Testing different methodologies for Granger causality estimation: A simulation study

Yuri Antonacci
Sapienza University of Rome
IRCCS Fondazione Santa Lucia
Rome, Italy
antonacci@diag.uniroma1.it

Laura Astolfi
Sapienza University of Rome
IRCCS Fondazione Santa Lucia
Rome, Italy
laura.astolfi@uniroma1.it

Luca Faes
University of Palermo
Palermo, Italy
luca.faes@unipa.it

Abstract—Granger causality (GC) is a method for determining whether and how two time series exert causal influences one over the other. As it is easy to implement through vector autoregressive (VAR) models and can be generalized to the multivariate case, GC has spread in many different areas of research such as neuroscience and network physiology. In its basic formulation, the computation of GC involves two different regressions, taking respectively into account the whole past history of the investigated multivariate time series (full model) and the past of all time series except the putatively causal time series (restricted model). However, the restricted model cannot be represented through a finite order VAR process and, when few data samples are available or the number of time series is very high, the estimation of GC exhibits a strong reduction in accuracy. To mitigate these problems, improved estimation strategies have been recently implemented, including state space (SS) models and partial conditioning (PC) approaches. In this work, we propose a new method to compute GC which combines SS and PC and tests it together with other four commonly used estimation approaches. In simulated networks of linearly interacting time series, we show the possibility to reconstruct the network structure even in challenging conditions of data samples available.

Index Terms—Granger Causality, Dynamical Networks, Vector Autoregressive Processes, Multivariate Time Series

I. INTRODUCTION

The evaluation of the direction and strength of the interactions among simultaneously observed dynamical systems is an important topic currently under investigation in many fields of science. Granger causality (GC) [1] is a very versatile tool for analysing the cause-effect relationships between different time series descriptive of the system dynamics. GC was firstly formulated in the framework of linear bivariate autoregressive modelling, stating that a time series X G-causes another series Y if the past of X contains information that helps predict the future of Y above and beyond the information already contained in the past of Y [1]. To account for the influence of other time series which are potentially affecting the two series under analysis, the bivariate formulation has been extended to the multivariate case through the use of vector autoregressive (VAR) models which lead to the computation of a conditional form of GC [1]. Being based entirely on a liner framework, GC is very easy to implement, with very few parameters to be estimated if compared with model-free approaches and with a reduced computational cost. These features have allowed its extensive employment in the neurosciences and in the emergent field of network physiology, where the time series to be analyzed are often highly multivariate [2].

GC from a driver to a target time series is typically quantified by comparing the prediction error variance obtained from two different linear regression models: a model where the present sample of the target series is regressed on the past samples of all time series in the dataset (full model), and a model where the present of the target is regressed on the past of all series excluding the driver (restricted model). Unfortunately, since the restricted model is theoretically of infinite order, the finite sample estimation of GC based on performing two separate regressions leads to strong bias or very large variability of the estimates, depending on whether small or high model orders are used to identify the restricted model [3]. To counteract this problem, approaches based on the identification of a single VAR model have been recently proposed to estimate the GC. The first approach exploits the relation between the prediction error variance of the restricted model and the autocovariance function (ACF) of the whole vector process estimated at arbitrarily high time lags [4], [5]. The second approach exploits the state space (SS) modelling of the observed VAR process, and the possibility to provide a closed form SS representation also for the restricted model [6]. Another relevant issue for the practical computation of the multivariate form of GC is related to the computational and numerical problems to be faced when dealing with a large number of time series and/or a small amount of data samples available. This issue can be tackled by strategies limiting the number of time series to be involved in the linear regressions. One of such strategies is the so-called partial conditioning GC (PCGC), which computes GC after selecting only the most informative processes for the driver series [6].

