Inferring the directionality of coupling with conditional mutual information

Martin Vejmelka*
Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic and Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Praha, Czech Republic

Milan Paluš
Institute of Computer Science, Academy of Sciences of the Czech Republic, Praha, Czech Republic
(Received 15 August 2007; published 21 February 2008)

Uncovering the directionality of coupling is a significant step in understanding drive-response relationships in complex systems. In this paper, we discuss a nonparametric method for detecting the directionality of coupling based on the estimation of information theoretic functionals. We consider several different methods for estimating conditional mutual information. The behavior of each estimator with respect to its free parameter is shown using a linear model where an analytical estimate of conditional mutual information is available. Numerical experiments in detecting coupling directionality are performed using chaotic oscillators, where the influence of the phase extraction method and relative frequency ratio is investigated.

DOI: 10.1103/PhysRevE.77.026214 PACS number(s): 05.45.Tp, 05.10.−a

I. INTRODUCTION

Cooperative behavior of coupled complex systems has recently raised much interest in the scientific community [1], since synchronization and related phenomena have been observed not only in physical but also in many biological systems. Examples include cardiorespiratory interaction [2–7] and the synchronization of neural signals [8–12]. In such physiological systems it is important not only to detect synchronized states, but also to identify drive-response relationships. Several indices detecting the directionality of coupling have been proposed. For example, Schreiber [13] proposes to compute the *transfer entropy*, based on the Kullback entropy measuring the deviation of the transition probability density function (PDF) from the generalized Markov property. In Refs. [14,15] the authors propose to approximate functional relationships between the instantaneous phases of interacting oscillators using their Fourier expansions, and compute a normalized "directionality index."

Other approaches include various cross-prediction methods which attempt to directly exploit Granger's ideas on mutual forecasting of series generated by coupled linear systems [16]. However, opposite opinions exist on how to interpret the cross-prediction accuracy, e.g., in Ref. [8] the authors hypothesize that the average cross-prediction error is smaller in the driving system while in Ref. [9] it is suggested that the cross-prediction error should be smaller when predicting the driven system.

In this paper, we present a nonparametric method of detecting the directionality of coupling based on information theory. The method involves estimating a well-known information theoretic functional—the conditional mutual information (CMI) [17]. There is a multitude of ways to estimate the CMI [18]. We select some of them and compare them from two standpoints: estimator bias on a linear model and correct detection rates on a pair of chaotic oscillators with known coupling parameters.

When processing experimental data the situation is quite complicated. Measurement noise and limited length of experimental time series can be sources of considerable additional variance in the estimates. Different statistical and dynamical properties (stochasticity, dominant frequencies) of the two underlying systems can cause severe bias in estimates of directionality indices. We show how it is possible to alleviate these effects to a large extent by testing the computed indices using sets of surrogate data [19,20]. Nevertheless, it should be noted that inferring directionality in systems with very different properties remains a much more difficult problem than inferring directionality in similar systems.

The paper is organized as follows. The Introduction continues with a brief overview of relevant concepts of information theory, the problem of detecting coupling directionality in general, exploiting phase dynamics in particular, and closes with the problem of significance testing. The next section introduces some prospective methods of estimating information theoretic functionals. Two experimental sections detailing the setup and results of several numerical experiments follow, and the last section contains a discussion of the experimental results.

A. Brief introduction to Information theory

Quantities based on information theoretic functionals have enjoyed an important position in detecting relationships between complex systems partly due to their nonparametric nature, which makes them widely applicable. In this section a brief introduction to the basic functionals is given.

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

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

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

*vejmelka@cs.cas.cz

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

for discrete sets Ξ and Υ. It is straightforward to extend the definition to more than two variables. The *conditional entropy* $H(Y|X)$ of Y given X is

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

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

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{3}$$

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

$$I(X;Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z), \tag{4}$$

$$I(X;Y|Z) = I(X;Y;Z) - I(X;Z) - I(Y;Z), \tag{5}$$

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

all of which are theoretically equivalent. However, depending on the method used to compute entropy or mutual information, some of the above formulas will be more appropriate than others in particular situations.

Entropy and mutual information are measured in *bits* if the base of the logarithms in their definitions is 2. In this work the natural logarithm is used and therefore the estimates are given in *nats*.

B. Detecting coupling directionality

Most causality detection methods available today are based on the Granger causality concept [16]. Given two time series, one of them Granger-causes the other one if the information provided by it helps in forecasting the second series. A generalized version of the above statement is that, if the time series generated by one process provides us with information on the time series generated by another process at some point in the future, the first process influences the second process. If only two processes are involved and coupling is detected exclusively in one direction it is inferred that the first process causally influences the second process. Granger has applied his principle to coupled linear models [16]. In Ref. [21] it was demonstrated that using changes in cross-prediction errors to indicate the directionality of coupling is not trivially extensible to nonlinear systems. For nonlinear systems, methods based on information theory have been shown to be widely applicable, especially when the estimators of the relevant information theoretic functionals are nonparametric and thus applicable to any PDFs (usually under some mild technical assumptions).

Let X and Y denote two stationary ergodic processes and $x(t)$ and $y(t)$ their time series. The presented method of detecting coupling directionality uses conditional mutual information as an indicator of the presence of net information flow [22] between the two analyzed systems characterized by their respective time series. The net information flow $I(X;\Delta_\tau Y|Y)$, where $\Delta_\tau Y$ is an observable derived from the state of the process Y τ steps in the future, is defined as the mutual information between X, Y, and $\Delta_\tau Y$ that is not a result of the action of the history of process Y on itself, i.e., excluding $I(Y;\Delta_\tau Y)$, and is also not the result of the common history of the two processes captured by $I(X;Y)$. A statistically significant information flow thus indicates that information is being transferred from the process X to the process Y at some later point in time. This can be readily interpreted as an influence of the process X on the process Y in the future. The detection criterion is based on two indices

$$i_{X\to Y} = \frac{1}{N}\sum_{\tau=1}^{N} I(x(t);\Delta_\tau y(t)|y(t)), \tag{7}$$

$$i_{Y\to X} = \frac{1}{N}\sum_{\tau=1}^{N} I(y(t);\Delta_\tau x(t)|x(t)), \tag{8}$$

where the notation

$$I(x(t);\Delta_\tau y(t)|y(t))$$

denotes mutual information between $x(t)$ and $\Delta_\tau y(t)$ conditioned on $y(t)$. The operator Δ_τ represents the difference

$$\Delta_\tau x(t) = x(t+\tau) - x(t).$$

The series $x(t)$ and $y(t)$ can contain the values generated by the respective systems or values which have been derived from the original time series. In general, each of the time series $x(t)$ and $y(t)$ should be considered multivalued. This is the case if state space reconstruction techniques [23] are applied prior to computing the information theoretic functionals [21].

C. Exploiting phase dynamics

In the special case where the systems generating the time series $x(t)$ and $y(t)$ can be considered oscillators, it is in many cases advantageous to extract the phase of each of the systems. The phase of a chaotic system is in general not uniquely defined [24], and several methods of extracting a phase exist. If phases form the basis for further analysis, it is usually not necessary to perform state space reconstruction of the observables: univariate series effectively describe the state of each system [21]. The series $x(t)$ and $y(t)$ in Eqs. (7) and (8) are univariate and the Δ_τ operator is the phase difference. In this paper the "wrapped" definition of the phase is used so that $x(t), y(t) \in \langle 0, 2\pi)$. In the following sections, we describe some frequently used methods for obtaining the phase from univariate signals.

