Comparison of Granger Causality and Phase Slope Index

Guido Nolte GUIDO.NOLTE@FIRST.FRAUNHOFER.DE
Intelligent Data Analysis Group, Fraunhofer FIRST
Kekuléstr. 7, 12489 Berlin, Germany

Andreas Ziehe ZIEHE@FIRST.FRAUNHOFER.DE
Intelligent Data Analysis Group, Fraunhofer FIRST
Kekuléstr. 7, 12489 Berlin, Germany

Nicole Krämer NKRAEMER@CS.TU-BERLIN.DE
Machine Learning Group, TU Berlin
Franklinstr. 28/29, 10587 Berlin, Germany

Florin Popescu FLORIN.POPESCU@FIRST.FRAUNHOFER.DE
Intelligent Data Analysis Group, Fraunhofer FIRST
Kekuléstr. 7, 12489 Berlin, Germany

Klaus-Robert Müller KRM@CS.TU-BERLIN.DE
Machine Learning Group, TU Berlin
Franklinstr. 28/29, 10587 Berlin, Germany

Editor: Isabelle Guyon, Dominik Janzing and Bernhard Schölkopf

Abstract

We recently proposed a new measure, termed Phase Slope Index (PSI), It estimates the causal direction of interactions robustly with respect to instantaneous mixtures of independent sources with arbitrary spectral content. We compared this method to Granger Causality for linear systems containing spatially and temporarily mixed noise and found that, in contrast to PSI, the latter was not able to properly distinguish truly interacting systems from mixed noise. Here, we extent this analysis with respect to two aspects: a) we analyze Granger causality and PSI also for non-mixed noise, and b) we analyze PSI for nonlinear interactions. We found a) that Granger causality, in contrast to PSI, fails also for non-mixed noise if the memory-time of the sender of information is long compared to the transmission time of the information, and b) that PSI, being a linear method, eventually misses nonlinear interactions but is unlikely to give false positive results.

Keywords: Phase Slope Index, Granger Causality, Noise, Nonlinearity

1. Introduction

To understand the direction of information flow in interacting systems, it is of fundamental importance to distinguish the driver from the recipient. Granger Causility proposed by Granger (1969) is probably the most prominent method to estimate the direction of causal influence in time series analysis.

Apart from Granger Causality, many other methods have been proposed to estimate the direction of information flow both for bivariate and multivariate data. Baccala and

Sameshima (1998) suggested to interpret autogressive matrices in the frequency domain to estimate directionality for bivariate data, which was generalized to multivariate data by Baccala and Sameshima (2001). The approach of Kaminski and Blinowska (1991) is equivalent to the preceeding ones for bivariate data, but differs for multivariate data most notably with regard to the question whether estimated information flux is direct or indirect. An information theoretic approach was taken by Schreiber (2000) by analyzing entropies of conditional probabilities (rather than the mean as implicitly done with Granger Causality). A model based method valid for nonlinear and weakly coupled oscillators was proposed by Rosenblum and Pikovsky (2001). With the notable exception of Rosenblum and Pikovsky (2001), all these methods are variations of the highly popular Granger causality, and this will serve as a comparison to our proposed method.

Granger Causality is based on asymmetric prediction accuracies of one time series on the future of another. The difficulty in realistic measurements is that asymmetries can also arise due to other factors, specifically independent background activity having nontrivial spectral properties and eventually being measured in unknown superposition in the channels. In this case the interpretation of the asymmetry as a direction of information flow can lead to significant albeit false results as demonstrated e.g. by Albo et al. (2004). To overcome this difficulty Nolte et al. (2008) recently proposed a method based on a frequency-average of the slope of the phase of coherence defined in such way that it is strictly robust with respect to instantaneous mixtures of independent sources of otherwise arbitrary nature.

In this paper we address two new aspects in more detail. First, in many situations one could argue that, while the measurements are noisy, this noise is not a mixture, and Granger Causality might work for this case. Second, the beneficial properties of PSI might disappear if interactions are nonlinear. We will first shortly recall both methods and then study both mentioned aspects with simulations.

