Journal of Neuroscience Methods 198 (2011) 344–358

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Reliability of multivariate causality measures for neural data

Esther Florin[a,b,*], Joachim Gross[c], Johannes Pfeifer[d], Gereon R. Fink[a,b], Lars Timmermann[a,**,1]

[a] Department of Neurology, University Hospital Cologne, Cologne, Germany
[b] Institute of Neuroscience and Medicine (INM-3), Cognitive Neurology Section, Research Center Jülich, Germany
[c] Centre for Cognitive Neuroimaging (CCNi), Department of Psychology, University of Glasgow, United Kingdom
[d] Bonn Graduate School of Economics (BGSE), Department of Economics, University of Bonn, Germany

ARTICLE INFO

Article history:
Received 21 April 2010
Received in revised form 12 March 2011
Accepted 5 April 2011

Keywords:
Partial directed coherence
Directed transfer function
Zero-and-one inflated beta regression
Pathophysiology of Parkinson's disease

ABSTRACT

In the past decade several multivariate causality measures based on Granger causality have been suggested to assess directionality of neural signals. To date, however, a detailed evaluation of the reliability of these measures is largely missing. We systematically evaluated the performance of five different causality measures (squared partial directed coherence (sPDC), partial directed coherence (PDC), directed transfer function (DTF), direct directed transfer function (dDTF) and transfer function) depending upon data length, noise level, coupling strength, and model order and performed simulations based on four different neural data recording procedures (magnetoencephalography, electroencephalography, electromyography, intraoperative local field potentials). Moreover, we analyzed the effect of two common numerical methods to determine the significance of the particular causality measure (random permutation and the leave one out method (LOOM)). The simulations showed the sPDC combined with the LOOM to be the most reliable and robust choice for assessing directionality in neural data. While DTF and H by construction were unable to distinguish between direct and indirect connections, the dDTF also failed this test. Finally, we applied the causality measures to a real data set. This showed the usefulness of our simulation results for practical applications in order to draw correct inferences and distinguish between conflicting evidence obtained with different causality measures.

1. Introduction

There have been exciting recent developments in functional connectivity analysis that allow investigating interactions between macroscopic brain areas and their consequences for behavior (Friston et al., 1997; Stephan et al., 2007). Connectivity analysis of electrophysiological data is typically performed using the classical coherence function (Schnitzler and Gross, 2005a), which does not allow to assess causality or directionality (Timmermann et al., 2007). Furthermore, coherence is incapable of distinguishing between direct and indirect interactions. In the latter case, the information transfer between two signals is mediated by a third signal. These constraints can be overcome, at least in part, by connectivity measures based on the idea of Granger causality (Granger, 1969), which formalizes Wiener's idea of causality for two simultaneously measured signals in a statistical framework (Wiener, 1956): if one can predict the first signal better by incorporating the past information from the second signal than by using only information from the first one, then the second signal can be called causal for the first one. However, traditional (bivariate) Granger causality can also not disentangle indirect influences (Kus et al., 2004). To overcome this drawback, multivariate methods have been introduced recently to the analysis of multi-channel neural data (Astolfi et al., 2006; Baccala and Sameshima, 2001; Kaminski and Blinowska, 1991; Korzeniewska et al., 2003). Before the introduction of truly multivariate methods, conditional Granger causality (CGC) was suggested (Geweke, 1984). With CGC the influence of a third signal can be taken into account and mutual tests are calculated to disentangle the various influences. However, with more channels this approach becomes cumbersome as either many

Abbreviations: AIC, Akaike information criterion; CGC, conditional Granger causality; DBS, deep brain stimulation; DGP, data generating process; DTF, directed transfer function; dDTF, direct directed transfer function; dl, data length; EDC, M. extensor digitorum communis; EEG, electroencephalography; EMG, electromyography; FDI, first dorsal interosseus; FDL, M. flexor digitorum superficialis; ffDTF, full frequency directed transfer function; FP, false positive detection; H, transfer function; LOOM, leave one out method; LFP, local field potential; MEG, magnetoencephalography; moa, model order above the true one; mob, model order below the true one; MVAR, multivariate autoregressive; PDC, partial directed coherence; RP, random permutation; SNR, signal to noise ratio; sPDC, squared partial directed coherence; STN, nucleus subthalamicus; ZI, zona incerta.

* Corresponding author at: Department of Neurology, University Hospital Cologne, Kerpener Str. 62, 50937 Köln, Germany. Tel.: +49 221 478 86416; fax: +49 221 89002.

** Corresponding author at: Department of Neurology, University Hospital Cologne, Kerpener Str. 62, 50937 Köln, Germany. Tel.: +49 221 478 7494; fax: +49 221 87512.

E-mail addresses: Esther.Florin@uk-koeln.de (E. Florin), Lars.Timmermann@uk-koeln.de (L. Timmermann).

0165-0270/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2011.04.005

triplets have to be tested or the reference signal has to be constructed as a composite vector of the remaining data channels (see, e.g. Zhou et al. (2009)). Thus, CGC is not considered any further in the following study.

As a first step, Astolfi et al. performed simulations with electroencephalographic (EEG) data, where they evaluated the influence of noise and data length on multivariate causality measures (Astolfi et al., 2007). Thus far, however, the performance of these measures with regard to both other data measurement methods such as magnetoencephalography (MEG) or electromyograms (EMG), as well as other parameters, such as model order or coupling strength, has not been tested.

