CHAOS2010 Conference
Rendering Statistical Significance of
Information Flow Measures

Angeliki Papana and Dimitris Kugiumtzis
Aristotle University of Thessaloniki
Department of Mathematical, Physical and Computational Sciences
Faculty of Engineering
54124 Thessaloniki, Greece
e-mail: agpapana@gen.auth.gr, dkugiu@gen.auth.gr

Abstract. Information causality measures, i.e. transfer entropy and symbolic transfer entropy, are modified using the concept of surrogate data in order to identify correctly the presence and direction of causal effects. The measures are evaluated on multiple bivariate time series of known coupled systems of varying complexity and on a range of embedding dimensions. The proposed modifications of the causality measures are found to reduce the bias in the estimation of the measures and preserve the zero level in the absence of coupling.

Keywords: Information flow, Causality, Transfer entropy.

1 Introduction

Given a set of time series observations, it is essential to assess whether they originate from coupled or uncoupled systems, detect the hidden causal dependencies between them and understand which system is the driver. For these purposes, many model-free measures have been developed based on phase synchronization [8], geometry in state space [1,7] and information theory [9,11,10].

There have been recently comparative studies on the different causality measures [2,5]. In a recent evaluation of different causality measures [6], we observed in many cases significant measure values in the absence of coupling that can be misinterpreted as weak coupling. Here, we investigate this on information measures. Their estimation involves the estimation of entropies that can have bias depending on the time series length, state space reconstruction and system complexity.

We attempt to establish the zero level of information measures when the systems are not coupled and for this we use surrogate data. The suggested surrogates are extracted by randomly shuffling the reconstructed points of the driving time series, so that the dynamical properties of each system are preserved in the point representation and only the coupling, if present, is destroyed. We believe that shuffling randomly the reconstructed points is more appropriate than shuffling the samples of the driving time series, as proposed
in the so-called effective transfer entropy [4]. The original, effective, and our corrected information measures are tested on simulated systems of varying complexity. We also study their performance for a range of embedding dimensions (different for the driving and the driven system) and in the presence of noise.

The structure of the paper is as follows. In Section 2, we present the information causality measures and the proposed corrections using surrogate data. In Section 3, we report the results of our simulation study, and in Section 4, we discuss the results and draw conclusions.

2 Information causality measures and proposed corrections

Let \( \{x_t\} \) and \( \{y_t\} \), \( t = 1, \ldots, n \), denote two simultaneously observed univariate time series derived from the dynamical systems \( X \) and \( Y \), respectively. The reconstructed vectors of the two state spaces are defined as

\[
x_t = (x_t, x_{t-\tau_x}, \ldots, x_{t-(m_x-1)\tau_x}), \quad y_t = (y_t, y_{t-\tau_y}, \ldots, y_{t-(m_y-1)\tau_y}),
\]

where \( t = 1, \ldots, n' \), \( n' = n - \max\{m_x - 1, m_y - 1\}\), \( \tau_x \) and \( \tau_y \) the delays for \( X \) and \( Y \), respectively.

We consider measures of the information flow from \( X \) to \( Y \), denoted \( X \rightarrow Y \). Transfer entropy (TE) quantifies the amount of information explained in \( Y \) at one step ahead from the state of \( X \), accounting for the concurrent state of \( Y \) [9]. In terms of entropy, TE is defined as

\[
\text{TE}_{X \rightarrow Y} = H(x_{t}, y_{t}) - H(y_{t+1}, x_{t}, y_{t}) + H(y_{t+1}) - H(y_{t}),
\]

(1)

where \( H(x) \) is the entropy of \( x \). We consider the differential entropy and estimate it from the correlation sum \( C(x) \) as \( H(X) \approx \ln C(x) + m \ln r \), where \( C(x) \) is the estimated cumulative density of inter-point distances at embedding dimension \( m \) and a given distance \( r \) [3]. To account for the different dimensions in each entropy term in (1), we use the standardized Euclidean norm in the computation of \( C(x) \). Thus TE is estimated by the correlation sums as

\[
\text{TE}_{X \rightarrow Y} = \log \frac{C(y_{t+1}, x_{t}, y_{t})C(y_{t})}{C(x_{t}, y_{t})C(y_{t+1}, y_{t})}.
\]

