

Article

Assessing Coupling Dynamics from an Ensemble of Time Series

Germán Gómez-Herrero ¹, Wei Wu ², Kalle Rutanen ³, Miguel C. Soriano ⁴, Gordon Pipa ⁵ and Raul Vicente ^{6,*}

¹ Netherlands Institute for Neuroscience, Meibergdreef 47, Amsterdam 1105 BA, The Netherlands; E-Mail: german.gomezherrero@gmail.com

² Lab of Neurophysics and Neurophysiology, Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, 96 JinZhai Rd., Hefei 230026, China; E-Mail: wu@fias.uni-frankfurt.de

³ Department of Mathematics, Tampere University of Technology, Korkeakoulunkatu 10, Tampere FI-33720, Finland; E-Mail: kalle.rutanen@tut.fi

⁴ Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), Campus Universitat de les Illes Balears E-07122 Palma de Mallorca, Spain; E-Mail: miguel@ifisc.uib-csic.es

⁵ Institut für Kognitionswissenschaft, University of Osnabrück, Albrechtstrasse 28, Osnabrück 49076, Germany; E-Mail: gpipa@uos.de

⁶ Institute of Computer Science, University of Tartu, J. Liivi 2, Tartu 50409, Estonia

* Author to whom correspondence should be addressed; E-Mail: raulvicente@gmail.com; Tel.: +372-737-5445; Fax: +372-737-5468.

Academic Editor: Deniz Gencaga

Received: 30 November 2014 / Accepted: 19 March 2015 / Published: 2 April 2015

Abstract: Finding interdependency relations between time series provides valuable knowledge about the processes that generated the signals. Information theory sets a natural framework for important classes of statistical dependencies. However, a reliable estimation from information-theoretic functionals is hampered when the dependency to be assessed is brief or evolves in time. Here, we show that these limitations can be partly alleviated when we have access to an ensemble of independent repetitions of the time series. In particular, we gear a data-efficient estimator of probability densities to make use of the full structure of trial-based measures. By doing so, we can obtain time-resolved estimates for a family of entropy combinations (including mutual information, transfer entropy and their conditional counterparts), which are more accurate than the simple average of individual estimates over trials. We show with simulated and real data generated by coupled electronic circuits that

the proposed approach allows one to recover the time-resolved dynamics of the coupling between different subsystems.

Keywords: entropy; transfer entropy; estimator; ensemble; trial; time series

1. Introduction

An important problem is that of detecting interdependency relations between simultaneously measured time series. Finding an interdependency is the first step in elucidating how the subsystems underlying the time series interact. Fruitful applications of this approach abound in different fields, including neuroscience [1], ecology [2] or econometrics [3]. In these examples, the discovery of certain statistical interdependency is usually taken as an indicator that some interrelation exists between subsystems, such as different brain regions [4], animal populations or economical indexes.

Classical measures to unveil an interdependency include linear techniques, such as cross-correlation, coherence or Granger causality [6]. These measures quantify the strength of different linear relations and, thus, belong to the larger class of parametric measures, which assume a specific form for the interdependency between two or more processes. In particular, parametric techniques are often data-efficient, generalizable to multivariate settings and easy to interpret.

In general, statistical relationships between processes are more naturally and generally formulated within the probabilistic framework, which relaxes the need to assume explicit models on how variables relate to each other. For this reason, when a model of the underlying dynamics and of the assumed interaction is not available, a sound non-parametric approach can be stated in terms of information theory [7]. For example, mutual information is widely used to quantify the information statically shared between two random variables. Growing interest in interdependency measures that capture information flow rather than information sharing lead to the definition of transfer entropy [9]. In particular, transfer entropy quantifies how much the present and past of a random variable condition the future transitions of another. Thus, transfer entropy embodies an operational principle of causality first championed by Norbert Wiener [8], which was explicitly formulated for linear models by Clive Granger [5]. However, it is important to note that transfer entropy should not be understood as a quantifier of interaction strength nor interventional causality. See [10–12] for a detailed discussion on the relation between transfer entropy and different notions of causality and information transfer. See also [13] for a detailed account of how spatially- and temporally-local versions of information theoretic functionals, including transfer entropy, can be used to study the dynamics of computation in complex systems.

A practical pitfall is that without simplifying assumptions, a robust estimation of information theoretic functionals might require a large number of data samples. This requisite directly confronts situations in which the dependency to be analyzed evolves in time or is subjected to fast transients. When the non-stationarity is only due to a slow change of a parameter, over-embedding techniques can partially solve the problem by capturing the slow dynamics of the parameter as an additional variable [14]. It is also habitual to de-trend the time series or divide them into small windows within which the signals can be considered as approximately stationary. However, the above-mentioned procedures become

unpractical when the relevant interactions change in a fast time scale. This is the common situation in brain responses and other complex systems where external stimuli elicit a rapid functional reorganization of information-processing pathways.

