

[Home](#) [Search](#) [Collections](#) [Journals](#) [About](#) [Contact us](#) [My IOPscience](#)

Synergy and redundancy in the Granger causal analysis of dynamical networks

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2014 New J. Phys. 16 105003

(<http://iopscience.iop.org/1367-2630/16/10/105003>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 157.193.116.31

This content was downloaded on 07/10/2014 at 07:12

Please note that [terms and conditions apply](#).

Synergy and redundancy in the Granger causal analysis of dynamical networks

Sebastiano Stramaglia^{1,2,3}, Jesus M Cortes^{2,3} and Daniele Marinazzo⁴

¹ Center of Innovative Technologies for Signal Detection and Processing TIRES, Dipartimento di Fisica, Università di Bari, Italy

² Computational Neuroimaging Lab, Biocruces Health Research Institute, Cruces University Hospital, Barakaldo, Spain

³ Ikerbasque, The Basque Foundation for Science, Bilbao, Spain

⁴ Faculty of Psychology and Educational Sciences, Department of Data Analysis, Ghent University, Henri Dunantlaan 1, B-9000, Ghent, Belgium

E-mail: sebastiano.stramaglia@ba.infn.it

Received 20 March 2014, revised 16 July 2014

Accepted for publication 30 July 2014

Published 6 October 2014

New Journal of Physics **16** (2014) 105003

doi:[10.1088/1367-2630/16/10/105003](https://doi.org/10.1088/1367-2630/16/10/105003)

Abstract

We analyze, by means of Granger causality (GC), the effect of synergy and redundancy in the inference (from time series data) of the information flow between subsystems of a complex network. While we show that fully conditioned GC (CGC) is not affected by synergy, the pairwise analysis fails to prove synergetic effects. In cases when the number of samples is low, thus making the fully conditioned approach unfeasible, we show that partially conditioned GC (PCGC) is an effective approach if the set of conditioning variables is properly chosen. Here we consider two different strategies (based either on informational content for the candidate driver or on selecting the variables with highest pairwise influences) for PCGC and show that, depending on the data structure, either one or the other might be equally valid. On the other hand, we observe that fully conditioned approaches do not work well in the presence of redundancy, thus suggesting the strategy of separating the pairwise links in two subsets: those corresponding to indirect connections of the CGC (which should thus be excluded) and links that can be ascribed to redundancy effects and, together with the results from the fully connected approach, provide a better description of the causality pattern in the presence of redundancy. Finally we apply these methods



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

to two different real datasets. First, analyzing electrophysiological data from an epileptic brain, we show that synergetic effects are dominant just before seizure occurrences. Second, our analysis applied to gene expression time series from HeLa culture shows that the underlying regulatory networks are characterized by both redundancy and synergy.

Keywords: Granger causality, synergy and redundancy, epilepsy, gene regulatory networks

1. Introduction

Living organisms can be modeled as an ensemble of complex physiological systems, each with its own regulatory mechanism and all continuously interacting between them [1]. Therefore, inferring the interactions within and between these modules is a crucial issue. Over the past few years, the interaction structure of many complex systems has been mapped in terms of networks, which have been successfully studied using tools from statistical physics [2]. Dynamical networks have modeled physiological behavior in many applications; examples range from networks of neurons [3], genetic networks [4], protein interaction nets [5], and metabolic networks [6–8].

The inference of dynamical networks from time series data is related to the estimation of the information flow between variables [9]; see also [10, 11]. Granger causality (GC) [12, 13] has emerged as a major tool to address this issue. This approach is based on prediction: if the prediction error of the first time series is reduced by including measurements from the second one in the linear regression model, then the second time series is said to have a Granger causal influence on the first one. It has been shown that GC is equivalent to transfer entropy [14] in the Gaussian approximation [15] and for other distributions [16]. See [17] for a discussion about the applicability of this notion in neuroscience and [18] for a discussion on the reliability of GC for continuous dynamical processes. It is worth stressing that several forms of coupling may mediate information flow in the brain, see [19, 20]. The combination of GC and complex networks theory is also a promising line of research [21].

The pairwise Granger analysis (PWGC) consists of assessing GC between each pair of variables independently from the rest of the system. It is well known that PWGC cannot disambiguate direct and indirect interactions among variables. The most straightforward extension, the conditioning approach [22], removes indirect influences by evaluating to what extent the predictive power of the driver on the target decreases when the conditioning variable is removed. It must be noted however that, even though its limitations are well known, the PWGC approach is still used in situations where the number of samples is limited and a fully conditioned (CGC) approach is unfeasible. As a convenient alternative to this suboptimal solution, a partially conditioned approach, consisting of conditioning on a small number of variables, chosen as the most informative one for the driver node, has been proposed [23]; this approach leads to results very close to those obtained with a fully conditioned analysis that are even more accurate in the presence of a small number of samples [24]. We remark that the use of partially conditioned GC (PCGC) may also be useful in non-stationary conditions whereas the GC pattern must be estimated on short time windows.

However, sometimes a CGC approach can encounter conceptual limitations in addition to the practical and computational ones: in the presence of redundant variables, the application of

the standard analysis leads to an underestimation of influences [25]. Redundancy and synergy are intuitive, yet elusive, concepts that have been investigated in different fields, from pure information theory [26–28], to machine learning [29] and neural systems [30, 31], with definitions that range from purely operative to conceptual. When analyzing interactions in a multivariate time series, redundancy may arise if some channels are all influenced by another signal that is not included in the regression; another source of redundancy may be the emergence of synchronization in a subgroup of variables, without the need for an external influence. Redundancy manifests itself through a high degree of correlation in a group of variables, both for instantaneous and lagged influences. Several approaches have been proposed to reduce dimensionality in multivariate sets that eliminate redundancy, which rely on generalized variance [32], principal components analysis [33], or GC itself [34].

