Exploring neural directed interactions with transfer entropy based on an adaptive kernel density estimator

Kai Zuo [1 2 *], **Jean-Jacques Bellanger** [2 3], **Chunfeng Yang** [1 2], **Huazhong Shu** [1 2], **Regine Le Bouquin Jeannes** [2 3 *]

[1] *LIST, Laboratory of Image Science and Technology* [*Nanjing*] *SouthEast University , School of Computer Science and Engineering , Si Pai Lou 2, Nanjing, 210096, CN*

[2] *CRIBS, Centre de Recherche en Information Biomédicale sino-français INSERM : Laboratoire International AssociÉ , Université de Rennes 1 , 9 Rue Jean Macé, 35700 Rennes, FR*

[3] *LTSI, Laboratoire Traitement du Signal et de l'Image INSERM : U1099 , Université de Rennes 1 , Campus Universitaire de Beaulieu - Bât 22 - 35042 Rennes, FR*

* Correspondence should be addressed to: Kai Zuo <zuokai1989@126.com >
* Correspondence should be addressed to: Regine Le Bouquin Jeannes <regine.le-bouquin.jeannes@univ-rennes1.fr >

Abstract

This paper aims at estimating causal relationships between signals to detect flow propagation in autoregressive and physiological models. The main challenge of the ongoing work is to discover whether neural activity in a given structure of the brain influences activity in another area during epileptic seizures. This question refers to the concept of effective connectivity in neuroscience, i.e. to the identification of information flows and oriented propagation graphs. Past efforts to determine effective connectivity rooted to Wiener causality definition adapted in a practical form by Granger with autoregressive models. A number of studies argue against such a linear approach when nonlinear dynamics are suspected in the relationship between signals. Consequently, nonlinear nonparametric approaches, such as transfer entropy (TE), have been introduced to overcome linear methods limitations and promoted in many studies dealing with electrophysiological signals. Until now, even though many TE estimators have been developed, further improvement can be expected. In this paper, we investigate a new strategy by introducing an adaptive kernel density estimator to improve TE estimation.

Author Keywords bioelectric phenomena ; brain models ; causality ; electroencephalography ; entropy ; medical signal detection ; neurophysiology ; nonlinear dynamical systems ; stochastic processes

Introduction

In neuroscience, recent works have been devoted to detecting effective connectivity [1] defined as a causal influence of the dynamics of a first system on the dynamics of a second one. In this context, two questions are commonly addressed: (i) how to choose a formal quantitative definition of effective connectivity and (ii) how to provide corresponding estimators defined as functions of signals recorded in both systems. Nowadays two approaches contrast. The first one does not rely on an underlying physiological model while the second one, namely dynamical causal modeling, does. In this contribution, we are only concerned with the first approach including linear and nonlinear methodologies, and we consider nonlinear nonparametric entropic characterization of this connectivity using the so-called transfer entropy (TE). When computed on a stationary bivariate time series (X, Y), this quantity measures the amount of information transferred from channel X (resp. Y) to channel Y (resp. X) and is denoted $TE_{\rightarrow}$ (resp. $TE_{\rightarrow}$) hereafter. It was introduced by Schreiber [2] and defined as the Kullback-Leibler divergence between two different predictive probability distributions of Y. The first distribution is defined conditionally to amplitudes of X , $n' < n$ and Y , $n' < n$, at time instants prior to n, and the second one is only defined conditionally to Y , $n' < n$. A simple exchange of X and Y leads to the definition of $TE_{\rightarrow}$. Formal definition is given by

$$TE_{X\to Y}=\int_{\mathbb{R}^{k+l+1}}\log\left(\frac{p_{Y_{n+1}/Y_n^k,X_n^l}(y,y^k,x^l)}{p_{Y_{n+1}/Y_n^k}(y,y^k)}\right)\times p_{Y_{n+1},Y_n^k,X_n^l}(y,y^k,x^l)\,d(y,y^k,x^l)$$

(1)

with $U_n^m \triangleq (U_n, \ldots, U_{n-m+1})$, p (u) denotes the probability density of a random vector U at u, k and l are the predictor dimensions. Let us note that the definition of TE is qualitatively consistent with Wiener and Granger approaches, which only compare mean square prediction errors. In theory, $TE_{\rightarrow}$ is never negative and is equal to zero iif

$$p_{Y_{n+1}/Y_n^k,X_n^l}(y,y^k,x^l)=p_{Y_{n+1}/Y_n^k}(y,y^k),\ y\in\mathbb{R}$$

(2)

