Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy

Jakob Runge,[1,2] Jobst Heitzig,[1] Vladimir Petoukhov,[1] and Jürgen Kurths[1,2,3]
[1]*Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany*
[2]*Department of Physics, Humboldt University, 12489 Berlin, Germany*
[3]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom*
(Received 18 February 2012; published 21 June 2012)

Multivariate transfer entropy (TE) is a model-free approach to detect causalities in multivariate time series. It is able to distinguish direct from indirect causality and common drivers without assuming any underlying model. But despite these advantages it has mostly been applied in a bivariate setting as it is hard to estimate reliably in high dimensions since its definition involves infinite vectors. To overcome this limitation, we propose to embed TE into the framework of graphical models and present a formula that decomposes TE into a sum of finite-dimensional contributions that we call decomposed transfer entropy. Graphical models further provide a richer picture because they also yield the causal coupling delays. To estimate the graphical model we suggest an iterative algorithm, a modified version of the PC-algorithm with a very low estimation dimension. We present an appropriate significance test and demonstrate the method's performance using examples of nonlinear stochastic delay-differential equations and observational climate data (sea level pressure).

DOI: 10.1103/PhysRevLett.108.258701 PACS numbers: 89.70.Cf, 02.50.−r, 05.45.Tp, 89.70.−a

Inferring causal relations between processes when only some time series of measurements are given is of very general interest in many fields of science, especially when the underlying mechanisms are poorly understood. Consider the climate system, where new observational technologies lead to an abundance of observational data and yet even the interaction mechanism between such important systems as the El Niño southern oscillation and the Indian monsoon system is not fully understood [1,2]. Also the inference of functional brain connectivity in neuroscience and the discovery of causal relationships in economic data are of great importance [3,4]. One can view the processes as nodes of a graph where the links denote interactions. Now the problem is whether these links can be interpreted as ''causal'' interactions. Towards a causal interpretation, statistical methods need to be able to (1) measure associations, (2) measure also time delays, and (3) exclude other influences [5]. Usually it is impossible to exclude all other influences, especially if the system cannot be experimentally manipulated, and therefore the term ''causal'' can only be understood to be meant relative to the system under study, i.e., in our framework the processes that comprise the nodes of the graph. The main caveats against a causal interpretation are the possibility of spurious causalities due to indirect influences or common drivers. For example, if a causal interaction is given by $X \to Y \to Z$, a bivariate analysis would give a significant link between X and Z that is detected as being only indirect in a multivariate analysis including Y.

Several existing measures fulfill requirements (1) and (2), e.g., classical product-moment correlation analysis or synchronization [6], if delays are taken into account. Some methods are also able to exclude spurious causalities due to other processes. In classical statistics this problem can be addressed by assuming some class of linear or nonlinear parametric models and estimating the model that best explains the time series. A similar model-based test is Granger causality [7] and the framework of methods using phase coherences [8]. There also exist methods that are less restrictive in that they do not require to specify a certain functional form, but still require *a priori* knowledge of the distributions or general structure of the model [3]. Therefore, the absence of a link in a graph inferred with these model-based methods does not imply that the processes are not interacting since only a certain class of causal mechanisms has been tested.

In the framework of information theory [9], no such model-based restrictions exist because interactions are viewed as transfers of information. The most widely applied information-theoretic functional is (conditional) mutual information (CMI) [4], especially in the form of (multivariate) transfer entropy (TE) [10]. As formally defined below, the common definition of TE leads to the problem of estimating infinite-dimensional densities, which is commonly called the ''curse of dimensionality'' and strongly affects the reliability of causal inference as demonstrated in Fig. 1.

