A STIMULI-RELEVANT DIRECTED DEPENDENCY INDEX FOR TIME SERIES

Payam Shahsavari Baboukani[1], *Sergios Theodoridis*[1], *Jan Østergaard*[1]

[1]Department of Electronic Systems, Aalborg University, Denmark
{pasba,sthe,jo}@es.aau.dk

ABSTRACT

Transfer entropy can to a certain degree assess the direction in addition to the strength of the couplings within dynamic time series. The greater the transfer entropy, the greater the strength of the dependency between time series. In this work, we are interested in quantifying the effect that a given time series (e.g., an external stimuli) has upon the coupling strength between other time series. Towards that end, we define a directed dependency index based on the difference of two causally conditioned transfer entropies. We then provide a lower bound for the dependency index, and demonstrate on synthetic data that this lower bound can be efficiently computed.

***Index Terms*—** Transfer entropy, directed dependecy, mutual information, intrinsic mutual information

1. INTRODUCTION

In time-series analysis, it is often desirable to assess the directed dependency between time series [1]. For example, the time series could be observations of physical or biological systems, where any possible directed dependencies in the data would mean that the systems are interacting with each other. One common application is the case of functional connectivity analysis of electroencephalography (EEG) signals [2, 3] where the observations obtained at different scalp electrodes are highly coupled.

Several time series analysis methods such as Granger causality [4, 5] and transfer entropy (TE) [1] have been proposed in order to quantify directed dependencies. While Granger causality is mainly used for analysing linear dependencies, TE is capable of determining both linear and non-linear interactions. In some cases, it is possible to distinguish between direct and indirect (implicit) couplings, where part of the information that two nodes share between them have been conveyed via a third node, see for example Fig. 1. By causally conditioning the estimation of TE between two nodes (say X and Y) on the third node (Z), it is possible to at least partially take the effect of indirect couplings into account [2]. In Fig. 1, nodes X and Y are directly coupled with the direction from X to Y, and they are also indirectly coupled via Z and S.

In this paper, we are interested in the general problem of quantifying to what degree a given time series is directly influencing the directed dependencies between other time series. Thus, referring to Fig. 1, we are interested in quantifying and computing a specific part of the information that describes the coupling between nodes X and Y. Specifically, we are only interested in the amount of information that exists within the directed coupling between X and Y and which is due to S but not due to Z. For example, consider the case, where a human subject is exposed to acoustic stimuli and the EEG response is being measured on the scalp. One might then be interested in observing possible changes in the directed dependency from data from one EEG electrode X to another electrode Y, and which is due to changes in the external acoustic stimuli S. However, it is not straight-forward to distinguish whether the changes in the couplings are due to the desired external stimuli, or whether it is due to another stimuli such as other EEG electrodes, visual stimuli or due to artefacts caused by muscle or eye movements or noise.

In [6], the information flow in the human auditory system due to external acoustic stimuli was quantified. In this case, the external stimuli A was first processed by the noisy acoustic environment yielding the output B, which then entered the human auditory system via the human eardrum. The output C of the human auditory system as a response to the stimuli A was then obtained through listening tests. The authors of [6] were then interested in how much information about the stimuli was lost in the human auditory system, and it was shown that this could be quantified via the difference $I(A;B) - I(A;C)$.

We note that [6] considered unidirectional measures and did not need to condition the mutual information. In our case, we consider directed information measures that are causally conditioned. This greatly complicates the problem. Specifically, we introduce a directed coupling index, which is defined as a difference between two causally conditional directed informations and which is on the form $I(X \to Y \| Z) - \min_{f(\hat{s}|s)} I(X \to Y \| Z, \hat{S})$, where the latter term includes a minimization over distributions. As was pointed out in [7], it is not known how to compute such a minimization for the case of continuous alphabet sources as is the case here. To circumvent this difficulty, we provide a lower bound to the dependency index, which involves a simpler minimization problem which can be directly computed using recent estimators for causally conditioned transfer entropy [8, 9].

