Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections

Milan Paluš* and Martin Vejmelka†

Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic

(Received 14 December 2006; revised manuscript received 1 March 2007; published 18 May 2007)

We discuss some problems encountered in inference of directionality of coupling, or, in the case of two interacting systems, in inference of causality from bivariate time series. We identify factors and influences that can lead to either decreased test sensitivity or false detections and propose ways to cope with them in order to perform tests with high sensitivity and a low rate of false positive results.

DOI: 10.1103/PhysRevE.75.056211 PACS number(s): 05.45.Tp, 05.45.Xt, 02.50.Ey, 87.80.Tq

I. INTRODUCTION

Cooperative behavior of coupled complex systems has recently attracted considerable interest from theoreticians as well as experimentalists (see, e.g., the monograph [1]), since synchronization and related phenomena have been observed not only in physical but also in many biological systems. Examples include cardiorespiratory interactions [2–5] and synchronization of neural signals [6–10]. In such systems it is important not only to detect synchronized states, but also to identify drive-response relationships between the systems studied. This problem is a special case of the general question of causality or causal relations between systems, processes, or phenomena. The mathematical formulation of causality in measurable terms of predictability was given by Wiener [11]. Granger [12] introduced a specific notion of causality into time series analysis by evaluation of predictability in bivariate autoregressive models. This linear framework for measuring and testing causality has been widely applied in economy and finance (see Geweke [13] for a comprehensive survey of the literature), but also in different sciences such as climatology (see [14] and references therein) or neurophysiology, where specific problems of multichannel electroencephalogram (EEG) recordings were solved by generalizing the Granger causality concept to multivariate cases [15,16]. Nevertheless, the limitation of the Granger causality concept to linear relations required further generalizations, which emerged especially in the intensively developing field of synchronization of complex systems. Considering the task of identification of drive-response relationships, a number of asymmetric dependence measures have been proposed [6,7,9,10,17–22] and applied in diverse scientific areas such as laser physics [23], climatology [24,25], cardiovascular physiology [22,24], neurophysiology [6,7,9,10,26–28], and finance [29]. In spite of these widespread applications of various coupling asymmetry measures, the task of correct inference of coupling asymmetry, i.e., the identification of the driving and driven systems from experimental time series, is far from resolved. In this paper we identify some problems encountered in this task and give some practical advice for avoiding false detections of coupling asymmetry or causality. We will consider two interacting systems, possibly one of them driving the other. Then the coupling asymmetry, or, as it is called, the directionality of coupling, also identifies causality, or causal relations between the studied systems. The problem of distinguishing the true causality from indirect influences in interactions of three or more systems is beyond the scope of this paper and will be addressed elsewhere.

In Sec. II we introduce three examples of unidirectionally coupled chaotic systems and analyze their coupling using three already published measures. In this way we demonstrate the importance of choice of an appropriate measure with known properties and a solid mathematical background. In Sec. III we review basic measures defined in information theory and specify applications of conditional mutual information (CMI) for detection of causality. Section IV introduces multidimensional conditional mutual information applicable to amplitudes of dynamical systems or stochastic processes, and a version of CMI for evaluation of coupling asymmetry using instantaneous phases of coupled oscillatory systems. Then, in Sec. V, we study bias and variance of CMI estimates and discuss statistical evaluation of the estimated CMI in order to assure correct inference of causality and/or coupling asymmetry from experimental time series. Further factors influencing the bias in the CMI estimates are discussed in Sec. VI, where also an example of assessing the direction of coupling in cardiorespiratory interaction is presented. The discussed topics are summarized and conclusions given in Sec. VII. Finally, the Appendix proves the equivalence of the conditional mutual information and the transfer entropy introduced by Schreiber [18].

II. ASYMMETRY IN COUPLING: SYSTEMS AND MEASURES

As the first example, let us consider the unidirectionally coupled Rössler and Lorenz systems, also studied in Refs. [7,9,19], described by the equations

$$\dot{x}_1 = -\alpha\{x_2 + x_3\},$$

$$\dot{x}_2 = \alpha\{x_1 + 0.2x_2\},$$

*Electronic address: mp@cs.cas.cz

†Also at Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Karlovo náměstí 13, 121 35 Praha 2—Nové Město, Czech Republic. Electronic address: vejmelka@cs.cas.cz

$$\dot{x}_3 = \alpha\{0.2 + x_3(x_1 - 5.7)\} \tag{1}$$

for the autonomous Rössler system, and

$$\dot{y}_1 = 10(-y_1 + y_2),$$

$$\dot{y}_2 = 28y_1 - y_2 - y_1y_3 + \epsilon x_2^{\beta},$$

$$\dot{y}_3 = y_1y_2 - \frac{8}{3}y_3 \tag{2}$$

for the driven Lorenz system, in which the equation for $\dot{y}_2$ is augmented by a driving term involving x_2. We will analyze the case with $\alpha=6$ and $\beta=2$.

For the second example we will use the unidirectionally coupled Hénon maps, also studied in Refs. [6,9,19], with equations

$$x_1' = 1.4 - x_1^2 + b_1x_2,$$

$$x_2' = x_1 \tag{3}$$

for the driving system $\{X\}$, and

$$y_1' = 1.4 - [\epsilon x_1y_1 + (1-\epsilon)y_1^2] + b_2y_2,$$

$$y_2' = y_1 \tag{4}$$

for the response system $\{Y\}$. Here we will study identical systems with $b_1=b_2=0.3$.

Our third example will be the unidirectionally coupled Rössler systems given by the equations

$$\dot{x}_1 = -\omega_1x_2 - x_3,$$

$$\dot{x}_2 = \omega_1x_1 + a_1x_2,$$

$$\dot{x}_3 = b_1 + x_3(x_1 - c_1) \tag{5}$$

for the autonomous system, and

$$\dot{y}_1 = -\omega_2y_2 - y_3 + \epsilon(x_1 - y_1)$$

$$\dot{y}_2 = \omega_2y_1 + a_2y_2$$

$$\dot{y}_3 = b_2 + y_3(y_1 - c_2) \tag{6}$$

for the response system. In this section we will use parameters $a_1=a_2=0.15$, $b_1=b_2=0.2$, $c_1=c_2=10.0$, and frequencies $\omega_1=1.015$ and $\omega_2=0.985$.

The data from continuous nonlinear dynamical systems were generated by numerical integration based on the adaptive Bulirsch-Stoer method [30] using the sampling interval 0.026 17 for the systems (1) and (2), and 0.1256 for the systems (5) and (6). In the latter case this gives 17–21 samples per one period. When Rössler systems with different frequencies were used, the sampling was updated in order to keep 19–20 samples per period of the faster system (Secs. V and VI).

In all the cases, we denote the driving, autonomous system by $\{X\}$, and the driven, response system by $\{Y\}$. For each of the above three examples we define a set of coupling strength parameters ϵ increasing from $\epsilon=0$ to an ϵ value before the synchronization threshold. As Paluš *et al.* [9] explain, the direction of coupling can be inferred from experimental data only when the underlying systems are coupled, but not yet synchronized. In the numerical examples, the synchronization threshold can be determined using the plot of Lyapunov exponents (LEs) of the coupled systems as the function of the coupling strength ϵ. With increasing ϵ, the positive Lyapunov exponent of the response system (also known as the conditional Lyapunov exponent [31]) decreases, and it becomes negative just at the ϵ value giving the synchronization threshold. The plots of the Lyapunov exponents for the Rössler-Lorenz systems (1) and (2) can be found in Refs. [9,19], the LE plots for the coupled Hénon systems (3) and (4) in Refs. [6,9,19], while further study of the coupled Rössler systems (5) and (6), including their LE plot, can be found in Sec. IV.

For each value of ϵ from the predefined range we numerically generate time series $\{x_i\}$ and $\{y_i\}$ as outputs of the systems $\{X\}$ and $\{Y\}$, obtained by recording the components x_1 and y_1, respectively, and analyze them by using the following three methods.

Two different methods exploit the approach suggested by Rulkov *et al.* [32] based on the assumption of the existence of a smooth map between the trajectories of $\{X\}$ and $\{Y\}$. If such a smooth map exists then closeness of points in the state space X of the system $\{X\}$ implies a closeness of points in the state space Y of the system $\{Y\}$.

One of the methods due to Le Van Quyen *et al.* [7] is based on cross prediction using the well-known idea of mutual neighbors. A known or reconstructed state space (e.g., using a time-delay embedding [33] $\mathbf{X}_n=[x_n,x_{n-\tau},x_{n-2\tau},\ldots]$) must be available. However, instead of using k nearest neighbors, a neighborhood size δ is preselected. Considering a map from X to Y, a prediction is made for the value of y_{n+1} one step ahead using the following formula:

$$\hat{y}_{n+1} = \frac{1}{|V_\delta(\mathbf{X}_n)|}\sum_{j:\mathbf{X}_j\in V_\delta(\mathbf{X}_n)} y_{j+1}. \tag{7}$$

The volume $V_\delta(\mathbf{X}_n)=\{\mathbf{X}_{n'}:|\mathbf{X}_{n'}-\mathbf{X}_n|<\delta\}$ is the δ neighborhood of $\mathbf{X}_n$ and $|V_\delta(\mathbf{X}_n)|$ denotes the number of points contained in the neighborhood. Using data rescaled to zero mean and unit variance, the authors define a cross-predictability index by subtracting the root-mean-square prediction error from 1

$$P(X\to Y) = 1 - \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{y}_{n+1} - y_{n+1})^2}, \tag{8}$$

which should measure how the system $\{X\}$ influences the evolution of the system $\{Y\}$. The cross-predictability index $P(Y\to X)$ in the opposite direction, characterizing the ability of the system $\{Y\}$ to influence the evolution of the system $\{X\}$, is defined in full analogy.

