Chapter 15
A Robust Estimation of Information Flow in Coupled Nonlinear Systems

Shivkumar Sabesan, Konstantinos Tsakalis, Andreas Spanias, and Leon Iasemidis

Abstract Transfer entropy (TE) is a recently proposed measure of the information flow between coupled linear or nonlinear systems. In this study, we first suggest improvements in the selection of parameters for the estimation of TE that significantly enhance its accuracy and robustness in identifying the direction and the level of information flow between observed data series generated by coupled complex systems. Second, a new measure, the net transfer of entropy (NTE), is defined based on TE. Third, we employ surrogate analysis to show the statistical significance of the measures. Fourth, the effect of measurement noise on the measures' performance is investigated up to $S/N = 3$ dB. We demonstrate the usefulness of the improved method by analyzing data series from coupled nonlinear chaotic oscillators. Our findings suggest that TE and NTE may play a critical role in elucidating the functional connectivity of complex networks of nonlinear systems.

15.1 Introduction

Recent advances in information theory and nonlinear dynamics have facilitated novel approaches for the study of the functional interactions between coupled

Shivkumar Sabesan
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Barrow Neurological Institute, Phoenix, AZ 85013, USA, e-mail: ssabesa@asu.edu

Konstantinos Tsakalis
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: tsakalis@asu.edu

Andreas Spanias
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA, e-mail: spanias@asu.edu

Leon Iasemidis
The Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85287, USA; Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA; Mayo Clinic, Phoenix, AZ 85054, USA, e-mail: leon.iasemidis@asu.edu

linear and nonlinear systems. The estimation of these interactions, especially when the systems' structure is unknown, holds promise for the understanding of the mechanisms of their interactions and for a subsequent design and implementation of appropriate schemes to control their behavior. Traditionally, cross-correlation and coherence measures have been the mainstay of assessing statistical interdependence among coupled systems. These measures, however, do not provide reliable information about directional interdependence, i.e., if one system drives the other.

To study the directional aspect of interactions, many other approaches have been employed [24, 22, 11, 18, 19]. One of these approaches is based on the improvement of the prediction of a series' future values by incorporating information from another time series. Such an approach was originally proposed by Wiener [24] and later formalized by Granger in the context of linear regression models of stochastic processes. Granger causality was initially formulated for linear models, and it was then extended to nonlinear systems by (a) applying to local linear models in reduced neighborhoods, estimating the resulting statistical quantity and then averaging it over the entire dataset [20] or (b) considering an error reduction that is triggered by added variables in global nonlinear models [2].

Despite the relative success of the above approaches in detecting the direction of interactions, they essentially are model-based (parametric) methods (linear or nonlinear), i.e., these approaches either make assumptions about the structure of the interacting systems or the nature of their interactions, and as such they may suffer from the shortcomings of modeling systems/signals of unknown structure. For a detailed review of parametric and nonparametric (linear and nonlinear) measures of causality, we refer the reader to [9, 15]. To overcome this problem, an information theoretic approach that identifies the direction of information flow and quantifies the strength of coupling between complex systems/signals has recently been suggested [22]. This method was based on the study of transitional probabilities of the states of systems under consideration. The resulted measure was termed transfer entropy (TE).

We have shown [18, 19] that the direct application of the method as proposed in [22] may not always give the expected results. We show that tuning of certain parameters involved in the TE estimation plays a critical role in detecting the correct direction of the information flow between time series. We propose a methodology to also test the significance of the TE values using surrogate data analysis and we demonstrate its robustness to measurement noise. We then employ the improved TE method to define a new measure, the net transfer entropy (NTE). Results from the application of the improved TE and NTE show that these measures are robust in detecting the direction and strength of coupling under noisy conditions.

The organization of the rest of this chapter is as follows. The measure of TE and the estimation problems we identified, as well as the improvements and practical adjustments that we introduced, are described in Section 15.2. In Section 15.3, results from the application of this method to a system of coupled *Rössler* oscillators are shown. These results are discussed and conclusions are drawn in Section 15.4.

