Measuring causality by taking the directional symbolic mutual information approach*

Chen Gui(陈 贵)†, Xie Lei(谢 磊), and Chu Jian(褚 健)

Institute of Cyber-System and Control, Zhejiang University, Hangzhou 310027, *China*

(Received 5 September 2012; revised manuscript received 24 October 2012)

We propose a novel measure to assess causality through the comparison of symbolic mutual information between the future of one random quantity and the past of the other. This provides a new perspective that is different from the conventional conceptions. Based on this point of view, a new causality index is derived that uses the definition of directional symbolic mutual information. This measure presents properties that are different from the time delayed mutual information since the symbolization captures the dynamic features of the analyzed time series. In addition to characterizing the direction and the amplitude of the information flow, it can also detect coupling delays. This method has the property of robustness, conceptual simplicity, and fast computational speed.

Keywords: causality measure, Bandt and Pompe method, mutual information, transfer entropy

PACS: 89.70.Cf, 02.50.–r, 05.45.Tp, 87.19.lo **DOI:** 10.1088/1674-1056/22/3/038902

1. Introduction

Synchronization and related phenomena exist in physical to biological systems.[1,2] To understand these phenomena, the time evolutions of the systems are studied. Hence, understanding the couplings between systems reduces to an investigation of the observable data presented as time series or variables. Tools used to unveil the nonlinear interdependency include mutual information,[3–5] non-linear synchrony measures,[6] and redundancy.[7] However, it is often more important to quantify the coupling direction or causality between systems in many fields.[10] The causality defined by Wiener[8] was firstly introduced into the analyses of observable time series by Granger[9] in the context of linear regression models. Plenty of work has been done based on the Granger causality (see Ref. [10] and references therein).

The definition of causality in philosophy is not so easy. From the practical point of view, it concerns the cause-and-effect relationship between events, objects, states of affairs, and variables. It is usually presumed that the cause chronologically precedes the effect. The probability theory seems to be the most used mathematical language in measuring or quantifying the relationship and strength. The probabilistic notion of causality was nicely described by Suppes.[11] That contribution was based on the three conditions of existence of causality given by Selltiz *et al.*[12] Current statistics consider the causal inference as one of the most important problems after the work of Granger.[9]

As an alternative, a non-parametric approach assessing the nonlinear causality termed as the transfer entropy (TE) was proposed by Schreiber.[13] In Ref. [10], the TE was shown to be equivalent to conditional mutual information (CMI).[14] The equivalence between the TE and the Granger causality for Gaussian variables was proved by Barnett *et al.*[15] Another measure identical to the TE and the CMI, called the partial mutual information (PMI), was proposed by Frenzel and Pompe.[14]

Let $x_t = x(t)$ and $y_t = y(t)$ $(t = 1, \ldots, T)$ denote the time series from systems X and Y. By considering the information flow from X to Y, the TE and the CMI can be expressed in terms of Shannon entropy $H(x) = -\sum p(x) \log p(x)$ as

$$\begin{aligned} T_{X\to Y} &= H(\boldsymbol{x}_t, \boldsymbol{y}_t) + H(y_{t+\tau}, \boldsymbol{y}_t) - H(\boldsymbol{y}_t) - H(y_{t+\tau}, \boldsymbol{x}_t, \boldsymbol{y}_t) \\ &= I(y_{t+\tau}; \boldsymbol{x}_t | \boldsymbol{y}_t), \end{aligned} \tag{1}$$

or directly in terms of the distribution functions as

$$T_{X\to Y} = \sum p(y_{t+\tau}, \boldsymbol{x}_t, \boldsymbol{y}_t) \log \frac{p(y_{t+\tau}|\boldsymbol{x}_t, \boldsymbol{y}_t)}{p(y_{t+\tau}|\boldsymbol{y}_t)}, \tag{2}$$

where $\boldsymbol{x}_t = \{x_{t+v_1}, \ldots, x_{t+v_{M_x}}\}$ and $\boldsymbol{y}_t = \{y_{t+v_1}, \ldots, y_{t+v_{M_y}}\}$ denote delay embedding vectors, $y_{t+\tau}$ is the τ-future value of time series $y(t)$, $\boldsymbol{v}_x = \{v_1, \ldots, v_{M_x}\}$ and $\boldsymbol{v}_y = \{v_1, \ldots, v_{M_y}\}$ are the embedding time combs, and M_x and M_y (≥ 2) are time-comb orders. Normally, $v_{m+1} - v_m = d$ for all $m = 1, \ldots, M_x$ (M_y), and d is a positive integer. However, in general, non-uniformly spaced time combs can be chosen if there is a good reason to do so.[16] Related research can be found in Refs. [17] and [18]. Then the directionality index $D_{XY} = T_{X\to Y} - T_{Y\to X}$ quantifies the relationship between X and Y.

