

MIT Open Access Articles

Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Lee, Joon et al. "Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series." *BioMedical Engineering OnLine* 11.1 (2012): 19. © 2012 BioMed Central Ltd

As Published: <http://dx.doi.org/10.1186/1475-925x-11-19>

Publisher: Biomed Central Ltd.

Persistent URL: <http://hdl.handle.net/1721.1/75000>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution



RESEARCH

Open Access

Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series

Joon Lee^{1*}, Shamim Nemati^{2,3}, Ikaro Silva¹, Bradley A Edwards³, James P Butler³ and Atul Malhotra³

* Correspondence: joonlee@mit.edu

¹Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

Full list of author information is available at the end of the article

Abstract

Background: The detection of change in magnitude of directional coupling between two non-linear time series is a common subject of interest in the biomedical domain, including studies involving the respiratory chemoreflex system. Although transfer entropy is a useful tool in this avenue, no study to date has investigated how different transfer entropy estimation methods perform in typical biomedical applications featuring small sample size and presence of outliers.

Methods: With respect to detection of increased coupling strength, we compared three transfer entropy estimation techniques using both simulated time series and respiratory recordings from lambs. The following estimation methods were analyzed: fixed-binning with ranking, kernel density estimation (KDE), and the Darbellay-Vajda (D-V) adaptive partitioning algorithm extended to three dimensions. In the simulated experiment, sample size was varied from 50 to 200, while coupling strength was increased. In order to introduce outliers, the heavy-tailed Laplace distribution was utilized. In the lamb experiment, the objective was to detect increased respiratory-related chemosensitivity to O_2 and CO_2 induced by a drug, domperidone. Specifically, the separate influence of end-tidal PO_2 and PCO_2 on minute ventilation (\dot{V}_E) before and after administration of domperidone was analyzed.

Results: In the simulation, KDE detected increased coupling strength at the lowest SNR among the three methods. In the lamb experiment, D-V partitioning resulted in the statistically strongest increase in transfer entropy post-domperidone for $PO_2 \rightarrow \dot{V}_E$. In addition, D-V partitioning was the only method that could detect an increase in transfer entropy for $PCO_2 \rightarrow \dot{V}_E$, in agreement with experimental findings.

Conclusions: Transfer entropy is capable of detecting directional coupling changes in non-linear biomedical time series analysis featuring a small number of observations and presence of outliers. The results of this study suggest that fixed-binning, even with ranking, is too primitive, and although there is no clear winner between KDE and D-V partitioning, the reader should note that KDE requires more computational time and extensive parameter selection than D-V partitioning. We hope this study provides a guideline for selection of an appropriate transfer entropy estimation method.

Introduction

In multi-variable time series analysis, a common subject of interest is the coupling among the variables. One promising measure of the coupling strength between two time series is transfer entropy [1,2], which quantifies the amount of information transfer from one variable to the other. Importantly, transfer entropy is non-parametric and can capture non-linear coupling effects. This property can be useful in analyzing complex systems where interactions among sub-systems are expected to be non-linear and where minimal a priori knowledge is available. Furthermore, transfer entropy is an asymmetric measure that conveys directional information. Unlike mutual information, which can only quantify the amount of shared information between two variables, transfer entropy can elucidate directional relationships between variables. In fact, it has been shown that transfer entropy is a non-linear extension of Granger causality [3].

Given all these advantages, transfer entropy has been gaining popularity as a powerful analytic tool for characterizing complex physiologic networks. For instance, this new method has been suggested [4] and successfully applied to electroencephalograms and magnetoencephalograms to help elucidate complex neural phenomena such as consciousness under anesthesia [5], sensory functional integration [6], auditory neural assembly [7], attention related processing [8], and sensor-motor connectivity [9]. In the area of cardiovascular physiology, Katura and colleagues [10] have utilized the technique of transfer entropy to elucidate (or reject) causal relationships that explain sources of variability in the regulation of cerebral hemodynamics. In addition, a number of studies have employed variations of transfer entropy, such as cross-conditional entropy and conditional mutual information, to investigate relationships between cardiovascular variables as indicator of baroreflex function [11-14]. Conditional mutual information has been shown to be equivalent to transfer entropy [15].

In practice, one major challenge with transfer entropy estimation using biomedical signals is the estimation of Probability Density Functions (PDFs) from finite data with outliers. This estimation challenge is well discussed in [2,16]. Small sample size can stem not only from experimental and/or technical constraints but also from the need for stationarity; a non-stationary time series may have to be broken into quasi-stationary segments before analysis. This sample size issue can be a major hindrance for biomedical studies where data collection is often costly. Also, outliers frequently originate from either noise sources or underlying physiology.

Therefore, our objective in this study was to investigate several transfer entropy estimation methodologies in the context of biomedical signals with small sample size and with outliers. This study can be viewed as an extension of the study on mutual information estimation (two dimensions) by Khan and colleagues [17] to transfer entropy estimation (three dimensions). Furthermore, unlike most other transfer entropy applications, we focused on the ability to detect increases in coupling strength, rather than on detection of the existence of a significant coupling or on identification of time lags at which significant couplings exist. This is a common research topic of interest when one wishes to study how a given coupling varies in magnitude under different conditions while the existence of the coupling is known a priori (our application to respiratory physiology in this study is one example). To the best of our knowledge, no study has attempted to investigate the sensitivity of various transfer entropy estimation methods to changes in coupling strength in the biomedical domain.

Specifically, we studied three non-parametric techniques in this endeavor: fixed-bins, Kernel Density Estimation (KDE), and adaptive partitioning. The fixed-bin approach, in which a given time series is quantized by dividing the dynamic range into equally-spaced bins, is the simplest method that boasts the least computational complexity at the cost of inefficient bin allocation [16]. KDE has been shown to perform well in mutual information estimation using short time series [17] and hence was included in our study. For adaptive partitioning, we introduce a novel method that applies the Darbellay-Vajda (D-V) partitioning algorithm [18] to transfer entropy estimation. The D-V partitioning algorithm is equivalent to the adaptive partitioning introduced by Fraser and Swinney [2,19] and has been shown to be effective in mutual information estimation [20].

The comparison among the three transfer entropy estimation techniques was accomplished via both simulated and real-life signals. In both experiments, we analyzed the sensitivity of the techniques to increased non-linear coupling strength, under small sample size and presence of outliers. The purpose of the simulated experiment was to perform a rigorous comparison in a controlled setting. In order to address any coupling characteristics that were not represented by the simulated data, we also applied the transfer entropy estimation methods to measurements of spontaneous breathing respiratory flow, end-tidal carbon dioxide pressure (PCO_2) and end tidal oxygen pressure (PO_2), collected from lambs [21]. The objective of the lamb experiment was to detect changes in chemosensitivity, which is of major clinical interest. The method of transfer entropy provides a natural framework for quantification of directional influence that breath-by-breath changes in blood gases have on ventilation (via peripheral chemoreceptors located in the aortic arch and carotid bodies), while taking into account the delayed nature of the response (due to lung-to-chemoreceptors transport time, typically on the order of 2-4 breaths) and possible non-linearities [22]. Ventilatory control instabilities are important in a variety of pathological conditions, including Cheyne-Stokes breathing in congestive heart failure [23,24] and obstructive sleep apnea [25-27]. Although the specific mechanisms underlying each condition may vary, increases in the ventilatory sensitivity to hypoxia/hypercapnia (controller gain) have been shown to play a critical role in the pathogenesis of unstable breathing. An elevated hypoxic ventilatory response at high altitude is associated with the presence of unstable breathing [28]. Moreover, heightened dynamic hypercapnic responses delineate heart failure patients with periodic breathing (in the form of Cheyne-Stokes respiration) from those patients that exhibit stable breathing [29-31]. Additionally, in heart failure patients, elevated hypoxic and/or hypercapnic sensitivity [32] and the consequent presence of Cheyne-Stokes respiration itself [33] are important predictors of mortality.