1. Hilbert transform

A commonly used approach involves the Hilbert transform, which extracts the phase of the system using the analytical signal concept [25],

$$\zeta(t) = s(t) + j\tilde{s}(t) = Ae^{j\phi(t)},$$

for which the phase can be unambiguously defined as $\phi(t)$. The function $\tilde{s}(t)$ is the Hilbert transform of $s(t)$,

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

The method has been applied directly to signals with broadband spectra [26]; however the physical meaning of the time series thus extracted is unclear [27]. The phase extracted using Hilbert transform can be readily interpreted if the source system has a clear main frequency of oscillation. Alternatively, the signal can be processed with a narrowband filter, but care must be taken not to disrupt the phase spectrum of the signal. This can be difficult in on-line or real-time settings, as common finite impulse response filters have a frequency-dependent phase shift (e.g., Butterworth filters have a phase shift that is linearly dependent on the frequency). The Hilbert transform can be computed efficiently in the frequency domain by manipulating the phases of the Fourier transform of the signal. Alternatively, the computation can be done in the time domain as a convolution

$$\tilde{s}(t) = s(t) \otimes \frac{1}{\pi t},$$

also yielding the complex part of the analytical signal.

2. Wavelet transform

Another method consists of extracting the complex part of the analytical signal by convolving a (possibly broadband) signal with a wavelet such as the complex Gabor wavelet [27–29], defined as

$$G(t,f) = \exp\left(\frac{t^2}{2\sigma_t^2}\right)\exp(j2\pi f t),$$

where σ_t defines the spread of the wavelet and f defines its frequency. Some authors recommend selecting σ_t based on the frequency, e.g., as $\sigma_t = 7/f$ [28], to obtain a fixed number of periods of the signal at the given frequency. Convolving the filtered signal with the complex Gabor wavelet above yields the complex part of the analytical signal similarly to the Hilbert transform approach,

$$\tilde{s}(t) = s(t) \otimes G(t,f).$$

Le Van Quyen *et al.* [27] have shown that in many cases both of the above phase extraction methods provide similar results. The phase series extracted using the analytical signal concept is sampled at the same frequency as the original signal; this is sometimes called the *instantaneous phase*.

3. Marked-events phase

A completely different method which provides information on the phase at discrete time instants is the marked-events method [1,7]. The marked-events method requires that an event which occurs once per period at a well-defined time instant is detected with the highest possible precision. The phase is linearly interpolated between each two neighboring detections:

$$\Phi(t) = 2\pi \frac{t - t_k}{t_{k+1} - t_k},$$

where $t_1, t_2, \ldots$ is the sequence of time instants when the selected event has been detected. Such an event may be, for example, the R peak of an electrocardiogram (ECG signal), which can be detected with high temporal precision. The marked-events phase is also useful if other definitions of phase prove difficult to apply (e.g., when analyzing a neural spike train).

D. Testing the significance

The indices $i_{X\to Y}$ and $i_{Y\to X}$ are generally difficult to interpret without additional information about the underlying systems. For structurally similar systems, it is possible to compare their numerical values; however, the indices may exhibit significant bias that is different in each direction if the generating systems vary widely in their dynamics or in some system parameter such as the main oscillation frequency [21].

A possible solution to this problem is provided by testing the computed indices using surrogate data sets [19,20,30]. Surrogate testing has been used in detecting nonlinearity [31] and synchronization [32]. A surrogate data set is a set of synthetically generated time series which ideally preserve all the properties of the underlying systems except the one that is being tested. It is usually a nontrivial task to generate surrogate data for a particular detection problem as the details of the procedure depend on the type of data available and on the detection problem. When testing for directionality, the surrogates should replicate the dynamics of both systems but break any existing coupling between them. The main technique applied today to reach this effect is altering some of the temporal structures of the time series so that (if the systems are coupled) cause and effect are separated.

Irrespective of the particular surrogate generation method, the significance test is a standard one-sided hypothesis test. The null hypothesis is that the two systems are independent and by using evidence in the time series an attempt is made to reject this hypothesis. Assuming that a bivariate time series is available, the test is performed as follows. The indices $i_{X\to Y}$ and $i_{Y\to X}$ are evaluated on a predetermined number of surrogate data sets generated from the bivariate time series, whence an estimate of the cumulative distribution function (CDF) of each index can be obtained. This distribution represents the variability of the directionality indices $i_{X\to Y}$ and $i_{Y\to X}$ for independent systems (i.e., under the null hypothesis). For an *a priori* selected level of significance, typically 5%, it is possible to obtain critical values for the indices from both of the CDFs. The critical values are used as a limit in a one-sided test for significance. If the value of the index obtained from the original (possibly coupled) bivariate time series is above this critical value, it is significant at the chosen significance level, and directional influence in the direction corresponding to the index has been detected. If the value of the index is not significant, then the evidence in the time series does not support rejecting the null hypothesis of no coupling in the given direction.

If the generating systems do not vary with time, it is possible to compute thresholds from the surrogates which ensure detection of directional influence at given levels of significance. Some frequently used surrogate generation methods are described in the following sections.

1. Fourier transform surrogates

Fourier transform surrogates are constructed by computing the Fourier transform of each signal, randomizing the phase of each frequency component (except the dc component), and then taking the inverse transform. In practice this is accomplished by computing the fast Fourier transform (FFT) [33] of the time series and there are many software packages available that provide required procedures. The autocorrelation function and spectrum are preserved but the distribution of amplitudes is usually slightly flattened [20]. Further improvements on the basic method try to iteratively converge upon surrogates that more accurately match the original frequency spectrum and the original amplitude distribution at the same time [19].

2. Permutation surrogates

Permutation surrogates have been previously applied to time series obtained using the marked-events method [7,34]. The sequence of durations (intervals) of each cycle of the time series is first extracted. The intervals are then randomly reordered and a new phase time series is constructed based on the reordered sequence. The process is repeated until the desired number of surrogate time series is obtained.

For instantaneous phase, essentially the same procedure applies but care must be taken to preserve the intraperiod information in the signal. The signal is first divided into periods and all the periods except the first and last are permuted randomly. The first and last period are not displaced, because this would create discontinuities in the surrogate signal, as usually only a part of the first and last period is available. The distribution of phases is preserved when this method is applied.

This procedure has been applied in the present work as the comparison includes methods which are dependent on the distances between data points, so that it is important to preserve the distribution of phases as accurately as possible. We note that permutation surrogates differ from the "white noise" or "scrambled" surrogates where the time-series samples are simply reshuffled randomly, yielding a time series which preserves the distribution but completely destroys any correlations in the time series.

3. Twin surrogates

Recently the technique of analyzing nonlinear dynamical systems using recurrence plots [35,36] has gained some popularity. Recurrence plots (RPs) indicate instances of time when the trajectory of the dynamical system passes close to a point it has visited previously. A recurrence plot of a process can be used to construct *twin surrogates* [37] by exploiting special points that are equivalent with respect to the recurrence plot. A twin surrogate is a time series which is constructed by reordering segments of the original time series so that the reordering does not alter the recurrence plot. Since several important quantities are derivable from recurrence plots, the equivalence of recurrence plots ensures that these quantities remain unchanged between the original data and the surrogate (up to a precision dependent on the resolution of the recurrence plot and the amount of data). The application of recurrence plots to univariate time series however requires that a state space reconstruction technique such as time-delay embedding [23] is applied prior to the construction of a RP.