In this Letter we show how this severe limitation can be overcome by embedding TE into the framework of graphical models. In this framework we derive a formula that decomposes TE into a sum of finite-dimensional contributions that we call decomposed transfer entropy (DTE). This can drastically reduce the estimation dimension and enables a much more reliable estimation of TE even from real-world data. Furthermore, graphical models provide a richer

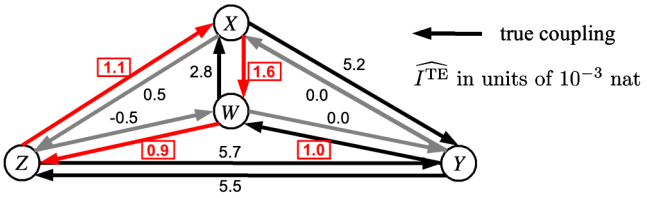

FIG. 1 (color online). Estimated TE between all subprocesses X, Y, Z, W for an example process [a discrete-time sampling of Eq. (5) that will be analyzed later] with true coupling structure given by the black arrows. If we truncate the infinite past vectors used in the common definition of TE at 15 lags, the estimation dimension is 61. The true TE of all gray and red links is zero, since they do not represent direct couplings. So of the four estimates of comparable size marked in red boxes, only the link $Y \to W$ is a correctly identified coupling.

picture of causal interactions because they also yield the causal coupling delays. We demonstrate the advantages of this approach on a model and on observational climate data of mean sea level pressure in Eastern Europe.

To derive DTE, we introduce the following notation: Given a stationary multivariate discrete-time stochastic process $\mathbf{X}$, we denote its uni- or multivariate subprocesses $X, Y, Z, W, \ldots$ and the values at time t as $\mathbf{X}_t, X_t, \ldots$. Their pasts are defined as $\mathbf{X}_t^- = (\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots)$ and $X_t^- = (X_{t-1}, X_{t-2}, \ldots)$ and with a slight abuse of notation X_t^- can be formally treated as a subset of $\mathbf{X}_t^-$. Now the TE $I^{\mathrm{TE}}_{X\to Y} = I(X_t^-; Y_t | \mathbf{X}_t^- \setminus X_t^-)$ is the reduction in uncertainty about Y_t when learning the past of X_t, if the rest of the past of $\mathbf{X}_t$, given by $\mathbf{X}_t^- \setminus X_t^-$, is already known. There are two infinite-dimensional parts in TE: X_t^- and $\mathbf{X}_t^- \setminus X_t^-$. We address the first by decomposing TE into contributions of individual lags of X via the chain rule (for detailed derivations, see the Supplemental Material [11]),

$$I(X_t^-; Y_t | \mathbf{X}_t^- \setminus X_t^-) = \sum_{\tau=1}^{\infty} I(X_{t-\tau}; Y_t | \mathbf{X}_t^- \setminus X_t^-, X_{t-\tau}^-). \tag{1}$$

Now the decisive step to escape the still infinite dimension of the condition in each term is done by utilizing the theory of graphical models [12,13]. Looking at Fig. 2(a), the main idea stems from a Markov property that relates the separation of nodes in the graph to the conditional independences in the process ([13], theorem 4.1). It implies that

$$I(X_{t-\tau}; Y_t | \mathbf{X}_t^- \setminus X_t^-, X_{t-\tau}^-) = I(X_{t-\tau}; Y_t | \mathcal{S}_{Y_t, X_{t-\tau}}), \tag{2}$$

for a certain finite subset $\mathcal{S}_{Y_t,X_{t-\tau}} \subset \mathbf{X}_t^- \setminus X_t^- \cup X_{t-\tau}^-$ of the conditions [see Fig. 2(a)]. Once the time series graph (defined below) of the process is known, suitable sets $\mathcal{S}_{Y_t,X_{t-\tau}}$ can be determined from it and the TE can be estimated using only low-dimensional densities. The remaining infinite sum can be truncated at some finite τ^* since the terms typically decay exponentially with τ,

