Causality, dynamical systems and the arrow of time

Milan Paluš, Anna Krakovská, Jozef Jakubík, and Martina Chvosteková

Citation: Chaos **28**, 075307 (2018); doi: 10.1063/1.5019944
View online: https://doi.org/10.1063/1.5019944
View Table of Contents: http://aip.scitation.org/toc/cha/28/7
Published by the American Institute of Physics

Causality, dynamical systems and the arrow of time

Milan Paluš,[1,a)] Anna Krakovská,[2] Jozef Jakubík,[2] and Martina Chvosteková[2]
[1]*Institute of Computer Science, Czech Academy of Sciences, Pod Vodárenskou věží 2, Praha 8 182 07, Czech Republic*
[2]*Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava 841 04, Slovak Republic*

(Received 18 December 2017; accepted 4 May 2018; published online 18 July 2018)

Using several methods for detection of causality in time series, we show in a numerical study that coupled chaotic dynamical systems violate the first principle of Granger causality that the cause precedes the effect. While such a violation can be observed in formal applications of time series analysis methods, it cannot occur in nature, due to the relation between entropy production and temporal irreversibility. The obtained knowledge, however, can help to understand the type of causal relations observed in experimental data, namely, it can help to distinguish linear transfer of time-delayed signals from nonlinear interactions. We illustrate these findings in causality detected in experimental time series from the climate system and mammalian cardio-respiratory interactions. *Published by AIP Publishing.* https://doi.org/10.1063/1.5019944

Any scientific discipline strives to explain causes of observed phenomena. Studying phenomena evolving in time and providing measurable quantities which can be registered in consecutive instants of time and stored in datasets called time series brings researchers a possibility to apply modern mathematical methods which can detect possible causal relations between different datasets. Methods based on the so-called Granger causality have been applied in diverse scientific fields from economics and finance, through Earth and climate sciences to research trying to understand the human brain. Chaotic dynamical systems are mathematical models reflecting very complicated behaviour. Recently, cooperative phenomena have been observed in coupled chaotic systems due to their ability to synchronize. On the way to synchronization, the question which system influences other systems emerges. To answer this question, research works successfully applied the Granger causality methods. In this study, we demonstrate that chaotic dynamical systems do not respect the principle of the effect following the cause. We explain, however, that such principle violation cannot occur in nature, only in mathematical models which, on the other hand, can help us to understand the mechanisms behind the experimentally observed causalities.

I. INTRODUCTION

The quest of causality, that is, the identification of cause–effect relationships among events, variables, or processes, is one of the fundamental challenges in natural and social sciences. In modern science, penetrated by computational approaches, a quantitative definition of causality is required. Probably, the first approach to describe causality in measurable, mathematically expressible terms can be traced to the 1950s work of the father of cybernetics, Wiener,[1] who wrote: *For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one.* Later, this concept has been introduced into time series analysis by Granger, the 2003 Nobel prize winner in economy. In his Nobel lecture,[2] he recalled the inspiration by Wiener's work and identified two components of the statement about causality:

1. The cause occurs before the effect.
2. The cause contains information about the effect that is unique, and is in no other variable.

According to Granger, a *consequence* of these statements is that the causal variable can help to forecast the effect variable after other data have been first used.[2] This restricted sense of causality, referred to as *Granger causality*, GC thereafter, characterizes the extent to which a process X_t is leading another process, Y_t, and builds upon the notion of incremental predictability. It is said that the *process* X_t *Granger causes process* Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t.

Granger has mathematically formalized these ideas using linear autoregressive (AR) models (see Sec. II A). Due to possible nonlinear dependence in time series from real-world processes, many authors have proposed various nonlinear generalizations[3] of the GC principle. In the following, we will particularly discuss the generalization of GC based on probability functionals from information theory. The information-theoretic functionals, in their general formulation, are applicable to a broad range of nonlinear processes; however, we will focus on time series generated by nonlinear, possibly chaotic dynamical systems. The observation that the chaotic dynamical systems generate information had led to an interesting and fruitful symbiosis of ergodic theory of dynamical systems and information theory.[4–6] Information theory has also been applied in the very intensive research field of the synchronization of chaotic dynamical systems.[7,8] We will remind information transfer between, and adjustment of information rates of dynamical systems on the route to synchronization.[9] In the case of

a) Electronic mail: mp@cs.cas.cz

unidirectionally coupled dynamical systems, the distinction between the driving and the driven systems has been targeted by a number of data analytic methods[10–14] which are usually also considered as methods for inference of causality.[15] We will consider examples of these methods together with the information-theoretic generalization of the GC principle. We will focus on bivariate time series generated by pairs of interacting dynamical systems; therefore, we will not discuss the above Granger statement 2. In this respect, we recommend readers become familiar with methods for inference of causality in multivariate time series.[16–18]

In this paper, we will focus on the above Granger statement 1 and will study behaviour of causality detection methods under the time reversal, as well as discuss quantitative characterization of time irreversibility of studied process.

Granger causality, as well as three methods for detecting causality in nonlinear systems, is introduced in Sec. II. The principle that the cause precedes the effect and its consequence for the time reversal in a simple linear process is studied in Sec. III. Time arrows and causality in nonlinear systems are analysed in Sec. IV. Time irreversibility is measured in Sec. V. Results are summarized and real-world examples presented in Sec. VI. A conclusion is given in Sec. VII.

II. METHODS

A. Granger causality

The standard test of GC developed by Granger[19] is based on a linear regression model

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \sum_{k=1}^{L} b_{2k} X_{t-k} + \xi_t, \tag{1}$$

where ξ_t are uncorrelated random variables with zero mean and variance σ^2, L is the specified number of time lags, and $t = L+1, \ldots, N$. The null hypothesis that "X_t does not Granger cause Y_t" is supported when $b_{2k} = 0$ for $k = 1, \ldots, L$, reducing Eq. (1) to

$$Y_t = a_o + \sum_{k=1}^{L} b_{1k} Y_{t-k} + \tilde{\xi}_t. \tag{2}$$

B. Information-theoretic approach to Granger causality

Let X be a discrete random variable that can acquire values $x_1, \ldots, x_m$, each with corresponding probability $p_i = p(x_i)$, $i = 1, \ldots m$. The average amount of information gained from a measurement that specifies one particular value x_i is given by the *entropy* $H(X)$:

$$H(X) = -\sum_{i=1}^{m} p_i \log p_i. \tag{3}$$

The *joint entropy* $H(X, Y)$ of two discrete random variables X and Y is defined analogously

$$H(X, Y) = -\sum_{i=1}^{m_X} \sum_{j=1}^{m_Y} p(x_i, y_j) \log p(x_i, y_j). \tag{4}$$

Here $p(x_i, y_j)$ denotes the joint probability that X is in state x_i and Y in state y_j.

The joint entropy may be expressed in terms of *conditional entropy* $H(X \mid Y)$ as $H(X, Y) = H(X \mid Y) + H(Y)$, where

$$H(X \mid Y) = -\sum_{i=1}^{m_X} \sum_{j=1}^{m_Y} p(x_i, y_j) \log p(x_i \mid y_j) \tag{5}$$

and $p(x_i \mid y_j)$ denotes the conditional probability.

Mutual information $I(X, Y)$ between two random variables X and Y is then defined as

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \tag{6}$$

The mutual information (MI) measures the strength of dependence in the sense that (1) $I(X, Y) = 0$ if and only if X is independent of Y, (2) For bivariate normal distributions, $I(X, Y) = \frac{1}{2} \log[1 - \rho^2(X, Y)]$, where ρ is the correlation coefficient between X and Y.