II. NONPARAMETRIC ESTIMATION OF CONDITIONAL MUTUAL INFORMATION FUNCTIONALS

There are several approaches to the estimation of conditional mutual information. It is possible to decompose the problem into multiple problems of estimating mutual information and compute the result as in Eq. (5) or into multiple entropy estimations and combine these using Eq. (6). Both mutual information and entropy are functionals of the PDF of the multivariate signal. Efficient estimation of the PDF from the time series is the main problem in estimating information theoretic quantities.

In this paper, we compare two basic classes of nonparametric methods for the estimation of conditional mutual information between pairs of time series. The first class can be best described as *binning methods*, as these methods discretize the space into regions usually called bins or boxes. Binning methods compute a coarse global profile of the PDF which is usually composed of regions of constant values. The second class of methods could be described as *metric methods*, because their PDF estimation algorithms depend exclusively on distances between samples computed using some metric. In contrast to the binning methods, metric methods compute a local approximation of the PDF on a set of points in the sample space.

A. Binning methods

Binning methods employ a scheme that divides the space into regions usually called bins. The simplest possible discretization scheme is *equidistant binning* [38] which splits the sample space into bins of equal size regardless of the distribution of the data. This is also called the histogram method. Alternative methods respecting the distribution of the sample data have been proposed, such as hierarchical subdivision schemes as employed by Fraser and Swinney [39] and later Darbellay and Vajda [40,41], and direct division schemes such as *equiquantal binning* [42,43].

1. Binning using B-spline functions

Daub *et al.* [44] have presented a method for estimating multidimensional entropies. Estimation of multidimensional entropies constitutes sufficient means to indirectly compute mutual information with Eq. (3) and also the conditional mutual information using Eq. (6). In classical binning approaches to computing mutual information, data points close to bin boundaries can cross over to a neighboring bin due to noise or fluctuations, thus introducing additional variance

into the computed estimate. To overcome this problem, Daub *et al.* have proposed a generalized binning method, which makes use of *B*-spline functions to assign data points to bins. The sample space is divided into equally sized bins as in equidistant binning. A major difference between classical binning and generalized binning is that in generalized binning, a data point is assigned to multiple bins simultaneously with weights given by (implicitly normalized) *B*-spline functions. The shape of the *B*-spline functions is determined by their order k, which is a parameter of the method. With *B*-spline order 1, each point is assigned to exactly one bin and the method is equivalent to simple equidistant binning. The proposed method is thus a fixed binning scheme extended with a preprocessing step designed to reduce the variance.

A *B*-spline function is defined with the help of a knot vector

$$t_i = \begin{cases} 0, & i < k, \\ i - k + 1, & k \le i \le M - 1, \\ M - 1 - k + 2, & i > M - 1, \end{cases}$$

where M is the total number of bins and i is an index into the knot vector. *B*-spline functions are defined (and evaluated) recursively [45] by

$$B_{i,1}(z) = \begin{cases} 1, & t_i < z < t_{i+1}, \\ 0 & \text{otherwise}, \end{cases}$$

$$B_{i,k}(z) = B_{i,k-1}(z)\frac{z - t_i}{t_{i+k-1} - t_i} + B_{i+1,k-1}\frac{t_{i+k} - z}{t_{i+k} - t_{i+1}}.$$

The standard definition for computing entropy (1) is used with

$$p(x_i) = \frac{1}{N}\sum_{j=1}^{N} B_{i,k}(f_{M,k}(x_j)), \tag{9}$$

where $f_{M,k}(x)$ is a linear transformation which maps the values of x onto the domain of the *B*-spline functions [44]. In two dimensions it is necessary to compute the joint PDF [44]

$$p(x_i, y_j) = \frac{1}{N}\sum_{l=1}^{N} B_{i,k}(f_x(x_l))B_{j,k}(f_y(y_l)). \tag{10}$$

This procedure can be readily generalized to three dimensions

$$p(x_i, y_j, z_l) = \frac{1}{N}\sum_{m=1}^{N} B_{i,k}(f_x(x_m))B_{j,k}(f_y(y_m))B_{l,k}(f_z(z_m)). \tag{11}$$

The computation of $I(X;Y|Z)$ can thus be written as

$$I_{M,k}(X;Y|Z) = H_{M,k}(X,Y) + H_{M,k}(X,Z) - H_{M,k}(Z) - H_{M,k}(X,Y,Z),$$

and each of the terms may be computed using the formulas (9)–(11) together with Eqs. (1) and (2). The notation $I_{M,k}(X;Y|Z)$ and $H_{M,k}(X,Y)$ indicates that the method has two parameters M, the number of bins, and k, the order of the *B* spline. For further analysis, the order of the *B* splines is fixed at $k=3$ as this was the order employed in Ref. [44].

2. Marginal equiquantization

Marginal equiquantization is an estimation method based on binning. The series are sorted in order of magnitude and divided evenly into a predefined number of bins. This defines the mapping to bins for each individual series. Henceforth the algorithm proceeds in exactly the same manner as for standard equidistant binning. The two- and three-dimensional bins are populated based on the bin assignments in each series. This method has been previously applied by Paluš [4,11,12,21,22,31,32,42,43].

An alternative is provided by hierarchical subdivision schemes such as that by Darbellay *et al.* [40,41] based on the work of Fraser and Swinney [39]. Adaptive schemes reduce the bias of PDF estimation because they adjust to the local density and offer a better approximation of the PDF using constant regions. However, depending on the particular method used in each scheme, the variance of these methods may be higher because of the additional number of degrees of freedom in the adaptive subdivision mechanisms. Numerical tests have confirmed that the above method produces estimates of mutual information with a very small bias. However, in a recent report by Celucci *et al.* [46], the Fraser and Swinney hierarchical subdivision scheme (essentially the same as used in Refs. [40,41]) has been shown to detect nonexistent structures when only a small amount of data points is available. In practice, it is imperative that good estimates of the CMI functional are provided with as little data as possible. For this reason, hierarchical subdivision schemes have not been included in the testing process.

B. Metric methods

The second type of method used to estimate CMI in this paper are *metric* methods. These methods require that the sample space be endowed with a metric. The algorithms depend exclusively on distances between points.

Nonparametric estimation of a local PDF in a metric space can be accomplished in two basic ways. One can begin by fixing a volume of a given shape (usually a hypersphere) around the point, where the PDF is to be estimated, and computing the number of points inside this volume. The other possibility consists of fixing the number of points and enclosing them in a hypersphere of minimum volume. The local PDF is estimated as the number of points divided by the enclosing volume in both cases.

1. k-nearest-neighbor method

The k-nearest-neighbor (kNN) method combines both approaches to estimating the local PDF. This estimate of mutual information is based on the original work by Kozachenko and Leonenko [47], who have derived an estimate of entropy based on nearest-neighbor distances in the sample space,

$$H(x) = \sum_{j=1}^{N} (N\rho_j) + \ln 2 + C_E,$$

where ρ_j is the distance from the jth sample to its nearest neighbor, N is the number of points, and C_E is the Euler-Mascheroni constant ($C_E \approx 0.5772$). Kozachenko and Leonenko have also proven the consistency of their estimator under some technical assumptions.