A complementary concept to redundancy is synergy. The synergetic effects that we address here, related to the analysis of dynamical influences in a multivariate time series, are similar to those encountered in sociological and psychological modeling, where *suppressors* are variables that increase the predictive validity of another variable after its inclusion into a linear regression equation [35]; see [36] for examples of easily explainable suppressor variables in multiple regression research. Redundancy and synergy have been further connected to information transfer in [37], where an expansion of the information flow has been proposed to prove redundant and synergetic multiplets of variables. Other information-based approaches have also addressed the issue of collective influence [28, 38]. The purpose of this paper is to provide evidence that, in addition to the problem related to indirect influence, PWGC shows another relevant pitfall: it fails to detect synergetic effects in the information flow. In other words, it does not account for the presence of subsets of variables that provide some information about the future of a given target only when all the variables are used in the regression model. We remark that, because it processes the system as a whole, CGC evidences synergetic effects; when the number of samples is low, PCGC can also detect synergetic effects after an adequate selection of the conditioning variables.

The paper is organized as follows. In the next section, we briefly recall the concepts of GC and PCGC. In section 3, we describe some toy systems that illustrate how redundancy can affect the results of CGC, whereas indirect interactions and synergy are the main problems inherent in PWGC. In section 4, we provide evidence of synergetic effects in epilepsy: we analyze EEG recordings of an epileptic patient corresponding to ten seconds before the seizure onset; we show that the two contacts that constitute the putative seizure onset act as synergetic variables driving the rest of the system. The pattern of information transfer proves the actual seizure onset only when synergy is correctly considered.

In section 5, we propose an approach that combines PWGC and GCC to prove the pairwise influences are only because of redundancy and not recognized by CGC. The conditioned GC pattern is used to partition the pairwise links in two sets: those that are indirect influences between the measured variables, according to CGC, and those that are not explained as indirect relationships. The unexplained pairwise links, presumably due to redundancy, are thus retained to complement the information transfer pattern discovered by CGC. In cases where the number of samples is so low that a fully multivariate approach is unfeasible, PCGC may be applied instead of CGC. We also address here the issue of variable selection for PCGC, and consider a novel strategy for the selection of variables: for each target variable, one selects the variables sending the highest amount of information to that node, as revealed by a pairwise analysis. By construction, this new selection strategy works more efficiently when the interaction graph has a

tree structure: indeed, in this case, conditioning on the parents of the target node ensures that indirect influences will be removed. In the epilepsy example, the selection based on the mutual information with the candidate driver provides results closer to those obtained by CGC. Finally we apply the proposed approach to time series of gene expressions, extracted from a data set from the HeLa culture. Section 6 summarizes our conclusions.

2. Insights into Granger causality

GC is a powerful and widespread data-driven approach used to determine whether and how two time series exert direct dynamical influences on each other [39]. A convenient nonlinear generalization of GC has been implemented in [40], exploiting the kernel trick, which makes computation of dot products in high-dimensional feature spaces possible using simple functions (kernels) defined on pairs of input patterns. This trick allows the formulation of nonlinear variants of any algorithm that can be cast in terms of dot products, for example support vector machines [41]. Hence, in [40], the idea is still to perform linear GC but in a space defined by the nonlinear features of the data. This projection is conveniently and implicitly performed through kernel functions [42] and a statistical procedure is used to avoid over-fitting.

Quantitatively, let us consider n time series $\{x_\alpha(t)\}_{\alpha=1,\dots,n}$; the lagged state vectors are denoted

$$X_\alpha(t) = (x_\alpha(t-m), \dots, x_\alpha(t-1)),$$

where m is the order of the model (window length). Let $\epsilon(x_\alpha | \mathbf{X})$ denote the mean squared error prediction of x_α on the basis of all the vectors $\mathbf{X} = \{X_\beta\}_{\beta=1}^n$ (corresponding to the kernel approach described in [43]). The multivariate GC index $\delta_{mv}(\beta \rightarrow \alpha)$ is defined as follows: consider the prediction of x_α on the basis of all variables except X_β and the prediction of x_α using all variables, then the GC is the (normalized) variation of the error in the two conditions, i.e.,

$$\delta_{mv}(\beta \rightarrow \alpha) = \log \frac{\epsilon(x_\alpha | \mathbf{X} \setminus X_\beta)}{\epsilon(x_\alpha | \mathbf{X})}; \quad (1)$$

In [44], it has been shown that not all the kernels are suitable to estimate GC. Two important classes of kernels that can be used to construct nonlinear GC measures are the *inhomogeneous polynomial kernel* (whose features are all the monomials in the input variables up to the p th degree; $p = 1$ corresponds to linear GC) and the *Gaussian kernel*.

The PCGC is given by:

$$\delta_{bv}(\beta \rightarrow \alpha) = \log \frac{\epsilon(x_\alpha | X_\alpha)}{\epsilon(x_\alpha | X_\alpha, X_\beta)}. \quad (2)$$

The PWGC is defined as follows. Let \mathbf{Y} denote the variables in \mathbf{X} , excluding X_α and X_β . Then equation (1) can be written as:

$$\delta_c^{\mathbf{Y}}(\beta \rightarrow \alpha) = \log \frac{\epsilon(x_\alpha | X_\alpha, \mathbf{Y})}{\epsilon(x_\alpha | X_\alpha, X_\beta, \mathbf{Y})}. \quad (3)$$