Fortunately, in several disciplines, the experiments leading to the multivariate time series can be systematically repeated. Thus, a typical experimental paradigm might render an ensemble of presumably independent repetitions or trials per experimental condition. In other cases, the processes under study display a natural cyclic variation and, thus, also render an ensemble of almost independent cycles or repetitions. This is often the case of seasonal time series that are common in economical and ecological studies and, more generally, of any cyclo-stationary process.

Here, we show how this multi-trial nature can be efficiently exploited to produce time-resolved estimates for a family of information-theoretic measures that we call entropy combinations. This family includes well-known functionals, such as mutual information, transfer entropy and their conditional counterparts: partial mutual information (PMI) [15,16] and partial transfer entropy (PTE) [17,18]. Heuristically, our approach can be motivated using the ergodic theorem. In other words, the time average of a measure converges to the space or ensemble average for an ergodic process. We can associate the conventional computation of entropies with a time average of log probabilities. Crucially, these should converge to the ensemble averages of the equivalent log probabilities, which we exploit with our (ensemble averaging) approach. In our case, the ensemble is constituted by multiple realizations of repeated trials. We use both simulations and experimental data to demonstrate that the proposed ensemble estimators of entropy combinations are more accurate than simple averaging of individual trial estimates.

2. Entropy Combinations

We consider three simultaneously measured time series generated from stochastic processes X , Y and Z , which can be approximated as stationary Markov processes [19] of finite order. The state space of X can then be reconstructed using the delay embedded vectors $\mathbf{x}(n) = (x(n), \dots, x(n - d_x + 1))$ for $n = 1, \dots, N$, where n is a discrete time index and d_x is the corresponding Markov order. Similarly, we could construct $\mathbf{y}(n)$ and $\mathbf{z}(n)$ for processes Y and Z , respectively. Let $V = (V_1, \dots, V_m)$ denote a random m -dimensional vector and $H(V)$ its Shannon entropy. Then, an entropy combination is defined by:

$$C(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = \sum_{i=1}^p s_i H(V_{\mathcal{L}_i}) - H(V) \quad (1)$$

where $\forall i \in [1, p] : \mathcal{L}_i \subset [1, m]$ and $s_i \in \{-1, 1\}$, such that $\sum_{i=1}^p s_i \chi_{\mathcal{L}_i} = \chi_{[1, m]}$, where $\chi_{\mathcal{S}}$ is the indicator function of a set \mathcal{S} (having the value one for elements in the set \mathcal{S} and zero for elements not in \mathcal{S}).

In particular, MI, TE, PMI and PTE all belong to the class of entropy combinations, since:

$$\begin{aligned} I_{X \leftrightarrow Y} &\equiv -H_{XY} + H_X + H_Y \\ T_{X \leftarrow Y} &\equiv -H_{WXY} + H_{WX} + H_{XY} - H_X \\ I_{X \leftrightarrow Y|Z} &\equiv -H_{XZY} + H_{XZ} + H_{ZY} - H_Z \\ T_{X \leftarrow Y|Z} &\equiv -H_{WXZY} + H_{WXZ} + H_{XZY} - H_{XZ} \end{aligned}$$

where random variable $W \equiv X^+ \equiv x(n+1)$, so that H_{WX} is the differential entropy of $p(x(n+1), \mathbf{x}(n))$. The latter denotes the joint probability of finding X at states $x(n+1), x(n), \dots, x(n-d_x+1)$ during time instants $n+1, n, n-1, \dots, n-d_x+1$. Notice that, due to stationarity, $p(x(n+1), \mathbf{x}(n))$ is invariant under variations of the time index n .

3. Ensemble Estimators for Entropy Combinations

A straightforward approach to the estimation of entropy combinations would be to add separate estimates of each of the multi-dimensional entropies appearing in combination. Popular estimators of differential entropy include plug-in estimators, as well as fixed and adaptive histogram or partition methods. However, other non-parametric techniques, such as kernel and nearest-neighbor estimators, have been shown to be extremely more data efficient [20,21]. An asymptotically unbiased estimator based on nearest-neighbor statistics is due to Kozachenko and Leonenko (KL) [22]. For N realizations $\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[N]$ of a d -dimensional random vector X , the KL estimator takes the form:

$$\hat{H}_X = -\psi(k) + \psi(N) + \log(v_d) + \frac{d}{N} \sum_{i=1}^N \log(\epsilon(i)) \quad (2)$$

where ψ is the digamma function, v_d is the volume of the d -dimensional unit ball and $\epsilon(i)$ is the distance from $\mathbf{x}[i]$ to its k -th nearest neighbor in the set $\{\mathbf{x}[j]\}_{\forall j \neq i}$. The KL estimator is based on the assumption that the density of the distribution of random vectors is constant within an ϵ -ball. The bias of the final entropy estimate depends on the validity of this assumption and, thus, on the values of $\epsilon(n)$. Since the size of the ϵ -balls depends directly on the dimensionality of the random vector, the biases of estimates for the differential entropies in Equation (1) will, in general, not cancel, leading to a poor estimator of the entropy combination. This problem can be partially overcome by noticing that Equation (2) holds for any value of k , so that we do not need to have a fixed k . Therefore, we can vary the value of k in each data point, so that the radius of the corresponding ϵ -balls would be approximately the same for the joint and the marginal spaces. This idea was originally proposed in [23] for estimating mutual information and was used in [16] to estimate PMI, and we generalize it here to the following estimator of entropy combinations:

$$\hat{C}(V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}) = F(k) - \sum_{i=1}^p s_i \langle F(k_i(n)) \rangle_n \quad (3)$$

where $F(k) = \psi(k) - \psi(N)$ and $\langle \dots \rangle_n = \frac{1}{N} \sum_{n=1}^N (\dots)$ denotes averaging with respect to the time index. The term $k_i(n)$ accounts for the number of neighbors of the n -th realization of the marginal vector $V_{\mathcal{L}_i}$ located at a distance strictly less than $\epsilon(n)$, where $\epsilon(n)$ denotes the radius of the ϵ -ball in the joint space. Note that the point itself is included in the counting neighbors in marginal spaces ($k_i(n)$), but not when selecting $\epsilon(n)$ from the k -th nearest neighbor in the full joint space. Furthermore, note that estimator Equation (3) corresponds to extending “Algorithm 1” in [23] to entropy combinations. Extensions to conditional mutual information and conditional transfer entropy using “Algorithm 2” in [23] have been discussed recently [12].

A fundamental limitation of estimator Equation (3) is the assumption that the involved multidimensional distributions are stationary. However, this is hardly the case in many real applications,

and time-adaptation becomes crucial in order to obtain meaningful estimates. A trivial solution is to use the following time-varying estimator of entropy combinations:

$$\hat{C}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = F(k) - \sum_{i=1}^p s_i F(k_i(n)) \quad (4)$$

This naive time-adaptive estimator is not useful in practice, due to its large variance, which stems from the fact that a single data point is used for producing the estimate at each time instant. More importantly, the neighbor searches in the former estimator run across the full time series and, thus, ignore possible non-stationary changes.

However, let us consider the case of an ensemble of r' repeated measurements (trials) from the dynamics of V . Let us also denote by $\{v^{(r)}[n]\}_r$ the measured dynamics for those trials ($r = 1, 2, \dots, r'$). Similarly, we denote by $\{v_i^{(r)}[n]\}_r$ the measured dynamics for the marginal vector $V_{\mathcal{L}_i}$. A straightforward approach for integrating the information from different trials is to average together estimates obtained from individual trials:

$$\hat{C}^{\text{avg}}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = \frac{1}{r'} \sum_{r=1}^{r'} \hat{C}^{(r)}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) \quad (5)$$

where $\hat{C}^{(r)}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n)$ is the estimate obtained from the r -th trial. However, this approach makes poor use of the available data and will typically produce useless estimates, as will be shown in the experimental section of this text.

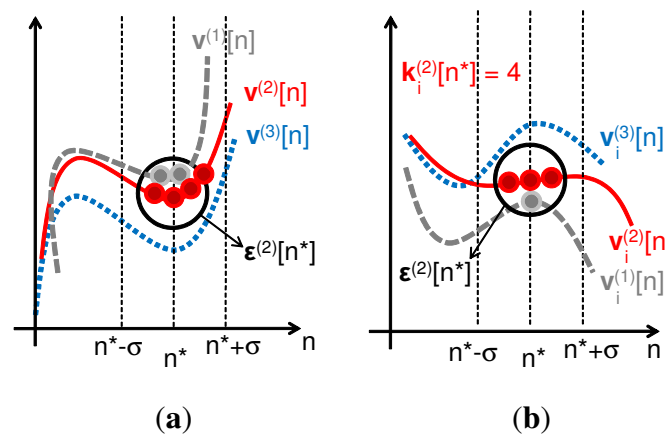


Figure 1. Nearest neighbor statistics across trials. **(a):** For each time instant $n = n^*$ and trial $r = r^*$, we compute the (maximum norm) distance $\epsilon^{(r^*)}(n^*)$ from $v^{(r^*)}[n^*]$ to its k -th nearest neighbor among all trials. Here, the procedure is illustrated for $k = 5$. **(b):** $k_i^{(r^*)}[n^*]$ counts how many neighbors of $v_i^{(r^*)}[n^*]$ are within a radius $\epsilon^{(r^*)}(n^*)$. The point itself (*i.e.*, $v_i^{(r^*)}[n^*]$) is also included in this count. These neighbor counts are obtained for all $i = 1, \dots, p$ marginal trajectories.