According to Frenzel and Pompe's PMI, the TE or CMI can be interpreted as the mutual information of $y_{t+\tau}$ and $\boldsymbol{x}_t$, which is not contained in $\boldsymbol{y}_t$. This, however, leads to the following problem. For causality assessment, some data points in $\boldsymbol{y}_t$ come from future relative to other part of the data in

*Project supported by the National Natural Science Foundation of China (Grant No. 60904039).

†Corresponding author. E-mail: gui.sam.chen@gmail.com

© 2013 Chinese Physical Society and IOP Publishing Ltd

$\boldsymbol{x}_t$, while the TE discards their mutual information. This may cause causality information loss.

Most of the existing estimation methods (e.g., Refs. [19] and [20]) have great demands on fine-tuning of parameters, and are highly sensitive to noise contributions and are computationally expensive. Symbolic transfer entropy (STE)[21] was proposed for solving the estimation problem and has been used in several areas.[22,23] It should be noted that the STE is not a direct analogue of the TE (see Refs. [24] and [25] using the same method on the CMI). The disadvantage of the STE was analyzed in Ref. [26], where an improved measure named transfer entropy on rank vectors (TERV) was provided. However, the TERV requires high-dimensional embedding vectors, which will lead to memory restrictions, as pointed in Ref. [27]. A detailed comparison of them was presented in Ref. [28].

To overcome the above problems, we suggest estimating the mutual information by adopting Bandt and Pompe's approach.[27] We will present how this permutation information theoretic measure, called symbolic mutual information (SMI), leads to a new point of view on causality quantification. By normalizing the SMI to $[0,1]$, our causality index ranges in $[-1,1]$, which offers a desired measure representing both the direction and the amplitude of the information flow.

2. Directional symbolic mutual information

A synopsis of permutation information approach can be described as follows. Assume that the time-comb orders for $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ are identical, i.e., $M_x = M_y = M$. Then the M real valued data in $\boldsymbol{x}_t$ can be sorted in an increasing order. In this way, we have $M!$ permutations that are considered in this study as possible order types. In the case of equal amplitude, we numerically break the equalities by adding small random perturbations. It is clear that every $\boldsymbol{x}_t$ is uniquely mapped onto one of the $M!$ possible permutations. When each permutation is considered as a symbol $\hat{x}_t$, then the reconstructed trajectory is represented by a symbol sequence. We estimate joint and conditional probabilities of the sequence with the relative frequencies of the symbols. Based on this coarse-grained symbolization method, Staniek and Lehnertz gave their STE as

$$\begin{aligned} T^{\mathrm{S}}_{X\to Y} &= H(\hat{x}_t,\hat{y}_t)+H(\hat{y}_{t+\tau},\hat{y}_t) \\ &\quad - H(\hat{y}_t)-H(\hat{y}_{t+\tau},\hat{x}_t,\hat{y}_t). \end{aligned} \tag{3}$$

Given symbol sequences $\{\hat{x}_t\}$ and $\{\hat{y}_t\}$, we define symbolic mutual information as

$$I^{\mathrm{S}}_{X,Y} = \sum p(\hat{x}_t,\hat{y}_t)\log\frac{p(\hat{x}_t,\hat{y}_t)}{p(\hat{x}_t)p(\hat{y}_t)}, \tag{4}$$

where the sum runs over all symbols. Thus $I^{\mathrm{S}}_{X,Y}$ is given in bit. In Ref. [25], the author named a similar measure as mutual sorting information. We prefer symbolic as sorting values of rank vector naturally is one symbolization method among different methods, e.g., fine-grained symbolization considering higher-order difference information or coarse-grained symbolization as a binary sequence. With the permutation entropy (PE) of $\{\hat{x}_t\}$ denoted as $H(\hat{x}_t) = -\sum p(\hat{x}_t)\log p(\hat{x}_t)$, we have $I^{\mathrm{S}}_{X,Y} = H(\hat{x}_t)+H(\hat{y}_t)-H(\hat{x}_t,\hat{y}_t) \geq 0$. Note that the PE is the metric entropy referred to the dynamic features[26] and the resulting SMI is not a direct analogue of the mutual information.