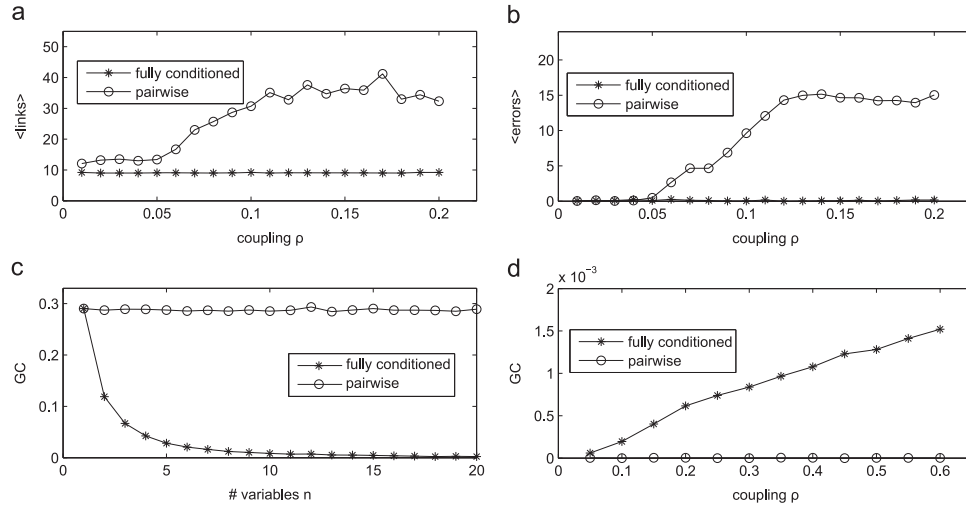


Figure 1. (a) The average number of links as a function of the coupling ρ , over 100 runs of 2000 time points, retrieved by PWGC and CGC on the coupled map lattice described in the text, equations (4)–(5). (b) On the coupled map lattice, the average error (sum of type I errors and type II errors in the recovery of causal interactions) by PCGC, obtained using the IB and PB strategy, is plotted versus the coupling ρ . Errors are averaged over 100 runs of 2000 time points. (c) For the example dealing with redundancy, CGC and PWGC are plotted versus the number of variables. Results are averaged over 100 runs of 1000 time points. (d) For the example dealing with synergy, CGC and PWGC are plotted versus the coupling ρ . Results are averaged over 100 runs of 1000 time points.

When \mathbf{Y} is only a subset of the total number of variables in \mathbf{X} not containing X_α and X_β , then $\delta_c^{\mathbf{Y}}$ is called the PCGC. In [23], the set \mathbf{Y} is chosen as the most informative for X_β . Here, we will also consider an alternative strategy: fixed a small number k , we select $\mathbf{Y} = \{\mathbf{X}_\gamma\}_{\gamma=1}^k$ as the k variables with the maximal pairwise GC $\delta_{bv}(\gamma \rightarrow \alpha)$ w.r.t. that target node, excluding X_β .

3. Examples

In this section, we provide some typical examples to note possible problems that pairwise and fully conditioned analysis may encounter.

3.1. Indirect GC among measured variables

We consider the following lattice of ten unidirectionally coupled noisy logistic maps, with

$$x_1(t) = f(x_1(t-1)) + 0.01\eta_1(t), \quad (4)$$

and

$$x_i(t) = (1 - \rho)f(x_i(t-1)) + \rho f(x_{i-1}(t-1)) + 0.01\eta_i(t), \quad (5)$$

with $i = 2, \dots, 10$. Variables η are unit variance Gaussian noise terms. The transfer function is given by $f(x) = 1 - 1.8x^2$. In this system, the first map is evolving autonomously, whereas the other maps are influenced by the previous maps with coupling ρ , thus forming a cascade of interactions. In figure 1(a), we plot, as a function of ρ , the number of GC interactions found by PWGC and CGC using the method described in [25] with the inhomogeneous polynomial

kernel of degree two. The CGC output is the correct one (nine links), whereas the PWGC output also accounts for indirect influences and therefore fails to provide the underlying network of interactions. In this example we also tested PCGC (see figure 1(b)). We considered only one conditioning variable, chosen according to the two strategies described previously. First, we consider the most informative w.r.t. the candidate driver, as described in [23]; we call this *information-based* (IB) strategy. Second, we choose the variable characterized by the maximal pairwise influence to the target node, a *pairwise-based* (PB) rule. The PB strategy yields the correct result in this example, whereas the IB strategy fails when only one conditioning variable is used and requires more than one conditioning variable to provide the correct output. This occurrence is due to the tree topology of the interactions in this example, which favors PB selecting by construction the parents of each node.

3.2. Redundancy due to a hidden source

Here, we show how redundancy constitutes a problem for CGC. Let $h(t)$ be a zero mean and unit variance hidden Gaussian variable, influencing n variables $x_i(t) = h(t - 1) + s\eta_i(t)$, and let $w(t) = h(t - 2) + s\eta_0(t)$ be another variable influenced by h but with a larger delay. The $\{\eta\}$ variables are unit variance Gaussian noise and s controls the noise level. In figure 1(c), we depict both the linear PWGC and the linear CGC from one of the x s , to w (note that h is not used in the regression model). As n increases, the conditioned GC vanishes as a consequence of redundancy. The GC relation that is found in the pairwise analysis is not revealed by CGC because $\{x\}$ variables are maximally correlated and thus x_i drives w only in the absence of any other variables.

The correct way to describe the information flow pattern in this example, where the true underlying source h is unknown, is that all $\{x\}$ variables are sending the same information to w , i.e., that variables $\{x\}$ constitute a redundant multiplet w.r.t. the causal influence to w . This pattern follows from observing that for all x s , CGC vanishes whereas PWGC does not vanish. This example shows that, in presence of redundancy, the CGC pattern alone is not sufficient to describe the information flow pattern of the system and PWGC should also be taken into account.

3.3. Synergetic contributions

Let us consider three unit variance i.i.d. Gaussian noise terms x_1 , x_2 , and x_3 . Let

$$x_4(t) = 0.1(x_1(t - 1) + x_2(t - 1)) + \rho x_2(t - 1)x_3(t - 1) + 0.1\eta(t).$$

Considering the influence $3 \rightarrow 4$, the CGC reveals that 3 influences 4, whereas PWGC fails to detect this causal relationship—see figure 1(d), in which we use the method described in [25] with the inhomogeneous polynomial kernel of degree two, where x_2 is a suppressor variable for x_3 w.r.t. the influence on x_4 . This example shows that PWGC fails to detect synergetic contributions. We note that use of nonlinear GC is mandatory in this case to verify the synergy between x_2 and x_3 .