The second approach, proposed by Arnhold *et al.* [17] and analyzed by Quian Quiroga *et al.* [19], uses mean square distances instead of the cross predictions in order to quantify the closeness of points in both spaces. We use the implemen-

tation according to Refs. [19,34] in which a time-delay embedding [33] is first constructed in order to obtain state space vectors $\mathbf{X}$ and $\mathbf{Y}$ for both time series $\{x_i\}$ and $\{y_i\}$, respectively; then the mean squared distance to k nearest neighbors is defined for each $\mathbf{X}$ as

$$R_n^{(k)}(\mathbf{X}) = \frac{1}{k}\sum_{j=1}^{k} |\mathbf{X}_n - \mathbf{X}_{r_{n,j}}|^2, \tag{9}$$

where $r_{n,j}$ denotes the index of the jth nearest neighbor of $\mathbf{X}_n$. The $\mathbf{Y}$-conditioned squared mean distance is defined by replacing the nearest neighbors of $\mathbf{X}_n$ by the equal time partners of the nearest neighbors of $\mathbf{Y}_n$ as

$$R_n^{(k)}(\mathbf{X}|\mathbf{Y}) = \frac{1}{k}\sum_{j=1}^{k} |\mathbf{X}_n - \mathbf{X}_{s_{n,j}}|^2, \tag{10}$$

where $s_{n,j}$ denotes the index of the jth nearest neighbor of $\mathbf{Y}_n$. Then the asymmetric measure

$$S^{(k)}(\mathbf{X}|\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N} \frac{R_n^{(k)}(\mathbf{X})}{R_n^{(k)}(\mathbf{X}|\mathbf{Y})} \tag{11}$$

should reflect interdependence in the sense that closeness of the points in Y implies closeness of their equal time partners in X and the values of $S^{(k)}(\mathbf{X}|\mathbf{Y})$ approach 1, while, in the case of X independent of Y, $S^{(k)}(\mathbf{X}|\mathbf{Y}) \ll 1$. The quantity $S^{(k)}(\mathbf{Y}|\mathbf{X})$ measuring the influence of $\{X\}$ on $\{Y\}$ is defined in full analogy.

The third measure, used in this section, the coarse-grained transinformation rate $i(X \to Y)$, is the average rate of the net amount of information "transferred" from the process $\{X\}$ to the process $\{Y\}$, or, in other words, the average rate of net information flow by which the process $\{X\}$ influences the process $\{Y\}$. The coarse-grained transinformation rate (CTIR), introduced by Paluš *et al.* [9], is based on the conditional mutual information and will be briefly reviewed in Sec. III.

The above three measures as functions of the coupling strength ϵ for the Rössler system (1) driving the Lorenz system (2) are plotted in Fig. 1. With the exception of the weakest coupling ($\epsilon \leqslant 0.6$) the cross predictability of the system $\{Y\}$ by the system $\{X\}$ [the solid line in Fig. 1(a)] is greater than the cross predictability of the system $\{X\}$ by the system $\{Y\}$ [the dashed line in Fig. 1(a)]. Our result in Fig. 1(a) agrees with that of Le Van Quyen *et al.* [7], who interpret the relation $P(X \to Y) > P(Y \to X)$ by the fact, that the autonomous Rössler system $\{X\}$ drives the response Lorenz system $\{Y\}$, and therefore the prediction of $\{Y\}$ from $\{X\}$ is better than the prediction in the opposite direction.

In a similar way, with a few exceptions, the relative average distance of the mutual nearest neighbors $S^{(k)}(\mathbf{Y}|\mathbf{X}) > S^{(k)}(\mathbf{X}|\mathbf{Y})$ [Fig. 1(b)] agrees with the results in [19], suggesting that the state of the response system $\{Y\}$ depends more on the state of the driver system $\{X\}$ than vice versa, as also claimed by Arnhold *et al.* [17]. (Note that the conditioning $X|Y$ reflects the influence $Y \to X$, and vice versa.) The same conclusion about $\{X\}$ driving $\{Y\}$ can be drawn from the CTIR $i(X \to Y) > i(Y \to X)$ [Fig. 1(c)]. The latter inequality holds for all positive values of ϵ but the ϵ values approaching the synchronization threshold which emerges for ϵ slightly above 2 [9,19].

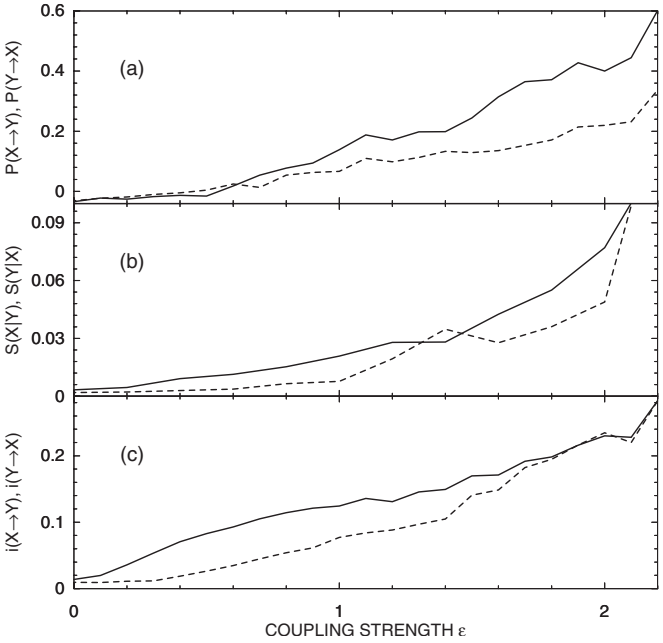

FIG. 1. (a) Cross predictability $P(X \to Y)$ (solid line) and $P(Y \to X)$ (dashed line), (b) relative average distance of the mutual nearest neighbors $S^{(k)}(\mathbf{Y}|\mathbf{X})$ (solid line) and $S^{(k)}(\mathbf{X}|\mathbf{Y})$ (dashed line), and (c) coarse-grained transinformation rate $i(X \to Y)$ (solid line) and $i(Y \to X)$ (dashed line) for the Rössler system (1) driving the Lorenz system (2), as functions of the coupling strength ϵ.

The same analyses as in Fig. 1, but for the unidirectionally coupled Hénon system (3) and (4), are presented in Fig. 2. One can immediately see that in this case $P(X \to Y) < P(Y \to X)$ [Fig. 2(a)]. This result agrees with that of Ref. [6]. Schiff *et al.* [6] offer an interpretation based on the Takens embedding theorem [33]: From the time series $\{x_i\}$ only the system $\{X\}$ can be reconstructed, while from the time series $\{y_i\}$ the whole system consisting of the coupled systems $\{X\}$ and $\{Y\}$ can be reconstructed, and therefore one can predict the driving system from the response system and not vice versa [6]. Also, the relation of the second interdependence measure reverses: In this case the inequality $S^{(k)}(\mathbf{Y}|\mathbf{X}) < S^{(k)}(\mathbf{X}|\mathbf{Y})$ holds [Fig. 2(b)]. Again, our result agrees with that of Ref. [19]. Quian Quiroga *et al.* [19] explain that the higher-dimensional system (obtained by the reconstruction from the time series $\{y_i\}$ which bears information about both the coupled systems) is "more active" than the lower-dimensional (autonomous, driving) system. Only the CTIR gives the same relation as in the previous case: $i(X \to Y) > i(Y \to X)$ [Fig. 2(c)] suggesting the fact that $\{X\}$ influences $\{Y\}$, while $\{X\}$ evolves autonomously.

Figure 3 presents the analysis of the unidirectionally coupled Rössler systems (5) and (6). We can see that the results are qualitatively the same as in the case of the coupled Hénon systems (Fig. 2), although these systems are more similar to the first example of the coupled Rössler-Lorenz systems. We can see that neither the cross predictability nor the mutual nearest neighbors statistics give con-

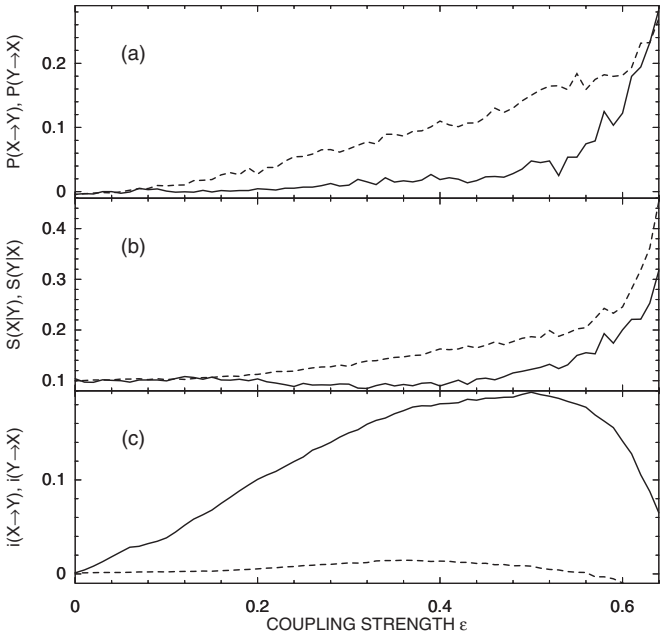

FIG. 2. (a) Cross predictability $P(X\rightarrow Y)$ (solid line) and $P(Y\rightarrow X)$ (dashed line), (b) relative average distance of the mutual nearest neighbors $S^{(k)}(\mathbf{Y}|\mathbf{X})$ (solid line) and $S^{(k)}(\mathbf{X}|\mathbf{Y})$ (dashed line), and (c) coarse-grained transinformation rate $i(X\rightarrow Y)$ (solid line) and $i(Y\rightarrow X)$ (dashed line) for the unidirectionally coupled Hénon system (3) and (4), as functions of the coupling strength ϵ.

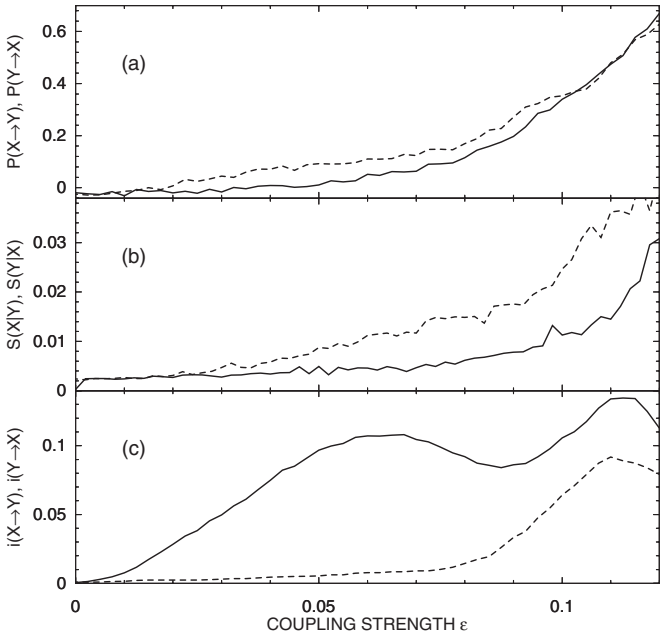

FIG. 3. (a) Cross predictability $P(X\rightarrow Y)$ (solid line) and $P(Y\rightarrow X)$ (dashed line), (b) relative average distance of the mutual nearest neighbors $S^{(k)}(\mathbf{Y}|\mathbf{X})$ (solid line) and $S^{(k)}(\mathbf{X}|\mathbf{Y})$ (dashed line), and (c) coarse-grained transinformation rate $i(X\rightarrow Y)$ (solid line) and $i(Y\rightarrow X)$ (dashed line) for the unidirectionally coupled Rössler systems (5) and (6), as functions of the coupling strength ϵ.

sistent results when using three different examples of unidirectionally coupled systems. Only the coarse-grained transinformation rate correctly identifies the direction of the causal influence in the above three examples, as well as in many other systems of different origins tested by the authors.

