

Article

## Simulation Study of Direct Causality Measures in Multivariate Time Series

Angeliki Papana <sup>1,\*</sup>, Catherine Kyrtsov <sup>1,2</sup>, Dimitris Kugiumtzis <sup>3</sup> and Cees Diks <sup>4</sup>

<sup>1</sup> Department of Economics, University of Macedonia, Egnatias 156, 54006, Thessaloniki, Greece; E-Mail: ckyrtsov@uom.gr

<sup>2</sup> University of Strasbourg, BETA, University of Paris 10, Economix, ISC-Paris, Ile-de-France

<sup>3</sup> Faculty of Engineering, Aristotle University of Thessaloniki, University Campus, 54124, Thessaloniki, Greece; E-Mail: dkugiu@gen.auth.gr

<sup>4</sup> Faculty of Economics, Department of Economics and Econometrics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE, Amsterdam, The Netherlands; E-Mail: c.g.h.diks@uva.nl

\* Author to whom correspondence should be addressed; E-Mail: angeliki.papana@gmail.com; Tel: +30-2310891764.

Received: 28 March 2013; in revised form: 5 June 2013 / Accepted: 27 June 2013 /

Published: 4 July 2013

---

**Abstract:** Measures of the direction and strength of the interdependence among time series from multivariate systems are evaluated based on their statistical significance and discrimination ability. The best-known measures estimating direct causal effects, both linear and nonlinear, are considered, *i.e.*, conditional Granger causality index (CGCI), partial Granger causality index (PGCI), partial directed coherence (PDC), partial transfer entropy (PTE), partial symbolic transfer entropy (PSTE) and partial mutual information on mixed embedding (PMIME). The performance of the multivariate coupling measures is assessed on stochastic and chaotic simulated uncoupled and coupled dynamical systems for different settings of embedding dimension and time series length. The CGCI, PGCI and PDC seem to outperform the other causality measures in the case of the linearly coupled systems, while the PGCI is the most effective one when latent and exogenous variables are present. The PMIME outweighs all others in the case of nonlinear simulation systems.

**Keywords:** direct Granger causality; multivariate time series; information measures

**Classification:** PACS 05.45.Tp 05.45.-a 02.70.-c

---

## 1. Introduction

The quantification of the causal effects among simultaneously observed systems from the analysis of time series recordings is essential in many scientific fields, ranging from economics to neurophysiology. Estimating the inter-dependence among the observed variables provides valuable knowledge about the processes that generate the time series. Granger causality has been the leading concept for the identification of directional interactions among variables from their time series, and it has been widely used in economics [1]. However, the last few years, it has become popular also in many different fields, e.g., for the analysis of electroencephalograms.

The mathematical formulation of linear Granger causality is based on linear regression modeling of stochastic processes. Many modifications and extensions of the Granger causality test have been developed; see e.g., [2–7]. Most of the non-causality tests, built on the Granger causality concept and applied in economics, are therefore based on the modeling of the multivariate time series. Despite the success of these strategies, the model-based methods may suffer from the shortcomings of model mis-specification.

The majority of the measures determining the interrelationships among variables that have been developed so far are for bivariate data, e.g., state-space based techniques [8,9], information measures [10–12] and techniques based on the concept of synchronization [13,14].

Bivariate causality tests may erroneously detect couplings when two variables are conditionally independent. To address this, techniques accounting for the effect of the confounding variables have been introduced, termed direct causality measures, which are more appropriate when dealing with multivariate time series [15–17]. Direct causality methods emerged as extensions of bivariate Granger causality. For example, the Granger causality index (GCI), implementing the initial idea for two variables in the time domain, has been extended to the conditional and partial Granger causality index (CGCI and PGCI) [2,18]. Directed coherence (DC) was introduced in the frequency domain, and being a bivariate measure, it cannot discriminate between direct and indirect coupling. The direct transfer function (DTF) is similarly defined as DC [19]. The partial directed coherence (PDC) is an extension of DC to multivariate time series measuring only the direct influences among the variables [20]. Similarly, direct Directed Transfer Function (dDTF) modified DTF to detect only direct information transfer [21].

Information theory sets a natural framework for non-parametric methodologies of several classes of statistical dependencies. Several techniques from information theory have been used in the last few years for the identification of causal relationships in multivariate systems, and the best known is transfer entropy (TE) [11]. Test for causality using the TE has also been suggested [22]. However, the TE is, again, bivariate and its natural extension to account for the presence of confounding variables has been recently introduced, namely, the partial TE (PTE), under different estimating schemes, using bins [23], correlation sums [24] and nearest neighbors [25]. The TE and PTE are actually expressions of conditional mutual information, and with this respect, an improved version of TE making use of a properly restricted non-uniform state space reconstruction was recently developed, termed mutual information on mixed embedding (MIME) [26]. Later, a similar approach to TE/PTE was implemented, which takes into consideration the conditional entropy [27]. Recently MIME was extended for multivariate time series to the partial MIME (PMIME) [28]. Other coupling methods have also

been suggested, such as Renyi's information transfer [29]. In a different approach, the TE has been defined on rank vectors instead of sample vectors, called the symbolic transfer entropy (STE) [30], and, respectively, to the multivariate case termed partial STE (PSTE) (for a correction of STE and PSTE, see, respectively, [31,32]).

Most comparative works on the effectiveness of causality measures concentrate on bivariate tests, e.g., [33–36], while some works evaluating multivariate methodologies include only model-based tests, see, e.g., [37–39], or compare direct and indirect causality measures, e.g., [36,40].

In this work, we compare model-based methods, both in the time and frequency domain, and information theoretic multivariate causality measures that are able to distinguish between direct and indirect causal effects. We include in the study most of the known direct causality measures of these classes, *i.e.*, CGCI and PGCI (linear in time domain), PDC (linear in frequency domain), PTE, PSTE and PMIME (from information theory). The statistical significance of the test statistics is assessed with resampling methods, bootstraps or randomization tests using appropriate surrogates, whenever it is not theoretically known.

The structure of the paper is as follows. The multivariate causality measures considered in this study are presented in Section 2. The statistical significance of the coupling measures is assessed on simulated systems. The simulation systems and the setup of the simulation study are presented in Section 3, while the results of this study and the performance of the causality measures are discussed in Section 4. Finally, the conclusions are drawn in Section 5.

## 2. Direct Causality Measures

Let  $\{x_{1,t}, \dots, x_{K,t}\}$ ,  $t = 1, \dots, n$ , denote a  $K$ -variate time series, consisting of  $K$  simultaneously observed variables,  $X_1, \dots, X_K$ , belonging to a dynamical system or representing respective subsystems of a global system. The reconstructed vectors of each  $X_i$  are formed as  $\mathbf{x}_{i,t} = (x_{i,t}, x_{i,t-\tau_i}, \dots, x_{i,t-(m_i-1)\tau_i})'$ , where  $t = 1, \dots, n'$ ,  $n' = n - \max_i\{(m_i - 1)\tau_i\}$ , and  $m_i$  and  $\tau_i$  are, respectively, the reconstruction parameters of embedding dimension and time delay for  $X_i$ . The notation,  $X_2 \rightarrow X_1$ , denotes the Granger causality from  $X_2$  to  $X_1$ , while  $X_2 \rightarrow X_1|Z$  denotes the direct Granger causality from  $X_2$  to  $X_1$ , accounting for the presence of the other (confounding) variables, *i.e.*,  $Z = \{X_3, \dots, X_K\}$ . The notation of Granger causality for other pairs of variables is analogous.

Almost all of the causality measures require the time series be stationary, *i.e.*, their mean and variance do not change over time. If the time series are non-stationary, then the data should be pre-processed, e.g., for time series that are non-stationary in mean, one can apply the measures on the first or higher order differences. Different transformations are needed in case of non-stationary data in variance or co-integrated time series.

### 2.1. Conditional Granger Causality Index

Granger causality is based on the concept that if the value of a time series,  $X_1$ , to be predicted is improved by using the values of  $X_2$ , then we say that  $X_2$  is driving  $X_1$ . A vector autoregressive model (VAR) in two variables and of order  $P$ , fitted to the time series,  $\{x_{1,t}\}$ , is:

$$x_{1,t+1} = \sum_{j=0}^{P-1} a_{1,j} x_{1,t-j} + \sum_{j=0}^{P-1} b_{1,j} x_{2,t-j} + \epsilon_{1,t+1} \quad (1)$$

where  $a_{1,j}, b_{1,j}$  are the coefficients of the model and  $\epsilon_1$  the residuals from fitting the model with variance  $s_{1U}^2$ . The model in Equation (1) is referred to as the unrestricted model, while the restricted model is obtained by omitting the terms regarding the driving variable [the second sum in Equation (1)] and has residual variance,  $s_{1R}^2$ . According to the concept of Granger causality, the variable,  $X_2$ , Granger causes  $X_1$  if  $s_{1R}^2 > s_{1U}^2$  [1]. The magnitude of the effect of  $X_2$  on  $X_1$  is given by the Granger Causality Index (GCI), defined as:

$$\text{GCI}_{X_2 \rightarrow X_1} = \ln(s_{1R}^2 / s_{1U}^2) \quad (2)$$

Considering all  $K$  variables, the unrestricted model for  $X_1$  is a VAR model in  $K$  variables and involves the  $P$  lags of all  $K$  variables [ $K$  sum terms instead of two in Equation (1)]; the restricted model will have all but the  $P$  lags of the driving variable,  $X_2$ . Likewise, the conditional Granger causality index (CGCI) is:

$$\text{CGCI}_{X_2 \rightarrow X_1 | Z} = \ln(s_{1R}^2 / s_{1U}^2) \quad (3)$$

where  $s_{1U}^2$  and  $s_{1R}^2$  are the residual variances for the unrestricted and restricted model defined for all  $K$  variables.

A parametric significance test for GCI and CGCI can be conducted for the null hypothesis that variable  $X_2$  is not driving  $X_1$ , making use of the  $F$ -significance test for all  $P$  coefficients,  $b_{1,j}$  [41]. When we want to assess collectively the causal effects among all pairs of the  $K$  variables, a correction for multiple testing should be performed, e.g., by means of the false discovery rate [42].

The order,  $P$ , of the VAR model is usually chosen using an information criterion, such as the Akaike Information Criterion (AIC) [43] and the Bayesian Information Criterion (BIC) [44]. The estimation of the coefficients of the VAR models and the residual variances of the models are described analytically in [45].

## 2.2. Partial Granger Causality Index

The partial Granger causality index (PGCI) is associated with the concept of Granger causality and partial correlation [18]. The PGCI addresses the problem of exogenous inputs and latent variables. The intuition is that the influence of exogenous and/or latent variables on a system will be reflected by correlations among the residuals of a VAR model of the measured variables. Thus, in the PGCI, one makes use of the residual covariance matrix of the VAR unrestricted and restricted model, denoted  $\Sigma$  and  $\rho$ , respectively, and not only of the residual variance of the response variable,  $X_1$ ,  $s_{1U}^2$  and  $s_{1R}^2$ , subsequently. For example, for  $X_2 \rightarrow X_1 | X_3$ , denoting the components of  $\Sigma$  as  $\Sigma_{ij}$ ,  $i, j = 1, 2, 3$  and the components of  $\rho$  as  $\rho_{ij}$ ,  $i, j = 1, 2$ , the PGCI is given as:

$$\text{PGCI}_{X_2 \rightarrow X_1 | X_3} = \ln \frac{\rho_{11} - \rho_{12} \rho_{22}^{-1} \rho_{21}}{\Sigma_{11} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{31}} \quad (4)$$

Note that  $\Sigma_{11} = s_{1U}^2$  and  $\rho_{11} = s_{1R}^2$ . The PGCI constitutes an improved estimation of the direct Granger causality as compared to the CGCI when the residuals of the VAR models are correlated; otherwise, it is identical to the CGCI. The estimation procedure for the PGCI is described analytically in [18].