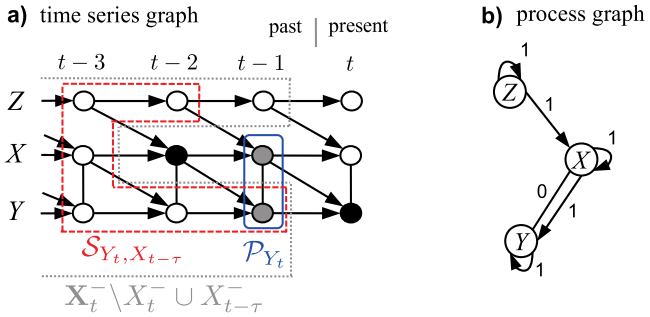

FIG. 2 (color online). Causal interactions in a multivariate process $\mathbf{X}$. (a) Time series graph (see definition in text). The boxes mark different sets of conditions for the CMI between $X_{t-\tau}$ at $\tau = 2$ and Y_t (marked by the black dots): the terms in Eq. (1) use the infinite set $\mathbf{X}_t^- \setminus X_t^- \cup X_{t-\tau}^-$ (gray dotted line). But only the finite set $\mathcal{S}_{Y_t,X_{t-\tau}}$ (red dashed line) is needed to satisfy Eq. (2). $\mathcal{S}_{Y_t,X_{t-\tau}}$ must be chosen so that it separates the skipped infinite conditions $(\mathbf{X}_t^- \setminus X_t^- \cup X_{t-\tau}^-) \setminus \mathcal{S}_{Y_t,X_{t-\tau}}$ from Y_t in the graph (for a formal definition of paths and separation, see [13]). The even smaller set of parents $\mathcal{P}_{Y_t}$ (blue box, gray nodes) separates Y_t from the past of the whole process $\mathbf{X}_t^- \setminus \mathcal{P}_{Y_t}$, which is used in the algorithm to estimate the graph. (b) Process graph, which aggregates the information in the time series graph for better visualization (labels denote the lags).

$$I^{\mathrm{TE}}_{X\to Y} \approx I^{\mathrm{DTE}}_{X\to Y} = \sum_{\tau=1}^{\tau^*} I(X_{t-\tau}; Y_t | \mathcal{S}_{Y_t,X_{t-\tau}}), \tag{3}$$

with $\tau^\star$ chosen as the smallest τ for which the estimated remainder is smaller than some given absolute tolerance (see the Supplemental Material for details [11]). This can improve the estimation of TE considerably as compared to the direct estimation, although the sets $\mathcal{S}_{Y_t,X_{t-\tau}}$ and the resulting dimensions can still be large.

But how can the required time series graph be estimated? As depicted in Fig. 2(a), each node in that graph represents a subprocess X at a certain time t. Nodes $X_{t-\tau}$ and Y_t are connected by a directed link $X_{t-\tau} \to Y_t$ pointing forward in time if and only if $\tau > 0$ and

$$I(X_{t-\tau}; Y_t | \mathbf{X}_t^- \setminus \{X_{t-\tau}\}) > 0, \tag{4}$$

and nodes X_t and Y_t are connected by an undirected contemporaneous link (visualized by a line) if and only if $I(X_t; Y_t | \mathbf{X}_{t+1}^- \setminus \{X_t, Y_t\}) > 0$ [13]. If $Y \neq X$, we say that the link $X_{t-\tau} \to Y_t$ represents a coupling at lag τ, while for $Y = X$ it represents an autodependency at lag τ, so the time series graph gives a richer picture of causal interrelations than TE, including the causal coupling delays and contemporaneous links as well, while the total influence of X on Y can be measured by TE. Figure 2(a) shows the time series graph for a small example of a multivariate process with couplings and auto-dependencies, which can be summarized in a process graph [see Fig. 2(b)].