In the above examples of unidirectionally coupled systems, we could see that the used measures are generally nonzero in both directions even before the systems become synchronized and comparison of the values of such measures does not always reflect the true causality given by the unidirectional coupling of the studied systems. The intuitively understandable implication lower prediction error (better predictability) $\Rightarrow$ stronger dependence cannot generally be applied for nonlinear systems. When the coupling of the systems is weaker than that necessary for the emergence of synchronization, as used in the above examples, no smooth deterministic function between the states of the systems exists yet. However, there is already some statistical relation valid on the coarse-grained description level. Although the deterministic quantities are based on the existence of a smooth functional relation, when estimated with finite precision they usually give nonzero values influenced not only by the existing statistical dependence but also by properties of the systems other than the coupling. It is therefore necessary to use quantities proposed for measuring statistical dependence, such as information-theoretic measures, which have solid mathematical background and whose properties have been thoroughly studied since their introduction in 1948 [35].

III. DEPENDENCE MEASURES FROM INFORMATION THEORY

In this section we review basic measures from information theory that we will need in further considerations. More details can be found, e.g., in Refs. [35,36]. Then we will describe how these measures can help in inference of causal relations or directionality of coupling.

Consider discrete random variables X and Y with sets of values Ξ and Υ, respectively, and probability distribution functions (PDFs) $p(x)$, $p(y)$ and joint PDF $p(x,y)$. The entropy $H(X)$ of a single variable, say X, is defined as

$$H(X) = -\sum_{x\in\Xi} p(x)\log p(x), \tag{12}$$

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

$$H(X,Y) = -\sum_{x\in\Xi}\sum_{y\in\Upsilon} p(x,y)\log p(x,y). \tag{13}$$

The conditional entropy $H(Y|X)$ of Y given X is

$$H(Y|X) = -\sum_{x\in\Xi}\sum_{y\in\Upsilon} p(x,y)\log p(y|x). \tag{14}$$

The average amount of common information, contained in the variables X and Y, is quantified by the mutual information $I(X;Y)$, defined as

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{15}$$

The mutual information normalized as

$$\iota(X;Y) = \frac{I(X;Y)}{\max(H(X),H(Y))} \tag{16}$$

attains values between 0 and 1, and can be used to define a distance measure $d(X,Y)$ as

$$d(X,Y) = 1 - \iota(X;Y), \tag{17}$$

which has all mathematical properties of a distance in the space of random variables [37]. Thus $d(.,.)$ defines a metric based on the strength of dependence. Independent variables have the maximum distance $[d(.,.)=1]$; functionally related variables have a zero distance.

The conditional mutual information $I(X;Y|Z)$ of the variables X, Y given the variable Z is given as

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z). \tag{18}$$

For Z independent of X and Y we have

$$I(X;Y|Z) = I(X;Y). \tag{19}$$

By a simple manipulation we obtain

$$I(X;Y|Z) = I(X;Y;Z) - I(X;Z) - I(Y;Z). \tag{20}$$

Thus the conditional mutual information $I(X;Y|Z)$ characterizes the net dependence between X and Y without the possible influence of another variable, Z.

The entropy and information are usually measured in bits if the base of the logarithms in their definitions is 2; here we use the natural logarithm and therefore the units are called nats.

Let $\{x(t)\}$ and $\{y(t)\}$ be time series considered as realizations of stationary and ergodic stochastic processes $\{X(t)\}$ and $\{Y(t)\}$, respectively, $t=1,2,3,\ldots$. In the following we will mark $x(t)$ as x and $x(t+\tau)$ as x_τ, and the same notation holds for the series $\{y(t)\}$.

The mutual information $I(y;x_\tau)$ measures the average amount of information contained in the process $\{Y\}$ about the process $\{X\}$ in its future τ time units ahead (τ-future thereafter). This measure, however, could also contain information about the τ-future of the process $\{X\}$ contained in this process itself if the processes $\{X\}$ and $\{Y\}$ are not independent, i.e., if $I(x;y)>0$. In order to obtain the net information about the τ-future of the process $\{X\}$ contained in the process $\{Y\}$ we need the conditional mutual information $I(y;x_\tau|x)$. Using the latter measure Paluš *et al.* [9] defined the coarse-grained transinformation rate as

$$i(Y \to X) = \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} I(y;x_\tau|x) - \frac{1}{2\tau_{max}} \sum_{\tau=-\tau_{max}}^{\tau_{max};\tau\neq 0} I(y;x_\tau). \tag{21}$$

In practical evaluation we do not use $I(y;x_\tau|x)$ for a particular time lag τ, but an average over a range of time lags. Theoretical reasons for this averaging are explained in detail in Ref. [38] and briefly reviewed in Ref. [9]. Here we stress only the practical point of view: Evaluation of $I(y;x_\tau|x)$ as an average over a number of time lags decreases the variance of the estimate. In the following, all results of various forms of the conditional mutual information will be obtained by averaging over time lags $\tau=1,\ldots,50$ samples.

In the original definition of the CTIR [9], also used in our examples in Sec. II, a symmetric dependence term is subtracted from the asymmetric conditional mutual information. Note that this subtraction does not change the relation between $i(Y\to X)$ and $i(X\to Y)$ since

$$\sum_{\tau=-\tau_{max}}^{\tau_{max};\tau\neq 0} I(y;x_\tau) = \sum_{\tau=-\tau_{max}}^{\tau_{max};\tau\neq 0} I(x;y_\tau).$$

IV. AMPLITUDES, PHASES, AND THE COURSE OF DIMENSIONALITY

In the standard statistical language we considered the time series $\{x(t)\}$ and $\{y(t)\}$ as realizations of stochastic processes $\{X(t)\}$ and $\{Y(t)\}$. If the processes $\{X(t)\}$ and $\{Y(t)\}$ are substituted by dynamical systems evolving in measurable spaces of dimensions m and n, respectively, the variables x and y in $I(y;x_\tau|x)$ and $I(x;y_\tau|y)$ should be considered as n- and m-dimensional vectors. In experimental practice, however, usually only one observable is recorded for each system. Then, instead of the original components of the vectors $\vec{X}(t)$ and $\vec{Y}(t)$, the time-delay embedding vectors according to Takens [33] are used. Then, back in time-series representation, we have

$$\begin{aligned} I(\vec{Y}(t);\vec{X}(t+\tau)|\vec{X}(t)) &= I((y(t),y(t-\rho),\ldots,y(t-(m-1)\rho));x(t \\ &\quad +\tau)|(x(t),x(t-\eta),\ldots,x(t-(n-1)\eta))), \end{aligned} \tag{22}$$

where η and ρ are time lags used for the embedding of trajectories $\vec{X}(t)$ and $\vec{Y}(t)$, respectively. For simplicity, only information about one component $x(t+\tau)$ in the τ-future of the system $\{X\}$ is used. The opposite CMI $I(\vec{X}(t);\vec{Y}(t+\tau)|\vec{Y}(t))$ is defined in full analogy. Exactly the same formulation can be used for Markov processes of finite orders m and n. Based on the idea of finite-order Markov processes, Schreiber [18] has proposed a "transfer entropy" which is in fact an equivalent expression for the conditional mutual information (22)—see the Appendix.

Let us return to the unidirectionally coupled Rössler systems (5) and (6). The dependence of their Lyapunov exponents (all but the two that are negative in the uncoupled case) on the coupling strength ϵ is plotted in Fig. 4(a). The change of the positive LE of the response system $\{Y\}$ to negative values slightly under $\epsilon=0.12$ gives the synchronization threshold for these systems. If we evaluate the simple CMI $I(y;x_\tau|x)$ and $I(x;y_\tau|y)$, without subtracting the symmetric term in (21), the CMI's in both direction are positive and increasing with the increasing coupling strength [Fig. 4(b)]. Before the synchronization threshold, the inequality $I(x;y_\tau|y)>I(y;x_\tau|x)$ indicates the correct direction of coupling; however, as we will see in the next section, for reliable inference in general it is desirable to obtain a zero value in the uncoupled direction $Y\to X$. This can be attained by a proper conditioning—the conditioning variable should con-

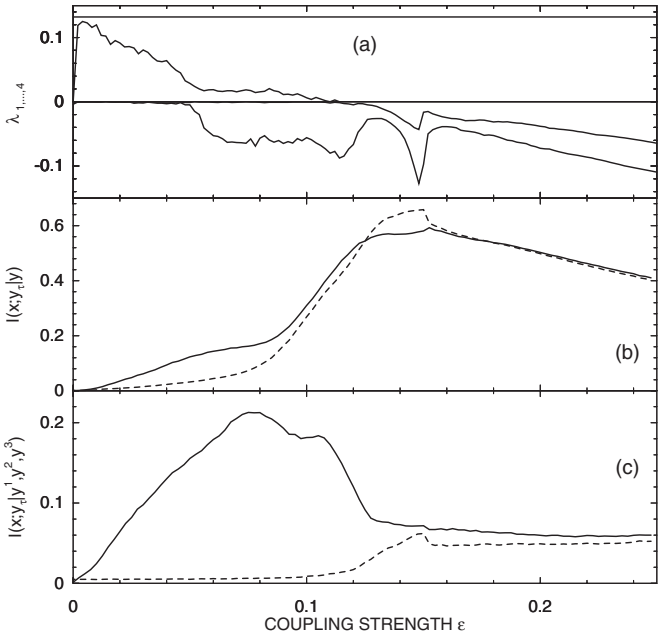

FIG. 4. (a) Two largest Lyapunov exponents of the drive $\{X\}$ (constant lines) and the response $\{Y\}$ (decreasing lines), (b) averaged conditional mutual information $I(x;y_\tau|y)$ (solid line) and $I(y;x_\tau|x)$ (dashed line), and (c) averaged CMI $I(x;y_\tau|\vec{Y})$ (solid line) and $I(y;x_\tau|\vec{X})$ (dashed line), where the vectors $\vec{X}$ and $\vec{Y}$ are the original components of the integrated systems, for the unidirectionally coupled Rössler systems (5) and (6), as functions of the coupling strength ϵ. The Lyapunov exponents are measured in nats per a time unit; the conditional mutual information in all figures is measured in units of nats.

