Simulation Study of Direct Causality Measures and Lag Estimations in Multivariate Time Series

Jolan Heyse[1], Pieter Van Mierlo[1]

-

[1]MEDISIP, Ghent University, Ghent, Belgium

Jolan Heyse – email jolan.heyse@ugent.be

STUDY OBJECTIVES

The detection and characterization of the causal effects among simultaneously observed systems provides valuable knowledge about the underlying processes and network, and is a topic of interests in many scientific fields. Many causality measures have been developed, each with their own advantages and disadvantages. In this work we consider some of the best-known multivariate causality measures, i.e. conditional Granger causality index (CGCI), partial directed coherence (PDC), directed transfer function (DTF) and partial mutual information on mixed embedding (PMIME). Their performance is assessed on stochastic and chaotic coupled and uncoupled dynamical systems for different settings of embedding dimension and time series length.

However, detection of the connectivity patterns alone is not enough. As the lag between communicating variables can have a significant impact on the network dynamics, it's important to provide an estimation of these lag values. This is often forgotten and could hold essential information. For each connectivity measure, we propose a novel method of estimating the lag values of the detected interactions.

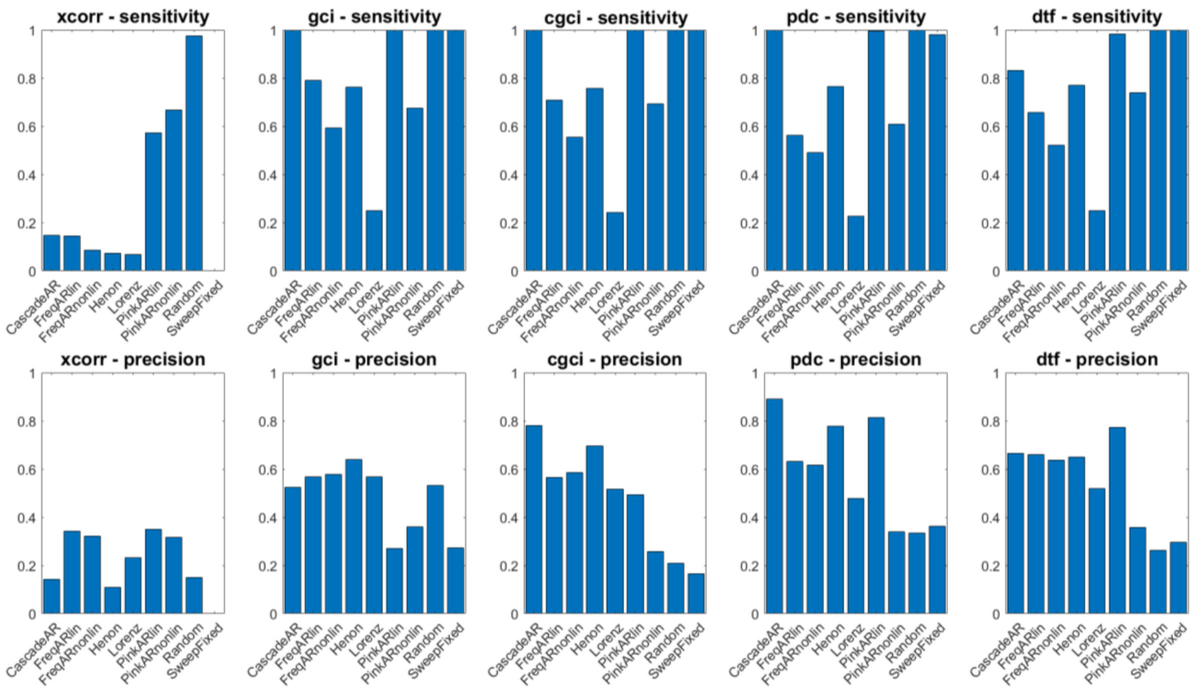
METHODS

Causality measures - The detected connectivity pattern depends strongly on the used connectivity measure. In this work we compare the performance of six well-known methods for detecting directed causal interactions: cross-correlation, (conditional) Granger causality index (CGCI), partial directed coherence (PDC), directed transfer function (DTF) and partial mutual information of a mixed embedding (PMIME).

CGCI is calculated from the coefficients of a multivariate autoregressive (MVAR) model fitted to the time series. [1] Both PDC and DTF are related to the same MVAR model, but are defined in the frequency domain. [2,3] Since AR models are inherently linear, these methods (as well as cross-correlation) will predominantly detect linear interactions. Detection of non-linear interactions is non-trivial and requires specialized techniques. Information theory sets a natural framework for non-parametric methodologies of statistical dependencies, which opens the possibility to detecting all types of interactions independent of their origin. To this end we also implement PMIME, which is based on conditional mutual information. [4]

Simulation models - The multivariate causality measures are evaluated in a simulation study, using a variety of models. All the considered coupling measures are computed on 100 realizations of multivariate uncoupled and coupled systems in various topologies. To provide a wide range of possible scenarios, the models contain linear as well as non-linear interactions and are generated through stochastic (i.e. autoregressive) processes and chaotic maps (Hénon maps and Lorenz systems).

Lag estimations - For each causality measure we developed a method for estimating the lag values of the interactions detected between the nodes of the network. The performance of these methods is also evaluated based on the 100 realizations of the simulation models mentioned above.



RESULTS

Performance of the causality measures is expressed in terms of sensitivity and precision. Preliminary results show that CGCI, PDC and DTF perform very well on linearly coupled systems and significantly outperform cross-correlation, as can be seen in the figure above. In case non-linear interaction terms are present, the sensitivity and precision drops, as was expected. Results for PMIME are not yet included because of its high computational cost.

Large variations exist for each methods performance in function of the simulation model. Especially the low sensitivity in detecting couplings between Lorenz systems is noteworthy. As these interactions are purely non-linear it is expected that PMIME will outperform the other measures for this specific case.

In the table below we summarize the accuracy of the estimated lag values for a selection of simulation models and estimation methods. The results are expressed as the absolute difference between the true lag and the estimated value (expressed in number of samples, f_s=256Hz), averaged over all realizations.

	gci	pdc
Hénon	0.83 +- 2.47	0.51 +- 1.32
Lorenz	0.11 +- 0.74	0.30 +- 1.29
PinkAR	0.99 +- 2.59	0.66 +- 1.46

CONCLUSIONS

Advanced causality measures such as CGCI, PDC and DTF prove to outperform simpler methods such as cross-correlation. However it should be noted that the accuracy is never perfect and results from causality studies should always be interpreted with caution. Furthermore, the performance of these measures shows to be strongly dependent on the simulation model, indicating the importance of choosing the best fitting method depending on the type of interactions that are suspected.

The newly developed lag estimation methods show great promise to be able to estimate the lag values with an accuracy of within a few samples. These results should however be replicated and further validated in different scenarios.

REFERENCES

[1] Brandt P.T. et al. (2007), Multiple Time Series Models, *Sage Publications*, chapter 2, pp. 32-34

[2] Baccala et al. (2001), Partial directed coherence: a new concept in neural structure determination, *Biol. Cybern.*, 84, 463-474

[3] Kaminski et al. (1991), A new method of the description of the information flow in the brain structures, *Biol. Cybergn.*, 65, 203-210

[4] Kugiumtzis D. et al. (2013), Direct coupling information measure from non-uniform embedding, *Phys. Rev. E*, 87