As shown in Fig. 2(a), there is an important difference between the conditions $\mathcal{S}_{Y_t,X_{t-\tau}}$ and the (direct) parents of Y_t, denoted $\mathcal{P}_{Y_t} = \{X_{t-\tau}: X \in \mathbf{X}, X_{t-\tau} \to Y_t\}$. This means

in particular that the terms in Eq. (3) can be nonzero also for lags τ for which there is no link in the graph and therefore they cannot be used directly as a measure of lag-specific coupling strength, a topic we explore further in a separate paper. Since $\mathcal{P}_{Y_t}$ (or any subset of $\mathbf{X}_t^-$ that contains $\mathcal{P}_{Y_t}$) will separate Y_t from $\mathbf{X}_t^- \setminus \mathcal{P}_{Y_t}$ in the graph, $I(X_{t-\tau}; Y_t | \mathcal{P}_{Y_t} \setminus \{X_{t-\tau}\}) > 0$ defines links equivalent to Eq. (4), which is used in the following algorithm to estimate the time series graph.

The estimation of graphical models is very similar to the problem of inferring the directed acyclic graph [14] of a set of random variables. To this end, the idea of the PC-algorithm (named after its inventors [14]) is to iteratively unveil the links by testing for conditional independence between all possible pairs of nodes conditioned on iteratively more conditions and testing all combinations among them. Thereby, the dimension stays as low as possible in every iteration step. We adapt this algorithm for time series and propose some modifications to speed up the performance. The algorithm starts with no *a priori* knowledge about the links and iteratively learns the set of parents for each Y. The union of parents together with the contemporaneous links then comprises the graph.

For every Y, first we estimate the MIs $I(X_{t-\tau}; Y_t)$ and initialize the preliminary parents $\tilde{\mathcal{P}}_{Y_t} = \{X_{t-\tau}{:}X \in \mathbf{X}, 0 < \tau \leq \tau_{\max}, I(X_{t-\tau}; Y_t) > 0\}$. This set also contains indirect links that are now iteratively removed by testing whether the CMI between Y_t and each $X_{t-\tau} \in \tilde{\mathcal{P}}_{Y_t}$ conditioned on the incrementally increased set of conditions $\tilde{\mathcal{P}}_{Y_t}^{n,i} \subseteq \tilde{\mathcal{P}}_{Y_t}$ vanishes. In the outer loop, iterate n over increasing number of conditions, starting with some $n_0 > 0$. In the inner loop, iterate i through all combinations of picking n nodes from $\tilde{\mathcal{P}}_{Y_t}$ to define the conditions $\tilde{\mathcal{P}}_{Y_t}^{n,i}$ in this step, and estimate the CMI $I(X_{t-\tau}; Y_t | \tilde{\mathcal{P}}_{Y_t}^{n,i})$ for all $X_{t-\tau} \in \tilde{\mathcal{P}}_{Y_t}$. After each step, the nodes $X_{t-\tau}$ with $I(X_{t-\tau}; Y_t | \tilde{\mathcal{P}}_{Y_t}^{n,i}) = 0$ are removed from $\tilde{\mathcal{P}}_{Y_t}$ and the i-iteration stops if all possible combinations have been tested. If the cardinality $|\tilde{\mathcal{P}}_{Y_t}| \leq n$, the algorithm converges, else, increase n by one and iterate again.

Once the parents of each process are known, the same algorithm for $\tau = 0$ can be used to infer the contemporaneous neighbors $\mathcal{N}_{Y_t} = \{X_t{:}X \in \mathbf{X}_t, X_t - Y_t\}$, where now undirected links are removed if $I(X_t; Y_t | \mathcal{P}_{Y_t}, \tilde{\mathcal{N}}_{Y_t}^{n,i}, \mathcal{P}(\tilde{\mathcal{N}}_{Y_t}^{n,i})) = 0$. The free parameters of this method are the maximum lag $\tau_{\max}$, the initial number of conditions n_0, the significance threshold I^* to determine whether $I(X_{t-\tau}; Y_t | \tilde{\mathcal{P}}_{Y_t}^{n,i}) > 0$, and any parameters of the estimator of CMI. A suitable significance test and details of the algorithm are discussed in the Supplemental Material [11]. For the estimation of CMI we use a k-nearest neighbor estimator [15,16].