Methods

Transfer Entropy Construction

Given two concurrently sampled time series $X = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_1, y_2, \dots, y_N\}$, the transfer entropy from X to Y , termed $T_{X \rightarrow Y}$, can be derived from conditional entropies as follows:

$$T_{X \rightarrow Y} = H(y_i | y_{i-t}^{(l)}) - H(y_i | y_{i-t}^{(l)}, x_{i-t}^{(k)}) \quad (1)$$

$$= \sum_{y_i, y_{i-t}^{(l)}, x_{i-\tau}^{(k)}} p(y_i, y_{i-t}^{(l)}, x_{i-\tau}^{(k)}) \log \frac{p(y_i | y_{i-t}^{(l)}, x_{i-\tau}^{(k)})}{p(y_i | y_{i-t}^{(l)})}, \quad (2)$$

where i indicates a given point in time, τ and t are the time lags in X and Y , respectively, and k and l are the block lengths of past values in X and Y , respectively (see Appendix I for a complete derivation). The past values on which the conditional probabilities in (2) are conditioned are $x_{i-\tau}^{(k)} = \{x_{i-\tau-k+1}, x_{i-\tau-k+2}, \dots, x_{i-\tau}\}$ and $y_{i-t}^{(l)} = \{y_{i-t-l+1}, y_{i-t-l+2}, \dots, y_{i-t}\}$. In words, (1) implies that transfer entropy measures the reduction in uncertainty in y_i given $x_{i-\tau}^{(k)}$ and $y_{i-t}^{(l)}$ in comparison with given only $y_{i-t}^{(l)}$. Note that $T_{X \rightarrow Y}$ cannot be negative because $H(y_i | y_{i-t}^{(l)}) \geq H(y_i | y_{i-t}^{(l)}, x_{i-\tau}^{(k)})$ in (1); conditioning on another variable, $x_{i-\tau}^{(k)}$, cannot increase the uncertainty in y_i . (Note that if base 2 logarithm is used, the unit of transfer entropy is bits, which has the interpretation of the amount of reduction in the average length of the optimal code needed to encode the target variable.)

Equation (2) is the most general definition of transfer entropy and provides maximum flexibility. However, the usual parameter choices in practice are $k = 1$ and $l = 1$ for computational reasons when sample size is small. In addition, here we set $t = 1$ under the assumption that the maximum auto-transfer of information occurs from the data point immediately before the target value in Y . For example, Francis et al. only considered lags of one breath in ventilation and CO_2 for the same reason [30]. These choices of $k = l = t = 1$ are especially appropriate in biomedical experiments where time series length is usually short and the absolute values of auto-correlation functions tend to decrease monotonically as time lag increases. Also, the original definition in [1] implicitly assumed $t = 1$ by using a Markov chain; our choice can be seen as a first-order Markov process. Hence, we restrict the focus of this paper to these usual parameter choices by simplifying (2) to

$$T_{X \rightarrow Y}(\tau) = \sum_{y_i, y_{i-1}, x_{i-\tau}} p(y_i, y_{i-1}, x_{i-\tau}) \log \frac{p(y_i | y_{i-1}, x_{i-\tau})}{p(y_i | y_{i-1})} \quad (3)$$

$$= \sum_{y_i, y_{i-1}, x_{i-\tau}} p(y_i, y_{i-1}, x_{i-\tau}) \log \frac{p(y_i, y_{i-1}, x_{i-\tau}) p(y_{i-1})}{p(y_{i-1}, x_{i-\tau}) p(y_i, y_{i-1})}. \quad (4)$$

Note that (3) and (4) explicitly show transfer entropy as a function of τ .

In practice, the foremost challenge in transfer entropy computation is estimating the joint probabilities in (4). The ensuing section describes the three different PDF estimation methodologies to be investigated in this study.

Probability Density Function Estimation Methodologies

Fixed Bins

The simplest estimation approach to obtain the PDFs in (4) is to allocate data points to fixed, equally-spaced bins. In order to enhance robustness against outliers and sparse regions in the underlying distribution, we combined fixed binning with ordinal sampling (ranking). In ordinal sampling, the time series values in X and Y are substituted with their ranks in sorted X and Y , similar to most non-parametric statistical

tests. The ranks are integers ranging from 1 (smallest value) to N (largest value). Let the transformed time series of X and Y be $U = \{u_1, u_2, \dots, u_N\}$ and $V = \{v_1, v_2, \dots, v_N\}$, respectively (U and V contain integers that represent ranks).

Then, (4) can be rewritten in terms of U and V as follows:

$$T_{U \rightarrow V}(\tau) = \sum_{v_i, v_{i-1}, u_{i-\tau}} p(v_i, v_{i-1}, u_{i-\tau}) \log \frac{p(v_i, v_{i-1}, u_{i-\tau})p(v_{i-1})}{p(v_{i-1}, u_{i-\tau})p(v_i, v_{i-1})}. \quad (5)$$

Using the same number of bins, Q , in each dimension for simplicity, the fixed-bin technique estimates (5) as follows:

$$T_{U \rightarrow V}(\tau) \approx \sum_{a=1, b=1, c=1}^Q \frac{m_{a,b,c}}{P} \log \frac{m_{a,b,c} m_b}{m_{b,c} m_{a,b}}, \quad (6)$$

where a , b , and c index bins along v_i , v_{i-1} , and $u_{i-\tau}$ respectively, and P is the total number of triplets of v_i , v_{i-1} , and $u_{i-\tau}$. Furthermore, $m_{a,b,c}$, $m_{a,b}$, and $m_{b,c}$ represent the number of data points in the intersection of the one-dimensional bins denoted by the subscript, whereas m_b is the number of data points in the b th bin along the v_{i-1} dimension.

Kernel Density Estimation

In KDE, the PDF is estimated by summing individual distributions centered at each data point. The shape of the individual distributions is dictated by a chosen kernel, the magnitude of which decreases as the distance from the center of the distribution increases. Since distance is a crucial measure in KDE, ordinal sampling is not employed. In the three-dimensional space of y_i , y_{i-1} , and $x_{i-\tau}$ the joint probability at an arbitrary point, $(\tilde{y}_i, \tilde{y}_{i-1}, \tilde{x}_{i-\tau})$, can be estimated by normalizing the following:

$$p(\tilde{y}_i, \tilde{y}_{i-1}, \tilde{x}_{i-\tau}) \approx \frac{1}{P} \sum_{j=1}^P \frac{1}{h_{y_i} h_{y_{i-1}} h_{x_{i-\tau}}} K\left(\frac{\tilde{y}_i - y_{i,j}}{h_{y_i}}\right) K\left(\frac{\tilde{y}_{i-1} - y_{i-1,j}}{h_{y_{i-1}}}\right) K\left(\frac{\tilde{x}_{i-\tau} - x_{i-\tau,j}}{h_{x_{i-\tau}}}\right), \quad (7)$$

where j indexes the P data points and $h_{(\cdot)}$ is the bandwidth for the dimension specified in the subscript. We utilized a scaled version of the rule of thumb specified in [34] to select the bandwidth:

$$h_{(\cdot)} = 1.06\alpha\hat{\sigma}P^{-1/5}, \quad (8)$$

where α is a multiplier for scaling and $\hat{\sigma}$ is the sample standard deviation in the given dimension. In terms of K , several possible kernels exist but we employed the widely used Gaussian kernel in this study, which is defined as follows:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-0.5u^2} \quad (9)$$

The rest of the probabilities in (4) can be computed by marginalizing $p(\tilde{y}_i, \tilde{y}_{i-1}, \tilde{x}_{i-\tau})$.