2. Methods

2.1 Granger Causality

The fundamental basis of estimates of causal relations using Granger Causality is the fact that a cause precedes the effect. Probably the simplest way to exploit this idea is to use linear prediction of future values of bivariate data $x_i(t)$ for $i = 1, 2$ with AR-modeling:

$$\mathbf{x}(t) = \sum_{p=1}^{P} A(p)\mathbf{x}(t-p) + \xi(t) \tag{1}$$

where $A(p)$ are the AR-matrices up to order P and $\xi(t)$ is white gaussian noise with estimated covariance matrix Σ.

The diagonal elements of Σ (i.e Σ_{ii} for $i = 1, 2$) measure the remaining error when future values of $x_i(t)$ are modeled with both time series', simultaneously. Instead of one multivariate model one can also model the data by two separate univariate models:

$$x_i(t) = \sum_{p=1}^{P} A_i(p)x_i(t-p) + \xi_i(t) \tag{2}$$

for $i = 1, 2$, and where $\xi_i(t)$ has estimated variance Σ_i.

Note that $\Sigma_i \leq \Sigma_{ii}$ because the univariate models do not use information contained in the other time series. The additional information contained in x_j about the future of x_i for $j \neq i$ can be quantified as

$$\Gamma_{j\to i} = \log\left(\frac{\Sigma_i}{\Sigma_{ii}}\right) \tag{3}$$

If $\Gamma_{j\to i} > 0$ one says that channel j 'Granger causes' channel i.

For a unidirectional information flow one has $\Gamma_{1\to 2} = 0$ or $\Gamma_{2\to 1} \neq 0$ or vice versa with obvious direction of information flux. In practice, results are rarely that clear and one can define the effective information flux from the first to the second channel as

$$\tilde{G} = \Gamma_{1\to 2} - \Gamma_{2\to 1} \tag{4}$$

We here normalize $\tilde{G}$ by its standard deviation, estimated by the Jackknife procedure (the validity was confirmed in simulations), and define the Granger Causality finally as

$$G = \frac{\tilde{G}}{std(\tilde{G})} \tag{5}$$

2.2 Phase Slope Index

In an alternative approch we first devide the whole data set into K segments of duration T (in physical units) and estimate the cross-spectral density as

$$S_{ij}(f) = \frac{1}{K}\sum_k z_i(f,k)z_j^*(f,k) \tag{6}$$

where $z_i(f,k)$ is the Fourier transform of the Hanning-windowed data in channel i and segment k. The 'Phase Slope Index' (PSI) is now defined as (Nolte et al. (2008))

$$\tilde{\Psi}_{ij} = \Im\left(\sum_{f\in F} C_{ij}^*(f)C_{ij}(f+\delta f)\right) \tag{7}$$

where

$$C_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}} \tag{8}$$

is the complex coherency, $\delta f = 1/T$ is the frequency resolution, and $\Im(\cdot)$ denotes taking the imaginary part. F is the set of frequencies over which the slope is summed. Typically, F contains all frequencies, but it can also be restricted to a specified band for rhythmic activities.

To see that the definition of $\tilde{\Psi}_{ij}$ corresponds to a meaningful estimate of the average slope it is convenient to rewrite it as

$$\tilde{\Psi}_{ij} = \sum_{f\in F}\alpha_{ij}(f)\alpha_{ij}(f+\delta f)\sin(\Phi(f+\delta f) - \Phi(f)) \tag{9}$$

with $C_{ij}(f) = \alpha_{ij}(f)\exp(i\Phi(f))$ and $\alpha_{ij}(f) = |C_{ij}(f)|$ being frequency dependent weights. For smooth phase spectra, $\sin(\Phi(f+\delta f)-\Phi(f)) \approx \Phi(f+\delta f)-\Phi(f)$ and hence $\tilde{\Psi}$ corresponds to a weighted average of the slope.

Let us list the most important properties of $\tilde{\Psi}$:

1. For an infinite amount of data and for arbitrary instantaneous mixtures of an arbitrary number of independent sources, $\tilde{\Psi}$ is exactly zero, because mixtures of independent sources do not induce an imaginary part of coherencies (Nolte et al. (2004)) which in turn is necessary to generate a non-vanishing $\tilde{\Psi}$. For finite data, $\tilde{\Psi}$ will then fluctuate in this case around zero within error bounds. A special case of this are phase jumps from 0 to $\pm\pi$ which can arise also for mixtures of independent sources.