The information-theoretic functional used for the inference of causality is called *conditional mutual information (CMI)*. The CMI between random variables X and Y given Z is then defined as

$$I(X, Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z). \tag{7}$$

For Z independent of X and Y $I(X, Y \mid Z) = I(X, Y)$ holds.

Another useful information-theoretic tool is the *Kullback-Leibler divergence (KLD)*. The KLD $K(p, q)$ quantifies difference between two probability distributions p and q, and is defined as

$$K(p, q) = \sum_{i=1}^{m} p_i \log \left(\frac{p_i}{q_i} \right). \tag{8}$$

This measure is not symmetric and therefore it is not a distance in the mathematical sense. The KLD is always nonnegative and it is zero if and only if the distributions p and q are identical. See Cover and Thomas[20] for details on information theory.

In the above formulas, either natural logarithm or binary logarithm may be considered and the above measures are then given in nats or bits, respectively.

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.*[9] 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 $I(\mathbf{Y}; \mathbf{X}_\tau \mid \mathbf{X})$, where $\mathbf{X} = \mathbf{X}(\mathrm{t})$ and $\mathbf{X}_\tau = \mathbf{X}(\mathrm{t} + \tau)$. Analogously, the information transferred from system $\{X\}$ to system $\{Y\}$ is measured by $I(\mathbf{X}; \mathbf{Y}_\tau \mid \mathbf{Y})$. In the case of unidirectionally coupled systems, Paluš *et al.*[9] interpreted the process of driving the slaved system by the master 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 idea of Markov processes, Schreiber[21] introduced a functional of conditional probability distributions called transfer entropy. Paluš and Vejmelka[22] show that the transfer entropy is equivalent to CMI $I(\mathbf{X};\mathbf{Y}_\tau \mid \mathbf{Y})$. Barnett *et al.*[23] have shown analytically that the transfer entropy (i.e., CMI $I(\mathbf{X};\mathbf{Y}_\tau \mid \mathbf{Y})$) is equivalent to Granger causality for Gaussian processes.

If the measurement of information about the future X_τ of the process $\{X\}$, shifted τ time units forward ("τ-future" thereafter), contained in the process $\{Y\}$ is used for testing the existence of a causal link from $\{Y\}$ to $\{X\}$, denoted as $Y \to X$, Paluš and Vejmelka[22] show that the vectors $\mathbf{X}$ and $\mathbf{Y}_\tau$ can be substituted by one-dimensional components x and y_τ, and the CMI in the time series representation reads as

$$I\{y(t);x(t+\tau) \mid x(t),x(t-\eta_1),\ldots x[t-(d_1-1)\eta_1]\}. \quad (9)$$

The condition in CMI (9) must contain complete information about the state of the system X. According to the Takens theorem,[24] the state of a d_1-dimensional dynamical system (a point in the state space) is mapped by the set of time-lagged coordinates $x(t),x(t-\eta_1),\ldots x[t-(d_1-1)\eta_1]$, where η_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.[25]

The causal link $X \to Y$ is tested in analogy with (9):

$$I\{x(t);y(t+\tau) \mid y(t),y(t-\eta_2),\ldots y[t-(d_2-1)\eta_2]\}. \quad (10)$$

Wibral *et al.*[26] introduced a slightly different formulation for CMI:

$$I\{x(t);y(t+\tau) \mid y(t+\tau-1), y(t+\tau-1-\eta_2),\ldots y[t+\tau-1-(d_2-1)\eta_2]\}, \quad (11)$$

in which the condition moves forward with increasing prediction horizon τ, while in the usual formulation, used also by Paluš and Vejmelka,[22] the condition is kept in the same position for all values of τ.

C. Convergent cross mapping

Convergent cross mapping (CCM)[15] is based on Takens' embedding theorem,[24] exploiting the geometry of attractors of coupled dynamical systems. While the CCM method constructs a map between mutual neighbourhoods in state spaces of the coupled dynamical systems under study, there has been a related research in which statistics of mutual nearest neighbour points have been developed.[10–14]

Let two dynamical systems X and Y be represented by two time series $\{x(t)\}_{t=1}^{L}$ and $\{y(t)\}_{t=1}^{L}$, respectively, having finite length $L \in \mathbb{N}$. For the cross mapping from X to Y, the attractor manifold M_X is constructed as a set of E-dimensional vectors $\mathbf{X}(t) = \{x(t),x(t-\eta),x(t-2\eta),\ldots,x[t-(E-1)\eta]\}$ for $t = 1+(E-1)\eta$ to $t = L$, i.e., $M_X = \{\mathbf{X}(t)\}_{t=1+(E-1)\eta}^{L}$. $E \in \mathbb{N}$ is the so-called embedding dimension, see the supplement of Ref. 15. We find $E+1$ nearest neighbours of $\mathbf{X}(t)$ in M_X and denote their time indices (from closest to farthest) by $t_1,\ldots,t_{E+1}$. These indices will be used in the construction of the cross mapping (12) as follows. We approximate $y(t)$, $t = 1+(E-1)\eta,\ldots,L$ by

$$\hat{y}(t) \mid M_X = \sum_{i=1}^{E+1} w_i y(t_i), \quad (12)$$

where $w_i = u_i / \sum_{j=1}^{E+1} u_j$, $u_i = \exp\{-\mathrm{d}[\mathbf{X}(t),\mathbf{X}(t_i)]/\mathrm{d}[\mathbf{X}(t),\mathbf{X}(t_1)]\}$, and $\mathrm{d}(.,.)$ is the Euclidean distance. The cross mapping from Y to X is defined analogously.

The skill of the cross-map estimates is quantified by the correlation coefficient (ρ) between the original $\{y(t)\}_{t=1+(E-1)\eta}^{L}$ and the approximated time series $\{\hat{y}(t) \mid M_X\}_{t=1+(E-1)\eta}^{L}$ (or between $\{x(t)\}_{t=1+(E-1)\eta}^{L}$ and $\{\hat{x}(t) \mid M_Y\}_{t=1+(E-1)\eta}^{L}$). Considering geometry of systems' attractors and the embedding theorem, it is argued that if a causal link from X to Y exists, then Y contains information about X and the states of X can be faithfully reconstructed from the mutual nearest neighbours on M_Y and $\rho(x,\hat{x}) \gg 0$. If the systems are coupled undirectionally, i.e., only the link $X \to Y$ exists, then $\rho(x,\hat{x}) \gg \rho(y,\hat{y})$.

D. Predictability improvement

Krakovská and Hanzely[27] proposed a method which also uses the Takens' embedding theorem,[24] however, is a direct generalization of the GC principle for dynamical systems. Again, two dynamical systems X and Y, represented by two time series $\{x(t)\}$ and $\{y(t)\}$, are considered. The manifold M_Y of states of the system Y consists of embedding vectors $\mathbf{Y}(t) = \{y(t),y(t-\eta_2),y(t-2\eta_2),\ldots,y[t-(d_2-1)\eta_2]\}$, where d_2 and η_2 are the embedding dimension and the time lag, respectively, for the system Y. M_Y provides the space for the predictions of Y without using additional information from X. The one-point predictions $\hat{\mathbf{Y}}$ of a large-enough statistical sample of points over the reconstructed trajectory are computed. The resulting errors $e_Y(t)$ are given by the difference between the actual and predicted values of the time series as $||y(t)-\hat{y}(t)||$.

Regarding the used method of prediction, the method of analogues[28] is applied that finds historical data similar to the current system state and assumes that the system will continue just as it did in the past. There are several ways to predict the follower of point $\mathbf{Y}(t)$, the simplest one being finding the time index i of its nearest neighbour from past states on the reconstructed trajectory and declaring $\hat{\mathbf{Y}}(t+1) = \mathbf{Y}(i+1)$. A modification, which was used here, improves the simplest version by averaging the followers of several neighbours while considering exponential weighting based on the distances of the neighbours from $\mathbf{Y}(t)$.