Adaptive Partitioning using the Darbellay-Vajda Algorithm

We introduce an extension of the D-V algorithm to three-dimensional space for transfer entropy estimation. After ordinal sampling, the D-V algorithm recursively partitions the three-dimensional space defined by v_i , v_{i-1} , and $u_{i-\tau}$ into cubes of varying sizes. Initially, the entire space is sliced into 8 equal cubes, where the boundaries are at the mid-points in the three dimensions. Using the 8 cubes, the following χ^2 statistic is

computed to test the null hypothesis that data points are evenly distributed across the 8 cubes:

$$s_{\chi^2} = \sum_{i=1}^8 (M_i - \mu_M)^2, \quad (10)$$

where M_1, M_2, \dots, M_8 are the numbers of data points contained in the 8 cubes and μ_M is the total number of data points divided by 8 (number of cubes). If $s_{\chi^2} > \chi_{95\%}^2$ (7) (at a 5% significance level and 7 degrees of freedom), then the null hypothesis is rejected and each of the 8 cubes is further partitioned into 8 smaller cubes in the same manner, and the recursion continues. If the null hypothesis is not rejected, the partitions in the current iteration are discarded and the current 8 cubes under scrutiny as a whole are taken as one partition. Cubes that contain no data point do not contribute to the transfer entropy estimation.

Appendix II shows a two-dimensional analogy of a partitioned space. It was generated based on 1,000 data points randomly sampled from a bivariate Gaussian distribution with covariance $\sigma_{xy} = 0.9$ and unit variance, $\sigma_x^2 = 1$ and $\sigma_y^2 = 1$. Note that many squares with smaller sizes are allocated to densely populated parts of the space, whereas fewer partitions are created in sparse areas.

The result of the partitioning process is a finite number of cubes, L , each of which contains nonzero data points. By approximating each probability in (5) by counting, $T_{U \rightarrow V}(\tau)$ can now be estimated using the L partitions as follows:

$$T_{U \rightarrow V}(\tau) \approx \sum_{k=1}^L \frac{n_k}{P} \log \frac{n_k n_k^{v_i-1}}{n_k^{v_{i-1}, u_{i-\tau}} n_k^{v_i, v_{i-1}}}, \quad (11)$$

where P is the total number of data triplets (v_i, v_{i-1} , and $u_{i-\tau}$), n_k is the number of data points in the k th partition, and $n_k^{v_i-1}$, $n_k^{v_i, v_{i-1}}$, and $n_k^{u_{i-\tau}, v_{i-1}}$ are the numbers of data points in the entire data that are greater than or equal to lower bounds and less than upper bounds of the k th partition with respect to the dimensions in the superscript. For example, if the k th partition's lower and upper limits in v_{i-1} are 5 and 10, then $n_k^{v_i-1}$ is the total number of data points such that $5 \leq v_{i-1} < 10$, regardless of their v_i and $u_{i-\tau}$ coordinates.

Our transfer entropy estimation method with D-V partitioning has been implemented in MATLAB, and we made the code and an example data set (from the simulated experiment described in the next section) publicly available at PhysioNet [35] (<http://physionet.org>).

Experiments

Simulation

The source time series, $X = \{x_1, x_2, \dots, x_{N+2}\}$ and the corresponding sink time series, $Y = \{y_1, y_2, \dots, y_{N+2}\}$, were generated as follows:

$$\begin{aligned} x_i &= S_{x,i} + v_{x,i} \\ y_i &= S_{y,i} + v_{y,i}, \end{aligned}$$

where the signal ($s_{x,i}$ and $s_{y,i}$) and noise components ($v_{x,i}$ and $v_{y,i}$) are

$$\begin{aligned}
 s_{x,i} &\sim \mathcal{N}(10, 1) \\
 s_{y,i} &= (a \cdot s_{x,i-2})^2 \\
 v_{x,i} &\sim \mathcal{L}(0, b_x) \\
 v_{y,i} &\sim \mathcal{L}(0, b_y),
 \end{aligned} \tag{12}$$

where \sim denotes sampling from a distribution, a is a coupling constant, and \mathcal{N} and \mathcal{L} represent the Gaussian (mean, variance) and Laplace (mean, scale) distributions, respectively. The Laplace distribution is an example of a heavy-tailed statistic and thus can better model the type of noise often encountered in physiological signals compared with Gaussian noise (e.g., occasional sighs and apneas in a ventilation time series or muscle artifacts in electrocardiography (ECG) recordings [36]). Equation (12) imposed a nonlinear relationship between X and Y (via the squaring) at a time lag of two. Sample size, N , was varied from 50 to 200 at a step of 50, representing the focus of this experiment on small sample size. Note that the lengths of X and Y were $N + 2$ due to the lag of 2; x_{N+1} , x_{N+2} , y_1 , and y_2 were discarded during transfer entropy estimation.

The coupling strength between X and Y was varied by increasing the signal-to-noise ratio (SNR) in both time series from 10 to 20 dB in steps of 1 dB. This increase in SNR was induced by increasing the coupling constant a . In order to maintain the same SNR in both X and Y , the noise power in X was gradually reduced by decreasing b_x while holding the signal power constant, whereas the signal power in Y was gradually increased by increasing a accordingly while holding the noise power (i.e., b_y) constant.

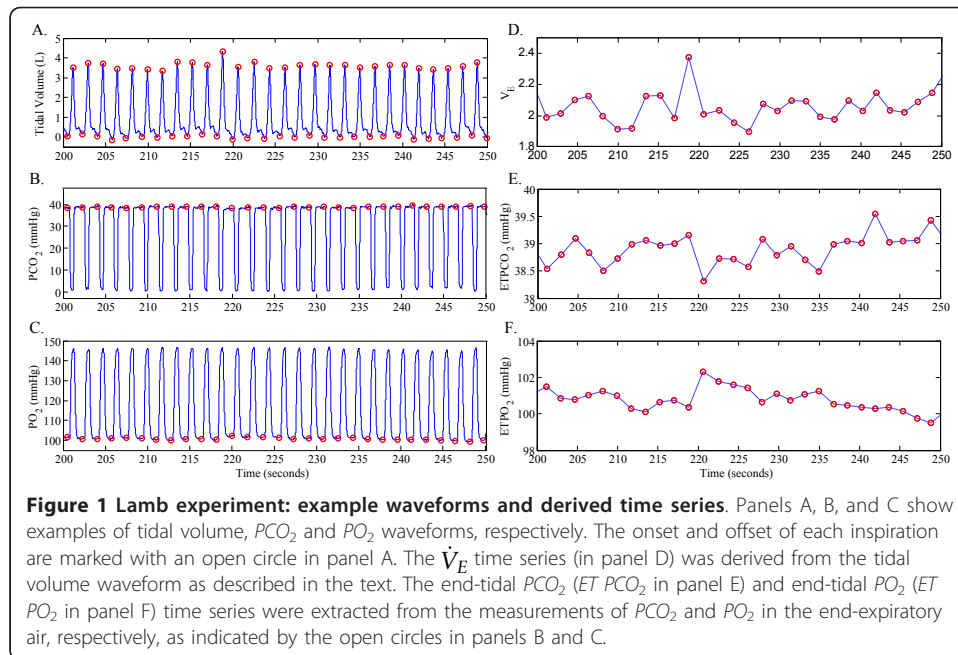
As the first step, we analyzed transfer entropy vs. time lag (τ , from 0 to 5 in steps of 1) with SNR = 20 dB and $N = 200$ using all three estimation methods. This analysis served not only as a comparison among the methods but also as a parameter selection step for the fixed-bin (Q was varied from 4 to 10 in steps of 2) and KDE (α was varied from 0.5 to 1.5 in steps of 0.25) methods.

With the selected values for the parameters Q and α , X and Y were randomly generated 100 times at each SNR level and the corresponding transfer entropies were computed using all three methods. Subsequently, one-sided Wilcoxon rank sum tests were conducted to judge whether there was a significant increase in transfer entropy at each step of increase in SNR with $p < 0.05$ considered to be statistically significant.

