Momentary information transfer as a coupling measure of time series

Bernd Pompe[1] and Jakob Runge[2]
[1]*Institute of Physics, University of Greifswald, Greifswald Germany,*
[2]*Potsdam Institute for Climate Impact Research and Department of Physics, Humboldt University Berlin, Berlin, Germany*
(Received 22 December 2010; published 19 May 2011)

We propose a method to analyze couplings between two simultaneously measured time series. Our approach is based on conditional mutual sorting information. It is related to other concepts for detecting coupling directions: the old idea of Marko for directed information and the more recent concept of Schreiber's transfer entropy. By setting suitable conditions we first of all consider momentary information in both time series. This enables the detection not only of coupling directions but also delays. Sorting information refers to ordinal properties of time series, which makes the analysis robust with respect to strictly monotonous distortions and thus very useful in the analysis of proxy data in climatology. Fortunately, ordinal analysis is easy and fast to compute. We consider also the problem of reliable estimation from finite time series.

DOI: 10.1103/PhysRevE.83.051122 PACS number(s): 02.50.−r, 05.45.Tp, 05.45.Xt

I. INTRODUCTION

The investigation of coupling and causality between systems is of great general interest in multivariate time-series analysis. Recently, nonlinear methods have been proposed, some of them on an information theoretical basis [1–9]. There are interesting applications, for example, in the analysis of the cardiac respiratory system [7], epileptic focus localization from electroencephalography (EEG) data of the human brain [8], functional magnetic resonance imaging of different brain regions in cognitive tasks [10], and the climate system [11]. Remember that the well-known cross correlation measures only linear statistical dependencies. The information theoretic quantity *mutual information* (MI) takes into consideration nonlinear dependencies as well. However, a proper *conditioned MI* (CMI) allows to study also causal relations (i.e., to distinguish between the driving and the driven subsystem).

The idea of using CMI for studying causal dependencies by setting suitable conditions is not very new. Already Marko [12,13] considered the so-called *directed information* in the context of data transmission over discrete memoryless channels with feedback. It had been discussed until recently in other applications like gambling, stock market portfolio strategies, and data compression with causal side information (e.g., Refs. [14–16]). Schreiber's *transfer entropy* (TE) [1] is similar to Marko's directed information. A problem with TE is that it does not well-detect coupling delays. We will argue that this can be solved by setting some other conditions leading to another special CMI which we call *momentary information transfer* (MIT). To be practicable we use some coarse- graining (partitioning) of state space which is related to order patterns of the time series, thus we consider only sorting information. It finally leads to a coupling measure which we call *momentary sorting information transfer* (MSIT). In this introductory section we give a rough explanation of our main ideas. Details and examples follow in the next sections.

A. Momentary information and noise

The existence of some momentary information in both time series is crucial for our concept of coupling analysis (i.e., we need series with nonvanishing source entropy). Remember that in the framework of Shannon's information theory *source entropy* h_x of a series of symbols is the uncertainty (also called entropy) on the next symbol x_t remaining if all former symbols $(\ldots,x_{t-2},x_{t-1})$ are known. If we get knowledge on this next symbol by a measurement, we would say that we get the *information* h_x. Thus, in the case of nonvanishing source entropy, complete knowledge on the next symbol would require some additional information. Here we call it also *momentary information* as it is the new information we would get from a measurement in this moment only. A series with positive source entropy is not predictable from its past with zero mean prediction error. Any noisy time series has momentary information as well as a series obtained from a chaotic dynamical system.

All errors in measurements we make due to nonperfect measuring equipment, which has, however, no influence on the time evolution of the considered system, are summarized as *observational noise*. In any measurement of a physical quantity with continuous amplitude we have some inevitable quantization error which is first of all due to a limited resolution of the analog-to-digital converter. This coarse-graining contributes to observational noise. In the framework of the ergodic theory of dynamical systems this quantization is considered as partitioning of state space. In contrast, the so-called *dynamical noise* has an influence on the time evolution. It is the random part of the time series that is not *observational noise*. Dynamical noise can occur due to the interaction of the considered system with another possibly very complex system. This might also be our measuring apparatus. It could be modeled by adding a random variable to the state space variables. However, also some multiplicative noise may be appropriate for modeling. As a result of the measurements we get a symbolic sequence (times series). The source entropy h^{symb} of this sequence depends on both the system under study, including dynamical noise, as well as the observational noise.

Consider a dynamical system (without dynamical noise), and the observational noise of the symbolic sequence should be only due to partitioning. Then the source entropy h^{symb} depends on the way of partitioning, however, it is bounded above by a finite value, the *Kolmogorov Sinai entropy* (KSE), h^{KSE}, also called *metric entropy*. Here $0 \leqslant h^{\mathrm{symb}} \leqslant h^{\mathrm{KSE}} < +\infty$ holds.

 ©2011 American Physical Society

A dynamical system is called chaotic if $0 < h^{\mathrm{KSE}} < +\infty$. If there is a partition of state space such that $h^{\mathrm{symb}} = h^{\mathrm{KSE}}$, it is called *generating partition*. Such partitions are "best" in the sense that, given all measurements in the past, the new measurement provides a maximum of new (momentary) information, which would be good for our coupling analysis. In general, we have $h^{\mathrm{symb}} \to h^{\mathrm{KSE}}$ if we use finer and finer partitions (i.e., for an idealized analog-to-digital converter with vanishing quantization step). Hence, source entropy h^{symb} approximates KSE for time series using high-precision measurements. This is, however, impractical. Later we come back to this point.

For a dynamical system with some dynamical noise or any typical random process (e.g., a random walk) KSE diverges, $h^{\mathrm{KSE}} = +\infty$. Nevertheless, in any case we have $h^{\mathrm{symb}} < +\infty$ if the number of possible different outputs of our measuring equipment is finite.

For our coupling analysis we need (i) nonvanishing symbolic source entropy, $h^{\mathrm{symb}} > 0$, and (ii) not all of this entropy should have its origin in observational noise only. Indeed, our method yields the best results if there is a large amount of dynamical noise and/or the system is highly chaotic (large KSE).

B. Basic idea of coupling analysis

In our coupling analysis we ask for mutual momentary information of both time series. More precisely, we search for momentary information in observations x_t of one time series $\{x_t\}$ that is contained in the momentary information of observations $y_{t+\tau}$ of the other series $\{y_t\}$, where the time lag $\tau > 0$ refers to coupling delays. The condition in this mutual information analysis is the *joint history* $[(\ldots,x_{t-2},x_{t-1}),(\ldots,y_{t+\tau-2},y_{t+\tau-1})]$ of both time series. Note that this is different from considering the histories of the x and y series separately. Thus, in our coupling analysis, we search for the common part of the momentary information in x_t and $y_{t+\tau}$ that is not in the joint history of both time series.

The variation of τ leads to the momentary information transfer (MIT) function $I^{\mathrm{MIT}}(\tau)$, which is our coupling measure. In practice we have only finite time series, thus we cannot consider the whole history. Instead we select a finite number of values in the past, and the set of corresponding time indices is called an *embedding time comb*.

If there is some coupling $x \to y$ with coupling delay $\tau_{x\to y} > 0$, we expect that $I^{\mathrm{MIT}}(\tau)$ significantly differs from zero for this lag $\tau = \tau_{x\to y}$. We consider only positive time lags as our analysis presumes that we have first the cause and then the effect. For the analysis of signals that are continuous in time we have to apply sampling periods that are less than the supposed minimum coupling delay. Otherwise we would observe the information transfer at $I^{\mathrm{MIT}}(\tau = 0)$ and thus could not resolve the coupling direction.

Coupling in the inverse direction $y \to x$ could be detected by the exchange of the series or, equivalently, by taking the first settings but negative time lags $\tau < 0$. Note that any coupling measure should be asymmetric to be able to distinguish between the driving and the driven subsystem. Our coupling measure has this asymmetry, as, in general, $I^{\mathrm{MIT}}(\tau) \neq I^{\mathrm{MIT}}(-\tau)$ for $\tau \neq 0$.

The basic idea behind our approach of coupling analysis can be illustrated as follows. Suppose we observe two real-world systems x and y by simultaneously recording any meaningful quantities x_t and y_t, respectively. We want to find out whether x can influence y. Suppose that this could be done by " kicking" system x and waiting for a response of y. The latter should be visible by any event in the y_t record that is not expected if y would run autonomously, this could be an unusual burst. If we see such a response at delay $\tau = \tau_{x\to y} > 0$, we would perhaps repeat this experiment again and again. When we always get the response at the same delay, we would be finally convinced that x indeed has an influence on y, with this delay.

This setup needs the opportunity to kick the system x. In real-world data analysis this is, however, often not possible. Fortunately, our approach is nonintervening. Momentary information $h_{x,t}$ of the x series plays the role of kicking. That part of the momentary information $h_{y,t+\tau}$ of the y series that is common with $h_{x,t}$ plays the role of the response (burst). The repeats are due to the fact that the considered information is an averaged quantity. This approach works also if there is a feedback $y \to x$, as the effect of feedbacks are suppressed by considering only momentary information.

C. Common driver problem

In our coupling analysis we assume that there are two systems x and y that can bidirectionally interact with delays $\tau_{x\to y} > 0$ and $\tau_{y\to x} > 0$. However, consider the case of a hidden common driving system c that couples to both other systems with delays $\tau_{c\to y} > \tau_{c\to x} > 0$, respectively. Then our bivariate analysis would erroneously yield a coupling $x \to y$ with coupling delay $\tau_{x\to y} = \tau_{c\to y} - \tau_{c\to x}$. From bivariate observations (time series) of systems x and y we are not able to distinguish between direct coupling, $x \to y$, and spurious coupling by a common driver, $c \to x$ and $c \to y$. To evaluate whether a special third system c acts as a common driver, we have to measure a third time series at system c. Then from the analysis of all six pairwise coupling delays we could get the whole picture. This is one reason why we are interested in detecting not only coupling directions, but also delays, and this makes the main difference of our MIT approach compared to TE. Such common driver analysis was also done in Ref. [6], however, with a coupling measure different from our present approach. A more detailed consideration of directed information and Granger causality between nodes in a complex network having more than two nodes is given in Refs. [17,18].