Kraskov *et al.* [48] were inspired by the Kozachenko-Leonenko estimate for entropy and derived a method for computing mutual information using kth-nearest-neighbor estimates. kth-nearest-neighbor distances exhibit better statistical properties than nearest-neighbor distances, and recently it has been proved by Goria *et al.* [49] that also the estimator of entropy based on kth-nearest-neighbor distances is consistent. Kraskov *et al.* [48] presented an estimator of mutual information and conducted numerical experiments which have indicated the lack of bias of the estimator for independent data even for small data sets. The estimator $I^{(1)}(X;Y)$ may be written as

$$I^{(1)}(X;Y) = \psi(k) + \psi(N) - \langle \psi(n_x + 1) + \psi(n_y + 1) \rangle, \quad (12)$$

where k is the rank of the neighbor used in distance computations and $\psi(\cdot)$ is the digamma function. The estimation algorithm involves iterating through the time series and for each point of the joint space the distance ϵ_k to its kth nearest neighbor is found. The subspaces X and Y are searched and the number of points in the hypersphere of radius ϵ_k around the current point is found. These counts are denoted as n_x and n_y respectively. The symbol $\langle \cdots \rangle$ denotes an average over all available data points. Any metric may be selected to evaluate the distances. For practical reasons, the maximum norm is the most frequently used metric.

Using the above formula, it is possible to compute the conditional mutual information using a decomposition into three mutual information computations. However, as Kraskov *et al.* state in their work [48], significant bias is incurred if different length scales are used in the computation of mutual information according to Eq. (5). Experiments in estimating conditional mutual information using the decomposition into multiple mutual information estimation problems have confirmed this expectation. The bias is caused by the isolation of the three mutual information computations in which, inevitably, different length scales in different spaces come into play. It is much more convenient to derive a new estimator similar to the formula (12), which estimates conditional mutual information directly using a single length scale. Expanding on the thought process of the authors in Ref. [48], we have derived a new estimator of conditional mutual information based on Eq. (12):

$$I(X;Y|Z) = \psi(k) - \langle \psi(n_{xz} + 1) + \psi(n_{yz} + 1) - \psi(n_z + 1) \rangle,$$

where n_A is the number of neighbors of the reference point in the space A with distance smaller than the distance to the kth neighbor of the given reference point in the joint space (X,Y,Z). It is generally not necessary that the metric is the same in all of the spaces, but we have elected to use the maximum norm in all cases.

2. Cross redundancies based on correlation integrals

The term redundancy is frequently used as a synonym for multidimensional mutual information in the context of dynamical systems. Redundancies based on correlation integrals (CIs) use a fixed volume approach to estimating the local PDF. The connection between the correlation integral and local probabilities is elucidated in Ref. [50]. The correlation integral has been introduced by Grassberger and Procaccia [51] in the context of estimating the correlation dimension of strange attractors. The correlation integral is given by

$$C(\mathbf{x}, \epsilon) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \Theta(\epsilon - \|x_i - x_j\|),$$

where $\|\cdot\|$ represents the selected metric, $\Theta(\cdot)$ is the Heaviside function, and ϵ is the radius of the hypersphere in which neighbors are sought. Savit *et al.* [52] show how it is possible to compute the conditional redundancy $R(x_1; x_m | x_2, \ldots, x_{m-1}, \epsilon)$ by quantifying the dependency between x_1 and x_m conditioned on $x_2, \ldots, x_{m-1}$:

$$R(x_1; x_m | x_2, \ldots, x_{m-1}, \epsilon) = -\log \frac{C(x_1, \ldots, x_{m-1}, \epsilon) C(x_2, \ldots, x_m, \epsilon)}{C(x_1, \ldots, x_m, \epsilon) C(x_2, \ldots, x_{m-1}, \epsilon)}.$$

Inspired by this approach, Prichard and Theiler [50] have extended the concept of conditional redundancies to multiple variables and propose computing the time-lagged mutual information between two time series with

$$I(x; y, l, \epsilon) = H(x(t), \epsilon) + H(y(t-l), \epsilon) - H(x(t), y(t-l), \epsilon),$$

where the Shannon entropy $H(x, \epsilon) \approx -\log[C(x, \epsilon)]$. In this work, the conditional mutual information is estimated using the formula

$$I(x; y, \tau, \epsilon) = I(x(t); \Delta_\tau y(t) | y(t), \epsilon) = -\log \frac{C(x(t), y(t), \epsilon) C(\Delta_\tau y(t), y(t), \epsilon)}{C(x(t), \Delta_\tau y(t), y(t), \epsilon) C(y(t), \epsilon)}.$$

It has been shown that the estimate of the mutual information using correlation integrals converges if $\epsilon \to 0$ [50]. However, in data sets of finite size, small sample effects are observed, which disrupt the convergence behavior of the estimator.

III. ESTIMATOR BIAS

In this section the estimators of conditional mutual information are evaluated using a linear system for which an analytical estimate of the conditional mutual information with negligible bias (compared to the tested estimators) is available. The convergence properties of each method are investigated.

The experiment tests the estimators using a stationary linear dynamical model with linear coupling: two coupled copies of the autoregressive moving average (ARMA) part of the Barnes sunspot model [53] defined by the equations

$$z_i^{[1]} = \alpha_1 z_{i-1}^{[1]} + \alpha_2 z_{i-2}^{[1]} + a_i^{[1]} - \beta_1 a_{i-1}^{[1]} - \beta_2 a_{i-2}^{[1]},$$

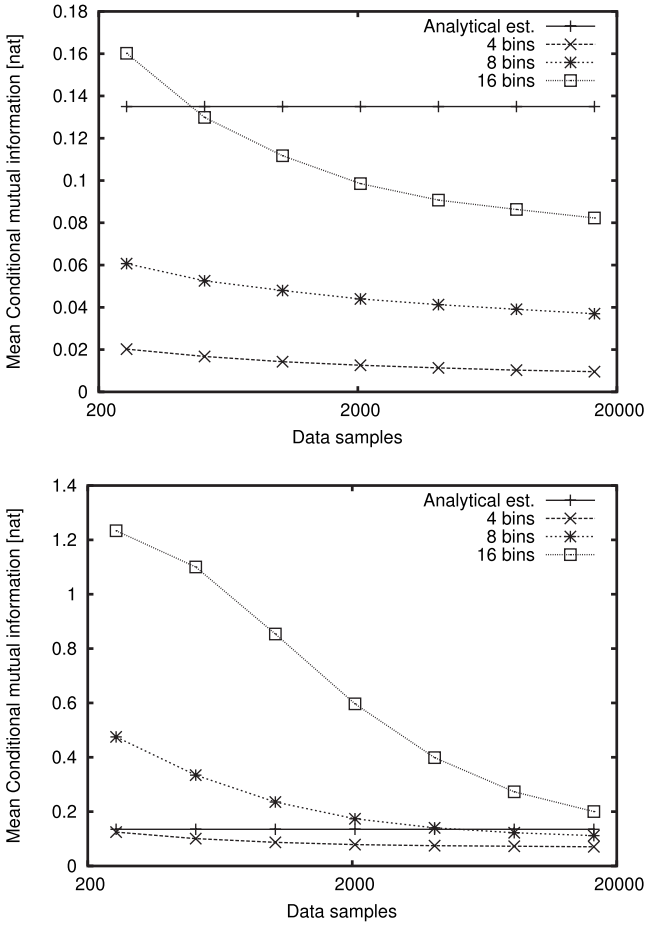

FIG. 1. Dependence of the mean conditional mutual information estimate by the B-spline method (top) and by the equiquantal estimator (bottom) on the number of samples in the time series for different numbers of bins.