### 2.3. Partial Directed Coherence

The partial directed coherence (PDC) is related to the same VAR model as the CGCI, but is defined in the frequency domain [20]. Denoting the  $K \times K$  matrix of the Fourier transform of the coefficients of the VAR model in  $K$  variables and order  $P$  by  $A(f)$ , the PDC from  $X_2$  to  $X_1$  at a frequency  $f$  is given by [20]:

$$\text{PDC}_{X_2 \rightarrow X_1|Z}(f) = \frac{|A_{1,2}(f)|}{\sqrt{\sum_{k=1}^K |A_{k,2}(f)|^2}} \quad (5)$$

where  $A_{i,j}(f)$  is the component at the position,  $(i, j)$ , in the matrix,  $A(f)$ .  $\text{PDC}_{X_2 \rightarrow X_1|Z}(f)$  provides a measure for the directed linear influence of  $X_2$  on  $X_1$  at frequency,  $f$ , conditioned on the other  $K - 2$  variables in  $Z$  and takes values in the interval,  $[0, 1]$ . The  $\text{PDC}_{X_2 \rightarrow X_1}(f)$  is computed at each frequency,  $f$ , within an appropriate range of frequencies. Parametric inference and significance tests for PDC have been studied in [46,47].

### 2.4. Partial Transfer Entropy

The transfer entropy (TE) is a nonlinear measure that quantifies the amount of information explained in  $X_1$  at  $h$  steps ahead from the state of  $X_2$ , accounting for the concurrent state of  $X_1$  [11]. The TE is given here in terms of entropies. For a discrete variable,  $X$  (scalar or vector), the Shannon entropy is  $H(X) = -\sum p(x_i) \ln p(x_i)$ , where  $p(x_i)$  is the probability mass function of variable,  $X$ , at the value,  $x_i$ . Further, the TE is expressed as:

$$\begin{aligned} \text{TE}_{X_2 \rightarrow X_1} &= I(x_{1,t+h}; \mathbf{x}_{2,t} | \mathbf{x}_{1,t}) = H(x_{1,t+h} | \mathbf{x}_{1,t}) - H(x_{1,t+h} | \mathbf{x}_{2,t}, \mathbf{x}_{1,t}) \\ &= H(\mathbf{x}_{2,t}, \mathbf{x}_{1,t}) - H(x_{1,t+h}, \mathbf{x}_{2,t}, \mathbf{x}_{1,t}) + H(x_{1,t+h}, \mathbf{x}_{1,t}) - H(\mathbf{x}_{1,t}) \end{aligned} \quad (6)$$

The first equality is inserted to show that the TE is equivalent to the conditional mutual information (CMI), where  $I(X, Y) = H(X) + H(Y) - H(X, Y)$  is the mutual information (MI) of two variables,  $X$  and  $Y$ . The time horizon,  $h$ , is introduced here instead of the single time step, originally used in the definition of TE.

The partial transfer entropy (PTE) is the extension of the TE designed for the direct causality of  $X_2$  to  $X_1$  conditioning on the remaining variables in  $Z$

$$\text{PTE}_{X_2 \rightarrow X_1|Z} = H(x_{1,t+h} | \mathbf{x}_{1,t}, \mathbf{z}_t) - H(x_{1,t+h} | \mathbf{x}_{2,t}, \mathbf{x}_{1,t}, \mathbf{z}_t) \quad (7)$$

The entropy terms of PTE are estimated here using the  $k$ -nearest neighbors method [48].

### 2.5. Symbolic Transfer Entropy

The symbolic transfer entropy (STE) is the continuation of the TE estimated on rank-points formed by the reconstructed vectors of the variables [30]. For each vector,  $\mathbf{x}_{2,t}$ , the ranks of its components

assign a rank-point,  $\hat{\mathbf{x}}_{2,t} = [r_1, r_2, \dots, r_{m_2}]$ , where  $r_j \in \{1, 2, \dots, m_2\}$  for  $j = 1, \dots, m_2$ . Following this sample-point to rank-point conversion, the sample,  $x_{1,t+h}$ , in Equation (7) is taken as the rank point at time,  $t + h$ ,  $\hat{\mathbf{x}}_{1,t+h}$ , and STE is defined as:

$$\text{STE}_{X_2 \rightarrow X_1} = H(\hat{\mathbf{x}}_{1,t+h} | \hat{\mathbf{x}}_{1,t}) - H(\hat{\mathbf{x}}_{1,t+h} | \hat{\mathbf{x}}_{2,t}, \hat{\mathbf{x}}_{1,t}) \quad (8)$$

where the entropies are computed based on the rank-points.

In complete analogy to the derivation of the PTE from the TE, the partial symbolic transfer entropy (PSTE) extends the STE for multivariate time series and is expressed as:

$$\text{PSTE}_{X_2 \rightarrow X_1 | Z} = H(\hat{\mathbf{x}}_{1,t+h} | \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t) - H(\hat{\mathbf{x}}_{1,t+h} | \hat{\mathbf{x}}_{2,t}, \hat{\mathbf{x}}_{1,t}, \hat{\mathbf{z}}_t) \quad (9)$$

where the rank vector,  $\hat{\mathbf{z}}_t$ , is the concatenation of the rank vectors for each of the embedding vectors of the variables in  $Z$ .

## 2.6. Partial Mutual Information on Mixed Embedding

The mutual information on mixed embedding (MIME) is derived directly from a mixed embedding scheme based on the conditional mutual information (CMI) criterion [26]. In the bivariate case and for the driving of  $X_2$  on  $X_1$ , the scheme gives a mixed embedding of varying delays from the variables,  $X_1$  and  $X_2$ , that explains best the future of  $X_1$ , defined as  $\mathbf{x}_{1,t}^h = [x_{1,t+1}, \dots, x_{1,t+h}]$ . The mixed embedding vector,  $\mathbf{w}_t$ , may contain lagged components of  $X_1$ , forming the subset,  $\mathbf{w}_t^{X_1}$ , and of  $X_2$ , forming  $\mathbf{w}_t^{X_2}$ , where  $\mathbf{w}_t = [\mathbf{w}_t^{X_1}, \mathbf{w}_t^{X_2}]$ . The MIME is then estimated as:

$$\text{MIME}_{X_2 \rightarrow X_1} = \frac{I(\mathbf{x}_{1,t}^h; \mathbf{w}_t^{X_2} | \mathbf{w}_t^{X_1})}{I(\mathbf{x}_{1,t}^h; \mathbf{w}_t)} \quad (10)$$

The numerator in Equation (10) is the CMI as for the TE in Equation (7), but for non-uniform embedding vectors of  $X_1$  and  $X_2$ . Therefore, the MIME can be considered as a normalized version of the TE for optimized non-uniform embedding of  $X_1$  and  $X_2$  [26].

For multivariate time series, the partial mutual information on mixed embedding (PMIME) has been developed in analogy to the MIME [28]. The mixed embedding vector that best describes the future of  $X_1$ ,  $\mathbf{x}_{1,t}^h$ , is now formed potentially by all  $K$  lagged variables, *i.e.*,  $X_1$ ,  $X_2$  and the other  $K - 2$  variables in  $Z$ , and it can be decomposed to the three respective subsets as  $\mathbf{w}_t = (\mathbf{w}_t^{X_1}, \mathbf{w}_t^{X_2}, \mathbf{w}_t^Z)$ . The PMIME is then estimated as:

$$\text{PMIME}_{X_2 \rightarrow X_1 | Z} = \frac{I(\mathbf{x}_{1,t}^h; \mathbf{w}_t^{X_2} | \mathbf{w}_t^{X_1}, \mathbf{w}_t^Z)}{I(\mathbf{x}_{1,t}^h; \mathbf{w}_t)} \quad (11)$$

Similarly to the MIME, the PMIME can be considered as a normalized version of the PTE for optimized non-uniform embedding of all  $K$  variables. Thus, the PMIME takes values between zero and one, where zero indicates the absence of components of  $X_2$  in the mixed embedding vector and, consequently, no direct Granger causality from  $X_2$  to  $X_1$ .

A maximum lag to search for components in the mixed embedding vector is set for each variable, here being the same maximum lag,  $L_{max}$ , for all variables.  $L_{max}$  can be set equal to a sufficiently large number without affecting the performance of the measure; however, the larger it is, the higher the computational cost is. For the estimation of the MI and the CMI, the  $k$ -nearest neighbors method is used [48].

### 3. Simulation Study

The multivariate causality measures are evaluated in a simulation study. All the considered direct coupling measures are computed on 100 realizations of multivariate uncoupled and coupled systems, for increasing coupling strengths and for all directions. The simulation systems that have been used in this study are the following.

- System 1: A vector autoregressive process of order one [VAR(1)] in three variables with  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$

$$\begin{aligned}x_{1,t} &= \theta_t \\x_{2,t} &= x_{1,t-1} + \eta_t \\x_{3,t} &= 0.5x_{3,t-1} + x_{2,t-1} + \epsilon_t\end{aligned}$$

where  $\theta_t$ ,  $\eta_t$  and  $\epsilon_t$  are independent to each other Gaussian white noise processes, with standard deviations one, 0.2 and 0.3, respectively.

- System 2: A VAR(5) process in four variables with  $X_1 \rightarrow X_3$ ,  $X_2 \rightarrow X_1$ ,  $X_2 \rightarrow X_3$  and  $X_4 \rightarrow X_2$  (Equation 12 in [49])

$$\begin{aligned}x_{1,t} &= 0.8x_{1,t-1} + 0.65x_{2,t-4} + \epsilon_{1,t} \\x_{2,t} &= 0.6x_{2,t-1} + 0.6x_{4,t-5} + \epsilon_{2,t} \\x_{3,t} &= 0.5x_{3,t-3} - 0.6x_{1,t-1} + 0.4x_{2,t-4} + \epsilon_{3,t} \\x_{4,t} &= 1.2x_{4,t-1} - 0.7x_{4,t-2} + \epsilon_{4,t}\end{aligned}$$

where  $\epsilon_{i,t}$ ,  $i = 1, \dots, 4$  are independent to each other Gaussian white noise processes with unit standard deviation.

- System 3: A VAR(4) process in five variables with  $X_1 \rightarrow X_2$ ,  $X_1 \rightarrow X_4$ ,  $X_2 \rightarrow X_4$ ,  $X_4 \rightarrow X_5$ ,  $X_5 \rightarrow X_1$ ,  $X_5 \rightarrow X_2$  and  $X_5 \rightarrow X_3$  [46]

$$\begin{aligned}x_{1,t} &= 0.4x_{1,t-1} - 0.5x_{1,t-2} + 0.4x_{5,t-1} + \epsilon_{1,t} \\x_{2,t} &= 0.4x_{2,t-1} - 0.3x_{1,t-4} + 0.4x_{5,t-2} + \epsilon_{2,t} \\x_{3,t} &= 0.5x_{3,t-1} - 0.7x_{3,t-2} - 0.3x_{5,t-3} + \epsilon_{3,t} \\x_{4,t} &= 0.8x_{4,t-3} + 0.4x_{1,t-2} + 0.3x_{2,t-3} + \epsilon_{4,t} \\x_{5,t} &= 0.7x_{5,t-1} - 0.5x_{5,t-2} - 0.4x_{4,t-1} + \epsilon_{5,t}\end{aligned}$$

and  $\epsilon_{i,t}$ ,  $i = 1, \dots, 5$ , as above.

- System 4: A coupled system of three variables with linear ( $X_2 \rightarrow X_3$ ) and nonlinear causal effects ( $X_1 \rightarrow X_2$  and  $X_1 \rightarrow X_3$ ) (Model 7 in [50])

$$\begin{aligned}x_{1,t} &= 3.4x_{1,t-1}(1 - x_{1,t-1})^2 \exp(-x_{1,t-1}^2) + 0.4\epsilon_{1,t} \\x_{2,t} &= 3.4x_{2,t-1}(1 - x_{2,t-1})^2 \exp(-x_{2,t-1}^2) + 0.5x_{1,t-1}x_{2,t-1} + 0.4\epsilon_{2,t} \\x_{3,t} &= 3.4x_{3,t-1}(1 - x_{3,t-1})^2 \exp(-x_{3,t-1}^2) + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4\epsilon_{3,t}\end{aligned}$$

and  $\epsilon_{i,t}$ ,  $i = 1, \dots, 3$ , as above.