The prediction errors are evaluated for various combinations of possible embedding parameters. Consequently, the lowest errors led us to the proper choices of d_2 and η_2. Analogously, we get the parameters d_1 and η_1 for the prediction of X.

Following the GC principle, the predictions of Y using information from both X and Y are obtained using

the so-called mixed-state space[29,30] manifold M_{X+Y}. The state-space points in M_{X+Y} contain some of the coordinates from M_Y and some from M_X. If we used the full number of coordinates, the state corresponding to time t would be $\{y(t), y(t-\eta_2), y(t-2\eta_2), \ldots, y[t-(d_2-1)\eta_2]\}, \{w.x(t), w.x(t-\eta_1), w.x(t-2\eta_1), \ldots, w.x[t-(d_1-1)\eta_1]\}$, where the weight w represents the impact of system X. The predictions of system Y in M_{X+Y}, denoted as $\hat{\mathbf{Y}}_{X+Y}(t)$ in time t, are computed again using the method of analogues.[28] The corresponding error e_{X+Y} is given by $||y(t) - \hat{y}_{X+Y}(t)||$. The latter is, however, chosen as an optimum when using different values of w.

In order to decide whether the addition of information from X improves the prediction of Y, the Welch test is used to test the null hypothesis H_0 that the errors come from independent random samples from normal distributions with equal means and equal but unknown variances against the alternative hypothesis that the mean of errors e_{X+Y} is less than that of e_Y. If H_0 is rejected on a 5 % significance level, then we accept that $e_{X+Y} < e_Y$ or, equivalently, that the inclusion of the knowledge of X significantly improves the prediction of Y, i.e., X causes Y ($X \to Y$) in the Granger sense.

Causality in the opposite direction, i.e., $Y \to X$, is investigated analogously—after exchanging the roles of X and Y in the above instructions.

The introduced methods will be applied to the detection of causality in idealized numerical experiments with sufficiently long, noise-free time series. Therefore, and also for the sake of simplicity, we will not present tests for statistical significance for these methods. In the idealized examples below, the detected direction of causality will be clearly distinguishable by visual inspection and comparison of values for the opposite directions. Details for significance testing for the PI method are given by Krakovská and Hanzely;[27] Paluš[31] and Paluš and Vejmelka[22] describe statistical testing for CMI using a surrogate data approach. Sugihara *et al.*[15] introduced the CCM method without statistical testing, however, in further applications, e.g., by Tsonis *et al.*,[32] also the surrogate data approach is used for statistical testing of CCM results.

III. THE CAUSE PRECEDES THE EFFECT

Consider probably the simplest demonstration of the Granger causality principle—a bivariate, order one autoregressive model

$$\begin{aligned} x(t) &= a_1 x(t-1) + \xi_1(t), \\ y(t) &= b_1 y(t-1) + c_1 x(t-1) + \xi_2(t), \end{aligned} \tag{13}$$

where $\xi_{1,2}$ are independent, independently distributed normal random deviates with zero mean and variance σ^2 given by $\sigma = 0.40662$, $a_1 = 0.90693$, $b_1 = 0.40693$, and $c_1 = 0.5$.

In this case, the time index t attains natural numbers $1, 2, \ldots$. The evolution of the process X depends only on its own past, i.e., it develops independently of Y, while the expression for $y(t)$ contains $x(t-1)$. Not surprisingly, the standard GC test gives significant causality only in the direction $X \to Y$. The mutual information $I[x(t); y(t+\tau)]$ [solid red line in Fig. 1(a)] and $I[y(t); x(t+\tau)]$ [dashed black line in Fig. 1(a)] are not able to indicate the direction of causality, since they both are nonzero for a range of forward time lags τ. The correct causality $X \to Y$ is indicated by CMI $I[x(t); y(t+\tau) \mid y(t)]$ [solid red line in Fig. 1(c)], or by CMI $I[x(t); y(t+\tau) \mid y(t+\tau-1)]$ [solid red line in Fig. 1(e)] which are positive for a range of τ's, while the CMI for the direction $Y \to X$ [dashed black lines in Figs. 1(c) and 1(e)] is kept near the zero value.

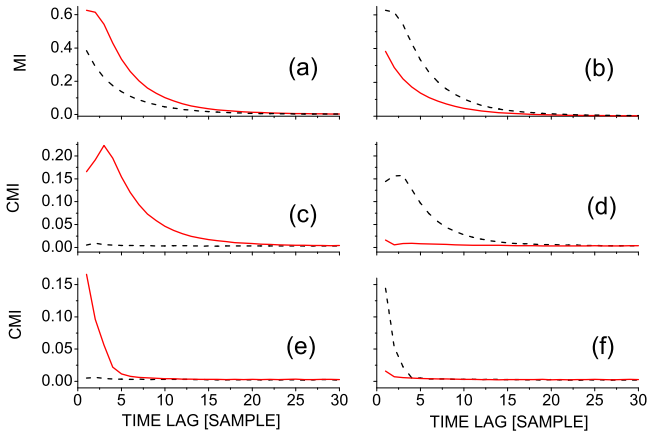

FIG. 1. Mutual information (a) and conditional mutual information (10) (c) and (11) (e) between the present of process X and the τ-future of process Y (solid red line) and between the present of process Y and the τ-future of process X (dashed black line) for the AR process (13). Graphs (b), (d), and (e) are the same as (a), (b), and (c), respectively, but for time-reversed time series $\{x(t)\}$ and $\{y(t)\}$.

In order to better understand the Granger's proposition 1 "The cause occurs before the effect," we repeat the above analysis using the time series $\{x(t)\}$ and $\{y(t)\}$ reversed in time. Having the original time series $\{x(i)\}$, $i = 1, \ldots, N$, its time-reverse $\{\tilde{x}(j)\}$ is defined as $\tilde{x}(j) = x(i)$ for $j = N - i + 1$. While the MI $I[x(t); y(t+\tau)]$ and $I[y(t); x(t+\tau)]$ of the time-reversed time series are just swapped [Fig. 1(b)], the CMI values were swapped and changed [Figs. 1(d) and 1(f)]. The swap of the causality directions is obvious—the existence of causality in the direction $Y \to X$ clearly emerged. It is consistent with the Granger's proposition 1: While going forward in time, present values $x(t)$ influence the future values $y(t+\tau)$. In reversed time series, the present value $y(t)$ contains information about the future values $x(t+\tau)$. Besides the clear effect of changing the order of the cause and the effect, there are also minor changes in the CMI values. The CMI for $Y \to X$ direction in the reversed time series [Figs. 1(d) and 1(f)] has slightly smaller values than the CMI for $X \to Y$ direction in the original time series [Figs. 1(c) and 1(e)]; and, in the reversed time series also a causality in the $X \to Y$ direction occurs for $\tau = 1$. Although the later causality is much "weaker," comparing, e.g., the related CMI values, it has been found significant in the standard GC test. This observation is apparently the effect of time-averaged noise terms. The reversed process X contains some information about its future due to averaged terms ξ_1 and therefore also about the future of Y. However, the "main" direction of causality, reflected in large CMI values spread over a number of lags τ, complies with the Granger's proposition 1: After reversing the order of

the cause and the effect, the direction of causality changed from $X \to Y$ to $Y \to X$.

