 0, \dots, 20$ ; which increase their rate of growth at approximately  $\varepsilon = 0.103$ . After a dip around  $\varepsilon = 0.15$ , both measures saturate on constant values, indicating stable synchronization. The entropy rates of the driven system  $Y$ , GS ER (Fig. 11c) and PE3 (Fig. 11d), initially track the behaviour of its positive LE (Fig. 11a). However, shortly before  $\varepsilon = 0.1$  their rates of change diverge, leading to mutual inconsistency between the two measures during the transient state. Then, for  $\varepsilon > 0.15$ , both GS ER and PE3 indicate that the entropy rate of the driven system  $Y$  converges to the entropy rate of the driver  $X$  (Fig. 11c, d).

#### IV. DISCUSSION

Using the route to synchronization of the unidirectionally coupled Rössler systems (18,19), we demonstrated in Sec. III A that, when conditional mutual information is employed as a measure of causality, the condition must adequately represent the state of the effect system. Specifically, for a  $d$ -dimensional effect system, the conditioning variable must also be  $d$ -dimensional. For the example using the three-dimensional Rössler system, CMI with the three-dimensional condition (CMI3) was identified as a reliable causality measure. We confirm this finding also for the permutation CMI of order 3 (PCMI3), following the recommendation of Staniek and Lehnertz<sup>44</sup> and other authors. On the other hand, Bahraminasab *et al.*<sup>45</sup> proposed the ordinal pattern encoding also for applications of CMI or transfer entropy to phase dynamics. Therefore, in Sec. III B we compared CMI1 and CMI3 with PCMI3 using the instantaneous phases of the unidirectionally coupled Rössler oscillators. While CMI1 and

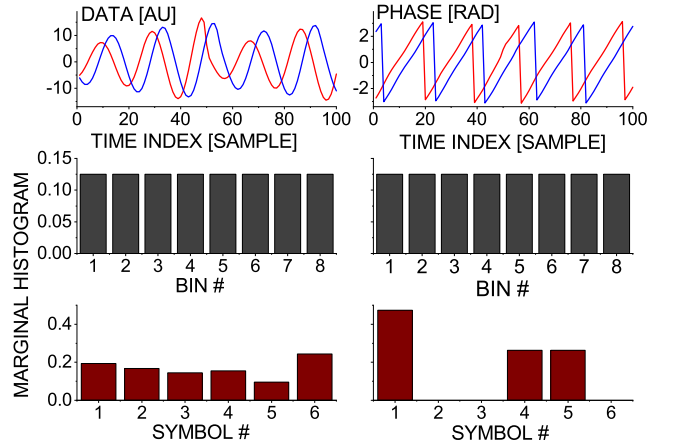


FIG. 12. (Top panel) Illustration of the Rössler data (left) and their instantaneous phases (right). (Middle panel) Equiquantal binning – marginal probabilities/histograms for the data (left) and for the phases (right). (Bottom panel) Ordinal patterns of order 3 – symbol probabilities/histograms for the data (left) and for the phases (right).

CMI3 correctly identified significant causality in one direction, the values of PCMI3 were statistically significant in both directions, including the direction without coupling. To better understand these results, we examine the Rössler data  $x(t)$  and  $y(t)$ , as well as their phases  $\phi_1(t)$  and  $\phi_2(t)$ , shown in Fig. 12 (top panel). Using the equiquantal estimator with 8 marginal bins, all bins are approximately equiprobable for both the raw data and the phases (Fig. 12, middle panel). When ordinal patterns of order 3 are used, the symbols corresponding to  $x(t)$  and  $y(t)$  are not equiprobable, however, all six possible symbols are present (Fig. 12, left bottom panel). In contrast, due to the simple and nearly monotonic dynamics of the phases (Fig. 12, right top panel), the ordinal pattern corresponding to a monotonically increasing sequence,  $\phi_k(t) < \phi_k(t+\eta) < \phi_k(t+2\eta)$  dominates the symbol distribution. Two additional symbols occur with lower probabilities, while the remaining three symbols are absent, i.e., have zero probability (Fig. 12, right bottom panel). Apparently, ordinal patterns do not efficiently encode the phase dynamics.