- System 5: Three coupled Hénon maps with nonlinear couplings,  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$

$$\begin{aligned}x_{1,t} &= 1.4 - x_{1,t-1}^2 + 0.3x_{1,t-2} \\x_{2,t} &= 1.4 - cx_{1,t-1}x_{2,t-1} - (1-c)x_{2,t-1}^2 + 0.3x_{2,t-2} \\x_{3,t} &= 1.4 - cx_{2,t-1}x_{3,t-1} - (1-c)x_{3,t-1}^2 + 0.3x_{3,t-2}\end{aligned}$$

with equal coupling strengths,  $c$ , and  $c = 0, 0.05, 0.3, 0.5$ .

The time series of this system become completely synchronized for coupling strengths,  $c \geq 0.7$ . In order to investigate the effect of noise on the causality measures, we also consider the case of addition of Gaussian white noise to each variable of System 5, with standard deviation 0.2 times their standard deviation.

- System 6: Three coupled Lorenz systems with nonlinear couplings,  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$

$$\begin{aligned}\dot{x}_1 &= 10(y_1 - x_1) & \dot{x}_2 &= 10(y_2 - x_2) + c(x_1 - x_2) & \dot{x}_3 &= 10(y_3 - x_3) + c(x_2 - x_3) \\ \dot{y}_1 &= 28x_1 - y_1 - x_1z_1, & \dot{y}_2 &= 28x_2 - y_2 - x_2z_2, & \dot{y}_3 &= 28x_3 - y_3 - x_3z_3 \\ \dot{z}_1 &= x_1y_1 - 8/3z_1 & \dot{z}_2 &= x_2y_2 - 8/3z_2 & \dot{z}_3 &= x_3y_3 - 8/3z_3\end{aligned}$$

The first variables of the three interacting systems are observed at a sampling time of 0.05 units. The couplings,  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$ , have the same strength,  $c$ , and  $c = 0, 1, 3, 5$ . The time series of the system become completely synchronized for coupling strengths,  $c \geq 8$ . For a more detailed description of the synchronization of the coupled Systems 5 and 6, see [51].

- System 7: A linear coupled system in five variables with  $X_1 \rightarrow X_2$ ,  $X_1 \rightarrow X_3$ ,  $X_1 \rightarrow X_4$ ,  $X_4 \leftrightarrow X_5$  with latent endogenous variables [18]

$$\begin{aligned}x_{1,t} &= 0.95\sqrt{2}x_{1,t-1} - 0.9025x_{1,t-2} + \epsilon_{1,t} + a_1\epsilon_{6,t} + b_1\epsilon_{7,t-1} + c_1\epsilon_{7,t-2} \\x_{2,t} &= 0.5x_{1,t-2} + \epsilon_{2,t} + a_2\epsilon_{6,t} + b_2\epsilon_{7,t-1} + c_2\epsilon_{7,t-2} \\x_{3,t} &= -0.4x_{1,t-3} + \epsilon_{3,t} + a_3\epsilon_{6,t} + b_3\epsilon_{7,t-1} + c_3\epsilon_{7,t-2} \\x_{4,t} &= -0.5x_{1,t-2} + 0.25\sqrt{2}x_{4,t-1} + 0.25\sqrt{2}x_{5,t-1} + \epsilon_{4,t} + a_4\epsilon_{6,t} + b_4\epsilon_{7,t-1} + c_4\epsilon_{7,t-2} \\x_{5,t} &= -0.25\sqrt{2}x_{4,t-1} + 0.25\sqrt{2}x_{5,t-1} + \epsilon_{5,t} + a_5\epsilon_{6,t} + b_5\epsilon_{7,t-1} + c_5\epsilon_{7,t-2}\end{aligned}$$

where  $\epsilon_{i,t}$  are zero mean uncorrelated processes with variances 0.8, 0.6, 1, 1.2, 1, 0.9, 1, respectively,  $a_1 = 5$ ,  $a_2 = a_3 = a_4 = a_5 = 1$  and  $b_i = 2$ ,  $c_i = 5$ ,  $i = 1, \dots, 5$ .

The time series lengths considered in the simulation study are  $n = 512$  and  $n = 2048$ . Regarding the CGCI, the PGCI and the PDC, the order,  $P$ , of the VAR model is selected by combining the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), as well as our knowledge for the degrees of freedom of each coupled system, as follows. The range of model orders, for which the AIC and the BIC are calculated, is selected to be at the level of the ‘true’ model order based on the equations of each system. Specifically, for Systems 1, 4, 5, 6 and 7, we considered the range of model orders,  $[1, 5]$ , for the calculation of AIC and BIC, and for Systems 2 and 3, we considered the range,  $[1, 10]$ . Further, we estimate the PDC for a range of frequencies determined by the power spectrum of the variables of each system. We specify this range by selecting those frequencies that display the highest values in the



auto-spectra of the variables [52]. The  $p$ -values from a non-parametric test for the PDC are estimated for the selected range of frequencies (using bootstraps [53]), and in order to decide whether a coupling is significant, at least 80% of the  $p$ -values from this range of frequencies should be significant.

The embedding dimension,  $m$ , for the PTE and the PSTE and the maximum lag,  $L_{max}$ , for the PMIME are set equal to  $P$  and  $\tau = 1$  for the PTE and the PSTE. Note that this choice of  $L_{max}$  may be very restrictive, and the PMIME may not be optimal; but, we adopt it here to make the choice for VAR order and embedding uniform. The time step ahead,  $h$ , for the estimation of the PTE, PSTE and PMIME is set to one for the first five systems and System 7 (the common choice for discrete-time systems), while for the continuous-time system (System 6),  $h$  is set to be equal to  $m$ . The number of nearest neighbors for the estimation of the PTE and the PMIME is set to  $k = 10$ . We note that the  $k$ -nearest neighbors methods for the estimation of the measures is found to be stable and not significantly affected by the choice of  $k$  [48]. The threshold for the stopping criterion for the mixed embedding scheme for the PMIME is set to  $A = 0.95$  (for details, see [26]).

### 3.1. Statistical Significance of the Causality Measures

In the simulation study, we assess the statistical significance of the causality measures by means of parametric tests, when applicable, and nonparametric (resampling) tests, otherwise, in the way these have been suggested in the literature for each measure. The correction for multiple testing regarding the significance of a measure on all possible variable pairs is not considered here, as the interest is in comparing the performance of the different direct causality measures rather than providing rigorous statistical evidence for the significance of each coupling.

Starting with the CGCI, it bears a parametric significance test, and this is the  $F$ -test for the null hypothesis that the coefficients of the lagged driving variables in the unrestricted VAR model are all zero [41]. If  $P_1$  and  $P_2$  are the numbers of variables in the restricted and the unrestricted autoregressive model, respectively, ( $P_2 > P_1$ ), and  $n$  is the length of the time series, then the test statistic is  $F = ((RSS_1 - RSS_2)/(P_1 - P_2))/(RSS_2/(n - P_2))$ , where  $RSS_i$  is the residual sum of squares of model,  $i$ . Under the null hypothesis that the unrestricted model does not provide a significantly better fit than the restricted model, the  $F$ -statistic follows the Fisher-Snedecor, or  $F$ , distribution with  $(P_2 - P_1, n - P_2)$  degrees of freedom. The null hypothesis is rejected if the  $F$ -statistic calculated on the data is greater than the critical value of the  $F$ -distribution for some desired false-rejection probability (here  $\alpha = 0.05$ ).

The statistical significance of the PGCI is assessed by means of confidence intervals formed by bootstrapping [53], since the null distribution is unknown. The empirical distribution of any statistic using bootstrapping is formed from the values of the statistic computed on a number of new samples obtained by random sampling with replacement from the observed data. In the context of vector autoregressive models, this can be realized by subdividing the data matrix (of the predictor and response jointly) into a number of windows, which are repeatedly sampled with replacement to generate bootstrap data matrices. By this procedure, the causal relationships within each window are not affected. The PGCI is computed for each bootstrapped data matrix. The confidence interval of the PGCI is formed by the lower and upper empirical quantiles of the bootstrap distribution of the PGCI for the significance

level,  $\alpha = 0.05$ . The bootstrap confidence interval for the PGCI can be considered as a significance test, where the test decision depends on whether zero is included in the confidence interval. The details for the estimation of the bootstrap confidence intervals of the PGCI can be found in [18].

The statistical significance of the PDC can be determined using both parametric testing [46,47], and randomization (surrogate) testing [54]. Here, we choose the parametric approach. The statistical significance of a nonzero value,  $\text{PDC}_{X_2 \rightarrow X_1}(f)$ , is investigated by means of a critical value,  $c_{PDC}$ . Under the null hypothesis that there is no Granger causality,  $X_2 \rightarrow X_1$ , it holds  $|A_{12}(f)| = 0$ , and  $c_{PDC}$  can be derived from theoretical considerations for each frequency,  $f$ , at a given  $\alpha$ -significance level by:

$$c_{PDC}(f) = ((\hat{C}_{ij}(f)\chi_{1,1-\alpha}^2)/(N \sum_k |\hat{A}_{kj}(f)|^2))^{1/2}$$

The term,  $\chi_{1,1-\alpha}^2$ , denotes the  $(1 - \alpha)$ -quantile of the  $\chi^2$  distribution with one degree of freedom, and  $\hat{C}_{ij}(f)$  is an estimate of the expression:

$$C_{ij}(f) = \Sigma_{ii}^{-1} \left( \sum_{k,l=1}^P \Sigma_{jj}^{-1} [\cos(kf) \cos(lf) + \sin(kf) \sin(lf)] \right)$$

where  $\Sigma_{jj}^{-1}$  denotes the entries of the inverse of the covariance matrix,  $\Sigma$ , of the VAR process [47].

The statistical significance of the PTE and the PSTE is evaluated assuming a randomization test with appropriate surrogate time series, as their null distribution is not known (for the PSTE, in [32], analytic approximations were built, but found to be inferior to approximations using surrogates). We create  $M$  surrogate time series consistent with the non-causality null hypotheses,  $H_0$ , i.e.,  $X_2$  does not Granger causes  $X_1$ . To destroy any causal effect of  $X_2$  on  $X_1$  without changing the dynamics in each time series, we randomly choose a number,  $d$ , less than the time series length,  $n$ , and the  $d$ -first values of the time series of  $X_2$  are moved to the end, while the other series remain unchanged. The random number,  $d$ , for the time-shifted surrogates is an integer within the range,  $[0.05n, 0.95n]$ , where  $n$  is the time series length. This scheme for generating surrogate time series is termed time-shifted surrogates [55]. We estimate the causality measure (PTE or PSTE) from the original multivariate time series, let us denote it  $q_0$ , and for each of the  $M$  multivariate surrogate time series, let us denote them  $q_1, q_2, \dots, q_M$ . If  $q_0$  is at the tail of the empirical null distribution formed by  $q_1, q_2, \dots, q_M$ , then  $H_0$  is rejected. For the two-sided test, if  $r_0$  is the rank of  $q_0$  in the ordered list of  $q_0, q_1, \dots, q_M$ , the  $p$ -value for the test is  $2(r_0 - 0.326)/(M + 1 + 0.348)$  if  $r_0 < (M + 1)/2$  and  $2[1 - (r_0 - 0.326)/(M + 1 + 0.348)]$  if  $r_0 \geq (M + 1)/2$ , by applying the correction for the empirical cumulative density function in [56].

Finally, PMIME does not rely on any significance test, as it gives zero values in the uncoupled case and positive values, otherwise. This was confirmed using time-shifted surrogates also for the PMIME in the simulation study, and the PMIME values of the surrogate time series were all zero.

For the estimation of CGCI and PGCI and their statistical significance, we used the 'Causal Connectivity Analysis' toolbox [57]. The programs for the computations of the remaining causality measures have been implemented in Matlab.

#### 4. Evaluation of Causality Measures

In order to evaluate the multivariate causality measures, the percentage of rejection of the null hypothesis of no causal effects ( $H_0$ ) in 100 realizations of the system is calculated for each possible pair

of variables and for different time series lengths and free parameters of the measures. The focus when presenting the results is on the sensitivity of the measure or, respectively, the power of the significance test (the percentage of rejection at the significance level 5% or  $\alpha = 0.05$  when there is true direct causality), as well as the specificity of the measure or size of the test (the percentage of rejection at  $\alpha = 0.05$  when there is no direct causality) and how these properties depend on the time series length and the measure-specific parameter.