The choice of k and l can impact drastically on theoretical TE value and, without a priori information on the hidden nonlinear dynamics generating (X, Y), this issue is not trivial and is not discussed in this paper. Given the theoretical index and an N point observation (X, Y), $n = 1..N$, we have to determine an estimation procedure to compute $\hat{TE}_{x \to y}$ from

$$(y_{n+1}, y_n^k, x_n^l),\ n = n_0, \ldots, N-1\ (n_0 = 1 + \max(k, l))$$

(3)

If all probability densities are known, the trivial Monte Carlo estimator could be:

$$\frac{1}{N - n_0} \sum_{n=n_0}^{N-1} \log \left(\frac{p_{y_{n+1}, y_n^k, x_n^l}(y_{n+1}, y_n^k, x_n^l) p_{y_n^k}(y_n^k)}{p_{y_{n+1}, y_n^k}(y_{n+1}, y_n^k) p_{y_n^k, x_n^l}(y_n^k, x_n^l)} \right)$$

(4)

Since the densities are unknown, a first method consists in replacing each density p by an estimation $\hat{p}$ possibly obtained by a fixed size kernel estimation approach as proposed in [3]. A second method computing estimations $\log \hat{(p_U)}$ of log (p) from K Nearest Neighbors (KNN) selection was developed for mutual information estimation in [4] and applied in [5]. It is also possible to compute the estimation of a density p with adaptive size kernels. We propose this improvement to compute TE and compare our results on linear Gaussian models (i) with corresponding theoretical values, and (ii) with TE estimated using a fixed bandwidth kernel and/or KNN kernel as in [5]. Then, these methods are compared on a neurophysiological model [11].

Methods and Materials

Kernel methods

In order to reconstruct a probability density p from N observed states u, the general form of a fixed kernel density estimator (FKDE) of bandwidth h is given by:

$$\hat{p}_U(u) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{h} K\left(\frac{u - u_n}{h}\right) = \frac{1}{N} \sum_{n=1}^{N} K_h(u - u_n)$$

(5)

where K denotes a kernel function. For joint density probability estimated at ($\boldsymbol{y}$, $\boldsymbol{y}$, $\boldsymbol{x}$), we write:

$$\hat{p}_{yyx}(\boldsymbol{y}, \boldsymbol{y}^k, \boldsymbol{x}^l) = \frac{1}{|\Delta|} \sum_{m \in \Delta} \frac{1}{h_y^{k+1} h_x^l} K\left(\frac{\boldsymbol{y} - y_{m+1}}{h_y}\right)$$
$$\prod_{j=1}^{k} K\left(\frac{\boldsymbol{y}^k(j) - y_{m-j+1}}{h_y}\right) \prod_{j=1}^{l} K\left(\frac{\boldsymbol{x}^l(j) - x_{m-j+1}}{h_x}\right)$$

(6)

where $\Delta = \{m / \max(k, l) < m \leq N - 1, |m - n| > \tau\}$, and τ is the decorrelation time defined as the minimum delay leading to a correlation coefficient equal to 0.1. Parameters h, h are the respective kernel bandwidths for signals x and y which are normalized. A fixed bandwidth (independent of m) is unable to deal satisfactorily with the tails of the distribution without over smoothing the main part of this distribution. To avoid this issue, two methods help in estimating a density f at a point x :

$$\hat{f}_1(x) = \frac{1}{N} \sum_{m=1}^{N} K_{h_x}(x - x_m) \text{ where } h_x \equiv h(x)$$

(7)

$$\hat{f}_2(x) = \frac{1}{N} \sum_{m=1}^{N} K_{h_m}(x - x_m) \text{ where } h_m \equiv h(x_m)$$

(8)

The first method (7) uses a x dependent bandwidth, this bandwidth being unchanged for different points x. One example of this method is a KNN estimator [6]. The second method (8) uses a x dependent bandwidth, which does not depend on x, leading to the adaptive kernel density estimator (AKDE) [7] we adopted.