Now, we demonstrate the method first on a model system and second on real data. In theory, CMI measures any statistical associations and the Markov property is fulfilled by a large class of model systems ([13], condition (S)). CMI has been applied to time-continuous chaotic models in [15] and in [17] it has been shown that dynamical noise actually helps in the coupling analysis. The latter work demonstrated that for a precise inference of coupling lags one needs enough dynamical noise in X that can be measured in Y. Therefore, we demonstrate the method with a system of four stochastic delay-differential equations and couple them linearly and nonlinearly, also in the stochastic terms. This system of Ornstein-Uhlenbeck processes can be interpreted as nonlinearly coupled particles, each fluctuating in its harmonic potential:

$$\begin{aligned}
\dot{X} &= -0.5X(t) + 0.6W(t-4)\eta_X(t), \\
\dot{Y} &= -0.9Y(t) - 1.0X(t-2) + 0.6Z(t-5) + \eta_Y(t), \\
\dot{Z} &= -0.7Z(t) - 0.5Y(t-6) + \eta_Z(t), \\
\dot{W} &= -0.8W(t) - 0.4Y(t-3)^2 + 0.05Y(t-3) + \eta_W(t),
\end{aligned} \tag{5}$$

with independent unit variance white noise processes $\eta_\cdot(t)$. Thus, we have a bidirectional feedback $Y \rightleftharpoons Z$ and a feedback loop $X \to Y \to W \to X$ in which $Y \to W$ is nonlinear and a stochastic coupling $W \to X$. This system would be hard to analyze using model-based approaches, especially given short sample lengths as used here ($T = 1000$). Throughout the analysis we have used a fixed significance threshold $I^* = 0.015$ and a maximum lag $\tau_{\max} = 15$. Figure 3 shows the iteration steps. Step (0.0) gives the result of an analysis using only MI, the first step of the algorithm. We would wrongly infer that Y drives X, X drives W, X and Z are coupled and conclude on a long-range memory process within Y and Z at $\tau \approx 12$. Also the precise coupling delays are buried under a broad range of significant lags. The CMI values for each lag can

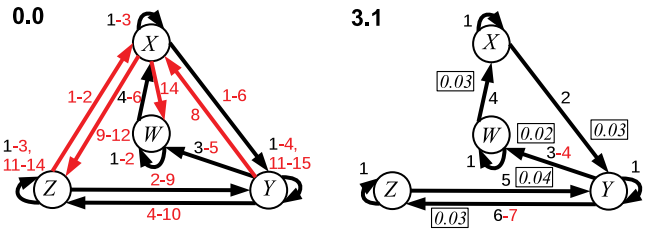

FIG. 3 (color online). Iterative steps in the analysis of model Eq. (5), time series length $T = 1000$, integration time step $dt = 0.01$, sampling interval $\Delta s = 100$, MI and CMI estimated using $k = 100$. The label $(n.i)$ indicates the iteration step. (0.0) shows the MI process graph. With an initial $n_0 = 3$ the next step (3.0) (see the Supplemental Material [11]) with three conditions is already almost identical to the converged graph in step (3.1), where the values in the boxes denote the estimate $\widehat{I^{\mathrm{DTE}}}$ of TE via Eq. (3) with τ^* chosen such that $I(X_{t-\tau^*-1}; Y_t | \mathcal{S}_{Y_t, X_{t-\tau^*-1}})$ has declined below significance. Incorrect links and lag labels are in red.