(2)

The so-called effective transfer entropy (ETE), is defined as the difference of TE computed on the original bivariate time series from the TE computed on a surrogate bivariate time series, where the driving time series \( X \) is randomly shuffled [4]. In order to gain better stability of the measure, we consider instead of a single surrogate an ensemble \( M \) surrogates and ETE is then defined as

\[
\text{ETE}_{X \rightarrow Y} = \frac{1}{M} \sum_{i=1}^{M} \text{TE}_{X_i \rightarrow Y},
\]

(3)
where \( l = 1, \ldots, M \) and \( \text{TE}_{X \rightarrow Y} \) is the TE for the \( l \)-th surrogate.

The random shuffling of the driving time series aims at destroying any possible coupling structure between systems \( X \) and \( Y \). Thus the mean of the TE on the surrogate bivariate time series estimates the significance level of TE and thus positive values of \( \text{ETE}_{X \rightarrow Y} \) indicate true coupling.

There is a serious shortcoming in ETE. The random permutation of the samples in \( \{x_t\} \) destroys not only the time correspondence between \( \{x_t\} \) and \( \{y_t\} \), but also the dynamics in \( X \), represented in the reconstructed points \( \{x_t\} \). The dynamics of the individual systems may fool the coupling measures and give false positive values in the absence of coupling. This form of bias in TE cannot be rendered by ETE and the significance level (for zero coupling) is not attained by the TE on these surrogate data. The remaining bias can be both positive and negative depending on, among others, the state space reconstruction. Therefore we suggest to shuffle the reconstructed points \( \{x_t\} \) instead. Moreover, to gain stability in the estimation of the significance level of TE, we propose to subtract the entropy on these surrogate data at each of the two terms in (1), where \( x_t \) occurs, i.e. \( H(x_t, y_t) \) and \( H(y_{t+h}, x_t, y_t) \). Thus considering the estimation of TE using correlation sums in (2), the terms \( C(y_{t+1}, x_t, y_t) \) and \( C(x_t, y_t) \) are replaced by the average on \( M \) replications of the random shufflings of the reconstructed vectors of \( X \)

\[
C_s(y_{t+1}, x_t, y_t) = \frac{1}{M} \sum_{l=1}^{M} C(y_{t+1, l}, x_{t, l}, y_{t, l}), \quad C_s(x_t, y_t) = \frac{1}{M} \sum_{l=1}^{M} C(x_{t, l}, y_{t, l})
\]

where \( t_l = 1, \ldots, n' \) are randomly chosen for each \( t = 1, \ldots, n' \) and this is repeated \( M \) times. Then, the TE in the absence of coupling between \( X \) and \( Y \) is

\[
\text{TE}_{X \rightarrow Y} = \log \frac{C_s(y_{t+1}, x_t, y_t)C_s(x_t, y_t)}{C_s(x_t, y_t)C_s(y_{t+1}, y_t)}.
\]

The suggested corrected transfer entropy (CTE) is thus derived as

\[
\text{CTE}_{X \rightarrow Y} = \text{TE}_{X \rightarrow Y} - \text{TE}_{S_{X \rightarrow Y}} = \log \frac{C_s(y_{t+1}, x_t, y_t)C_s(x_t, y_t)}{C_s(y_{t+1}, x_t, y_t)C_s(x_t, y_t)}.
\] (4)

The symbolic transfer entropy (STE) is a recently suggested version of TE using rank points instead of reconstructed points [10]. For each vector, say \( y_t \), the ranks of its components assign a rank-point \( \hat{y}_j = [r_1, r_2, \ldots, r_{m_y}] \), where \( r_j \in \{1, 2, \ldots, m_y\} \) for \( j = 1, \ldots, m_y \). Following this sample-point to rank-point conversion, the sample \( y_{t+1} \) in Eq.(1) is taken as a rank point at time \( t+1 \), \( \hat{y}_{t+1} \), and STE is defined in terms of entropy as

\[
\text{STE}_{X \rightarrow Y} = H(\hat{x}_t, \hat{y}_t) - H(\hat{y}_{t+h}, \hat{x}_t, \hat{y}_t) + H(\hat{y}_{t+h}, \hat{y}_t) - H(\hat{y}_t).
\] (5)