As a correlation measure, the MI is normalized into $[0,1]$ in lots of literature. It is clear that the PE has lower bound 0 and upper bound $\log M!$. With the assumption that $x(t)$ and $y(t)$ have the same delay-embedding dimension M, the SMI in Eq. (4) ranges from 0 to 1 if the log is with base $M!$. A similar approach was used by Wang *et al.*[29]

Recall the definition of the Wiener–Granger causality, it contains two critical components: the cause occurs before its effect; the effect can be better predicted using the past information of its cause than the past information of itself only. The first component seems to be apodeictic, but the comparison of predictability is not an easy task especially when employing model-free approaches for nonlinear processes. Based on, but not entirely using the previous causality definition, we state our causality as:

1) the cause occurs before its effect;

2) the future of effect contains more information about the past of its cause than that contained by the future of cause about the past of its effect.

According to this, we use asymmetry or time ordering MI between the bivariate random processes to assess their relationship (which one is causing the other). Since the MI is the reduction in the uncertainty of one random process,[30] our causality can also be interpreted as the comparison of mutual predictability. Considering the SMI defined in Eq. (4), we define the directional SMI through introducing a time delay

$$I^{\mathrm{S}}_{X\to Y}(d) = \sum p(\hat{x}_t,\hat{y}^d_t)\log\frac{p(\hat{x}_t,\hat{y}^d_t)}{p(\hat{x}_t)p(\hat{y}^d_t)}, \tag{5}$$

where $\hat{y}^d_t$ is the symbol obtained by sorting time-delayed vector $\boldsymbol{y}^d_t = \{y_{t+d+v_1},\ldots,y_{t+d+v_M}\}$. An arrow is used here for expressing the time order relationship. We then define a causality index (CI) as

$$C_{X,Y} = I^{\mathrm{S}}_{X\to Y} - I^{\mathrm{S}}_{Y\to X}, \tag{6}$$

which is expected to attain positive values for unidirectional couplings with X as the driver and negative values for Y driving X. It is expected to be 0 for symmetric bidirectional couplings. Notice that $C_{X,Y}$ is a measure with direction and $C_{Y,X}$. But we will get the same conclusion no matter whether $C_{X,Y}$ or $C_{Y,X}$ is used. If $I^{\mathrm{S}}_{X\to Y}$ and $I^{\mathrm{S}}_{Y\to X}$ can be normalized into $[0,1]$, then causality index $C_{X,Y}$ can be normalized into $[-1,1]$, which contains both coupling strength and direction information.

The estimation of our SMI and STE boils down to the calculation of the probability of occurrence of a rank vector such as $\hat{x}_t$. Measuring the computational efficiency by the dimensions of rank vectors, we list the results for STE and SMI in Table. 1. It is clearly that the sum of the dimension of the SMI is much smaller than the other two.

Table 1. Computational efficiencies (dimensions) of three measures.

STE	TERV	SMI
$2M_x+6M_y+2$	$2M_x+6M_y+2\tau$	$2M_x+2M_y$

3. Properties demonstration

To demonstrate the use of our causality index, we study three examples: (i) a coupled nonlinear stochastic system; (ii) a pair of delay coupled Lorenz systems; and (iii) electroencephalogram (EEG) recordings from rats.

First, the following three coupled maps are considered:[31,32]

$$\begin{aligned} x_t &= 3.4x_{t-1}(1-x_{t-1}^2)\,\mathrm{e}^{-x_{t-1}^2}+0.4w_{1t},\\ y_t &= 3.4y_{t-1}(1-y_{t-1}^2)\,\mathrm{e}^{-y_{t-1}^2}\\ &\quad +0.5x_{t-1}y_{t-1}+0.4w_{2t},\\ z_t &= 3.4z_{t-1}(1-y_{t-1}^2)\,\mathrm{e}^{-z_{t-1}^2}\\ &\quad +0.3y_{t-1}+0.5x_{t-1}^2+0.4w_{3t}, \end{aligned} \tag{7}$$

where w_{1t}, w_{2t}, and w_{3t} are standard Gaussian noise processes. The correct directed causal effects are $X\to Y$, $Y\to Z$, and $X\to Z$. We show the estimated CIs with different time series lengths in Fig. 1. Positive values of $C_{X,Y}, C_{Y,Z}$, and $C_{X,Z}$ are obtained. This is in agreement with the theoretical result.