Adaptive kernel density estimator

AKDE is an improved alternative to FKDE. Given an initial bandwidth h and a first FKDE based estimation $\hat{f}_0$, Abramson [8] adapted the bandwidth according to these initial quantities:

$$h_m \propto h_0 / \sqrt{\hat{f}_0(x_m)}$$

(9)

In [9], Hwang extended this procedure:

$$h_m = (\hat{f}_0(x_m)/g)^{-r} h_0$$

(10)

where g is the geometric mean of $(\hat{f}_0(x))$, i.e.

$$\log g = \frac{1}{N}\sum_{m=1}^{N} \log \hat{f}_0(x_m)$$

(11)

r is an user defined sensitivity parameter generally satisfying $0 < r < 1$. For different dimensions, the value of r should change. However, it is difficult to choose the proper r for four different probability density estimations. We suggest to compute $\hat{p}_{YYX}(y_{n+1}, y_n^k, x_n^l)$ before its substitution in (4) using (10) and to leave the other densities unchanged. The three steps of the proposed algorithm are as follows:

- *Step 1*: for n (≈10) values v of h, $0 < v_1 < .. < v_{nh} \leq 3$, compute fixed kernel density estimations at (y_{n+1}, y_n^k), (y_n^k, x_n^l), y_n^k and (y_{+1}, y_n^k, x_n^l) ;

- *Step 2*: compute h using (10) only to update $\hat{p}_{YYX}(y_{n+1}, y_n^k, x_n^l)$ leaving the other densities unchanged, and search the set V of values v leading to unimodality of $f_{v_i} : r \in \{0 < r_1, \ldots, r_{20} < 1\} \to \hat{TE}_{x\to y}(h_m(v, r)) = f_{v_i}(r)$ for $v \in \{v_1, .., v_{nh}\}$ hence eliminating monotonic curves (no multimodal curve was observed);

- *Step 3*: finally retain the maximum value $\hat{TE}_{x\to y} = \max_{(h,r)\in V_h \times R} (\hat{f}_h(r))$.

Experimentally, the selected value h in the last step is often close to the initial bandwidth h_0. Clearly, the computation time is increased with AKDE (multiplied by 20) for updating the density in step 2.

Experimental Results

We tested our method with Gaussian kernels to compute $TE_{\rightarrow}$ and $TE_{\rightarrow}$ on two kinds of signals. The first kind included two toy linear AutoRegressive (AR) models and the second one was a realistic EEG model. Predictor dimensions k and l were chosen equal to the corresponding AR models orders estimated by the generalized Bayesian Information Criterion as in [10]. For AR models, the decorrelation time was τ=20 and experiments were repeated 200 times on 1024-point signals to get averaged values.

Unidirectional linear model

For the first linear stochastic system (model 1), the following two signals were generated:

$$\begin{cases} x(n) = 1.3435x(n-1) - 0.9025x(n-2) + e_1(n) \\ y(n) = 0.5x(n-3) - 0.4y(n-2) + e_2(n) \end{cases}$$

(12)

where e_1 and e_2 were independent white Gaussian noises with zero means and unit variances.

Fig. 1 displays TE computed with a fixed bandwidth h. Experimental transfer entropy $\hat{TE}_{x\to y}$ is smaller than the theoretical value which is equal to Granger Causality index divided by 2 (GC/2 was computed from the model coefficients) in the case of Gaussian signals (see Table I). Fig. 2 corresponds to TE values vs. r using AKDE. The flow propagation from signal x to signal y was correctly established whereas the estimated influence from signal y to signal x was not significant. Step 3 of the algorithm led to h = 0.45. Comparing Fig. 1 and Fig. 2, TE estimated using AKDE is much closer to the exact value (0.41). We also compared our estimator with Trentool toolbox [5] and concluded to its relevant behavior as seen in Table I which allows to compare the different estimators in terms of mean and standard deviation. It reveals visible improvement in $\hat{TE}_{x\to y}$ all other performance with Gaussian AKDE over estimators.

Bidirectional linear model

For the second linear stochastic system (model 2), we generated the following signals:

$$\begin{cases} x(n) = 0.5x(n-1) + 0.3y(n-2) + e_1(n) \\ y(n) = 0.5x(n-3) - 0.4y(n-1) + e_2(n) \end{cases}$$

(13)

where e_1 and e_2 were as in (12). Fig. 3 and Fig. 4 allow to compare TE values using either a fixed bandwidth (Fig. 3) or AKDE (Fig. 4, h = 0.45). For this model, the exact value of TE from signal x to signal y (resp. from signal y to signal x) given in Table II is represented by a solid grey line (resp. a dashed grey line) in Fig. 3 and 4. Focusing on Fig. 4, the bidirectional flow propagation was correctly detected using AKDE, the mean values of TE being close to the exact ones (see also Table II). This figure reveals that the bias of AKDE estimator

is negligible. As for TE estimated with a fixed bandwidth (Fig. 3 and Table II), its values remain lower than the exact ones, similarly as those estimated with Trentool toolbox. For all estimators tested, the standard deviation is 5 to 10 times lower than the corresponding mean value.