3.4. Redundancy due to synchronization

As another example, we consider a toy system made of five variables $\{x_i\}$. The first four variables constitute a multiplet made of a fully coupled lattice of noisy logistic maps with

coupling ρ , evolving independently of the fifth variable. The fifth variable is influenced by the mean field from the coupled map lattice. The equations are, for $i = 1, \dots, 4$:

$$x_i(t) = (1 - \rho)f(x_i(t-1)) + \rho \sum_{j=1, j \neq i}^4 f(x_j(t-1)) + 0.01\eta_i(t), \quad (6)$$

and

$$x_5(t) = \sum_{i=1}^4 \frac{x_i(t-1)}{8} + \eta_5(t), \quad (7)$$

where η denotes unit variance Gaussian noise terms. Increasing the coupling ρ among the variables in the multiplet $\{x_1, x_2, x_3, x_4\}$, the synchronization of these variables (measured by Pearson correlations) increases and they become nearly synchronized for ρ greater than 0.1 (perfect synchronization cannot be achieved due to the noise terms); redundancy, in this example, arises due to the complex inherent dynamics of the units. In figure 2, we depict both the causality from one variable in the multiplet (x_1 ; the same results hold for x_2, x_3 , and x_4) to x_5 , and the causality between pairs of variables in the multiplet: both linear and nonlinear PWGC and CGC are shown for the two quantities.

Concerning the causality $x_1 \rightarrow x_5$, we note that, for low coupling, both PWGC and CGC, with a linear or nonlinear kernel, correctly detect the causal interaction. Around the transition to synchronization, in a window centered at $\rho = 0.05$, all algorithms fail to detect the causality $x_1 \rightarrow x_5$. In the *almost* synchronized regime, $\rho > 0.1$, CGC continues to fail due to redundancy, whereas the PWGC correctly provides the causal influence, using both the linear and the nonlinear algorithm.

As far as the causal interactions within the multiplet are concerned, we note that using the linear approach, we obtain small values of causality just at the transition, whereas we obtain zero values far from the transition. Using the nonlinear algorithm, which is the correct one in this example because the system is nonlinear, we obtain a nonzero causality among the variables in the multiplet, using both PWGC or CGC: the resulting curves are non-monotonous, as one may expect, due to the inherent nonlinear dynamics. For $\rho > 1$, nonzero GC is observed because of the noise that prevents the system to go in the complete synchronized state.

This example again shows that, in the presence of redundancy, one should consider both CGC and PWGC results. Moreover, it also shows how nonlinearity may render extremely difficult the inference of interactions: in this system, there is a range of values, corresponding to the onset of synchronization, in which all methods fail to provide the correct causal interaction.

4. Synergetic effects in the epileptic brain

As a real example, we consider intracranial EEG recordings from a patient with drug-resistant epilepsy with an implanted array of 8×8 cortical electrodes (CE) and two depth electrodes (DE) with six contacts each. The data are available in [45] and further details on the dataset are given in [46]. Data were sampled at 400 Hz. Here we consider a portion of data recorded in the preictal period, 10 seconds preceding the seizure onset. To handle this data, we use linear GC with m equal to five. In figure 3, we depict the PWGC between DEs (panel a), from DEs to CEs (panel b), between CEs (panel c), and from CEs to DEs (panel d). We note a complex pattern of bivariate interactions among CEs, whereas the first DE seems to be the subcortical drive to the cortex. We note that there is no PWGC from the last two contacts of the second DE (channels