2. BACKGROUND

2.1. Conditioning can increase the mutual information

It is well known that conditioning cannot increase entropy, i.e., $H(X|Y) \leq H(X)$. This is, however, not always the case for mutual information, where we could have that $I(X;Y|Z) \geq I(X;Y)$, as the following simple example shows. Let $Y = Z \oplus X$, where Z and X are mutually independent binary random variables, and where $\oplus$ denotes XOR. Clearly X and Y are mutually independent and $I(X;Y) = 0$. On the other hand, when conditioning upon Z, we obtain:

$$\begin{aligned} I(X;Y|Z) &= I(X;Z \oplus X|Z) = I(X;X|Z) \\ &= I(X;X) = H(X) \geq 0. \end{aligned} \tag{1}$$

Conditioning cannot increase the mutual information if $X - Z - Y$ forms a Markov chain, in that order [10]. To see this, we note that in this case $I(X;Y|Z) = 0$, whereas $I(X;Y) \geq 0$.

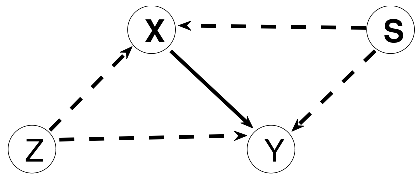

Fig. 1: The nodes X and Y are directly coupled with each other and indirectly via Z and S.

2.2. Synergistic, intrinsic, and shared information

As pointed out in e.g. [7], the information measured by the directed information methods and its conditional versions, can be interpreted[1] as consisting of intrinsic information and in addition there could potentially be shared and synergistic information. Thus, the directed information might overestimate the amount of information being conveyed between systems (or signals). Specifically, let Y_0 denote the current sample of the process Y and let X_{-1} denote the past sample of the process X. Then, $I(X_{-1}; Y_0)$ quantifies the intrinsic information from the past of X to the current Y in addition to the shared information, which already exist between the variables. Let us now assume that the shared information between X_{-1} and Y_0 is also contained in Y_{-1}. Then, we can simply remove the shared information by conditioning, i.e.: $I(X_{-1}; Y_0|Y_{-1})$. Whilst the shared information has been removed, we have now potentially introduced synergistic information, as shown in (1).

In [7, 15], the intrinsic mutual information $\tilde{I}(X;Y|Z)$ between between two random variables X and Y given another random variable Z was identified as an upper bound to the secret key agreement rate which is given as:

$$\tilde{I}(X;Y|Z) \triangleq \min_{f_{\hat{Z}|Z}(\hat{Z}|Z)} I(X;Y|\hat{Z}), \tag{2}$$

which can be interpreted as the minimum conditional mutual information between X and Y given any possible (probabilistic or deterministic) function of Z. In [7], the intrinsic directed dependency between X and Y was defined based on (2) as $\tilde{I}(X_{-1}; Y_0|Y_{-1})$.

For the case where Z has a discrete and finite alphabet, it was shown in [7] that the minimization in (2) can easily be computed. For the case where Z has a continuous alphabet, it is not clear how to compute (2). If $f_{\hat{Z}|Z}$ is constant, then it can be shown that $\tilde{I}(X;Y|Z) \leq I(X;Y)$ [7]. On the other hand, since one of the possible mappings from Z to $\hat{Z}$ is the identity function, it clearly follows that $\tilde{I}(X;Y|Z) \leq I(X;Y|Z)$.

2.3. Transfer Entropy

We are interested in determining the directed dependency from X to Y in the network of Fig. 1. We denote by X_i the random variable obtained by sampling the process X at the present time i. Furthermore, let X_i^- be the past of the process X up to but not including X_i. A similar notation applies to Y_i^- and Z_i^-. The transfer entropy from X to Y causally conditioned on Z is then defined as [16, 1]:

$$TE(X \to Y||Z) \triangleq I(X_i^-; Y_i|Y_i^-, Z_i^-). \tag{3}$$

[1] There exists a wealth of interpretations and decompositions of the conditional mutual information, cf. [11, 12, 13, 14]

3. QUANTIFYING STIMULI-RELEVANT DIRECTED DEPENDENCY (SRDD) INDEX

Let X and Y in Fig. 1 be the nodes of interest, let S be the external stimuli, and let Z be a (super) node that describes the consensus of the remaining nodes in a potentially complex network. Both S and Z could potentially affect both X and Y.

Let $I(X;Y|Z)$ be the conditional mutual information between X and Y given Z. Due to potential *synergistic* information, we might have that $I(X;Y|Z) > I(X;Y)$. Thus, $I(X;Y|Z)$ does not necessarily mean that we "remove" the information about Z, which is in either X or Y. However, consider now the difference:

$$I(X;Y|Z) - I(X;Y|Z,S). \tag{4}$$

In (4), any potential synergistic information, which is *only* due to conditioning upon Z, will vanish. In addition, new synergistic information due to conditioning upon S, and jointly upon S, Z could occur. To remove this effect, one can minimize over all possible functions of S as in (2). The intuition is now that by using the difference in (4), we remove the effect of Z, while at the same time avoiding overestimating the effect of S. Turning this into a problem that involves dynamic time series, we replace the conditional mutual information expressions in (4) with their causally conditioned transfer entropy counterparts [16].