IV. CAUSALITY DETECTION METHODS AND THE ARROW OF TIME

The predictability improvement (PI) method of Krakovská and Hanzely[27] is a direct generalization of the GC principle for nonlinear dynamical systems in the sense that it infers the existence of causality in the direction $X \to Y$ by testing the improvement of prediction of Y by using also the knowledge of the present and past states of X. The conditional mutual information (CMI) (10) or (11) is used to infer the causal relation in the direction $X \to Y$ by measuring the conditional dependence between the present and past states of X and a future state of Y, i.e., the CMI evaluates the ability to predict the future of Y using the present and past states of X. Using the convergent cross-map (CCM) method, for the inference of the same direction of causality, $X \to Y$, Y should cross-map X. At first sight, it seems contradictory; however, note that CCM asks whether the present state of Y contains information about the present state of X. Considering coupled dynamical systems, we can say that while the methods, which generalize the standard Granger causality principle (PI and CMI here), evaluate the ability of the driver ("master") system to *forecast* the driven ("slave") system, the CCM method evaluates the ability of the slave to *nowcast* the master. Thus the CCM lacks any arrow of time in its formulation. (The time lags η are used for embedding of scalar time series into d-dimensional state spaces.) What it means for the Granger's proposition 1? Since the CCM is tailored for dynamical systems, let us have a look at the unidirectionally coupled Rössler systems studied in detail by Paluš and Vejmelka.[22] The driving, master system X is defined as

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

and the driven system Y as

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

where $\omega_1 = 1.015$ and $\omega_2 = 0.985$. The first component of Y contains the diffusive coupling term $\epsilon[x_1(t-\delta) - y_1(t)]$, where ϵ is the coupling strength and δ is a delay time in which the driving information from X reaches the driven system Y. We will start with $\delta = 0$, which is the standard setting in many synchronization studies and can be considered as physically relevant if the sampling time [the time between the measurements $x(t)$ and $x(t+1)$] is greater than δ. We integrate (see Appendix for details) the above systems for a range of coupling strengths ϵ, record the components $x_1(t)$ and $y_1(t)$, and use them as an input into the CCM analysis. The results are presented in Fig. 2(a). We can see that there is a range of ϵ for which the CCM skill for the direction $X \to Y$ is clearly larger than for the opposite direction. We will explain the ϵ-dependence of causality measure shortly below. Now we just conclude that for a range of ϵ, the CCM is able to distinguish the correct causality direction $X \to Y$. The results of the same CCM analysis, but using the time-reversed time series, are presented in Fig. 2(b). We can see that the result is practically the same, indicating the causality direction $X \to Y$. This should not be surprising considering the above discussion about the lack of any time arrow in the CCM approach.

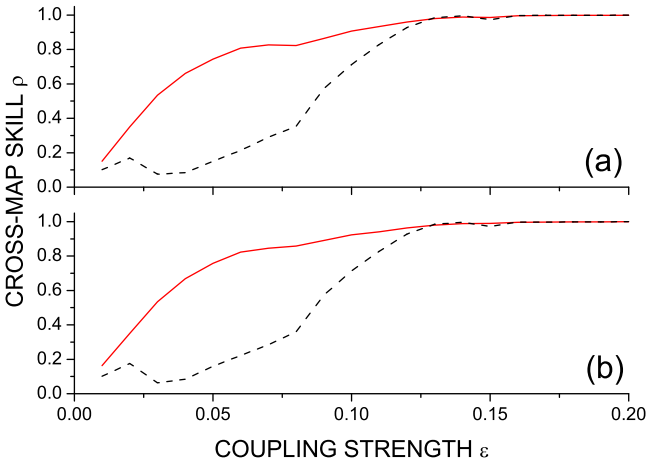

FIG. 2. (a) Cross-mapping skills for the $X \to Y$ direction ("Y cross-maps X"—solid red line) and for the opposite direction (dashed black line) as a function of coupling strength ϵ for the unidirectionally coupled Rössler systems, $\delta = 0$. (b) The same as (a), but for time-reversed time series.

Paluš and Vejmelka[22] also studied the unidirectionally coupled Rössler systems with $\delta = 0$. We refresh their results in Figs. 3(a) and 3(b). Three-dimensional chaotic systems, such as each of the above Rössler systems, are characterized by three Lyapunov exponents[33]—one positive, one zero, and one negative LE. The positive and zero LEs for both Rössler systems are plotted as functions of the coupling strength ϵ in Fig. 3(a). For the driving system X, they remain constant for all values of ϵ, while for the driven system Y they decrease (although nonmonotonously) with increasing ϵ. When the positive LE of the driven system [the solid red line in Fig. 3(a)] becomes negative, the systems synchronize. Since in the synchronized state trajectories of both systems are topologically equivalent, it is impossible to detect the direction of coupling, see the discussion and references in Coufal *et al.*[34] Thus for the systems able to synchronize, there is a limited range of the coupling strength ϵ for which the direction of coupling can be reliably determined from time series. We could observe this phenomenon in the case of CCM in Fig. 2 as well as for the CMI analysis in Fig. 3(b).

Now let us repeat the CMI analysis for the time-reversed time series. We can see in Fig. 3(c) that after the time reversal, the CMI, just like the CCM, detects the same causality direction $X \to Y$.

The CMI (10) measures the ability to improve the prediction of $y(t+\tau)$ using $x(t)$ [or vice versa using CMI (9)]. The results in Figs. 3(b) and 3(c) present the mean CMI for lags $\tau = 1, \ldots, 50$ samples. The τ-dependence of the CMI for $\epsilon = 0.07$ (and $\delta = 0$) is presented in Fig. 4(a) for $\tau = 1, \ldots, 530$. Any dependence measure between $x(t)$ and $y(t+\tau)$ [or $x(t+\tau)$] for chaotic systems vanishes for large

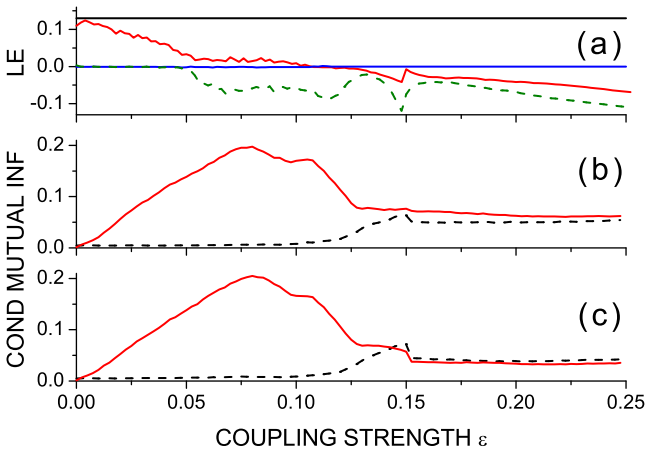

FIG. 3. (a) Two largest Lyapunov exponents of the driver X (black and blue constant lines) and the response system Y (decreasing curves, originally positive LE is marked by a solid red line, originally zero LE by a dashed green line). (b) Averaged conditional mutual information (10) for the $X \to Y$ direction (solid red line) and for the opposite direction (dashed black line). (c) The same as (b) but for time-reversed time series. All results for the unidirectionally coupled Rössler systems with $\delta = 0$.

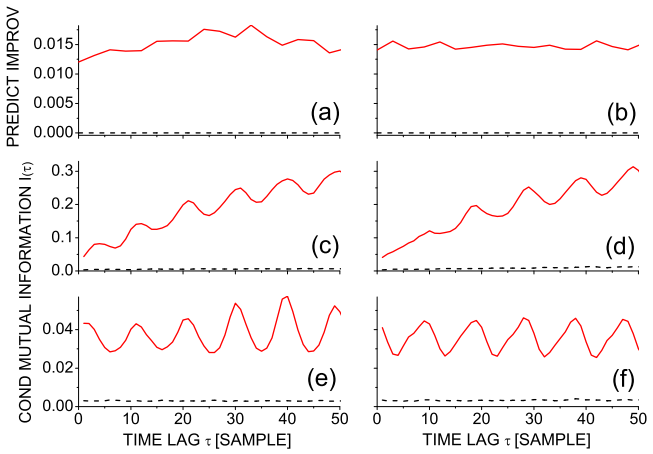

FIG. 5. (a) Predictability improvement, (c) conditional mutual information (10), and (e) CMI (11) as functions of forward time lag τ for the $X \to Y$ direction (solid red line) and for the opposite direction (dashed black line). [(b), (d), and (f)] The same as (a), (c), and (d), respectively, but for time-reversed time series. All results for the unidirectionally coupled Rössler systems with $\delta = 30$ and $\epsilon = 0.07$.