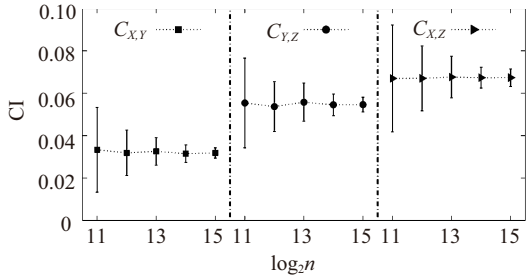


Fig. 1. Mean CI (error bars denote the standard deviations) from 100 realizations of the coupled nonlinear stochastic system versus time-series length $\log_2 n$. The embedding time comb is a continuous integral vector with dimension 4, and $d=2$.

For illustrating the coupling delay detection of our method, we consider two coupled Lorenz systems with a time delay described by the following differential equations:

$$\begin{aligned} \dot{x}_i &= \sigma(y_i-x_i),\\ \dot{y}_i &= rx_i-y_i-x_iz_i+c_jy_j^2(t-\tau_j),\\ \dot{z}_i &= x_iy_i-bz_i, \end{aligned} \tag{8}$$

where $i,j=1,2$. Here we use standard parameters $\sigma=10$, $r=28$, and $b=8/3$. To demonstrate the impact of the time delay, we choose different coupling delays with $\tau_1=12$ and $\tau_2=5$. We fix all the parameters except the coupling from system 1 to 2, c_1, and vice versa c_2. The conventional MI as a dependence measure can also reveal the coupling time delay. Our SMI has a similar function through varying time lag d defined in Eq. (4). Thus, in this example, the causality analysis is expected to represent inverse results at two coupling delays when both coupling strengths are significant. We consider time series $\{y_{1,2}\}$ with sample rate 0.2 s and data length 5000 points. The directional SMI $I^{\mathrm{S}}_{1\to2}$ and $I^{\mathrm{S}}_{2\to1}$ shown in Figs. 2(a) and 2(b) reach the maximum values when time lag d equals to the corresponding coupling delay $\tau_{1,2}$.

In Fig. 2(c), we illustrate the causality index $C_{1,2}$ with time lag d equals to 5 and 12, respectively. Although the coupling strength c_2 is fixed at 0.2, both the directional SMI $I_{2\to1}$ and the causality index have slight fluctuation, which seems to be the effect of varying c_1. The causality index with time lag $d=5,12$ reveals the correct relationships. To unveil the coupling strength at different time lag, we calculate the difference of the maximum values of $I^{\mathrm{S}}_{1\to2}$ and $I^{\mathrm{S}}_{2\to1}$, which is shown in Fig. 2(d) as a function of c_1. It gives the correct trend with bias when $c_1=c_2$ and $c_1=0.35$. For comparison, we set $\tau_1=\tau_2=5$ s, the causality indices with time lag $d=5$ give similar results.

For an additional validation, we test the sensitivity of our approach in detecting the causality change with time. We set $\tau_1=\tau_2=5$ s. In the beginning, $c_2=0.1$ and $c_1=0.4$, then the values of c_1 and c_2 are exchanged at time 2000 s (corresponding to the data point 10000). We use the moving window technique. The window size is 5000 samples (1000 s), and the overlap is 4500 samples (900 s). The results are shown in Fig. 3, which provides a useful indication of the variation in time of the directional SMI and CI.

Our third example consists of five pairs of rat EEG signals on the right (R) and the left (L) cortical intracranial electrodes of male adult WAG/Rjj rats, which have been studied in Refs. [8] and [33]. They are referenced to an electrode placed in the cerebellum, filtered between 1 Hz and 100 Hz, and digitized at 200 Hz. The five-pair data are labeled from (i) to (v). The first pair (i) is a normal EEG, and all the others contain spike discharges which are the landmark of epileptic activity. The directional SMIs from left to right $I^{\mathrm{S}}_{\mathrm{L}\to\mathrm{R}}(d)$ are estimated at time lag d from 1 to 5, and vice versa $I^{\mathrm{S}}_{\mathrm{R}\to\mathrm{L}}(d)$. We construct delay-embedding vectors with length 4 using the embedded time comb as the continuous natural number. figure 4 shows the causality indices $C_{\mathrm{L,R}}$ with different time lags d. Our results are qualitatively consistent with the results reported in Ref. [33], where the synchronization level rank (iv) > (v) > (ii) > (i) > (iii) was indicated. Usually epilepsy is

