

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/265466743>

MuTE: A new Matlab toolbox for estimating the multivariate Transfer entropy in physiological variability series

Conference Paper · May 2014

DOI: 10.1109/ESGCO.2014.6847518

CITATION

1

READS

272

3 authors:



[Alessandro Montalto](#)

Ghent University

12 PUBLICATIONS 44 CITATIONS

SEE PROFILE



[Luca Faes](#)

Bruno Kessler Foundation, and Università ...

126 PUBLICATIONS 1,523 CITATIONS

SEE PROFILE



[Daniele Marinazzo](#)

Ghent University

104 PUBLICATIONS 1,232 CITATIONS

SEE PROFILE

MuTE: a New Matlab Toolbox for Estimating the Multivariate Transfer Entropy in Physiological Variability Series

Alessandro Montalto*, Luca Faes, and Daniele Marinazzo

Abstract—We present a new time series analysis toolbox, developed in Matlab, for the estimation of the Transfer entropy (TE) between time series taken from a multivariate dataset. The main feature of the toolbox is its fully multivariate implementation, that is made possible by the design of an approach for the non-uniform embedding (NUE) of the observed time series. The toolbox is equipped with parametric (linear) and non-parametric (based on binning or nearest neighbors) entropy estimators. All these estimators, implemented using the NUE approach in comparison with the classical approach based on uniform embedding, are tested on RR interval, systolic pressure and respiration variability series measured from healthy subjects during head-up tilt. The results support the necessity of resorting to NUE for obtaining reliable estimates of the multivariate TE in short-term cardiovascular and cardiorespiratory variability.

I. INTRODUCTION

The Transfer entropy (TE) is a powerful tool to detect the transfer of information between joint processes [1]. Although its underlying concept has been increasingly used to assess directional dependencies between physiological time series [2,10], the practical computation of the TE is hampered by the difficulty that it requires the estimation of entropies in high-dimensional spaces. This is a daunting task, especially in the challenging conditions posed by physiological time series analysis where usually only short time series are available due to experimental constraints. For this reason, reliable practical estimators of the TE have been proposed mostly in their bivariate version where only two time series are considered at a time [2,4].

However, bivariate TE analyses may be seriously limited as they may lead to the inference of wrong interactions in the presence of unobserved confounders; an example in physiological variability is the confounding effect of respiration on the interactions between arterial pressure and heart rate variability [5]. To overcome these limitations, a data-efficient approach has been proposed recently which tackles the problem of high-dimensional entropy estimation [3,6]. The present study introduces a Matlab toolbox that implements this approach exploiting different entropy estimators, thus allowing to estimate the TE in the multivariate case. In the following, the toolbox is described in its main features and then used for assessing the multivariate TE in short-term cardiovascular variability.

This work is supported by: the Belgian Science Policy (IUAP VII project CEREBNET P7 11); the University of Gent (Special Research Funds for visiting researchers); copyright: 978-1-4799-3969-5/14/ \$31.00 ©2014 IEEE

Alessandro Montalto is with Gent University, Gent, Belgium (*corresponding author: alessandro.montalto@ugent.be)

Luca Faes is with University of Trento, Trento, Italy

Daniele Marinazzo is with Gent University, Gent, Belgium

II. THE MUTE TOOLBOX

The MuTE toolbox implements several approaches for the estimation of the multivariate TE. The underlying theoretical framework considers a set of M interacting dynamical systems, visiting states which are described by stationary stochastic processes. We are interested in evaluating the information flow from the source system X to the destination system Y , collecting the remaining $M - 2$ processes in the vector \mathbf{Z} . Denoting as X_n , Y_n and \mathbf{Z}_n the variables describing the present of the processes at time n , and as $X_n^- = [X_{n-1} X_{n-2} \dots]$, $Y_n^- = [Y_{n-1} Y_{n-2} \dots]$ and $\mathbf{Z}_n^- = [\mathbf{Z}_{n-1} \mathbf{Z}_{n-2} \dots]$ the variables describing their past up to time $n - 1$, the multivariate TE from X to Y conditioned to \mathbf{Z} is

$$\begin{aligned} TE &= H(Y_n | Y_n^-, \mathbf{Z}_n^-) - H(Y_n | Y_n^-, X_n^-, \mathbf{Z}_n^-) \\ &= H(Y_n, Y_n^-, \mathbf{Z}_n^-) - H(Y_n^-, \mathbf{Z}_n^-) \\ &\quad - H(Y_n, Y_n^-, X_n^-, \mathbf{Z}_n^-) + H(Y_n^-, X_n^-, \mathbf{Z}_n^-) \quad (1) \end{aligned}$$

where $H(\cdot)$ and $H(\cdot|\cdot)$ denote respectively entropy and conditional entropy (CE).