τ; however, the oscillatory character of the Rössler system makes this decay very slow. The CMI τ-dependence reflects basic quasi-oscillatory period around 20 samples (see Appendix for the sampling time) and a slower quasi-periodicity. Figure 4(c) zooms the pattern from Fig. 4(a) for $\tau = 1, \ldots, 50$ and for the same range of values of the prediction horizon τ the results for the predictability improvement method are given in Fig. 4(e). Although the τ-dependence for CMI and PI differs, both CMI and PI clearly detect the causality in the $X \to Y$ directions. For Figs. 4(a), 4(c), and 4(e), their counterparts for the time-reversed time series are Figs. 4(b), 4(d), and 4(f), respectively. Again, the time reversal did not change the detected direction of causality in either method.

One can think that the above result could be caused by the instantaneous driving with $\delta = 0$. Let us integrate

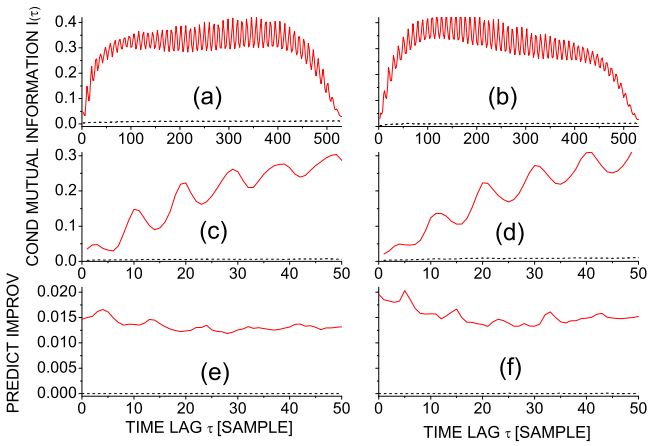

FIG. 4. [(a) and (c)] Conditional mutual information (10) and (e) predictability improvement as functions of forward time lag τ for the $X \to Y$ direction (solid red line) and for the opposite direction (dashed black line). [(b), (d), and (f)] The same as (a), (c), and (d), respectively, but for time-reversed time series. All results for the unidirectionally coupled Rössler systems with $\delta = 0$ and $\epsilon = 0.07$.

the unidirectionally coupled Rössler systems with a distinctively nonzero coupling delay $\delta = 30$ samples. The results for $\epsilon = 0.07$ for the predictability improvement and the conditional mutual information (10) and (11) are presented in Fig. 5. Wibral *et al.*[26] introduced the CMI (11) as a measure for inferring the coupling delay. The discussion of this problem can be found in Coufal *et al.*[34] and will not be repeated here. We just conclude that both CMI formulations give qualitatively the same result: The causality direction $X \to Y$ does not change after the time reversal. And the same result is obtained using the PI method.

Figure 6 summarizes the results for the Rössler systems with the time-delayed unidirectional coupling $X \to Y$ for the studied range of coupling strength ϵ. The CMI is again presented as the mean for lags $\tau = 1, \ldots, 50$, PI used $\tau = 1$ and CCM cross-maps the states of both systems at the same time. For all the methods, even for the Rössler systems with the time-delayed coupling, the direction of causality after the time reversal remains the same as in the original time series recoded in forward time evolution. This behaviour has been confirmed also in other unidirectionally coupled chaotic dynamical systems, e.g., identical and non-identical Henon systems or the Rössler system driving the Lorenz system, examples defined and studied by Paluš and Vejmelka.[22]

The violation of the Granger causality principle 1 that the cause precedes the effect, in the case of the coupled Rössler systems, has also been observed using the information flow and causality method of Liang[35,36] which is entirely independent of the methods considered above. A number of methods for inference of causality from experimental data, based on the estimation of a predefined model, have been proposed, for instance, dynamic causal modelling,[37] fitting of a phase-dynamic model,[38] dynamical Bayesian inference,[39] or maximum likelihood methods.[40] It is an interesting question how would such methods evaluate causality in time-reversed time series; however, we leave this question for future research.

An interested reader would naturally ask how would the standard GC test behave if applied to time-reversed time series

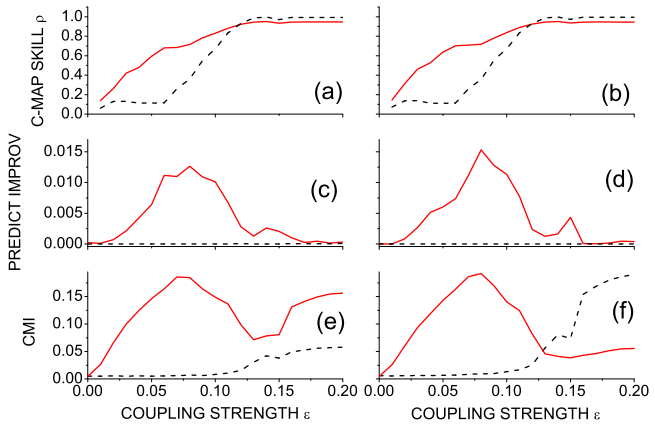

FIG. 6. (a) CCM skill, (c) predictability improvement, and (e) conditional mutual information (10) as functions of coupling strength ϵ for the $X \to Y$ direction (solid red line) and for the opposite direction (dashed black line). [(b), (d), and (f)] The same as (a), (c), and (d), respectively, but for time-reversed time series. All results for the unidirectionally coupled Rössler systems with $\delta = 30$.

from chaotic systems, or, vice versa, what would be the results of the CCM and PI methods applied to the AR model and time-reversed AR model time-series? Unfortunately, we cannot report such results, since the standard GC test typically fails when applied to nonlinear time series such as the used output of the coupled Rössler systems. On the other hand, the CCM and PI methods employ the geometry of attractors of dynamical systems and cannot bring consistently correct results when applied to stochastic systems such as the used AR model. Recently, Krakovská *et al.*[41] have analysed six methods for causality detection, including the three used here, and evaluated their performance when applied on different types of data, including an AR model and coupled chaotic systems. The failures of methods designed for a specific type of a system and applied to a different one are described in terms of false positive or false negative results.

V. TEMPORAL ASYMMETRY AND IRREVERSIBILITY

In 1996, Paluš[42] proposed to evaluate the Kullback-Leibler divergence $K(p^{\rightarrow}, p^{\leftarrow})$ between the probability distributions $p^{\rightarrow} = p\{x(t), x(t+\tau), \ldots, x[t+(m-1)\tau], x(t+m\tau)\}$ and $p^{\leftarrow} = p\{x(t+m\tau), x[t+(m-1)\tau], \ldots, x(t+\tau), x(t)\}$ in order to quantify the temporal asymmetry of a time series $\{x(t)\}$. This measure indicated that the temporal asymmetry might be one of nonlinear properties of normal human electroencephalogram.[42]

Recently, an exact relationship has been derived between dissipation and the distinguishability of a process from its time reverse, quantified by the KLD between probability densities of forward and backward system states.[43,44] This relation says that the dissipation results from the asymmetry between the forward and backward evolutions of a system: it is zero only when $p^{\rightarrow} = p^{\leftarrow}$. The above expression is also consistent with a proposal, linking the time-asymmetry of the Kolmogorov-Sinai entropy to the entropy production of the dynamical system.[45] The result of Gaspard[45] provides an interpretation of the entropy production as a manifestation of the time-reversal symmetry breaking. In the following development, Roldán and Parrondo[46] showed that the above defined KLD $K(p^{\rightarrow}, p^{\leftarrow})$ applied to stationary time series provides information about the entropy production of the physical mechanism generating the series, even if one ignores any detail of that mechanism.