Of the GC estimation strategies above presented, the methods based on ACF and SS models do not deal with the numerical problems related to high dimensionality and small data samples, while PCGC does not overcome the problem related with identifying the restricted model by means of finite

order regressions. In this paper, to overcome the limitations of existing GC estimation strategies, we propose to combine them by including the partial conditioning into a framework based on SS models. The performance of this combined estimator, denoted as PCGC-SS, is assessed in comparison with the standard double regression (GC-DR), the autocovariance-based method (GC-ACF), the state-space method (GC-SS) and the partial conditioning method (PCGC), in a simulation study reproducing linear GC interactions among multivariate time series under different conditions of data samples available for the estimation process.

II. GRANGER CAUSALITY

Let us consider a discrete-time, stationary vector stochastic process composed of M real-valued zero-mean scalar processes, $Y_n = [y_{1,n} \ldots y_{M,n}]^T$. Assuming that Y_n is a Markov process of order p, in the linear signal processing framework it can be completely described by the VAR model:

$$Y_n = \sum_{k=1}^{p} \mathbf{A}_k Y_{n-k} + U_n, \tag{1}$$

where $\mathbf{A}_k$ is an $M \times M$ matrix containing the autoregressive (AR) coefficients $a_{ji,k}$ that relate $y_{j,n}$ to $y_{i,n-k}$ $(i, j \in (1, \ldots, M), k \in (1, \ldots, p))$, and $U_n = [u_{1,n} \ldots u_{M,n}]^T$ is a vector of M zero-mean gaussian innovation processes with covariance matrix $\mathbf{\Sigma} \equiv \mathbb{E}[U_n U_n^T]$ (where $\mathbb{E}$ is the expectation value). The problem (1) can be solved by means of ordinary least squares (OLS), computing the matrix of coefficients that minimizes the residual error term [8].

Let us assume the process $y_{j,n}$ as the *target* and the process $y_{i,n}$ as the *driver* process, with the remaining $M-2$ processes collected in the vector $Y_{k,n}$, where $k = \{1, .., M\} \setminus \{i, j\}$. Then, denoting $y_{m,n}^- = [y_{m,n-1} y_{m,n-2} \ldots]^T$ as the past history of the generic process y_m, we state that the ith process G-causes the jth process (conditional on the other k processes), if $y_{i,n}^-$ conveys information about $y_{j,n}$ above and beyond all information contained in $y_{j,n}^-$ and $Y_{k,n}^-$. This definition leads to perform a regression of the present of the target on the past of all processes, yielding to the prediction error $e_{j|ijk,n} = y_{j,n} - \mathbb{E}[y_{j,n}|Y_n^-]$, and on the past of all processes except the driver, yielding to the prediction error $e_{j|jk,n} = y_{j,n} - \mathbb{E}[y_{j,n}|y_{j,n}^-, Y_{k,n}^-]$. The prediction error variances resulting from these "full" and "restricted" regressions, $\lambda_{j|ijk} = \mathbb{E}[e_{j|ijk,n}^2]$ and $\lambda_{j|jk} = \mathbb{E}[e_{j|jk,n}^2]$ are then combined to obtain the definition of GC from y_i to y_j [8]:

$$F_{i \to j} = ln \frac{\lambda_{j|jk}}{\lambda_{j|ijk}}. \tag{2}$$

III. COMPUTATION OF GRANGER CAUSALITY

In this Section we describe five different approaches for the computation of $F_{i \to j}$ defined in (2). For all of them, given that $e_{j|ijk,n} = u_{j,n}$ the error variance of the full regression can be obtained as the jth diagonal element of the error covariance matrix, $\lambda_{j|ijk} = \mathbf{\Sigma}(j, j)$, solving (typically through OLS estimation) the identification problem (1). The

approaches differ in the computation of the error variance of the restricted regression, $\lambda_{j|jk}$, whose evaluation is described in the following.