2. $\tilde{\Psi}$ is expressed in terms of coherencies, only. The standard deviation of a coherency is approximately constant and approximately only depends on the number of averages which is equal for all frequencies. Thus, large but meaningless phase fluctuations in frequency bands containing essentially independent signals are implicitly suppressed.

3. If the phase $\Phi(f)$ is linear in f and provided that the frequency resolution is sufficient (i.e. δf is sufficiently small), the argument in the sum has the same sign across all frequencies and then $\tilde{\Psi}$ will have the same sign as the slope of $\Phi(f)$.

Finally, as for Granger Causality it is convenient to normalize $\tilde{\Psi}$ by an estimate of its standard deviation

$$\Psi = \frac{\tilde{\Psi}}{std(\tilde{\Psi})} \tag{10}$$

with $std(\tilde{\Psi})$ being estimated by the Jackknife method, which was validated in simulations. In the examples below we consider absolute values of each larger than 2 as significant.

3. Simulations and Causality Challenge

3.1 Uncorrelated noise

Granger Causality is based on the assumption that a sender possesses information about the future of the recipient which is not available at the recipient itself, because, roughly speaking, this information has not yet arrived. In contrast, the recipient cannot access any information about the sender other than that already contained in the present and past of the sender because causal interactions are necessarily forward in time.

However, the situation changes when the measurements, especially of the sender, are noisy. In that case the signal of the recipient contains delayed but cleaner information about the sender which is masked/hidden to the sender itself due to the noise. Thus, the slightly outdated information of the receiver may help to predict the future of the sender and yet lead to wrong results in a Granger test. In other words, the disadvantage of the recipient of receiving only old information might have been compensated or even overcompensated by the advantage of being measured in a much cleaner way.

Apparently, the impact of this trade-off depends on the memory time of the sender: If the sender has a long memory and the transmission time is short then the time delay of the interaction is largely irrelevant.

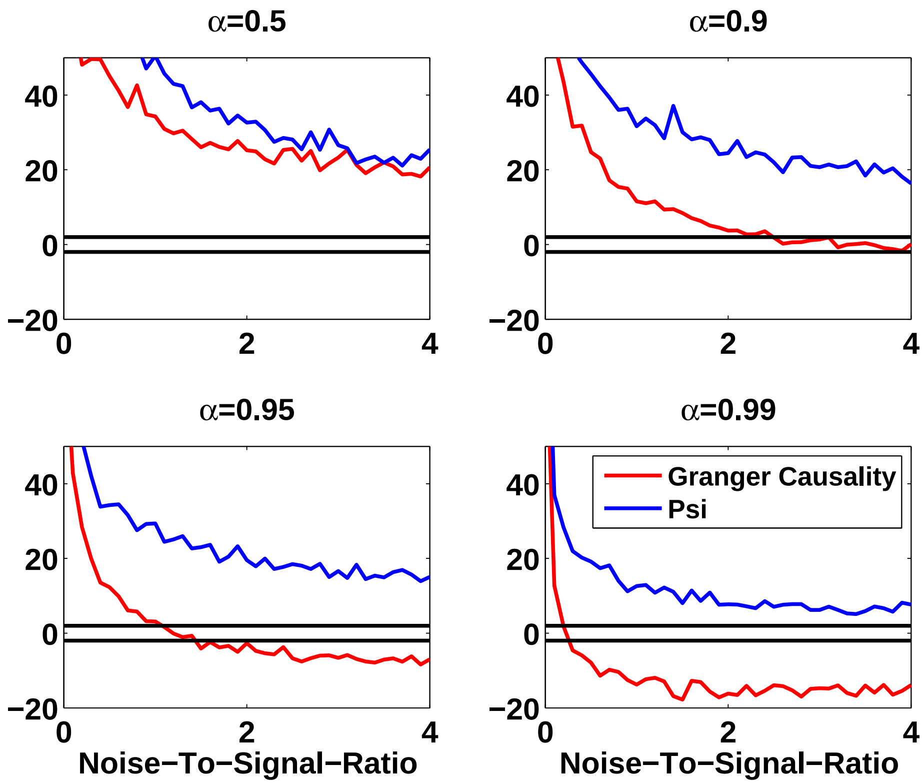

Figure 1: Granger Causality and PSI as a function of noise level for systems with different memory. The memory time in bins is roughly given by $1/(1-\alpha)$. Values outside the narrow horizontal strip are statistically significant.