#### 4.1. Results for System 1

For the estimation of the linear measures, the order of the model,  $P$ , is set to one, as indicated from the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC), while for the estimation of PTE,  $m$  is also set to one. The PDC is estimated for the range of frequencies  $[0, 0.5]$ , since the auto-spectra of the variables do not suggest a narrower range. Indeed, the  $p$ -values of PDC are all significant in  $[0, 0.5]$ , when there is direct causality, and not significant, when there is no direct causality. The CGCI, PGCI, PDC and PTE correctly detect the direct causal effect for both time series lengths,  $n = 512$  and 2048. All the aforementioned measures indicate 100% rejection of  $H_0$  for the true couplings,  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$ , and low percentages for all other couplings. Their performance is not affected by the time series length, for the time series lengths considered. The estimated percentages are displayed for both  $n$  in Table 1.

**Table 1.** Percentage of statistically significant values of the causality measures for System 1,  $P = m = L_{max} = 1$  [ $m = 2$  for partial symbolic transfer entropy (PSTE)]. The directions of direct causal effects are pointed out in bold. When the same percentage has been found for both  $n$ , a single number is displayed in the cell.

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
<b><math>X_1 \rightarrow X_2</math></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>1 / 0</b>
$X_2 \rightarrow X_1$	4 / 3	2 / 1	2 / 3	8 / 5	58 / 100	2 / 7
<b><math>X_2 \rightarrow X_3</math></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
$X_3 \rightarrow X_2$	7 / 6	8 / 1	3	3 / 5	7 / 25	0
$X_1 \rightarrow X_3$	3 / 5	0 / 2	2 / 3	5 / 7	93 / 100	0
$X_3 \rightarrow X_1$	2 / 7	0	3	3 / 2	14 / 43	7 / 7

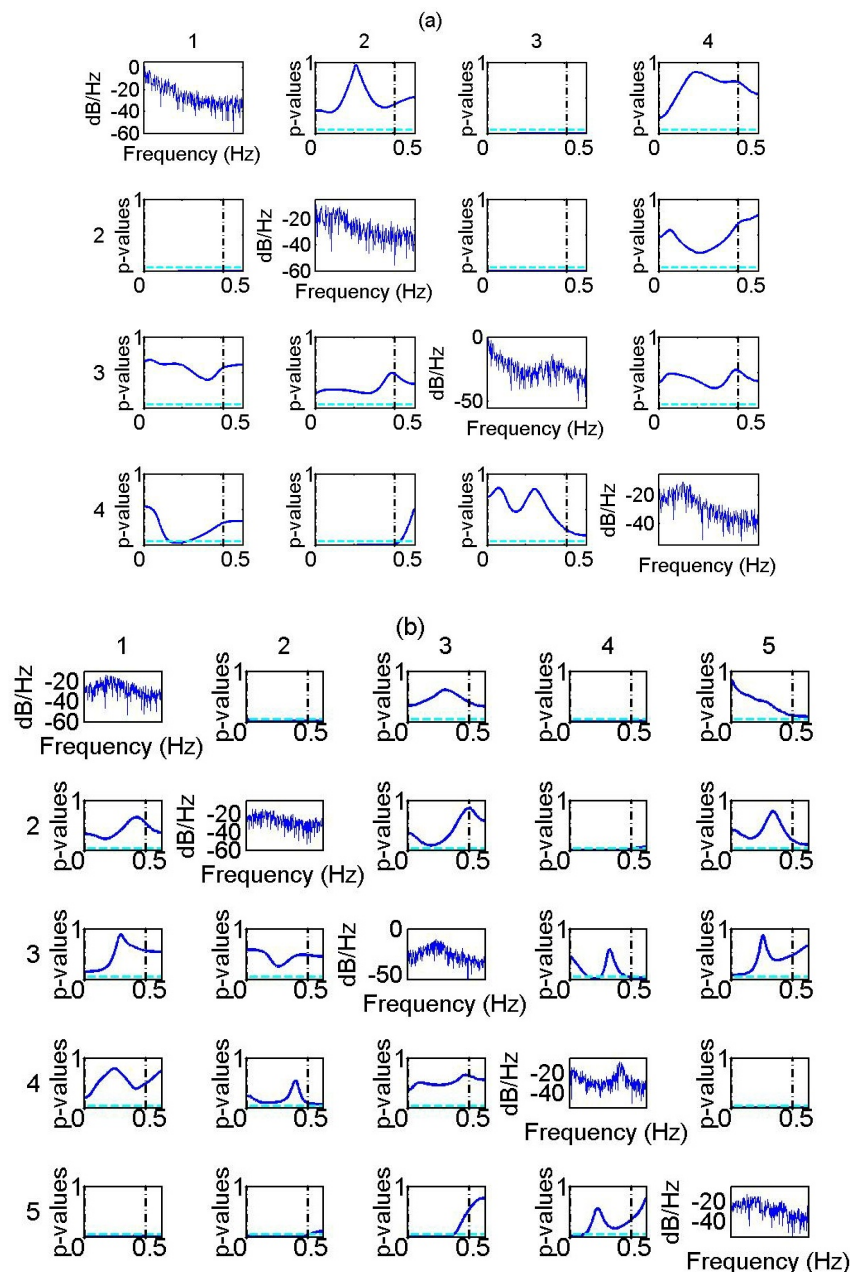
The PSTE can only be estimated for  $m \geq 2$ , and therefore, results are obtained for  $m = 2$ . The PSTE correctly detects the direct causalities for  $m = 2$ ; however, it also indicates the indirect effect,  $X_1 \rightarrow X_3$ , and the spurious causal effect,  $X_2 \rightarrow X_1$ .

Only for this system, the PMIME with the threshold of  $A = 0.95$  failed to detect the true direct effects, and the randomization test gave partial improvement (detection of one of the two true direct effects,  $X_2 \rightarrow X_3$ ). This is merely a problem of using the fixed threshold,  $A = 0.95$ , in this system, and following the adapted threshold proposed in [28], the two true direct effects could be detected for all realizations with  $n = 512$  and  $n = 2048$  with the largest rate of false rejection being 8%.

#### 4.2. Results for System 2

For the second simulation system, the model order is set to  $P = 5$ , as indicated both by BIC and AIC. The embedding dimension,  $m$ , and  $L_{max}$  are also set to five. The PDC is estimated for the range of frequencies,  $[0, 0.4]$ , since the auto-spectra of all the variables are higher in the range,  $[0, 0.2]$ , while variable,  $X_3$ , exhibits a peak in the range,  $[0.2, 0.4]$ . Indicatively, the  $p$ -values from one realization of the system for the range of frequencies,  $[0, 0.5]$ , is displayed in Figure 1a.

**Figure 1.** Graph summarizing the causal influences for one realization of System 2 (rows  $\rightarrow$  columns) in (a) and System 3 in (b). The  $p$ -values from the partial directed coherence (PDC) are displayed for the range of frequencies,  $[0, 0.5]$ , while the dotted vertical lines indicate the frequency, 0.4. The horizontal cyan lines indicate the 5%-significance level. The auto-spectra are shown on the diagonal.



The CGCI, PGCI, PDC and PMIME correctly detect the direct couplings ( $X_2 \rightarrow X_1$ ,  $X_1 \rightarrow X_3$ ,  $X_2 \rightarrow X_3$ ,  $X_4 \rightarrow X_2$ ), as shown in Table 2. The performance of the CGCI and PDC is not affected by the time series length. The PGCI is also not affected by  $n$ , except for the causal effect,  $X_2 \rightarrow X_4$ , where the PGCI falsely indicates causality for  $n = 2048$  (20%). The PMIME indicates lower power of the test compared to the linear measures only for  $X_2 \rightarrow X_3$  and  $n = 512$ .

**Table 2.** Percentage of statistically significant values of the causality measures for System 2,  $P = m = L_{max} = 5$ . The directions of direct causal effects are displayed in bold face. CGCI, conditional Granger causality index; PGCI, partial Granger causality index; PDC, partial directed coherence; PTE, partial transfer entropy; PSTE, partial symbolic transfer entropy; PMIME, partial mutual information on mixed embedding.

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$X_1 \rightarrow X_2$	6 / 2	1	0	6 / 11	7 / 20	0
<b><math>X_2 \rightarrow X_1</math></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>5 / 11</b>	<b>100</b>
<b><math>X_1 \rightarrow X_3</math></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>7 / 15</b>	<b>100</b>
$X_3 \rightarrow X_1$	5 / 3	0	0	6 / 14	6 / 9	0
$X_1 \rightarrow X_4$	6 / 7	0	0	7 / 50	2 / 24	3 / 0
$X_4 \rightarrow X_1$	3 / 2	0	0	2 / 5	5	0
<b><math>X_2 \rightarrow X_3</math></b>	<b>100</b>	<b>98 / 100</b>	<b>94 / 100</b>	<b>14 / 39</b>	<b>9 / 18</b>	<b>64 / 99</b>
$X_3 \rightarrow X_2$	8 / 5	0	0	4 / 16	5 / 3	1 / 0
$X_2 \rightarrow X_4$	7 / 6	3 / 20	0	5 / 8	1 / 20	2 / 0
<b><math>X_4 \rightarrow X_2</math></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>7 / 2</b>	<b>100</b>
$X_3 \rightarrow X_4$	4 / 3	0 / 2	0	8 / 29	2 / 8	3 / 0
$X_4 \rightarrow X_3$	4 / 5	0	0	7	5 / 6	0

**Table 3.** Mean PTE values from 100 realizations of System 2, for  $P = 5$  and  $n = 512, 2048$ . The values of the true direct couplings are highlighted.

mean	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$	$X_1 \rightarrow X_4$	$X_4 \rightarrow X_1$
$n = 512$	0.0042	<b>0.0920</b>	<b>0.0772</b>	0.0034	0.0067	0.0043
$n = 2048$	0.0029	<b>0.1221</b>	<b>0.0965</b>	0.0016	0.0034	0.0020
	<b><math>X_2 \rightarrow X_3</math></b>	$X_3 \rightarrow X_2$	$X_2 \rightarrow X_4$	<b><math>X_4 \rightarrow X_2</math></b>	$X_3 \rightarrow X_4$	$X_4 \rightarrow X_3$
$n = 512$	<b>0.0060</b>	0.0052	0.0095	<b>0.0998</b>	0.0071	0.0033
$n = 2048$	<b>0.0059</b>	0.0032	0.0061	<b>0.1355</b>	0.0042	0.0013

The PTE detects the direct causal relationships, apart from the coupling,  $X_2 \rightarrow X_3$ , although the percentage of rejection in this direction increases with  $n$  (from 14% for  $n = 512$  to 39% for  $n = 2048$ ). Further, the erroneous relationships,  $X_1 \rightarrow X_4$  (50%) and  $X_3 \rightarrow X_4$  (29%), are observed for  $n = 2048$ . Focusing on the PTE values, it can be observed that they are much higher for the directions of direct

couplings than for the remaining directions. Moreover, the percentages of significant PTE values increase with  $n$  for the directions with direct couplings and decrease with  $n$  for all other couplings (see Table 3).

We note that the standard deviation of the estimated PTE values from the 100 realizations are low (on the order of  $10^{-2}$ ). Thus, the result of having falsely statistically significant PTE values for  $X_1 \rightarrow X_4$  and  $X_3 \rightarrow X_4$  is likely due to insufficiency of the randomization test.

PSTE fails to detect the causal effects for the second coupled system for both time series lengths, giving rejections at a rate between 1% and 24% for all directions. The failure of PSTE may be due to the high dimensionality of the rank vectors (the joint rank vector has dimension 21).

#### 4.3. Results for System 3

The CGCI, PGCI, PDC, PTE and PMIME correctly detect all direct causal effects ( $X_1 \rightarrow X_2$ ,  $X_1 \rightarrow X_4$ ,  $X_5 \rightarrow X_1$ ,  $X_2 \rightarrow X_4$ ,  $X_5 \rightarrow X_2$ ,  $X_5 \rightarrow X_3$ ,  $X_4 \rightarrow X_5$ ) for  $P = m = L_{max} = 4$  (based on BIC and AIC), as shown in Table 4.

**Table 4.** Percentages of statistically significant values of causality measures for System 3, for  $P = m = L_{max} = 4$ .