A. GC based on double regression (GC-DR) [1]

The restricted model is formulated as follows:

$$y_{j,n} = \sum_{m=1, m \neq i}^{M} \sum_{k=1}^{p} c_{mk} y_{m,n-k} + e_{j|jk,n}, \tag{3}$$

from which the $(M-1)p$ coefficients c_{mk} and the residuals $e_{j|jk,n}$ can be obtained from a realization of length N of the multivariate process using OLS estimation. Then, the desired error variance $\lambda_{j|jk}$ is computed simply as the variance of the estimated residuals. Note that, with this approach, $\lambda_{j|jk}$ is computed after approximation of the whole past history of the observed processes with p lags, i.e., $y_{m,n}^- \approx y_{m,n}^p = [y_{m,n-1} \ldots y_{m,n-p}]$.

B. GC based on Autocovariance Function (GC-ACF) [4]

The restricted error variance can be obtained starting from the ACF of the process Y_n, which is a matrix containing the covariance between time-lagged variables taken from the entire process at each lag $k \geqslant 0$: $\mathbf{\Gamma}_k = \mathbb{E}[Y_n Y_{n-k}^T]$. Applying this definition to (1) it is easy to show that the ACF is related to the VAR parameters via the well-known Yule-Walker equations:

$$\mathbf{\Gamma}_k = \sum_{l=1}^{p} \mathbf{A}_l \mathbf{\Gamma}_{k-l} + \delta_{k0} \mathbf{\Sigma}, \tag{4}$$

where δ_{k0} is the Kronecher product. The ACF matrices $\mathbf{\Gamma}_0, ..., \mathbf{\Gamma}_{p-1}$ can be computed from the estimated VAR parameters $\mathbf{A}_k$ and $\mathbf{\Sigma}$ by solving a discrete-time Lyapunov equation as demonstrated in [9]. Then, the ACF matrices can be calculated recursively for any lag $k \geqslant p$ by repeatedly applying (4); the recursion is typically performed up to a lag q such that the spectral radius of the VAR process is smaller than 10^{-8} [4]. Then, the desired prediction error variance is computed exploiting a result which states that, given a scalar variable V and a vector variable W the variance of a linear regression of V on W is given by [5], [9]:

$$\sigma(V|W) = \sigma(V) - \Sigma(V; W)\Sigma(W)^{-1}\Sigma(V; W)^T, \tag{5}$$

where $\Sigma(\cdot)$, $\Sigma(\cdot; \cdot)$ and $\sigma(\cdot)$ denote respectively covariance matrix, cross-covariance matrix and variance. Here, (5) is exploited assuming $V = y_{j,n}$ and $W = [y_{j,n}^q Y_{k,n}^q]$, so that the covariance and cross-covariance matrices can be obtained by rearranging the elements of the ACF matrices $\mathbf{\Gamma}_0, ..., \mathbf{\Gamma}_q$, and the partial variance $\sigma(V|W)$ is equivalent to the desired error variance $\lambda_{j|jk}$ [5].

C. GC based on State Space models (GC-SS) [6]

The VAR model (1) can be represented equivalently as an SS model which relates the observed process Y_n to an unobserved state process Z_n through the equations [6]

$$Z_{n+1} = \mathbf{A} Z_n + \mathbf{K} E_n \tag{6}$$

$$Y_n = \mathbf{C}Z_n + E_n, \qquad (7)$$

where the innovations $E_n = Y_n - \mathbb{E}[Y_n|Y_n^-]$ are equivalent to the innovations U_n in (1) and thus have covariance matrix $\mathbf{\Phi} \equiv \mathbb{E}[E_n E_n^T] = \mathbf{\Sigma}$. This representation, typically denoted as "innovation form" SS model (ISS) [6], also evidences the Kalman Gain matrix $\mathbf{K}$, the state matrix $\mathbf{A}$ and the observation matrix $\mathbf{C}$, which can all be computed from the original VAR parameters in (1). The advantage of this representation is that it allows to form "submodels" which exclude one or more scalar processes from the observation equation (7) leaving the state equation (6) unaltered. In particular, the submodel excluding the driver process y_i has observation equation:

$$Y_{jk,n} = \mathbf{C}_{(jk)}Z_n + E_{jk,n}, \qquad (8)$$

where the subscript (jk) denotes the selection of the rows with indices j and k in a matrix. As demonstrated in [6], the submodel (6-8) is not in ISS form, but can be converted into ISS by solving a Discrete Algebric Riccati equation (DARE). Then, the covariance matrix of the innovations $E_{jk,n}$ includes the desired error variance $\lambda_{j|jk}$ as the first diagonal element.