$$z_i^{[2]} = \alpha_1[\epsilon z_{i-1}^{[1]} + (1-\epsilon)z_{i-1}^{[2]}] + \alpha_2 z_{i-2}^{[2]} + a_i^{[2]} - \beta_1 a_{i-1}^{[2]} - \beta_2 a_{i-2}^{[2]},$$

where $\alpha_1 = 1.906\,93$, $\alpha_2 = -0.987\,51$, $\beta_1 = 0.785\,12$, $\beta_2 = -0.406\,62$, and $a_i^{[j]}$ are independent and identically distributed Gaussian random variables with zero mean and standard deviation 0.4. The conditional mutual information $I(z^{[1]}; z_T^{[2]} | z^{[2]})$ and $I(z^{[2]}; z_T^{[1]} | z^{[1]})$ can be computed analytically [43], where $z_T^{[i]}$ denotes the time series $z^{[i]}$ shifted by T samples into the future. For a set of variables $X_1, X_2, \ldots, X_n$ having zero mean, unit variance, and correlation matrix C, their mutual information can be computed as

$$I(X_1, X_2, \ldots, X_N) = -\frac{1}{2}\sum_{i=1}^{N} \log(\sigma_i),$$

where σ_i are eigenvalues of the correlation matrix C. The conditional mutual information can be computed using mutual information according to Eq. (5). Long time series ($>10^6$ samples) were used to compute the analytical estimates ensuring that they have a negligible bias and variance with respect to the nonparametric methods.

One hundred realizations of the above model have been generated and the estimators of conditional mutual information presented above have been applied to each realization in turn. The results of the analysis are shown in Figs. 1–3. The numerical results are interpreted in the context of the theoretical expectations of each estimator. In this experiment the estimators were applied directly to the time series normalized to zero mean and unit variance.

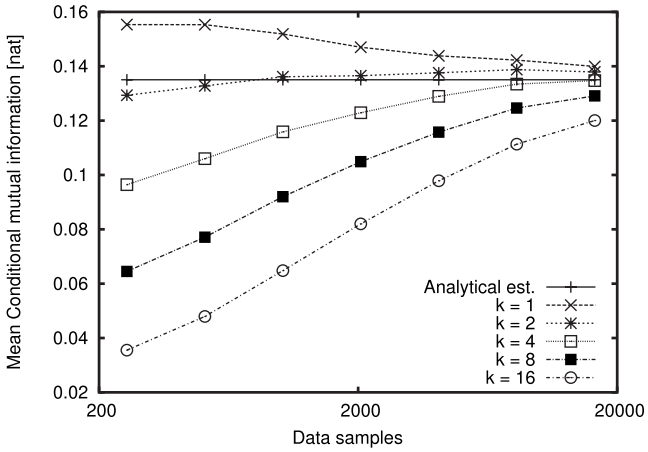

FIG. 2. Dependence of the mean conditional mutual information estimated with the k-nearest-neighbor estimator on the number of samples in the time series for different k.

1. Binning estimators

There are no consistency results for the B-spline estimator; however, Fig. 1 shows that the estimators converge to some value dependent on the free parameter—the number of bins. The theoretical result for classical histogram binning (the B-spline estimator with splines of order 1 is equivalent to the histogram estimator) states that the histogram estimator converges to the true PDF if

$$N \to \infty \quad \text{and} \quad h \to 0 \quad \text{with} \quad hN \to \infty, \tag{13}$$

where h is the width of each bin [54]. The B-spline estimator does not include any additional mechanisms to improve con-

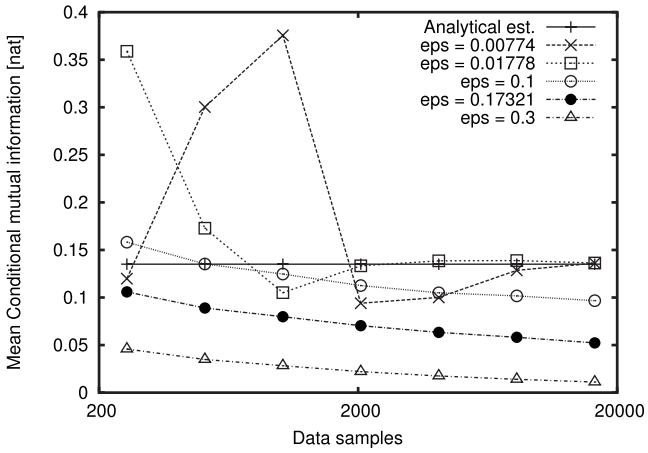

FIG. 3. Dependence of the mean conditional mutual information estimated with the cross-redundancies estimator on the number of samples in the time series for different neighborhood sizes.

vergence and a better result cannot be expected. The experiments also indicate that in practical settings one cannot expect to converge to the true value of the conditional mutual information regardless of the length of the data set provided. The number of bins would have to be adjusted with respect to the length of the data set and the nature of the series. Figure 1 shows the convergence behavior for the B-spline estimator and for the equiquantal estimator. The equiquantal estimator displays similar convergence behavior (with respect to the number of bins) although the absolute values differ from those of the B-spline estimator.

2. k nearest neighbors

The kNN estimator of mutual information is consistent and should asymptotically converge to the true value of the mutual information regardless of the value of k [49]. The conditional mutual information estimator presented in this paper is based on the same theoretical assumptions, and it was expected that the estimator would also exhibit convergence in the experiments. The results of the experiment on the Barnes model are in agreement with this hypothesis and the convergence behavior of the kNN CMI estimator is clearly seen in Fig. 2.

3. Cross redundancies

The estimate of mutual information using the correlation integral as defined by Grassberger and Procaccia [51] is known to be consistent and converges to the true value of mutual information if the size of the neighborhood $\epsilon \to 0$ [50]. The estimate of conditional mutual information seems to behave in a similar fashion, as shown by the results on the Barnes model computed using the cross-redundancy method. In general larger neighborhoods increase negative bias, as seen in Fig. 3. For short time series, however, small neighborhoods do not provide a good estimate as the variance of the estimator is high due to the sparsity of samples.

IV. EXPERIMENTS IN DETECTING DIRECTIONALITY OF COUPLING

The subsequent sections describe numerical experiments with a unidirectionally coupled pair of chaotic Rössler oscillators. Four distinct cases are investigated. For each case, 1000 realizations and 1000 permutation surrogates of the coupled systems were generated and analyzed to obtain dependable statistics.

The first experiment shows the effectiveness of each estimator on a pair of Rössler oscillators with a small frequency mismatch. Phase is extracted by means of the Hilbert transform. The second experiment simulates the marked-events phase extraction method by destroying the intrawave information in the instantaneous phase time series from the first experiment. In this way, the effect of a reduction of the amount of information in the time series can be investigated. In the third experiment the relative frequencies of the systems were changed to 5:1 (faster system drives slower system) and finally in the fourth experiment to 1:5 (slower system drives faster system).

TABLE I. Rössler 1:1 experiment: The first column in the table identifies the applied method, the second column the range of parameters providing good results, the third column the smallest amount of samples from which more than 95% true positives and fewer than 5% false positives were obtained, and the fourth column shows the number of false positives (FP).