The present work thus aims at extending the work by Astolfi et al. by carrying out simulations to evaluate the performance of five different causality measures (sPDC, PDC, DTF, dDTF, transfer function (*H*)) in combination with two commonly used numerical significance computation approaches – a permutation approach (Kaminski et al., 2001) and the leave-one-out method (LOOM) (Schlögl and Supp, 2006) – and the dependence of these measures on the parameters *noise level*, *data length*, *model order*, and *coupling strength*. We generated a system of 7 data channels with a predefined causality structure (Kus et al., 2004), where the first data channel was a real data channel from either EEG, EMG, local field potentials or MEG recordings. Furthermore a data type independent model suggested by Schelter et al. (2006) was employed. With both models we tested whether the different causality measures were able to detect the underlying causality structure. To gauge the effect of data length, noise level, model order, coupling strength, and underlying real data type we systematically varied these parameters when generating the model and performed a regression analysis to obtain a statistically valid estimate of the parameters' influence on the causality measures. Finally an application to LFP and EMG data was performed to demonstrate the usefulness and applicability of the simulation results.

2. Methods

We evaluated the performance of 5 multivariate causality measures in simulations, which are all based on the notion of Granger causality (Granger, 1969): PDC (Baccala and Sameshima, 2001), sPDC (Astolfi et al., 2006), DTF (Kaminski and Blinowska, 1991), dDTF (Korzeniewska et al., 2003),[1] and *H* (Kaminski and Blinowska, 1991). The mathematics for all causality measures is given in Appendix A. For each measure the influence of noise level, data length, model order, and coupling strength was determined for 4 different data types (MEG, EEG, EMG and LFP). Models suggested by Kus et al. (2004) and by Schelter et al. (2006) were used to impose a predefined causality structure on the data.

The MEG recordings were obtained with a whole-head Neuromag 122 MEG-system during rest (Ahonen et al., 1993). For the analysis only channel 96 was used. As signal for the EEG recording the tutorial dataset from EEGLAB was taken (http://www.sccn.ucsd.edu/eeglab). The LFP and EMG were obtained intra-operatively with the INOMED ISIS MER-system (INOMED corp., Teningen, Germany) from a patient with Parkinson's disease and tremor while the patient was holding his arm upwards.

[1] Eichler has shown in 2006 that the dDTF is not a measure of Granger causality (Eichler, 2006), because it will detect connections not detected with Granger causality. We nevertheless test the dDTF in the present study, as it is has been used to detect causality (Giannakakis and Nikita, 2008; Korzeniewska et al., 2008) and we want to gauge whether the dDTF shows performance advantages that might justify its use in real applications despite its known shortcomings.

For each data type and model two ways of testing for significance were used: (i) The *leave one out method* (LOOM) and (ii) *random permutation* of the data. An analytic approach to determine the significance of the causality measures was not pursued, because for some of the multivariate autoregressive (MVAR) measures analytic significance criteria do not yet exist or only exist for large samples, which limits their applicability for smaller data sets often encountered in practical applications (Davidson and MacKinnon, 2004).

2.1. Computation of the level of significance (RP and LOOM)

All multivariate methods require a level of significance in order to differentiate between true connections and noise. We used two different methods to overcome this problem, the *LOOM* (Schlögl and Supp, 2006), which is based on the jackknife (Quenouille, 1949; Tukey, 1958), and the random permutation (*RP*) (Kaminski et al., 2001) of the data, based on the classical bootstrap (Efron, 1979; Efron and Tibshirani, 1993).[2]

Random permutation of the data is a numerical method to obtain a significance estimate by generating surrogate data under the null hypothesis in order to trace out the unknown distribution of a test statistic under this null hypothesis. The random permutation derives its name from randomly interchanging observations of a time series, whereby surrogate data based on the original observations are produced, but where any causality should be removed by randomly changing the time ordering. This has the effect of generating data with the same mean and variance, but where all temporal (correlation) structure and hence any causality is removed from the data. Thus, resampling the data determines the distribution of the causality measures under the null hypothesis of no causation.[3] Note that while our random permutation approach is only valid for the case of i.i.d. noise, it can easily be adapted to more complicated settings using the more general residual bootstrap.[4]

After randomizing the data, an autoregressive model was fitted to the surrogate data and the respective causality measure was calculated. Kaminski et al. (2001) used 100 repetitions in their study. As our simulations still showed a large variability in the results with this value, we adopted 200 repetitions as a compromise between computational speed and accuracy. As the PDC, sPDC, and DTF are bounded by 0 and 1, they can by definition not be normally distributed, so that it is more appropriate to choose a percentile instead of a parametric *p*-value derived from assumed normality.

[2] Kaminski et al. (2001), who first used this method in the context of biological causality analysis, credited Theiler et al.'s "method of surrogate data" (Theiler et al., 1992) as their conceptual predecessor. However, in the basic form applied in their study and the present one, Theiler et al. credited Scheinkman and LeBaron (1989), who based their approach on the bootstrap.

[3] In the context of our particular application to a VAR-structure with i.i.d. noise of the form

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}(k)\,\mathbf{X}(t-k) + \Sigma(t)$$

where $\mathbf{X}(t) = [X_1(t), X_2(t), \ldots, X_n(t)]^T$ is the data vector of all signals at time t, $\Sigma(t)$ represents an error term consisting i.i.d. noise, $A(k)$ is the matrix of autoregressive coefficients for the kth time lag, and p is the maximum number of time lags, this method is equivalent to the residual bootstrap (Lütkepohl, 2005, Appendix D). The reason is that under the null hypothesis of no causation

$$H_0 : A(k) = 0 \quad \forall k.$$

Hence, resampling the observations $X(t)$ is equal to using residual bootstrap to subsequently generate the model (8) under H_0.

[4] For a good overview over the possible methods see the special issue of Econometric Reviews on "Bootstrapping time series models" (Li and Maddala, 1996) and in particular Li and Maddala (1996) who also discuss the potential pitfalls of naively sampling the $X(t)$ instead of properly using the residual bootstrap.