D. Partial Conditioning GC (PCGC) [7]

Partial conditioning is an approach whereby GC is computed, for any assigned pair of driver and target processes, including only a subset of the M observed processes in the VAR representation. Specifically, to compute GC from the ith to the jth process in Y, a VAR model in the form of (1) is identified starting from a vector process which comprises y_i, y_j and the n_d processes which are deemed as most informative for the driver y_i. Such other processes, which are collected in the vector Y_c, are identified maximizing the mutual information between $y_{i,n}^-$ and $Y_{c,n}^-$ [7]; maximal mutual information is obtained, under the Gaussian assumption, when the matrix formulation of the partial variance in (5) is minimized [5]. Moreover, the number of conditioning processes is selected automatically finding the knee of the curve that measures the MI values as a function of the number of processes in Y_c (up to $M-1$). Then, a double regression limited to the subset of identified processes is performed defining the full and restricted regression models:

$$y_{j,n} = \sum_{m=1}^{n_d+2} \sum_{k=1}^{p} c_{mk}y_{m,n-k} + e_{j|ijc,n}, \qquad (9)$$

$$y_{j,n} = \sum_{m=1,m\neq i}^{n_d+2} \sum_{k=1}^{p} \tilde{c}_{mk}y_{m,n-k} + e_{j|jc,n}. \qquad (10)$$

After identification of these models through the OLS method, the variance of the residuals $\lambda_{j|ijc} = \mathbb{E}[e_{j|ijc,n}^2]$ and $\lambda_{j|jc} = \mathbb{E}[e_{j|jc,n}^2]$ are estimated and used to compute the GC measure

$$F_{i\to j}^C = ln\frac{\lambda_{j|jc}}{\lambda_{j|ijc}}. \qquad (11)$$

E. Partial Conditioning GC based on SS models (PCGC-SS)

Starting from Y_c and from the same n_d processes selected for the computation of PCGC, the idea here is to evaluate $F_{i\to j}^C$ with an SS approach. In this case the full model is described by (9) and the full error variance is $\lambda_{j|ijc} = \mathbb{E}[e_{j|ijc,n}^2]$. The error variance of the restricted regression is obtained considering an SS model with state and observation equations as follows:

$$Z_{ijc,n+1} = \mathbf{A}Z_{ijc,n} + \mathbf{K}E_{ijc,n} \qquad (12)$$

$$Y_{jc,n} = \mathbf{C}_{(jc)}Z_{ijc,n} + E_{jc,n} \qquad (13)$$

Also in this case, by solving a DARE equation is it possibile to bring the model (12)-(13) into an ISS form, so that the covariance matrix of the innovations $E_{jc,n}$ includes the desired $\lambda_{j|jc}$ as the first diagonal element.

F. Testing the significance of GC values

Since the multivariate GC is a measure of the information transferred directly (i.e., not through indirect paths) from the driver to the target process, the assessment of its statistical significance is useful establish the existence of a directed link between the two network nodes generating the driver and target dynamics. In this work, the significance of GC computed with the five methods described above was tested generating surrogate driver and target series which share the same power spectrum of the original time series but are otherwise uncorrelated [10]. Specifically, 100 sets of surrogate time series were generated, the GC estimated for each pair of processes was compared with a threshold set at the 95th percentile of its distribution on the surrogates, and was deemed at statistically significant if it exceeded the threshold.