15.2 Methodology

15.2.1 Transfer Entropy *(TE)*

Consider a kth order Markov process [10] described by

$$P(x_{n+1}|x_n, x_{n-1}, \cdots, x_{n-k+1}) = P(x_{n+1}|x_n, x_{n-1}, \cdots, x_{n-k}), \tag{15.1}$$

where P represents the conditional probability of state x_{n+1} of a random process X at time $n+1$. Equation (15.1) implies that the probability of occurrence of a particular state x_{n+1} depends only on the past k states $[x_n, \cdots, x_{n-k+1}] \equiv x_n^{(k)}$ of the system. The definition given in Equation (15.1) can be extended to the case of Markov interdependence of two random processes X and Y as

$$P(x_{n+1}|x_n^{(k)}) = P(x_{n+1}|(x_n^{(k)}, y_n^{(l)})), \tag{15.2}$$

where $x_n^{(k)}$ are the past k states of the first random process X and $y_n^{(l)}$ are the past l states of the second random process Y. This generalized Markov property implies that the state x_{n+1} of the process X depends only on the past k states of the process X and not on the past l states of the process Y. However, if the process X also depends on the past states (values) of process Y, the divergence of the hypothesized transition probability $P(x_{n+1}|x_n^{(k)})$ (L.H.S. of Equation (15.2)), from the true underlying transition probability of the system $P(x_{n+1}|(x_n^{(k)}, y_n^{(l)}))$ (R.H.S of Equation (15.2)), can be quantified using the Kullback–Leibler measure [11]. Then, the Kullback–Leibler measure quantifies the transfer of entropy from the driving process Y to the driven process X, and if it is denoted by TE(Y→X), we have

$$\text{TE}(Y \rightarrow X) = \sum_{n=1}^{N} P(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{P(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{P(x_{n+1}|x_n^{(k)})}. \tag{15.3}$$

The values of the parameters k and l are the orders of the Markov process for the two coupled processes X and Y, respectively. The value of N denotes the total number of the available points per process in the state space.

In search of optimal k, it would generally be desirable to choose the parameter k as large as possible in order to find an invariant value (e.g., for conditional entropies to converge as k increases), but in practice the finite size of any real data set imposes the need to find a reasonable compromise between finite sample effects and approximation of the actual value of probabilities. Therefore, the selection of k and l plays a critical role in obtaining reliable values for the transfer of entropy from real data. The estimation of TE as suggested in [22] also depends on the neighborhood size (radius r) used in the state space for the calculation of the involved joint and conditional probabilities. The value of radius r in the state space defines the maximum norm distance in the search for neighboring state space points. Intuitively,

different radius values in the estimation of the multidimensional probabilities in the state space correspond to different probability bins. The values of radius for which the probabilities are not accurately estimated (typically large r values) may eventually lead to an erroneous estimate of TE.

15.2.2 Improved Computation of Transfer Entropy

15.2.2.1 Selection of k

The value of k (order of the driven process) used in the calculation of TE $(Y \rightarrow X)$ (see Equation (15.3)) represents the dependence of the state x_{n+1} of the system on its past k states. A classical linear approach to autoregressive (AR) model order selection, namely the Akaike information criterion (AIC), has been applied to the selection of the order of Markov processes. Evidently, AIC suffers from substantial overestimation of the order of the Markov process order in nonlinear systems and, therefore, is not a consistent estimator [12]. Arguably, a method to estimate this parameter is the delayed mutual information [13]. The delay d at which the mutual information of X reaches its first minimum can be taken as the estimate of the interval within which two states of X are dynamically correlated with each other. In essence, this value of d minimizes the Kullback–Leibler divergence between the dth and higher order corresponding probabilities of the driven process X (see Equation (15.1)), i.e., there is minimum information gain about the future state of X by using its values that are more than d steps in the past. Thus, in units of the sampling period, d would be equal to the order k of the Markov process.

If the value of k is severely underestimated, the information gained about x_{n+1} will erroneously increase due to the presence of y_n and would result to an incorrect estimation of TE. A straightforward extension of this method for estimation of k from real-world data may not be possible, especially when the selected value of k is large (i.e., the embedding dimension of state space would be too large for finite duration data in the time domain). This may thus lead to an erroneous calculation of TE. From a practical point of view, a statistic that may be used is the correlation time constant t_e, which is defined as the time required for the autocorrelation function (AF) to decrease to $1/e$ of its maximum value (maximum value of AF is 1) (see Fig. 15.1d) [13]. AF is an easy metric to compute over time, has been found to be robust in many simulations, but detects only linear dependencies in the data. As we show below and elsewhere [18, 19], the derived results from the detection of the direction and level of interactions justify such a compromise in the estimation of k.