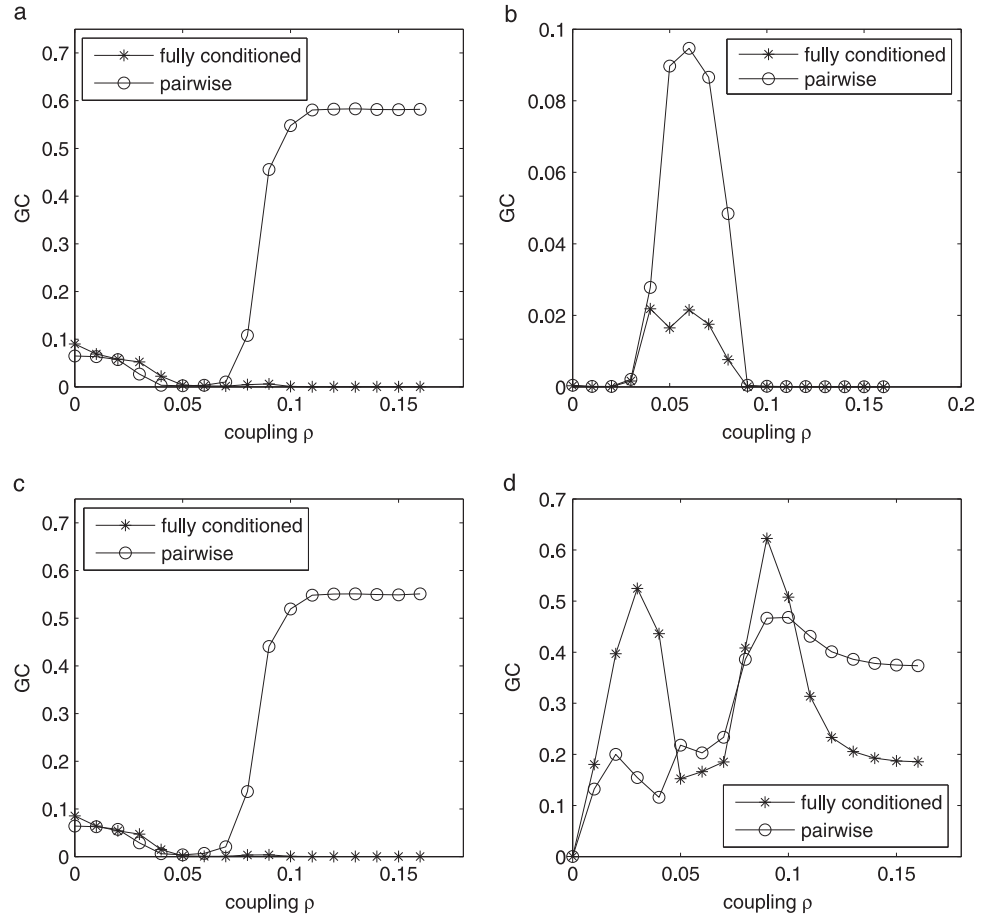


Figure 2. (a) The CGC and PWGC of the causal interaction $x_1 \rightarrow x_5$ are plotted as functions of the coupling ρ for the example dealing with redundancy due to synchronization, equations (6)–(7). The linear algorithms are used here and results are averaged over 100 runs of 2000 time points. (b) The CGC and PWGC of the causal interaction between two variables in the multiplet are plotted as functions of the coupling ρ for the example dealing with redundancy due to synchronization. The linear algorithms are used here and results are averaged over 100 runs of 2000 time points. (c) The nonlinear CGC and PWGC of the causal interaction $x_1 \rightarrow x_5$ are plotted as a function of the coupling ρ for the example dealing with redundancy due to synchronization. The algorithm with the polynomial kernel of order 2 is used here and results are averaged over 100 runs of 2000 time points. (d) The nonlinear CGC and PWGC of the causal interaction between two variables in the multiplet are plotted as a function of the coupling ρ for the example dealing with redundancy due to synchronization. The algorithm with the polynomial kernel of order 2 is used here and results are averaged over 100 runs of 2000 time points.

11 and 12) to CEs and neither to the contacts of the first DE. In figure 4, we depict the CGC among DEs (panel a), from DEs to CEs (panel b), among CEs (panel c), and from CEs to DEs (panel d). The scenario in the conditioned case is clear: contacts 11 and 12, from the second DE, are the drivers both for the cortex and for the subcortical region associated to the first DE. These two contacts can be then associated to the seizure onset zone (SOZ). The PWGC strength among CEs is due to redundancy, as these are all driven by the same subcortical source.

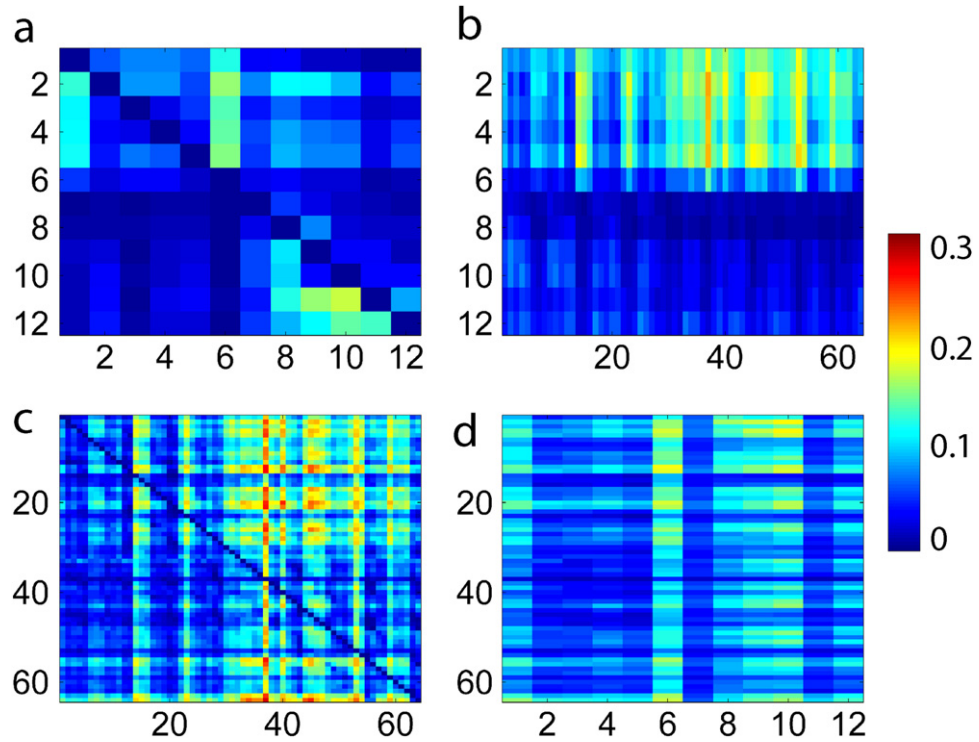


Figure 3. The PWGC is depicted for the epilepsy data. (a) PWGC between the contacts of the two DEs. (b) PWGC from DEs to CEs. (c) PWGC between CEs. (d) PWGC from CEs to DEs.

Because contact 12 is also driving contact 11 (see figure 4(a)), we conclude that contact 12 is the closest to the SOZ and that contact 11 is a *suppressor* variable for it, because it is necessary to include it in the regression model to verify the influence of contact 12 on the rest of the system. Conversely, contact 12 acts as a suppressor for contact 11. We stress that the influence from contacts 11 and 12 to the rest of the system emerges only in the CGC and is neglected by PWGC: these variables are synergetically influencing the dynamics of the system. To our knowledge, this is the first time that synergetic effects are found in relation with epilepsy.

We also apply PCGC on this data using one conditioning variable. The results are depicted in figure 5: using the IB strategy, we obtain a pattern very close to the one from CGC, whereas this is not the case of PB. These results seem to suggest that IB works better in the presence of redundancy; however, we do not offer arguments to claim that this is a general rule. It is worth mentioning that in the presence of synergy, the selection of variables for partial conditioning is equivalent to the search of suppressor variables.