A more effective procedure takes into account the multi-trial nature of our data by searching for neighbors across ensemble members, rather than from within each individual trial. This nearest ensemble

neighbors [24] approach is illustrated in Figure 1 and leads to the following ensemble estimator of entropy combinations:

$$\hat{C}^{\text{en}}(\{V_{\mathcal{L}_1}, \dots, V_{\mathcal{L}_p}\}, n) = F(k) - \frac{1}{r'} \sum_{r=1}^{r'} \sum_{i=1}^p s_i F(k_i^{(r)}(n)) \quad (6)$$

where the counts of marginal neighbors $\{k_i^{(r)}(n)\}_{\forall i=1, \dots, p}^{\forall r=1, \dots, r'}$ are computed using overlapping time windows of size 2σ , as shown in Figure 1. For rapidly changing dynamics, small values of σ might be needed to increase the temporal resolution, thus, being able to track more volatile non-stationarities. On the other hand, larger values of σ will lead to lower estimator variance and are useful when non-stationarities develop over slow temporal scales.

4. Tests on Simulated and Experimental Data

To demonstrate that \hat{C}^{en} can be used to characterize dynamic coupling patterns, we apply the ensemble estimator of PTE to multivariate time series from coupled processes.

In particular, we simulated three non-linearly-coupled autoregressive processes with a time-varying coupling factor:

$$\begin{aligned} x^r[n] &= 0.4x^r[n-1] + \eta_x, \\ y^r[n] &= 0.5y^r[n-1] + \kappa_{yx}[n] \sin(x^r[n - \tau_{yx}]) + \eta_y, \\ z^r[n] &= 0.5z^r[n-1] + \kappa_{zy}[n] \sin(y^r[n - \tau_{zy}]) + \eta_z. \end{aligned}$$

during 1,500 time steps and repeated $R = 50$ trials with new initial conditions. The terms η_x , η_y and η_z represent normally-distributed noise processes, which are mutually independent across trials and time instants. The coupling delays amount to $\tau_{yx} = 10$, $\tau_{zy} = 15$, while the dynamics of the coupling follows a sinusoidal variation:

$$\begin{aligned} k_{yx}[n] &= \begin{cases} \sin\left(\frac{2\pi n}{500}\right) & \text{for } 250 \leq n < 750 \\ 0 & \text{otherwise} \end{cases} \\ k_{zy}[n] &= \begin{cases} \cos\left(\frac{2\pi n}{500}\right) & \text{for } 750 \leq n < 1250 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Before PTE estimation, each time series was mapped via a delay embedding to its approximate state space. The dimension of the embedding was set using the Cao criterion [25], while the embedding delay time was set as the autocorrelation decay time. Other criteria to obtain embedding parameters, such as described in [19], provide similar results. Furthermore, each time series was time-delayed, so that they had maximal mutual information with the destination of the flow. That is, before computing some $T_{a \leftarrow b|c}(n)$, the time series b and c were delayed, so that they shared maximum information with the time series a , as suggested in [16]. For a rigorous and formal way to investigate the lag in the information flow between systems, we refer to [26,27].

To assess the statistical significance of the PTE values (at each time instant) we applied a permutation test with surrogate data generated by randomly shuffling trials [28]. Figure 2 shows the time-varying PTEs obtained for these data with the ensemble estimator of entropy combinations given in Equation (6). Indeed, the PTE analysis accurately describes the underlying interaction dynamics. In particular, it

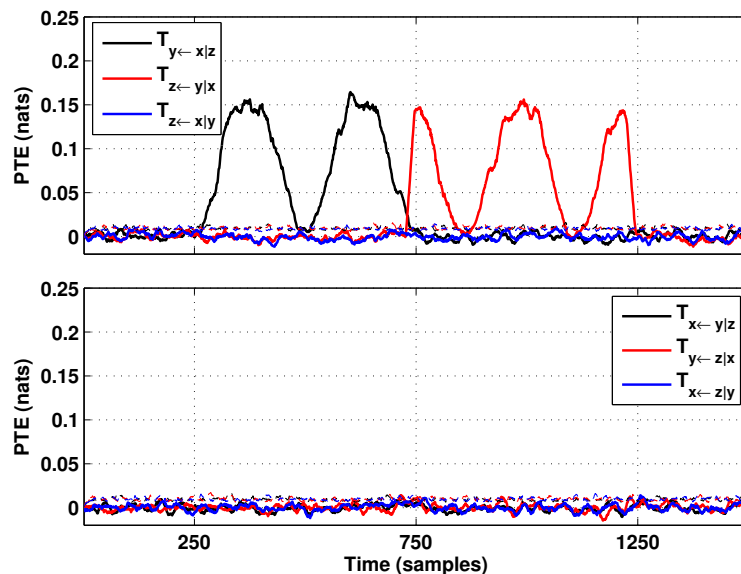


Figure 2. Partial transfer entropy between three non-linearly coupled Gaussian processes. The upper panel displays the partial transfer entropy (PTE) in directions compatible with the structural coupling of Gaussian processes (X to Y to Z). The lower panel displays the PTE values in directions non-compatible with the structural coupling. The solid lines represent PTE values, while the color-matched dashed lines denote corresponding $p = 0.05$ significance levels. $k = 20$. The time window for the search of neighbors is $2\sigma = 10$. The temporal variance of the PTE estimates was reduced with a post-processing moving average filter of order 20.

captures both the onset/offset and the oscillatory profile of the effective coupling across the three processes. On the other hand, the naive average estimator Equation (5) did not reveal any significant flow of information between the three time series.