To show this explicitly we simulated clean data of the sources using an $AR(1)$ model with coefficient matrix

$$A(1) = \begin{pmatrix} \alpha & 0 \\ 1 & .5 \end{pmatrix} \tag{11}$$

This system models a unidirectional information flow from channel 1 to channel 2. The memory of the first channel, the sender, is controlled by α: any input decays after n time points as $\alpha^n = \exp(-n \log(1/\alpha))$ and has hence a decay rate of $-1/\log(\alpha)$.

Let us denote the output of this clean system for the $i.th$ source (i.e. true signal of interest) as $x_i(t)$. Then we assume the measurements $y_i(t)$ to be $y_1(t) = x_1(t) + \beta\eta(t)$ and $y_2(t) = x_2(t)$ with $\eta(t)$ being white gaussian noise and β a free parameter which controlls the relative strength of true signal and noise. Results for these systems are shown in Fig.1 for various values of α. We observe that Granger Causality results in significant wrong direction estimates for long memory times of the sender. In contrast, the Phase Slope Index always results in the correct directionality. We note, that with α also the magnitude of the sender changes which also has an impact on the results. However, normalizing the sender leads to essentially identical results provided that influence of the sender on the recipient is at least

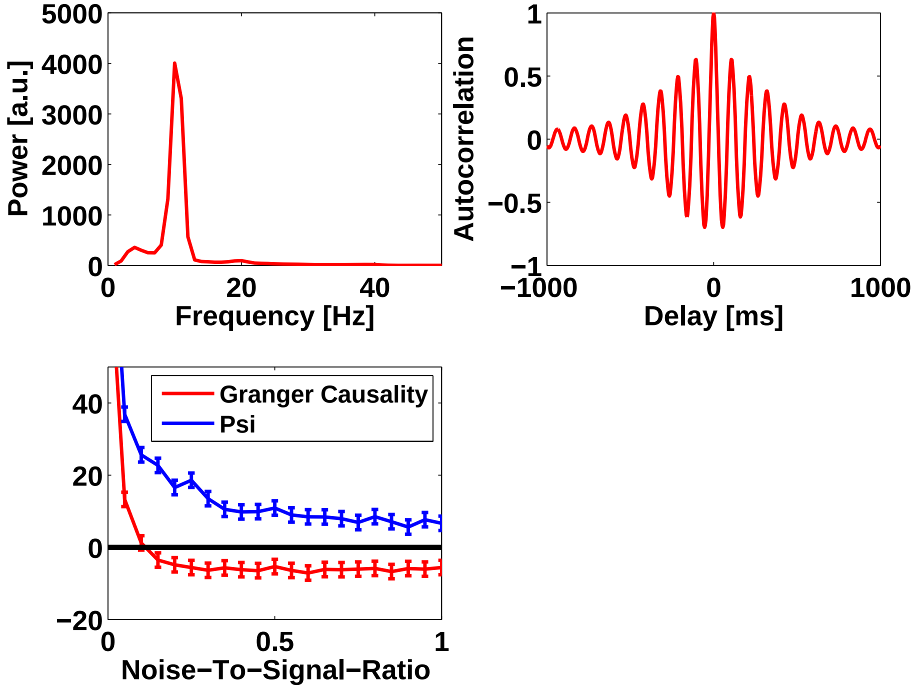

Figure 2: Top panels: Power and autocorrelation of a single channel for a real EEG experiment. Bottom: Granger Causality and PSI as a function of noise level using the real EEG data as driver and adding noise to the sender. The noise level is measured as the power ratio at the peak frequency. The vertical bars correspond to two estimated standard deviations and indicate significance if they do not cross the zero line.

as large as the innovation process of the recipient, i.e. of $\xi_2(t)$. This leads to the somewhat paradoxical situation that for noisy measurements the larger the causal drive from A to B the more likely Granger Causality estimates a drive from B to A.