15.2.2.2 Selection of l

The value of l (order of the driving system) was chosen to be equal to 1. The justification for the selection of this value of l is the assumption that the current state
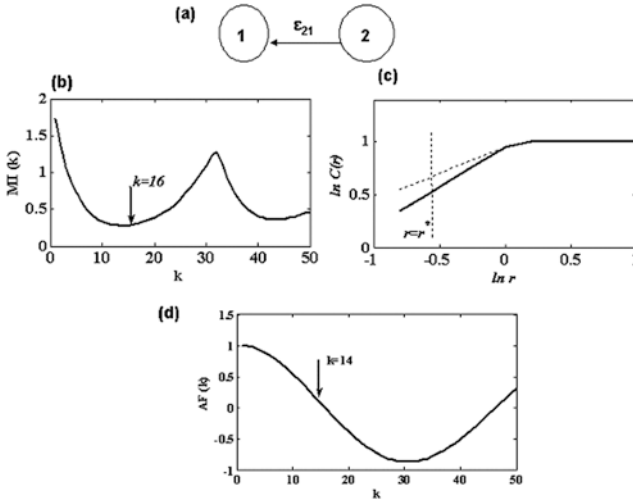
Fig. 15.1: (**a**) Unidirectionally coupled oscillators with $\varepsilon_{21} = 0.05$. (**b**) Mutual information MI vs. k (the first minimum of the mutual information between the X variables of the oscillators is denoted by a downward arrow at $k = 16$). (**c**) $\ln C$ vs. $\ln r$ with $k = 16$, $l = 1$, where C and r denote average joint probability and radius, respectively (*dotted line* for direction of flow $1 \rightarrow 2$; *solid line* for direction of flow $2 \rightarrow 1$). (**d**) Autocorrelation function (AF) vs. k (the delay at which the AF decreases to $1/e$ of its maximum value is denoted by a *downward arrow* at $k = 14$).

of the driving system is sufficient to produce a considerable change in the dynamics of the driven system within one time step (and hence only immediate interactions between X and Y are assumed to be detected in the analysis herein). When larger values for l were employed (i.e., a delayed influence of Y on X), detection of information flow from Y to X was also possible. These results are not presented in this chapter.

15.2.2.3 Selection of Radius r

The multi-dimensional transitional probabilities involved in the definition of transfer entropy (Equation (15.3)) are calculated by joint probabilities using the conditional probability formula $P(A|B) = P(A,B)/P(B)$. One can then reformulate the transfer entropy as

$$
\begin{aligned}
&\mathrm{TE}(Y \rightarrow X) = \\
&\sum_{n=1}^{N} P(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{P(x_{n+1}, x_n^{(k)}, y_n^{(l)}) P(x_n^{(k)})}{P(x_{n+1}, x_n^{(k)}) P(x_n^{(k)}, y_n^{(l)})}.
\end{aligned}
\tag{15.4}
$$

From the above formulation, it is clear that probabilities of a vector in the state space at the nth time step are compared with ones of vectors in the state space at the $(n+1)$th time step, and, therefore, the units of TE are in bits/time step, where time step in simulation studies is the algorithm's (e.g., Runge–Kutta) iteration step (or a multiple of it if one downsamples the generated raw data before the calculation of TE). In real life applications (like in electroencephalographic (EEG) data), the time step corresponds to the sampling period of the sampled (digital) data. In this sense, the units of TE denote that TE actually estimates the rate of the flow of information. The multidimensional joint probabilities in Equation (15.4) are estimated through the generalized correlation integrals $C_n(r)$ in the state space of embedding dimension $p = k + l + 1$ [14] as

$$P_r(x_{n+1}, x_n^{(k)}, y_n^{(l)}) =$$

$$\frac{1}{N} \sum_{m=0}^{N-1} \Theta \left(r - \left| \begin{matrix} x_{n+1} - x_{m+1} \\ x_n^{(k)} - x_m^{(k)} \\ y_n^{(l)} - y_m^{(l)} \end{matrix} \right| \right) \tag{15.5}$$

$$= C_{n+1}(r),$$

where $\Theta(x > 0) = 1$; $\Theta(x = 0) = 0$, $|\cdot|$ is the maximum distance norm, and the subscript $(n+1)$ is included in C to signify the dependence of C on the time index n (note that averaging over n is performed in the estimation of TE, using Equation (15.5) into Equation (15.3)). In the rest of the chapter we use the notation $C_n(r)$ or $C_{n+1}(r)$ interchangeably. Equation (15.5) is in fact a simple form of a kernel density estimator, where the kernel is the Heaviside function Θ. It has been shown that this approach may present some practical advantages over the box-counting methods for estimating probabilities in a higher dimensional space. We also found that the use of a more elaborate kernel (e.g., a Gaussian or one which takes into account the local density of the states in the state space) than the Heaviside function does not necessarily improve the ability of the measure to detect direction and strength of coupling. Distance metrics other than the maximum norm, such as the Euclidean norm, may also be considered, however, at the cost of increased computation time. In order to avoid a bias in the estimation of the multidimensional probabilities, temporally correlated pairs of points are excluded from the computation of $C_n(r)$ by means of the Theiler correction and a window of $(p-1) * l = k$ points in duration [23].