To evaluate the robustness and performance of the entropy combination estimator to real levels of noise and measurements variability, we also present a second example derived from experimental data on electronic circuits. The system consists of two nonlinear Mackey–Glass circuits unidirectionally coupled through their voltage variables. The master circuit is subject to a feedback loop responsible for generating high dimensional chaotic dynamics. A time-varying effective coupling is then induced by periodically modulating the strength of the coupling between circuits as controlled by an external CPU. Thus, the voltage variables of Circuits 1 and 2 are assumed to follow a stochastic dynamics of the type:

$$\begin{aligned}\frac{dx_1}{dt} &= \beta_1 \frac{x_{1\delta}}{1 + x_{1\delta}^n} - \gamma_1 x_1 + \eta_1, \\ \frac{dx_2}{dt} &= \beta_2 \frac{(1/2 + 1/4 \sin(\omega t)) x_{1\tau}}{1 + x_{1\tau}^n} - \gamma_2 x_2 + \eta_2,\end{aligned}\quad (7)$$

where x_δ represents the value of the variable x at time $t - \delta$, γ , β and n are positive numbers and η represent noise sources. The feedback loop of the first circuit and time-varying coupling between the two circuits are represented by the first terms of each equation, respectively. We note that the former set of equations was not used to sample data. Instead, time series were directly obtained from the voltage

variables of the electronic circuits. The equations above just serve to illustrate in mathematical terms the type of dynamics expected from the electronic circuits.

Thus, we applied transfer entropy between the voltage signals directly generated from the two electric circuits for 180 trials, each 1,000 sampling times long. Delay embedding and statistical significance analysis proceeded as in the previous example. Figure 3 shows the TE ensemble estimates between the master and slave circuit obtained with Equation (6) *versus* the temporal lag introduced between the two voltage signals (intended to scan the unknown coupling delay τ). Clearly, there is a directional flow of information time-locked at lag $\tau = 20$ samples, which is significant for all time instants ($p < 0.01$).

The results show that the TE ensemble estimates accurately capture the dynamics of the effect exerted by the master circuit on the slave circuit. On the other hand, the flow of information in the opposite direction was much smaller ($T_{1 \leftarrow 2} < 0.0795$ nats $\forall(t, \tau)$) and only reached significance ($p < 0.01$) for about 1% of the tuples (n, τ) Figure 4. Both the period of the coupling dynamics (100 samples) and the coupling delay (20 samples) can be accurately recovered from Figure 3.

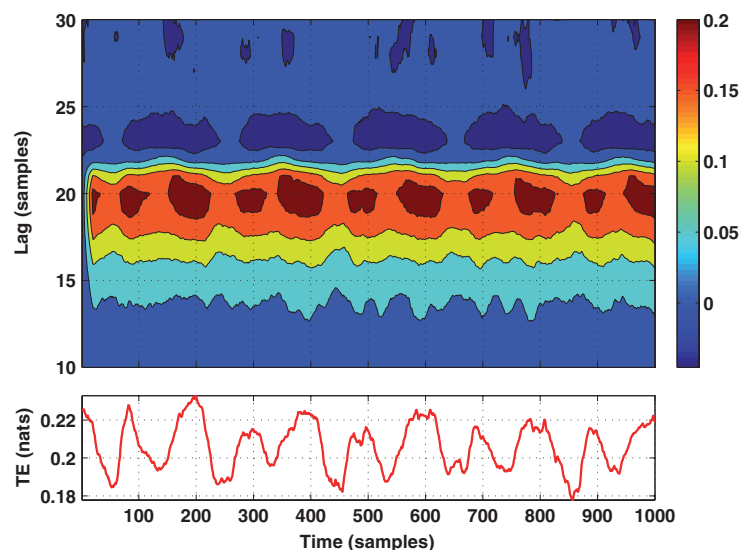


Figure 3. Transfer entropy from the first electronic circuit towards the second. The upper figure shows time-varying TE *versus* the lag introduced in the temporal activation of the first circuit. The lower figure shows that the temporal pattern of information flow for $\tau = 20$, *i.e.*, $T_{2 \leftarrow 1}(n, \tau = 20)$, which resembles a sinusoid with a period of roughly 100 data samples.