D. Symbolic analysis

In the framework of Shannon's information theory, entropies are functionals of probability distributions. Especially in the analysis of conditional mutual information considered here, we usually need rather high-dimensional distributions of some joint random variables that are supposed to generate the time series we are dealing with. The latter have finite length and thus it may be hard to get reliable entropy estimates with moderate computational efforts, and without additional assumptions on the underlying stochastic processes.

For series that are continuous in amplitude we could work with some approximations of probability densities via kernel estimators as proposed in Ref. [3]. We could use also some nearest-neighbor statistics as in Ref. [6]. The existence of densities is well motivated in real-world data analysis due to

the omnipresent noise. However, in the framework of nonlinear dynamics we often have to deal with invariant measures that are fractal, and thus no such density exists.

Symbolic analysis does not presume probability densities and thus represents an alternative approach. In general, the symbols are obtained from some coarse-graining of state space. For instance, due to analog-to-digital conversion of a continuous signal we already have a symbolic representation. However, usually the corresponding number of different symbols is still too large for reliable statistics. Hence some additional coarse-graining (binning) must be performed. In the so-called *adaptive binning* the bins are defined by (nearly) equidistant quantiles of the one-dimensional (marginal) empirical probability distribution. However, in general there is much ambiguity of binning.

We derive a symbolic representation of the time series in another way — we consider order patterns of the delay embedding vectors (" words"). Each order pattern is equivalent to a permutation of the coordinates of this vector such that they are arranged in an ascending order. In this way only ordinal properties of the data are considered. Indeed, each order pattern (permutation) represents a subset of the embedding state space. Hence, the basis of our method of symbolization is nothing but a special partition of state space, the *permutation partition*.

Note also that ordinal time-series analysis has advantages in the analysis of proxy data in climatology. Think, for instance, of historical temperature data that can be measured only indirectly from color intensities of sediments [19]. If the relation between both quantities is strictly monotone, order relations are the same in each series, and thus temperature analysis can be based directly on color intensities.

E. Permutation and sorting entropy

Permutation entropy (PE) [20] is derived from permutation partition. In the framework of (nonlinear) dynamics, PE is related to the metric entropy of the dynamical system [21–23]. That is why the method of order patterns is considered as a rather natural way of coarse-graining, and we expect that it keeps essential parts of momentary information in the symbol sequences — information that is needed for our coupling analysis as described above.

PE measures the information on the order relations in a (real-valued) delay embedding vector (i.e., the information needed to sort this vector in an ascending sequence). *Mutual sorting information* (MSI) is the information on sorting one vector contained in PE of another vector. Finally, *conditional MSI* (CMSI) is MSI under the condition that sorting of some other delay embedding vectors is given. The condition has to be carefully chosen to consider (first of all) momentary information in our coupling analysis. An advantage of our approach is that no explicit partitioning of state space has to be done. For a recent review on permutation entropy and ordinal patterns we refer to Amigo [23].

Order patterns have already been used for coupling analysis [7,8]. However, the methods proposed there consider CMI of the series of permutation indices. As they are constructed from delay embedding vectors, the series of permutation indices have some inevitable autodependencies even when the original series is perfectly random [i.e., if it can be considered as the realizations of an *independent identically distributed* (iid) stochastic process]. This implies disadvantages for coupling analysis as it was discussed in Ref. [9]. The methods are compared in Fig. 1. Moreover, the concept in Ref. [7] is not purely ordinal as a difference of vectors is involved.

F. Overview

In Sec. II we review basic definitions for delay embedding, order statistics, information, and entropy. In Sec. III we introduce MSI and CMSI. For appropriate conditions this leads to MSIT, which is our measure for coupling of time series. The relation of MIT to Schreiber's transfer entropy or Marko's directed information are discussed as well as that of MSIT to other coupling measures *symbolic transfer entropy* (STE) and *transfer entropy on rank vectros* (TERV), which are defined on the basis of sorting entropies introduced in Refs. [8,9]. The problem of CMSI estimation from finite time series is considered in Sec. IV. In Sec. V we give several examples, among them bidirectionally and unidirectionally coupled nonlinear dynamical systems. We also discuss the influence of dynamical and observational noise. Finally, we consider some real-world data from climatology.

II. BASIC DEFINITIONS

A. Embedding time comb

For a (stationary, scalar) time series $\{x_t\}_{t=1}^{T}$ we consider the M-dimensional delay-embedding vector ($m = 1,2,\ldots,M$),

$$x_t^{\vartheta} = (x_{t+\vartheta_1},\ldots,x_{t+\vartheta_m},\ldots,x_{t+\vartheta_M}). \tag{1}$$

We use the time delays, called *embedding time comb*,

$$\vartheta = \{\vartheta_1,\ldots,\vartheta_m,\ldots,\vartheta_M\},$$

with integers $\vartheta_{m^*} \neq \vartheta_m$ for all $m^* \neq m$, and embedding dimension $M \geqslant 2$. We suppose $|\min\vartheta|, |\max\vartheta| \ll T$, and have in mind $M \ll T$. Time t is running in a range such that $1 \leqslant t + \vartheta_m \leqslant T$ for all m. In this paper we use only time combs $\{\vartheta_1,\ldots,\vartheta_M\}$ formed by multiples of an integer *embedding delay* d (i.e., $|\vartheta_m - \vartheta_{m+1}| = d$, for all $m = 1,\ldots,M-1$). However, in general, nonuniformly spaced time combs could be chosen as well, if there is good reason to do so [6].

B. Order patterns

We sort the coordinates according to their ranks

$$x_{t+\vartheta_{\pi(1)}} < \ldots < x_{t+\vartheta_{\pi(m)}} < \ldots < x_{t+\vartheta_{\pi(M)}}.$$

Sorting is related to a permutation π of the time comb indices

$$\{1,\ldots,m,\ldots,M\}$$
$$\downarrow$$
$$\{\pi(1),\ldots,\pi(m),\ldots,\pi(M)\}.$$

The *order pattern* of a vector x_t^{ϑ} is given by

$$R_t^{\vartheta} = (R_{t+\vartheta_1},\ldots,R_{t+\vartheta_m},\ldots,R_{t+\vartheta_M}), \tag{2}$$

where $R_{t+\vartheta_m} = \pi(m) - 1$ is the rank of $x_{t+\vartheta_m}$ among the coordinates of x_t^{ϑ}. It can be written also as

$$R_{t+\vartheta_m} = \#\{x_{t+\vartheta_{m^*}} < x_{t+\vartheta_m} : m^* = 1,2,\ldots,M\}, \tag{3}$$

where #{·} denotes the cardinality (number of elements) of the set {·}. Thus, the rank $R_{t+\vartheta_m}$ of the coordinate $x_{t+\vartheta_m}$ of the delay embedding vector (1) is given by the number of coordinates of this vector that are less than $x_{t+\vartheta_m}$. In this way, the smallest coordinate gets the rank 0 and the largest coordinate has rank $M-1$. Note that R_t^ϑ is a permutation of $[0,1,\ldots,(M-1)]$. In this way, a permutation π encodes the order pattern R_t^ϑ of the delay-embedding vector x_t^ϑ. Note that there may occur, at most, $M!$ different order patterns.

To be well defined also for tied ranks (i.e., for cases where there are some equal coordinates $x_t = x_{t^*}$) we set $x_t < x_{t^*}$ if $t < t^*$. However, we have in mind situations without tied ranks, which is almost surely the case for time series with absolutely continuously distributed amplitude, and rather likely for series obtained from real-world signals with a high resolution of analog-to-digital conversion.

C. Numbering of order patterns

The order pattern $R^\vartheta = (R_{\vartheta_1}, R_{\vartheta_2}, \ldots, R_{\vartheta_M})$ of any vector $x^\vartheta = (x_{\vartheta_1}, x_{\vartheta_2}, \ldots, x_{\vartheta_M})$ (omitting time index t for convenience) can be unambiguously numbered in the following way:

$$i = 1 + \sum_{m=2}^{M} (m-1)!\, r_m,$$

where $r_m \in \{0,1,2,\ldots,m-1\}$ is the rank of x_{ϑ_m} among $(x_{\vartheta_1}, x_{\vartheta_2}, \ldots, x_{\vartheta_m})$.

Consider, for example, the case $M = 3$. Here we have $M! = 6$ order patterns $(R_{\vartheta_1}, R_{\vartheta_2}, R_{\vartheta_3})$ which are numbered as

$$\begin{aligned}
(R_{\vartheta_1}, R_{\vartheta_2}, R_{\vartheta_3}) \to (r_2, r_3) &\to i = 1 + 1!\, r_2 + 2!\, r_3, \\
(2,1,0) &\to (0,0) \to i = 1, \\
(1,2,0) &\to (1,0) \to i = 2, \\
(2,0,1) &\to (0,1) \to i = 3, \\
(0,2,1) &\to (1,1) \to i = 4, \\
(1,0,2) &\to (0,2) \to i = 5, \\
(0,1,2) &\to (1,2) \to i = 6.
\end{aligned}$$

Note that this numbering is very useful for programming.

D. Permutation series

In the described way, the M-dimensional vector x_t^ϑ is mapped to a permutation

$$x_t^\vartheta \to \pi_t^\vartheta \in \{\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_{M!}\},$$

and a time series is transformed to the symbolic series of permutations

$$\{x_t\} \overset{\text{embedding}}{\longrightarrow} \{x_t^\vartheta\} \overset{\text{sorting}}{\longrightarrow} \{\pi_t^\vartheta\}. \tag{4}$$

Consider, for example, the time series

$$\{x_t\} = \{0.6, -1.4, -2.5, 0.9, 4.6, -3.6, \ldots\},$$

and the three-dimensional time comb $\vartheta = \{-4, -2, 0\}$, using embedding delay $d = 2$. Sorting of the delay-embedding vector $x_5^\vartheta = (x_1, x_3, x_5) = (0.6, -2.5, 4.6)$ yields the order pattern $(1,0,2)$. At the next time step we get $x_6^\vartheta = (x_2, x_4, x_6) = (-1.4, 0.9, -3.6)$ with the order pattern $(1,2,0)$.