IV. SIMULATION STUDY

Simulated multivariate time series ($M = 16$) were generated as realizations of a VAR(3) process fed by Gaussian noises with variance equal to 0.1. The simulated network had the tree structure of Fig. 1, previously used in another study [7]. 100 realizations of the process were generated under different values of the parameter K, defined as the ratio between the number of data samples and the number of AR coefficients to be estimated [11]. For each realization, the connections between nodes were randomly assigned from delay 1 to delay 3, with values of the coefficient equal to 0.5.
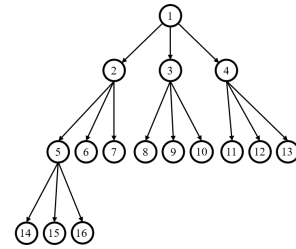

Fig. 1. Structure of the directed tree with 16 nodes used for the simulation.

For each realization of the process, the performances of the five methodologies were tested by computing the following

parameters [11]: (i) the mean absolute error (MAE), obtained as the average relative difference between the estimated GC value $\widehat{F}$ and the true GC value F assessed over the non-null links (it represents the average normalized bias in the estimation of the coupling strength over the active network links); (ii) false positive rate (FPR), false negative rate (FNR) and Area Under Curve (AUC) in the detection of the network structure (it represents the performance in detecting the network structure, where positive and negative outcomes are associated respectively with non-null and null links); (iii) computational time required for the entire estimation process for each method tested (average over K values and realizations).

V. RESULTS

MAE, FNR, FPR and AUC were subjected to a two-way repeated measures ANOVA considering as within factors K and the method used (TYPE). The results reported in Table 1 show that both the within-factors and their interactions have an effect on MAE, FNR, FPR and AUC parameters.

TABLE I
ANOVA RESULTS

	Parameters			
Factors	MAE	FNR	FPR	AUC
K	584**	610**	2.31	612**
TYPE	146**	277**	33.3**	284**
K x TYPE	66.1**	93.3**	5.72**	104**

ANOVA resutls (F values).** means $p < 0.001$.

Fig. 2 shows the distributions of the MAE assessed for the five measures computed in different conditions of data samples available (parameter K). For all methods, the MAE decreases and its distribution is less dispersed at increasing K, documenting that the availability of more data reduces the bias and the variance of GC estimates. The Tukey's post-hoc test highlights no significant differences between GC-DR and PCGC for K=5. In the most challenging condition (K=1.5) PCGC-SS and GC-DR exhibit the highest and lowest performance in terms of MAE.
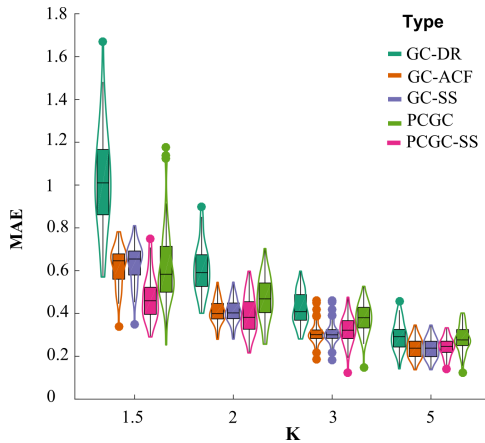

Fig. 2. Violin plot for the MAE. Dots represent the ouliers and the probability density function is estimated with the kernel density estimator.