Finally, we also performed numerical simulations to study the behavior of the bias and variance of the ensemble estimator with respect to the number of neighbors chosen and the sample size. In particular, we simulated two unidirectionally-coupled Gaussian linear autoregressive processes ($Y \rightarrow X$) for which the analytical values of TE can be known [29], so that we could compare the numerical and expected values. Then, we systematically varied the level of nominal TE (which was controlled by the chosen level of correlation coefficient between $X(t + 1)$ and $X(t)$), the number of neighbors chosen and the sample size and compute measures of bias and variance. Figure 5 and 6 display in a color-coded manner the quantities $-20 \times \log_{10}(\text{bias})$ and $-20 \times \log_{10}(\text{var})$, so large values of these quantities correspond to small bias and variances, respectively. In particular, Figure 5 shows the bias and variance of the estimator as a function of the number of samples and cross-correlation coefficient. As observed in the plot, the

smaller the value of the underlying TE (smaller cross-correlation), the better its estimation (smaller bias and variance). For a given value of TE, the estimation improves as more samples are included, as is expected. Regarding the number of neighbors (Figure 6), we obtain that beyond a minimum number of samples, the accuracy obtained increased by either increasing the sample size or the number of neighbors.

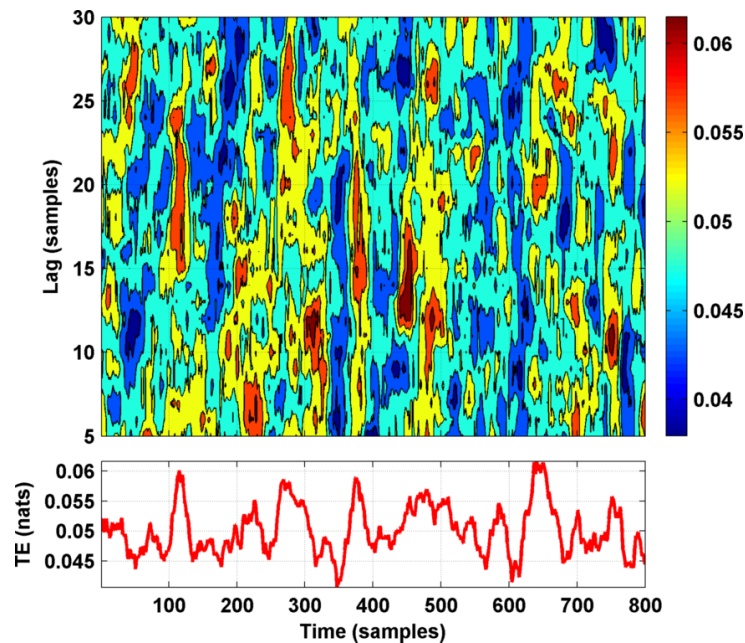


Figure 4. Transfer entropy from the second electronic circuit towards the first. The upper figure shows time-varying TE *versus* the lag introduced in the temporal activation of the first circuit. The lower figure shows that the temporal pattern of information flow for $\tau = 20$, *i.e.*, $T_{1 \leftarrow 2}(n, \tau = 20)$.

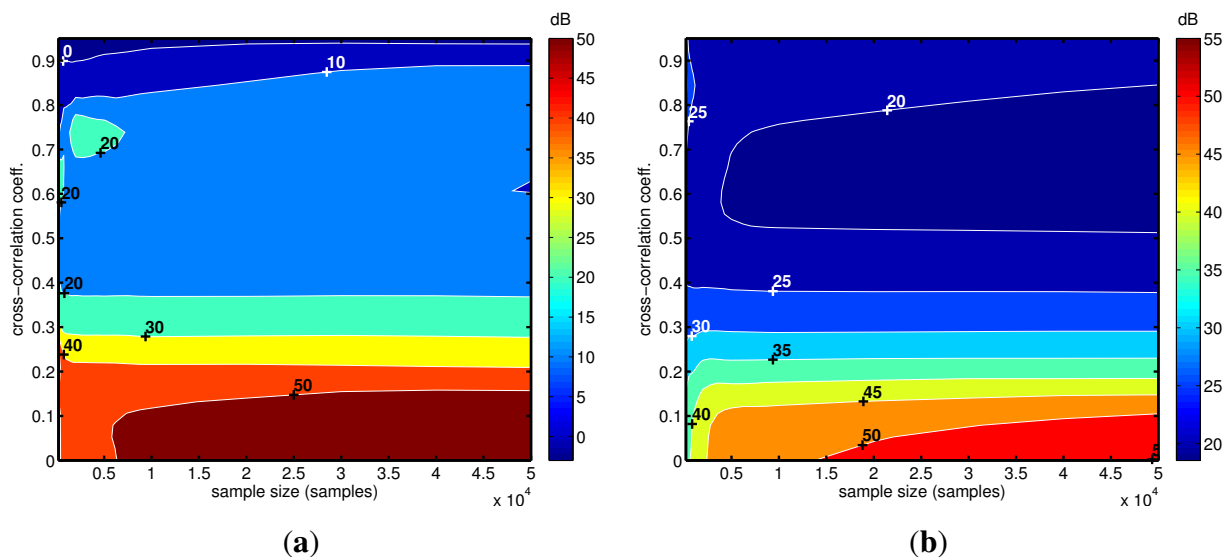


Figure 5. (a): $-20 \times \log_{10}(\text{bias})$ of ensemble estimator $\text{TE}(Y \rightarrow X)$ as a function of the number of samples and cross-correlation coefficient for X (which controls the nominal TE value for $(Y \rightarrow X)$). (b): $-20 \times \log_{10}(\text{variance})$ as a function of the number of samples and cross-correlation coefficient for X (which controls the nominal TE value for $(Y \rightarrow X)$).