Definition 1 (Stimuli-relevant Directed Dependency (SRDD) Index): We define the stimuli-relevant directed dependency index for $(X^N \to Y^N)$ given Z^N and with stimuli S^N as:

$$\begin{aligned} I_{S^N}(X^N \to Y^N|Z^N) &\triangleq \sum_{i=1}^{N} I_S(X^{i-1}; Y_i|Y^{i-1}, Z^{i-1}) \\ I_{S^N}(X^{i-1}; Y_i|Y^{i-1}, Z^{i-1}) &\triangleq I(X^{i-1}; Y_i|Y^{i-1}, Z^{i-1}) \\ &\quad - \min_{f(\hat{s}^N|s^N)} I(X^{i-1}; Y_i|Y^{i-1}, Z^{i-1}, \hat{S}^N). \end{aligned} \tag{5}$$

where $X^{i-1} = (X_1, X_2, \dots, X_{i-1})$.

When the processes X, Y, Z, and S are all jointly stationary but otherwise arbitrarily distributed, we ignore the sum in (5), and with a slight abuse of notation, we simply define the SRDD index to be given by:

$$\begin{aligned} I_S(X \to Y|Z) &\triangleq I(X_i^-; Y_i|Y_i^-, Z_i^-) \\ &\quad - \min_{f(\hat{s}|s)} I(X_i^-; Y_i|Z_i^-, Y_i^- \hat{S}), \end{aligned} \tag{6}$$

where the minimization is over all jointly *stationary* distributions $f(\hat{s}, s)$ satisfying $f(\hat{s}, s) = f(\hat{s}|s)f(s)$.

Lemma 1: Let X, Y, Z, S be jointly stationary but otherwise arbitrarily distributed random processes. Then, the SRDD index given in (6) can be lower bounded by:

$$\begin{aligned} I_S(X \to Y|Z) &\geq I(X_i^-; Y_i|Y_i^-, Z_i^-) \\ &\quad - \min_{\phi \subseteq S_i^-} I(X_i^-; Y_i|Y_i^-, Z_i^-, \phi), \end{aligned} \tag{7}$$

where the minimization is over all subsets of the past of S.

Proof: The proof follows immediately since the minimization in (7) is over a subset of the set from (6). ■

We note that (7) forms a tighter bound than if the minimization is simply replaced by the full amount of information, $\phi = S_i^-$. Moreover, the motivation for restricting the minimization to be over subsets of the sequence S_i^- is to provide the means for efficient estimation of a non-trivial lower bound.

It is possible to move the stimuli away from the conditioning, which can be useful depending upon the choice of estimator. In this case, one needs to minimize over a difference of mutual information terms as is shown below:

Lemma 2: The lower bound in Lemma 1 can be equivalently expressed as:

$$
\begin{aligned}
&I(X_i^-;Y_i|Y_i^-,Z_i^-) - \min_{\phi\subseteq S_i^-} I(X_i^-;Y_i|Y_i^-,Z_i^-,\phi) \\
&= \min_{\phi\subseteq S_i^-} \left[I(\phi;y_i|Y_i^-,Z_i^-,X_i^-) - I(\phi;y_i|Z_i^-,y_i^-) \right].
\end{aligned}
$$

Proof: We omit super- and subscripts when it is clear from context to simplify the notation. Assuming that the differential entropies are well defined, we establish the following equivalence by expanding the mutual information in terms of entropies:

$$
\begin{aligned}
&I(X;Y_i|Y_i^-,Z,S) = h(Y_i|Y_i^-,Z,S) - h(Y_i|Y_i^-,Z,S,X) \\
&= h(Y_i,Y_i^-,Z,S) - h(Y_i^-,Z,S) \\
&\quad - h(Y_i,Y_i^-,Z,S,X) + h(Y_i^-,Z,S,X) \\
&= h(S|Y_i,Y_i^i,Z) + h(S,Y_i,Y_i^-) - h(S|Y_i^-,Z) - h(Y_i^-,Z) \\
&\quad - h(S|Y_i,Y_i^-,X,Z) - h(Y_i,Y_i^-,X,Z) \\
&\quad + h(S|Y_i^-,Z,X) + h(Y_i^-,X,Z) \\
&= I(S;Y_i|Z,Y_i^-,X) - I(S;Y_i|Y_i^-,Z) \\
&\quad + h(Y_i|Y_i^-,Z) - h(Y_i|Y_i^-,Z,X) \\
&= I(S;Y_i|Z,Y_i^-,X) - I(S;Y_i|Y_i^-,Z) + I(X;Y_i|Y_i^-,Z_i^-).
\end{aligned}
$$