Physiology based model

We simulated EEG signals with a nonlinear SDE (stochastic differential equation) model [11] of order 20 to represent the activities of two neuronal populations *Pop* and *Pop*:

$$d[X_I^T X_E^T] = g_X(X_I, X_E, \theta)dt + dW_X$$
$$d[Y_I^T Y_E^T] = g_Y(Y_I, Y_E, \theta, K_{XY} Sig(X))dt + dW_Y$$
$$X(t) = L(X_I(t), X_E(t)), Y(t) = L(Y_I(t), Y_E(t))$$

(14)

where the line vectors $X_I^T X_E^T$ and $Y_I^T Y_E^T$ are in $\mathbb{R}^{10}$ (*T* is the transpose operator) and denote dynamical state evolutions for *Pop* and *Pop* aggregating components [] and [] modeling respectively interacting inhibitory and excitatory activities. Parameterized nonlinear functions (*u*, *v*) → *g* (*u*, *v*, θ) drive the state evolutions. Sig(.) is a nonlinear sigmoidal function. The coupling parameter *K* allows effective connectivity adjustment from the first population to the second one. *X* and *Y* are computed with a linear function *L* and are interpretable as two intracranial EEG signals recorded from proximal field electrodes. The independent Brownian processes *W* and *W* represent random surrounding populations influences. The parameters vector θ includes three scalar components *A*, *B* and *G*, allowing to modify the type of activity (normal/epileptic). The model was time discretized by Euler scheme to produce two discrete time outputs. As in [11], we fixed these parameters to 5, 3, 20 in *Pop* and to 3.5, 3.5, 84 in *Pop*. This resulted in a narrow band fast activity around 25 Hz (similar to that observed at seizure onset) in each population. *K* was set to 1500. Fifty blocks of 8-second length signals were simulated with a sampling rate of 256 Hz. In this experiment, as the influence from one physiological signal to another one may be largely delayed, and to get a not too large predictor dimension *l*, we first shifted signal *Y* as proposed in [1]. The delay corresponded to 33 sampling time instants and was determined from cross covariance maximization. The maximum order in the model (after shifting) was set to 2 and τ was set to 500. According to Fig. 5 and Table III, we conclude to the relevance of the new estimator compared to the “references” given by Granger causality index (GC/2) and Trentool toolbox. As a matter of fact, $\hat{TE}_{x\to y}$ and $\hat{TE}_{y\to x}$ are sensibly more contrasted (considering means and standard deviations) with Gaussian AKDE than with the two other methods (Table III). Moreover, when comparing the mean values of $\hat{TE}$ obtained for the different methods with this physiological model, a larger dispersion was observed than with previous linear models 1 and 2. The difference between the AKDE based estimator and GC/2 could be expected due to the nonrobustness of Granger index to nonlinearities. On the other hand, the difference between the AKDE based estimator and Trentool estimator (which even failed in detecting the flow direction) was unexpected.

Conclusion

In this paper, we focused on information propagation between two observations using TE and introduced an adaptive kernel density estimator to improve fixed kernel TE estimator. Results on simulated AR models revealed a very low bias with AKDE approach and proved the relevance of this new method in detecting uni/bi-directional propagation flows. Using a fixed bandwidth or Trentool approach led to much more biased values. For physiological signals, even if we had no ground-truth, the causal effects were perfectly identified and allowed characterizing the driving system and the responding one. In the future, the AKDE method will be tested on real EEG signals and on more complex scenarios including stronger nonlinearities and/or multivariate observations. A validation phase including statistical hypothesis tests based on surrogate data will complete this work.