The estimation of the Shannon entropy terms is straightforward by estimating the probabilities from the relative frequencies of the observed ranks. The effective symbolic transfer entropy (ESTE) and corrected symbolic transfer entropy (CSTE) are defined analogously to ETE and CTE, respectively.
3 Results

In all simulations, the time series were normalized to have mean zero and standard deviation one and the distance \( r \) in the calculation of correlation sums was set to 0.2. The first simulation system is comprised of two Henon maps coupled in one direction

\[
x_{t+1} = 1.4 - x_t^2 + 0.3x_{t-1} \\
y_{t+1} = 1.4 - cx_t y_t + (1-c)y_t^2 + 0.3y_{t-1}
\] (6)

with coupling strengths \( c = 0, 0.15, 0.1, 0.2, 0.3, 0.4 \) and 0.5. The second system is comprised by two Mackey-Glass systems with unidirectional coupling

\[
\frac{dx}{dt} = \frac{0.2x_t - \Delta_1}{1 + x_t^{10} - \Delta_1} - 0.1x_t \\
\frac{dy}{dt} = \frac{0.2y_t - \Delta_2}{1 + y_t^{10} - \Delta_2} + c\frac{0.2x_t - \Delta_1}{1 + x_t^{10} - \Delta_1} - 0.1y_t.
\] (7)

The two systems can have different complexity determined by \( \Delta_1 \) and \( \Delta_2 \) taking all combinations of the values 17, 30, 100. The coupling strength is \( c = 0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4 \) and 0.5.

The time series lengths for the coupled Henon maps are \( n = 512, 1024, 2048 \) and for the Mackey-Glass systems \( n = 2048 \). Gaussian white noise is added on the time series with mean zero and standard deviation (SD) 0.05 and 0.2. We also investigate the influence of varying embedding dimensions \( m_x \) and \( m_y \) on the performance of the measures. For each system, coupling strength and condition \((n, \text{noise SD}, m_x \text{ and } m_y)\), all measures presented in Section 2 are computed on 100 realizations for both directions, \( X \rightarrow Y \) (the correct dependence direction) and \( Y \rightarrow X \).

For the coupled Henon maps, all TE measures (including ETE and CTE) correctly detect the direction of the information flow for all combinations of \( m_x \) and \( m_y \), even for small time series lengths (see Fig.1a and b). Only for \( m_x = 1 \) and \( m_y > 1 \), the three measures spuriously increase with the coupling strength in the direction \( Y \rightarrow X \). TE is positively biased and ETE is negatively biased for \( c = 0 \) for a number of combinations of the embedding dimensions, whereas CTE attains always the zero level. ETE is affected by the selection of the embedding dimensions more than TE and CTE and has also larger variance. With the addition of noise, the direction of the information flow is still correctly detected (see Fig.1c for noise SD 0.2), with ETE being more affected than the other measures especially for large embedding dimensions. The variance of all information measures gets also much larger. Also for the noisy data, CTE improves the performance of TE giving values at the zero level for \( c = 0 \) and for varying embedding dimensions and time series lengths. The best results for all measures are obtained when the embedding dimensions are equal.
Fig. 1. (a) Mean estimated values of TE, ETE and CTE for both directions from 100 realizations of the noise-free unidirectionally coupled Henon maps, with \( n = 512 \) and \( m_x = m_y = 2 \). (b) As in (a) but for \( m_x = 2, m_y = 4 \). (c) As in (a) but for noise SD 0.2.

STE, ESTE and CSTE correctly detect the direction of information flow for \( m_x \geq m_y > 2 \), for all time series lengths (see Fig.2). STE tends to increase in both directions with the embedding dimensions \( m_x \) and \( m_y \), while ESTE and CSTE tend to decrease. Thus, for embedding dimensions larger than 3, STE is positively biased, ESTE is negatively biased, whereas CSTE preserves again the zero level for \( c = 0 \). As for the TE measures, STE measures are also stable to noise (see Fig.2c).