An Application to Respiratory Physiology

Using the three transfer entropy estimation methods, we quantified the impact of the administration of domperidone, a dopamine D_2 -receptor antagonist that increases carotid body sensitivity to O_2 and CO_2 [21]. We performed our analysis on recordings from 14 newborn lambs obtained at Monash University, Australia [21]. All surgical and experimental procedures conformed to the guidelines of the National Health and Medical Research Council of Australia and had the approval of the Standing Committee in Ethics in Animal Experimentation of Monash University.

Measurements of spontaneous breathing respiratory flow, end-tidal PCO_2 and end-tidal PO_2 were obtained for a period of approximately 10 minutes before and following the intravenous administration of domperidone. Minute ventilation (\dot{V}_E) was calculated for each breath as V_T/T_{tot} , where V_T is tidal volume and T_{tot} is breath duration (see Figure 1 for example waveforms and derived time series). Breath-to-breath time series



data of \dot{V}_E , end-tidal PCO_2 and end-tidal PO_2 were further band-pass filtered to remove any oscillations slower than (i.e., more breaths/cycle than) 20 breaths/cycle or faster than (i.e., fewer breaths/cycle than) 5 breaths/cycle (using a 7th-order Butterworth digital filter with band-pass frequency of 0.05-0.2 cycles/breath). The frequencies included in the resulting band-passed time series are consistent with the range of oscillations often observed during periodic breathing [21,25]. All extracted time series included at least ~ 200 data points. In a few cases where longer observations were available, the time series were divided into non-overlapping windows of size ~ 200 data points and the window with the largest variance was selected for further analysis. Our rationale behind this choice of data segment selection was to characterize the system at its most variable (or unstable) state during spontaneous breathing (see [37] for more information).

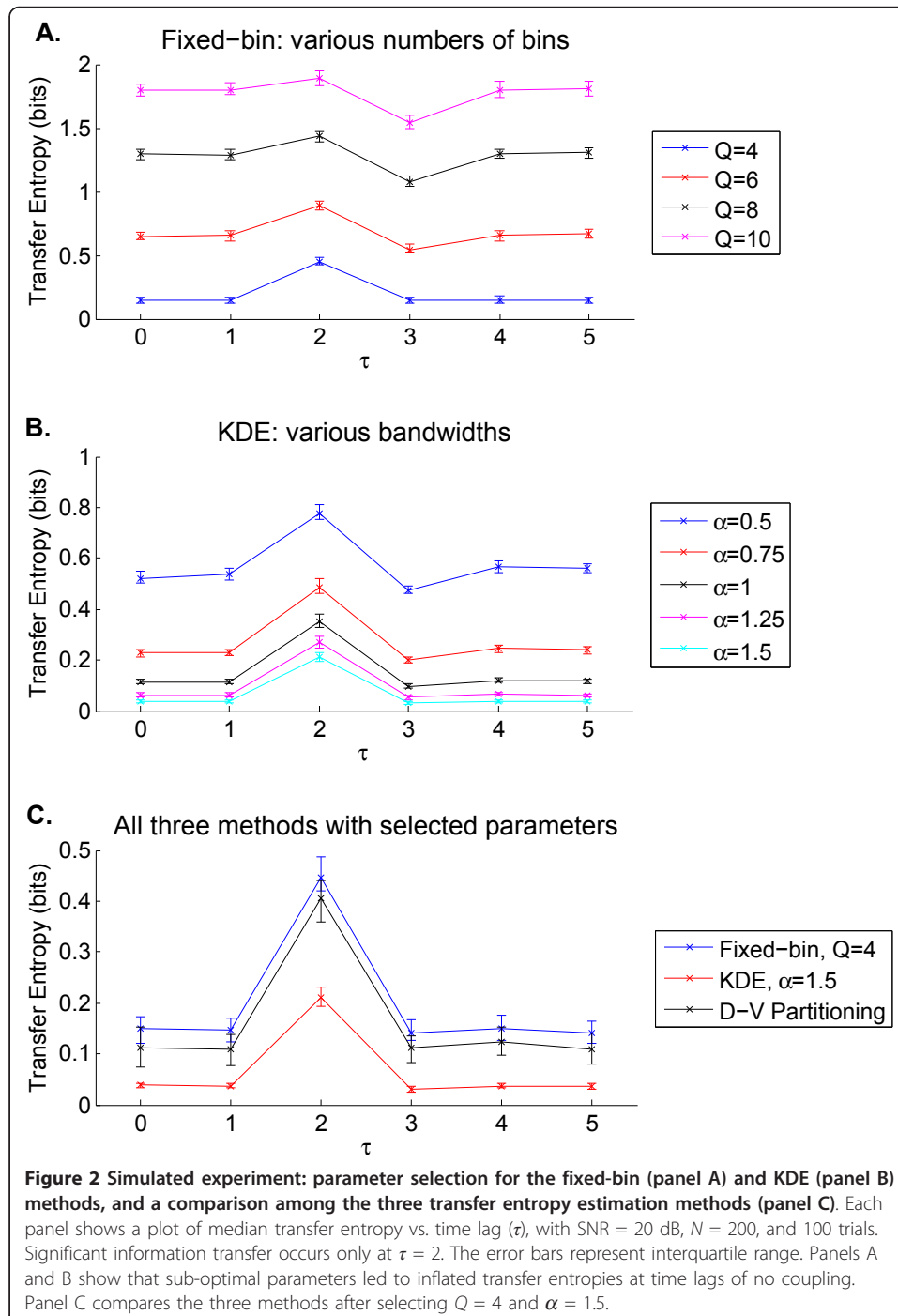
The three transfer entropy estimation methods were applied to the \dot{V}_E , PO_2 , and PCO_2 time series from each lamb. Specifically, the following two directional information flows were investigated before and after the administration of domperidone: $PO_2 \rightarrow \dot{V}_E$ and $PCO_2 \rightarrow \dot{V}_E$. Unlike the simulation experiment, the time lag τ at which a significant coupling exists was unknown a priori. Hence, τ was varied between one and five breaths and the results for the time lag at which the value of transfer entropy was the largest and statistically significant were reported. Significant information flows were determined by Monte Carlo surrogates, i.e., temporally shuffled time series. In this experiment, transfer entropies were computed for 100 surrogates of the source time series (PO_2 or PCO_2), and the transfer entropy from the original source time series was deemed to be significant if it was greater than the 95th percentile of the surrogate results (as described in [38]). If no τ value resulted in a significant information flow, that particular time series pair was removed from further analysis. There was difficulty in selecting optimal Q and α values for the fixed-bin and KDE methods, respectively, due to the lack of a priori knowledge regarding coupling time lag. Therefore, arbitrary, but reasonable, choices of $Q = 5$ and $\alpha = 1$ were made. Finally, increases in

coupling strength caused by domperidone were tested using the Wilcoxon signed rank sum test.

Results

Simulated Experiment

Figure 2 shows how the three transfer entropy estimation methods performed when τ was varied from 0 to 5 with $N = 200$ and SNR = 20 dB in the simulated experiment.



Panels A and B illustrate transfer entropy estimation for different values of Q and α for the fixed-bin and KDE methods, respectively. Since one expects to see near zero transfer entropies at τ values other than 2, $Q = 4$ and $\alpha = 1.5$ were selected for the rest of the experiment. Panel C of Figure 2 compares the three methods using the selected parameter values. All three methods allowed one to visually distinguish $\tau = 2$ from other time lags. KDE was better at estimating low transfer entropies at $\tau \neq 2$ at the cost of a reduced transfer entropy at $\tau = 2$. Fixed-binning and D-V partitioning performed comparably, although D-V partitioning yielded lower transfer entropies at $\tau \neq 2$.