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$X_1 \rightarrow X_2$	100	86 / 100	92 / 100	81 / 100	8 / 7	100
$X_2 \rightarrow X_1$	7 / 3	2 / 0	0	7 / 2	3 / 8	18 / 0
$X_1 \rightarrow X_3$	6 / 5	2 / 0	0	3 / 4	4 / 10	2 / 0
$X_3 \rightarrow X_1$	1 / 2	0	0	5 / 4	2 / 4	14 / 0
$X_1 \rightarrow X_4$	100	100	100	52 / 100	2 / 9	100
$X_4 \rightarrow X_1$	7 / 4	1 / 0	0	6 / 4	4 / 10	12 / 0
$X_1 \rightarrow X_5$	5	2	0	6 / 10	7 / 12	16 / 0
$X_5 \rightarrow X_1$	100	98 / 100	99 / 100	100	2 / 16	100
$X_2 \rightarrow X_3$	4	2 / 0	0	4 / 6	9 / 5	2 / 0
$X_3 \rightarrow X_2$	6 / 2	1 / 0	0	4 / 2	5 / 4	9 / 1
$X_2 \rightarrow X_4$	100	99 / 100	96 / 100	18 / 77	7	94 / 100
$X_4 \rightarrow X_2$	5 / 9	0	0	5 / 4	2 / 4	9 / 0
$X_2 \rightarrow X_5$	6 / 4	2 / 3	0	5 / 4	4 / 6	22 / 0
$X_5 \rightarrow X_2$	100	96 / 100	99 / 100	99 / 100	3 / 8	100
$X_3 \rightarrow X_4$	3 / 7	1 / 0	0	3 / 4	5 / 8	0
$X_4 \rightarrow X_3$	4	1 / 0	0	3 / 6	4 / 5	4 / 0
$X_3 \rightarrow X_5$	4 / 3	1	0	7 / 4	6	14 / 0
$X_5 \rightarrow X_3$	100	87 / 100	84 / 100	49 / 97	3 / 15	100
$X_4 \rightarrow X_5$	100	100	100	100	6 / 4	100
$X_5 \rightarrow X_4$	5 / 2	0	0	17 / 37	6 / 14	1 / 0

The PDC is again estimated in the range of frequencies,  $[0, 0.4]$ , (see in Figure 1b the auto-spectra of the variables and the  $p$ -values from the parametric test of PDC from one realization of the system). The CGCI, PGCI and PDC perform similarly for the two time series lengths. The PTE indicates 100% significant values for  $n = 2048$  when direct causality exists. However, the PTE also indicates the spurious causality,  $X_5 \rightarrow X_4$ , for  $n = 2048$  (37%). The specificity of the PMIME is improved by the



increase of  $n$ , and the percentage of positive PMIME values in case of no direct causal effects varies from 0% to 22% for  $n = 512$ , while for  $n = 2048$ , it varies from 0% to 1%. The PSTE again fails to detect the causal effects, giving very low percentage of rejection at all directions (2% to 16%).

Since the linear causality measures CGCI, PGCI and PDC have been developed for the detection of direct causality in linear coupled systems, it was expected that these methods would be successfully applied to all linear systems. The nonlinear measures PMIME and PTE also seem to be able to capture the direct linear couplings in most cases, with PMIME following close the linear measures both in specificity and sensitivity.

In the following systems, we investigate the ability of the causality measures to correctly detect direct causal effects when nonlinearities are present.

#### 4.4. Results for System 4

For the fourth coupled system, the BIC and AIC suggest to set  $P = 1, 2$  and  $3$ . The performance of the linear measures does not seem to be affected by the choice of  $P$ . The PDC is estimated for frequencies in  $[0.1, 0.4]$ . The auto-spectra of the three variables do not display any peaks or any upward/downward trends. No significant differences in the results are observed if a wider or narrower range of frequencies is considered. The linear measures, CGCI, PGCI and PDC, capture only the linear direct causal effect,  $X_2 \rightarrow X_3$ , while they fail to detect the nonlinear relationships,  $X_1 \rightarrow X_2$  and  $X_1 \rightarrow X_3$ , for both time series lengths.

The PTE and the PMIME correctly detect all the direct couplings for the fourth coupled system for  $m = L_{max} = 1, 2$  and  $3$ . The percentage of significant values of the causality measures are displayed in Table 5. The PTE gives equivalent results for  $m = 1$  and  $m = 2$ . The PTE correctly detects the causalities for  $m = 3$ , but at a smaller power of the significance test for  $n = 512$  (63% for  $X_1 \rightarrow X_2$ , 46% for  $X_2 \rightarrow X_3$  and 43% for  $X_1 \rightarrow X_3$ ). The percentage of significant PMIME values is 100% for the directions of direct couplings, and falls between 0% and 6% for all other couplings, and this holds for any  $L_{max} = 1, 2$  or  $3$  and for both  $n$ .

**Table 5.** Percentage of statistically significant values of the causality measures for System 4,  $P = m = L_{max} = 2$ .

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$X_1 \rightarrow X_2$	12 / 7	1	2	97 / 100	10 / 69	100
$X_2 \rightarrow X_1$	2 / 7	0 / 1	1 / 0	8 / 9	4 / 8	3 / 0
$X_2 \rightarrow X_3$	100	73 / 100	100	76 / 100	69 / 100	100
$X_3 \rightarrow X_2$	7 / 4	3 / 1	1	4	3 / 9	4 / 0
$X_1 \rightarrow X_3$	7	1 / 0	0 / 1	86 / 100	1 / 7	100
$X_3 \rightarrow X_1$	4 / 5	0	0	4 / 6	8 / 21	0

The PSTE indicates the link  $X_2 \rightarrow X_3$  for both time series lengths, while  $X_1 \rightarrow X_2$  is detected only for  $n = 2048$ . The PSTE fails to point out the causality,  $X_1 \rightarrow X_3$ . The results for  $m = 2$  and  $3$



are equivalent. In order to investigate whether the failure of PSTE to show  $X_1 \rightarrow X_3$  is due to finite sample data, we estimate the PSTE also for  $n = 4096$ . For  $m = 2$ , it indicates the same results as for  $n = 2048$ . For  $m = 3$ , the PSTE detects all the direct causal effects,  $X_1 \rightarrow X_2$  (99%),  $X_2 \rightarrow X_3$  (100%),  $X_1 \rightarrow X_3$  (86%), but  $X_3 \rightarrow X_1$  (62%) is also erroneously detected.

#### 4.5. Results for System 5

For the fifth coupled simulation system, we set the model order,  $P = 2$ , based on the complexity of the system, and  $P = 3, 4$  and  $5$  using the AIC and BIC. The auto-spectra of the variables display peaks in  $[0.1, 0.2]$  and  $[0.4, 0.5]$ . The PDC is estimated for different ranges of frequencies to check its sensitivity with respect to the selection of the frequency range. When small frequencies are considered, the PDC seems to indicate larger percentages of spurious couplings; however, also, the percentages of significant PDC values at the directions of true causal effects are smaller. The results are presented for System 5 considering the range of frequencies,  $[0.4, 0.5]$ .

The CGCI seems to be sensitive to the selection of the model order  $P$ , indicating some spurious couplings for the different  $P$ . The best performance for CGCI is achieved for  $P = 3$ ; therefore, only results for  $P = 3$  are shown. On the other hand, the PGCI turns out to be less dependent on  $P$ , giving similar results for  $P = 2, 3, 4$  and  $5$ . The PTE is not substantially affected by the selection of the embedding dimension,  $m$  (at least for the examined coupling strengths); therefore, only results for  $m = 2$  are discussed. The PSTE is sensitive to the selection of  $m$ , performing best for  $m = 2$  and  $3$ , while for  $m = 4$  and  $5$ , it indicates spurious and indirect causal effects. The PMIME does not seem to depend on  $L_{max}$ . Results are displayed for  $L_{max} = 5$ . The percentage of significant values for each measure are displayed in Figure 2, for all directions, for increasing coupling strength and for both  $n$ .

**Figure 2.** Percentage of significant (a) CGCI ( $P = 3$ ); (b) PGCI ( $P = 3$ ); (c) PDC ( $P = 3$ ); (d) PTE ( $m = 2$ ); (e) PSTE ( $m = 2$ ); and (f) PMIME ( $L_{max} = 5$ ) values, for System 5, for increasing coupling strengths,  $c$ , at all directions and for both  $n$  (for  $n = 512$ , solid lines, for  $n = 2048$ , dotted lines).

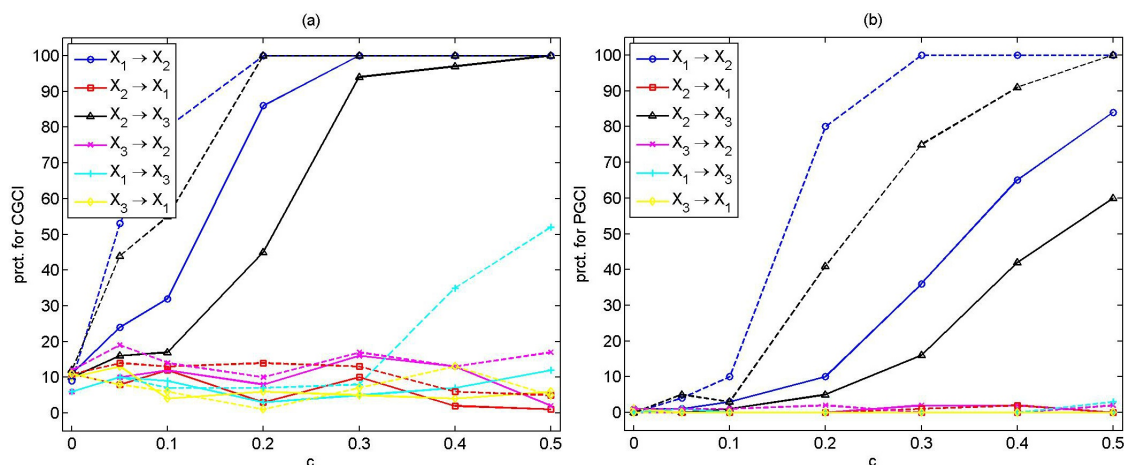
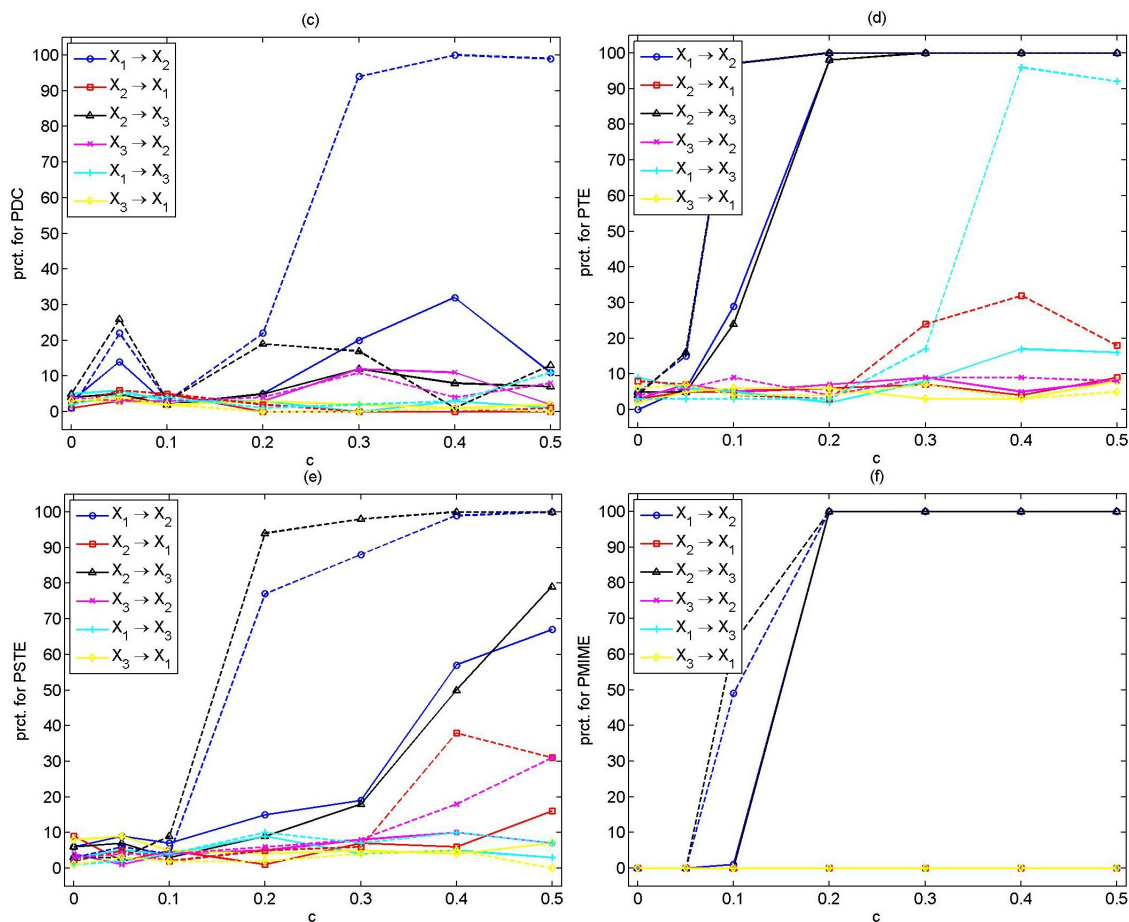


Figure 2. Cont.