Let us estimate $K(p^{\rightarrow}, p^{\leftarrow})$ for the above unidirectionally coupled Rössler systems using a simple equiquantal binning algorithm.[3] The estimates for n-tuples $x(t), x(t+\tau), \ldots, x[t+(n-1)\tau]$ for $n = 3, 4, 5$ and different values of τ give qualitatively consistent results (the same shape of curves although different values). Because of the inherent estimator bias, the "instrumental zero," or the range for $K(p^{\rightarrow}, p^{\leftarrow})$ values for time symmetric processes, was estimated using the Fourirer transform surrogate data, used also for testing the statistical significance of CMI estimates by Paluš and Vejmelka.[22] $K(p^{\rightarrow}, p^{\leftarrow})$ for both systems as the function of the coupling strength ϵ is plotted in Fig. 7. We see that all values for the Rössler systems are distinctively higher than the range of the surrogate data, that is, the observed dynamics breaks the time-reversal symmetry. Due to dissipation and a positive entropy production rate, the studied chaotic dynamics is not reversible in time so that the backward run of the process cannot occur naturally. $K(p^{\rightarrow}, p^{\leftarrow})$ for the driving system X is constant for all values of ϵ (just the straight line is perturbed by the variance of the estimator). $K(p^{\rightarrow}, p^{\leftarrow})$ for the driven system Y reflects the decreasing behaviour of its positive LE (Fig. 3) till the synchronization threshold and then it starts to rise and reach the $K(p^{\rightarrow}, p^{\leftarrow})$ value of the system X when the system Y becomes fully slaved to X. This is a nice example how information-theoretic measures relate various properties of chaotic dynamical systems—the level of time irreversibility is related to the entropy production which is related to the exponential divergence of trajectories measured by the positive LE according to the theorem of Pesin.[47] The levels of irreversibility of the synchronized systems adjust to a common value, in compliance with our original concept of synchronization as a process of adjustment of information or entropy production rates.[9,48]

VI. SUMMARY OF RESULTS AND THEIR APPLICATION

In the above numerical study, we have demonstrated that the chaotic dynamical systems violated the Granger causality principle 1 that the cause precedes the effect. On the other hand, a short excursion to measuring time-irreversibility explained that chaotic processes are not reversible in time. In other words, such a process, obtained by the time reversal of a chaotic process, cannot occur in nature. Then, what is the scientific value of the presented results? Besides better understanding of different causality detection methods, the time reversal can help researchers in understanding mechanisms underlying causal interactions observed in experimental data. Let us have a look at two examples of detected causality already described in scientific literature.

Runge *et al.*[49] presented a method for detecting causality in multivariate time series and identified a number of causal links in the Earth climate system based on the analysis of global near-surface air pressure field. The continuous spatio-temporal pressure field is originally represented by a very

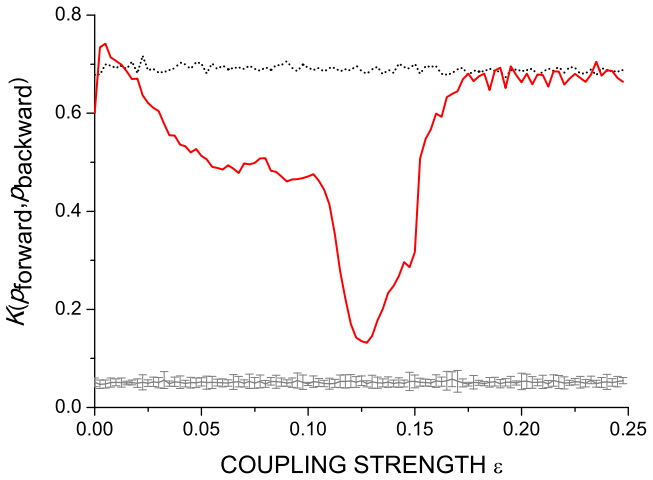

FIG. 7. The Kullback-Leibler divergence $K(p^{\rightarrow}, p^{\leftarrow})$ between the forward and backward joint probability distributions for the driving system X (dotted black line) and the driven system system Y (solid red line) as a function of coupling strength ϵ for the unidirectionally coupled Rössler systems. The grey line and whiskers present mean $\mp$ 2 standard deviations of KLD for a set of Fourier transform surrogate data.

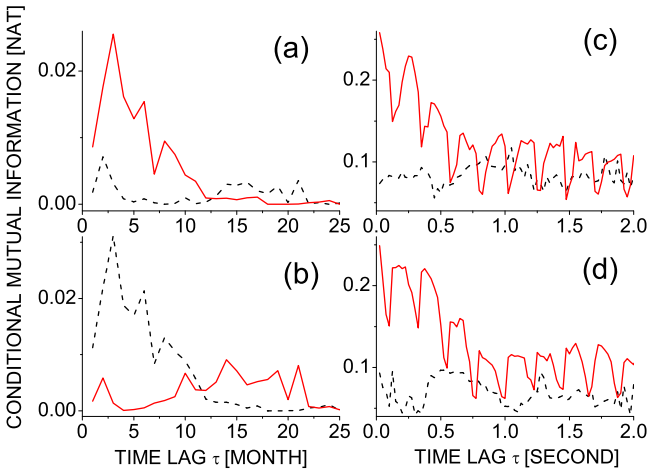

FIG. 8. (a) Conditional mutual information (10) as a function of time lag τ, characterizing the causal influence of [(a) and (b)] the climatic El Niño mode on the Arabian Sea mode (solid red line) and in the opposite direction (dashed black line); and [(c) and (d)] the respiratory rhythm on the heart rhythm (solid red line) and in the opposite direction (dashed black line). The results for the original (forward time) data are in (a) and (c), the results for the time reversed time series in (b) and (d).

high-dimensional data with approximately 40 thousand variables. Using a dimensionality reduction algorithm according to Vejmelka *et al.*,[50] about 60 *climate variability modes* were extracted. Each mode, represented by a single time series, is localized in a particular Earth region and is typically related to a climate phenomenon occurring in that region. The mode related to the El Niño Southern Oscillation (ENSO) main area and phenomenon has been found influencing many other regions worldwide. One such causal link leads to the mode located in the Arabian Sea (see Figs. 5 and 6 in Vejmelka *et al.*[50] and Fig. 3 in Runge *et al.*[49]). Let us analyse the monthly time series[50] related to these two climate variability modes. The CMI analysis confirms that the main direction of causality leads from the ENSO mode to the Arabian Sea mode [solid red line in Fig. 8(a)]. After the time reversal [Fig. 8(b)] also the causality direction reversed. This behaviour has been observed in the linear AR model (13), i.e., the observed causality corresponds to a linear transfer of a time-delayed signal. Hlinka *et al.*[51] add to our result their observation, that the direction of causality in the global air temperature field basically agrees with the prevailing direction of winds in the same areas. Thus the causality, or information transfer[51] in the global climate system, is a consequence of the transport of air masses and related energy. (There are, however, highly nonlinear phenomena in the Earth climate, which might behave differently, but will not be discussed here; e.g., the ENSO dynamics itself, or interactions of the annual cycle[52] with slower climate oscillations[53] in mid-latitudes.)