The main results of the simulated experiment are shown in Figure 3 which illustrates the estimated transfer entropies from the 100 trials as a function of SNR. Following from Figure 2, $Q = 4$ and $\alpha = 1.5$ for the fixed-bin and KDE methods, respectively. Overall, the fixed-bin method estimated higher transfer entropies at all lags than the

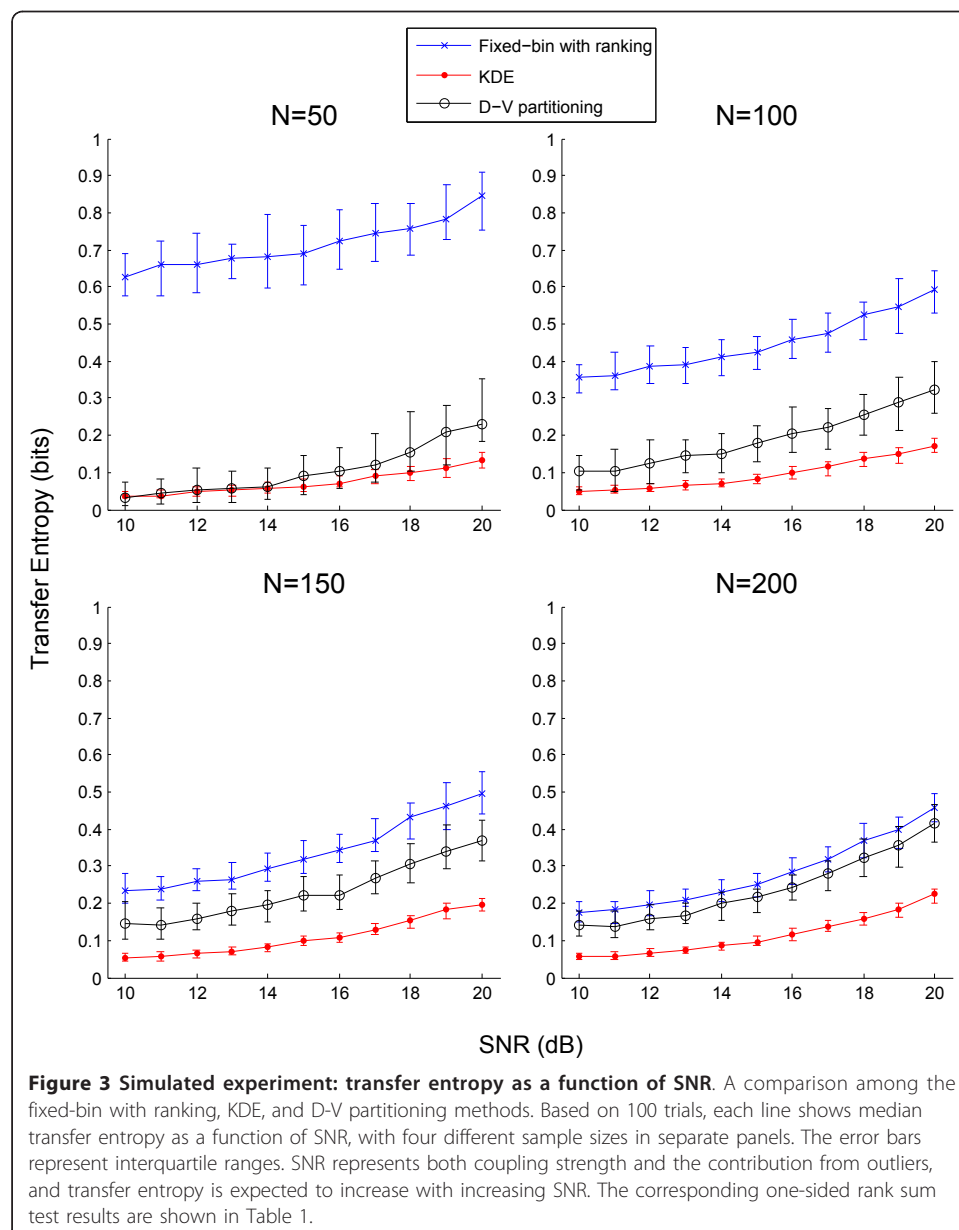


Table 1 Simulated experiment: minimum SNR associated with sustained detection of statistically significant increases in transfer entropy

<i>N</i>	Fixed-bin	KDE	D-V partitioning
50	18 → 19 dB	15 → 16 dB	17 → 18 dB
100	15 → 16 dB	11 → 12 dB	17 → 18 dB
150	13 → 14 dB	11 → 12 dB	16 → 17 dB
200	11 → 12 dB	11 → 12 dB	13 → 14 dB

The minimum SNR at and beyond which statistically significant increases in transfer entropy are sustained according to the one-sided rank sum test. The two compared SNR levels are tabulated. See Figure 3 for visualization.

other two methods, which was improved as *N* increased. KDE showed the smallest dynamic range in transfer entropy as SNR increased but was also associated with the smallest estimation variance at any given SNR. In fact, estimation variance with KDE was similar across different sample sizes, whereas the other two methods resulted in improved estimation variance as *N* increased.

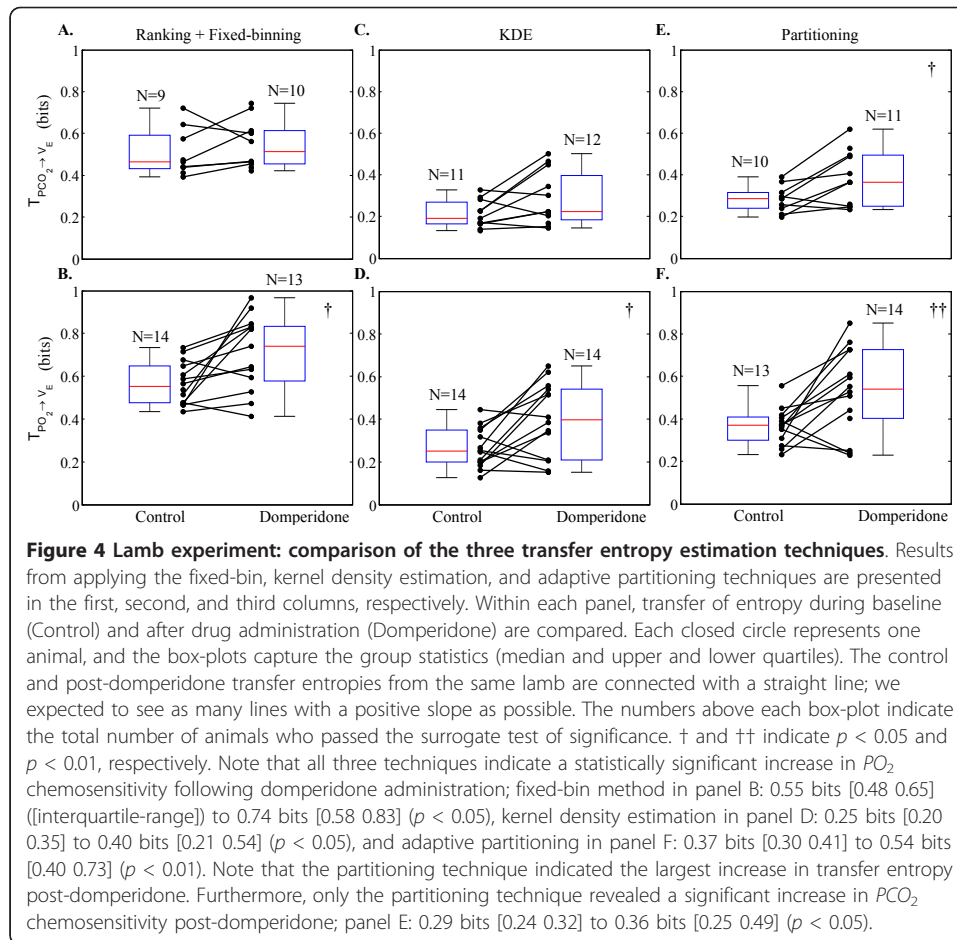
Table 1 tabulates the one-sided rank sum test results of Figure 3 for detection of statistically significant increases in coupling strength. The two compared SNR levels are shown and statistically significant increases in transfer entropy were found to be sustained at and beyond the indicated SNR levels. In Table 1 all methods show the trend of detecting increased coupling at a lower SNR as *N* increases. Overall, KDE outperformed the other two methods in sensitivity to increasing coupling strength; for $N \geq 100$, KDE was able to distinguish 11 dB ($\alpha = 1.059$) and 12 dB ($\alpha = 1.122$). Although D-V partitioning was able to detect increases in transfer entropy at a lower SNR than the fixed-bin method when $N = 50$, the fixed-bin method slightly outperformed D-V partitioning at other *N* values.