E. Permutation entropy

Let p_i denote the probability for $\pi_t^\vartheta = \pi_i$ in the permutation series (4). Then order M *permutation entropy* is defined as [20]

$$H_x = \sum_{i=1}^{M!} p_i \ln \frac{1}{p_i}. \tag{5}$$

We always use the natural logarithm $\ln = \log_e$, thus entropies are measured in units of *nit* henceforth.

Consider two time combs ϑ_x and ϑ_z with cardinalities M_x and M_z, respectively. Now we ask for the conditional uncertainty on the order pattern of $x_t^{\vartheta_x}$, if we already know the order pattern of $x_t^{\vartheta_z}$

$$H_{x|z} = H_{xz} - H_z = \sum_{k=1}^{M_z!} p_k \sum_{i=1}^{M_x!} p_{i|k} \ln \frac{1}{p_{i|k}}, \tag{6}$$

where H_{xz} is the permutation entropy of the joint probabilities p_{ik} for $(\pi_t^{\vartheta_x}, \pi_t^{\vartheta_z}) = (\pi_{i\cdot}, \pi_{\cdot k})$. (Note that, in general, $\pi_{i\cdot} \neq \pi_{\cdot k}$ if ϑ_x and ϑ_z have different cardinality.) The marginal probabilities are $p_{\cdot k} = \sum_i p_{ik}$, and the conditional probabilities are given by $p_{i|k} = p_{ik}/p_{\cdot k}$, provided $p_{\cdot k} > 0$, $p_{i|k} = 0$ otherwise.

For the special time combs $\vartheta_z = \{-M_z, \ldots, -1\}$, $\vartheta_x = \vartheta_z \cup \{0\}$, $H_{x|z}$ is the so-called *sorting entropy* (SE) [20], with $H_{x|z} \leqslant \ln(M_z + 1)$. Equality $H_{x|z} = \ln(M_z + 1)$ holds for any iid series. For chaotic series, SE is related to the metric entropy of the underlying ergodic dynamical system [21,22]. For piecewise monotone one-dimensional (1D) maps SE converges to the metric entropy for $M_z \to \infty$, however, for noisy series SE diverges for this limit. In any case, for sufficiently large M_z, SE represents more or less *momentary information*.

Consider an example, $\vartheta_z = \{-2, -1\}$ and $\vartheta_x = \{-2, -1, 0, +1\}$. Here we have $H_{x|z} \leqslant \ln 3 + \ln 4$. This is the uncertainty to sort (x_t, x_{t+1}) among $(x_{t-2}, x_{t-1}, x_t, x_{t+1})$ remaining if (x_{t-2}, x_{t-1}) is already sorted. Again, equality holds for iid series.

For $\vartheta_z \subseteq \vartheta_x$ we get $H_{x|z} \leqslant \ln M_x! - \ln M_z!$. For any time combs we have $H_{x|z} \leqslant H_x \leqslant \ln M_x!$, where equalities hold for iid series if $\vartheta_z \cap \vartheta_x = \varnothing$. However, for $\vartheta_z \cap \vartheta_x \neq \varnothing$ we get $H_{x|z} < \ln M_x!$, even if the series is iid.

III. COUPLING MEASURES

A. Mutual sorting information

Suppose there are two time series $\{x_t\}$ and $\{y_t\}$ of any underlying systems x and y. We want to know whether these systems are coupled, with some positive delays $\tau_{x\to y}$ and $\tau_{y\to x}$. Somehow, naively we could consider *mutual sorting information* for varying time lag τ of the permutation series $\{\pi_t^{\vartheta_x}\}$ and $\{\pi_t^{\vartheta_y}\}$ obtained from the x and y series, respectively, as described by Eq. (4)

$$\begin{aligned}
I_{xy}^{\mathrm{MSI}}(\tau) &= H_x + H_y - H_{xy} \\
&= H_x - H_{x|y}
\end{aligned}$$

$$= H_y - H_{y|x}$$
$$= \sum_{i,j=1}^{M_x!,\,M_y!} p_{ij} \ln \frac{p_{ij}}{p_{i\cdot}\, p_{\cdot j}}. \quad (7)$$

The joint entropy H_{xy} is obtained from the joint probabilities p_{ij} for $(\pi_t^{\vartheta_x},\pi_{t+\tau}^{\vartheta_y}) = (\pi_{i\cdot},\pi_{\cdot j})$. The marginals are $p_{i\cdot} = \sum_j p_{ij}$, and $p_{\cdot j} = \sum_i p_{ij}$. The joint probabilities p_{ij} and hence also H_{xy} depend on time lag τ, however, the marginal entropies H_x and H_y do not due to supposed stationarity. For coupling analysis we would use embedding time combs ϑ_x and ϑ_y that have negative entries, but one entry being zero

$$\vartheta = \{0,\vartheta_2,\ldots,\vartheta_m,\ldots,\vartheta_M\},$$
$$\text{with} \quad \vartheta_m < 0 \quad \text{for all m} = 2,\ldots,\text{M}.$$

However, from this analysis we would get, in general, no detailed knowledge on coupling delays as the corresponding order patterns contain information smeared over a time interval given by the embedding time comb. Autodependencies of the time series could be misleading as well.

B. Conditional mutual sorting information

In our coupling analysis we want to restrict (as well as possible) to momentary mutual sorting information. This is achieved by setting suitable conditions z. We ask for MSI of x and y under the condition z which will be specified below. This leads to the *conditional mutual sorting information*

$$I_{xy|z}^{\text{CMSI}}(\tau) = H_{x|z} + H_{y|z} - H_{xy|z} \quad (8)$$
$$= H_{xz} + H_{yz} - H_z - H_{xyz} \quad (9)$$
$$= H_{x|z} - H_{x|yz}$$
$$= H_{y|z} - H_{y|xz}$$
$$= \sum_{k=1}^{M_z^*} p_{\cdot\cdot k} \sum_{i,j=1}^{M_x!,\,M_y!} p_{ij|k} \ln \frac{p_{ij|k}}{p_{i\cdot|k}\, p_{\cdot j|k}} \quad (10)$$
$$= \sum_{i,j,k=1}^{M_x!,\,M_y!,\,M_z^*} p_{ijk} \ln \frac{p_{i\cdot|jk}}{p_{i\cdot|k}} \quad (11)$$
$$= \sum_{i,j,k=1}^{M_x!,\,M_y!,\,M_z^*} p_{ijk} \ln \frac{p_{\cdot j|ik}}{p_{\cdot j|k}}. \quad (12)$$

In general, all entropies on the right-hand side depend on time lag τ. The joint entropy $H_{xyz} = H_{x|yz} + H_{yz}$ is obtained from the joint probabilities p_{ijk} for $[\pi_t^{\vartheta_x},\pi_{t+\tau}^{\vartheta_y},(\pi_t^{\vartheta_{zx}},\pi_{t+\tau}^{\vartheta_{zy}})] = (\pi_{i\cdot\cdot},\pi_{\cdot j\cdot},\pi_{\cdot\cdot k})$. The condition z is constructed from the joint permutations $(\pi_t^{\vartheta_{zx}},\pi_{t+\tau}^{\vartheta_{zy}})$, which can be numbered by $k = 1,2,\ldots,M_z^*$, with $M_z^* = M_{z_x}!\,M_{z_y}!$. The marginal distributions are obtained from $p_{i\cdot k} = \sum_j p_{ijk}$, $p_{\cdot jk} = \sum_i p_{ijk}$, and $p_{\cdot\cdot k} = \sum_{i,j} p_{ijk}$. Moreover, we use the notations

$$p_{i\cdot|k} = p_{i\cdot k}/p_{\cdot\cdot k}, \qquad p_{\cdot j|k} = p_{\cdot jk}/p_{\cdot\cdot k},$$
$$p_{i\cdot|jk} = p_{ijk}/p_{\cdot jk}, \qquad p_{\cdot j|ik} = p_{ijk}/p_{i\cdot k}.$$

In general, we have $0 \leqslant I_{xy|z}^{\text{CMSI}} \leqslant \min\{H_{x|z}, H_{y|z}\}$. Equality $I_{xy|z}^{\text{CMSI}} = 0$ holds if and only if x and y are independent under the condition z (i.e., if and only if $p_{ij|k} = p_{i\cdot|k}\, p_{\cdot j|k}$ for all i, j, k). For empty time combs $\vartheta_{z_x} = \vartheta_{z_y} = \oslash$, CMSI (8) reduces to MSI (7).

C. Momentary sorting information transfer

For our coupling analysis we take embedding time combs ϑ_{z_*}, where $*$ stands for x or y,

$$\vartheta_{z_*} = \{\vartheta_1,\ldots,\vartheta_m,\ldots,\vartheta_{M_{z_*}}\},$$
$$\text{with} \quad \vartheta_m < 0 \quad \text{for all} \quad m = 1,\ldots,M_{z_*}, \quad \text{and}$$
$$\vartheta_* = \vartheta_{z_*} \cup \{0\}. \quad (13)$$

If we use in Eq. (8) these special embedding time combs and positive time lags τ, we would call CMSI *momentary sorting information transfer*. Note that for large embedding dimensions M_{z_x} and M_{z_y}, MSIT regards first of all to momentary sorting information. Thus we write for all $\tau > 0$

$$I_{x\to y}^{\text{MSIT}}(\tau) = I_{xy|z}^{\text{CMSI}}(\tau),$$
$$I_{y\to x}^{\text{MSIT}}(\tau) = I_{xy|z}^{\text{CMSI}}(-\tau) = I_{\text{yx|z}}^{\text{CMSI}}(\tau). \quad (14)$$

Consider an example: We set $\vartheta_{z_x} = \vartheta_{z_y} = \{-2,-1\}$ and $\vartheta_x = \vartheta_y = \{-2,-1,0\}$. These time combs match the conditions in Eq. (13), hence CMSI is considered as MSIT. Then $I_{x\to y}^{\text{MSIT}}(\tau)$ is the information to sort x_t among (x_{t-2},x_{t-1}) that is contained in the information to sort $y_{t+\tau}$ among $(y_{t+\tau-2},y_{t+\tau-1})$, provided that the condition z is already known, which is the joint information of sorting (x_{t-2},x_{t-1}) and of sorting $(y_{t+\tau-2},y_{t+\tau-1})$. Here we get $I_{x\to y}^{\text{MSIT}}(\tau) \leqslant \ln 3$.