be represented via lag functions as shown in the Supplemental Material [11]. The algorithm proceeds as follows. Considering the estimation of the parents of X, we choose an initial $n_0 = 3$ and start with the links with the weakest MI, Z at $\tau = 1, 2$, conditioned on the three preliminary parents with the largest MI, $\tilde{\mathcal{P}}^{3,0}_{X_t} = \{X_{t-1}, X_{t-2}, W_{t-4}\}$. As these links are due to Y which drives Z and with one step delay also X via W (see process graphs in Fig. 3), $I(Z_{t-\tau}; X_t|\tilde{\mathcal{P}}^{3,0}_{X_t})$ for $\tau = 1, 2$ vanishes and we remove these links from $\tilde{\mathcal{P}}_{X_t}$. The second weakest link, $Y_{t-8} \to X_t$, is indirectly mediated via W and thus $I(Y_{t-8}; X_t|\tilde{\mathcal{P}}^{3,0}_{X_t})$ vanishes and we remove this link as well. Next, we check the coupling from W and the autodependency from X_{t-2} conditioned on the same nodes, whereupon the links from W_{t-5} and X_{t-2} vanish. Now the set $\tilde{\mathcal{P}}_{X_t}$ is smaller than $n = 3$, i.e., all possible conditions have been tested, and the algorithm for X converges already in the second step. The analysis for the remaining subprocesses is discussed in the Supplemental Material [11]. Apart from some inaccuracy in the coupling lags, which is due to the continuous nature of the system, this yields the correct graph. While the dimension for the direct estimation of TE is $D = \tau_{\max} \cdot 4 + 1 = 61$, the dimension using the graph and Eq. (3) is between 5 and 24 (depending on $\mathcal{S}_{Y_t, X_{t-\tau}}$). It is interesting to compare TE with the model parameters in Eq. (5): $Z \to Y$ has a higher $\widehat{I^{\mathrm{DTE}}}$ than $X \to Y$, while the corresponding parameters are 0.6 and 1.0, respectively. TE as a measure of the total influence between two processes, therefore, cannot be simply related to the parameters of the underlying model.

To study the performance of the algorithm using a significance test from shuffle surrogates, we ran 1000 numerical experiments with the class of nonlinear discrete stochastic models. The results are shown in the Supplemental Material [11] and indicate that, compared to MI, the rate of false positives is strongly reduced, while correct linear and nonlinear links are well detected. Further, we found that a larger initial n_0 significantly speeds up the performance.

We now analyze a climatological data set of daily mean sea level pressure anomalies in the winter months of 1997–2003 [18] at four locations in Eastern Europe indicated on the map in Fig. 4. First, we give the statistical analysis and then provide a climatological interpretation. From MI in step (0.0), one would infer an almost fully connected graph with a broad range of lags (see lag functions in the Supplemental Material [11]). For example, we found a strong $Y \to Z$ "link" and a "link" $W \to X$ with a delay of about 2 days. The iteration using an initial $n_0 = 2$ converges in the third step (2.1). The link $Y \to Z$ is now much weaker (even below our significance threshold), because a lot of the shared entropy is due to the common driver W. Even more apparent, the $W \to X$ link vanishes due to the condition on Y. Note that the contemporaneous links $X - Z$ and $W - Y$ are possibly due to spatial proximity. This causal picture of a south-eastward "flow of entropy" is consistent with the dynamical processes governing the lower and middle atmosphere circulation in the considered area. One usually observes a superposition of westerly winds with traveling extratropical cyclones that traverse the area and whose trajectories are regulated by the aforementioned westerlies [19]. Consistent with the causal lags of 1 or 2 days these processes act on short daily time scales. We note that this causal structure might change in the high troposphere where the influence of quasistationary planetary waves and the Ferrel cell might noticeably modify an above-mentioned causality. This analysis underlines the importance of knowing coupling delays for physical interpretations and serves as a first step to study more complex systems like the Indian monsoon and El Niño.

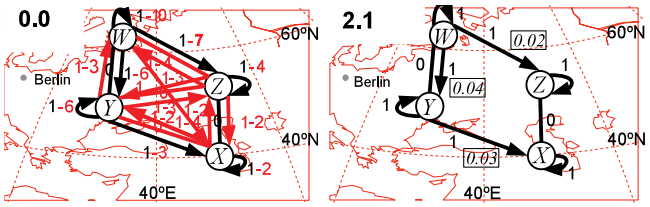

FIG. 4 (color online). Analysis of time series of mean sea level pressure with $T = 1268$ using the same threshold as before and $\tau_{\max} = 10$. (0.0) depicts the process graph for MI. The iteration converges in step (2.1); step (2.0) and the lag functions are shown in the Supplemental Material [11]. Labels are as in Fig. 3.