Respiratory Physiology Experiment

Figure 4 illustrates how the estimated transfer entropy values changed for $PCO_2 \rightarrow \dot{V}_E$ and $PO_2 \rightarrow \dot{V}_E$ (first and second row, respectively), before ("Control") and after ("Domperidone") the administration of domperidone. The first, second, and third columns correspond to the fixed-bin, KDE, and D-V partitioning methods, respectively. Each closed circle represents one lamb, and the pre- and post-domperidone transfer entropies from the same lamb are connected with a straight line. Hence, we expect to see as many lines with a positive slope as possible. The number of subjects (*N*) varied among the different estimation methods as well as between pre- and post- domperidone since some cases did not pass the surrogate significance test. The group statistics are shown by the box plots.

Figure 4 shows that all three techniques were able to detect a statistically significant increase in transfer entropy for $PO_2 \rightarrow \dot{V}_E$ (interpreted as increased chemosensitivity) post-domperidone. However, the D-V partitioning technique indicated the statistically strongest increase in transfer entropy following domperidone administration ($p < 0.01$). Moreover, only D-V partitioning detected a significant increase in the $PCO_2 \rightarrow \dot{V}_E$ chemosensitivity post-domperidone ($p < 0.05$).

There are three other notable findings in Figure 4. First, fixed-binning led to higher transfer entropy values for both control and domperidone cases than the other two techniques. Second, unlike the simulation results, KDE did not show smaller estimation variance than the other two methods. Third, domperidone administration increased the range in transfer entropy in most panels in Figure 4.



Discussion

In this paper, D-V partitioning was extended to three dimensions and, along with ranking, was used in the framework of transfer entropy estimation. Also, a comparison among three transfer entropy estimation techniques (namely traditional fixed-binning with ranking, state-of-the-art KDE, and adaptive partitioning) was conducted via a simulated experiment and an application to respiratory physiology. The focus of both studies was to investigate the methods' ability to detect increases in directional coupling strength under small sample size and outliers which are commonly encountered in biomedical studies. Although KDE detected increases in coupling strength at the lowest SNR level across different sample sizes in the simulation, D-V partitioning outperformed the other two methods in detecting increased chemosensitivity caused by domperidone in the lamb experiment. The respiratory chemoreflex system was a suitable example of non-linearity and non-stationarity. Several authors have reported evidence for non-linearity in the respiratory chemoreflex system [22,39]. Non-stationarity may also arise as a result of fluctuations in other variables/parameters such as cardiac output, respiratory rate, sleep-state, arousal-related changes in respiratory mechanics and the chemical control system, behavioral factors, or interventions. The application to respiratory physiology in this article demonstrated that the adaptive-partitioning-based transfer entropy estimation technique was more sensitive to increases in

coupling strength for both the PCO_2 and PO_2 chemosensitivity than the other two methods. These results, which are based on non-invasive measurements of spontaneous breathing ventilatory variables, are in agreement with the experimental findings of Edwards and colleagues [21] based on the traditional hypercapnic and hypoxic responses (i.e., ratio of change in ventilation in response to an incremental increase in PCO_2 or decrease in PO_2 in the inspired air). Therefore, transfer entropy has the potential to be a useful surrogate marker of chemosensitivity, and may provide physiologists and physicians with a non-invasive tool to track patient response to pharmacological interventions (e.g., Acetazolamide).

Strictly speaking, transfer entropy measures the coupling strength of a causal link between two time series at a specific time lag and thus makes it difficult to attribute any observed change in transfer entropy to changes in chemosensitivity. For example, Porta et al. have also shown that cross-conditional entropy can be high with a relatively low level of baroreflex sensitivity [11]. Thus, any interchangeability in the current study between transfer entropy and chemosensitivity requires further investigation, since it is possible to increase transfer entropy by, for example, improving the SNR in the data acquisition system under the same level of chemosensitivity or changing the operating point of the feedback system in the direction of an improved SNR. The simulation corroborated the lamb experiment by enabling a controlled comparative analysis. For example, unlike the lamb experiment, the simulated experiment allowed us to control the amount of noise and forced the signals to be stationary throughout. Furthermore, the square function and Laplace noise simulated a non-linear coupling relationship and real-life noise, respectively, that are easy to interpret and control.

Both fixed-binning and D-V partitioning incorporated ordinal sampling, or ranking, which should have at least partially dealt with outliers and multi-modal PDFs. The only difference between the two methods is in how the three-dimensional space is partitioned, and here the efficient bin allocation of the D-V partitioning algorithm becomes a substantial advantage. Under non-uniform distributions that exhibit multiple peaks or varying concentrations of data points, fixed-binning is destined to function inefficiently by allocating fixed-size bins from the minimum to maximum observed values. On the other hand, the D-V partitioning adaptively adjusts bin size according to how evenly data are distributed in a given sub-region of the data space. The expected effectiveness of adaptive partitioning increases with greater deviation from a uniform distribution, and this advantage could partially explain the improved performance of the D-V partitioning method in the lamb experiment in comparison with the simulation.

Moreover, there are minimal adjustable parameters in the D-V partitioning algorithm (one can choose to change the significance level of the χ^2 test), whereas the fixed-bin and KDE methods require one to choose the number of bins and the bandwidth of the kernel a priori. In fact, KDE allows one more degree of freedom: selection of the kernel function, which was not analyzed in this study. Parameter selection can be challenging especially with real-life data where the true coupling characteristics (e.g., strength, direction, time lag, etc.) are often unknown. With the D-V partitioning technique, one can bypass the often arbitrary step of model parameter selection, and this can be a substantial advantage in practice.

The block lengths k and l as well as the time lag t were fixed at 1 in the current study similar to previous studies quoting computational reasons, despite the known limitation that transfer entropy is expected to be underestimated [1,38]. In cases where data size is sufficiently large and computational tractability is not a concern, one can attempt to optimize k , l , and t . For instance, Faes et al. discussed a greedy approach for conditional entropy that selects an optimal conditional pattern consisting of past data from two or more related signals [12]. Porta et al. studied the effects of conditional pattern length on cross-conditional entropy [40]. With respect to computational time, the three different methods are ranked in the following ascending order: fixed-bin, D-V partitioning, and KDE. Although KDE performed very well in the simulated experiment, the necessity to compute the distance between a given point and every point in the data set along each dimension makes it computationally intensive. However, it is worthwhile noting that computational time is not a major limiting factor in most research studies that only require retrospective analysis.

In a number of biomedical studies, small sample size and presence of outliers are often closely associated with physiological dynamics. A common focus of many research studies is to investigate how physiological systems behave under different conditions or respond to external interventions, naturally formulating a non-stationary problem. One way to analyze a non-stationary time series is to segment them into quasi-stationary intervals. This limits the number of data points in each segment and leads to the small sample size problem. This can be a major challenge to many parameter estimation algorithms in presence of noise, because they often require sufficiently long observations. In addition, physiological systems tend to exhibit non-linear behaviors which may cause the state of the system to abruptly change from its steady-state baseline, producing outliers. These outliers, originating from underlying physiological phenomena rather than noise, convey meaningful value and hence should not be removed for mere computational convenience. We remind the reader that our simulated experiment explicitly investigated the issues of small sample size and of outliers.