D. Transfer entropy and momentary information transfer

From Eqs. (11) and (12) we see that CMSI can be considered as a Kullback-Leibler entropy (divergence), that is, as the information we gain if the probability distribution $\{p_{i\cdot|k}\}$ is replaced by $\{p_{i\cdot|jk}\}$, or if $\{p_{\cdot j|k}\}$ is replaced by $\{p_{\cdot j|ik}\}$, respectively. Note that Schreiber's *transfer entropy* introduced in Ref. [1] was also motivated by interpreting it as an information gain. Using our notations for embedding dimensions, he considered the coupling measure

$$I_{x\to y}^{\text{TE}} = \sum p\left(x_t^{M_x}, y_{t+1}^{M_{z_y}+1}\right) \ln \frac{p\left(y_{t+1}|x_t^{M_x}, y_t^{M_{z_y}}\right)}{p\left(y_{t+1}|y_t^{M_{z_y}}\right)} \quad (15)$$
$$= \sum p\left(x_t^{M_x}, y_{t+1}, z\right) \ln \frac{p\left(x_t^{M_x}, y_{t+1}|z\right)}{p\left(x_t^{M_x}|z\right) p\left(y_{t+1}|z\right)}, \quad (16)$$

$$\text{with the condition} \quad z = y_t^{M_{z_y}},$$

where we use the delay embedding vectors

$$x_t^{M_x} = (x_{t-M_{z_x}+1},\ldots,x_t),$$
$$y_t^{M_{z_y}} = (y_{t-M_{z_y}+1},\ldots,y_t),$$

and $p(\ldots)$ being the corresponding probabilities. Hence TE could be considered also as a conditional mutual information between the future value y_{t+1} of the y series and the past $x_t^{M_x}$ of the x series under the condition $z = y_t^{M_{z_y}}$, which is the past of the y series. Note also that Marko's directed information [12,13] is equal to transfer entropy for embedding dimensions $M_x, M_{z_y} \to \infty$.

Suppose that we do not consider sorting information in our MSIT approach, but instead the probability distributions of

the time series in the same way as in TE. Moreover, we take embedding time combs as in Eq. (13), however, with $\vartheta_m = m$ for all $m = 1, \ldots, M_{z_*}$. Then instead of MSIT in Eqs. (10), (11), and (12) we obtain a quantity that we call *momentary information transfer* of x_t to y_{t+1}

$$I^{\mathrm{MIT}}_{x\to y}(\tau) = \sum p\left(x_t^{M_{z_x}+1}, y_{t+\tau}^{M_{z_y}+1}\right) \ln \frac{p\left(y_{t+\tau}|x_t^{M_{z_x}+1}, y_{t+\tau-1}^{M_{z_y}}\right)}{p\left(y_{t+\tau}|x_{t-1}^{M_{z_x}}, y_{t+\tau-1}^{M_{z_y}}\right)} \tag{17}$$

$$= \sum p(x_t, y_{t+\tau}, z) \ln \frac{p(x_t, y_{t+\tau}|z)}{p(x_t|z)p(y_{t+\tau}|z)},$$

$$\text{with the condition } z = \left(x_{t-1}^{M_{z_x}}, y_{t+\tau-1}^{M_{z_y}}\right),$$

$$\text{and time lag } \tau = 1,2,\ldots. \tag{18}$$

A comparision of Eqs. (15) and (16) with Eqs. (17) and (18), respectively, makes obvious the difference between both approaches: TE considers the mutual information of y_{t+1} and x_t under the condition z, which is the past of y_{t+1}. MIT represents the mutual information of $y_{t+\tau}$ and x_t under the condition z, which is the joint past of both, x_t and $y_{t+\tau}$, for varying time lag $\tau = 1,2,\ldots$. The idea for considering MSIT is just the same as considering MIT instead of TE. However, in MSIT only sorting information of $y_{t+\tau}$ among $y_{t+\tau-1}^{M_{z_y}}$ and that of x_t among $x_{t-1}^{M_{z_x}}$ is considered.

From $I^{\mathrm{M(S)IT}}_{x\to y}(\tau)$, $\tau = 1,2,\ldots$, we can expect a better resolution of coupling delays because we consider, first of all, momentary information not only of x_t but also of $y_{t+\tau}$. For large embedding dimensions (i.e., for $M_{z_x}, M_{z_y} \to \infty$) M(S)IT accounts for momentary information in both time series. In real-world data analysis this limit cannot be attained. That is why it may be rather useful to look for nontrivial low-dimensional embedding time combs such that, first of all, momentary information is used for coupling analysis. In some of our examples given below we vary the time combs.

E. Coinformation

In general, the difference between conditional and unconditional mutual information is known as *interaction information* or *coinformation* [24,25]. Here we get the *sorting coinformation* (SCO)

$$I^{\mathrm{SCO}}_{xyz} = I^{\mathrm{CMSI}}_{xy|z} - I^{\mathrm{MSI}}_{xy}$$
$$= H_{xy} + H_{xz} + H_{yz} - H_x - H_y - H_z - H_{xyz}.$$

Obviously, coinformation is completely symmetric (e.g., $I^{\mathrm{SCO}}_{xyz} = I^{\mathrm{SCO}}_{xzy} = I^{\mathrm{SCO}}_{yzx}$). In general, we get

$$I^{\mathrm{SCO,min}}_{xyz} \leqslant I^{\mathrm{SCO}}_{xyz} \leqslant I^{\mathrm{SCO,max}}_{xyz},$$

where

$$I^{\mathrm{SCO,min}}_{xyz} = -\min\left\{I^{\mathrm{MSI}}_{xy}, I^{\mathrm{MSI}}_{xz}, I^{\mathrm{MSI}}_{yz}\right\} \leqslant 0,$$
$$I^{\mathrm{SCO,max}}_{xyz} = \min\left\{I^{\mathrm{CMSI}}_{xy|z}, I^{\mathrm{CMSI}}_{xz|y}, I^{\mathrm{CMSI}}_{yz|x}\right\} \geqslant 0.$$

Hence coinformation may be positive, negative, or zero. However, in our examples discussed below we always get nonpositive coinformation, which is the more intuitive (natural) case where the condition z decreases mutual information between x and y.

A simple example for positive coinformation is given by the joint probability distribution $\{p_{ijk}\}$, where $p_{111} = p_{122} = p_{212} = p_{221} = 1/4$ and all other probabilities equal to zero. Here we get $H_x = H_y = H_z = \ln 2$, $H_{xy} = H_{xz} = H_{yz} = H_{xyz} = \ln 4$. Thus we have $I^{\mathrm{MSI}}_{xy} = 0$ (i.e., x and y are independent). However, $I^{\mathrm{CMSI}}_{xy|z} = \ln 2 > 0$ (i.e., x and y are not independent under the condition z). Hence, $I^{\mathrm{SCO}}_{xyz} = \ln 2 > 0$.

F. Relations of MSIT to other concepts

For comparison of our MSIT approach with that in Refs. [8,9] we refer to Fig. 1. We discuss it for embedding time delay $d = 1$. All three approaches represent special cases

(a) $I^{\mathrm{STE}}_{\mathrm{x}\to\mathrm{y}}(\tau)$ for $\tau > 0$

$$x_t^{M_{z_x}} = x_{t-M_{z_x}+1}, \ldots, x_t$$
$$\downarrow$$
$$\tau \quad y_{t+\tau}^{M_y} = y_{t+\tau-M_y+1}, \ldots, y_{t+\tau}$$

condition:

$$y_t^{M_y} = y_{t-M_y+1}, \ldots, y_t$$

(b) $I^{\mathrm{TERV}}_{\mathrm{x}\to\mathrm{y}}$

$$x_t^{M_{z_x}} = x_{t-M_{z_x}+1}, \ldots, x_t$$
$$\downarrow$$
$$y_{t+M_y^*}^{M_{z_y}+M_y^*} = y_{t-M_{z_y}+1}, \ldots, y_{t-1}, y_t, y_{t+1}, \ldots, y_{t+M_y^*}$$

condition:

$$y_t^{M_{z_y}} = y_{t-M_{z_y}+1}, \ldots, y_t$$

(c) $I^{\mathrm{MSIT}}_{\mathrm{x}\to\mathrm{y}}(\tau)$ for $\tau > 0$

$$x_t^{M_{z_x}+1} = x_{t-M_{z_x}}, \ldots, x_{t-1}, x_t \quad \pi_t^{\vartheta_x}$$
$$\downarrow$$
$$\tau \quad y_{t+\tau}^{M_{z_y}+1} = y_{t+\tau-M_{z_y}}, \ldots, y_{t+\tau-1}, y_{t+\tau} \quad \pi_{t+\tau}^{\vartheta_y}$$

condition:

$$\tau \quad y_{t+\tau-1}^{M_{z_y}} = y_{t+\tau-M_{z_y}}, \ldots, y_{t+\tau-1} \quad \pi_{t+\tau}^{\vartheta_{z_y}}$$
$$x_{t-1}^{M_{z_x}} = x_{t-M_{z_x}}, \ldots, x_{t-1} \quad \pi_t^{\vartheta_{z_x}}$$

FIG. 1. Setups of three different approaches for coupling analysis. Each box symbolizes the permutation (order pattern) π of the corresponding random vectors. (a) Symbolic transfer entropy (STE) according to Ref. [8]. (b) Transfer entropy on rank vectors (TERV) according to Ref. [9]. (c) Momentary sorting information transfer (MSIT, our approach). All three approaches represent special cases of conditional mutual sorting information (CMSI), Eq. (8).

of CMSI. One of the main differences is that in MSIT the condition z is derived from the joint past of both time series. In STE and TERV introduced in Refs. [8] and [9], respectively, the condition is only derived from the past of y. This could be expressed formally by $\vartheta_x = \oslash$ and hence $z = z_y$. Another striking difference of our CMSI approach is that for increasing time lag τ of the order pattern $\pi_{t+\tau}^{\vartheta_y}$ we also shift the conditional pattern $\pi_{t+\tau}^{\vartheta_{z_y}}$. All this has the effect that for varying time lag τ we can more precisely detect the right coupling delays as we consider first of all only momentary information.