Most of the measures show good specificity, and the percentage of rejection for all pairs of the variables of the uncoupled system ( $c = 0$ ) is at the significance level,  $\alpha = 0.05$ , with only CGCI scoring a somehow larger percentage of rejection up to 17%.

For the weak coupling strengths,  $c = 0.05$  and  $0.1$ , the causality measures cannot effectively detect the causal relationships or have a low sensitivity. The CGCI and the PTE seem to have the best performance, while the PMIME seems to be effective only for  $n = 2048$  and  $c = 0.1$ .

As the coupling strength increases, the sensitivity of the causality measures is improved. For  $c = 0.2$ , the CGCI, PTE and PMIME correctly indicate the true couplings for both  $n$ , while the PGCI and the PSTE do this only for  $n = 2048$ . The PDC has low power, even for  $n = 2048$ . For  $c = 0.3$ , nearly all measures correctly point out the direct causal effects (see Table 6). The best results are obtained with the PMIME, while the CGCI and PTE display similar performance. The PGCI and the PSTE are sensitive to the time series length and have a high power only for  $n = 2048$ . The PDC performs poorly, giving low percentage of significant PDC values, even for  $n = 2048$ . All measures have good specificity, with CGCI and PTE giving rejections well above the nominal level for some non-existing couplings.

Considering larger coupling strengths, the causality measures correctly indicate the true couplings, but also some spurious ones. The PMIME outperforms the other measures giving 100% positive values for both  $n$  for  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$  and 0% at the remaining directions for  $c \geq 0.2$ . Indicative results for all measures are displayed for the strong coupling strength  $c = 0.5$  in Table 7.

In order to investigate the effect of noise on each measure, we consider the coupled Hénon map (System 5) with the addition of Gaussian white noise with standard deviation 0.2 times the standard deviation of the original time series. Each measure is estimated again from 100 realizations from the noisy system for the same free parameters as considered in the noise-free case.

The CGCI is not significantly affected by the addition of noise, giving equivalent results for  $P = 3$  as for the noise-free system. The CGCI detects the true causal effects even for weak coupling strength ( $c \geq 0.05$ ). For different  $P$  values ( $P = 2, 4$  or  $5$ ), some spurious and/or indirect causal effects are observed for  $c > 0.3$ .

**Table 6.** Percentage of statistically significant values of the causality measures for System 5 for  $c = 0.3$ , where  $P = 3$ ,  $m = 2$  and  $L_{max} = 5$ .

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$\mathbf{X_1 \rightarrow X_2}$	<b>100</b>	<b>36 / 100</b>	<b>20 / 94</b>	<b>100</b>	<b>19 / 88</b>	<b>100</b>
$X_2 \rightarrow X_1$	10 / 13	0 / 1	0 / 2	7 / 24	7 / 6	0
$\mathbf{X_2 \rightarrow X_3}$	<b>94 / 100</b>	<b>16 / 75</b>	<b>12 / 19</b>	<b>100</b>	<b>18 / 98</b>	<b>100</b>
$X_3 \rightarrow X_2$	16 / 17	2 / 0	12 / 4	9	8	0
$X_1 \rightarrow X_3$	5 / 8	0	0 / 1	8 / 17	4 / 7	0
$X_3 \rightarrow X_1$	5 / 7	0	2 / 0	3 / 7	5 / 4	0

**Table 7.** Percentages of statistically significant values of the causality measures for System 5 for  $c = 0.5$ , where  $P = 3$ ,  $m = 2$  and  $L_{max} = 5$ .

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$\mathbf{X_1 \rightarrow X_2}$	<b>100</b>	<b>84 / 100</b>	<b>11 / 99</b>	<b>100</b>	<b>67 / 100</b>	<b>100</b>
$X_2 \rightarrow X_1$	1 / 5	0	0 / 1	9 / 18	16 / 31	0
$\mathbf{X_2 \rightarrow X_3}$	<b>100</b>	<b>60 / 100</b>	<b>7 / 13</b>	<b>100</b>	<b>79 / 100</b>	<b>100</b>
$X_3 \rightarrow X_2$	2 / 17	0 / 2	2 / 8	8	7 / 31	0
$X_1 \rightarrow X_3$	12 / 52	0 / 3	1 / 11	16 / 92	3 / 7	0
$X_3 \rightarrow X_1$	6 / 5	0	2 / 0	8 / 5	7 / 0	0

The PGCI is also not considerably affected by the addition of noise. The causal effects are detected only for coupling strengths,  $c \geq 0.3$ , for  $n = 512$ , and for  $c \geq 0.2$ , for  $n = 2048$ , while the power of the test increases with  $c$  and with  $n$  (see Figure 3a).

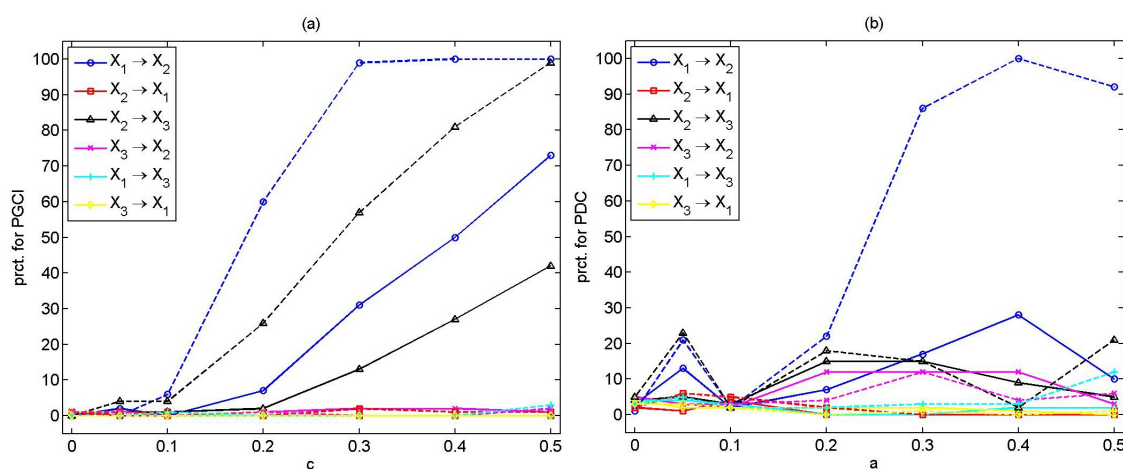
The PDC fails in the case of the noisy coupled Hénon maps, detecting only the coupling  $X_1 \rightarrow X_2$ , for coupling strengths,  $c \geq 0.2$  and  $n = 2048$  (see Figure 3b).

The PTE seems to be significantly affected by the addition of noise, falsely detecting the coupling,  $X_2 \rightarrow X_1$ ,  $X_3 \rightarrow X_2$ , and the indirect coupling,  $X_1 \rightarrow X_3$ , for strong coupling strengths. The performance of PTE is not significantly influenced by the choice of  $m$ . Indicative results are presented in Table 8 for  $m = 2$ .

Noise addition does not seem to affect the performance of PSTE. Results for  $m = 2$  are equivalent to the results obtained for the noise-free case. The power of the significance test increases with  $c$  and  $n$ . The PSTE is sensitive to the selection of  $m$ ; as  $m$  increases, the percentage of significant PSTE values in the directions of no causal effects also increases.

The PMIME outperforms the other measures also for the noisy coupled Hénon maps, detecting the true couplings for  $c \geq 0.2$  for  $n = 512$  (100%) and for  $c \geq 0.1$  for  $n = 2048$  (for coupling strength  $c = 0.1$  the percentages are 22% and 23% for  $X_1 \rightarrow X_2$ ,  $X_2 \rightarrow X_3$ , respectively, and for  $c \geq 0.2$ , the percentages are 100%, for both couplings).

**Figure 3.** Percentage of significant (a) PGCI ( $P = 3$ ) and (b) PDC ( $P = 3$ ) values, for System 5 with addition of noise (solid lines for  $n = 512$ , dotted lines for  $n = 2048$ ).



**Table 8.** Percentages of statistically significant PTE ( $m = 2$ ) values for System 5 with the addition of noise.

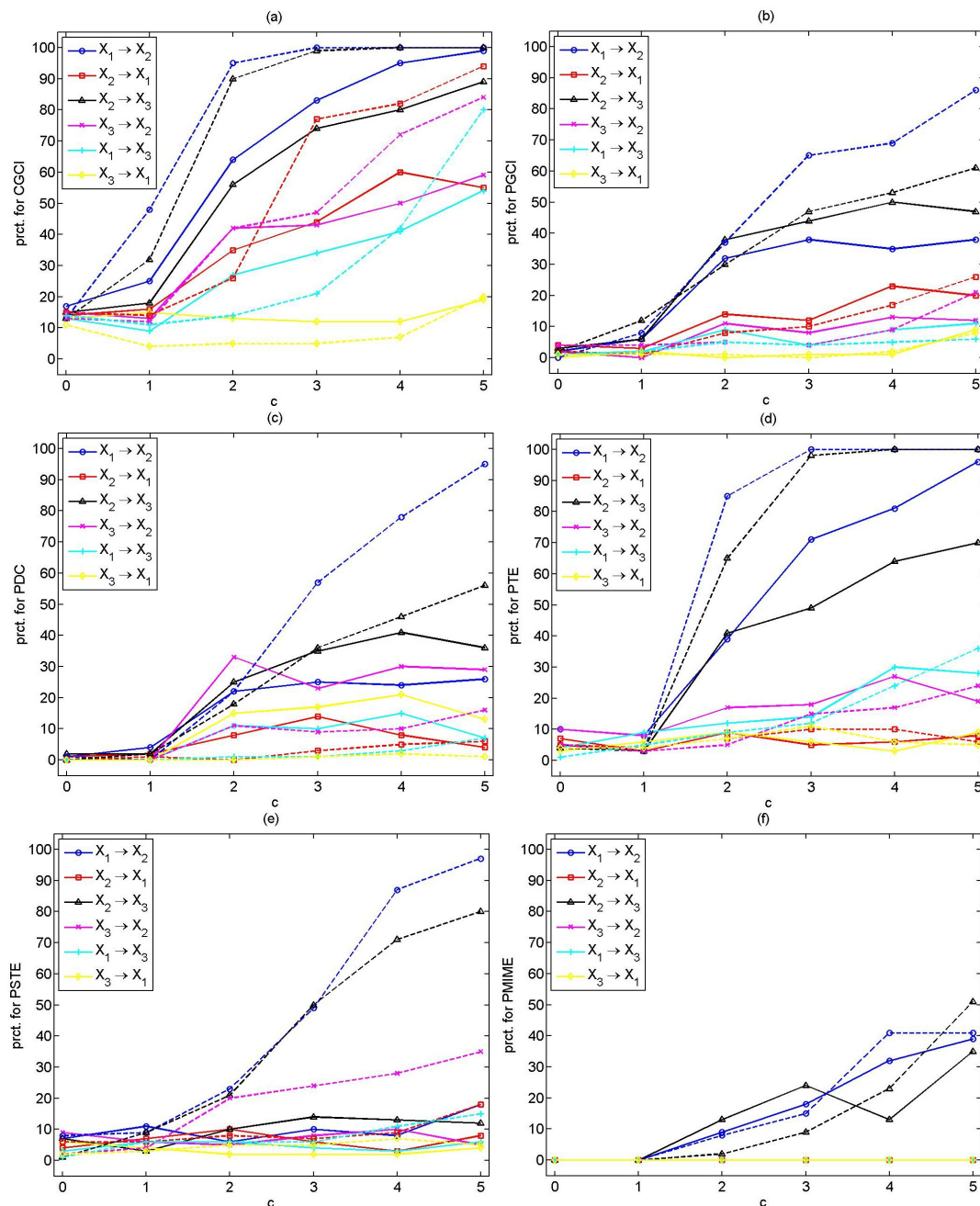
$n = 512/2048$	$X_1 \rightarrow X_2$	$X_2 \rightarrow X_1$	$X_2 \rightarrow X_3$	$X_3 \rightarrow X_2$	$X_1 \rightarrow X_3$	$X_3 \rightarrow X_1$
$c = 0$	2 / 5	5 / 6	4	5 / 7	10 / 1	7 / 6
$c = 0.05$	6 / 17	3	5 / 20	4 / 5	3 / 6	7 / 1
$c = 0.1$	22 / 98	6 / 2	22 / 98	3 / 7	4 / 5	5 / 8
$c = 0.2$	100	6 / 11	99 / 100	4 / 7	1 / 5	4 / 2
$c = 0.3$	100	10 / 52	100	8 / 22	12 / 27	4 / 8
$c = 0.4$	100	9 / 79	100	6 / 50	24 / 97	7 / 10
$c = 0.5$	100	23 / 95	100	7 / 48	39 / 100	8 / 13