The estimation of joint probabilities between two different time series requires concurrent calculation of distances in both state spaces (see Equation (15.4)). Therefore, in the computation of $C_n(r)$, the use of a common value of radius r in both state spaces is desirable. In order to establish a common radius r in the state space of X and Y, the data are first normalized to zero mean ($\mu = 0$) and unit variance ($\sigma = 1$). In previous publications [18, 19], using simulation examples (unidirectional as well as bidirectional coupling in two and three coupled oscillator model configurations), we have found that the TE values obtained for only a certain range of r accurately detect the direction and strength of coupling. In general, when any of the joint probabilities ($C_n(r)$) in log scale is plotted against the corresponding radius r in log scale,

it initially increases with increase in the value of the radius (linear increase for small values of r) and then saturates (for large values of r) [3]. It was found that using a value of r^* within the quasilinear region of the $\ln C_n(r)$ vs. $\ln r$ curve produces consistent changes in TE with changes in directional coupling.

Although such an estimation of r^* is possible in noiseless simulation data, for physiological data sets that are always noisy, and the underlying functional description is unknown, it is difficult to estimate an optimal value r^* simply because a linear region of $\ln C_n(r)$ vs. $\ln r$ may not be apparent or even exist. It is known that the presence of noise in the data will be predominant for small r values [10, 8] and over the entire space (high dimensional). This causes the distance between neighborhood points to increase. Consequently, the number of neighbors available to estimate the multidimensional probabilities at the smaller scales may decrease and it would lead to a severely biased estimate of TE. On the other hand, at large values of r, a flat region in $\ln C_n(r)$ may be observed (saturation). In order to avoid the above shortcomings in the practical application of this method (e.g., in simulation models with added noise or in the EEG), we approximated TE as the average of TEs estimated over an intermediate range of r values (from $\sigma/5$ to $2\sigma/5$). The decision to use this range for r was made on the practical basis that r less than $\sigma/2$ typically (well-behaved data) avoids saturation and r larger than $\sigma/10$ typically filters a large portion of A/D-generated noise (simulation examples offer corroborative evidence for such a claim). Even though these criteria are soft for r (no exhaustive testing of the influence of the range of r on the final results), it appears that the proposed range constitutes a very good compromise (sensitivity and specificity-wise) for the subsequent detection of the direction and magnitude of flow of entropy (see Section 15.3). Finally, to either a larger or lesser degree, all existing measures of causality suffer from the finite sample effect. Therefore, it is important to always test their statistical significance using surrogate techniques (see next subsection).

15.2.3 Statistical Significance of Transfer Entropy

Since TE calculates the direction of information transfer between systems by quantifying their conditional statistical dependence, a random shuffling applied to the original driver data series Y destroys the temporal correlation and significantly reduces the information flow $TE(Y \rightarrow X)$. Thus, in order to estimate the statistically significant values of $TE(Y \rightarrow X)$, the null hypothesis that the current state of the driver process Y does not contain any additional information about the future state of the driven process X was tested against the alternate hypothesis of a significant time dependence between the future state of X and the current state of Y. One way to achieve this is to compare the estimated values of $TE(Y \rightarrow X)$ (i.e., the $TE(x_{n+1}|x_n^{(k)}, y_n^{(l)})$), thereafter denoted by TE_o, with the TE values estimated by studying the dependence of future state of X on the values of Y at randomly shuffled time instants (i.e., $TE(x_{n+1}|x_n^{(k)}, y_p^{(l)})$), thereafter denoted by TE_s, where $p \in 1, \ldots, N$ is selected from the shuffled time instants of Y. The above described surrogate

analysis is valid when $l = 1$; for $l > 1$, tuples from original Y, each of length l, should be shuffled instead.

The shuffling was based on generation of white Gaussian noise and reordering of the original data samples of the driver data series according to the order indicated by the generated noise values (i.e., random permutation of all indices $1, \ldots, N$ and reordering of the Y time series accordingly). Transfer entropy TE_s values of the shuffled datasets were calculated at the optimal radius r^* from the original data. If the TE values obtained from the original time series (TE_o) were greater than Tth standard deviations from the mean of the TE_s values, the null hypothesis was rejected at the $\alpha = 0.01$ level (The value of Tth depends on the desired level of confidence $1 - \alpha$ and the number of the shuffled data segments generated, i.e., the degrees of freedom of the test). Similar surrogate methods have been employed to assess uncertainty in other empirical distributions [4, 21, 16].