References:

- 1 . Vicente R , Wibral M , Lindner M , Pipa G . Transfer entropy - a model-free measure of effective connectivity for the neurosciences . J of Computational Neuroscience . 30 : 45 - 67 2011 ;
- 2 . Schreiber T . Measuring information transfer . Physical Review Letters . 85 : 461 - 464 2000 ;
- 3 . Kaiser A , Schreiber T . Information transfer in continuous processes . Physica D . 166 : 43 - 62 2002 ;
- 4 . Kraskov A , Stögbauer H , Grassberger P . Estimating mutual information . Physical review E . 69 : (6) 066138:1 - 16 2004 ;
- 5 . Lindner M , Vicente R , Priesemann V , Wibral M . TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy . BMC Neuroscience . 12 : 119 - 2011 ;
- 6 . Loftsgaarden DO , Quesenberry CP . A nonparametric estimate of a multivariate density function . The Annals of Mathematical Statistics . 36 : (3) 1049 - 1051 1965 ;
- 7 . Silverman BW . Density estimation for statistics and data analysis, Monographs on Statistics and Applied Probability . London Chapman & Hall ; 1986 ;
- 8 . Abramson IS . On bandwidth variation in kernel estimate-A square root law . Annals Statist . 10 : 1217 - 1223 1982 ;
- 9 . Hwang JN . Nonparametric multivariate density estimation: a comparative study . IEEE Trans Signal Processing . 42 : 2795 - 2810 1994 ;
- 10 . Yang C , Le Bouquin Jeannès R , Bellanger JJ , Shu H . A new strategy for model order identification and its application to transfer entropy for EEG signals analysis . IEEE Trans on Biomedical Engineering . (under press)
- 11 . Wendling F , Hernandez A , Bellanger JJ , Chauvel P , Bartolomei F . Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG . J of Clinical Neurophysiology . 22 : 343 - 356 2005 ;

Figure 1

Results of TE between two time series *x* and *y* (model 1) using a Gaussian kernel and a fixed bandwidth *h* Solid line: $\hat{TE}_{x\to y}$, dashed line: $\hat{TE}_{y\to x}$ Horizontal line: exact value $TE_{\rightarrow}$=0.41

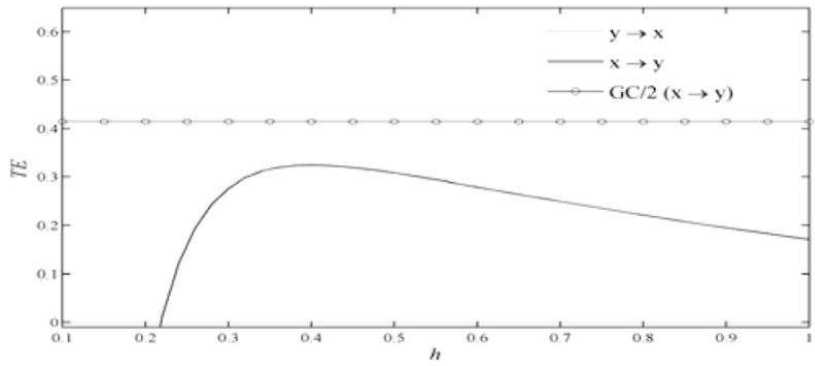

Figure 2

Results of TE between two time series *x* and *y* (model 1) using a Gaussian kernel and AKDE (10) Solid line: $\hat{TE}_{x\to y}$, dashed line: $\hat{TE}_{y\to x}$ Horizontal line: exact value $TE_{\rightarrow}$=0.41

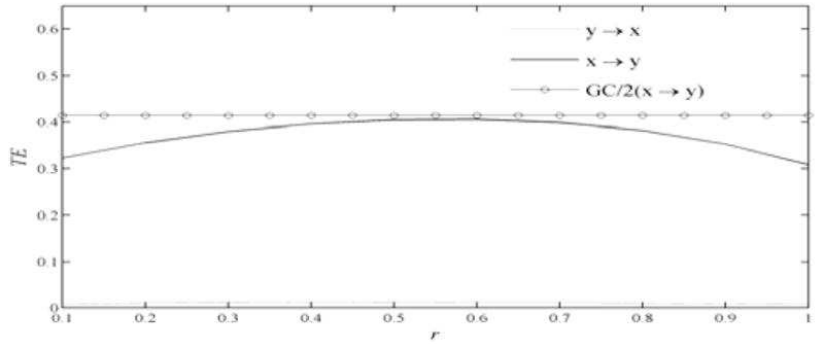

Figure 3

Results of TE between two time series *x* and *y* (model 2) using a Gaussian kernel and a fixed bandwidth *h* Solid line: $\hat{TE}_{x\to y}$, dashed line: $\hat{TE}_{y\to x}$, Horizontal lines: exact values