Figure 13 illustrates how the CMI1 applied to phases  $[I(\phi_1; \Delta_\tau \phi_2 | \phi_2)$  and  $I(\phi_2; \Delta_\tau \phi_1 | \phi_1)]$  distinguishes the correct direction of coupling depending on the number of marginal bins used in the equiquantal estimator. The binary partition (Fig. 13a) clearly detects the increasing coupling with increasing  $\varepsilon$ , however, does not distinguish the coupling direction. Neither the partition with 3 marginal bins would provide a reliable detection of the coupling direction. The latter result (Fig. 13b) is very similar to that obtained when PCMI3 was applied on the instantaneous phases  $\phi_1(t)$  and  $\phi_2(t)$  (Fig. 5b). Using four marginal bins, the two CMI curves are better discernible, however  $I(\phi_2; \Delta_\tau \phi_1 | \phi_1)$  in the uncoupled direction (dashed blue line in Fig. 13c) increases too early,

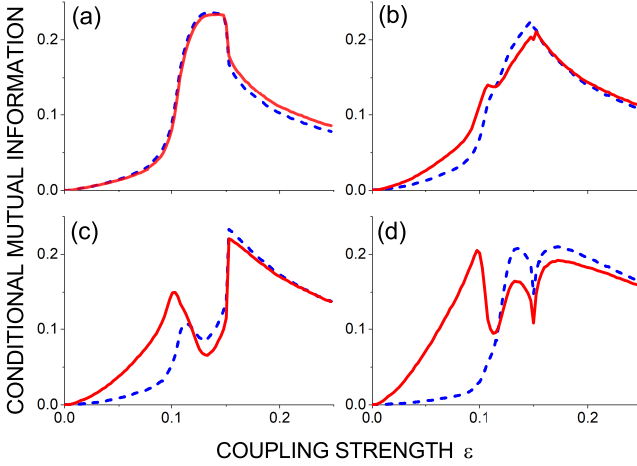


FIG. 13. Averaged conditional mutual information for the instantaneous phases  $I(\phi_1; \Delta_\tau \phi_2 | \phi_2)$  in the direction  $X \rightarrow Y$  with the 1-dimensional condition (solid red curve) and  $I(\phi_2; \Delta_\tau \phi_1 | \phi_1)$  in the opposite direction (dashed blue line), estimated from the equiprobable marginal partitions with 2 (a); 3 (b); 4 (c); and 6 (d) equiquantal bins for the unidirectionally coupled Rössler systems, plotted as functions of the coupling parameter  $\varepsilon$ .

making it prone to false positive results. Only for six marginal bins (Fig. 13d), CMI is a sensitive and specific measure of causality for the coupling strength before the synchronization threshold  $\varepsilon = 0.12$ .

The inefficiency of the ordinal encoding of the phase dynamics is likely only one of the reasons for the false positive results observed with PCMI. The failure of PCMI to correctly identify the causal direction in the case of the simple linear autoregressive model (20,21) (Fig. 7) suggests the presence of an additional, likely more fundamental factor. Ordinal encoding yields a meaningful estimate of a multivariate probability distribution only if the encoded time series corresponds to a component or a projection of a dynamical system's trajectory. This is the same principle as for obtaining “Geometry from a Time Series,” as proposed by Packard et al.<sup>4</sup> and formalized by Takens.<sup>28</sup>

The results in Figure 13 suggests that CMI equiquantal estimator is a reliable indicator of causality when at least six marginal bins are used. Based on our previous experience,<sup>26</sup> we consider eight marginal bins to be a safe choice. This selection allows us to estimate the total number of bins required in a multivariate state space to be adequately covered by the data when inferring causality in  $d$ -dimensional systems. In our formulation of CMI (10,11) we use 2 scalars and the  $d$ -dimensional condition. Then, using  $m$  marginal bins, the total number of bins is  $m^{d+2}$ . The standard permutation CMI contains three symbol variables, then, considering order  $d$ , the total number of symbols is  $(d!)^3$ . Figure 14 illustrates the increase of the total number of bins or symbols with the increasing system dimension. In principle, the ordinal encoding could provide more efficient CMI estimator