The last term does not include S in the conditioning and is identical to the first term on the right hand side of (7), which proves the first equality in the lemma. The second equality follows by symmetry in X_i^- and Y_{i-1}. ∎

We will also define a relative SRDD index, which does not depend on the absolute strength of the couplings. We denote this the normalized SRDD index:

$$
\hat{I}_S(X \to Y|Z) \triangleq 1 - \frac{\min_{\phi\subseteq S_i^-} I(X_i^-;Y_i|Y_i^-,Z_i^-,\phi)}{I(X_i^-;Y_i|Y_i^-,Z_i^-)}. \quad (8)
$$

It can easily be seen that $\hat{I}_S(X \to Y|Z) \in [0,1]$, where values close to 1 indicate that the external stimuli S has a great effect upon the coupling strength between X and Y, and a value close to 0 indicates that S has no effect.

3.1. Examples

The following examples show that as long as (partial) information about the stimuli S is shared between the sensors' data, the lower bound is (7) is non-trivially bounded away from zero. In these examples, we let X, Y, Z, and S be mutually independent standard Normal random variables.

3.1.1. Example 1

Assume that a sensor is measuring $\xi_1 = X+Y$ and another sensor is measuring $\xi_2 = X+Z+S$. We are interested in comparing the mutual information between the two sensor measurements conditioned upon knowledge of Z or (Z,S). We therefore obtain:

$$
\begin{aligned}
&I(\xi_1;\xi_2|Z) = I(X+Y;X+Z+S|Z) \\
&\quad = I(X+Y;X+S) = h(X+Y) - h(X+Y|X+S), \\
&I(\xi_1;\xi_2|Z,S) = I(X+Y;X+Z+S|Z,S) \\
&\quad = I(X+Y;X) = h(X+Y) - h(Y).
\end{aligned}
$$

Since X, Y, and S are jointly Gaussian, the conditional variance of $X+Y$ given $X+S$ is:

$$
\operatorname{var}(X+Y|X+S) = \operatorname{var}((X+Y) - \alpha(X+S)) \quad (9)
$$

$$
= \operatorname{var}(Y) + (1-\alpha)^2\operatorname{var}(X) + \alpha^2\operatorname{var}(S) > \operatorname{var}(Y), \quad (10)
$$

where $\alpha \geq 0$ denotes a linear estimator.

Thus, $h(Y) < h(X+Y|X+S)$, which implies that $I(\xi_1;\xi_2|Z)$ is less than $I(\xi_1;\xi_2|Z,S)$ and conditioning upon S therefore leads to an increase of mutual information. If the two sensor measurements represents $\xi_1 = X_i^-$ and $\xi_2 = Y_{i-1}$ in (7), and Z and S represent Z_i^-, S_i^-, respectively, then the minimum in (7) is zero, and it is achieved by letting $\phi = \emptyset$.

3.1.2. Example 2

Let us now assume that $\xi_1 = X+Y+S$ and $\xi_2 = X+Z+S$ so that the external stimuli S is affecting the information captured by both sensors. Then we get

$$
I(\xi_1;\xi_2|Z) = h(X+Y+S) - h(Y), \quad (11)
$$

$$
I(\xi_1;\xi_2|Z,S) = h(X+Y) - h(Y), \quad (12)
$$

where clearly $h(X+Y+S) \geq h(X+Y)$, which implies that $I(\xi_1;\xi_2|Z,S) \leq I(\xi_1;\xi_2|Z)$.

4. COMPUTING STIMULI-RELEVANT DIRECTED DEPENDENCY (SRDD) INDEX

Let us consider the $M = 4$ nodes network shown in Fig. 1. Let node S be the stimuli and we are interested in estimating the SRDD between nodes X, Y and Z. It is generally infeasible to take the infinite past of the processes into account, when estimating the directed dependencies. In this work, we suggest to use Takens' delay embedding approach [17, 18] whereby the infinite past X_i^- is approximated by:

$$
X_i^- \approx [X_{i-m}, X_{i-2m}, \ldots, X_{i-dm}],
$$

where m and d are the embedding delay and dimension, respectively. Similarly for Y_i^-, S_i^- and Z_i^-.