5. Combination of pairwise and conditioned Granger causality

In the preceding sections, we have shown that CGC encounters issues resulting in poor performance in presence of redundancy, and that information about redundancy may be obtained from the PWGC pattern. Here we develop a strategy to combine the two approaches: some links inferred from PWGC are retained and added to those obtained from CGC. The

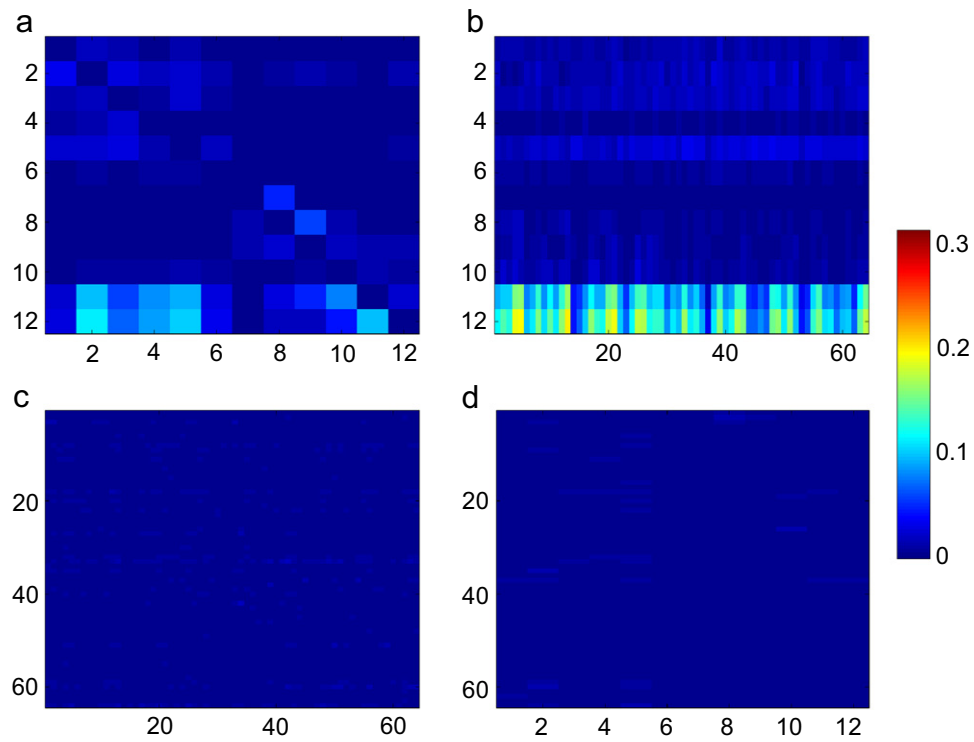


Figure 4. The CGC is depicted for the epilepsy data. (a) CGC between the contacts of the two DEs. (b) CGC from DEs to CEs. (c) CGC between CEs. (d) CGC from CEs to DEs.

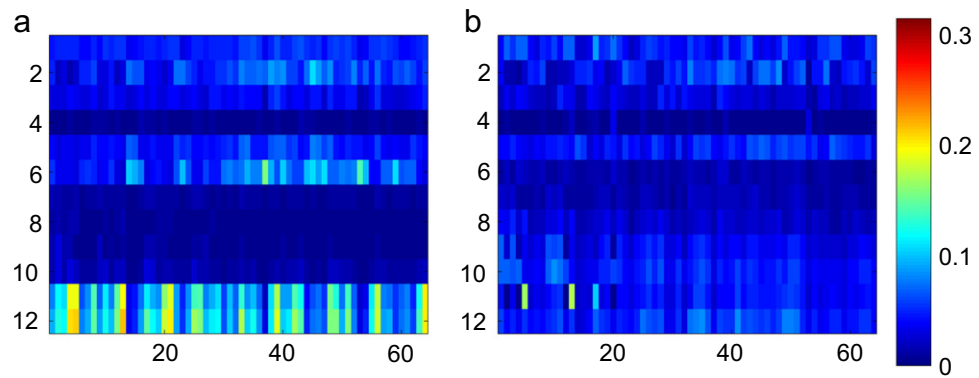


Figure 5. The PCGC is depicted for the epilepsy data. (a) PCGC from DEs to CEs, with IB strategy for variable selection. Note that the influences from DE11 are conditioned here on DE12, and the influences from DE11 are conditioned on DE12, thus showing that these variables are suppressors among themselves. (b) PCGC from DEs to CEs, with PB strategy for variable selection.

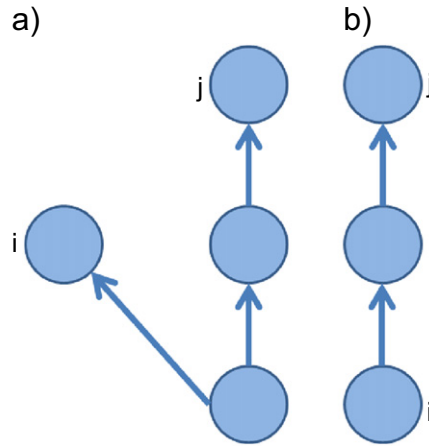


Figure 6. (a) The indirect influence $i \rightarrow j$ corresponding to a nonzero element of the matrix $\Delta^2 \Delta^\top$. If $\Delta^2 \Delta^\top(i, j) \neq 0$, then a common source influences i and j but with different lags. (b) The indirect causality $i \rightarrow j$ corresponding to nonzero elements of the matrix Δ^2 . If $\Delta^2(i, j) \neq 0$, then a third node acts as a mediator of the interaction $i \rightarrow j$.

PWGC links that are discarded are those that can be derived as indirect links from the CGC pattern. In the following, we describe the proposed approach in detail.

Let Δ denote the matrix of influences from CGC (or PCGC). Let Δ^* denote the matrix from PWGC. Nonzero elements of Δ and Δ^* correspond to the estimated influences. Let these matrices be evaluated using a model of order m .

The matrix

$$M_{\alpha\beta} = \Delta^\alpha (\Delta^\top)^\beta$$

contains paths of length $\alpha + \beta$ with delays in the range $[-\beta m + \alpha, \dots, -\beta + \alpha m]$. Indeed:

$$M_{\alpha\beta}(i, j) = \sum_{i_1} \sum_{i_2} \dots \sum_{i_{\alpha+\beta-1}} \Delta(i, i_1) \Delta(i_1, i_2) \dots \Delta(j, i_{\alpha+\beta-1}); \quad (8)$$

because all elements of Δ are non-negative, it follows that $M_{\alpha\beta}(i, j)$ is non-vanishing if, and only if it is possible, in matrix Δ , to go from node i to node j moving β steps backward and α steps forward, where a step is allowed if the corresponding element of Δ is not zero. Therefore, the nonzero elements of the matrix $M_{\alpha\beta}$ describe a situation where two nodes receive a common input from a third node that is α steps backward in time from one node and β steps backward in time from the other node. In other words, if the element $M_{\alpha\beta}(i, j)$ does not vanish, then there exists an indirect interaction between nodes i and j due to a common input. The circuit corresponding to M_{21} is represented in figure 6(a): if the element $M_{2,1}(i, j)$ is non-vanishing, then i and j are connected as shown in figure 6(a).