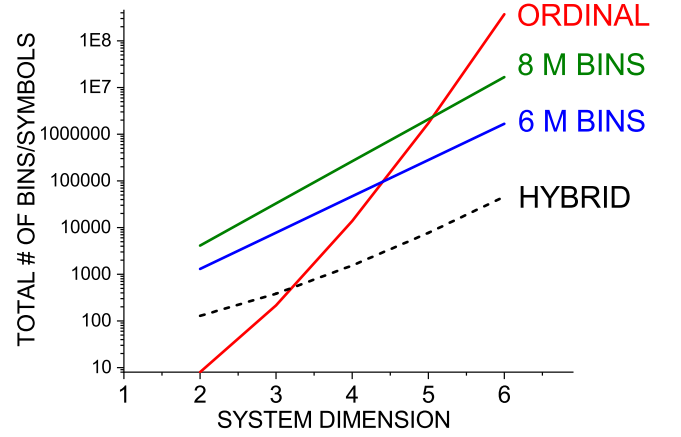


FIG. 14. Total number of symbols or bins required for estimating conditional mutual information for systems of given dimension, using estimators based on ordinal encoding, binning with 8 or 6 marginal bins, and a hybrid estimator combining both approaches.

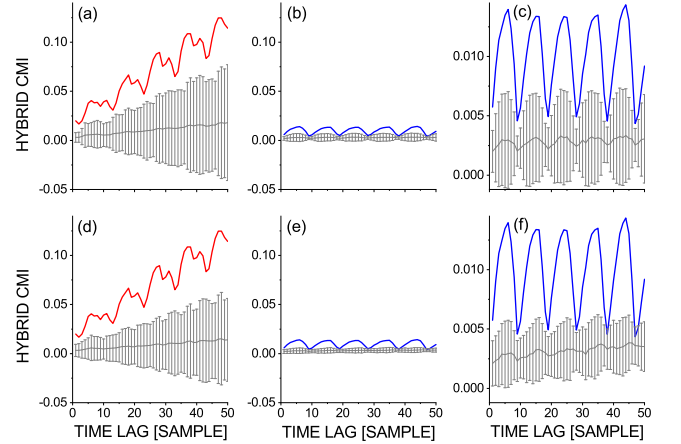


FIG. 15. Conditional mutual information hybrid estimator  $I(x(t); y(t + \tau) | S_y^3(t))$  in the direction  $X \rightarrow Y$  with ordinal symbol in the condition (red curve, a) and  $I(y(t); x(t + \tau) | S_x^3(t))$  in the opposite direction (blue, (b) on the common scale with (a); (c) on the individual scale); (d, e, f) the same as (a, b, c), respectively, but with the first series circularly shifted. All CMI estimates plotted as functions of time lag  $\tau$  for the unidirectionally coupled Rössler systems with the coupling strength  $\varepsilon = 0.06$  and the time series length 65536 samples. Gray curves and whiskers represent the mean and mean  $\pm 2$ SD of the set of 30 circularly shifted surrogate data realizations.

than the partition based on eight marginal bins only for dimensions less than five.

One could consider applying the same “trick” as in our CMI approach: using two scalars and  $d$ -dimensional vector only for the conditioning variable. If the scalars are encoded using 8 bins and the condition is represented as an ordinal symbol, the total number of bins/symbols for which probabilities are estimated is  $8 \cdot 8 \cdot d!$  (dashed black line in Fig. 14). Such a “hybrid” estimator would require

|            |            | BIN | ORD | GAUSS |
|------------|------------|-----|-----|-------|
| Causality  | trajectory | ✓   | ✓   | n/t   |
|            | phase      | ✓   | x   | n/t   |
|            | stochastic | ✓   | x   | n/t   |
| Complexity | trajectory | n/t | ✓   | ✓     |
|            | phase      | n/t | n/t | n/t   |
|            | stochastic | n/t | (✓) | ✓     |