G. Proper embedding time combs

The choice of proper embedding time combs plays a central role in our analysis. For our analysis it is crucial that, first of all, *momentary* information of the x and y series contributes to CMSI. Moreover, if we want to apply our method to short time series, we should use low embedding dimensions $M_x, M_y, M_{z_x}, M_{z_y}$, to get reliable CMSI estimates. A method to select perhaps nonuniformly spaced time combs for nonlinear autoregressive predictions of time series was described in Ref. [26]. Similar techniques to detect best time combs could be applied here as well. A more detailed consideration of time comb selection is beyond the scope of this paper. Nevertheless, in some of our examples given below we systematically vary time combs.

If the signal $x(t)$ is time continuous, a proper choice of the sampling period Δt is important. Suppose the signal is bandlimited. Then, for any embedding dimension M_z and sufficiently small Δt, most of the embedding vectors $[x(t - M_z\,\Delta t), \ldots, x(t - \Delta t), x(t)]$ represent a continuously increasing or decreasing episode, and it is rather likely that $x(t)$ is the largest or smallest coordinate, respectively. Hence, for fixed $M_z > 1$ and $\Delta t \to 0$, SE in Eq. (6) would vanish and our coupling analysis becomes impossible.

If the signal is noisy (dynamical noise), SE reaches its maximum $\ln(M_z + 1)$ for $\Delta t \to 0$, even for very low noise levels. A large value of SE is, in principle, good for our coupling analysis. However, this is no guarantee to detect all couplings. Some effects of noise are illustrated in the examples given below.

H. Coupling indices

In analogy to Refs. [2,5], we could consider also "coupling indices"

$$I_{x\to y}^{\sum \mathrm{MSIT}} = \frac{1}{\tau_{\max}} \sum_{\tau=1}^{\tau_{\max}} I_{x\to y}^{\mathrm{MSIT}}(\tau), \tag{19}$$

$$I_{y\to x}^{\sum \mathrm{MSIT}} = \frac{1}{\tau_{\max}} \sum_{\tau=-1}^{-\tau_{\max}} I_{x\to y}^{\mathrm{MSIT}}(\tau). \tag{20}$$

They might be considered as a coupling strength. The delay $\tau_{\max} > 0$ is the presumed maximal coupling delay, describing the coupling of x to y and y to x, respectively. However, from these indices we cannot derive coupling delays. That is why we rather consider the function $I_{x\to y}^{\mathrm{MSIT}}(\tau)$. We should also be aware that the magnitude of our coupling measure depends on the embedding dimensions and time combs. Hence its interpretation is somehow arbitrary, and we should look first of all for time lags τ where $I_{xy|z}^{\mathrm{MSIT}}(\tau)$ has significant peaks.

Note also that if we would take coupling indices (19) and (20) based on transfer (sorting) entropy, T(S)E, instead of momentary (sorting) information transfer, M(S)IT, the indices would typically yield a larger coupling strength as T(S)E typically has broader peaks near the coupling delays than M(S)IT has.

IV. ESTIMATION

The proposed CMSI method is based on permutation entropy (5) which is a (discrete) Shannon entropy. It has to be estimated from finite permutation series $\{\pi_t^{\vartheta}\}_{t=0}^{T}$, derived from the time series as described in Eq. (4). The same holds for the joint permutation entropies at the right hand side of Eq. (8). (Note that length T here is a bit smaller than that of the time series due to embedding.)

For the estimation of entropy (5) we could somehow naively replace the probability p_i by the relative frequencies t_i/T, $t_i = \#\{\pi_t^{\vartheta} = \pi_i \,:\, t = 1, \ldots, T\}$. The resulting naive estimator

$$\widehat{H}^n = \ln T - \frac{1}{T} \sum_{i=1}^{M!} t_i \ln t_i, \tag{21}$$

would work for $T \gg M!$ and $t_i \gg 1$ for all $i = 1,2, \ldots, M!$. However, for small t_i it is strongly biased. Here Grassberger's estimator [27,28] works much better, which is defined by

$$\widehat{H}^G = \ln T - \frac{1}{T} \sum_{i=1}^{M!} t_i\, G(t_i), \tag{22}$$

where

$$\begin{gathered} G(0) = 0, \\ G(1) = -\gamma - \ln 2, \quad \text{with Euler's constant } \gamma \approx 0.577\,215, \\ G(2) = 2 + G(1), \\ G(t) = G(t-1) \quad \text{for all odd } t = 3,5\ldots, \\ G(t) = G(t-1) + 2/(t-1) \quad \text{for all even } t = 4,6\ldots. \end{gathered}$$

The bias of $\widehat{H}^G$ might also be negative which is in contrast to that of $\widehat{H}^n$. This is, however, irrelevant for our CMSI estimator decribed below.

We consider the CMSI estimator

$$\widehat{I^{\mathrm{CMSI}}}^G_{xy|z}(\tau) = \widehat{H_{xz}}^G + \widehat{H_{yz}}^G - \widehat{H_z}^G - \widehat{H_{xyz}}^G, \tag{23}$$

which is defined by replacing H in Eq. (8) with $\widehat{H}^G$ defined in Eq. (22). Analogously, we define the naive CMSI estimator $\widehat{I^{\mathrm{CMSI}}}^n$ by replacing H in Eq. (8) with $\widehat{H}^n$. Some values for the bias b and standard deviation (SD) s of both CMSI estimators are given in Table I.

We found that $\widehat{I^{\mathrm{CMSI}}}^G$ has lower bias than $\widehat{I^{\mathrm{CMSI}}}^n$, however, usually larger SD. In our examples we always use $\widehat{H}^G$. The uncoupled iid series provide much larger bias than the uncoupled logistic maps. Thus we cannot simply derive a universal threshold $I^{\mathrm{CMSI}*}$ such that coupling could be considered to be significant if $\widehat{I^{\mathrm{CMSI}}} > I^{\mathrm{CMSI}*}$.

We obtain this threshold from a shuffle test: Let $\{\pi_{\lambda(t)}^{\vartheta_y}\}$ and $\{\pi_{\lambda(t)}^{\vartheta_{z_y}}\}$ be the shuffled permutation series corresponding to

TABLE I. Bias b and standard deviation s of the naive and Grassberger's CMSI estimator (23) for uncoupled iid sequences and uncoupled logistic maps [Eq. (25), for $c_{y\to x} = c_{x\to y} = 0$]. The time combs are $\vartheta_x = \vartheta_y = \{-2,-1,0\}$, and for the condition $\vartheta_{z_x} = \vartheta_{z_y} = \{-2,-1\}$. The values are obtained from $N = 1000$ runs for each series length T. The error of the bias represents $b \pm (2s/\sqrt{N})$.

	$\widehat{I^{\mathrm{CMSI}^n}}$		$\widehat{I^{CMSI^G}}$	
$\log_2 T$	b	s	b	s
	Uncoupled iid series			
6	0.1737(33)	0.052	0.0063(65)	0.10
7	0.0754(16)	0.026	0.0013(28)	0.044
8	0.03375(74)	0.011	0.0006(12)	0.020
9	0.01639(36)	0.0058	0.00032(63)	0.0099
10	0.00805(18)	0.0028	0.00012(30)	0.0048
11	0.004000(88)	0.0013	0.00009(14)	0.0023
12	0.001980(43)	0.00069	0.000057(76)	0.0012
	Uncoupled logistic maps			
6	0.0888(26)	0.041	0.0118(52)	0.083
7	0.0435(13)	0.020	0.0032(24)	0.039
8	0.02037(64)	0.010	0.0014(11)	0.018
9	0.01000(33)	0.0053	0.00085(55)	0.0088
10	0.00481(16)	0.0025	0.00029(26)	0.0042
11	0.002367(79)	0.0012	0.00016(13)	0.0021
12	0.001169(38)	0.00060	0.000050(68)	0.0010

the y series, using a random permutation λ, $[1,2,\ldots,T] \to [\lambda(1),\lambda(2),\ldots,\lambda(T)]$. This kind of shuffling keeps the probability distribution $\{p_{yz_y}\}$, and hence the corresponding marginal distributions $\{p_{y\cdot}\}$ and $\{p_{\cdot z_y}\}$, but it destroys (with high probability) any coupling between the x and y series. We shuffle using many different random permutations λ_l, $l = 1,2,\ldots,L$, and get for each λ_l a CMSI estimate I_l^{CMSI} according to Eq. (23). Finally, the threshold $I^{\mathrm{CMSI}*}$ for coupling decision is derived from the corresponding empirical 95% quantile, that is, $I^{\mathrm{CMSI}*}$ is chosen such that

$$I^{\mathrm{CMSI}*} \in \left\{I_l^{\mathrm{CMSI}}\right\}, \quad L^* - 1 < 0.95L \leqslant L^*, \tag{24}$$

where $L^* = \#\{I_l^{\mathrm{CMSI}} < I^{\mathrm{CMSI}*}\}$.

Note that for the time combs defined in Eq. (13), CMSI is equal to our coupling measure MSIT, Eq. (14).