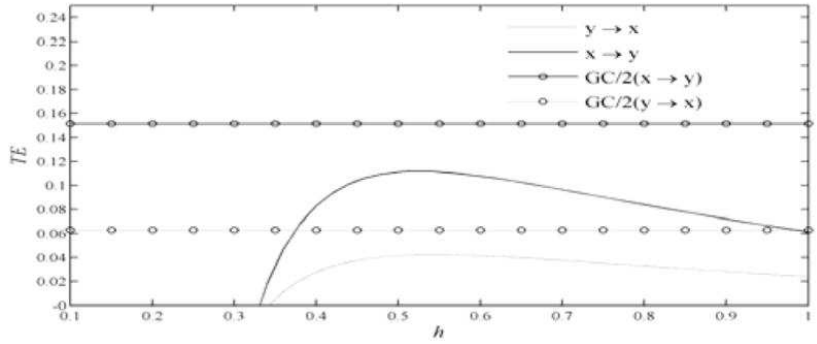

Figure 4

Results of TE between two time series *x* and *y* (model 2) using a Gaussian kernel and AKDE (10) Solid line: $\hat{TE}_{x\to y}$, dashed line: $\hat{TE}_{y\to x}$, Horizontal lines: exact values

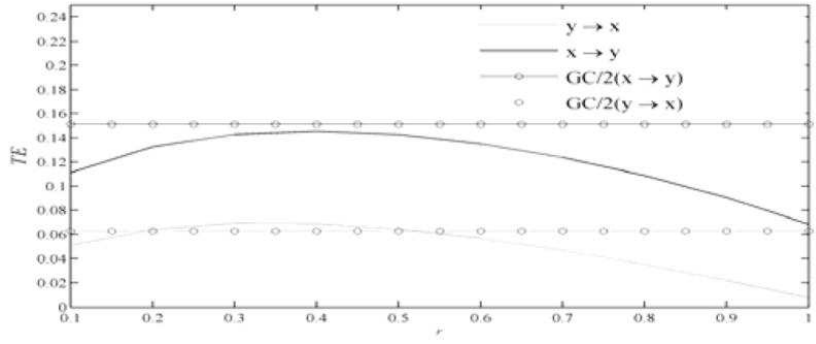

Figure 5

Results of TE between two delayed simulated EEG signals using a Gaussian kernel and AKDE (10) Solid line: $\hat{TE}_{x-y}$, dashed line: $\hat{TE}_{y-x}$

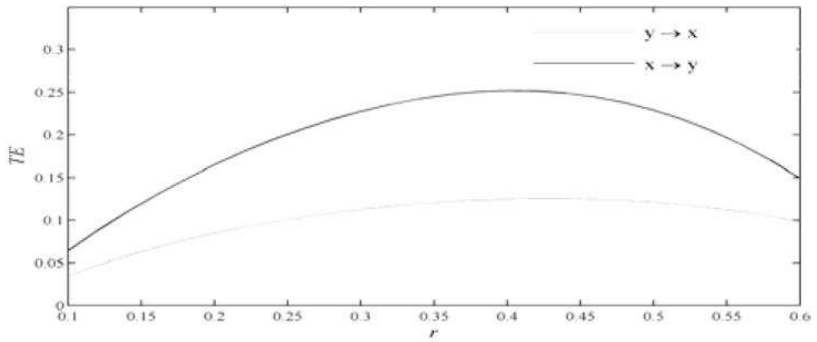

TABLE I

Mean values (and standard deviation in parentheses) of transfer entropy using the different estimators on model 1

Estimator	$X \to Y$	$Y \to X$
GC/2	0.4146	0
TE (fixed *h*)	0.3242 (0.0158)	0 (0)
TE (AKDE)	0.4063 (0.0179)	0.01267(0.0045)
Trentool	0.3484 (0.0115)	−0.0158 (0.0070)

TABLE II

Mean values (and standard deviation in parentheses) of transfer entropy using the different estimators on model 2

Estimator	$X \to Y$	$Y \to X$
GC/2	0.1511	0.0630
TE (fixed *h*)	0.1118 (0.0123)	0.0422 (0.0083)
TE (AKDE)	0.1457 (0.0133)	0.0689 (0.0087)
Trentool	0.1120 (0.0091)	0.0446 (0.0079)

TABLE III

Mean values (and standard deviation in parentheses) of transfer entropy using the different estimators on the physiological model

Estimator	$X \to Y$	$Y \to X$
GC/2	0.011 (0.0193)	0.0028 (0.0009)
TE (AKDE)	0.2521 (0.1143)	0.1249 (0.0615)
Trentool	0.0049 (0.3182)	0.0091 (0.0083)