Fig. 3 reports the mean and confidence intervals across realizations of FNR, FPR and AUC (percentage values) assessed for the five measures and for different values of K. While FPR assumes, for each method, the expected steady value of $\sim 5\%$, FNR shows a sharp decrease with the increase of K, reaching values close to zero for all methods when $K = 5$. In the challenging conditions $K = [1.5, 2]$ PCGC and PCGC-SS outperform the other methods with no significant differences between them, as highlighted by the Tukey's post-hoc test for each condition analyzed. All methods show an increase of the AUC value with K, showing that the performance in recovering the network structure strongly depends on the amount of data. While with $K = 5$ all methods exhibit AUC $\sim 100\%$, in the conditions $K = [1.5, 2]$, PCGC and PCGC-SS show AUC values $\sim 20\%$ higher than the other methods. Also in this case, the two methods behave in a very similar way.
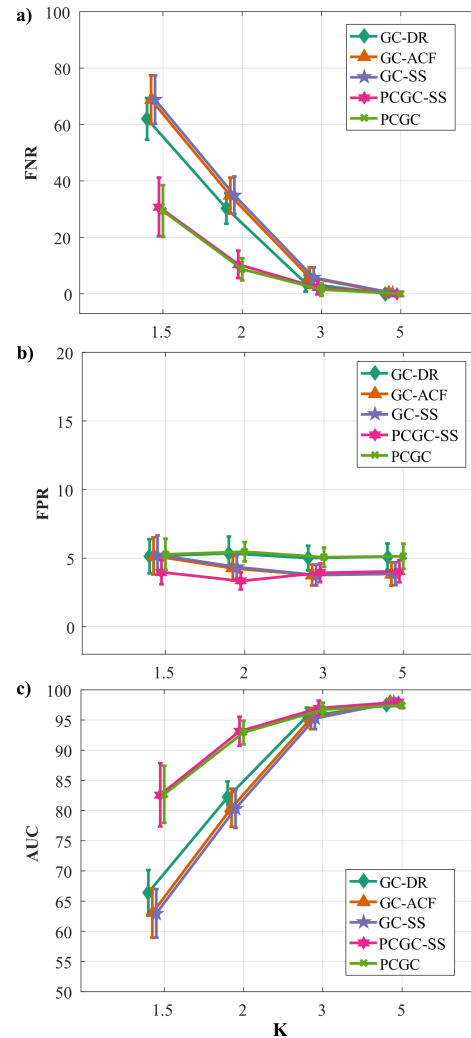

Fig. 3. Plot of means of FNR, FPR and AUC index for the interaction KxTYPE (respectively panel a, b and c). The diagram shows the mean values of the indeces for different values of K and different methods. The bars represent the 95% confidence interval.

The average computational times of the five GC estimation methods were calculated after their implementation in the

Matlab® environment with parallel computing on a PC with a 6 cores Intel Xeon (clock speed $3.7GHz$), $128GB$ RAM DDR4. Results show that the fastest method is GC-DR (computational time of 0.73 sec), followed by PCGC (4 sec), GC-ACF (76.9 sec), GC-SS (151.9 sec) and PCGC-SS (274.9 sec). Notably, the implementation of PCGC-SS almost doubled the computational time compared to the exclusive use of SS models, and increased it of more than 60 times compared to simple partial conditioning.

VI. DISCUSSION AND CONCLUSIONS

The results of our simulation show that the bias and variance of the estimates of the multivariate GC strongly depend on the amount of data samples available relative to the number of model coefficients regulated by the parameter K (Fig. 2). This aspect has already been explored in the context of EEG signals [11], [12] and the present results confirm that a solution to a linear regression problem with OLS is not feasible when the amount of observations is close to the amount of VAR parameters to be estimated (K=1.5). The deterioration of the performances with lower values of K is observed also as regards the detection of the network structure (Fig. 3), confirming previous studies reporting an increased rate of false negative detection of causality in unfavorable conditions of data samples available [12].