TABLE I. Summary of the results. Estimators: BIN: equiquantal binning; ORD: ordinal encoding; GAUSS: formula for Gaussian processes. Data: trajectory: components or projections of trajectories of low-dimensional dynamical systems; phase: instantaneous phases of oscillatory systems; stochastic: data from stochastic processes. Results: ✓ - recommended; x - not recommended; (✓) - use with caution; n/t - not tested in this study.

far less data than the other methods considered, however, it does not constitute a reliable measure of causality. Figure 15 demonstrates that the “hybrid” estimator suffers from the same false positive results in the uncoupled direction as PCMI does for the cases of the phases and AR1 processes. In order to have “fair” conditions for randomization of the conditioning symbols in the surrogate data, we perform the surrogate test first randomizing (randomly shifting) the series  $y(t)$  (Fig. 15a, b, c) and then the series  $x(t)$  (Fig. 15d, e, f). Some effects on the surrogate ranges in Fig. 15c and Fig. 15f are apparent, however, the result indicating the false positive detection of causality is not changed.

Ordinal patterns, when used to characterize dynamics through permutation entropy, yield more optimistic results than their application in causality analysis. Already, the permutation entropy of order three (PE3) closely tracks the behavior of the positive Lyapunov exponents and the Kolmogorov–Sinai entropy of dynamical systems, as well as the entropy rate of a simple stochastic process. It should be noted, however, that similar results are obtained using the Gaussian process entropy rate (GS ER). Perhaps the most interesting findings arise from the application of PE3 and GS ER to study the adjustment of entropy rates in the unidirectionally coupled Rössler systems along their route to synchronization. The transient to synchronization resembles a second-order phase transition,<sup>46</sup> in which changes in the control parameter  $\varepsilon$  lead to continuous changes in the order parameter – here quantified by mutual information as a measure of synchronization. GS ER and PE3, which reflect the entropy rate of the driven system, exhibit a change in their rate of variation prior to the transition. The change is particularly pronounced for PE3, indicating that permutation entropy can act as a reliable indicator of an impending transition.

We summarize the results of this study in Tab. I. Not all combinations of estimators and data types were evaluated; these cases are marked as n/t in Tab. I. Although a Gaussian estimator for mutual information (6) can be derived<sup>36</sup> and has been used in causality analysis,<sup>34,47</sup> it is not suitable for the nonlinear dynamical systems con-

sidered here. Coarse-grained entropy rates based on the equiquantal binning estimator have also been proposed for characterizing dynamical complexity.<sup>35</sup> For brevity, however, we report only the results obtained using the Gaussian entropy rate estimator.

Readers may expect recommendations on which estimator is most appropriate for particular types of data. Apart from cases where ordinal encoding is unsuitable, we refrain from making universal recommendations. In practice, it is advisable to analyze experimental data using at least two different methods. Agreement between independent approaches increases confidence in the inferred causal relationships. Conversely, when results disagree, identifying the source of the discrepancy and understanding why a particular method fails can provide valuable insight into both the limitations of causality detection techniques and the underlying mechanisms generating the observed data.

## V. CONCLUSION

Inference of causality from time series is a challenging problem in which tools from information theory<sup>15</sup> brought considerable progress. Ordinal patterns were proposed as an efficient alternative to naive estimators based on state space partitions. Our results, however, do not confirm majority of the optimistic claims. Ordinal patterns cannot solve the curse of dimensionality, since from the dimension five the number of ordinal symbols increases faster than the number of bins of coarse-grained partitions found effective in the inference of causality.<sup>26</sup> Ordinal patterns can approximate multivariate probability distribution only in the case of trajectories of dynamical systems, owing to the Takens embedding theorem.<sup>28</sup> Estimators based on the ordinal patterns are less robust to noise than simple binning estimators, and fail if the stochastic dynamics dominates. Ordinal patterns are not suitable using other aspects of dynamical systems than the trajectories, in particular, instantaneous phases.

In contrast to their limited performance in causality analysis, ordinal patterns are highly effective for characterizing dynamics through entropy rates computed using permutation entropy. Permutation entropy, possibly in combination with other entropy rate estimators, appears to be a promising tool for constructing early-warning signals that indicate an approaching system transition. This observation, however, requires further thorough investigation.