Method	Parameter	Min. pts.	FP
B spline	Order 3, eight bins	512	≈ 0
Eqiquantal binning	Eight bins	512	≈ 0
Cross redundancies	$\epsilon \in \langle 0.22, 0.3 \rangle$	512	≈ 0
kNN	$k \in \{8, 16, 32, 64\}$	1024	≈ 0

Each of the estimators has one free parameter; however, the binning estimators were tested only with eight bins as this number of bins was already used in previous work [21] and is known to be effective for a wide range of systems. This has also reduced the amount of computations performed. The cross-redundancy estimator was tested with neighborhood size varying from 0.007 to 0.3 in ten logarithmically spaced steps. The maximum neighborhood size was fixed to this value because most spatial indexing algorithms lose their effectiveness if large hyperspheres must be scanned for neighbors, thereby rendering the correlation integral expensive to compute. The kNN estimator was tested with the number of neighbors from 1 to 64 in powers of two, as higher neighbor counts increase the computational cost considerably.

The coupled Rössler systems used in this study can be described by the following set of equations:

$$\dot{x}_{1,2} = -\omega_{1,2} y_{1,2} - z_{1,2} + \epsilon_{1,2}(x_{2,1} - x_{1,2}),$$

$$\dot{y}_{1,2} = \omega_{1,2} x_{1,2} + a_{1,2} y_{1,2},$$

$$\dot{z}_{1,2} = b + z_{1,2}(x_{1,2} - c), \tag{14}$$

where $\epsilon_1 = 0$, $b = 0.2$, and $c = 10$. The parameters $\omega_{1,2}$, $a_{1,2}$, and ϵ_2 are adjusted in the experiments.

A. Coupled 1:1 Rössler models: Instantaneous phase

The first numerical experiment was performed with both oscillators having very close frequencies $\omega_{1,2} = 1 \pm 0.015$. 1000 realizations with $a_{1,2} = 0.15$ and $\epsilon_2 = 0.05$ were generated using Eqs. (14) by sampling the systems with about 20 samples per period. The phase time series were extracted from the x_1 and x_2 variables using the Hilbert transform approach by convolving the original series in the time domain with the Hilbert transform impulse response to obtain the complex part of the analytical signal. The convolution window covered approximately ten periods of the signal. The significance of the results has been tested using a distribution obtained from permutation surrogates where both of the phase time series were subjected to the permutation procedure.

The results of the experiment are summarized in Table I. The summary shows that all of the methods perform rather

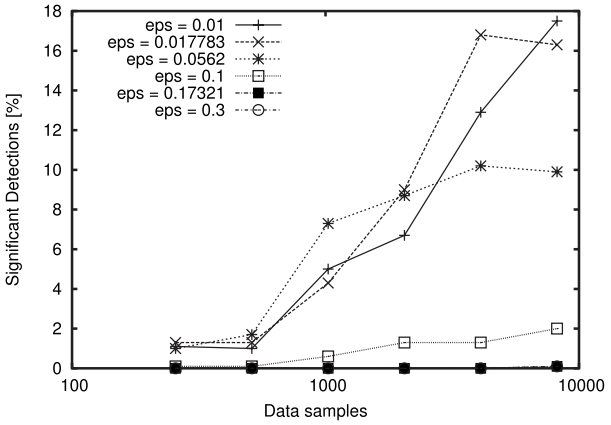

FIG. 4. Rössler 1:1 experiment. Dependence of the false positive rate of the cross-redundancy method for various neighbor sizes. Only $\epsilon \geq 0.1$ yielded acceptable false positive levels in the experiment.

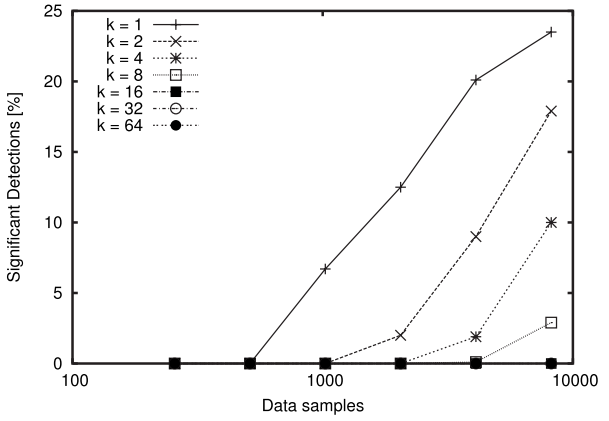

FIG. 5. Rössler 1:1 experiment. Dependence of the false positive rate of the kNN method for various nearest-neighbor counts. The false positive rate is zero for $k \in \{16, 32, 64\}$.

well as the required detection rate of 95% in the direction of coupling is obtained for all methods for short time series (512 points correspond to ≈25 periods) except for the kNN method, which required 1024 points to reach the same detection rates.

The B-spline and equiquantal binning estimators are very effective on the basic version of the coupled Rössler oscillators. There are almost no false positives for any of the methods. This confirms the results in Ref. [21] where the authors used equiquantal binning with Fourier transform surrogates on the same Rössler pair and obtained no false positives. The cross-redundancy estimator seems to perform better when larger neighborhood sizes are used during the estimation of the correlation integrals. Choosing a neighborhood size too small increases the number of false positives beyond acceptable levels quickly, as seen in Fig. 4. With more data available, the width of the distribution of the permutation surrogates drops faster than that of the distribution of the coupled systems. This behavior may be caused by the fact that in long time series intercycle (long-term) dependencies become more important, and permutation surrogates in their basic version do not preserve dependencies across period boundaries. Under these conditions more sophisticated surrogate generation methods may be necessary to adequately represent the distribution under the null hypothesis.

In the case of the kNN estimator, increasing the number of neighbors increases the specificity at the cost of decreasing the sensitivity (cf. Fig. 5). For $k \geq 8$ the false positives drop below the required level of 5%; however, the sensitivity of the detector is decreased as the number of neighbors grows, and this is the reason that the required true positive rate is reached at 1024 points. An analysis of the histograms of the distributions for mean CMI in the uncoupled direction arising from the data set and from the surrogate data set show that the reason for higher false positives with small k is that the variance of the CMI of the surrogates drops faster than the variance of the CMI estimates from the data. This is similar to the situation of the cross-redundancy estimator. Use of a higher number of neighbors causes the variances of CMI on the data set and on the surrogate set to equalize. However, a higher number of neighbors causes the CMI estimates to be negatively biased, and thus the true positive rate (sensitivity) of the method decreases as a result. This negative bias can also be observed in Fig. 2, where the curves for larger k converge to the analytical estimate from below.

Summarizing the results, all of the methods perform well on the basic version of linearly coupled Rössler systems with permutation surrogates and correctly detect the directionality for short time series. However, there are some caveats when applying metric methods: care must be taken when choosing the value of the free parameter. The binning estimators with the default number of bins provide excellent results on the given data set.

B. Coupled 1:1 Rössler models: Marked-events phase

In this section, we repeat the previous experiment but the intrawave information is first destroyed by interpolating the phase linearly between period limits. Surrogates were generated in the same manner as for the above experiment. This experiment setup mimics marked-events phase extraction while allowing us to find how a reduction in information content of the phase time series affects the detection statistics.