15.2.4 Detecting Causality Using Transfer Entropy

Since it is difficult to expect a truly unidirectional flow of information in real-world data (where flow is typically bidirectional), we have defined the causality measure net transfer entropy (NTE) that quantifies the driving of X by Y as

$$\text{NTE}(Y \to X) = \text{TE}(Y \to X) - \text{TE}(X \to Y). \qquad (15.6)$$

Positive values of $\text{NTE}(Y \to X)$ denote that Y drives (causes) X, while negative values denote the reverse case. Values of NTE close to 0 may imply either equal bidirectional flow or no flow of information (then, the values of TE will help decide between these two plausible scenarios). Since NTE is based on the difference between the TEs per direction, we expect this metric to generally be less biased than TE in the detection of the driver. In the next section, we test the ability of TE and NTE to detect direction and causality in coupled nonlinear systems and also test their performance against measurement (observation) noise.

15.3 Simulation Example

In this section, we show the application of the method of TE to nonlinear data generated from two coupled, nonidentical, *Rössler*-type oscillators i and j [7], each governed by the following general differential equations:

$$\begin{aligned}
\dot{x}_i &= -\omega_i y_i - z_i + \sum_{j=1, i \neq j}^{2} \varepsilon_{ji} x_j - \varepsilon_{ii} x_i, \\
\dot{y}_i &= \omega_i x_i + \alpha_i y_i, \\
\dot{z}_i &= \beta_i x_i + z_i (x_i - \gamma_i),
\end{aligned} \qquad (15.7)$$

where $i, j = 1, 2$, and $\alpha_i = 0.38$, $\beta_i = 0.3$, $\gamma_i = 4.5$ are the standard parameters used for the oscillators to be in the chaotic regime, while we introduce a mismatch in their parameter ω (i.e., $\omega_1 = 1$ and $\omega_2 = 0.9$) to make them nonidentical, ε_{ji} denotes the strength of the diffusive coupling from oscillator j to oscillator i; ε_{ii} denotes self-coupling in the ith oscillator (it is taken to be 0 in this example). Also, in this example, $\varepsilon_{12} = 0$ (unidirectional coupling) so that the direction of information flow is from oscillator $2 \rightarrow 1$ (see Fig. 15.1a for the coupling configuration). The data were generated using an integration step of 0.01 and a fourth-order Runge–Kutta integration method. The coupling strength ε_{21} is progressively increased in steps of 0.01 from a value of 0 (where the two systems are uncoupled) to a value of 0.25 (where the systems become highly synchronized). Per value of ε_{21}, a total of 10,000 points from the x time series of each oscillator were considered for the estimation of each value of the TE after downsampling the data produced by Runge–Kutta by a factor of 10 (common practice to speed up calculations after making sure the integration of the differential equations involved is made at a high enough precision). The last data point generated at one value of ε_{21} was used as the initial condition to generate data at a higher value of ε_{21}. Results from the application of the TE method to this system, with and without our improvements, are shown next.

Figure 15.1b shows the time-delayed mutual information MI of oscillator 1 (driven oscillator) at one value of ε_{21} ($\varepsilon_{21} = 0.05$) and for different values of k. The first minimum of MI occurs at $k=16$ (see the downward arrow in Fig. 15.1b). The state spaces were reconstructed from the x time series of each oscillator with embedding dimension $p = k + l + 1$. Figure 15.1c shows the $\ln C_{2 \rightarrow 1}(r)$ vs. $\ln r$ (dotted line), and $\ln C_{1 \rightarrow 2}(r)$ vs. $\ln r$ (solid line), estimated according to Equation (15.5). TE was then estimated according to Equation (15.4) at this value of ε_{21}. The same procedure was followed for the estimation of TE at the other values of ε_{21} in the range $[0, 0.25]$. Figure 15.1d shows the lags of the autocorrelation function AF of oscillator 1 (driven oscillator – see Figure 15.1a) at one value of ε_{21} (that is, $\varepsilon_{21} = 0.05$) and for different values of k. The value of k at which AF drops to $1/e$ of its maximum value was found equal to 14 (see the downward arrow in Fig. 15.1b), that is close to 16 that MI provides us with. Thus, it appears that AF could be used instead of MI in the estimation of k, an approximation that can speed up calculations, as well as end up with an accurate estimate for the direction of information flow.