## ACKNOWLEDGMENTS

The author would like to thank the anonymous referees for their valuable comments and constructive suggestions. Special thanks are due to Annika Betken and Alexander Schnurr for organizing the inspiring workshop

“Classical Ordinal Patterns and Beyond,” which stimulated many of the ideas presented in this work.

This study was supported by the Czech Science Foundation, project No. 25-18105S.

## DATA AVAILABILITY STATEMENT

All data used in this study were numerically generated. Mathematical formulas are given and numerical methods described in Sec. II G Data.

## AUTHOR DECLARATIONS

### Conflict of Interest

The author has no conflicts to disclose.

- <sup>1</sup>M. Priestley, *Spectral Analysis and Time Series* (Academic Press, New York, 1981).
- <sup>2</sup>M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis* (Academic Press, New York, 1988).
- <sup>3</sup>H. Tong, *Non-Linear Time Series Analysis* (Oxford University Press, Oxford, 1990).
- <sup>4</sup>N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, “Geometry from a time series,” *Phys. Rev. Lett.* **45**, 712–716 (1980).
- <sup>5</sup>K. E. Petersen, *Ergodic theory* (Cambridge University Press, 1989).
- <sup>6</sup>C. E. Shannon, “A mathematical theory of communication,” *The Bell system technical journal* **27**, 379–423 (1948).
- <sup>7</sup>R. Shaw, “Strange attractors, chaotic behavior, and information flow,” *Zeitschrift für Naturforschung A* **36**, 80–112 (1981).
- <sup>8</sup>P. Martien, S. Pope, P. Scott, and R. Shaw, “The chaotic behavior of the leaky faucet,” *Phys. Lett. A* **110**, 399–404 (1985).
- <sup>9</sup>A. M. Fraser, “Information and entropy in strange attractors,” *IEEE Trans. Inf. Theory* **35**, 245–262 (1989).
- <sup>10</sup>A. N. Kolmogorov, “Entropy per unit time as a metric invariant of automorphisms,” in *Dokl. Akad. Nauk SSSR*, Vol. 124 (1959) pp. 754–755.
- <sup>11</sup>L. M. Pecora and T. L. Carroll, “Synchronization in chaotic systems,” *Phys. Rev. Lett.* **64**, 821–824 (1990).
- <sup>12</sup>A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: a universal concept in nonlinear sciences*, Vol. 12 (Cambridge University Press, 2003).
- <sup>13</sup>S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, “The synchronization of chaotic systems,” *Phys. Rep.* **366**, 1–101 (2002).
- <sup>14</sup>M. Paluš, V. Komarek, Z. Hrnčíř, and K. Sterbová, “Synchronization as adjustment of information rates: Detection from bivariate time series,” *Phys. Rev. E* **63**, 046211 (2001).
- <sup>15</sup>K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhat-tacharya, “Causality detection based on information-theoretic approaches in time series analysis,” *Phys. Rep.* **441**, 1–46 (2007).
- <sup>16</sup>C. Bandt and B. Pompe, “The entropy profile—a function describing statistical dependences,” *J. Stat. Phys.* **70**, 967–983 (1993).
- <sup>17</sup>B. Pompe, “Ranking and entropy estimation in nonlinear time series analysis,” in *Nonlinear Analysis of Physiological Data*, edited by H. Kantz, J. Kurths, and G. Mayer-Kress (Springer Berlin Heidelberg, Berlin, Heidelberg, 1998) pp. 67–90.
- <sup>18</sup>C. Bandt and B. Pompe, “Permutation entropy: A natural complexity measure for time series,” *Phys. Rev. Lett.* **88**, 174102 (2002).
- <sup>19</sup>C. Bandt, G. Keller, and B. Pompe, “Entropy of interval maps via permutations,” *Nonlinearity* **15**, 1595 (2002).
- <sup>20</sup>K. Keller and M. Sinn, “Ordinal analysis of time series,” *Physica A* **356**, 114–120 (2005).
- <sup>21</sup>T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, NY, USA, 1991).