The results in Table II surprisingly show that destroying

TABLE II. Rössler 1:1 marked-events phase: The first column in the table identifies the applied method, the second column the range of parameters providing good results, the third column the smallest amount of samples from which more than 95% true positives and fewer than 5% false positives were obtained, and the fourth column shows the number of false positives.

Method	Parameter	Min. pts.	FP
B spline	Order 3, eight bins	512	≈0
Eqiquantal binning	Eight bins	512	≈0
Cross redundancies	$\epsilon \in \langle 0.22, 0.3 \rangle$	512	≈0
kNN	$k \in \{32, 64\}$	1024	≈0

the intrawave information in the phase time series does not affect the detection methods in any significant way. The *k*NN estimator is affected by a quantization effect: the distribution of phases and phase differences is heavily discretized (because of the low sampling rate of 20 points per period). This has been anticipated by Kraskov *et al.* [48], who recommend using a higher number of neighbors to resolve this problem or adding low-amplitude noise to the time series. Adding low-amplitude noise is a potentially dangerous practice with very short time series as it may create spurious structures not inherent in the data. However, as is seen from the results of the experiment, the *k*NN estimator succeeds in detecting the directionality correctly with a higher number of neighbors.

C. Coupled 1:5 Rössler models: Instantaneous phase

In the next experiment, a pair of Rössler oscillators was set up to have a substantial relative ratio of frequencies (1:5) to investigate its effect on the quality of directionality detection. The parameters of the coupled oscillator pair (14) were set to: $\omega_1=0.5$, $\omega_2=2.515$, $a_1=0.15$, $a_2=0.72$, and $\epsilon_2=0.1$. The sampling frequency was changed so that the faster system was again sampled at approximately 20 samples per period and phase was extracted using the Hilbert transform.

Using permutation surrogates in the same way as for the experiments above produced directionality estimates with a high rate of false positives for all of the tested CMI estimation algorithms. An analysis of the distributions of CMI shows that randomization of the phase time series corresponding to the fast system causes a significant reduction in variance of the surrogates relative to the variance of the data. This is clearly seen in the histogram for the *k*NN estimator in Fig. 6. The histograms for the all of the other methods are similar and display the same effect. We note that this happens only for the estimates in the uncoupled direction. In the direction of coupling, all three possibilities produce very similar distributions of CMI in the surrogate data set. In light of this evidence, it is recommended to use a more elaborate surrogate generation scheme that preserves dependency structures longer than one period or to leave the time series of the fast system unchanged when constructing surrogates.

Binning methods are able to correctly determine the directionality even when the relative frequencies are unequal, but it is clearly seen that detection is much more difficult than in the experiments with comparable frequencies. None of the methods can correctly detect the directionality for time series shorter than 8192 samples (approximately 220 periods of the faster system).

The parameters of the metric methods can be set to provide results on par with those of the binning methods; however, the space of parameters where they provide good results is reduced compared to the experiments with comparable frequencies. The *k*NN method loses some of its sensitivity already with $k=16$ neighbors, and does not satisfy the requirements of the hypothesis test as the true positive rate drops below 95% for the investigated data sample sizes (see Table III).

D. Coupled 5:1 Rössler models: Instantaneous phase

The last numerical experiment tests the detection of directionality when the direction of coupling is from the faster system to the slower system. The parameters of the coupled oscillator pair (14) were set to $\omega_1=2.515$, $\omega_2=0.5$, $a_1=0.72$, $a_2=0.15$, and $\epsilon_2=0.1$. The surrogate data sets were created by applying the permutation procedure to the slower system as in the last experiment.

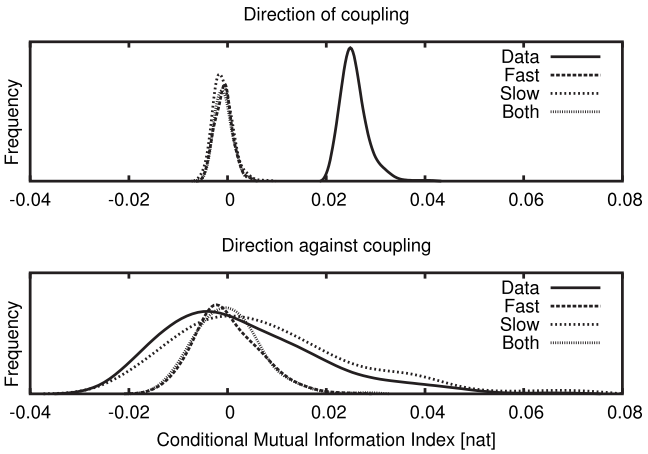

FIG. 6. Rössler 5:1 experiment. Histogram of conditional mutual information estimates for the model data and for the permutation surrogates where either one of the time series or both of them were randomized. The coupling is unidirectional from the faster system to the slower system. In the direction of coupling all the surrogates generate very similar distributions of the CMI index (top). However, in the reverse direction when the fast system is also randomized, the width of the distribution of the surrogates is clearly smaller than that of the distribution arising from the original data (bottom).

Table IV now shows a clear difference between methods dependent on distances and binning methods. Metric methods require many more data points to satisfy the requirements of the hypothesis test. The table shows that both of the metric methods seem to have a lower false positive rate than the binning methods, but this is because the metric methods need many more data points to gain sufficient sensitivity. The binning methods have comparable false positive rates when the number of data points is 8192.

It should be noted that even for smaller amounts of data than the ones listed in Tables I–IV, the false positive rate satisfies the hypothesis test and only the sensitivity is adversely impacted. Using smaller than recommended amounts of data should not therefore cause spurious detections.

TABLE III. Rössler 1:5 experiment: The first column in the table identifies the applied method, the second column the range of parameters providing good results, the third column the smallest amount of samples from which more than 95% true positives and fewer than 5% false positives were obtained, and the fourth column shows the number of false positives.

Method	Parameter	Min. pts.	FP
B spline	Order 3, eight bins	8192	≈5%
Eqiquantal binning	Eight bins	8192	≈5%
Cross redundancies	$\epsilon\in\langle 0.22,0.3\rangle$	8192	≈5%
*k*NN	$k\in\{4,8\}$	8192	≈5%

TABLE IV. Rössler 5:1 experiment: The first column in the table identifies the applied method, the second column the range of parameters providing good results, the third column the smallest amount of samples from which more than 95% true positives and fewer than 5% false positives were obtained, and the fourth column shows the number of false positives.

Method	Parameter	Min. pts.	FP
B spline	Order 3, eight bins	1024	$\approx 5\%$
Eqiquantal binning	Eight bins	1024	$\approx 5\%$
Cross redundancies	$\epsilon \in \langle 0.1, 0.3 \rangle$	8192	$<5\%$
kNN	$k \in \{1,2,4,8,16\}$	8192	$<5\%$

V. DISCUSSION AND CONCLUSION

The problem of inferring directionality of coupling from time series was examined in this paper. A directionality detection method based on the framework of information theory was described. Several methods of estimating information theoretic functionals were presented. The behavior of these methods on a linear model for which an analytical estimate is available was shown. The methods showed convergence characteristics in accordance with expectations based on available theory. Detailed numerical experiments in detecting the directionality of coupling in pairs of Rössler oscillators were performed under various conditions. Based on the results of the experiments, it is clear that metric methods provide performance on par with the binning methods; however, their parameters may need to be adjusted. This leads us to recommend using binning methods on experimental data.