Musizza *et al.*[54] analysed interactions between brain, heart, and respiration in rats undergoing anaesthesia. In the vigilant state, the most pronounced causal link has been found in cardio-respiratory interactions in which the phase of the respiratory rhythm influences the phase of the heart rhythm. An example of this causality together with establishing its statistical significance has also been presented by Paluš and Vejmelka.[22] The CMI applied to the respiratory and heart rhythms phases shows a causal link in the direction respiration $\rightarrow$ heart [solid red line in Fig. 8(c)] persisting up to the lag about half a second. After the time reversal [Fig. 8(d)], the direction of causality did not change. As this behaviour has been observed in nonlinear, chaotic dynamical systems, the present result can be considered as an additional support for understanding the cardio-respiratory interactions as a system of coupled nonlinear oscillators, coined by Stefanovska *et al.*[55,56] However, we stress "support," not evidence. Even linear AR processes of higher order can produce different behaviour than the example (13). For instance, consider an AR2 process in which X drives Y by the term $x(t-1)$. For modelling the time reversed process, we express $x(t-2)$ and $y(t-2)$ as functions of $x(t), x(t-1), y(t), y(t-1)$. Then $X \rightarrow Y$ also in the time reversed version. For the evidence of a nonlinear dynamical origin of a process and its interactions, suitable nonlinearity tests[57,58] should be also included in the analysis.

We can also observe the "linear transfer" behaviour of coupled dynamical systems, however, in a very specific condition. Such a condition happened in the above Rössler systems. A close inspection of CMI in Figs. 3(b) and 3(c), and especially in Figs. 6(e) and 6(f) where the effect is amplified by the time delayed coupling, shows that the direction of causality, established by CMI, reversed after the time reversal of the time series for coupling strength beyond the synchronization threshold. The explanation of this observation lies in a specific type of synchronization reached by the analysed Rössler oscillators. It is called the lag synchronization[59] in which the states of two oscillators are nearly identical, but one system lags in time to the other. Thus, in the lag synchronization state for $\epsilon > 0.125$, in spite of the nonlinear dynamical origin of the signals, their relation corresponds to a transfer of a time-delayed signal.

VII. CONCLUSION

Inference of causality from time series is a challenging problem when analysing behaviour of various complex systems. In many cases, the working hypothesis is that the studied data have been generated by coupled dynamical systems. Therefore, we focussed on three known causality detection methods applied to dynamical systems: conditional mutual information[22] (CMI, also known as transfer entropy[21]), convergent cross-mapping[15] (CCM), and the predictability improvement[27] (PI) method. Two methods (CMI, PI) have been proposed as a nonlinear generalization of Granger causality[19] (GC) since the unidirectional coupling in which one system drives another has been considered as a special case of GC.[9] Comparing the methods on a theoretical level, we concluded that the CCM is a method for detecting coupling between dynamical systems, but it is not a method for detecting causality in the Granger sense, since it ignores the time sequence of the cause preceding the effect. Indeed, the application of the CCM on the time series reversed in time brought the same results as the original time series recorded forward in time. The CMI method, when applied to a canonical example of GC in an autoregressive model of order 1, confirmed its roots in the GC principle: After time-reversal of the AR time series, the direction of causality also reversed. However, when both CMI and PI were applied to time series from chaotic dynamical systems, they behaved as the CCM: the same direction of causality as for the original data was also detected for the reversed time series. This violation of the Granger causality principle that the cause precedes the effect is probably due to dynamical memory of dynamical systems. Analysis of time irreversibility of the studied processes, however, showed that chaotic systems are not reversible in time. Therefore, the observed violation of the causality principle can occur only in a numerical study but not in real-world systems. The time reversal in causality analysis can help to distinguish between a linear transfer of a time-delayed signal and nonlinear interactions of dynamical systems. Any detection of causality, however, should be accompanied by a battery of time series analysis methods, namely, tests for nonlinearity and synchronization should be performed, as well as standard spectral analysis enhanced by time-frequency analysis since causal links can occur in or between different time scales of multiscale processes.[60]

ACKNOWLEDGMENTS

The authors would like to thank two anonymous referees for their constructive comments which helped to improve the manuscript. Very special thanks are due to the referee who included in the report the analysis of the Rössler systems using the causality method of Liang.[35,36] The authors would also like to thank D. Coufal for integrating the dynamical systems with the delayed coupling, and B. Musizza and the BRACCIA (EC FP6 Project No 517133 NEST) team for providing the animal cardiorespiratory data. The climate data were provided as a supplement to Ref. 50.

This study was supported by the bilateral cooperation of the Czech and Slovak Academies of Sciences Project SAV-15-18. A.K., J.J., and M.C. were also supported by the Slovak Grant Agency for Science (Grant No. 2/0011/16) and by the Slovak Research and Development Agency (Grant No. APVV-15-0295), M.P. by the Czech Health Research Council (Project No. NV15-33250A).

APPENDIX

The parameters of the AR model (13) were chosen such that the model provides long runs of nontrivial stationary time series. Equivalent results can be also obtained for different sets of parameters. The first 50 000 iterations were discarded as possible transients and following 131 072 samples were recorded and used for the CMI computations.

The unidirectionally coupled Rössler systems were integrated using the ode8 (Dormand-Prince) solver in the MATLAB® Simulink® environment. The fixed step size 0.0785 was used in the integration, and the data were downsampled to each fourth giving the sampling step 0.314. For the case $\delta = 0$, also, the alternative numerical integration based on the adaptive Bulirsch-Stoer method[61] was used. The latter method uses an adaptive integration step, however, the final sampling time 0.314 was prescribed, which leads to approximately 20 samples per period. Data from both integration methods give equivalent results. After the initialization, 5000 samples were discarded and the following 131 072 samples were recorded and used for the CMI computations. The subset of the first 50 000 samples was used for PI and CCM computations. 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., $y_1(0) = 11.120979 * [ran(iseed) + 0.5]$}. 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.[22]

[1]N. Wiener, "The theory of prediction," in *Modern Mathematics for the Engineer*, edited by E. F. Beckenbach (McGraw-Hill, New York, NY, 1956), pp. 125–139.
[2]C. W. J. Granger, "Time series analysis, cointegration, and applications. Nobel Lecture, December 8, 2003," in *Les Prix Nobel. The Nobel Prizes 2003*, edited by T. Frängsmyr (Nobel Foundation, Stockholm, 2004), pp. 360–366.
[3]K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," Phys. Rep. **441**, 1–46 (2007).
[4]R. Shaw, "Strange attractors, chaotic behavior, and information flow," Z. Naturforsch. A **36**, 80–112 (1981).
[5]A. M. Fraser, "Information and entropy in strange attractors," IEEE Trans. Inf. Theory **35**, 245–262 (1989).
[6]M. Paluš, "Kolmogorov entropy from time series using information-theoretic functionals," Neural Netw. World **7**, 269–292 (1997). http://www.cs.cas.cz/mp/papers/rd1a.pdf
[7]A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge university press, 2003), Vol. **12**.
[8]S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, "The synchronization of chaotic systems," Phys. Rep. **366**, 1–101 (2002).
[9]M. Paluš, V. Komarek, Z. Hrncir, and K. Sterbova, "Synchronization as adjustment of information rates: Detection from bivariate time series," Phys. Rev. E **63**, 046211 (2001).