- <sup>22</sup>M. S. Pinsker, *Information and information stability of random variables and processes* (Holden-Day, San Francisco, 1964).
- <sup>23</sup>M. Paluš, “On entropy rates of dynamical systems and Gaussian processes,” *Phys. Lett. A* **227**, 301–308 (1997).
- <sup>24</sup>Y. B. Pesin, “Characteristic Lyapunov exponents and smooth ergodic theory,” *Russian Mathematical Surveys* **32**, 55 (1977).
- <sup>25</sup>T. Schreiber, “Measuring information transfer,” *Phys. Rev. Lett.* **85**, 461–464 (2000).
- <sup>26</sup>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).
- <sup>27</sup>L. Barnett, A. B. Barrett, and A. K. Seth, “Granger causality and transfer entropy are equivalent for Gaussian variables,” *Phys. Rev. Lett.* **103**, 238701 (2009).
- <sup>28</sup>F. Takens, “Detecting strange attractors in turbulence,” in *Dynamical systems and turbulence, Warwick 1980* (Springer, 1981) pp. 366–381.
- <sup>29</sup>A. M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” *Phys. Rev. A* **33**, 1134–1140 (1986).
- <sup>30</sup>D. Gabor, “Theory of communication,” *J. IEE London* **93**, 429–457 (1946).
- <sup>31</sup>M. Paluš and A. Stefanovska, “Direction of coupling from phases of interacting oscillators: An information-theoretic approach,” *Phys. Rev. E* **67**, 055201 (2003).
- <sup>32</sup>M. Paluš, “From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics,” *Contemp. Phys.* **48**, 307–348 (2007).
- <sup>33</sup>G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, “Surrogate data for hypothesis testing of physical systems,” *Phys. Rep.* **748**, 1–60 (2018).
- <sup>34</sup>P. Manshour, G. Balasis, G. Consolini, C. Papadimitriou, and M. Paluš, “Causality and information transfer between the solar wind and the magnetosphere-ionosphere system,” *Entropy* **23**, 390 (2021).
- <sup>35</sup>M. Paluš, “Coarse-grained entropy rates for characterization of complex time series,” *Physica D* **93**, 64–77 (1996).
- <sup>36</sup>M. Paluš, “Testing for nonlinearity using redundancies - quantitative and qualitative aspects,” *Physica D* **80**, 186–205 (1995).
- <sup>37</sup>C. J. Cellucci, A. M. Albano, and P. E. Rapp, “Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms,” *Phys. Rev. E* **71**, 066208 (2005).
- <sup>38</sup>W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical recipes. The art of scientific computations* (Cambridge University Press, 1989).
- <sup>39</sup>H. G. Schuster and J. Wolfram, *Deterministic Chaos: An Introduction* (Wiley-VCH, Weinheim, 2005).
- <sup>40</sup>E. Lorenz, “Deterministic nonperiodic flow,” *J. Atmos. Sci.* **20**, 130–141 (1963).
- <sup>41</sup>D. Coufal, J. Jakubík, N. Jajcay, J. Hlinka, A. Krakovská, and M. Paluš, “Detection of coupling delay: A problem not yet solved,” *Chaos* **27**, 083109 (2017).
- <sup>42</sup>M. G. Rosenblum and A. S. Pikovsky, “Detecting direction of coupling in interacting oscillators,” *Phys. Rev. E* **64**, 045202 (2001).
- <sup>43</sup>I. Vlachos and D. Kugiumtzis, “Nonuniform state-space reconstruction and coupling detection,” *Phys. Rev. E* **82**, 016207 (2010).
- <sup>44</sup>M. Staniek and K. Lehnertz, “Symbolic transfer entropy,” *Phys. Rev. Lett.* **100**, 158101 (2008).
- <sup>45</sup>A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. E. McClintock, and H. Kantz, “Direction of coupling from phases of interacting oscillators: A permutation information approach,” *Phys. Rev. Lett.* **100**, 084101 (2008).

<sup>46</sup>P. Jizba, H. Lavička, and Z. Tabachová, “Causal inference in time series in terms of Rényi transfer entropy,” *Entropy* **24**

(2022), 10.3390/e24070855.

<sup>47</sup>M. Paluš, “Cross-scale interactions and information transfer,” *Entropy* **16**, 5263–5289 (2014).