15.3.1 Statistical Significance of TE and NTE

A total of 50 surrogate data series for each original data series at each ε_{21} coupling value were produced. The null hypothesis that the obtained values of TE_o are not statistically significant was then tested at $\alpha = 0.005$ for each value of ε_{21}. For every ε_{21}, if the TE_o values were greater than 2.68 standard deviations from the mean of the TE_s values, the null hypothesis was rejected (one-tailed t-test; $\alpha = 0.005$). Figure 15.2a depicts the TE_o and the corresponding mean of 50 surrogate TE_s values along with 99% confidence interval error bars in the directions $1 \rightarrow 2$ and $2 \rightarrow 1$ (using
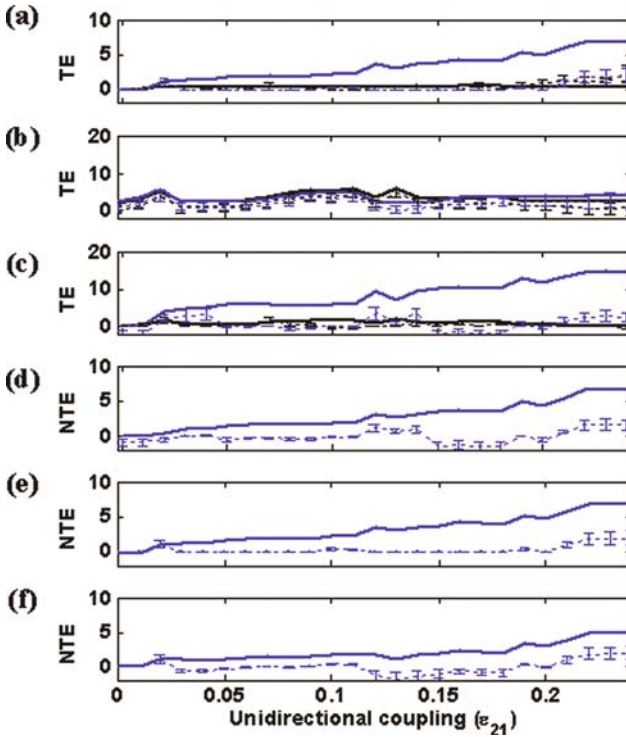
Fig. 15.2: Transfer entropy TE and net transfer of entropy NTE between coupled oscillators 1 and 2 (1→2 *black line*, 2→1 *blue line*) and mean ± 99.5% error bars of their corresponding 50 surrogate values as a function of the systems' underlying unidirectional coupling ε_{21} (from 0 to 0.25). Each TE value was estimated from $N = 10,000$ data points at each ε_{21}. The value of ε_{21} was increased by a step of 0.01. (**a**) TE_o (original data), mean, and 99.5% error bars from the distribution of TE_s (surrogate data). With $k = 16$, $l = 1$ (i.e., the suggested values by our methodology), TE is estimated at radius r^* within the linear region of $\ln C(r)$ vs. $\ln r$ from the original data (see Fig. 15.1c). $TE_o(2{\rightarrow}1)$ (*solid blue line*) is statistically significant ($p < 0.01$) and progressively increases in value with an increase in ε_{21}, whereas $TE_o(1{\rightarrow}2)$ (*solid black line*) is only locally statistically significant and remains constant and very close to 0 despite the increase in ε_{21}. (**b**) TEs estimated with $k = 5$, $l = 5$ as an average of the TEs at intermediate values of the radius r $[\sigma/5 < \ln r < 2\sigma/5]$. Neither $TE_o(2{\rightarrow}1)$ nor $TE_o(1{\rightarrow} 2)$ is statistically significant ($p > 0.01$) and does not progressively increase in value with an increase in ε_{21}. (**c**) TE estimated with the optimal values $k = 16$, $l = 1$. The picture is very similar to the one in (a) above, suggesting that use of r^* is not critical in the estimation of TEs. (**d**) $NTE_o(2{\rightarrow}1)$ and their corresponding $NTE_s(2{\rightarrow}1)$ estimated from the TE values in (**c**). (**e**) As in (**d**) above with noise of SNR = 10 dB added to the data. (**f**) As in (**d**) above with more noise (SNR = 3 dB) added to the data. Detection of direction of information flow is possible at all ε_{21} values ($p < 0.01$), except at very small ε_{21} values ($\varepsilon_{21} < 0.02$). Units of the estimated measures TE and NTE are in bits per iteration (time step was 0.1, i.e., Runge's time step 0.01 *times* 10, because of the 10:1 decimation we applied on the generated data before analysis).