To conclude, we have addressed the problem of inferring causal relations among multiple processes within the model-free information-theoretic framework. For the commonly used TE, we have derived an exact decomposition formula that enables an estimation using finite vectors. Although the estimation dimension is drastically reduced, it can still be quite high. The graphical model, on the other hand, can be very efficiently estimated by a modification of the PC-algorithm with much lower dimension. It gives a complementary picture of causal relations among multiple processes with precise coupling delays and contemporaneous links. Thus, time resolved causal relations are much easier to estimate than TE as a measure of the total influence between two processes. Now the limiting factor in the construction of "causal" networks from brain or climate data [2] is no longer the network size, but only the maximum degree, i.e., the number of parents.

We acknowledge the support by the German Environmental Foundation (DBU), the DFG Grant No. KU34-1, the DFG research group 1380 "HIMPAC," and the BMBF project "PROGRESS". We thank J. Zscheischler for helpful comments.

[1] G. Pant and K. Kumar, *Climates of South Asia* (John Wiley & Sons, New York, 1997).

[2] J. Donges, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **87**, 48007 (2009).
[3] C. Hiemstra and J. D. Jones, J. Finance **49**, 1639 (1994); D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**, 144103 (2008); J. Prusseit and K. Lehnertz, Phys. Rev. E **77**, 041914 (2008); W.-X. Wang, R. Yang, Y. C. Lai, V. Kovanis, and C. Grebogi, Phys. Rev. Lett. **106**, 154101 (2011).
[4] K. Hlavackova-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[5] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, England, 2000).
[6] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, England, 2003).
[7] C. Granger, Econometrica **37**, 424 (1969); J. Geweke, in *Handbook of Econometrics II*, edited by Z. Griliches and M. D. Intriligator (Elsevier, Amsterdam, 1984), Chap. 19.
[8] B. Schelter, M. Winterhalder, R. Dahlhaus, J. Kurths, and J. Timmer, Phys. Rev. Lett. **96**, 208103 (2006); L. Sommerlade, M. Eichler, M. Jachan, K. Henschel, J. Timmer, and B. Schelter, Phys. Rev. E **80**, 051128 (2009); J. Nawrath, M. Carmen Romano, M. Thiel, I. Z. Kiss, M. Wickramasinghe, J. Timmer, J. Kurths, and B. Schelter, Phys. Rev. Lett. **104**, 038701 (2010).
[9] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2006).
[10] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000); M. Paluš, V. Komárek, Z. Hrnčíř, and K. Štěrbová, Phys. Rev. E **63**, 046211 (2001).
[11] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.108.258701 for formal derivations, remarks on the algorithm, numerical experiments, and a comprehensive discussion of the model and application analyses.
[12] S. L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series, Vol. 16 (Clarendon, Oxford, 1996); R. Dahlhaus, Metrika **51**, 157 (2000).
[13] M. Eichler, Probab. Theory Relat. Fields **1**, 233 (2012).
[14] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search* (MIT, Cambridge, MA, 2000).
[15] S. Frenzel and B. Pompe, Phys. Rev. Lett. **99**, 204101 (2007).
[16] M. Vejmelka and M. Paluš, Phys. Rev. E **77**, 026214 (2008).
[17] B. Pompe and J. Runge, Phys. Rev. E **83**, 051122 (2011).
[18] T. Ansell *et al.*, J. Climate **19**, 2717 (2006).
[19] E. Palmen and C. W. Newton, *Atmospheric Circulation Systems: Their Structure and Physical Interpretation* (Academic, New York, 1969).