Because the order of the model is m , a simple comparison between the delays from the common source to i and j demonstrates that the indirect influence corresponding to the nonzero element $M_{\alpha,\beta}(i, j)$ might be detected by PWGC only if

$$[-\beta m + \alpha, \dots, -\beta + \alpha m] \cap [1, \dots, m] \neq \emptyset;$$

this is equivalent to

$$\frac{\beta + 1}{m} \leq \alpha \leq (\beta + 1)m.$$

Now, the matrix $F_\alpha = \Delta^\alpha$, with $\alpha \geq 1$, contains paths of length α with delays in the range $[\alpha, \dots, \alpha m]$, indeed:

$$F_\alpha(i, j) = \sum_{i_1} \sum_{i_2} \dots \sum_{i_{\alpha-1}} \Delta(i, i_1) \Delta(i_1, i_2) \dots \Delta(i_{\alpha-1}, j); \quad (9)$$

Any nonzero element of the matrix $F_\alpha(i, j)$ describes an indirect causal interaction between nodes i and j , where i sends information to j through a cascade of α links: $i \rightarrow i_1, i_1 \rightarrow i_2, \dots, i_{\alpha-1} \rightarrow j$. The circuit corresponding to F_2 is depicted in figure 6(b). The indirect causal interaction $i \rightarrow j$, corresponding to the nonzero element $F_\alpha(i, j)$, might be detected by PWGC if $\alpha \leq m$.

Let us now consider the matrix

$$B = \sum_{\alpha, \beta} M_{\alpha\beta} + \sum_{\alpha'=1}^m F_{\alpha'},$$

where the first sum is over pairs $\{\alpha, \beta\}$ satisfying $\frac{\beta+1}{m} \leq \alpha \leq (\beta+1)m$. If B_{ij} is non-vanishing, then according to CGC, there is an indirect causal interaction between i and j : therefore, PWGC might misleadingly reveal such interact, considering it a direct one. In the approach previously described, we discard (as indirect) the links found by PWGC for which $B(i, j) \neq 0$. Therefore, in the pairwise matrix Δ^* , we set to zero all the elements such that $B_{ij} > 0$ (pruning). The resulting matrix Δ^* contains links which cannot be interpreted as indirect links of the multivariate pattern, and will be retained and ascribed to redundancy effects,

For $m = 1$, we have that the only terms in the first sum are those with $\alpha = \beta + 1$, so the first non-trivial terms are

$$B_1 = \Delta + \Delta^2 \Delta^\top.$$

For $m = 2$, the simplest terms are:

$$B_2 = \Delta + \Delta^2 + \Delta \Delta^\top.$$

Because, due to the finite number of samples, a mediated interaction is more unlikely to be detected (by the pairwise analysis) if it corresponds to a long path, we limit the sum in the matrix B to the simplest terms.

As a toy example to illustrate an application of the proposed approach, we consider a system made of five variables $\{x_i\}$. The first four variables constitute a multiplet made of unidirectionally coupled logistic maps, equations (4)–(5) with i ranging in $\{1, 2, 3, 4\}$, coupling ρ , and interactions $1 \rightarrow 2, 2 \rightarrow 3$, and $3 \rightarrow 4$. The fifth variable is influenced by the mean field from the coupled map lattice, see equation (7). The four variables in the multiplet become *almost* synchronized for $\rho > 0.4$. In figure 7, we depict both the average influence from the variables in the multiplet to x_5 and the average influence between pairs of variables in the multiplet: both linear and nonlinear PWGC and CGC are shown for the two quantities. Note that only the nonlinear algorithm correctly verifies the causal interactions within the multiplet of four variables, whereas the linear algorithm detects a very low causal interdependency among them. The driving influence from the multiplet to x_5 detected by CGC vanishes at high coupling