a *black* and *blue lines* respectively) estimated over ε_{21} at a value of r^* chosen in the linear region of the $\ln C_n(r)$ vs. $\ln r$ of the original data, using MI-suggested k values (k values decrease from 16 to 14 as ε_{21} increases) and with $l = 1$. (The corresponding values for k through the use of AF changed from 14 to 12 with the increase of ε_{21}.) From this figure, it is clear that the $TE_o(2\rightarrow1)$ is significantly greater than $\mu(TE_s) + 2.68 \times \sigma(TE_s)$ almost over the entire range of coupling, where $\mu(TE_s)$ is the mean of TE_s over 50 surrogate values and $2.68 \times \sigma(TE_s)$ is the error bar on the distribution of TE_s (49 degrees of freedom) at the $\alpha = 0.005$ level. [For very small values of coupling ($\varepsilon_{21} < 0.02$), detection of the direction of information flow is not possible ($p > 0.05$).] Also, TE_o shows a progressive increase in the direction $2\rightarrow1$, proportional to the increase of coupling in that direction, and no significant change in the direction $1\rightarrow2$. In Fig. 15.2b, the $TE_o(\varepsilon_{21})$ and the mean and 99.5% error bars on the distribution of $TE_s(\varepsilon_{21})$ are illustrated for a pair of arbitrary chosen values for k and l (e.g., $k = l = 5$). Neither a statistically significant preferential direction of information flow ($1\rightarrow2$ or $2\rightarrow1$) nor a statistically significant progressive increase in TE_o values with the increase of coupling ε_{21} were observed, due to erroneous selection of k and l for the estimation of TE.

In Fig. 15.2c, we present the same quantities as in Fig. 15.2a, but they now are estimated as averages of TEs over an intermediate range of values of r [$\sigma/5 < \ln r < 2\sigma/5$] (that is, not at r^*). We also observe that the TE values in Fig. 15.2c are larger than the ones in Fig. 15.2a with $r = r^*$ and that it is possible from Fig. 15.2c to detect the correct direction of flow and its significant changes with the strength of coupling. This result is very important for the estimation of TE in practical applications, where an optimal r^* is difficult to obtain. In Fig. 15.2d, we show the values of the measure of causality NTE and its statistical significance for the detection of direction and strength of coupling in the two coupled oscillator system over a range of ε_{21}. NTE was also estimated as an average of NTEs over intermediate values of r [$\sigma/5 < \ln r < 2\sigma/5$]. From the statistically significant values of NTE, it is clear that oscillator 2 drives oscillator 1 and the degree of driving increases proportional to the increase in their coupling.

15.3.2 Robustness to Noise

In order to assess the practical usefulness of this methodology for the detection of causality in noise-corrupted data, Gaussian noise with variance corresponding to a 10 or 3 dB signal-to-noise ratio (SNR) was added independently to the X series of the original data from each of the two coupled *Rössler* systems. The noisy data were then processed in the same way as the noise-free data, including testing against the null hypothesis that an obtained value of TE at each coupling value ε_{21} is not statistically significant. The corresponding TE_o and TE_s values are illustrated in Fig. 15.2e, f for each of the two SNR values, respectively. It is noteworthy that only at extremely low values of coupling ($\varepsilon_{21} < 0.02$) NTE cannot detect the direction of information flow. Thus, it appears that the NTE, along with the suggested

improvements and modifications for the estimation of TE, is a robust measure for detecting the direction and the rate of the net information flow in coupled nonlinear systems even under severe noise conditions (e.g., SNR = 3 dB).

15.4 Discussion and Conclusion

In this study, we suggested and implemented improvements for the estimation of transfer entropy (TE), a measure of the direction and the level of information flow between coupled subsystems, built upon it to introduce a new measure of information flow, and showed their application to a simulation example. The two innovations we introduced in the TE estimation were: (a) the distance in the state space at which the required probabilities should be estimated and (b) the use of surrogate data to evaluate the statistical significance of the estimated TE values. The new estimator for TE was shown to be consistent and reliable when applied to complex signals generated by systems in their chaotic regime. A more practical estimator of TE, that averages the values of TE produced in an intermediate range of distances r in the state space, was shown to be robust to additive noise up to S/N=3 dB, and could reliably and significantly detect the direction of information flow for a wide range of coupling strengths, even for coupling strengths close to 0. Our analysis in this chapter dealt with only pairwise (bivariate) interactions between subsystems and as such, it does not detect both direct and indirect interactions among multiple subsystems at the time resolution of the sampling period of the data involved. A multivariate extension of TE to detect information flow between more than two subsystems is straightforward. Such an extension could also be proven useful in distinguishing between direct and indirect interactions [6,5], and thus further enhance TE's capability to detect causal interactions from experimental data.