Conclusions

In this article, we have shown that transfer entropy can detect changes in directional coupling between two biomedical time series. We have extended D-V partitioning to transfer entropy estimation and compared the performance of three transfer entropy estimation methods in detection of increased coupling strength: fixed-binning with ranking, KDE, and D-V partitioning with ranking. Based on the results of this study, fixed-binning, even with ranking, failed to clearly outperform the other two methods, while the comparison between KDE and D-V partitioning was inconclusive. However, D-V partitioning may be the most attractive option after taking into account computational time and the difficulty associated with parameter selection. We hope that this study provides a helpful guideline in selecting an appropriate transfer entropy estimation method.

Appendix I

Given two concurrently sampled time series $X = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_1, y_2, \dots, y_N\}$, the conditional entropy $H(x_i | y_i)$ is defined as:

$$H(x_i | y_i) = \sum_{x_i, y_i} p(x_i, y_i) \log \frac{p(y_i)}{p(x_i, y_i)}.$$

Note that $H(x_i | y_i) \geq 0$. Then, the transfer entropy from X to Y , termed $T_{X \rightarrow Y}$, can be derived from conditional entropies as follows:

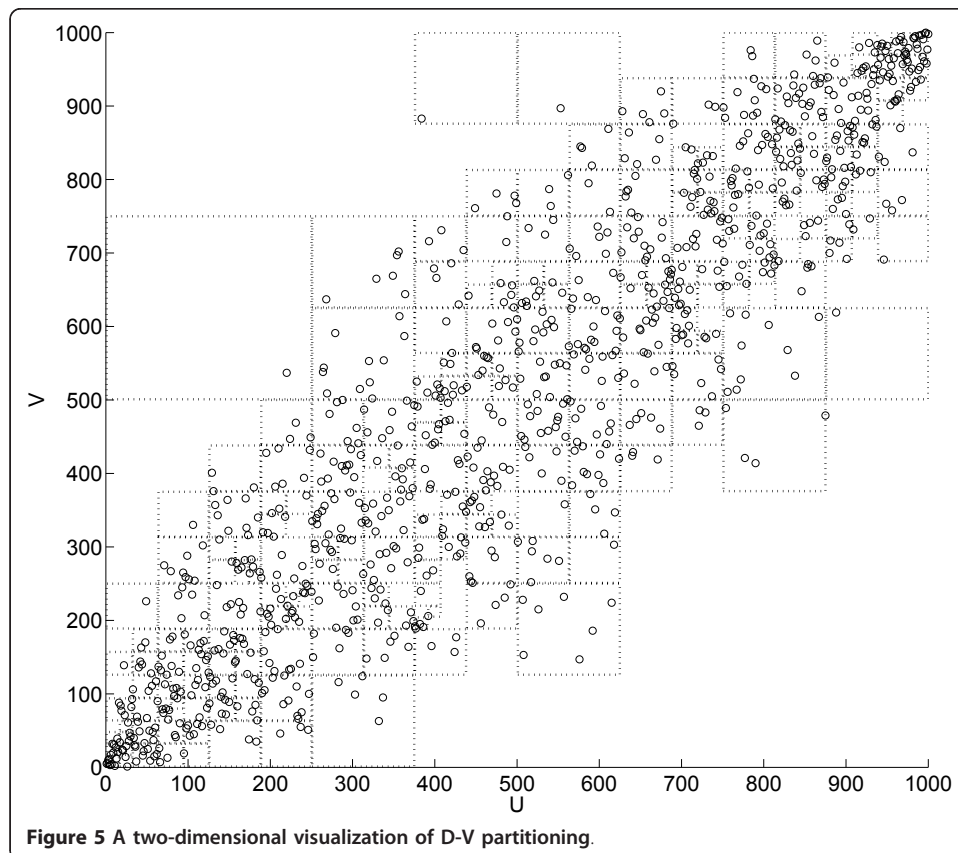
$$\begin{aligned}
 T_{X \rightarrow Y} &= H(y_i | \gamma_{i-t}^{(l)}) - H(y_i | \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}) \\
 &= \sum_{\gamma_i, \gamma_{i-t}^{(l)}} p(\gamma_i, \gamma_{i-t}^{(l)}) \log \frac{p(\gamma_{i-t}^{(l)})}{p(\gamma_i, \gamma_{i-t}^{(l)})} - \sum_{\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}} p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}) \log \frac{p(x_{i-t}^{(k)}, \gamma_{i-t}^{(l)})}{p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)})}. \quad (13)
 \end{aligned}$$

But since $p(\gamma_i, \gamma_{i-t}^{(l)}) = \sum_{x_{i-t}^{(k)}} p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)})$, the two arguments in the separate summations can be combined. Thus,

$$\begin{aligned}
 T_{X \rightarrow Y} &= \sum_{\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}} p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}) \log \frac{p(\gamma_{i-t}^{(l)}) p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)})}{p(\gamma_i, \gamma_{i-t}^{(l)}) p(\gamma_{i-t}^{(l)}, x_{i-t}^{(k)})} \\
 &= \sum_{\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}} p(\gamma_i, \gamma_{i-t}^{(l)}, x_{i-t}^{(k)}) \log \frac{p(\gamma_i | \gamma_{i-t}^{(l)}, x_{i-t}^{(k)})}{p(\gamma_i | \gamma_{i-t}^{(l)})}.
 \end{aligned}$$

Appendix II

Figure 5 shows a two-dimensional visualization of D-V partitioning. The space contains 1,000 data points sampled from a bivariate Gaussian distribution with $\sigma_{xy} = 0.9$, $\sigma_x^2 = 1$, and $\sigma_y^2 = 1$. This figure shows the transformed data space after ordinal



sampling. The proposed transfer entropy estimation with the D-V partitioning algorithm extends this kind of partitioning to three dimensions.

Acknowledgements

This research work was funded by the National Institutes of Health (NIH) (through grant numbers R01-EB001659, R01-HL73146, HL085188-01A2, HL090897-01A2, K24 HL093218-01A1, cooperative agreement U01-EB-008577, and training grant T32-HL07901) and American Heart Association (AHA) (grant 0840159N). Dr. Lee holds a Postdoctoral Fellowship funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. Edwards is the recipient of the Thoracic Society of Australia and New Zealand/Allen and Hanburys Respiratory Research Fellowship. The content of this document is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, AHA, or NSERC. The authors would also like to thank Professor George Verghese for his helpful comments.

Author details

¹Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA. ²Massachusetts Institute of Technology, Cambridge, MA, USA. ³Harvard Medical School, Division of Sleep Medicine, Brigham and Women's Hospital, Boston, MA, USA.

Authors' contributions

JL and SN conceived of the study, developed and implemented the D-V partitioning transfer entropy estimation algorithm, conducted the experiments, and wrote the entire manuscript. IS conceived of the study, contributed to the algorithm development, and wrote parts of the manuscript. BE supplied the lamb data and rigorously revised the manuscript. JB and AM thoroughly revised the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 30 November 2011 Accepted: 13 April 2012 Published: 13 April 2012