[10]J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependences: Application to intracranially recorded EEG," Phys. D: Nonlin. Phenom. **134**, 419–430 (1999).
[11]R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: A case study on electroencephalographic signals," Phys. Rev. E **65**, 041903 (2002).
[12]R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, "Bivariate surrogate techniques: Necessity, strengths, and caveats," Phys. Rev. E **68**, 066202 (2003).
[13]D. A. Smirnov and R. G. Andrzejak, "Detection of weak directional coupling: Phase-dynamics approach versus state-space approach," Phys. Rev. E **71**, 036207 (2005).
[14]D. Chicharro and R. G. Andrzejak, "Reliable detection of directional couplings using rank statistics," Phys. Rev. E **80**, 026217 (2009).
[15]G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," Science **338**, 496–500 (2012).
[16]J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, "Escaping the curse of dimensionality in estimating multivariate transfer entropy," Phys. Rev. Lett. **108**, 258701 (2012).
[17]I. Vlachos and D. Kugiumtzis, "Nonuniform state-space reconstruction and coupling detection," Phys. Rev. E **82**, 016207 (2010).
[18]J. Sun, D. Taylor, and E. M. Bollt, "Causal network inference by optimal causation entropy," SIAM J. Appl. Dyn. Syst. **14**, 73–106 (2015).
[19]C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," Econometrica **37**, 424–438 (1969).
[20]T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, NY, USA, 1991).
[21]T. Schreiber, "Measuring information transfer," Phys. Rev. Lett. **85**, 461–464 (2000).
[22]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).
[23]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).
[24]F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Springer, 1981), pp. 366–381.
[25]A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134–1140 (1986).
[26]M. Wibral, N. Pampu, V. Priesemann, F. Siebenhühner, H. Seiwert, M. Lindner, and J. T. Lizier, "Measuring information-transfer delays," PLoS ONE. **2**, e55809 (2013).
[27]A. Krakovská and F. Hanzely, "Testing for causality in reconstructed state spaces by an optimized mixed prediction method," Phys. Rev. E **94**, 052203 (2016).
[28]E. N. Lorenz, "Atmospheric predictability as revealed by naturally occurring analogues," J. Atmos. Aci. **26**, 636–646 (1969).
[29]M. Wiesenfeldt, U. Parlitz, and W. Lauterborn, "Mixed state analysis of multivariate time series," Int. J. Bif. Chaos **11**, 2217–2226 (2001).
[30]U. Feldmann and J. Bhattacharya, "Predictability improvement as an asymmetrical measure of interdependence in bivariate time series," Int. J. Bif. Chaos **14**, 505–514 (2004).
[31]M. Paluš, "From nonlinearity to causality: Statistical testing and inference of physical mechanisms underlying complex dynamics," Contemp. Phys. **48**, 307–348 (2007).
[32]A. A. Tsonis, E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang, "Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature," Proc. Natl. Acad. Sci. **112**, 3253–3256 (2015).
[33]H. G. Schuster and J. Wolfram, *Deterministic Chaos: An Introduction* (Wiley-VCH, 2005).
[34]D. Coufal, J. Jakubík, N. Jajcay, J. Hlinka, A. Krakovská, and M. Paluš, "Detection of coupling delay: A problem not yet solved," Chaos: Interdiscip. J. Nonlin. Sci. **27**, 083109 (2017).
[35]X. S. Liang, "Unraveling the cause-effect relation between time series," Phys. Rev. E **90**, 052150 (2014).
[36]X. S. Liang, "Information flow and causality as rigorous notions ab initio," Phys. Rev. E **94**, 052201 (2016).
[37]K. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," NeuroImage **19**, 1273–1302 (2003).
[38]M. G. Rosenblum and A. S. Pikovsky, "Detecting direction of coupling in interacting oscillators," Phys. Rev. E **64**, 045202 (2001).
[39]T. Stankovski, A. Duggento, P. V. E. McClintock, and A. Stefanovska, "Inference of time-evolving coupled dynamical systems in the presence of noise," Phys. Rev. Lett. **109**, 024101 (2012).
[40]I. T. Tokuda, S. Jain, I. Z. Kiss, and J. L. Hudson, "Inferring phase equations from multivariate time series," Phys. Rev. Lett. **99**, 064101 (2007).
[41]A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, "Comparison of six methods for the detection of causality in a bivariate time series," Phys. Rev. E **97**, 042207 (2018).
[42]M. Paluš, "Nonlinearity in normal human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos," Biol. Cybern. **75**, 389–396 (1996).
[43]R. Kawai, J. M. R. Parrondo, and C. V. den Broeck, "Dissipation: The phase-space perspective," Phys. Rev. Lett. **98**, 080602 (2007).
[44]J. M. R. Parrondo, C. V. den Broeck, and R. Kawai, "Entropy production and the arrow of time," New. J. Phys. **11**, 073008 (2009).
[45]P. Gaspard, "Time-reversed dynamical entropy and irreversibility in Markovian random processes," J. Stat. Phys. **117**, 599–615 (2004).
[46]E. Roldán and J. M. R. Parrondo, "Estimating dissipation from single stationary trajectories," Phys. Rev. Lett. **105**, 150607 (2010).
[47]Y. B. Pesin, "Characteristic Lyapunov exponents and smooth ergodic theory," Russ. Math. Surv. **32**, 55 (1977).
[48]M. Paluš, "Linked by dynamics: Wavelet-based mutual information rate as a connectivity measure and scale-specific networks," in *Advances in Nonlinear Geosciences*, edited by A. A. Tsonis (Springer, 2018), pp. 427–463.
[49]J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, "Identifying causal gateways and mediators in complex spatio-temporal systems," Nat. Commun. **6**, 8502 (2015).
[50]M. Vejmelka, L. Pokorná, J. Hlinka, D. Hartman, N. Jajcay, and M. Paluš, "Non-random correlation structures and dimensionality reduction in multivariate climate data," Clim. Dyn. **44**, 2663–2682 (2015).
[51]J. Hlinka, N. Jajcay, D. Hartman, and M. Paluš, "Smooth information flow in temperature climate network reflects mass transport," Chaos: Interdiscip. J. Nonlin. Sci. **27**, 035811 (2017).
[52]M. Paluš, D. Novotná, and P. Tichavský, "Shifts of seasons at the European mid-latitudes: Natural fluctuations correlated with the North Atlantic Oscillation," Geophys. Res. Lett. **32**, L12805 (2005).
[53]M. Paluš and D. Novotná, "Enhanced Monte Carlo singular system analysis and detection of period 7.8 years oscillatory modes in the monthly NAO index and temperature records," Nonlin. Process. Geophys. **11**, 721–729 (2004).
[54]B. Musizza, A. Stefanovska, P. V. E. McClintock, M. Paluš, J. Petrovčič, S. Ribarič, and F. F. Bajrović, "Interactions between cardiac, respiratory and EEG-δ oscillations in rats during anaesthesia," J. Physiol. (Lond.) **580**, 315–326 (2007).
[55]A. Stefanovska and M. Bračič, "Physics of the human cardiovascular system," Contemp. Phys. **40**, 31–55 (1999).
[56]A. Stefanovska, M. Bračič-Lotrič, S. Strle, and H. Haken, "The cardiovascular system as coupled oscillators?," Physiol. Meas. **22**, 535 (2001).
[57]M. Paluš, "Testing for nonlinearity using redundancies: Quantitative and qualitative aspects," Phys. D: Nonlin. Phenom. **80**, 186–205 (1995).
[58]M. Paluš, "Detecting nonlinearity in multivariate time series," Phys. Lett. A **213**, 138–147 (1996b).
[59]M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "From phase to lag synchronization in coupled chaotic oscillators," Phys. Rev. Lett. **78**, 4193–4196 (1997).
[60]M. Paluš, "Multiscale atmospheric dynamics: Cross-frequency phase-amplitude coupling in the air temperature," Phys. Rev. Lett. **112**, 078702 (2014).
[61]W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes. The Art of Scientific Computations* (Cambridge University Press, 1989).