V. EXAMPLES

A. Bidirectionally coupled logistic maps

We study a bidirectionally delay-coupled logistic map $f : [0,1] \to [0,1]$, $f(x) = 4x(1-x)$,

$$\begin{aligned} x_t &= f(g_{y\to x} \bmod 1), \\ g_{y\to x} &= c_{y\to x} y_{t-\tau_{y\to x}} + (1-c_{y\to x}) x_{t-1}, \\ y_t &= f(g_{x\to y} \bmod 1), \\ g_{x\to y} &= c_{x\to y} x_{t-\tau_{x\to y}} + (1-c_{x\to y}) y_{t-1}. \end{aligned} \tag{25}$$

In our examples we always take only $T = 512$ samples. First we examine the role of conditioning. We test whether CMSI can detect the correct coupling lag which allows to draw

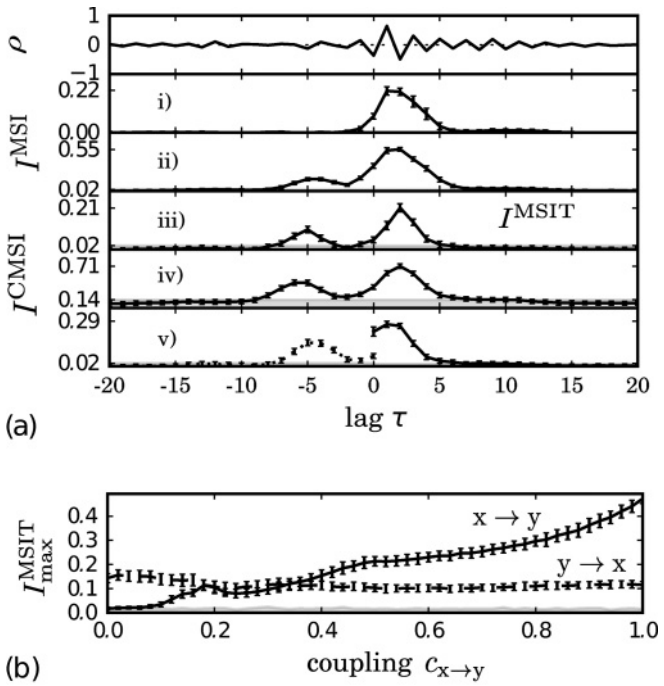

FIG. 2. Estimates of mutual sorting information (MSI), Eq. (7), and several conditional MSI (CMSI), Eq. (8), according to the estimator of Eq. (23) for the bidirectionally coupled logistic map (25). The coupling strengths amount to $c_{y\to x} = 0.2$ and $c_{x\to y} = 0.5$, the coupling delays to $\tau_{y\to x} = 5$ and $\tau_{x\to y} = 2$. The sample size is $T = 512$. Error bars denote standard deviations of 1000 trials. The gray band denotes the 95% significance level $I^{\mathrm{CMSI}*}$, Eq. (24). Panel (a) shows the cross correlation function ρ in the top plot as well as MSI and CMSI functions for different time combs: (i) MSI for $\vartheta_x = \vartheta_y = \{-1,0\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \oslash$, (ii) MSI for $\vartheta_x = \vartheta_y = \{-2,-1,0\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \oslash$, and (iii) CMSI for $\vartheta_x = \vartheta_y = \{-2,-1,0\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \{-2,-1\}$. Note that for these time combs CMSI is equivalent to MSIT [cf. Eqs. (13) and (14)]. (iv) CMSI for $\vartheta_x = \vartheta_y = \{-2,-1,0,1\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \{-2,-1\}$, (v) CMSI for $\vartheta_x = \vartheta_y = \{-2,-1,0\}$ for all lags τ, but $\vartheta_{z_x} = \oslash$, $\vartheta_{z_y} = \{-2,-1\}$ for positive $\tau > 0$, and $\vartheta_{z_x} = \{-2,-1\}$, $\vartheta_{z_y} = \oslash$ for $\tau < 0$. The ticks at the ordinate mark the significance threshold $I^{\mathrm{CMSI}*}$, Eq. (24), and the maximum value $I_{\max}^{\mathrm{CMSI}} = \max_\tau I^{\mathrm{CMSI}}$. The figures illustrate that best results for coupling analysis are obtained by the MSIT approach of (a) (iii). (b) shows the maximum MSIT value for positive delays, $I_{\max,\tau>0}^{\mathrm{MSIT}} = \max_{\tau>0} I^{\mathrm{MSIT}}$ (solid line), and negative delays, $I_{\max,\tau<0}^{\mathrm{MSIT}} = \max_{\tau<0} I^{\mathrm{MSIT}}$ (dashed line). The coupling strength $c_{y\to x} = 0.2$ is constant, and $c_{x\to y}$ is varied. We use here the time combs of (a) (iii), hence CMSI is indeed MSIT, Eqs.(13) and (14). The 95% significance level is separate for each coupling.

conclusions about a coupling direction. In Fig. 2(a) we have set the coupling delays to $\tau_{y\to x} = 5$ and $\tau_{x\to y} = 2$, and the coupling strengths to $c_{y\to x} = 0.2$ and $c_{x\to y} = 0.5$.

Cross correlation ρ does not detect any coupling $y \to x$ for negative lags τ. Moreover, for positive lags the magnitude of ρ detects the wrong coupling delay. CMSI is shown for different time combs. Subfigure (i) and (ii) correspond to MSI with no conditions as in Eq. (7). Due to autodependencies the peaks are quite broad, the right peak extends to the side of negative delays which is misleading for the detection of causal direction, especially when we would consider sums

over lags like in Eqs. (19) and (20). In (i) the weaker coupling is not detected. Subfigure (iii) shows CMSI as in Eq. (8) for the smallest possible time comb. The peaks are much sharper and allow for a correct detection of coupling directions and delays because these embedding time combs quite well select momentary information only. Note that the time combs fulfill the conditions given in Eq. (13), and thus this special CMSI is called MSIT, defined by Eq. (14). In subfigure (iv) CMSI measures the sorting of the two next values leading to a greater sensitivity in coupling strength. However, now no longer first of all momentary information is used which makes the peaks broader. Nevertheless, the information used is quasimomentary, as it comes from two adjacent samples. In subfigure (v) the choice of time combs corresponds to transfer sorting entropy as the condition in CMSI is performed only from one time series. It yields an asymmetric coupling measure, but for the logistic map it fails to detect the right delay $\tau_{x\to y}$, and the peaks are broad. Generally the variance decreases for larger samples. Note that for $\tau = 1$ this is similar to STE and TERV introduced in Refs. [8] and [9], respectively (Fig. 1). Thus the figures illustrate that best results for coupling analysis are obtained by the MSIT approach. As we did argue above this is due to considering first of all momentary information.

Figure 2(b) shows the maximum MSIT value for positive delays, $I^{\mathrm{MSIT}}_{\max,\tau>0} = \max_{\tau>0} I^{\mathrm{MSIT}}$ (solid line), and negative delays, $I^{\mathrm{MSIT}}_{\max,\tau<0} = \max_{\tau<0} I^{\mathrm{MSIT}}$ (dashed line). The coupling strength $c_{y\to x} = 0.2$ is constant, and $c_{x\to y}$ is varied. $I^{\mathrm{MSIT}}_{\max,\tau>0}$ describes the coupling direction $x \to y$, and $I^{\mathrm{MSIT}}_{\max,\tau<0}$ that of the inverse direction, $y \to x$. Note that $I^{\mathrm{MSIT}}_{\max,\tau<0}$ is nearly constant which well reflects that $c_{y\to x}$ is constant. On the other hand, $I^{\mathrm{MSIT}}_{\max,\tau>0}$ is varied, however, it does not increase monotonously with $c_{x\to y}$. In general, such a monotonous relation cannot be expected as the coupled systems may undergo several bifurcations.

At $c_{x\to y} = c_{y\to x} = 0.2$, MSIT reproduces the equal coupling strength. This equality is expected because now the whole system is completely symmetric with respect to the subsystems. At larger values of $c_{x\to y}$ the system undergoes some bifurcations, and beyond $c_{x\to y} = 0.4$ the directions can again be well distinguished.

This analysis is rather robust if we add dynamical noise $\xi_{*,t}$, where $*$ stands for x or y. We take noise that is iid, zero mean normally distributed with standard deviation $\sigma_* \lesssim 0.1$ [i.e., $\xi_{*,t} \sim \mathcal{N}(0,\sigma_*)$]. In each iteration we first map $(x_{t-1}, y_{t-\tau_{x\to y}}) \to (x_t, y_t)$ according to Eq. (25) and then we add the noise

$$x_t \to x_t + \sigma_x \xi_{x,t}, \quad y_t \to y_t + \sigma_y \xi_{y,t}. \tag{26}$$

The random variables $\xi_{x,t}$ and $\xi_{y,t}$ are independent, and we take the same noise levels, $\sigma = \sigma_x = \sigma_y$. For increasing σ, however, both values $I^{\mathrm{MSIT}}_{\max,\tau<0}$ and $I^{\mathrm{MSIT}}_{\max,\tau>0}$ decrease, the latter more than the first. For $\sigma \gtrsim 0.1$ longer time series are necessary to get significant results.

B. Unidirectionally coupled Lorenz systems

Next, we demonstrate the importance of a suitable time comb in two delay-coupled Lorenz systems, defined for $i, j = 1,2$ by

$$\begin{aligned}
\dot{X}_i(t) &= \sigma\,(Y_i(t) - X_i(t)) + \sigma_{\mathrm{dyn}}\xi_{X,i},\\
\dot{Y}_i(t) &= r\,X_i(t) - Y_i(t) - X_i(t)\,Z_i(t)\\
&\quad + \sum_{j\neq i} c_{ij}\,Y_j^2(t-\tau_{ij}) + \sigma_{dyn}\xi_{Y,i},\\
\dot{Z}_i(t) &= X_i(t)\,Y_i(t) - b\,Z_i(t) + \sigma_{\mathrm{dyn}}\xi_{Z,i}.
\end{aligned} \tag{27}$$

We take standard parameters, $\sigma = 10$, $b = 8/3$, $r = 28$ coupling strength $c_{12} = 1$, $c_{21} = 0$, and coupling delay $\tau_{12} = 1$. Hence, system $i = 1$ operates autonomously, and couples unidirectionally to system $j = 2$. The integration step was 0.001, but we recorded only every 20th point leading to a sampling period $\Delta t = 0.02$. Dynamical noise was added each integration step, with $\xi_{X,i}$, $\xi_{Y,i}$, $\xi_{Z,i}$ being completely independent and standard normally distributed [i.e., each being $\sim \mathcal{N}(0,1)$]. For our coupling analysis we take, $x_t = Z_1(t\,\Delta t)$ and $y_t = Z_2(t\,\Delta t)$. The corresponding power spectra have a strong periodic component, with a period of about $38\,\Delta t$.