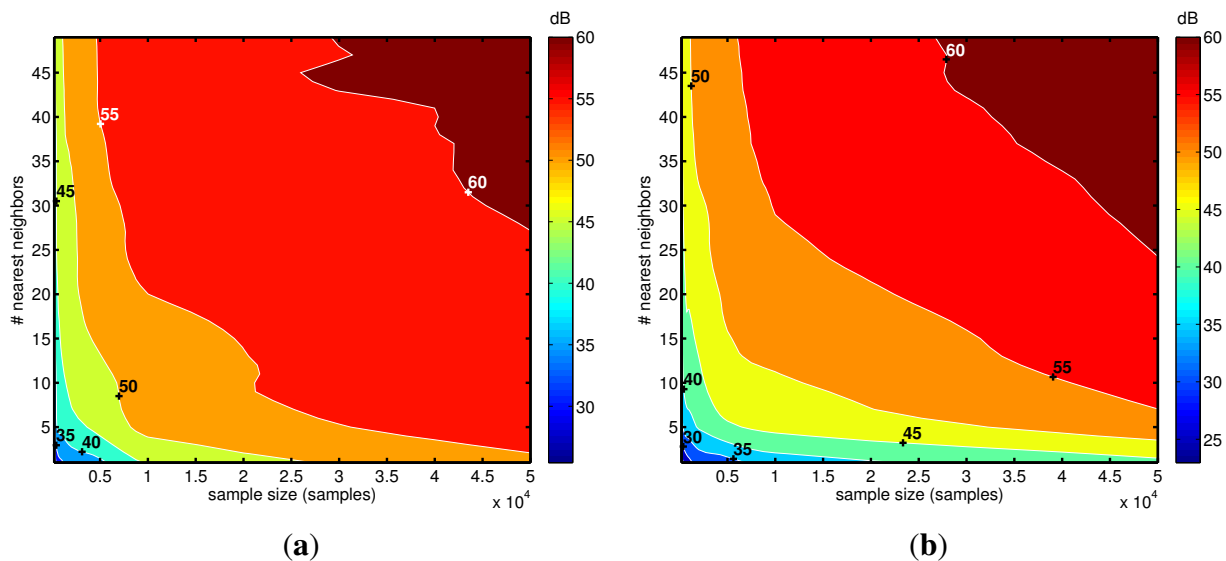


Figure 6. (a): $-20 \times \log_{10}(\text{bias})$ of ensemble estimator $\text{TE}(Y \rightarrow X)$ as a function of the number of samples and the number of nearest neighbors used in the estimator. (b): $-20 \times \log_{10}(\text{variance})$ as a function of the number of samples and the number of nearest neighbors used in the estimator.

5. Conclusions

In conclusion, we have introduced an ensemble estimator of entropy combinations that is able to detect time-varying information flow between dynamical systems, provided that an ensemble of repeated measurements is available for each system. The proposed approach allows one to construct time-adaptive estimators of MI, PMI, TE and PTE, which are the most common information-theoretic measures for dynamical coupling analyses. Using simulations and real physical measurements from electronic circuits, we showed that these new estimators can accurately describe multivariate coupling dynamics. However, strict causal interpretations of the transfer entropy analyses are discouraged [10].

It is also important to mention that intrinsic to our approach is the assumption that the evolution of the interdependencies to be detected are to some degree “locked” to the trial onset. In the setting of electrophysiology and the analysis of event-related potentials, the dispersion of the dynamics with respect to their onset is clearly an acute issue. Indeed, the key distinction between evoked and induced responses rests upon time-locking to a stimulus onset. In principle, one could apply the ensemble average entropic measures to induced responses as measured in terms of the power of the signals, even when they are not phase-locked to a stimulus. In general, the degree of locking determines the maximum temporal resolution achievable by the method (which is controlled via σ). Nevertheless, it is possible to use some alignment techniques [30] to reduce the possible jitter across trials and, thus, increase the resolution.

The methods presented here are general, but we anticipate that a potential application might be the analysis of the mechanisms underlying the generation of event-related brain responses and the seasonal variations of geophysical, ecological or economic variables. Efficient implementations of the ensemble estimators for several information-theoretic methods can be found in [31,32].

Acknowledgments

We are indebted to the anonymous referees for their constructive and valuable comments and discussions that helped to improve this manuscript. This work has been supported by the EU project GABA(FP6-2005-NEST-Path 043309), the Finnish Foundation for Technology Promotion, the Estonian Research Council through the personal research grants P.U.T.program (PUT438 grant), the Estonian Center of Excellence in Computer Science (EXCS) and a grant from the Estonian Ministry of Science and Education (SF0180008s12).