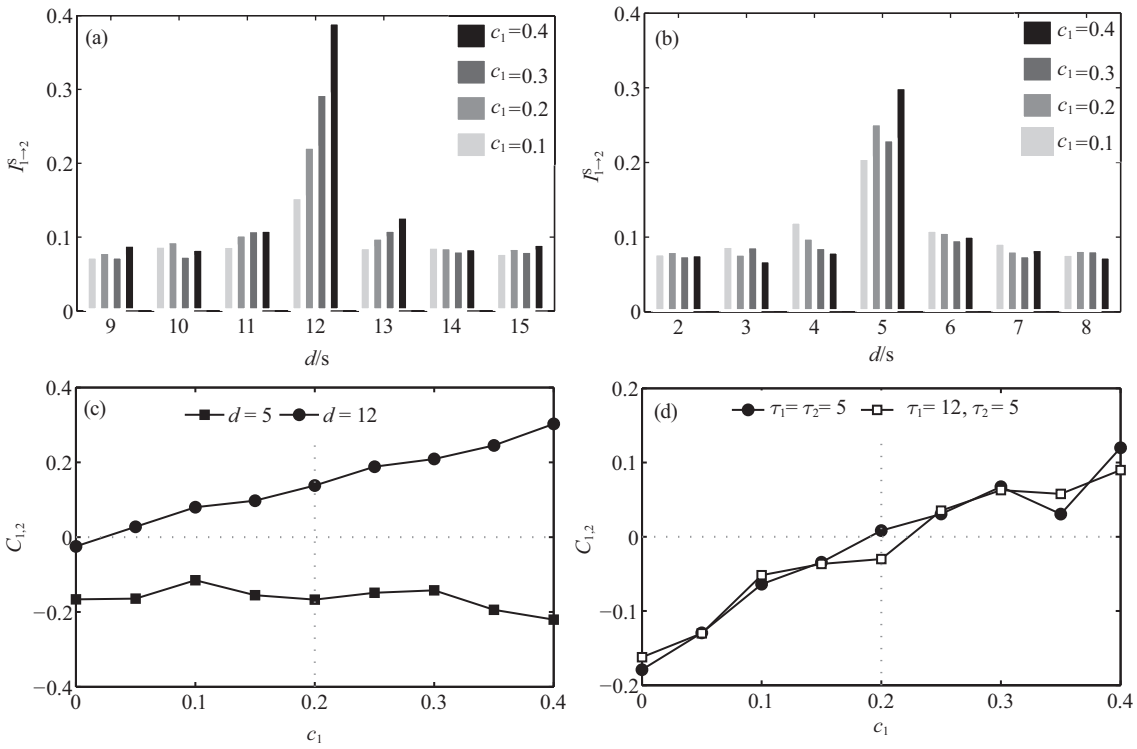


Fig. 2. (color online) Causality results of coupled Lorenz systems. (a) Directional SMI $I^{\rm S}_{1\to2}$ versus delayed time lag d, with coupling strength c_1 varying from 0.1 to 0.4 while $c_2 = 0.2$; (b) directional SMI $I^{\rm S}_{2\to1}$ as in panel (a); (c) causality index $C_{1,2}$ versus c_1 ($c_2 = 0.2$), with time lag $d = 5, 12$; (d) causality index $C_{1,2}$ ($d = 5$) versus c_1 ($c_2 = 0.2$) when both coupled time delays are 5 s (line with circle), and the line with square denotes $\{I^{\rm S}_{1\to2}(12) - I^{\rm S}_{2\to1}(5)\}$ versus c_1 ($c_2 = 0.2$).