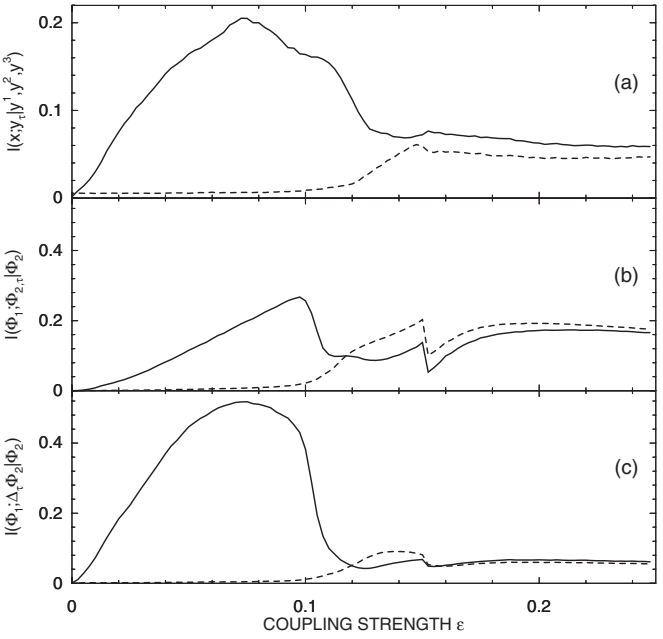

FIG. 5. (a) Averaged conditional mutual information $I(x;y_\tau|\vec{Y})$ (solid line) and $I(y;x_\tau|\vec{X})$ (dashed line), using the time-delay embedding vectors $\vec{X}=[x_1(t),x_1(t-\eta),x_1(t-2\eta)]$ and analogously for $\vec{Y}$, (b) averaged CMI $I(\phi_1(t);\phi_2(t+\tau)|\phi_2(t))$ (solid line) and $I(\phi_2(t);\phi_1(t+\tau)|\phi_1(t))$ (dashed line), and (c) averaged CMI $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid line) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed line), for the unidirectionally coupled Rössler systems (5) and (6), as functions of the coupling strength ϵ.

tain full information about future values of the system or processes generating this variable in the uncoupled case. So it should be a three-dimensional vector $\vec{X}$ or $\vec{Y}$ for the studied Rössler systems. On the other hand, it is sufficient to have just one component of each vector variable for establishing the presence of coupling, i.e., the appropriate measures for inference of coupling directions are the CMI's $I(x;y_\tau|\vec{Y})$ and $I(y;x_\tau|\vec{X})$. Evaluation of the latter quantities brings a five-dimensional estimation problem, which might be hard to solve with a limited amount of available data, not to speak about the seven or nine dimensions if the formal definition (22) is used.

The CMI's $I(x;y_\tau|\vec{Y})$ and $I(y;x_\tau|\vec{X})$ where for the conditioning vectors $\vec{X}$ and $\vec{Y}$ the original components $x_1(t)$, $x_2(t)$, $x_3(t)$ and $y_1(t)$, $y_2(t)$, $y_3(t)$, respectively, were used, are displayed in Fig. 4(c). We can see that $I(y;x_\tau|\vec{X})$ in the uncoupled direction stays at the zero value up to ϵ close to the synchronization threshold, while $I(x;y_\tau|\vec{Y})$ is distinctly positive [Fig. 4(c)]. The CMI's $I(x;y_\tau|\vec{Y})$ and $I(y;x_\tau|\vec{X})$ with the conditioning vectors $\vec{X}$ and $\vec{Y}$ obtained as time-delay embedding [33] from the components $x_1(t)$ and $y_1(t)$, respectively, are displayed in Fig. 5(a). We can see a quite good agreement of the results in Figs. 4(c) and 5(a).

Many interesting processes in physics and biology can be modeled by weakly coupled oscillators and their interactions can be inferred by analyzing the dynamics of their instantaneous phases [1,20,21]. The instantaneous phase of a signal $s(t)$ can be determined by using the analytic signal concept of Gabor [39], recently introduced into the field of nonlinear dynamics within the context of phase synchronization [40,41]. The analytic signal $\psi(t)$ is a complex function of time defined as

$$\psi(t) = s(t) + j\hat{s}(t) = A(t)e^{j\phi(t)}. \tag{23}$$

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

$$\hat{s}(t) = \frac{1}{\pi}\mathrm{P}\int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}d\tau. \tag{24}$$

(P 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

$$\phi(t) = \arctan\frac{\hat{s}(t)}{s(t)}. \tag{25}$$

Since in this paper we are not interested in the instantaneous amplitude $A(t)$, we reserve the word "amplitude" for the "raw" signal (time series) $s(t)$, as opposed to the instantaneous phase $\phi(t)$.

Paluš and Stefanovska [22] have shown that the conditional mutual information can be applied also in inference of coupling of systems from their instantaneous phases, confined in interval $[0,2\pi)$ or $[-\pi,\pi)$ (so-called wrapped phases). Thus we can come back to the time series $\{x(t)\}$ and $\{y(t)\}$ generated by the unidirectionally coupled Rössler systems (5) and (6) and compute their instantaneous phases $\phi_1(t)$ and $\phi_2(t)$, respectively, according to Eqs. (24) and (25). Then we evaluate the conditional mutual information $I(\phi_1(t);\phi_2(t+\tau)|\phi_2(t))$ and $I(\phi_2(t);\phi_1(t+\tau)|\phi_1(t))$ and plot the results in Fig. 5(b). We can see that the CMI evaluated from the phases again distinguishes the driving from the driven system. Moreover, the application of the phase dynamics decreases the dimensionality of the problem—already $I(\phi_2(t);\phi_1(t+\tau)|\phi_1(t))$ with the one-dimensional condition is zero in the uncoupled direction. Even better distinction [Fig. 5(c)] can be obtained when we study dependence between the phase of one system and the phase increment

$$\Delta_\tau\phi_{1,2}(t) = \phi_{1,2}(t+\tau) - \phi_{1,2}(t) \tag{26}$$

of the second system instead of the dependence between $\phi_{1,2}(t)$ and $\phi_{2,1}(t+\tau)$; so that we evaluate the conditional mutual information $I(\phi_1(t);\Delta_\tau\phi_2(t)|\phi_2(t))$ and $I(\phi_2(t);\Delta_\tau\phi_1(t)|\phi_1(t))$; in a shorter notation $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$, respectively. A different approach to detection of coupling direction from the instantaneous phase has been introduced by Rosenblum *et al.* [20,21]. The two approaches are compared in Refs. [22,42].

V. ESTIMATION FROM DATA: BIAS, VARIANCE, AND INFERENCE

Every quantity descriptive of the state of a system or process under study suffers from bias and variance when estimated from noisy, nonstationary experimental data. Using limited, relatively short time series, estimates of complicated quantities such as the conditional mutual information can have non-negligible bias and variance even if evaluated from noise-free, stationary model data. It is necessary to know the behavior of the used estimator of any measure before it is applied in analysis of real data. In order to study the bias and variance of the CMI estimates, we choose a particular coupling strength (ϵ=0.05) and evaluate $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ from 1000 realizations of the unidirectionally coupled Rössler systems (5) and (6), starting in different initial conditions, for various time series length.

In this study we evaluate the CMI using a simple box-counting algorithm based on marginal equiquantization [38,43,44], i.e., a partition is generated adaptively in one dimension (for each variable) so that the marginal bins become equiprobable. This means that the marginal boxes are not defined equidistantly but so that there is approximately the same number of data points in each marginal bin. The only parameter of this method is the number Q of the marginal bins. Paluš [43] proposed that computing mutual information of n variables, the number of marginal bins should not exceed the $(n+1)$st root of the number of the data samples, i.e., $Q \leqslant \sqrt[n+1]{N}$. Cellucci *et al.* [45], who use the same binning procedure, determine the number of bins using the minimum description length criterion. In this paper we use a simple pragmatic choice Q=8 in all computations.

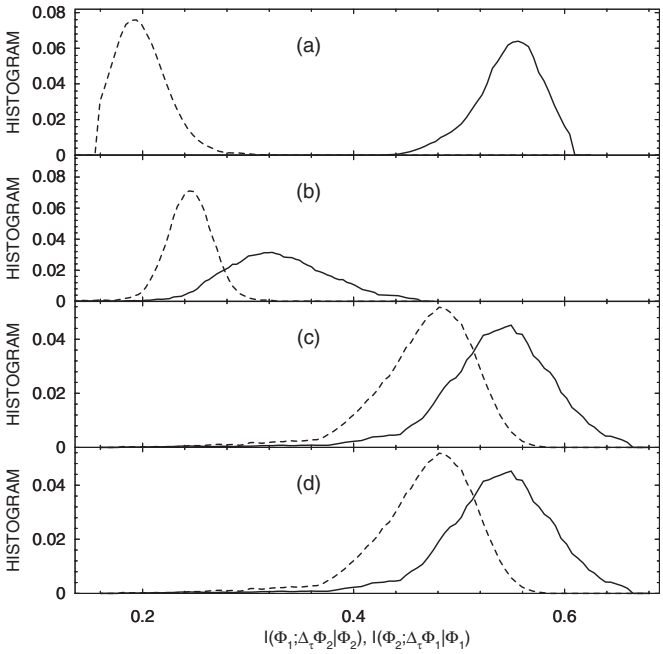

FIG. 6. Histograms of estimates of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed lines) from 1000 realizations of (a) the unidirectionally coupled Rössler systems (5) and (6), the coupling strength ϵ=0.05, and the number of samples N=1024; (b) fast Fourier transform (FFT) surrogate data for the data used in (a), N=1024; (c) the same as in (a), but the number of samples N=512; and (d) FFT surrogate data for the data used in (c), N=512.

The equiquantization method effectively transforms each variable into a uniform distribution, i.e., the individual (marginal) entropies are maximized. This type of mutual information estimate, even its coarse-grained version, is invariant against any monotonous (possibly nonlinear) transformation of data [38].

Histograms obtained from the 1000 CMI estimates, using the series length N=1024 samples, are plotted in Fig. 6(a). We can see the relatively large variance of the estimates and the clear bias of $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ in the uncoupled direction; however, the distinction between the coupled and the uncoupled directions is still clear. When we use the time series length N=512 samples [Fig. 6(c)], not only does the variance increase, but the bias in the uncoupled direction rises, so that the values of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed lines) partially overlap. It is clear that we need some statistical approach to establish critical values of the CMI estimates from which we could infer that the CMI is nonzero due to a coupling and not due to the estimator bias. In other words, we need to find out what bias and variance we can expect from our data, if there is no coupling. For this purpose we use so-called surrogate data [43,46], i.e., artificially generated data that preserve statistical properties of the original data but are randomized so that any possible coupling is removed. Since we evaluate the CMI from the instantaneous

phases, derivation of which gives instantaneous frequencies, as the surrogate data we construct time series with the same frequency distribution as the series under study. The surrogate data with the same sample spectrum as the tested time series can be constructed using the fast Fourier transform (FFT). The FFT of the series is computed, the magnitudes of the (complex) Fourier coefficients are kept unchanged, but their phases are randomized. The surrogate series is then obtained by computing the inverse transform into the time domain. Different realizations of the process are obtained using different sets of the random Fourier phases. In this study, we obtained perfectly consistent results when we either constructed one surrogate realization for each of the 1000 realizations of the coupled Rössler series, or constructed 1000 surrogate realizations using just one realization of the coupled Rössler series.