In a second example, we simulated the situation based on real EEG data. Results for power and autocorrelation function are shown in Fig.2. The memory time of the system is about 0.5 seconds which is large compared to typical transmission times along neuronal fibers. Neuronal signals in axons in white brain matter, which are relevant for long distance information transfer, travel with a speed of about 1cm/msec and need only a few milliseconds to cross the whole brain. The simulation was identical to the previous one with the exception that the real data $x(t)$, normalized to unit standard deviation, were taken as sender and the recipient was assumed to be $y(t) = 2x(t-3) - .5y(t-1) + \eta(t)$ with $\eta(t)$ being white gaussian noise with unit standard deviation. Since the sampling rate was 256 Hz, the delay corresponds to a transmission time of about $12ms$. Results again showed that already a fairly small amount of noise put on the measurement of the sender is sufficient to result in significant false direction estimates of Granger Causality while PSI always predicted the correct direction.

3.2 Nonlinear interactions

To test Granger Causality and PSI for bivariate nonlinear systems we included a nonlinearity of specific order into the interaction term and generated 500 examples as randomly as possible. The data $\mathbf{z}(t)$ were generated as

$$\mathbf{z}(t) = (1-\gamma)\frac{\mathbf{x}(t)}{||X||} + \gamma\frac{B\mathbf{y}(t)}{||BY||} \tag{12}$$

where $\mathbf{x}$ is a unidirectional and in general nonlinear system and $\mathbf{y}$ are two independent noise sources which are mixed into channels by a random matrix B. The parameter γ was set randomly between 0 and 1, and $||\cdot||$ denotes Frobenius matrix norm. The noise $\mathbf{y}(t)$ was generated with an AR(10)-model with diagonal but otherwise random parameters. The signal $\mathbf{x}(t)$ was generated in the following way. If, e.g., the first channel was the sender then $x_1(t)$ was generated with a random AR-model of order 10, and $x_2(t)$ was generated as

$$x_2(t) = \sum_p A_{22}(p)x_2(t-p) + f(x_1(t-1), ..., x_1(t-P)) \tag{13}$$

where P was set to 10 and f is a in general nonlinear function of specific order chosen in the most general way. E.g., for order 4 the function f was given by

$$f(x_1(t-1), ..., x_1(t-P)) = \sum_{ijkl} a_{ijkl}x_1(t-i)x_1(t-j)x_1(t-k)x_1(t-l) \tag{14}$$

with random parameters a_{ijkl}. The construction for other orders is analoguous.

Results for PSI and Granger Causality are shown in Fig.3 and Fig.4, respectively. We observe that PSI, in contrast to Granger Causality, hardly ever results in false significant direction estimates. We also oberve that for even order of nonlinearity PSI is also not able to detect any intercation at all. However, this can be explained by the sign symmetry of the interaction and is due to the linear nature of PSI.

3.3 Causality Challenge

We submitted a dataset to the Causality Challenge[1] which consists of 1000 examples identical to the ones in the previous section for the $order = 1$ case except for two minor details: for the challenge we chose uniformly distributed innovation processes (i.e. $\xi(t)$ in Eq.(1)) instead of gaussian distributed input, and we chose three noise sources instead of two.

The task is to estimate the causal direction for as many examples as possible. The counting is as follows: +1 point for each correct result, -10 points for each wrong results, and 0 points for each missed example. For the top left panel of Fig.3 this means that one gets +1 point for each dot in the lower left or upper right box, -10 points for each point in the lower right or upper left box, and 0 points for each point in the narrow horizontal stripe.