The first crucial step for the estimation of the TE is to approximate the infinite-dimensional variables representing the past of the processes. This problem can be seen in terms of performing suitable multivariate embedding of the considered time series [7]. The most obvious approach is uniform embedding (UE), whereby the vector Y_n^- is approximated using the embedding vector $V_n^Y = [Y_{n-m} Y_{n-2m} \dots Y_{n-dm}]$, where d and m are the embedding dimension and embedding delay (the same for X_n^- and \mathbf{Z}_n^- , approximated by V_n^X and V_n^Z). As an alternative to UE, the toolbox implements an approach based on non-uniform embedding (NUE) [3,6]. The approach is based on the progressive selection, from a set of candidate variables including the past of X , Y , and \mathbf{Z} considered up to a maximum lag, of the lagged variables which are more informative about the target variable Y_n . At each step, selection is performed maximizing the mutual information between the candidate and the target variable, conditioned to the variables selected up to that step. This achieves a criterion for maximum relevance and minimum redundancy for candidate selection, so that the resulting embedding vector $V = [V_n^X V_n^Y V_n^Z]$ includes only the components of X_n^- , Y_n^- and \mathbf{Z}_n^- , which contribute most to the description of Y_n .

After multivariate embedding, estimation of the TE results from the application of estimators of entropy and CE to the various terms in (1). The toolbox contains three of such estimators. The first is the linear estimator (LIN) that assumes that data are Gaussianly distributed. Under this assumption, the two CE terms defining the TE can

be quantified by means of linear regressions involving the relevant variables taken from the embedding vector [8]. The second estimator is the classical binning estimator (BIN), which consists in coarse-graining the observed dynamics using Q quantization levels, and then computing entropies by approximating probability distributions with the frequencies of occurrence of the quantized values. The third estimator is based on k -nearest neighbor techniques (NN) which exploit the statistics of distances between neighboring data points in the embedding space to estimate entropy terms. We adopted the bias-reduction method of estimating entropies through neighbor search in the space of higher dimension and range searches in the subspaces of lower dimension [4,9].

The toolbox is designed to run in parallel mode across methods or realizations, setting for each case the source and destination time series to be passed to TE estimation. It allows to take instantaneous effects among the observed time series into account, based on their physiological or non-physiological meaning [5]. Moreover, it provides tools for the selection of embedding parameters and the assessment of the statistical significance of each estimated TE. The latter issue is addressed exploiting the procedure for candidate selection in case of NUE, using F-statistics for the LIN UE method, and using surrogate data for the BIN UE and NN UE methods. The execution time for a single run of the dataset analysed here ranged from 0.01 s for the LIN UE to 17.2 s for NN NUE.

III. APPLICATION

This Section reports the application of the toolbox on cardiovascular and cardiorespiratory time series measured from 15 young healthy subjects undergoing a head-up tilt testing protocol [10]. The series consisted in the beat-to-beat variability sequences (300 points) of the cardiac RR intervals, systolic arterial pressure (SAP) and respiration (Resp) measured in the resting supine position (SU) and in the 60° upright position (UP) reached through passive head-up tilt. The TE was computed between all pairs of normalized series, using all estimators based on UE (parameters: $d = 5, m = 1$) and NUE (parameters: $Q = 6, k$ nearest neighbours = 10). Physiologically relevant instantaneous effects were set from Resp to SAP and RR, and from SAP to RR [5].