The FFT surrogates were originally proposed for testing nonlinearity, and in addition to preserving the spectrum also the preservation of the histogram is usually solved [43,46]. As noted above, our CMI estimator is invariant against invertible nonlinear transformations, including histogram transformations. Therefore, we use the simple FFT surrogates without any amplitude adjustment. The two time series $\{x(t)\}$ and $\{y(t)\}$ are randomized independently.

Histograms of estimates of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed lines) from the FFT surrogate data using series lengths $N=1024$ and $N=512$ samples are plotted in Figs. 6(b) and 6(d), respectively. We can see that the average bias of the CMI $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ in the direction $Y\rightarrow X$ in the surrogate data is even larger than in the original data [cf. Figs. 6(a) and 6(b)]. This fact helps us to avoid false detections of causality (positive information flow) in the uncoupled direction: Even though $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ from the data gains positive values, these values are not greater than the values from the (uncoupled) surrogates, and thus such positive CMI values cannot be considered as evidence for causality, nor for a directional interaction. In order to translate these considerations into a statistical test, we integrate the histogram of the surrogate CMI values into a cumulative histogram and find the CMI value (known as the critical value) giving 95% of the CMI distribution, counting from the left side. If a CMI value from the tested data is greater than this critical value, we say that this result is significant at the level $\alpha=0.05$. The meaning of the statistical significance is that the positive value of CMI was obtained by chance (for other reasons than true causality) with probability $p<0.05$. Of course, one can use a stricter test by setting the nominal level to, e.g., $\alpha=0.01$.

Using the cumulative histograms obtained from the 1000 surrogate realizations and having set the nominal value for the significance, $\alpha=0.05$, leading to a critical value for each test, we can use the 1000 realizations of the data from the Rössler systems (5) and (6) for the evaluation of the performance of our test. Comparing the values of $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ in the uncoupled direction with their critical values we obtain the rate of false positive results, while using $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ in the coupled, causal direction we count the rate of correctly positive results, so that we evaluate the sensitivity of the test. The distribution of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ from the surrogate data

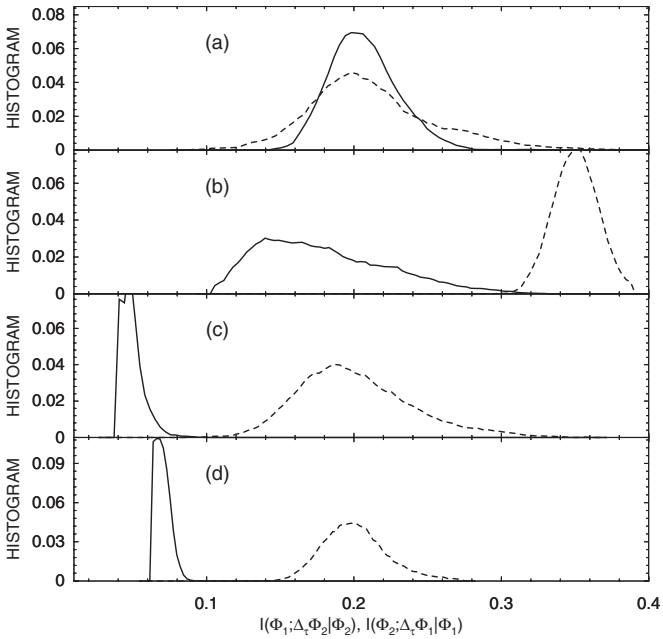

FIG. 7. Histograms of estimates of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed lines) from 1000 realizations of (a) the unidirectionally coupled Rössler systems (5) and (6) with the frequency ratio 1:5, the coupling strength $\epsilon=0.1$, and the number of samples $N=1024$; (b) FFT surrogate data for the data used in (a), $N=1024$; (c) the uncoupled Rössler systems (5) and (6) with the frequency ratio 1:5, $\epsilon=0$, $N=1024$; (d) the uncoupled Rössler systems (5) and (6) with the frequency ratio 1.015:0.985, $N=1024$. In all the cases $a_1=a_2=0.15$.

with $N=1024$ [Fig. 6(b), solid line] allows 100% sensitivity, i.e., values of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ from all 1000 realizations of the original Rössler time series were correctly detected as significant, reflecting truly nonzero causal information flow from $\{X\}$ to $\{Y\}$. In the opposite direction [cf. the histograms plotted by the dashed lines in Figs. 6(a) and 6(b)] we have eight false detections from 1000 realization, i.e., the false detection rate is 0.008, still well under the nominal $\alpha=0.05$. Using $N=512$ samples the sensitivity is worse, giving the value 0.866, i.e., 134 realizations from 1000 were not recognized by the test. The false detection rate, however, was 0.001. With insufficient amount of data the sensitivity of the test could be lowered; however, the surrogate test prevents false detections very well.

Even though there is no coupling in our surrogate data, we can see in Fig. 6(b) that the CMI in the opposite directions, estimated from 1024 samples, have different biases. The only difference between the original Rössler systems (5) and (6) is the small mismatch in their frequencies $\omega_1=1.015$, $\omega_2=0.985$. Now, let us study the same systems, but with $\omega_1=0.5$ and $\omega_2=2.515$, i.e., with the approximate frequency ratio 1:5. Such frequency ratios are typical in studies of cardiorespiratory interactions. We find that in such case the problem of the correct inference of causality is tougher: With the series length $N=1024$ the CMI estimates in the opposite directions overlap [Fig. 7(a)] and the surrogate data [Fig. 7(b)] cannot help to distinguish the coupled from the uncoupled direction. The surrogate data, however, prevent

the false detection of causality [cf. Figs. 7(a) and 7(b)]. In order to demonstrate the "pure" bias, we integrate the systems with the frequency ratio 1:5 without coupling, i.e., setting $\epsilon=0$. We can see that the bias is larger in the direction from the system with $\omega_2=2.515$ to the system with $\omega_1=0.5$ [Fig. 7(c)]. The surrogate data obtained from the uncoupled systems give results similar to the surrogate data for the coupled case [Fig. 7(b)], and again they prevent the false detection of causality in both directions. It seems that the bias is larger in the direction from the faster system ($\omega_2=2.515$) to the slower system ($\omega_1=0.5$). Is this a general rule? Let us return to the previous example of the systems with the approximate frequency ratio 1:1 and analyze them in the uncoupled case ($\epsilon=0$). Here we obtain a larger bias in the direction from the system with $\omega_2=0.985$ to the system with $\omega_1=1.015$ [Fig. 7(d)], i.e., from the slightly slower to the slightly faster system. Let us return to Fig. 4(a) and notice that the system with $\omega_2=0.985$ in the uncoupled case ($\epsilon=0$) is not chaotic (its largest Lyapunov exponent $\lambda_1\approx 0$) and only becomes chaotic due to a weak coupling with the chaotic system with $\omega_1=1.015$. Then, with increasing coupling, its largest Lyapunov exponent λ_1 decreases to negative values at the synchronization threshold. The behavior of the Rössler systems with $\omega_1=0.5$ and $\omega_2=2.515$ is quite similar. The autonomous system with $\omega=0.5$ is chaotic, while the autonomous system with $\omega_2=2.515$ is again quasiperiodic. From these examples we can conclude that the larger bias in the CMI estimates can be observed in the direction from less complex (periodic, quasiperiodic) systems to systems with more complex dynamics.

VI. VARIABILITY OF BIAS, INFLUENCE OF NOISE, AND INFERENCE IN REAL DATA

We attributed the positive departures of the CMI values from zero, in the cases when no information flow existed, to the bias due to insufficient amounts of data. In order to support this statement and to demonstrate that in the uncoupled direction the CMI asymptotically vanishes, we plot the mean CMI estimate (the mean from the 1000 realizations of either the Rössler series or the surrogate data) as a function of the time series length in Fig. 8. In Fig. 8(a) we can see that the convergence to zero of the $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ estimates in the uncoupled direction of the unidirectionally coupled systems (the thick dotted line for the case of the frequency ratio 1:5, and the thick solid line for the case 5:1) is quite similar to that of the uncoupled systems (thin solid and dashed lines) in the same direction, considering the frequencies ω_1 and ω_2 of the systems. Specifically, the bias is larger and the convergence to zero slower in the direction from the system with $\omega=2.515$ to the system with $\omega=0.5$. The latter behavior is preserved in the surrogate data [Fig. 8(b)]: the (overlapping) dashed and dotted lines show the slower convergence to zero of $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ in the direction from the faster to the slower system, in comparison with $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ in the opposite direction (solid lines). Above we have noted that in these cases the property leading to different biases is not the frequency, but the complexity of dynamics of the two systems. In order to study the possible influence of different