As anticipated in the Introduction, the high bias and variance exhibited by the GC-DR method for each value of K can be ascribed to the fact that, even if the full model is of finite order, the restricted model is theoretically of infinite order and thus it cannot be identified properly [3]. This problem is tackled successfully by the GC-ACF and GC-SS methods, which offer more appropriate identification of the restricted model (respectively, through exploitation of the ACF computed up to lags much larger than the model order [4] and through a closed form parametric representation of the restricted model [9]) and obtain lower values of MAE than the classical DR approach (Fig. 2). On the other hand, PCGC performs worse than GC-ACF and GC-SS in favorable conditions of data samples. While this is expected because PCGC is based on a double regression, its accuracy improves when K decreases, up to the point that it outperforms GC-ACF and GC-SS when K=1.5 (Fig.2). This behavior, related to the prior selection of a subset of processes [7], is even improved by the use of PCGC-SS, which inherits the advantages of proper VAR formulation and partial conditioning and yield the best accuracy in the estimation of GC values, especially in conditions of data paucity (Fig.2, K=1.5). The advantages of the partial conditioning scheme appear more evidently when the focus is on the network structure rather than on the GC values (Fig. 3). In this analysis, PCGC and PCGC-SS exhibit overlapping trends with better performance compared to the other methods in terms of FNR and AUC. This highlights how conditioning on a small number of processes allows a more correct detection of network links.

The reported results document how partial conditioning can be efficiently integrated in the context of SS modelling,

providing a new tool for GC analysis. However, the advantage provided in the accuracy of GC estimates comes at the cost of higher computational times. This drawback, related to the need of implementing several SS models and submodels, may become critical when analyses need to be performed on large networks (such as brain networks) without the proper computational power.

In conclusion, we showed that an accurate estimation of the multivariate GC should account both for appropriate modeling schemes (like the SS framework) and for strategies to limit the number of variables involved into practical estimation (like the partial conditioning). The use of the PCGC-SS estimator can provide, when computational time is not a critical issue, an accurate way to assess the structure and the link strength in dynamical networks explored in unfavorable analysis conditions. A possible application, desirable both in neuroscience and in the field of network physiology, is the evaluation of multiscale GC where computation is challenging at coarse time scales due to substantial reduction of the time series length [8]. Future works will be directed to test these estimators on real physiological signals in the context of information dynamics [13], and to compare them with other variable reduction methods such as those based on sparse identification [11], [14].

REFERENCES

[1] J.F Geweke,"Measures of conditional linear dependence and feedback between time series", Journal of the American Statistical Association, vol.79, pp. 907–915, 1984.
[2] R.P.Bartsch,"Network physiology: how organ systems dynamically interact",PloS one, vol.10, 2015.
[3] L. Faes,"On the interpretability and computational reliability of frequency-domain Granger causality", F1000Research, vol.6, 2017.
[4] L.Barnett,"The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference", Journal of neuroscience methods, vol. 223, pp. 50–68, 2014.
[5] L. Barnett,"Granger causality and transfer entropy are equivalent for Gaussian variables", Phys. Rev. L.,vol.10, pp. 238701, 2009.
[6] L. Barnett,"Granger causality for state-space models",Physical Review E, vol.91, pp. 040101, 2015.
[7] D.Marinazzo,"Causal information approach to partial conditioning in multivariate data sets", Computational and mathematical methods in medicine, vol. 2012, 2012.
[8] L.Faes,"Multiscale granger causality",Physical Review E, vol. 96,pp. 042150, 2017.
[9] L.Faes,"Information decomposition in multivariate systems: definitions, implementation and application to cardiovascular networks",Entropy, vol.19, 2017.
[10] T. Schreiber, "Improved surrogate data for nonlinearity tests," Physicaln review letters, vol. 77, no. 4, p. 635, 1996.
[11] Y. Antonacci,"Single-trial Connectivity Estimation through the Least Absolute Shrinkage and Selection Operator", 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6422–6425, 2019.
[12] Y. Antonacci,"Estimating brain connectivity when few data points are available: Perspectives and limitations",2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4351–4354, 2017.
[13] M.Zanetti,"Information dynamics of the brain, cardiovascular and respiratory network during different levels of mental stress",Entropy, vol. 21, pp. 275,2019.
[14] E. Siggiridou, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model",IEEE-TBME, vol.64, pp.1759–1773, 2015.