The problem of bias and variance of the conditional mutual information estimators merits a discussion of its own. In general, it can be stated that low variance of an estimator should always be preferred over low bias if the estimates are to be subjected to surrogate testing. This is because any bias of the estimator itself tends to be almost identical in the original data and in the appropriate surrogates, and thus has practically no effect when testing for the significance of directionality detections. The variance of an estimator causes the distribution of the CMI estimates from the surrogates to widen further than is merited by their intrinsic variance. This effect becomes important if the additional variance is large enough to distort the results of the surrogate tests. As a consequence, binning estimators do not suffer from worse detection statistics than the kNN method, although binning methods have a significantly higher bias (cf. Figs. 1 and 2). Estimators that have displayed a higher bias (or a substantial dependence of the bias on the value of the free parameter) can thus prove more robust when applied to detection problems, as was the case with the equiquantal estimator. There is one additional important source of bias in the testing problem: the surrogates themselves. Because in practice there are no perfect surrogates, the distribution of the directionality indices on the surrogate data does not exactly represent the null hypothesis of uncoupled systems. This bias is much more difficult to cope with and can be decreased only with improved surrogate generation methods that would allow for a more accurate estimation of the distribution of the CMI indices under the null hypothesis. This effect is seen in Figs. 4 and 5. The width of the CMI index distribution on permutation surrogates drops too quickly with larger data sets for some parameter values, causing false positive detections.

Several statements can be made about the character of the problem of directionality estimation using information theoretic functionals. Surprisingly, removing the intrawave information from the phase time series does not impact the performance of the estimators if the system parameters are similar. This may, however, require that the discrete nature of the time series is taken into account when selecting the value of the free parameter of a particular estimator. It is clear from the above experiments that the problem of detecting directionality between dissimilar systems is much more difficult than detecting directionality in systems with similar dynamics. However, many interesting systems such as the cardiorespiratory system are composed of subsystems with quite different dynamics: the frequency ratio of the lung oscillator to the heart oscillator is approximately 1:4. The conducted experiments clearly indicate that good performance on a simple balanced model does not necessarily imply that a particular method is fit to be applied to practical problems with mismatched parameters or differing dynamics. New methods for detecting directionality should be tested precisely on such unbalanced systems to support the claim that they correctly detect directionality in complex cases.

ACKNOWLEDGMENTS

This study was supported by the EC FP6 project BRACCIA Contract No. 517133 NEST and in part by the Institutional Research Plan No. AV0Z10300504.

[1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization. A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, U.K., 2003).
[2] C. Schäfer, M. G. Rosenblum, J. Kurths, and H.-H. Abel, Nature (London) **392**, 239 (1998).
[3] C. Schäfer, M. G. Rosenblum, H.-H. Abel, and J. Kurths, Phys. Rev. E **60**, 857 (1999).
[4] M. Paluš and D. Hoyer, IEEE Eng. Med. Biol. Mag. **17**, 40 (1998).
[5] M. Bračič Lotrič and A. Stefanovska, Physica A **283**, 451 (2000).
[6] J. Jamšek, A. Stefanovska, and P. V. E. McClintock, Phys. Med. Biol. **49**, 4407 (2004).
[7] A. Stefanovska, H. Haken, P. V. E. McClintock, M. Hožič, F. Bajrovič, and S. Ribarič, Phys. Rev. Lett. **85**, 4831 (2000).
[8] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, Phys. Rev. E **54**, 6708 (1996).
[9] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela,

Physica D **127**, 250 (1999).
[10] 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).
[11] M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[12] M. Paluš, V. Komárek, T. Procházka, Z. Hrnčíř, and K. Štěrbová, IEEE Eng. Med. Biol. Mag. **20**, 65 (2001).
[13] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[14] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001).
[15] M. G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, and R. Mrowka, Phys. Rev. E **65**, 041909 (2002).
[16] C. W. J. Granger, Econometrica **37**, 424 (1969).
[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).
[18] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[19] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[20] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer, Physica D **58**, 77 (1992).
[21] M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[22] M. Paluš and A. Stefanovska, Phys. Rev. E **67**, 055201(R) (2003).
[23] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. Rand and L. Young (Springer, Berlin, 1981), Vol. 898, pp. 366–381.
[24] A. Pikovsky, M. Rosenblum, G. Osipov, and J. Kurths, Physica D **104**, 219 (1997).
[25] D. Gabor, J. IEE London **93**, 429 (1946).
[26] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, Physica D **144**, 358 (2000).
[27] M. Le Van Quyen, J. Foucher, J.-P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F. Varela, J. Neurosci. Methods **111**, 83 (2001).
[28] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, Hum. Brain Mapp. **8**, 194 (1999).
[29] M. Paluš, D. Novotná, and P. Tichavský, Geophys. Res. Lett. **32**, L12805 (2005).
[30] D. Prichard and J. Theiler, Phys. Rev. Lett. **73**, 951 (1994).
[31] M. Paluš, Phys. Lett. A **213**, 138 (1996).
[32] M. Paluš, Phys. Lett. A **235**, 341 (1997).
[33] J. W. Cooley and J. W. Tukey, Math. Comput. **19**, 297 (1965).
[34] B. Musizza, A. Stefanovska, P. V. E. McClintock, M. Palus, J. Petrovcic, S. Ribaric, and F. F. Bajrovič, J. Physiol. **580**, 315 (2007).
[35] N. Marwan, M. Romano, M. Thiel, and J. Kurths, Phys. Rep. **438**, 237 (2007).
[36] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, Europhys. Lett. **4**, 973 (1987).
[37] M. Thiel, M. Romano, J. Kurths, M. Rolfs, and R. Kliegl, Europhys. Lett. **75**, 535 (2006).
[38] R. Moddemeijer, Signal Process. **16**, 233 (1989).
[39] A. M. Fraser and H. L. Swinney, Phys. Rev. A **33**, 1134 (1986).
[40] G. Darbellay and I. Vajda, IEEE Trans. Inf. Theory **45**, 1315 (1999).
[41] G. Darbellay, Comput. Stat. Data Anal. **32**, 1 (1999).
[42] M. Paluš, Physica D **93**, 64 (1996).
[43] M. Paluš, Physica D **80**, 186 (1995).
[44] C. O. Daub, R. Steuer, J. Selbig, and S. Kloska, BMC Bioinf. **5**, 118 (2004).
[45] C. D. Boor, *A Practical Guide to Splines* (Springer, Berlin, 1978).
[46] C. J. Cellucci, A. M. Albano, and P. E. Rapp, Phys. Rev. E **71**, 066208 (2005).
[47] L. F. Kozachenko and N. N. Leonenko, Probl. Inf. Transm. **23**, 9 (1987).
[48] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[49] M. Goria, N. Leonenko, V. Mergel, and P. Inverardi, J. Nonparametric Stat. **17**, 277 (2005).
[50] D. Prichard and J. Theiler, Physica D **84**, 476 (1995).
[51] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).
[52] R. Savit and M. Green, Physica D **50**, 95 (1991).
[53] J. Barnes, H. Sargent, and P. Tryon, in *The Ancient Sun*, edited by R. Pepin, J. Eddy, and R. Merrill (Pergamon Press, New York, 1980), pp. 159–163.
[54] R. O. Duda and P. P. E. Hart, *Pattern Classification and Scene Analysis* (John Wiley, New York, 1973).