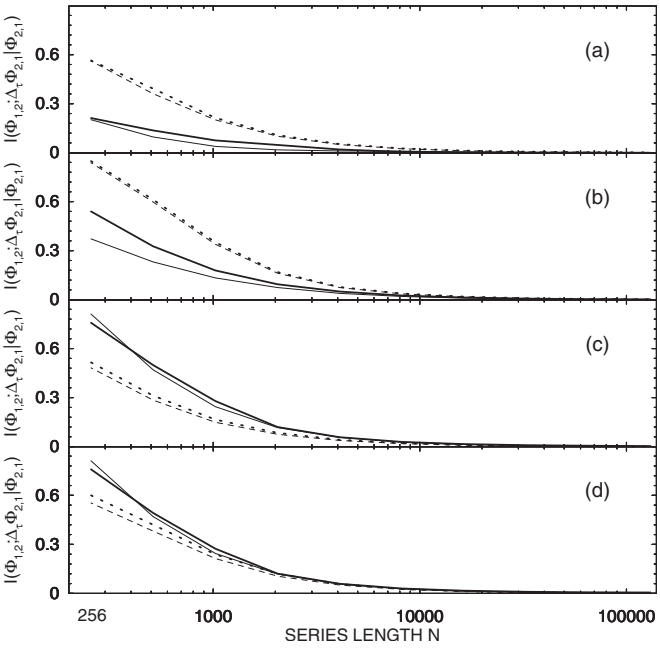

FIG. 8. (a) Convergence with the time series length of mean CMI estimates for $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (thin solid line) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thin dashed line) for the uncoupled Rössler systems (5) and (6) with the frequency ratio 1:5, $a_2=0.15$; $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thick dotted line) for the same systems with the frequency ratio 1:5 and the coupling strength $\epsilon=0.1$ and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thick solid line) for the systems with the frequency ratio 5:1, $a_1=0.15$ and $\epsilon=0.1$. (b) Convergence of the CMI means for the FFT surrogate data of the Rössler systems (5) and (6) with the frequency ratio 1:5, $a_2=0.15$ and the coupled ($\epsilon=0.1$, thick lines) and the uncoupled ($\epsilon=0$, thin lines) cases, $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines), and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed or dotted lines). (c) Convergence with the time series length of mean CMI estimates for $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (thin solid line) and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thin dashed line) for the uncoupled Rössler systems (5) and (6) with the frequency ratio 1:5, $a_2=0.72$; $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thick dotted line) for the same systems with the frequency ratio 1:5 and the coupling strength $\epsilon=0.11$ and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (thick solid line) for the systems with the frequency ratio 5:1, $a_1=0.72$ and $\epsilon=0.11$. (d) Convergence of the CMI means for the FFT surrogate data of the Rössler systems (5) and (6) with the frequency ratio 1:5, $a_2=0.72$, and the coupled ($\epsilon=0.11$, thick lines) and uncoupled ($\epsilon=0$, thin lines) cases, $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$ (solid lines), and $I(\phi_2;\Delta_\tau\phi_1|\phi_1)$ (dashed or dotted lines).

frequencies in the case of systems with comparable complexity, we have found that for $\omega=2.515$ the autonomous Rössler system has a suitable chaotic solution ($\lambda_1=0.12$) when the parameter a is set to $a=0.72$. The autonomous Rössler system with $\omega=0.5$ and the "standard" $a=0.15$ gives $\lambda_1=0.10$. Using these two systems with comparable "chaoticity" we repeated the same convergence studies as presented in Figs. 8(a) and 8(b) and, using the same line codes, we present the results in Figs. 8(c) and 8(d). We can see that, when the two systems have comparable complexity of dynamics, the bias is determined by the main system frequency, and, typically, the bias is larger and the convergence to zero in the uncoupled direction is slower in the direction from the slower to the faster system. We have observed this phenomenon also

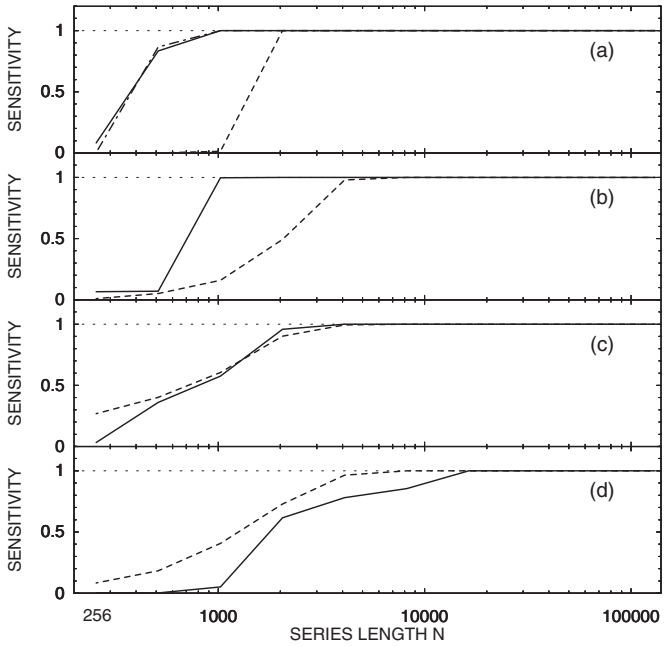

FIG. 9. (a) Sensitivity as function of time series length N for the tests using the coupled Rössler systems (5) and (6) with the frequency ratio 1:1 (dash-dotted line), 5:1 (solid line), and 1:5 (dashed line) for $a_1=a_2=0.15$. The Hilbert phases were used. (b) Sensitivity of the tests for the coupled Rössler systems (5) and (6) with the frequency ratio 5:1 (solid line) and 1:5 (dashed line) for $a_1=a_2=0.15$, as a function of the series length N. The marked event phases were used. (c) Sensitivity as function of time series length N for the tests using the coupled Rössler systems (5) and (6) with the frequency ratio 5:1, $a_1=0.72$, $a_2=0.15$ (solid line), and 1:5, $a_1=0.15$, $a_2=0.72$ (dashed line). The Hilbert phases were used. (d) The same as in (c), but the marked event phases were used.

in studies of noisy periodic oscillators or simple noisy phase oscillators (such as Eq. (3) in [22]) with different frequencies.

Do the different biases influence the performance of the tests for the inference of the causal direction? The sensitivity of the tests, based on evaluation of $I(\phi_1;\Delta_\tau\phi_2|\phi_2)$, as a function of the time series length N is illustrated in Fig. 9. As already mentioned above, the rate of the detection of the causal direction for the data from the Rössler systems (5) and (6) with the frequency ratio 1:1 was 0.87 for the series length $N=512$. For shorter time series, the test completely loses its sensitivity, while for $N=1024$ and more the test has 100% sensitivity [Fig. 9(a), dash-dotted line partially overlaping with the solid line], while the false detection rate is in the range from 0.001 to 0.008, well under the nominal $\alpha=0.05$. Analyzing the data from the Rössler systems (5) and (6) with the frequency ratio 1:5 and $a_1=a_2=0.15$, the sensitivity is still close to zero for $N=1024$, and rises to 100% from $N=2048$ [Fig. 9(a), dashed line], while the rate of false positives ranges from 0.001 to 0.05. Setting the frequency ratio to 5:1, the sensitivity [Fig. 9(a), solid line] behaves similarly to that of the system with the frequency ratio 1:1, while the rate of false positives is always zero. In the case 5:1 the bias is in the same direction as the true causality, so the test has slightly better performance than in the case 1:1, while in the case 1:5 the bias is directed in the opposite direction than the true causality and the test requires more data for the reliable detection of the true causality.

Estimation of the test critical values from the empirical cumulative histograms, as we have done above, is a generally correct approach, independent of the actual distribution of the CMI (or other quantity) values. This approach, however, requires a large number (1000 in this study) of the surrogate realizations. In many published applications of surrogate tests, authors evaluate their test statistics from a small number of surrogate realizations and express the difference between the tested values and the surrogate mean in the number of standard deviations of the surrogates. Then, for assessing significance of the result, they use the critical values theoretically derived from the normal distribution. As we have seen in the above Figs. 6 and 7, the surrogate distribution may be quite different from the normal one. Comparing the two approaches, we have found that once we had enough data to obtain 100% sensitivity of the test ($N\geqslant 1024$ for the cases 1:1 and 5:1), both approaches gave equivalent results, even using so few as 30 surrogate realizations in the test based on the normal distribution. The difference is important when the test is done "on the edge" of its sensitivity, i.e., with $N=512$ in this case. Then we had the sensitivity 0.86 using the histogram approach and 1000 surrogate realizations. The normal distribution approach gives the sensitivity 0.48 using all 1000 surrogate realizations and only 0.29 when 30 surrogate realizations were used for estimating the surrogate mean and standard deviation. One must keep in mind that this "edge" amount of data increases when the studied systems have different dynamics or the analyzed data are contaminated by noise, as is demonstrated below.

The above used "Hilbert" phases (25), i.e., the instantaneous phases obtained by using the Hilbert transform (24) from continuous signals, are not always available in experimental practice. In studies of cardiorespiratory interactions usually the instantaneous phases of the cardiac and respiratory oscillations are computed by the so-called marked events method: Let t_k and t_{k+1} be the times of two consecutive events, here peaks in the signal (ECG or respiratory signal). The instantaneous phases are then linearly interpolated as [5]

$$\phi(t)=2\pi\frac{t-t_k}{t_{k+1}-t_k},\quad t_k\leqslant t<t_{k+1}. \tag{27}$$

Using such phases in the CMI estimation, especially considering the distribution of the phase increments (26), could certainly influence the performance of the above causality tests. Therefore we integrated the above studied Rössler systems (5) and (6) with ten times higher sampling and defined the marked events as times when the coordinate x_1 or y_1 passed the Poincaré section given by $x_1=0$ or $y_1=0$, respectively. Then we generated the marked events phases (27) and subsampled them in order to have the same sampling as when using the Hilbert phases. In the marked event phases the only information about the underlying dynamics is the duration of individual cycles, while the intracycle dynamics is lost. This information reduction leads to a decrease of the CMI values in the coupled direction and, consequently, to

decreased sensitivity of the causality tests. In other words, for a test with the sensitivity of 100%, we need more data when we use the marked event phases than using the Hilbert phases [cf. Figs. 9(a) and 9(b)]. The same conclusion can be drawn for the case when both the systems are chaotic, i.e., having the frequency ratio 5:1 and $a_1=0.72$, $a_2=0.15$ (the solid line); and 1:5, $a_1=0.15$, $a_2=0.72$ [the dashed line in Fig. 9(c) for the Hilbert phases and Fig. 9(d) for the marked event phases]. Especially in the latter case [Fig. 9(d)] there is a difference in sensitivities caused by different biases due to different system main frequencies. In systems with comparable complexity (chaoticity), the bias is larger in the direction from the slower to the faster system. Therefore, in the case 5:1, the bias is oriented against the true causality in the direction from the faster to the slower system, and the test in this direction requires at least $N=16\,384$ data samples in order to gain 100% sensitivity [Fig. 9(d), the solid line].

In applications of the marked event phases, the only available data are the event times $\ldots,t_k,t_{k+1},\ldots$, or the interevent intervals $\ldots,t_{k+1}-t_k,\ldots$. Then a simple way to construct surrogate data is random permutation of the interevent intervals before the surrogate marked event phases are computed according to (27). This type of surrogate data was used with the marked event phases for the above Rössler systems, as well as in the real data example presented at the end of this section.