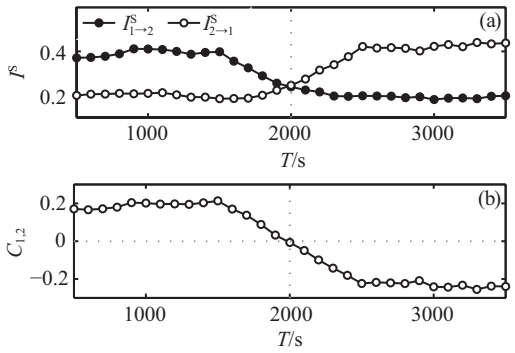


Fig. 3. The results of sensitivity for causality change. The window size is 5000 samples and the time labeled corresponds the center of a data window. (a) The directional SMI $I^{\rm S}_{1\to2}$ and $I^{\rm S}_{2\to1}$; (b) the corresponding evolution of $C_{1,2}$. The change can be detected accurately by the smooth transition.

related to an abnormal synchronization in the brain. A localized appearance of spikes can delimit a zone with abnormal discharges. On the contrary, if the spikes are observed over the whole set of electrodes, the abnormal synchronization is said to be global. For cases (ii) and (iii), spikes at the left and the right electrodes can be easily seen, which points towards a global synchronization behavior. However, in case (ii) they are well synchronized, but they are not in case (iii). For cases (iv) and (v), significant synchronizations can be observed, and our approach also reaches the results. A similar result with the same data analyzed using the nonlinear Granger causality method was reported in Ref. [34].

For comparison, the mutual information with different time lags is calculated, and the values of STE at different τ are obtained. The mutual information is estimated using the k-nearest neighbor method.[35] From Fig. 5(a), the difference of mutual information presents the information flow, but fails to reveal the coupling strength. The results of STE showed in Fig. 5(b) are similar to those in Fig. 4. However, for data sets (i), (ii), and (iii), the order of coupling strength is blurry owing to the bias. In our test with these data sets, the STE shows a slightly unstable numerical estimation.

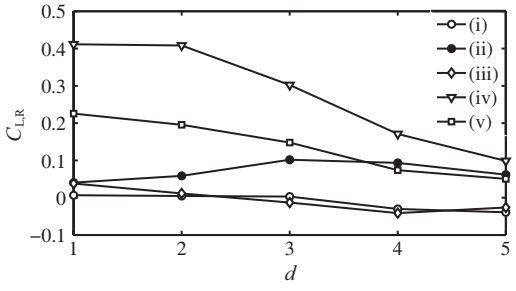


Fig. 4. Results of causality of the five pairs of rats' EEG data: CIs obtained by $I^{\rm S}_{\rm L\to R}(d) - I^{\rm S}_{\rm R\to L}(d)$ as a function of time lag d, positive values indicate causality L $\to$ R.

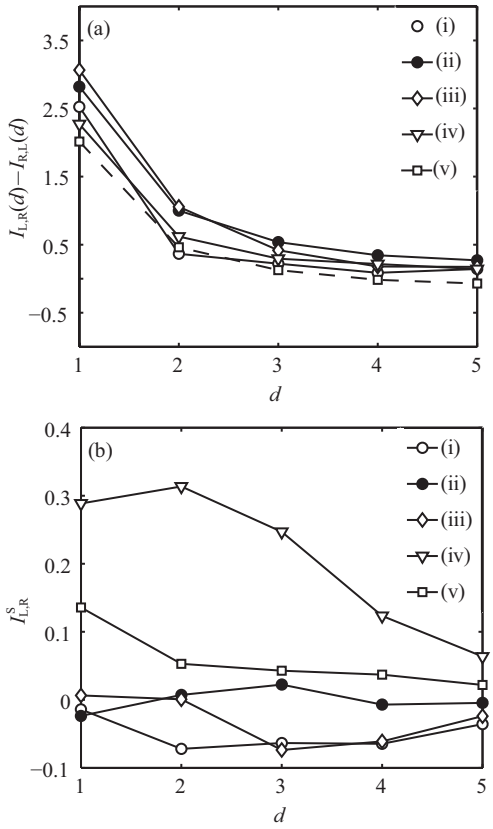


Fig. 5. Mutual information and STE for the rats' EEG data. (a) Mutual information denoted by $I_{\mathrm{L,R}}(d)$, the plot shows the difference of $I_{\mathrm{L,R}}(d) - I_{\mathrm{R,L}}(d)$; (b) calculated symbolic transfer entropy $T^{\mathrm{S}}_{\mathrm{L,R}}$ and the direction index $D^{\mathrm{S}}_{\mathrm{L,R}} = T^{\mathrm{S}}_{\mathrm{L}\to\mathrm{R}} - T^{\mathrm{S}}_{\mathrm{R}\to\mathrm{L}}$. The dimensions of used rank vectors are $M_x = M_y = 4$.