#### 4.6. Results for System 6

For System 6, we set  $P = 3$  based on the complexity of the system and  $P = 5$  regarding the AIC and BIC. The PTE, PSTE and PMIME are estimated for four different combinations of the free parameters,  $h$  and  $m$  ( $L_{max}$  for PMIME), i.e., for  $h = 1$  and  $m = 3$ , for  $h = 3$  and  $m = 3$ , for  $h = 1$  and  $m = 5$  and for  $h = 5$  and  $m = 5$ . The PDC is computed for the range of frequencies,  $[0, 0.2]$ , based on the auto-spectra

of the variables. As this system is a nonlinear flow, the detection of causal effects is more challenging compared to stochastic systems and nonlinear coupled maps. Indicative results for all causality measures are displayed for increasing coupling strengths in Figure 4.

**Figure 4.** Percentage of significant (a) CGCI ( $P = 5$ ); (b) PGCI ( $P = 5$ ); (c) PDC ( $P = 5$ ); (d) PTE ( $h = 1, m = 5$ ); (e) PSTE ( $h = 3, m = 3$ ); and (f) PMIME ( $h = 1, L_{max} = 5$ ) values, for System 6, for increasing coupling strengths,  $c$ , at all directions and for both  $n$  (solid lines for  $n = 512$ , dotted lines for  $n = 2048$ ).



The CGCI has poor performance, indicating many spurious causalities. The PGCI improves the specificity of the CGCI, but still, the percentages of statistically significant PGCI values increase with  $c$  for non-existing direct couplings (less for larger  $n$ ). Similar results are obtained for  $P = 3$  and 5. On



the other hand, the PDC is sensitive to the selection of  $P$ , indicating spurious causal effects for all  $P$ . As  $P$  increases, the percentage of significant PDC values at the directions of no causal effects is reduced. However, the power of the test is also reduced.

The PTE is sensitive to the embedding dimension  $m$  and the number of steps ahead  $h$ , performing best for  $h = 1$  and  $m = 5$ . It fails to detect the causal effects for small  $c$ ; however, for  $c > 1$ , it effectively indicates the true couplings. The size of the test increases with  $c$  (up to 36% for  $c = 5$  and  $n = 2048$ ), while the power of the test increases with  $n$ .

The PSTE is also affected by its free parameters, performing best for  $h = 3$  and  $m = 3$ . It is unable to detect the true causal effects for weak coupling strengths ( $c \leq 1$ ) and for small time series lengths. The PSTE is effective only for  $c > 2$  and  $n = 2048$ . Spurious couplings are observed for strong coupling strengths  $c$  and  $n = 2048$ .

The PMIME is also influenced by the choice of  $h$  and  $L_{max}$ , indicating low sensitivity when setting  $h = 1$ , but no spurious couplings, while for  $h = L_{max}$ , the percentage of significant PMIME values for  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$  is higher, but the indirect coupling  $X_1 \rightarrow X_3$  is detected for strong coupling strengths. The PMIME has a poor performance for weak coupling strength ( $c < 2$ ).

As  $c$  increases, the percentages of significant values of almost all the causality measures increase, but not only at the true directions,  $X_1 \rightarrow X_2$  and  $X_2 \rightarrow X_3$ . Indicative results are presented in Table 9 for strongly coupled systems ( $c = 5$ ). The CGCI gives high percentages of rejection of  $H_0$  for all couplings (very low specificity). This also holds for the PGCI, but at a lower significance level. The PTE correctly detects the two true direct causal effects for  $h = 1$  and  $m = 5$ , but at some significant degree, also the indirect coupling,  $X_1 \rightarrow X_3$ , and the non-existing coupling,  $X_3 \rightarrow X_2$ . The PSTE does not detect the direct couplings for  $n = 512$ , but it does when  $n = 2048$  (97% for  $X_1 \rightarrow X_2$  and 80% for  $X_2 \rightarrow X_3$ ), but then it detects also spurious couplings, most notably  $X_3 \rightarrow X_2$  (35%). The PMIME points out only the direct causal effects, giving, however, a lower percentage than the other measures for  $h = 1$ ,  $L_{max} = 5$ . Its performance seems to be affected by the selection of  $h$  and  $L_{max}$ . The nonlinear measures turn out to be more sensitive to their free parameters.

**Table 9.** Percentage of statistically significant values of the causality measures for System 6 with  $c = 5$ , where  $P = 5$ ,  $h = 1$  and  $m = 5$  for PTE,  $h = 3$  and  $m = 3$  for PSTE and  $h = 1$  and  $L_{max} = 5$  for PMIME.

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$X_1 \rightarrow X_2$	99 / 100	38 / 86	26 / 95	96 / 100	18 / 97	39 / 41
$X_2 \rightarrow X_1$	55 / 94	20 / 26	4 / 6	8 / 6	8 / 18	0
$X_2 \rightarrow X_3$	89 / 100	47 / 61	36 / 56	70 / 100	12 / 80	35 / 51
$X_3 \rightarrow X_2$	59 / 84	12 / 21	29 / 16	19 / 24	5 / 35	0
$X_1 \rightarrow X_3$	54 / 80	11 / 6	7 / 7	28 / 36	6 / 15	0
$X_3 \rightarrow X_1$	19 / 20	9 / 8	13 / 1	9 / 5	4 / 5	0

## 4.7. Results for System 7

For the last coupled simulation system, we set the model order  $P = 2, 3, 4$  based on AIC and BIC, while the PDC is estimated in the range of frequencies,  $[0.1, 0.2]$ . The embedding dimension,  $m$ , for the estimation of PTE and PSTE, as well as  $L_{max}$  for the estimation of PMIME, are set equal to  $P$ . Results for all causality measures are displayed in Table 10.

The CGCI (for  $P = 5$ ) correctly indicates the causal effects for  $n = 512$ , giving 100% percentage of significant values at the direction  $X_1 \rightarrow X_2$  and  $X_1 \rightarrow X_4$ , but lower percentage at the directions  $X_1 \rightarrow X_3$  (63%),  $X_4 \rightarrow X_5$  (37%) and  $X_5 \rightarrow X_4$  (42%). The power of the test increases with  $n$ , but spurious couplings are also detected for  $n = 2048$ . For  $P = 2, 3$  and 4, the CGCI indicates more spurious couplings than for  $P = 5$ .

**Table 10.** Percentage of statistically significant values of the causality measures for System 7, where  $P = m = L_{max} = 3$ .

$n = 512/2048$	CGCI	PGCI	PDC	PTE	PSTE	PMIME
$\mathbf{X_1 \rightarrow X_2}$	<b>100</b>	<b>100</b>	<b>47 / 100</b>	<b>100</b>	<b>47 / 31</b>	<b>100</b>
$X_2 \rightarrow X_1$	8 / 26	0	7 / 23	6 / 9	86 / 100	0
$\mathbf{X_1 \rightarrow X_3}$	<b>63 / 100</b>	<b>100</b>	<b>57 / 100</b>	<b>36 / 91</b>	<b>100</b>	<b>100</b>
$X_3 \rightarrow X_1$	6 / 8	1 / 8	1 / 2	98 / 100	32 / 26	100
$\mathbf{X_1 \rightarrow X_4}$	<b>100</b>	<b>100</b>	<b>73 / 100</b>	<b>100</b>	<b>31 / 84</b>	<b>100</b>
$X_4 \rightarrow X_1$	6	0	2 / 0	54 / 96	73 / 100	87 / 23
$X_1 \rightarrow X_5$	13 / 75	3 / 8	2 / 11	0 / 1	100	20 / 9
$X_5 \rightarrow X_1$	9 / 5	0	1 / 2	5 / 16	100	1 / 0
$X_2 \rightarrow X_3$	14 / 36	0	7 / 30	49 / 75	40 / 24	3 / 41
$X_3 \rightarrow X_2$	9 / 12	0	4 / 3	40 / 66	100	1 / 0
$X_2 \rightarrow X_4$	11 / 38	3 / 0	7 / 29	36 / 81	91 / 92	0
$X_4 \rightarrow X_2$	7 / 10	18 / 22	2 / 1	0 / 2	100	0
$X_2 \rightarrow X_5$	10 / 37	0	5 / 29	47 / 69	100	86 / 94
$X_5 \rightarrow X_2$	5 / 7	0	1 / 3	20 / 49	72 / 96	1 / 0
$X_3 \rightarrow X_4$	12	0	5 / 3	88 / 100	100	0
$X_4 \rightarrow X_3$	5 / 10	0	2 / 2	14 / 59	98 / 100	15 / 0
$X_3 \rightarrow X_5$	8 / 13	0	4 / 1	1 / 2	100	0
$X_5 \rightarrow X_3$	7 / 11	0	1 / 3	14 / 43	100	5 / 0
$\mathbf{X_4 \rightarrow X_5}$	<b>37 / 94</b>	<b>83 / 100</b>	<b>23 / 88</b>	<b>6 / 7</b>	<b>94 / 100</b>	<b>100</b>
$\mathbf{X_5 \rightarrow X_4}$	<b>42 / 100</b>	<b>6 / 4</b>	<b>20 / 76</b>	<b>100</b>	<b>78 / 75</b>	<b>100</b>

System 7 favors the PGCI, as it has been specifically defined for systems with latent and exogenous variables. The PGCI denotes the couplings,  $X_1 \rightarrow X_2$ ,  $X_1 \rightarrow X_3$ ,  $X_1 \rightarrow X_4$  and  $X_4 \rightarrow X_5$ , even for  $n = 512$  with a high percentage; however, it fails to detect the coupling,  $X_5 \rightarrow X_4$ , for both  $n$ .



The PDC detects the true couplings at low percentage for  $n = 512$ . The percentage increases with  $n = 2048$  at the directions of the true couplings. However, there are also false indications of directed couplings.

The PTE does not seem to be effective in this setting for any of the considered  $m$  values, since it indicates many spurious causal effects, while it fails to detect  $X_4 \rightarrow X_5$ . The PSTE completely fails in this case, suggesting significant couplings at all directions. The true causal effects are indicated by the PMIME, but here, as well, many spurious causal effects are also observed.

## 5. Discussion

In this paper, we have presented six multivariate causality measures that are able to detect the direct causal effects among simultaneously measured time series. The multivariate direct coupling measures are tested on simulated data from coupled and uncoupled systems of different complexity, linear and nonlinear, maps and flows. The linear causality measures and the PMIME can be used in complex systems with a large number of observed variables, but the PTE and the PSTE fail, because they involve estimation of probability distributions of high dimensional variables.