Considering applications to real data, influence of noise on the presented test should also be evaluated. Gaussian noise has been added to the raw data from the coupled Rössler systems (5) and (6) with the frequency ratio 1:1 (exactly 1.015:0.985). The amount of added noise is characterized by the noise standard deviation (SD) expressed as a percentage of the SD of the original data, e.g., 10% of noise, means that $\mathrm{SD(noise)}=0.1\times\mathrm{SD(data)}$, or 100% of noise means $\mathrm{SD(noise)}=\mathrm{SD(data)}$. The noisy data were processed and tested in the same way as the noise-free data above. The test sensitivity, i.e., the rate of true positive detections of causality, as well as the rate of false positives, i.e., the rate of formal detections of causality in uncoupled directions, as functions of time series length N are illustrated in Fig. 10. The higher the amount of noise in the data, the more data samples are required in order to obtain 100% sensitivity of the test [Fig. 10(a)]. For moderate amounts of noise, the rate of false positives remains well under or about the nominal value $\alpha=0.05$ [Fig. 10(b)]. With large amounts of noise, however, the attainment of the 100% sensitivity is followed by an increase of the rate of false positives. With 100% of noise in the data, the rate of false positives goes to 1 (i.e., to 100%) even before the sensitivity rises from 0 to 1, i.e., the detection ability of the test is completely lost. For amounts from 50% of noise there is a bounded range of time series lengths for which the test is reliable, e.g., from 8192 to 32 768 samples for 50%, but only around 16 384 samples for 70% of noise. The applicability of the test is limited when the data are contaminated by a large amount of noise. An improvement is possible not only by preprocessing and denoising of the data, but potentially also by considering alternative approaches in the three stages of the test: the phase estimation, the surrogate data construction, and the conditional mutual information estimation. Various approaches to the latter will be discussed separately.

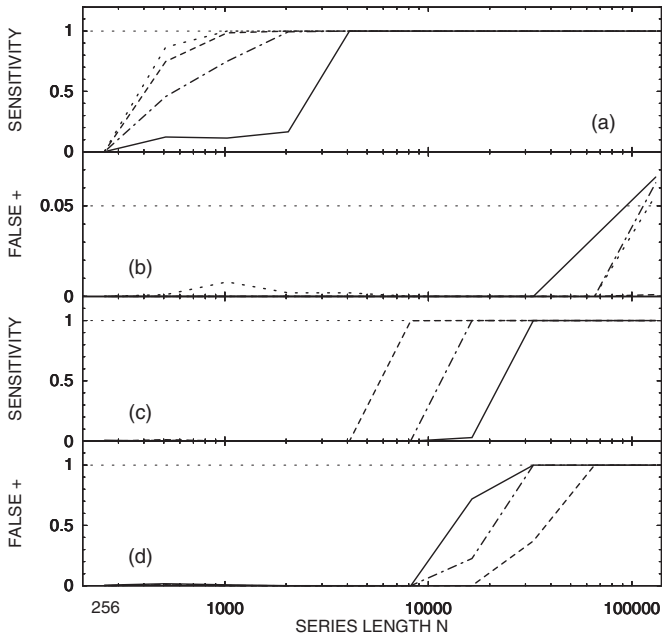

FIG. 10. Sensitivity (a),(c) and the rate of false positives (b),(d) as functions of time series length N for the tests using the coupled Rössler systems (5) and (6) with the frequency ratio 1:1 for different amounts of noise in the data. The portions of noise are in (a),(b) 0% (dotted line), 10% (dashed line), 20% (dash-dotted line), and 30% (solid line); in (c),(d) 50% (dashed line), 70% (dash-dotted line), and 100% (solid line).

In order to demonstrate how the discussed approach can be applied to real data, we use cardiac and respiratory data from an animal experiment described in [5]. Detailed account of the causality analysis in the cardiorespiratory interactions of anesthetized rats will be given elsewhere [42]. Using the interbeat and interbreath intervals, we construct the marked event phases ϕ_C, ϕ_R for the cardiac and respiratory dynamics, respectively. We estimated $I(\phi_C;\Delta_\tau\phi_R|\phi_R)$ and $I(\phi_R;\Delta_\tau\phi_C|\phi_C)$ from 33-min recordings which gave, after subsampling to 40 Hz, the series length $N=80\,000$ samples. Thus we can expect a good performance of the tests using the marked event phases from experimental, possibly noisy data. In the first test, presented in Figs. 11(a) and 11(b) we tried to evaluate the bias and the ability of the surrogate data to prevent possible false detections. For this test we used the cardiac data from one animal and the respiratory data from another animal, so true causality cannot exist in this case. The CMI estimates are positive (although small, but this is typical using the marked events phases) in both the directions and the value of $I(\phi_R;\Delta_\tau\phi_C|\phi_C)$ [Fig. 11(b)] reflecting the influence of the (slower) respiratory rhythm on the (faster) cardiac dynamics is larger than $I(\phi_C;\Delta_\tau\phi_R|\phi_R)$ [Fig. 11(a)] in the opposite direction. Both values from the tested data, however, lie inside the surrogate histograms. The latter result means that the CMI values are not significantly larger than zero and no causality, or no information flow, exists in either direction, as we expected using the data from two different animals. The situation is different when analyzing data from a single animal. While there is no significant influence of the cardiac dynamics on the respiration [Fig.

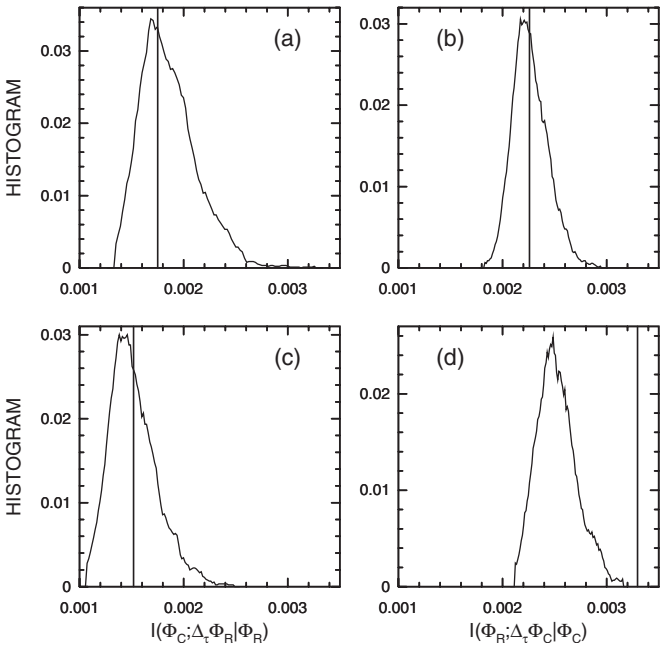

FIG. 11. Tests of causal influences of cardiac oscillations on respiratory oscillations measured by $I(\phi_C;\Delta_\tau\phi_R|\phi_R)$ (a),(c) and of the influence of the respiratory rhythm on the cardiac oscillations given by $I(\phi_R;\Delta_\tau\phi_C|\phi_C)$ (b),(d). Values from the tested data are marked by the vertical lines, the surrogate ranges are illustrated by the histograms obtained from 2500 surrogate realizations. (a),(b) The test of bias using data from two animals; (c),(d) a real test for one of the animals.

11(c)], the influence of the respiratory rhythm on the cardiac dynamics is clearly significant, since $I(\phi_R;\Delta_\tau\phi_C|\phi_C)$ from the tested data lies outside the surrogate distribution [Fig. 11(d)] and thus even this small value $I(\phi_R;\Delta_\tau\phi_C|\phi_C)=0.0033$ nat is significantly positive.

In order to assure reliable tests, we used a large amount of data. Due to the anesthetized state we can expect an acceptable level of stationarity in this case. In other applications, however, only shorter data series are available. It is therefore necessary to find out how to improve the performance of the tests. One way is to use more sophisticated estimators of the dependence measure. The behavior of several different estimators of the conditional mutual information will be discussed in a separate presentation. Another way is the application of different randomization schemes for the interevents intervals, or better, to prefer the "continuous" Hilbert phases and more sophisticated surrogate data, such as the recurrence (twin) surrogates [47].

VII. CONCLUSION

Inference of direction of coupling, or causality, as we can say when we consider two possibly coupled systems, is not a trivial problem. In this paper we identified some problems that could lead to incorrect inference of causality from experimental bivariate time series. Let us summarize the basic problems and ways of coping with them.

A. Dependence measures

As noted by Paluš *et al.* [9], the direction of coupling can be inferred when two systems are coupled, but not yet fully synchronized. This can be understood by considering the example of identical synchronization. Once the (identical) systems are synchronized, they produce identical time series, and there is no way to infer the correct causality relation just from the measured data. In the case of generalized synchronization, there is a one-to-one relation between the states of the systems. Time series $\{x(t)\}$ can be predicted from time series $\{y(t)\}$, and vice versa. Although some dependence measures, including those based on prediction errors, can give different values for the relations $x\rightarrow y$ and $y\rightarrow x$, these values are not given by the causality relations but rather by properties of the functional relation between the states of the systems, e.g., by its Jacobian. The causal relation can be inferred only when coupling is weaker than that necessary for emergence of synchronization, or when the synchronized state is frequently perturbed by variability in coupling or by internal or external noise driving the systems out of the synchronized state. Then the relation between the system states is not deterministic, but probabilistic, and can be measured by measures of statistical dependence, such as the above introduced information-theoretic measures.

B. Asymmetric dependence measures in uncoupled direction

As we observed above, asymmetric measures of dependence can have nonzero values even in the uncoupled direction in cases of unidirectional coupling. This holds for both probabilistic and deterministic measures. Even though no deterministic relation exists before the systems are synchronized, the deterministic measures, estimated in a coarse-grained approximation, reflect the statistical dependence that occurs in both direction even in the case of unidirectional coupling. Mutual comparison of these positive values or the positivity/(negativity) of their difference (sometimes rescaled to some "directionality indices") does not necessarily indicate the correct causal direction. For a correct inference of causality it is desirable to have a measure that vanishes in the uncoupled direction in the case of unidirectional coupling, so that we can identify the causal direction by its statistically significant digression from zero, while in the uncoupled direction the measure does not cross the borders of a "statistical zero." The latter is given by the range obtained from appropriate surrogate data, separately for each direction. As a measure satisfying this requirement we introduced the conditional mutual information.

C. Amplitudes, phases, and the course of dimensionality

The proper conditioning which assures vanishing CMI in the uncoupled direction should contain full information about the future in the uncoupled state of the system, the influence on which is evaluated. This means that in a case of m-dimensional dynamical system, or a variable which can be modeled by a (possibly nonlinear) autoregressive process of order m, the proper condition is an m-dimensional vector. The order m should be estimated from studied data before causality tests are applied. Then the estimation of the $(m+2)$-dimensional probability distribution functional can also be a nontrivial problem. It can be helpful if the studied cou-

pling can be reflected in the dynamics of instantaneous phases, since, in the case of phase dynamics, one-dimensional conditioning is sufficient in many cases.