1. "NOISE", http://www.causality.inf.ethz.ch/repository.php?id=17

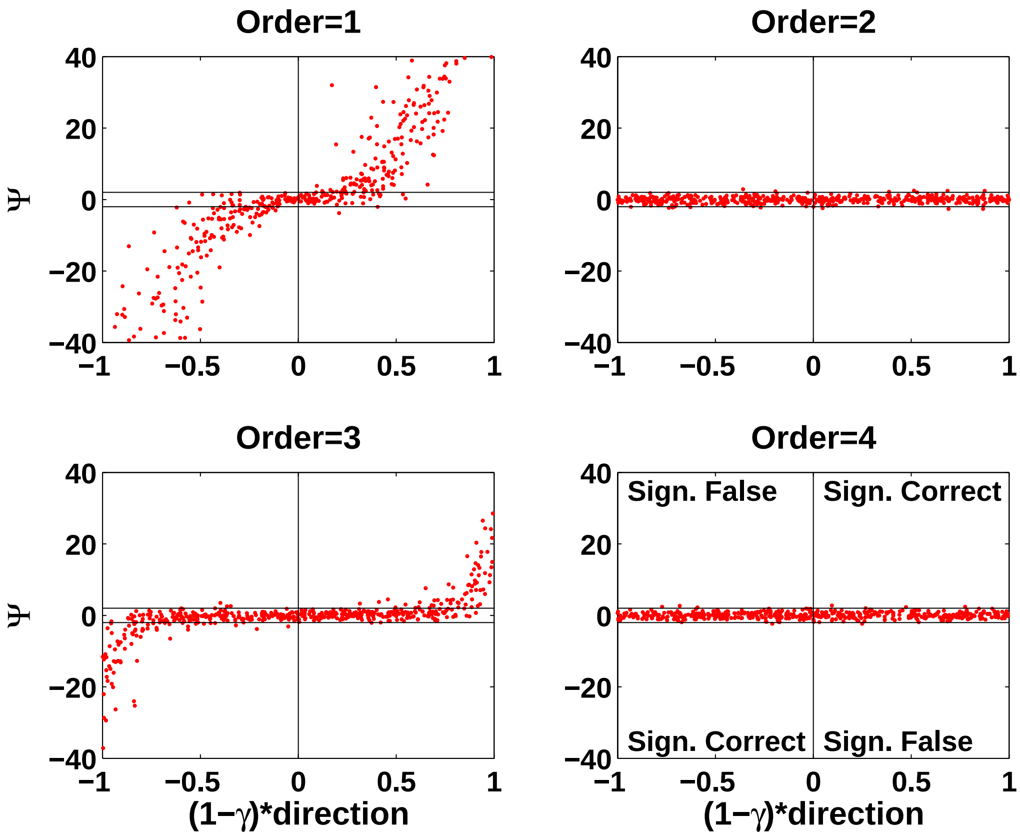

Figure 3: Results for PSI for 500 random systems as a function of noise level times the sign of true direction for different orders of nonlinearity (order=1 is linear). All results in the narrow horizontal strip are insignificant, and the others are as indicated in the lower right panel. For each panel, left and right borders, i.e. $1 - \gamma = 1$, correspond to zero noise and the center, i.e. $\gamma = 1$, corresponds to only noise.

For the challenge data, Granger causality leads to 736 correct and 100 wrong results scoring a total of -264 points. Note, that 164 insignificant results are not counted. In comparison, PSI[2] leads to 638 correct and 6 wrong results scoring a total of +578 points.

This counting was introduced to address the importance of evidence for scientific claims. A finding which was just guessed right has little value. In many cases conclusions cannot be drawn with the given data measured in a specific situation. Researchers must be able to also recognize these cases and should then not draw conclusions at all.

In a second set of data sets we provided real EEG data for 10 subjects measured at rest in eyes closed condition. A specific feature of this measurement is a strong 10Hz rhythm predominantly in the back part of the brain. Using our methods we found information flow