The embedding dimension affects more the information causality measures on the unidirectionally coupled Mackey-Glass systems. TE and CTE detect correctly the direction of information flow only for \( m_x \geq m_y \), while ETE varies a lot with \( m_x \) and \( m_y \) and takes negative values for small \( c \) (see Fig.4a and b). For example, in Fig.4a where \( m_x = 2, m_y = 5 \) all measures vary with \( c \), but in the same way, giving larger values for the correct direction when \( c \) is small and the opposite for larger \( c \). This contradicting behavior was found for all combinations of \( \Delta_1 \) and \( \Delta_2 \). However, in all cases CTE
CTE stays close to zero for all but very large \( c \) values. Other sources of bias for \( c > 0 \), probably due to mixed dynamics, cannot be accounted for by the random shuffling of the reconstructed points. When noise is added to the time series, TE is more sensitive to the selection of the embedding dimensions and gives positive values for \( c = 0 \), whereas CTE stays close to zero for all but very large \( m_x \) and \( m_y \) (see Fig. 4c).

The problem of inconclusive estimation of coupling direction for \( m_x < m_y \) holds also for STE, as shown in Fig. 4a and b. For all scenarios of complexity of the coupled Mackey-Glass systems, STE obtains significant positive values also for \( c = 0 \) when \( m_x < m_y \), while ESTE obtains often negative values. CSTE follows the same dependence on \( c \) as STE but displaced so that CSTE falls to zero for \( c = 0 \). It is interesting that in the presence of noise, STE gets even larger value for \( c = 0 \) and increases faster for larger \( c \), and CSTE has the same course with \( c \) as STE but starts at the zero level for \( c = 0 \) (see Fig. 4c). We illustrate this nice property of CTE and CSTE for \( c = 0 \) in Fig. 5.
as a function of \( m_x = m_y \). First, we note that STE measures have much less

\[
\begin{align*}
\text{STE} & \xrightarrow{X \rightarrow Y} \\
\text{ESTE} & \xrightarrow{X \rightarrow Y} \\
\text{CSTE} & \xrightarrow{X \rightarrow Y}
\end{align*}
\]

variance than the respective TE measures and attain the zero level for small \( m_x = m_y \) (in Fig. 5a only the distribution of CTE contains zero for varying \( m_x = m_y \)). In Fig. 5c, where the systems have different complexity, ETE gets more affected by the individual system complexity as the embedding dimension increases, TE and CTE do not differ much in the two directions, but only CTE is at the zero level.

4 Discussion

We propose to correct the entropy terms in the definition of transfer entropy (TE) and symbolic transfer entropy (STE) by subtracting the average entropy computed on an ensemble of the trajectories of the two systems, where for the driving system the points of the reconstructed trajectory are randomly shuffled. We have found that this correction gives more stable estimates than when taking the overall average over all entropy terms. This can be explained by the variation in the estimation of each entropy term on the surrogate data. The corrected TE and STE, termed CTE and CSTE, respectively, attain the correct statistical significance, i.e., their values are within the zero level when the two systems are uncoupled. This has been illustrated with simulations on the unidirectionally coupled Henon map and Mackey-Glass system of varying complexity for the driving and the driven system. In the same simulations, we compared the so-called effective TE (ETE), and introduced also effective STE (ESTE), which use random shuffling of the samples of the time series instead of the reconstructed points. Both ETE and ESTE are found to be rather unstable and very sensitive to variations of the embedding dimensions. It is interesting that the correct significance of CTE and CSTE holds for a
range of varying embedding dimensions and in the presence of noise, as well as for different data sizes.

In all works on coupling measures, the embedding dimension of the driving system $m_x$ and driven system $m_y$ are always equal and often the two systems $X$ and $Y$ are of the same complexity (though not identical in parameters). One would expect that different complexity of $X$ and $Y$ would suggest different $m_x$ and $m_y$. Our simulation results confirm that the popular choice $m_x = m_y$ gives indeed the best results and identifies best the correct direction of information flow even when the systems differ in complexity.

**Acknowledgments** This research project is implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and is co-financed at 90% jointly by E.U.-European Social Fund (75%) and the Greek Ministry of Development-GSRT (25%) and at 10% by "The Epilepsy Centre - SSE", Oslo University Hospital, Oslo, Norway.

**References**