The results, shown in Fig. 1, document the capability of the estimators to reflect, in terms of TE magnitude and number of subjects with statistically significant TE, expected cardiovascular and cardiorespiratory regulation mechanisms [10]. Examples are the detection of an unidirectional interaction from Resp to RR and from Resp to SAP (elements 3,1 and 3,2 vs. 1,3 and 3,1 of each TE matrix), reflecting respiration-related influences on cardiovascular variability; the decrease from SU to UP of the TE from Resp to RR (element 3,1), documenting a weakening of respiratory sinus arrhythmia; and the increase from SU to UP of the TE from SAP to RR (element 2,1), documenting an involvement of the baroreflex. Note that UE used with BIN and NN estimators was unable to detect these mechanisms with respect to NUE; this is due to the curse of dimensionality which prevents reliable non-parametric estimation of entropies in high-dimensional spaces. The results were described equally well using non-parametric estimators based on NUE and

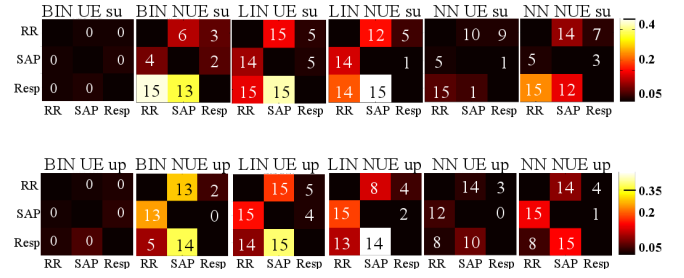


Fig. 1. TE matrices obtained for each estimator applied to cardiovascular data in the SU and UP conditions of head-up tilt. The i, j element of each matrix depicts color-coded values of the TE from i to j conditioned to k , and contains the number of subjects for which the TE was detected as significant (i, j and k are any of RR, SAP and Resp time series).

the LIN estimator, suggesting a linear nature for the mechanisms governing the information flows in the considered tilt protocol. Nevertheless, a better discrimination capability for BIN and NN estimators based on NUE was observed when looking at the number of subjects with significant TE, suggesting the potential of the scheme for candidate selection as to the evaluation of the statistical significance of TE.

IV. CONCLUSIONS

The most distinctive feature of the proposed MuTE toolbox, compared with existing toolboxes for TE estimation such as the TRENTTOOL [4], is its fully multivariate implementation. The practical estimation of the multivariate TE is made possible by the proposed NUE scheme, which performs a parsimonious reconstruction of the past states of the observed systems to limit the curse of dimensionality. The usefulness of this approach is confirmed by the reported application, which showed the necessity of NUE to perform non-parametric estimation of the multivariate TE in short term cardiovascular and cardiorespiratory variability.

REFERENCES

- [1] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, p. 461, 2000.
- [2] J. Lee, S. Nemat, I. Silva, B. A. Edwards, J. P. Butler, A. Malhotra et al., "Transfer entropy estimation and directional coupling change detection in biomedical time series," *Biomed. Eng. Online*, vol. 11, no. 1, pp. 117, 2012.
- [3] D. Kugiumtzis, "Direct-coupling information measure from nonuniform embedding," *Phys. Rev. E*, vol. 87, p. 062918, Jun 2013.
- [4] M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, "TRENTTOOL: A matlab open source toolbox to analyse information flow in time series data with transfer entropy," *BMC Neuroscience*, vol. 12, no. 1, p. 119, 2011.
- [5] L. Faes, G. Nollo, and A. Porta, "Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series," *Entropy*, vol. 15, no. 1, pp. 198219, 2013.
- [6] L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique," *Phys. Rev. E*, vol. 83, p. 051112, 2011.
- [7] I. Vlachos and D. Kugiumtzis, "Nonuniform state-space reconstruction and coupling detection," *Phys. Rev. E*, vol. 82, no. 1, p. 016207, 2010.
- [8] L. Barnett, A. Barrett, and A. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," *Phys. Rev. Lett.*, vol. 103, no. 23, p. 238701, 2009.
- [9] A. Kraskov, H. Stogbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, p. 066138, 2004.
- [10] L. Faes, G. Nollo, and A. Porta, "Information domain approach to the investigation of cardio-vascular, cardio-pulmonary and vasculo-pulmonary causal couplings," *Front. Physiol.*, vol. 2, no. 80, 2011.