After replacing the approximated past states of the processes in (7) or (8), two conditional mutual informations (or transfer entropies) are estimated by using the nearest neighbor-based approach proposed in [9] (which is referred to Kraskov–Stögbauer–Grassberger (KSG) estimator in the literature). The combination of the uniform embedding approach and the KSG estimator has been widely used in the literature to estimate transfer entropy [18, 3, 19].

The conditional mutual informations in (7) can be expressed as the sum/difference of four differential joint entropies. The entropy of the set of variables with the highest dimensionality is estimated via a neighbor search and the other three entropies are estimated using range searches [18, 3, 8]. The conditional mutual information is finally estimated as [8, 18, 3]:

$$I(X_i^-; Y_i|Y_i^-, Z_i^-) \approx \psi(K) + \frac{1}{N}\sum_{i=1}^{N}\psi\left(N_{[Y_i^-,Z_i^-]}+1\right)$$
$$-\psi\left(N_{[Y_i,Y_i^-,Z_i^-]}+1\right) - \psi\left(N_{[X_i^-,Y_i^-,Z_i^-]}+1\right),$$

where $N_{[Y_i^-,Z_i^-]}$ denotes the number of realizations (or points) whose maximum norm from $[Y_i^-, Z_i^-]$ is strictly less than twice the maximum norm of $[Y_i, Y_i^-, X_i^-, Z_i^-]$ from its K^{th} neighbor. Similarly for $N_{[Y_i,Y_i^-,Z_i^-]}$ and $N_{[X_i^-,Y_i^-,Z_i^-]}$.

To solve the minimization in (7), we search over all possible subsets (including the empty and the full sets) of $\phi \subseteq [S_{i-m}, S_{i-2m}, \dots, S_{i-dm}]$.

5. SIMULATION STUDY

We use simulated data in order to evaluate our proposed SRDD index in a controlled way. A non-linear autoregressive (AR) model is used to generate multivariate data with controlled directed dependency strengths and stimuli effect. The model reflects the network of Fig. 1 and is given as:

$$X_i = 0.35X_{i-1} + 0.5Z_{i-2} + 0.6S_{i-1} + \varepsilon_X$$
$$Y_i = 0.35Y_{i-1} + 0.5X_{i-2}^2 + 0.5Z_{i-2} + 0.6S_{i-1} + \varepsilon_Y$$
$$Z_i = 0.35Z_{i-1} + \varepsilon_Z$$
$$S_i = 0.25\sqrt{2}S_{i-1} - 0.2025S_{i-2} + \varepsilon_S,$$

where ε_X, ε_Y, ε_Z and ε_S are mutually independent zero mean and unit variance white Gaussian processes. We assume that S is the (external) stimuli, which is affecting nodes X and Y. Node X is non-linearly and causally coupled to node Y. There are also directed dependencies from Z to both X and Y.

The SRDD index and the transfer entropy causally conditioned on Z and Z, S (for the sake of comparison) averaged over 50 realizations of the AR model at length $N = 5000$ are reported in Table 1. The embedding delay m and dimension d were chosen as 1 and 5, respectively. For estimating the transfer entropy and the SRDD index, $K = 10$ neighbors were considered.

As expected, there are non-zero transfer entropies from node X to Y, Z to X and Z to Y in both tables 1a and 1b. Adding the external stimuli S to the conditioning process decreases the directed dependency only from node X to Y and it does not affect other couplings. On the other hand, the stimuli can only affect the directed dependency between nodes X and Y, which should lead to a non-zero SRDD index only from node X to Y. This is confirmed in Table 1c, where only between nodes X and Y there is detected a non-zero directed dependency due to the effect of the stimuli.

Table 1: Dependency matrices: a) conditioned transfer entropy on Z , b) on Z, S and c) SRDD index. The direction of dependencies are from columns to the rows. The reported numbers are rounded to two decimal places.