D. Estimator bias and significance testing

Having an appropriate asymmetric dependence measure, asymptotically vanishing in the uncoupled direction, the inference of causality can be complicated by a bias in estimation from a limited amount of possibly noisy data. Therefore we need to establish a statistical significance of the obtained result, e.g., by the application of the surrogate data testing approach. The surrogate data, however, should reflect statistical and dynamical properties of the tested data, since those can be the source of bias. It is necessary to test that the surrogate data preserve the frequency distribution of the original data, which might be more important than the amplitude distribution. For instance, it might be entirely incorrect to make tests against white noise surrogates, obtained by random permutation of amplitude time series, even though they preserve amplitude distributions. Exceptional care must be applied when we study relations between systems that have different main frequencies, or different complexity of dynamics, or even different variability.

E. Test critical values

In many applications of surrogate data, the significance of the departure of the tested values from the surrogate range is based on the (not always explicitly stated) assumption of a normal distribution of the test measure estimated from the surrogate data. As we have observed, it is necessary to study the surrogate distributions from large enough surrogate ensembles in order to establish the critical test values independently of the form of the distribution and compare them with the critical values based on the normal distribution and estimated from a small number of surrogate realizations.

F. Test performance and data amounts

Before real data applications, it is always useful to assess the performance of any test using appropriate model data in order to estimate the amount of data necessary for reliable inference. Subsampled data can cause problems, while increasing the data amount by oversampling does not improve the test performance. Inference of causality usually requires more data than detection of synchronization.

G. Reliable inference of causality

In this paper we have discussed several problems encountered in inference of causal relations between two systems, i.e., in the identification of the driving and driven systems from experimental time series. We have summarized our extensive experience from development of a reliable causality test, suitable for practical applications. We hope that the above information will help other authors to avoid interpreting numerical and statistical artifacts as causality and to detect true directional interactions.

ACKNOWLEDGMENTS

The authors would like to thank B. Musizza and the BRACCIA team for providing the animal cardiorespiratory data. This study was supported by the EC FP6 project BRACCIA (Contract No. 517133 NEST), and in part by the Institutional Research Plan no. AV0Z10300504.

APPENDIX

Using the idea of finite-order Markov processes, Schreiber [18] introduced a measure quantifying causal information transfer between systems evolving in time, based on appropriately conditioned transition probabilities. Assuming that the system under study can be approximated by a stationary Markov process of order k, the transition probabilities describing the evolution of the system are $p(i_{n+1}|i_n,\ldots,i_{n-k+1})$. If two processes I and J are independent, then the generalized Markov property

$$p(i_{n+1}|i_n,\ldots,i_{n-k+1}) = p(i_{n+1}|i_n^{(k)},j_n^{(l)}) \tag{A1}$$

holds, where $i_n^{(k)}=(i_n,\ldots,i_{n-k+1})$ and $j_n^{(l)}=(j_n,\ldots,j_{n-l+1})$ and l is the number of conditioning states from process J. Schreiber [18] proposed using the Kullback-Leibler divergence to measure the deviation of the transition probabilities from the generalized Markov property (A1). This results in the definition

$$T_{J\to I} = \sum p(i_{n+1},i_n^{(k)},j_n^{(l)})\log\frac{p(i_{n+1}|i_n^{(k)},j_n^{(l)})}{p(i_{n+1}|i_n^{(k)})}, \tag{A2}$$

denoted the *transfer entropy*. The transfer entropy can be understood as the excess amount of bits that must be used to encode information on the state of the process by erroneously assuming that the actual transition probability distribution function is $p(i_{n+1}|i_n^{(k)})$, instead of $p(i_{n+1}|i^{(k)},j_n^{(l)})$.

Let us do a few simple manipulations with the conditional probabilities in (A2):

$$\begin{aligned} T_{J\to I} &= \sum p(i_{n+1},i_n^{(k)},j_n^{(l)})\log\frac{p(i_{n+1},i_n^{(k)},j_n^{(l)})}{p(i_{n+1}|i_n^{(k)})p(i_n^{(k)},j_n^{(l)})} \\ &= \sum p(i_{n+1},i_n^{(k)},j_n^{(l)})\log\frac{p(i_{n+1},i_n^{(k)},j_n^{(l)})}{p(i_{n+1}|i_n^{(k)})p(i_n^{(k)},j_n^{(l)})}\frac{p(i_n^{(k)})}{p(i_n^{(k)})} \\ &= \sum p(i_{n+1},i_n^{(k)},j_n^{(l)})\log\frac{p(i_{n+1},j_n^{(l)}|i_n^{(k)})}{p(i_{n+1}|i_n^{(k)})p(j_n^{(l)}\ |i_n^{(k)})}. \end{aligned}$$

Finally,

$$\begin{aligned} T_{J\to I} = &\sum p(i_{n+1},i_n^{(k)},j_n^{(l)})\log p(i_{n+1},j_n^{(l)}|i_n^{(k)}) \\ &- \sum p(i_{n+1},i_n^{(k)})\log p(i_{n+1}|i_n^{(k)}) \\ &- \sum p(i_n^{(k)},j_n^{(l)})\log p(j_n^{(l)}\ |i_n^{(k)}). \end{aligned} \tag{A3}$$

Now, considering Eq. (18), let us go back to the expression for conditional mutual information (22) and express it using conditional entropies as

$$I(\vec{Y}(t);\vec{X}(t+\tau)|\vec{X}(t)) = H((y(t),y(t-\rho),\ldots,y(t-(m-1)\rho))|(x(t),x(t-\eta),\ldots,x(t-(n-1)\eta)))$$
$$+ H(x(t+\tau)|(x(t),x(t-\eta),\ldots,x(t-(n-1)\eta)))$$
$$- H((y(t),y(t-\rho),\ldots,y(t-(m-1)\rho)),x(t+\tau)|(x(t),x(t-\eta),\ldots,x(t-(n-1)\eta))). \quad \text{(A4)}$$

Next, let us express the conditional entropies using the probability distributions; however, let us change our notations according to Schreiber [18] by equating $I \equiv \{X(t)\}$, $m=k$, and $J \equiv \{Y(t)\}$, $n=l$, substituting t for n, and setting $\eta=\rho=\tau=1$. We can see that we obtain the same expression as Eq. (A3) for the transfer entropy. Thus the transfer entropy is an equivalent expression for the conditional mutual information (22).

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization. A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, U.K., 2001).
[2] C. Schäfer, M. G. Rosenblum, J. Kurths, and H.-H. Abel, Nature (London) **392**, 239 (1998); C. Schäfer, M. G. Rosenblum, H.-H. Abel, and J. Kurths, Phys. Rev. E **60**, 857 (1999).
[3] M. Paluš and D. Hoyer, IEEE Eng. Med. Biol. Mag. **17**, 40 (1998).
[4] M. Bračič Lotrič and A. Stefanovska, Physica A **283**, 451 (2000).
[5] A. Stefanovska, H. Haken, P. V. E. McClintock, M. Hožič, F. Bajrović, and S. Ribarič, Phys. Rev. Lett. **85**, 4831 (2000).
[6] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996).
[7] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela, Physica D **127**, 250 (1999).
[8] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H.-J. Freund, Phys. Rev. Lett. **81**, 3291 (1998).
[9] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[10] M. Paluš, V. Komárek, T. Procházka, Z. Hrnčíř, and K. Štěrbová, IEEE Eng. Med. Biol. Mag. **20**, 65 (2001).
[11] N. Wiener, in *Modern Mathematics for Engineers*, edited by E. F. Beckenbach (McGraw-Hill, New York, 1956).
[12] C. W. J. Granger, Econometrica **37**, 424 (1969).
[13] J. Geweke, in *Handbook of Econometrics*, edited by Z. Griliches and M. D. Intriligator (North-Holland, Amsterdam, 1984), Vol. 2, 1101–1144.
[14] U. Triacca, Theor. Appl. Climatol. **81**, 133 (2005).
[15] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, Biol. Cybern. **85**, 145 (2001).
[16] K. J. Blinowska, R. Kuś, and M. Kamiński, Phys. Rev. E **70**, 050902(R) (2004).
[17] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, Physica D **134**, 419 (1999).
[18] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[19] R. Quian Quiroga, J. Arnhold, and P. Grassberger, Phys. Rev. E **61**, 5142 (2000).
[20] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001).
[21] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, Phys. Rev. E **65**, 041909 (2002).
[22] M. Paluš and A. Stefanovska, Phys. Rev. E **67**, 055201(R), (2003).
[23] K. Otsuka, Y. Miyasaka, T. Kubota, and J. Y. Ko, Phys. Rev. E **69**, 046201 (2004).
[24] P. F. Verdes Phys. Rev. E **72**, 026222 (2005).
[25] I. I. Mokhov and D. A. Smirnov, Geophys. Res. Lett. **33**, L03708 (2006).
[26] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Phys. Rev. E **65**, 041903 (2002).
[27] M. Chávez, J. Martinerie, and M. Le Van Quyen, J. Neurosci. Methods **124**, 113 (2003).
[28] J. Brea, D. F. Russell, and A. B. Neiman Chaos, Chaos **16**, 026111 (2006).
[29] R. Marschinski and H. Kantz, Eur. Phys. J. B **30**, 275 (2002).
[30] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, U.K. 1986).
[31] K. Pyragas, Phys. Rev. E **56**, 5183 (1996).
[32] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, Phys. Rev. E **51**, 980 (1995).
[33] F. Takens, *Dynamical Systems and Turbulence (Warwick 1980)*, edited by D. A. Rand and L. S. Young, Lecture Notes in Mathematics Vol. 898, (Springer, Berlin, 1981), p. 366.
[34] R. Quian Quiroga, FORTRAN computer code, http://www.vis.caltech.edu/~rodri/programs/synchro.for
[35] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (J. Wiley and Sons, New York, 1991).
[37] Y. Horibe, IEEE Trans. Syst. Man Cybern. **15**, 641 (1985).
[38] M. Paluš, Physica D **93**, 64 (1996).
[39] D. Gabor, J. IEE London **93**, 429 (1946).
[40] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **76**, 1804 (1996).
[41] M. Paluš, Phys. Lett. A **235**, 341 (1997).
[42] B. Musizza *et al.*, J. Physiol. (London) **580**(1), 315 (2007).
[43] M. Paluš, Physica D **80**, 186 (1995).
[44] M. Paluš, Neural Network World **7**(3), 269 (1997).
[45] C. J. Cellucci, A. M. Albano, and P. E. Rapp, Phys. Rev. E **71**, 066208 (2005).
[46] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Physica D **58**, 77 (1992).
[47] M. Thiel, M. C. Romano, J. Kurths, M. Rolfs, and R. Kliegl, Europhys. Lett. **75**, 535 (2006).