4. Conclusion

In summary, we suggested a new method for causality quantification. Through Bandt and Pompe's permutation information approach, the directional symbolic mutual information was proposed to measure the time ordering information flow. From this conception, a causality index was derived, which can also be normalized into $[-1, 1]$. Furthermore, for coupling systems with time delays, it can unveil the delay time. All the properties were illustrated by a stochastic multivariate coupled system, time delay coupled Lorenz systems, and real EEG data. These examples illustrated that the mutual information using Bandt and Pompe's method show different information from the proper mutual information. It is conceptional intuitive, robust, and fast, which is convenient for field data.

References

[1] Pikovsky A, Rosenblum M and Kurths J 2003 *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge: Cambridge University Press)
[2] Pereda E, Quiroga R Q and Bhattacharya J 2005 *Prog. Neurobiol.* **77** 1
[3] Bell C B 1962 *Ann. Math. Stat.* **33** 587
[4] Fraser A M and Swinney H L 1986 *Phys. Rev. A* **33** 1134
[5] Matsuda H 2000 *Phys. Rev. E* **62** 3096
[6] Quiroga R Q, Kraskov A, Kreuz T and Grassberger P 2002 *Phys. Rev. E* **65** 041903
[7] Paluš M, Albrecht V and Dvorak I 1993 *Phys. Lett. A* **175** 203
[8] Wiener N 1956 *The Theory of Prediction, Modern Mathematics for the Engineer* (New York: McGraw-Hill)
[9] Granger C 1969 *Econometrica* **37** 424
[10] Hlaváčcková-Schindler K, Paluš M, Vejmelka M, and Bhattacharya J 2007 *Phys. Rep.* **441** 1
[11] Suppes P 1970 *A Probilistic Theory of Causality* (Amsterdam: North-Holland Publishing Company)
[12] Selltiz C, Wrightsman L S and Cook S W 1959 *Research Methods in Social Relations* (New York: Holt, Rinehart andWinston)
[13] Schreiber T 2000 *Phys. Rev. Lett.* **85** 461
[14] Paluš M, Komarek V, Hrncir Z and Sterbova K 2001 *Phys. Rev. E* **63** 046211
[15] Barnett L, Barrett A B and Seth A K 2009 *Phys. Rev. Lett.* **103** 238701
[16] Frenzel S and Pompe B 2007 *Phys. Rev. Lett.* **99** 204101
[17] Zhang S Q, Jia J, Gao M and Han X 2010 *Acta Phys. Sin.* **59** 1576 (in Chinese)
[18] Xiao F H, Yan G R and Han Y H 2005 *Acta Phys. Sin.* **54** 550 (in Chinese)
[19] Kaiser A and Schreiber T 2002 *Physica D* **166** 43
[20] Lungarella M, Pitti A and Kuniyoshi Y 2007 *Phys. Rev. E* **76** 056117
[21] Staniek M and Lehnertz K 2008 *Phys. Rev. Lett.* **100** 158101
[22] Wang J and Yu Z F 2012 *Chin. Phys. B* **21** 018702
[23] Li N Q, Pan W, Yan L S, Luo B, Xu M F and Tang Y L 2012 *Chin. Phys. Lett.* **29** 030502
[24] Bahraminasab A, Ghasemi F, Stefanovska A, McClintock P V E and Kantz H 2008 *Phys. Rev. Lett.* **100** 084101
[25] Pompe B and Runge J 2011 *Phys. Rev. E* **83** 051122
[26] Kugiumtzis D 2010 arXiv: 1007.0357v1
[27] Bandt C and Pompe B 2002 *Phys. Rev. Lett.* **88** 174102
[28] Amigó J M 2010 *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and all that* (Berlin Heidelberg: Springer-Verlag)
[29] Wang Q, Shen Y and Zhang J Q 2005 *Physica D* **200** 287
[30] Cover T and Thomas J 2004 *Elements of Information Theory* (2nd edn.) (New Jersey: Wiley-Interscience)
[31] Chen Y H, Rangarajan G, Feng J F and Ding M Z 2004 *Phys. Lett. A* **324** 26
[32] Gourevitch B, Le Bouquin-Jeannes R and Faucon G 2006 *Biol. Cybern.* **95** 349
[33] Quiroga R Q, Kreuz T and Grassberger P 2002 *Phys. Rev. E* **66** 041904
[34] Marinazzo D, Pellicoro M and Stramaglia S 22008 *Phys. Rev. Lett.* **100** 144103
[35] Kraskov A, Stögbauer H and Grassberger P 2004 *Phys. Rev. E* **69** 066138