To \ From	X	Y	Z
X	0.00	0.00	0.04
Y	0.09	0.00	0.02
Z	0.00	0.00	0.00

(a) $I(X \to Y||Z)$

To \ From	X	Y	Z
X	0.00	0.00	0.04
Y	0.05	0.00	0.02
Z	0.00	0.00	0.00

(b) $I(X \to Y||Z, S)$

To \ From	X	Y	Z
X	0.00	0.00	0.00
Y	0.04	0.00	0.00
Z	0.00	0.00	0.00

(c) SRDD index

6. CONCLUSIONS AND FUTURE WORKS

In this paper, we introduced a new directed dependency index based on the difference of two transfer entropies for quantifying the amount of the directed dependency which is due to a given time series (e.g. stimuli). We provided a lower bound for the proposed index and showed that the lower bound could be calculated using existing transfer entropy estimators. In a simulation study, we utilized the KSG estimator and demonstrated that the dependency index could be efficiently computed and it was quite accurately able to predict the true dependencies in the data. Other estimators based on binning technique [19] and linear estimators [19] could potentielly also be used to compute the stimuli-relevant directed dependency index.

As an example, we considered a small network containing only a few nodes. However, our proposed index does not depend upon the size of the network, and can therefore be applied to more complex networks. The difficulty is to reliably estimate the transfer entropy for high dimensional problems. A possible solution in the literature to overcome the problem of entropy estimation of high-dimensional data is to select only the most informative subset of candidates among the full set of variables [19]. For the problems considered in this paper, we could potentially use the greedy selection technique of [8, 3, 20, 21] to approximately solve the minimization problem in (8), and at the same time reduce the dimensionality of the problem.

7. REFERENCES

[1] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461 – 464, 2000.

[2] K. Mehta and J. Kliewer, "Directional and causal information flow in eeg for assessing perceived audio quality," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 3, no. 3, pp. 150–165, 2017.

[3] P.S. Baboukani, C. Graversen, E. Alickovic, and J. Østergaard, "Estimating conditional transfer entropy in time series using mutual information and nonlinear prediction," *Entropy*, vol. 22, no. 10, pp. 1124, 2020.

[4] C.W.J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424 – 438, 1969.

[5] C.W.J. Granger, "Testing for causality: a personal viewpoint," *Journal of Economic Dynamics and control*, vol. 2, pp. 329–352, 1980.

[6] M.Z. Jahromi, A. Zahedi, J. Jensen, and J. Østergaard, "Information loss in the human auditory system," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 472–481, 2018.

[7] R.G. James, B.D.M. Ayala, B. Zakirov, and J. Crutchfield, "Modes of information flow," 2018.

[8] P.S. Baboukani, G. Graversen, and J. Østergaard, "Estimation of directed dependencies in time series using conditional mutual information and non-linear prediction," pp. 2388–2392, 2021.

[9] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, pp. 066138, 2004.

[10] R.W. Yeung, *A First Course in Information Theory*, Springer, 2006.

[11] N. Ay and D. Polani, "Information flows in causal networks," *Advances in complex systems*, vol. 11, no. 1, pp. 17 – 41, 2008.

[12] L. Faes, D. Marinazzo, and S. Stramaglia, "Multiscale information decomposition: Exact computation for multivariate gaussian processes," *Entropy*, vol. 19, no. 8, pp. 408, 2017.

[13] A.B. Barrett, "Exploration of synergistic and redundant information sharing in static and dynamical gaussian systems," *Physical Review E*, vol. 91, no. 5, pp. 052802, 2015.

[14] P.L. Williams and R.D. Beer, "Nonnegative decomposition of multivariate information," *arXiv preprint arXiv:1004.2515*, 2010.

[15] U.M. Maurer and S. Wolf, "Unconditionally secure key agreement and the intrinsic conditional information," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 499–514, 1999.

[16] J. Massey, "Causality, feedback and directed information," pp. 303–305, 1990.

[17] F. Takens, "In dynamical systems of turbulence," *Lecture notes in mathematics*, vol. 898, pp. 366, 1981.

[18] M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, "Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy," *BMC neuroscience*, vol. 12, no. 1, pp. 1–22, 2011.

[19] A. Montalto and L. Faesand D. Marinazzo, "Mute: a matlab toolbox to compare established and novel estimators of the multivariate transfer entropy," *PloS one*, vol. 9, no. 10, pp. e109462, 2014.

[20] Z. Jia, Y. Lin, Z. Jiao, Y. Ma, and J. Wang, "Detecting causality in multivariate time series via non-uniform embedding," *Entropy*, vol. 21, no. 12, pp. 1233, 2019.

[21] J. Zhang, "Low-dimensional approximation searching strategy for transfer entropy from non-uniform embedding," *PloS one*, vol. 13, no. 3, pp. e0194382, 2018.