Author Contributions

Germán Gómez-Herrero and Raul Vicente conceptualized the problem and the technical framework. Germán Gómez-Herrero, Gordon Pipa, and Raul Vicente managed the project. Wei Wu, Kalle Rütanen, Germán Gómez-Herrero, and Raul Vicente developed and tested the algorithms. Miguel C. Soriano collected the experimental data. All authors wrote the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Gray, C.; König, P.; Engel, A.; Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **1989**, *338*, 334.
2. Bjornstad, O.; Grenfell, B. Noisy clockwork: Time series analysis of population fluctuations in animals. *Science* **2001**, *293*, 638.
3. Granger, C.; Hatanaka, M. *Spectral Analysis of Economic Time Series*; Princeton University Press: Princeton, NJ, USA, 1964.
4. Valdes-Sosa, P.A.; Roebroeck, A.; Daunizeau, J.; Friston, K. Effective connectivity: Influence, causality and biophysical modeling. *Neuroimage* **2011**, *58*, 339.
5. Granger, C. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424.
6. Pereda, E.; Quiñero, R.; Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobio.* **2005**, *77*, 1.
7. Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley: Hoboken, NJ, USA, 2006.
8. Wiener, N. The theory of prediction. In *Modern Mathematics for Engineers*; McGraw-Hill: New York, NY, USA, 1956.
9. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461.
10. Chicharro, D.; Ledberg, A. When two become one: the limits of causality analysis of brain dynamics. *PLoS One* **2012**, *7*, e32466.
11. Wibral, M.; Vicente, R.; Lizier, J.T. *Directed Information Measures in Neuroscience*; Springer: Berlin, Germany, 2014.

12. Wibral, M.; Vicente, R.; Lindner, M. Transfer entropy in Neuroscience. In *Directed Information Measures in Neuroscience*; Wibral, M., Vicente, R., Lizier, J.T., Eds.; Springer: Berlin, Germany, 2014.
13. Lizier, J.T. *The Local Information Dynamics of Distributed Computation in Complex Systems*; Springer: Berlin, Germany, 2013.
14. Kantz, H.; Schreiber, T. *Nonlinear Time Series Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.
15. Wyner, A.D. A Definition of Conditional Mutual Information for Arbitrary Ensembles. *Inf. Control* **1978**, *38*, 51.
16. Frenzel, S.; Pompe, B. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **2007**, *99*, 204101.
17. Verdes, P.F. Assessing causality from multivariate time series. *Phys. Rev. E* **2005**, *72*, 026222.
18. Gómez-Herrero, G. Ph.D. thesis, Department of Signal Processing, Tampere University of Technology, Finland, 2010.
19. Ragwitz, M.; Kantz, H. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys. Rev. E* **2002**, *65*, 056201.
20. Victor, J.D. Binless strategies for estimation of information from neural data. *Phys. Rev. E* **2002**, *66*, 051903.
21. Vicente, R.; Wibral, M. Efficient estimation of information transfer. In *Directed Information Measures in Neuroscience*; Wibral, M., Vicente, R., Lizier, J.T., Eds.; Springer: Berlin, Germany, 2014.
22. Kozachenko, L.; Leonenko, N. Sample Estimate of the Entropy of a Random Vector. *Problemy Peredachi Informatsii* **1987**, *23*, 9.
23. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
24. Kramer, M.A.; Edwards, E.; Soltani, M.; Berger, M.S.; Knight, R.T.; Szeri, A.J. Synchronization measures of bursting data: application to the electrocorticogram of an auditory event-related experiment. *Phys. Rev. E* **2004**, *70*, 011914.
25. Cao, L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D* **1997**, *110*, 43.
26. Wibral, M.; Pampu, N.; Priesemann, V.; Siebenhühner, S.; Seiwert, S.; Lindner, M.; Lizier, J.T.; Vicente, R. Measuring information-transfer delays. *PLoS One* **2013**, *8*, e55809.
27. Wollstadt, P.; Martínez-Zarzuela, M.; Vicente, R.; Diaz-Pernas, F.J.; Wibral, M. Efficient transfer entropy analysis of non-stationary neural time series. *PLoS One* **2014**, *9*, e102833.
28. Pesarin, F. *Multivariate Permutation Tests*; John Wiley and Sons: Hoboken, NJ, USA, 2001.
29. Kaiser, A.; Schreiber, T. Information transfer in continuous processes. *Physica D* **2002**, *166*, 43.
30. Kantz, H.; Ragwitz, M. Phase space reconstruction and nonlinear predictions for stationary and nonstationary Markovian processes. *Int. J. Bifurc. Chaos* **2004**, *14*, 1935.
31. Rutanen, K. TIM 1.2.0. Available online: <http://www.tut.fi/tim> (accessed on 2 April 2015).

32. 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 Neurosci.* **2011**, *12*, 119.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).