Figure 3(b) shows MSIT for varying embedding delays $d = 1, \ldots, 30$, which is in units of the sampling period Δt. The embedding dimensions are fixed at $M_x = M_y = 3$, $M_{z_x} = M_{z_y} = 2$. The time combs $\vartheta_x = \vartheta_y = \{-2d, -d, 0\}$ cover a period of $2d$ which we refer to as *time scale*. Remember that for these time combs CMSI is the same as MSIT, see Eqs. (13) and (14). The bottom plot shows the cumulated MSIT over all time scales, $\sum_d I^{\mathrm{MSIT}}$. For comparison, we also show mutual information (MI) which is obtained from adaptive binning according to the $(n/8)$ quantiles, $n = 1,2,\ldots,7$, of the empirical one-dimensional distribution of the series. For time scales $2d = 20$ and $2d = 38$, the MSIT functions are plotted separately in Fig. 3(c).

Obviously the choice of proper time combs is crucial here. If the time scale is similar to the period of the dominant harmonic component (i.e., for $2d \approx 38$), $I^{\mathrm{MSIT}}(\tau)$ much better reflects the right coupling at time delay $\tau = 50 \times 0.02$ equal to the coupling delay $\tau_{12} = 1$. However, for most time lags and time scales MSIT exceeds our significance level, which might be misleading. The reason is that, in general, time series with strong (nearly) periodic components require higher-dimensional embedding covering many periods to ensure that, first of all, momentary information is used in the coupling analysis. This is unfeasible for short time series. The lack of momentary information in the shuffle surrogates leads to an underestimation of the autodependencies and thus yields too small values of the decision threshold $I^{\mathrm{MSIT}*}$. Nevertheless, if we add dynamical noise (Fig. 4), the right coupling delay and coupling direction is detected very well for all time scales. Thus, dynamical noise improves our coupling analysis.

Figure 5 shows results for the influence of observational noise. For this we first generate the whole time series $\{x_t^*\}$ and $\{y_t^*\}$ as in Fig. 3 (without any noise) and then we add each sampling period $\Delta t = 0.02$ some observational noise (iid) $\xi_{\mathrm{obs},*,t} \sim \mathcal{N}(0,\sigma_{\mathrm{obs},*})$, where $* = x, y$, with noise level $\sigma_{\mathrm{obs},*} = 20$. Thus we consider in our coupling analysis the series $\{x_t = x_t^* + \xi_{\mathrm{obs,x};t}\}$ and $\{y_t = y_t^* + \xi_{\mathrm{obs,y};t}\}$.

Obviously, observational noise does not improve our coupling analysis, which is in contrast to dynamical noise. This is not surprising, as observational noise does not increases

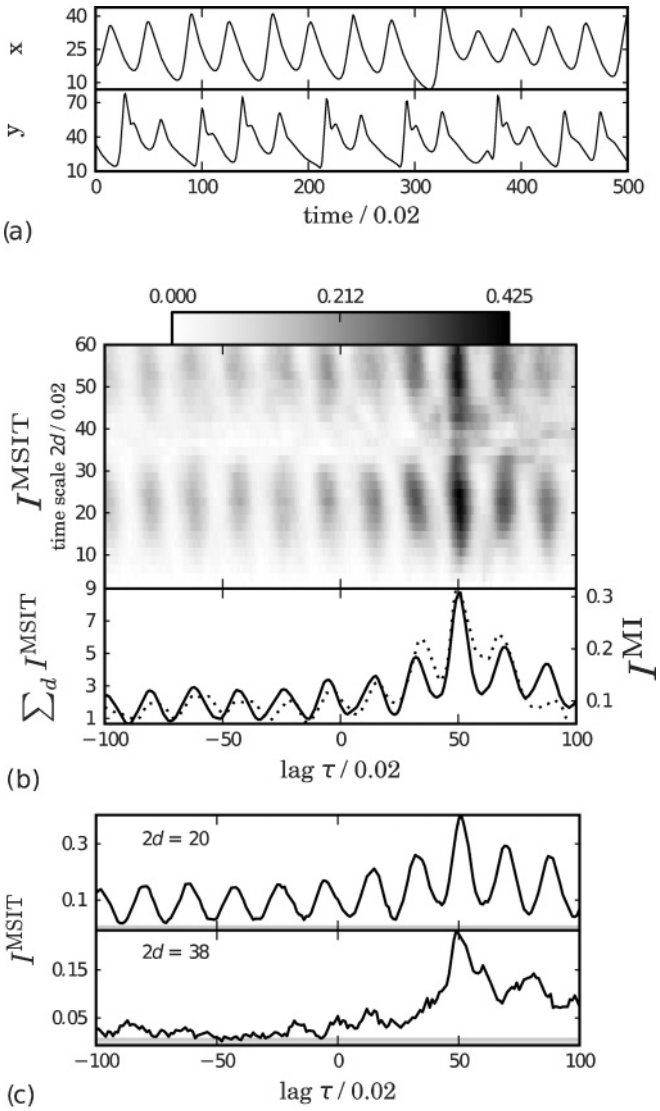

FIG. 3. Results for the Lorenz system, Eq. (27), for dynamical noise level $\sigma_{\mathrm{dyn}} = 0$. (a) The signals x and y represent Z components of two unidirectionally delay-coupled Lorenz systems as in Eq. (27), with $T = 1500$ samples (only 500 shown), using sampling period $\Delta t = 0.02$. (b) MSIT function, gray encoded, for varying time combs $\vartheta_x = \vartheta_y = \{-2d, -d, 0\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \{-2d, -d\}$ for different embedding delays $d = 1, 2, \ldots, 30$. Lower plot shows the cumulated MSIT over all time scales (solid), and mutual information (MI, dotted). (c) MSIT for time scales $2d = 20$ (top) and $2d = 38$ (bottom) where the gray band denotes the 95% significance level $\mathrm{MSIT}^* = I^{\mathrm{CMSI}*}$, Eq. (24).

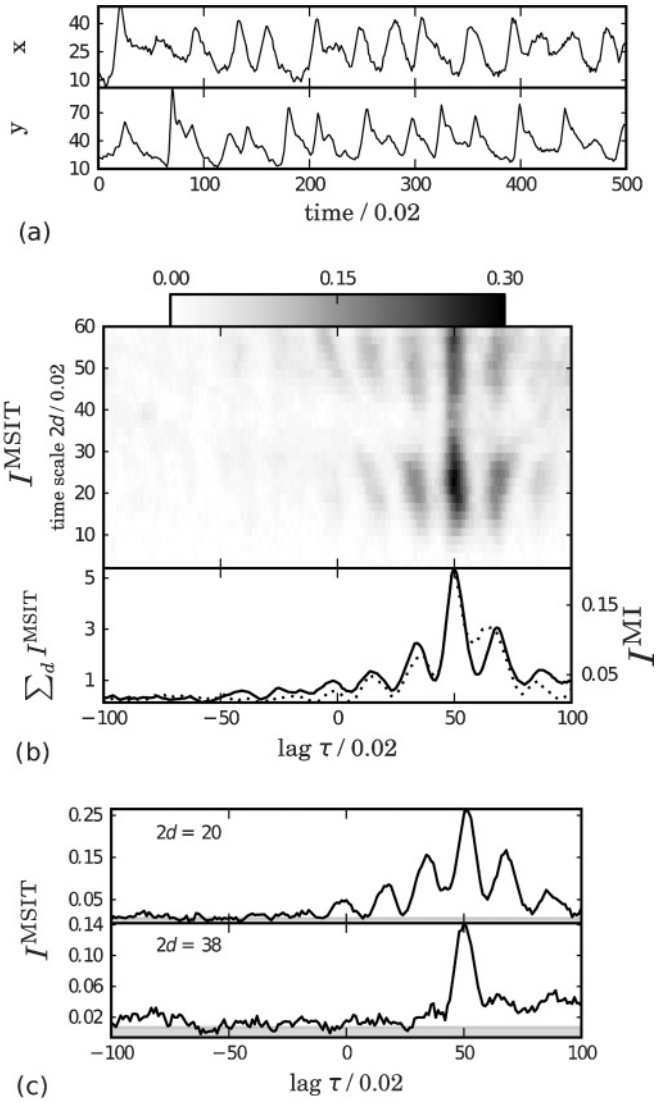

FIG. 4. Results as in Fig. 3, but with dynamical noise $\sigma_{\mathrm{dyn}} = 0.3$.

momentary dynamical information which is used in our analysis. Observational noise only increases irrelevant information in a measurement.

C. Real-world data from the climate system

We apply now our MSIT approach of coupling analysis to real-world data. In detail, we analyze the sea surface temperature anomalies of the North Atlantic at positions **x** (40 °N, 50 °W), and **y** (60 °N, 20 °W). The data are taken from the Hadley Center SST Reanalysis [29] from 1870 to 2010. The path of the North Atlantic Current (NAC), the time series from positions **x** and **y**, and MSIT on different time scales $2d$ are shown in Fig. 6. MSIT reveals couplings at two lags, one at about four years only on larger time scales, and one with about zero lag on all time scales.

The first one might be explained as caused by the NAC. The path from **x** to **y** has a distance of about 5000 . . . 6000 km and the detected coupling delay of about four years implies an average velocity of about 5 cm/s which is of the same order as the mean velocity measured by drift experiments [30]. This long-range mechanism with a large delay is only detected in the MSIT functions for larger time scales, as high frequency noise destroys the coupling information on smaller time scales.