A new measure of causality, namely net transfer of entropy [NTE($i \rightarrow j$)], was then introduced for a system i driving a system j (see Equation (15.6)). NTE of the system i measures the outgoing net flow of information from the driving i to the driven j system, that is, it takes into consideration both incoming TE to and outgoing TE from the driving system i. Our simulation example herein also showed the importance of NTE for the identification of the driving system in a pair of coupled systems for a range of coupling strengths and noise levels. We believe that our approach to estimating information flow between coupled systems can have several potential applications to coupled complex systems in diverse scientific fields, from medicine and biology, to physics and engineering.

References

1. Bharucha-Reid, A. Elements of the Theory of Markov Processes and Their Applications. Courier Dover Publications, Chemsford, MA (1997)
2. Chen, Y., Rangarajan, G., Feng, J., Ding, M. Analyzing multiple nonlinear time series with extended Granger causality. Phys Lett A **324**(1), 26–35 (2004)
3. Eckmann, J., Ruelle, D. Ergodic theory of chaos and strange attractors. In: Ruelle, D. (ed.) Turbulence, Strange Attractors, and Chaos, pp. 365–404. World Scientific, Singapore (1995)
4. Efron, B., Tibshirani, R. An Introduction to the Bootstrap. CRC Press, Boca Raton (1993)
5. Franaszczuk, P., Bergey, G. Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. Brain Topogr **11**(1), 13–21 (1998)
6. Friston, K. Brain function, nonlinear coupling, and neuronal transients. Neuroscientist **7**(5), 406–418 (2001)
7. Gaspard, P., Nicolis, G. What can we learn from homoclinic orbits in chaotic dynamics? J Stat Phys **31**(3), 499–518 (1983)
8. Grassberger, P. Finite sample corrections to entropy and dimension estimates. Phys Lett A **128**(6–7), 369–373 (1988)
9. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. Phys Rep **441**(1), 1–46 (2007)
10. Iasemidis, L.D., Sackellares, J.C., Savit, R. Quantification of hidden time dependencies in the EEG within the framework of nonlinear dynamics. In: Jansen, B., Brandt, M. (eds.) Nonlinear Dynamical Analysis of the EEG, pp. 30–47. World Scientific, Singapore (1993)
11. Kaiser, A., Schreiber, T. Information transfer in continuous processes. Physica D **166**, 43–62 (2002)
12. Katz, R. On some criteria for estimating the order of a Markov chain. Technometrics **23**(3), 243–256 (1981)
13. Martinerie, J., Albano, A., Mees, A., Rapp, P. Mutual information, strange attractors, and the optimal estimation of dimension. Phys Rev A **45**(10), 7058–7064 (1992)
14. Pawelzik, K., Schuster, H. Generalized dimensions and entropies from a measured time series. Phys Rev A **35**(1), 481–484 (1987)
15. Pereda, E., Quiroga, R.Q., Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. Prog Neurobiol **77**(1–2), 1–37 (2005)
16. Politis, D., Romano, J., Wolf, M. Subsampling, Springer Series in Statistics, Springer Verlag, New York (1999)
17. Quiroga, R.Q., Arnhold, J., Lehnertz, K., Grassberger, P. Kulback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients. Phys Rev E **62**(6), 8380–8386 (2000)
18. Sabesan, S., Narayanan, K., Prasad, A., Spanias, A. and Iasemidis, L. Improved measure of information flow in coupled nonlinear systems. In: Proceedings of International Association of Science and Technology for Development, pp. 24–26 (2003)
19. Sabesan, S., Narayanan, K., Prasad, A., Tsakalis, K., Spanias, A., Iasemidis, L. Information flow in coupled nonlinear systems: Application to the epileptic human brain. In: Pardalos, P., Boginski, V., Vazacopoulos, A. (eds.) Data Mining in Biomedicine, Springer Optimization and Its Applications Series, Springer, New York, pp. 483–504 (2007)
20. Schiff, S., So, P., Chang, T., Burke, R., Sauer, T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. Phys Rev E **54**(6), 6708–6724 (1996)
21. Schreiber, T. Determination of the noise level of chaotic time series. Phys Rev E **48**(1), 13–16 (1993)
22. Schreiber, T. Measuring information transfer. Phys Rev Lett **85**(2), 461–464 (2000)
23. Theiler, J. Spurious dimension from correlation algorithms applied to limited time-series data. Phys Rev A **34**(3), 2427–2432 (1986)
24. Wiener, N. Modern Mathematics for the Engineers [Z]. Series 1. McGraw-Hill, New York (1956)