2. The Matlab code can be downloaded at http://ml.cs.tu-berlin.de/causality/

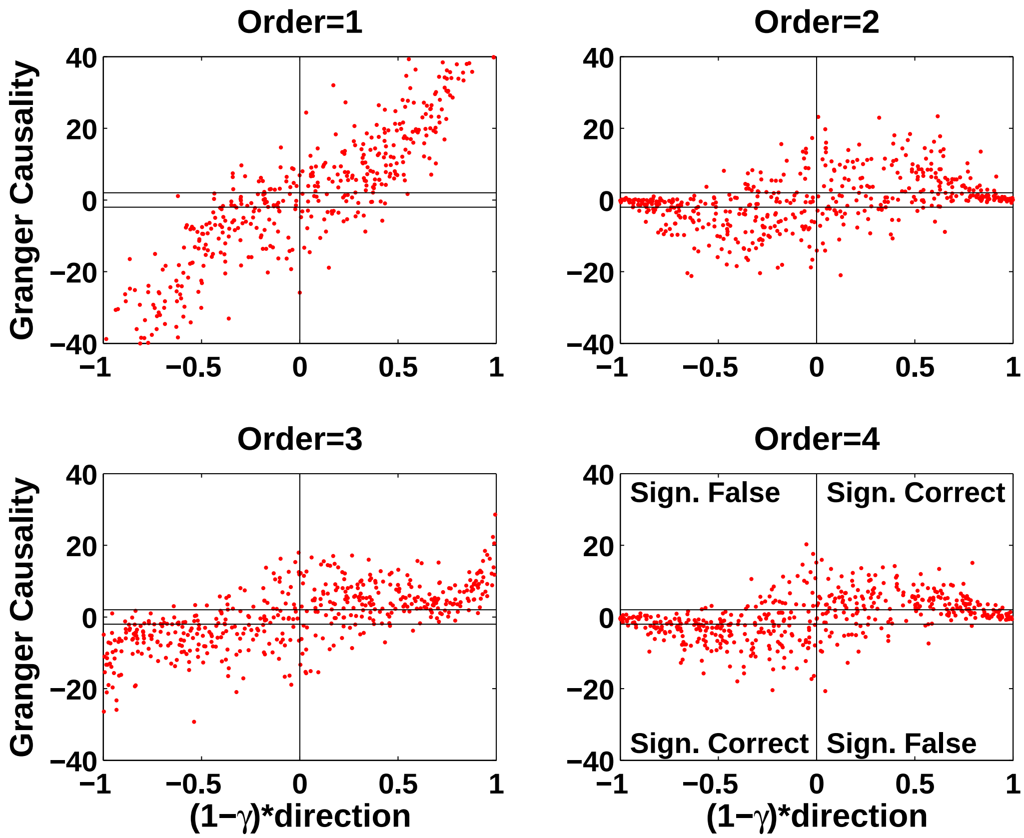

Figure 4: Same as Fig.3 for Granger Causality.

from front to back, i.e. from channels with low signal to ratio to channels with high signal to ratio.

The data used in section 3.1 were taken from one of these subjects. We showed in that section that Granger Causality has a bias to estimate direction from clean to noisy signals, and a finding using Granger Causality stating that information is flowing from back to front is possibly caused by the different signal to noise ratios rather than by true information flow.

For the challenge we can only put this to discussion since the ground truth is not known. We therefore just presented our own results and provided excellent data sets to let people apply their own methods to this case.

4. Conclusion

The paper presents novel insights on causality measures and carefully evaluates their domain of applicability. In particular, we present simulations that contrast Granger Causality and our new Phase Slope Index.

Interestingly, under noise the classical Granger Causality can fail, even to an extent that a wrong causal direction is inferred with a high significance level and even if noise is uncorrelated.

We could show that the PSI approach does not suffer from such a shortcoming including in simulations modeling random and highly nonlinear interactions. Clearly real-world data are always noisy and many complex technical or biological systems contain nonlinear elements. Therefore inference on causal structure in data is required to be robust, a property that is inherent to our proposed new method.

References

Z. Albo, G.V. Di Prisco, Y. Chen, G. Rangarajan, W. Truccolo, J. Feng, R.P. Vertes, and M. Ding. Is partial coherence a viable technique for identifying generators of neural oscillations? *Biol. Cybern.*, 90:318–326, 2004.

L.A. Baccala and K. Sameshima. Directed coherence: a tool for exploring functional interactions among brain structures. *Methods for Simultaneous Neuronal Ensemble Recordings, CRC Press, Boca Raton*, pages 179–192, 1998.

L.A. Baccala and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern.*, 84:463–74, 2001.

C.W.J. Granger. Investigating causal relations by economic models and cross-sprectal methods. *Econometrica*, 37:424–438, 1969.

M. Kaminski and K.J. Blinowska. A new method of the description on information flow. *Biol.Cybern.*, 65:203–210, 1991.

G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett. Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clin. Neurophysiol.*, 115:2292–2307, 2004.

G. Nolte, A. Ziehe, V.V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K.R. Müller. Robustly estimating the flow direction of information in complex physical systems. *Phys Rev Lett*, 100: 234101, 2008.

M.G. Rosenblum and A.S. Pikovsky. Detecting direction of coupling in interacting oscillators. *Phys. Rev. E*, 64:045202, 2001.

T. Schreiber. Measuring information transfer. *Phys. Rev Let.*, 85:461–4, 2000.