The maximum at zero lag is probably caused by atmospheric processes affecting both **x** and **y** like the North Atlantic Oscillation (NAO) which are typically on a time scale of about one month. It is known [31] that the Atlantic surface winds are strongest during winter when they average near 5 m/s from the eastern United States across the Atlantic onto northern Europe. For distances 5000 . . . 6000 km between **x** and **y** this yields an expected coupling delay of 0.38 . . . 0.46 months. As

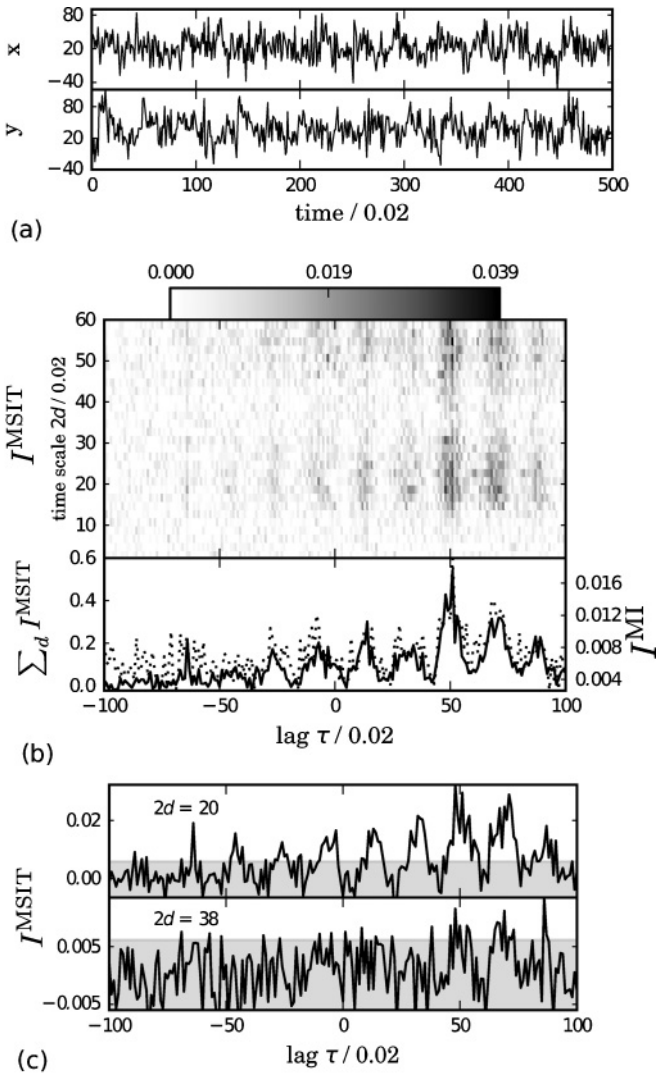

FIG. 5. Results as in Fig. 3, but with observational noise. Remember that the MSIT (CMSI) estimator (23) can be negative as we take Grassbergers entropy estimator.

we have only monthly data, our MSIT function cannot resolve this delay, instead we get a peak at delay $\tau = 0$.

The correlation coefficient between **x** and **y** at lag zero is $\rho = -0.13$, implying an anticorrelation. (Note that MSIT does not distinguish correlation from anticorrelation.) The correlation of the NAO index (data from Ref. [31]) and **x** is $\rho = +0.15$, between NAO and **y** it is $\rho = -0.24$ (anticorrelated). Thus the common effect of the NAO might be the cause of the anticorrelation between **x** and **y** at zero lag.

We note that the precise coupling delays are much better detected with MSIT than with MI and cross correlation (not shown, very similar to MI). Thus the condition leading to the restriction on momentary information is very helpful in determining the time delays of a physical coupling mechanism.

VI. OUTLOOK

The proposed momentary sorting information transfer is able to detect the direction and the coupling delays of information exchange in coupled systems. We expect that it

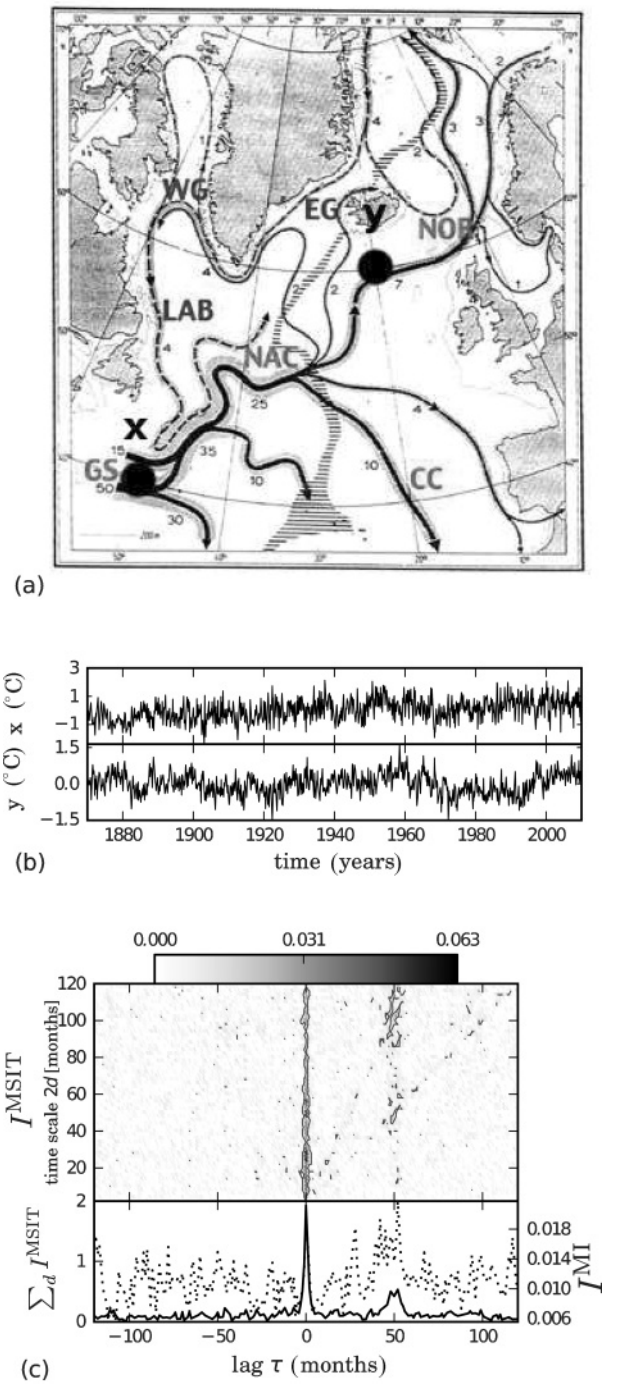

FIG. 6. (a) Path of the North Atlantic Current (NAC) with the locations marked as **x** and **y** (from Ref. [32]). (b) Anomalized time series (1680 months). (c) MSIT time scales, significant estimates according to the 99.9% significance level (calculated from 10 000 samples) are above the contours. MSIT is shown for varying time combs $\vartheta_x = \vartheta_y = \{-2d, -d, 0\}$, $\vartheta_{z_x} = \vartheta_{z_y} = \{-2d, -d\}$ for $d = 1, \ldots, 60$ covering time scales up to ten years. Lower plot shows cumulated MSIT over all time scales (solid) and MI (dotted, estimated using adaptive binning with eight quantiles). Our MSIT coupling analysis detects two significant coupling delays: The first is about 50 months, it could be related to the surface water current from **x** and **y** with mean velocity of about 5 cm/s. The second is less than one month, it could be related to the Atlantic surface winds from **x** and **y** with mean velocity in winter of about 5 m/s.

works even for rather short time series, which needs, however, a more detailed analysis that is beyond the scope of this paper.

Another interesting aspect would be the application nonstationary cases. We expect that our method is robust against some trends which is due to considering an ordinal pattern only. This would make the analysis appropriate for sliding window analysis. However, this is part of future work.

Moreover, the method is based on suitable time combs for delay embedding. The main rule here is to find a low-dimensional embedding such that first of all momentary information in each series contributes to mutual sorting information. Our examples give some hints how to do so, however, a more systematic approach is part of future work as well.

ACKNOWLEDGMENTS

We thank J. Kurths, J. Heitzig, and J. F. Donges for useful comments. J. R. has been financially supported by the German Academic Foundation, the project ("ECONS") of the Leibniz association, and the (DFG) research group 1380 "HIMPAC".

[1] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[2] M. Paluš, V. Komarek, Z. Hrncir, and K. Sterbova, Phys. Rev. E **63**, 046211 (2001).
[3] A. Kaiser and T. Schreiber, Physica D **166**, 43 (2002).
[4] M. Paluš and A. Stefanovska, Phys. Rev. E **67**, 055201 (2003).
[5] M. Paluš and M. Vejmelka, Phys. Rev. E **75**, 056211 (2007).
[6] S. Frenzel and B. Pompe, Phys. Rev. Lett. **99**, 204101 (2007).
[7] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. E. McClintock, and H. Kantz, Phys. Rev. Lett. **100**, 084101 (2008).
[8] M. Staniek and K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008).
[9] D. Kugiumtzis, e-print arXiv:1007.0357v1.
[10] J. T. Lizier *et al*., J. Comput. Neurosci. **30**, 85 (2011).
[11] J. Runge, Master thesis, Humboldt University Berlin, Berlin, 2010.
[12] H. Marko, Kybernetik **3**, 128 (1966).
[13] H. Marko, IEEE Trans. Comput. **21**, 1345 (1973).
[14] J. L. Massey, in *Proceedings of the International Symposium on Information Theory and its Application, Waikiki, Hawaii, Nov. 27–30*, 1990.
[15] G. Kramer, Phd thesis, ETH Zürich, Zürich, 1998.
[16] H. H. Permuter *et al*., e-print arXiv:0802.1383v1.
[17] P. O. Amblard and O. J. J. Michel, J. Comput. Neurosci. **30**, 7 (2011).
[18] P. O. Amblard and O. J. J. Michel, e-print arXiv:0911.2873v3.
[19] P. M. Saco *et al*., Physica A **389**, 5022 (2010).
[20] Ch. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[21] Ch. Bandt *et al*., Nonlinearity **15**, 1595 (2002).
[22] J. M. Amigó *et al*., Physica D **77**, 77 (2005).
[23] J. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That*. Springer Series in Synergetics (Springer, New York, 2010).
[24] W. J. McGill, Psychometrica **19**, 97 (1954).
[25] A. Jakulin and I. Bratko, e-print arXiv:cs/0308002v1.
[26] B. Pompe, in *Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics Studies in Computational Finance*, edited by A. S. Soofi and L. Cao (Kluwer Acadmic Publishers, Boston, Dordrecht, London, 2002).
[27] P. Grassberger, Phys. Lett. A **128**, 369 (1988).
[28] P. Grassberger, e-print arXiv:physics/0307138v2.
[29] N. A. Rayner *et al*., J. Geophys. Res. **108**, 4407 (2003).
[30] D. M. Fratantoni, J. Geophys. Res. **106**, 22067 (2001).
[31] J. W. Hurrell and C. Deser, Journal of Marine Systems **78**, 28 (2009).
[32] G. Dietrich and K. Kalle, *General Oceanography* (Wiley, New York, 1975).