References

- Schreiber T: **Measuring information transfer.** *Physical Review Letters* 2000, **85**(2):461-464.
- Hlavackova-Schindler K, Palus M, Vejmelka M, Bhattacharya J: **Causality detection based on information-theoretic approaches in time series analysis.** *Physics Reports* 2007, **441**:1-46.
- Barnett L, Barrett AB, Seth AK: **Granger causality and transfer entropy are equivalent for Gaussian variables.** *Physical Review Letters* 2009, **103**(23):238701.
- Rubinov M, Sporns O: **Complex network measures of brain connectivity: uses and interpretations.** *NeuroImage* 2010, **52**(3):1059-1069, [http://www.ncbi.nlm.nih.gov/pubmed/19819337].
- Alkire MT, Hudetz AG, Tononi G: **Consciousness and anesthesia.** *Science* 2008, **322**:876-880, [http://dx.doi.org/10.1126/science.1149213].
- Vakorin VA, Kovacevic N, McIntosh AR: **Exploring transient transfer entropy based on a group-wise ICA decomposition of EEG data.** *NeuroImage* 2010, **49**(2):1593-1600, [http://www.ncbi.nlm.nih.gov/pubmed/19698792].
- Gourevitch B, Eggermont JJ: **Evaluating information transfer between auditory cortical neurons.** *Journal of Neurophysiology* 2007, **97**(3):2533-2543, [http://www.ncbi.nlm.nih.gov/pubmed/17202243].
- Hinrichs H, Noesselt T, Heinze HJ: **Directed information flow: a model free measure to analyze causal interactions in event related EEG-MEG-experiments.** *Human Brain Mapping* 2008, **29**(2):193-206, [http://www.ncbi.nlm.nih.gov/pubmed/17390316].
- Vicente R, Wibral M, Lindner M, Pipa G: **Transfer entropy—a model-free measure of effective connectivity for the neurosciences.** *Journal of Computational Neuroscience* 2011, **30**:45-67, [http://www.ncbi.nlm.nih.gov/pubmed/20706781].
- Katura T, Tanaka N, Obata A, Sato H, Maki A: **Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics.** *NeuroImage* 2006, **31**(4):1592-1600, [http://www.ncbi.nlm.nih.gov/pubmed/16549367].
- Porta A, Catai AM, Takahashi ACM, Magagnin V, Bassani T, Tobaldini E, van de Borne P, Montano N: **Causal relationships between heart period and systolic arterial pressure during graded head-up tilt.** *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* 2011, **300**:R378-R386.
- Faes L, Nollo G, Porta A: **Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique.** *Physical Review E* 2011, **83**:051112.
- Vejmelka M, Palus M: **Inferring the directionality of coupling with conditional mutual information.** *Physical Review E* 2008, **77**:026214.
- Palus M, Stefanovska A: **Direction of coupling from phases of interacting oscillators: an information-theoretic approach.** *Physical Review E* 2003, **67**:055201.
- Palus M, Vejmelka M: **Directionality of coupling from bivariate time series: how to avoid false causalities and missed connections.** *Physical Review E* 2007, **75**:056211.
- Kaiser A, Schreiber T: **Information transfer in continuous processes.** *Physica D* 2002, **166**:43-62.
- Khan S, Bandyopadhyay S, Ganguly AR, Saigal S, Erickson DJ, Protopoulos V, Ostrouchov G: **Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data.** *Physical Review E* 2007, **76**:026209.
- Darbellay AG, Vajda I: **Estimation of the information by an adaptive partitioning of the observation space.** *IEEE Transactions on Information Theory* 1999, **45**(4):1315-1321.
- Fraser AM, Swinney HL: **Independent coordinates for strange attractors from mutual information.** *Physical Review A* 1986, **33**(2):1134-1140.

20. Hudson JE: **Signal processing using mutual information.** *IEEE Signal Processing Magazine* 2006, **23**(6):50-53.
21. Edwards B, Sands S, Skuza E, Stockx E, Brodecky V, Wilkinson M, Berger P: **Increased peripheral chemosensitivity via dopaminergic manipulation promotes respiratory instability in lambs.** *Respiratory physiology & neurobiology* 2008, **164**(3):419-428.
22. Bruce EN: **Temporal variations in the pattern of breathing.** *J Appl Physiol* 1996, **80**(4):1079-1087.
23. Xie A, Skatrud JB, Puleo DS, Rahko PS, Dempsey JA: **Apnea-hypopnea threshold for CO₂ in patients with congestive heart failure.** *Am J Respir Crit Care Med* 2002, **165**(9):1245-1250.
24. Bradley TD, Floras JS: **Sleep apnea and heart failure: Part I: obstructive sleep apnea.** *Circulation* 2003, **107**(12):1671-1678, [http://dx.doi.org/10.1161/01.CIR.0000061757.12581.15].
25. Khoo MC, Kronauer RE, Strohl KP, Slutsky AS: **Factors inducing periodic breathing in humans: a general model.** *J Appl Physiol* 1982, **53**(3):644-659.
26. Younes M, Ostrowski M, Atkar R, Laprairie J, Siemens A, Hanly P: **Mechanisms of breathing instability in patients with obstructive sleep apnea.** *J Appl Physiol* 2007, **103**(6):1929-1941, [http://dx.doi.org/10.1152/jappphysiol.00561.2007].
27. Wellman A, Malhotra A, Jordan AS, Stevenson KE, Gautam S, White DP: **Effect of oxygen in obstructive sleep apnea: role of loop gain.** *Respir Physiol Neurobiol* 2008, **162**(2):144-151, [http://dx.doi.org/10.1016/j.resp.2008.05.019].
28. Lahiri S, Maret K, Sherpa MG: **Dependence of high altitude sleep apnea on ventilatory sensitivity to hypoxia.** *Respir Physiol* 1983, **52**(3):281-301.
29. Solin P, Roebuck T, Johns DP, Walters EH, Naughton MT: **Peripheral and central ventilatory responses in central sleep apnea with and without congestive heart failure.** *Am J Respir Crit Care Med* 2000, **162**(6):2194-2200.
30. Francis DP, Willson K, Davies LC, Coats AJ, Piepoli M: **Quantitative general theory for periodic breathing in chronic heart failure and its clinical implications.** *Circulation* 2000, **102**(18):2214-2221.
31. Topor ZL, Johannson L, Kasprzyk J, Remmers JE: **Dynamic ventilatory response to CO₂ in congestive heart failure patients with and without central sleep apnea.** *J Appl Physiol* 2001, **91**:408-416.
32. Ponikowski P, Chua TP, Anker SD, Francis DP, Doehner W, Banasiak W, Poole-Wilson PA, Piepoli MF, Coats AJ: **Peripheral chemoreceptor hypersensitivity: an ominous sign in patients with chronic heart failure.** *Circulation* 2001, **104**(5):544-549.
33. Lanfranchi PA, Bagnhiroli A, Bosimini E, Mazzuero G, Colombo R, Donner CF, Giannuzzi P: **Prognostic value of nocturnal Cheyne-Stokes respiration in chronic heart failure.** *Circulation* 1999, **99**(11):1435-1440.
34. Silverman BW: *Circulation, Volume 26 of Monographs on Statistics and Applied Probability.* Chapman and Hall.
35. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE: **PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.** *Circulation* 2000, **101**(23):E215-E220.
36. Martinez J, Olmos S: **A robust T wave alternans detector based on the GLRT for Laplacian noise distribution.** *Computers in Cardiology* 2002, **2002**:677-680.
37. Nemati S, Edwards BA, Sands SA, Berger PJ, Wellman A, Verghese GC, Malhotra A, Butler JP: **Model-based characterization of ventilatory stability using spontaneous breathing.** *Journal of Applied Physiology* 2011, **111**:55-67, [http://jap.physiology.org/content/111/1/55.abstract].
38. Ruddell BL, Kumar P: **Ecohydrologic process networks: 1. identification.** *Water Resources Research* 2009, **45**:W03419.
39. Fiamma MN, Straus C, Thibault S, Wysocki M, Baconnier P, Similowski T: **Effects of hypercapnia and hypocapnia on ventilatory variability and the chaotic dynamics of ventilatory flow in humans.** *Am J Physiol Regul Integr Comp Physiol* 2007, **292**(5):R1985-R1993.
40. Porta A, Baselli G, Lombardi F, Montano N, Malliani A, Cerutti S: **Conditional entropy approach for the evaluation of the coupling strength.** *Biological Cybernetics* 1999, **81**:119-129.

doi:10.1186/1475-925X-11-19

Cite this article as: Lee et al.: Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series. *BioMedical Engineering OnLine* 2012 11:19.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