The simulation results suggest that for real world data, it is crucial to investigate the presence of nonlinearities and confirm the existence of causal effects by estimating more than one causality measure, sensitive to linear, as well as nonlinear causalities. Concerning the specificity of the coupling measures (in absence of direct causality), the PMIME outperforms the other measures, but for weak coupling, it is generally less sensitive than the PTE. In general, the PMIME indicated fewer spurious causal effects. Here, we considered only systems of a few variables, and for larger systems, the PMIME was found to outperform the PTE and, also, the CGCI [28].

Regarding the three first linear coupled systems in the simulation study, the CGCI, PGCI and PDC are superior to the nonlinear causality measures, both in sensitivity and specificity. The PMIME correctly indicates the coupling among the variables, but tends to have smaller sensitivity than the linear tests. The PTE cannot detect the true direct causality in all the examined linear systems and gives also some spurious results. The PSTE was the less effective one. For the last simulation system with the exogenous and latent variables (System 7), all but PGCI measures had low specificity, indicating only the true direct causal effects.

Concerning the nonlinear coupled system, the PTE and the PMIME outperform the other methods. The linear measures (CGCI, PGCI and PDC) fail to consistently detect the true direct causal effects (low sensitivity). The failure of the linear measures may not only be due to the fact that the system is nonlinear, but also due to the small time series lengths and the low model order. For example, the PDC correctly indicated the causal effects on simulated data from the coupled Rössler system for  $n = 50,000$  and model order,  $P = 200$  (see [49]). The PSTE requires large data sets to have a good power, while it gives spurious couplings at many cases. Though the PSTE performed overall worst in the simulation study, there are other settings in which it can be useful, e.g., in the presence of outliers or non-stationarity in mean, as slow drifts do not have a direct effect on the ranks. The addition of noise does not seem to affect the causality measures, CGCI, PGCI, PSTE and PMIME.

The free parameters were not optimized separately for each measure. For all systems, the parameters of model order,  $P$ , embedding dimension,  $m$ , and maximum lag,  $L_{max}$ , were treated as one free parameter, the values of which were selected according to the complexity of each system and the standard criteria of AIC and BIC. The linear measures tend to be less sensitive to changes on this free parameter than the nonlinear ones. The PTE gave more consistent results than the PSTE for varying  $m$ , whereas the PMIME was not dependent on  $L_{max}$ . For the nonlinear measures and the continuous-time system (three coupled Lorenz systems), we considered also the causalities at more than one step ahead, and the PTE, PSTE and PMIME were found to be sensitive to the selection of the steps ahead.

A point of concern regarding all direct causality measures, but the PMIME, is that the size of the significance test was high in many settings. This was observed for both types of spurious direct causal effect, *i.e.*, when there is indirect coupling or when there is no causal effect. In many cases of non-existing direct causalities, although the observed test size was large, the estimated values of the measure were low compared to those in the presence of direct causalities. This raises also the question of the validity of the significance tests. The randomization test used time-shifted surrogates. Although it is simple and straightforward to implement, it may not always be sufficient, and further investigation for other randomization techniques is due for future work.

In conclusion, we considered six of the best-known measures of direct causality and studied their performance for different systems, time series lengths and free parameters. The worst performance was observed for the PSTE, since it completely failed in the case of the linear coupled systems, while for nonlinear systems, it required large data sets. The other measures scored differently in terms of sensitivity and specificity in the different settings. The CGCI, PGCI and PDC outperformed the nonlinear ones in the case of the linear coupled simulation systems, while in the presence of exogenous and latent variables, the PGCI seems to be the most effective one. The PMIME seems to have the best performance for nonlinear and noisy systems, while always obtaining the highest specificity, indicating no spurious effects. It is the intention of the authors to pursue the comparative study on selected real applications.

## Acknowledgements

The research project is implemented within the framework of the Action “Supporting Postdoctoral Researchers” of the Operational Program, “Education and Lifelong Learning” (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
2. Geweke, J. Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* **1982**, *77*, 304–313.

3. Baek, E.; Brock, W. A general test for nonlinear Granger causality: Bivariate model, 1992. Working paper, University of Wisconsin, Madison, WI, USA.
4. Hiemstra, C.; Jones, J.D. Testing for linear and nonlinear Granger causality in the stock Price-Volume Relation. *J. Financ.* **1994**, *49*, 1639–1664.
5. Aparicio, F.; Escribano, A. Information-theoretic analysis of serial correlation and cointegration. *Stud. Nonlinear Dyn. Econom.* **1998**, *3*, 119–140.
6. Freiwald, V.A.; Valdes, P.; Bosch, J.; Biscay, R.; Jimenez, J.; Rodriguez, L. Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *J. Neurosci. Methods* **1999**, *94*, 105–119.
7. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103.
8. Arnhold, J.; Grassberger, P.; Lehnertz, K.; Elger, C.E. A robust method for detecting interdependences: Application to intracranially recorded EEG. *Physica D: Nonlinear Phenom.* **1999**, *134*, 419–430.
9. Romano, M.; Thiel, M.; Kurths, J.; Grebogi, C. Estimation of the direction of the coupling by conditional probabilities of recurrence. *Phys. Rev. E* **2007**, *76*, 036211.
10. Marko, H. The bidirectional communication theory—a generalization of information theory. *IEEE Trans. Commun.* **1973**, *21*, 1345–1351.
11. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
12. Paluš, M. Coarse-grained entropy rates for characterization of complex time series. *Physica D: Nonlinear Phenom.* **1996**, *93*, 64–77.
13. Rosenblum, M.G.; Pikovsky, A.S. Detecting direction of coupling in interacting oscillators. *Phys. Rev. E* **2001**, *64*, 045202.
14. Nolte, G.; Ziehe, A.; Nikulin, V.; Schlogl, A.; Kramer, N.; Brismar, T.; Muller, K.R. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* **2008**, *100*, 234101.
15. Blinowska, K.J.; Kus, R.; Kaminski, M. Granger causality and information flow in multivariate processes. *Phys. Rev. E* **2004**, *70*, 050902.
16. Kus, R.; Kaminski, M.; Blinowska, K.J. Determination of EEG activity propagation: Pair-wise versus multichannel estimate. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1501–1510.
17. Eichler, M. Graphical modelling of multivariate time series. *Probab. Theory Relat. Fields* **2012**, *153*, 233–268.
18. Guo, S.; Seth, A.K.; Kendrick, K.M.; Zhou, C.; Feng, J. Partial granger causality—eliminating exogenous inputs and latent variables. *J. Neurosci. Methods* **2008**, *172*, 79–93.
19. Kaminski, M.J.; Blinowska, K.J. A new method of the description of the information flow in the brain structures. *Biol. Cybern.* **1991**, *65*, 203–210.
20. Baccala, L.A.; Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.* **2001**, *84*, 463–474.
21. Korzeniewska, A.; Manczak, M.; Kaminski, M.; Blinowska, K.; Kasicki, S. Determination of information flow direction between brain structures by a modified directed transfer function (dDTF) method. *J. Neurosci. Methods* **2003**, *125*, 195–207.

22. Chavez, M.; Martinerie, J.; Quyen, M. Statistical assessment of non-linear causality: Application to epileptic EEG signals. *J. Neurosci. Methods* **2003**, *124*, 113–128.
23. Verdes, P. Assessing causality from multivariate time series. *Phys. Rev. E* **2005**, *72*, 026222.
24. Vakorin, V.A.; Krakovska, O.A.; McIntosh, A.R. Confounding effects of indirect connections on causality estimation. *J. Neurosci. Methods* **2009**, *184*, 152–160.
25. Papana, A.; Kugiumtzis, D.; Larsson, P.G. Detection of direct causal effects and application in the analysis of electroencephalograms from patients with epilepsy. *Int. J. Bifurc. Chaos* **2012**, *22*, doi: 10.1142/S0218127412502227.
26. Vlachos, I.; Kugiumtzis, D. Non-uniform state space reconstruction and coupling detection. *Phys. Rev. E* **2010**, *82*, 016207.
27. Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E* **2011**, *83*, 051112.
28. Kugiumtzis, D. Direct coupling information measure from non-uniform embedding. *Phys. Rev. E* **2013**, *87*, 062918.
29. Jizba, P.; Kleinert, H.; Shefaat, M. Rényi's information transfer between financial time series. *Physica A: Stat. Mech. Appl.* **2012**, *391*, 2971–2989.
30. Staniek, M.; Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett.* **2008**, *100*, 158101.
31. Kugiumtzis, D. Transfer entropy on rank vectors. *J. Nonlinear Syst. Appl.* **2012**, *3*, 73–81.
32. Kugiumtzis, D. Partial transfer entropy on rank vectors. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 401–420.
33. Smirnov, D.A.; Andrzejak, R.G. Detection of weak directional coupling: Phase-dynamics approach versus state-space approach. *Phys. Rev. E* **2005**, *71*, 036207.
34. Faes, L.; Porta, A.; Nollo, G. Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: Comparison among different strategies based on k nearest neighbors. *Phys. Rev. E* **2004**, *78*, 026201.
35. Papana, A.; Kugiumtzis, D.; Larsson, P.G. Reducing the bias of causality measures. *Phys. Rev. E* **2011**, *83*, 036207.
36. Silfverhuth, M.J.; Hintsala, H.; Kortelainen, J.; Seppanen, T. Experimental comparison of connectivity measures with simulated EEG signals. *Med. Biol. Eng. Comput.* **2012**, *50*, 683–688.
37. Dufour, J.M.; Taamouti, A. Short and long run causality measures: Theory and inference. *J. Econ.* **2010**, *154*, 42–58.
38. Florin, E.; Gross, J.; Pfeifer, J.; Fink, G.; Timmermann, L. Reliability of multivariate causality measures for neural data. *J. Neurosci. Methods* **2011**, *198*, 344–358.
39. Wu, M.-H.; Frye, R.E.; Zouridakis, G. A comparison of multivariate causality based measures of effective connectivity. *Comput. Biol. Med.* **2011**, *21*, 1132–1141.
40. Blinowska, K.J. Review of the methods of determination of directed connectivity from multichannel data. *Med. Biol. Eng. Comput.* **2011**, *49*, 521–529.
41. Brandt, P.T.; Williams, J.T. *Multiple Time Series Models*; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2007; chapter 2, pp. 32–34.
42. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300.

43. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
44. Schwartz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *5*, 461–464.
45. Ding, M.; Chen, Y.; Bressler, S.L. Granger Causality: Basic Theory and Applications to Neuroscience. In *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*; Schelter, B., Winterhalder, M., Timmer, J., Eds.; Wiley-VCH Verlag: Berlin, Germany, 2006; Chapter 17, pp. 437–460.
46. Schelter, B.; Winterhalder, M.; Eichler, M.; Peifer, M.; Hellwig, B.; Guschlbauer, B.; Lücking, C.H.; Dahlhaus, R.; Timmer, J. Testing for directed influences in neuroscience using partial directed coherence. *J. Neurosci. Methods* **2006**, *152*, 210–219.
47. Takahashi, D.Y.; Baccala, L.A.; Sameshima, K. Connectivity inference between neural structures via partial directed coherence. *J. Appl. Stat.* **2007**, *34*, 1255–1269.
48. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
49. Winterhalder, M.; Schelter, B.; Hesse, W.; Schwab, K.; Leistriz, L.; Klan, D.; Bauer, R.; Timmer, J.; Witte, H. Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. *Signal Process.* **2005**, *85*, 2137–2160.
50. Gourévitch, B.; Le Bouquin-Jeannés, R.; Faucon, G. Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biol. Cybern.* **2006**, *95*, 349–369.
51. Stefanski, A. *Determining Thresholds of Complete Synchronization, and Application*; World Scientific: Singapore, Singapore, 2009.
52. Wilke, C.; van Drongelen, W.; Kohrman, M.; Hea, B. Identification of epileptogenic foci from causal analysis of ECoG interictal spike activity. *Clin. Neurophys.* **2009**, *120*, 1449–1456.
53. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: New York, NY, USA, 1994.
54. Faes, L.; Porta, A.; Nollo, G. Testing frequency-domain causality in multivariate time series. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1897–1906.
55. Quian Quiroga, R.; Kraskov, A.; Kreuz, T.; Grassberger, P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E* **2002**, *65*, 041903.
56. Yu, G.H.; Huang, C.C. A distribution free plotting position. *Stoch. Environ. Res. Risk Assess.* **2001**, *15*, 462–476.
57. Seth, A.K. A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* **2010**, *186*, 262–273.