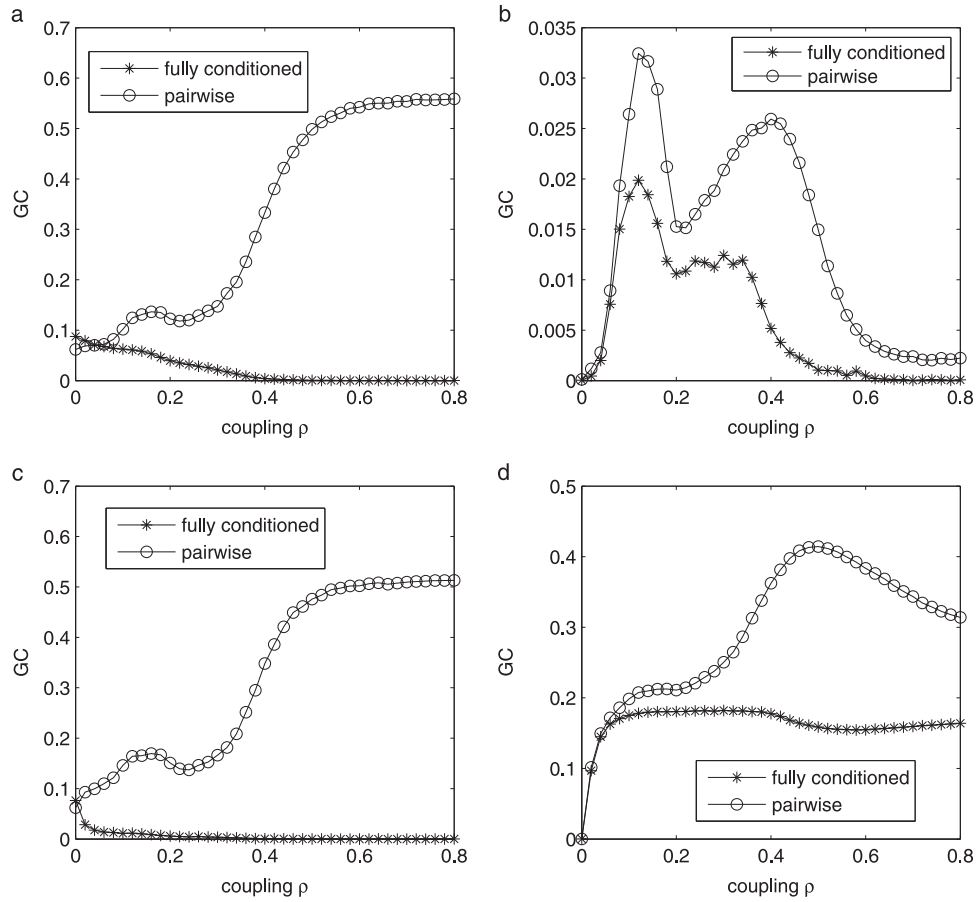


Figure 7. (a) The CGC and PWGC of the causal interaction $x_1 \rightarrow x_5$ are plotted as functions of the coupling ρ for the toy model for the proposed approach combining PWGC and CGC. The linear algorithms are used here and results are averaged over 100 runs of 2000 time points. (b) The CGC and PWGC of the causal interaction between two variables in the multiplet are plotted as functions of the coupling ρ for the toy model for the proposed approach combining PWGC and CGC. The linear algorithms are used here and results are averaged over 100 runs of 2000 time points. (c) The nonlinear CGC and PWGC of the causal interaction $x_1 \rightarrow x_5$ are plotted as functions of the coupling ρ for the toy model for the proposed approach combining PWGC and CGC. The algorithm with the polynomial kernel of order 2 is used here and results are averaged over 100 runs of 2000 time points. (d) The nonlinear CGC and PWGC of the causal interaction between two variables in the multiplet are plotted as functions of the coupling ρ for the toy model for the proposed approach combining PWGC and CGC. The algorithm with the polynomial kernel of order 2 is used here and results are averaged over 100 runs of 2000 time points.

redundancy, both in the linear and nonlinear approach, due to the redundancy induced by synchronization.

To explain how the proposed approach works, we describe two situations corresponding to low and high coupling, respectively. At low coupling, the CGC approach estimates the correct causal pattern in the system, and the nonzero elements of Δ are $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 4$, $1 \rightarrow 5$, $2 \rightarrow 5$, $3 \rightarrow 5$, and $4 \rightarrow 5$. The nonzero elements of the matrix Δ^* , from PWGC analysis, are

the same as Δ plus $1 \rightarrow 3$, $1 \rightarrow 4$, $2 \rightarrow 4$, corresponding to indirect causalities; however, these three interactions lead to nonzero elements of Δ^2 (and, therefore, of B), hence they must be discarded. It follows that Δ^* does not provide further information than Δ at low coupling.

On the contrary, at high coupling, due to synchronization, the CGC approach does not reveal the causal interactions $1 \rightarrow 5$, $2 \rightarrow 5$, $3 \rightarrow 5$, and $4 \rightarrow 5$, although still they are recognized by PWGC. Here, Δ^* is still nonzero in correspondence to $1 \rightarrow 5$, $2 \rightarrow 5$, $3 \rightarrow 5$, and $4 \rightarrow 5$, while the corresponding elements of B are vanishing. According to our previous discussion, the interactions $1 \rightarrow 5$, $2 \rightarrow 5$, $3 \rightarrow 5$, and $4 \rightarrow 5$, detected by PWGC, should not be discarded: combining the results from CGC and PWGC, we obtain the correct causal pattern even in presence of strong synchronization.

6. Application to gene expression data

HeLa [47] is a famous cell culture, isolated from a human uterine cervical carcinoma in 1951. HeLa cells have acquired cellular immortality, in that the normal mechanisms of programmed cell death after a certain number of divisions have somehow been switched off. We consider the HeLa cell gene expression data of [48]. Data corresponds to 94 genes and 48 time points, with an hour interval separating two successive readings (the HeLa cell cycle lasts 16 hours). The 94 genes were selected from the full data set described in [49], on the basis of the association with cell cycle regulation and tumor development. We apply linear PWGC and linear CGC (using only another conditioning variable and using both IB and PB selection strategies, described in section 3). We note that the CGC approach is unfeasible in this case due to the limited number of samples. In this case, due to the limited number of samples, we do not use statistical testing to assess the significance of the retrieved links. Rather, we introduce a threshold for the influence and analyze the pattern as the threshold is varied. In figure 8, results are reported as a function of the number of links found by PWGC $n_{pairwise}$ (which increases as the threshold is decreased); we plot (1) the number of links found by PGC $n_{partial}$, (2) the number of links found by PGC and not by PWGC n_{novel} , which are thus a signature of synergy, (3) the percentage of pairwise links that can be explained as direct or indirect causalities of the PGC pattern (thus being consistent with the partial causality pattern), found using the matrix B_1 to detect the indirect links, which correspond to circuits like the one described in figure 6(a) and (4) the number of causality links found by PWGC and not consistent with PWGC $n_{unexplained}$, corresponding to redundancy. The two curves refer to the two selection strategies for partial conditioning.

The low number of samples here allowed us to use only one conditioning variable, and therefore, to analyze the circuits of only three variables; a closely related analysis [50] has been proposed to study how a gene modulates the interaction between two other genes. Conversely, the true underlying gene regulatory network being unknown, we cannot assess the performances of the algorithms in terms of correctly detected links.

We note that both n_{novel} and $n_{unexplained}$ assume relatively large values; hence, both redundancy and synergy characterize this data set. The PB selection strategy yields slightly higher values of n_{novel} and $n_{unexplained}$, emerging as a better discriminator of synergy and redundancy than IB. A comparison with the fully conditioned approach is not possible in this case. On the other hand, as far as the search for synergetic effects is concerned, we find that the synergetic interactions found by PCGC with the two strategies are not coinciding; indeed, only 10% of all synergetic interactions are found by both strategies. This suggests that, when

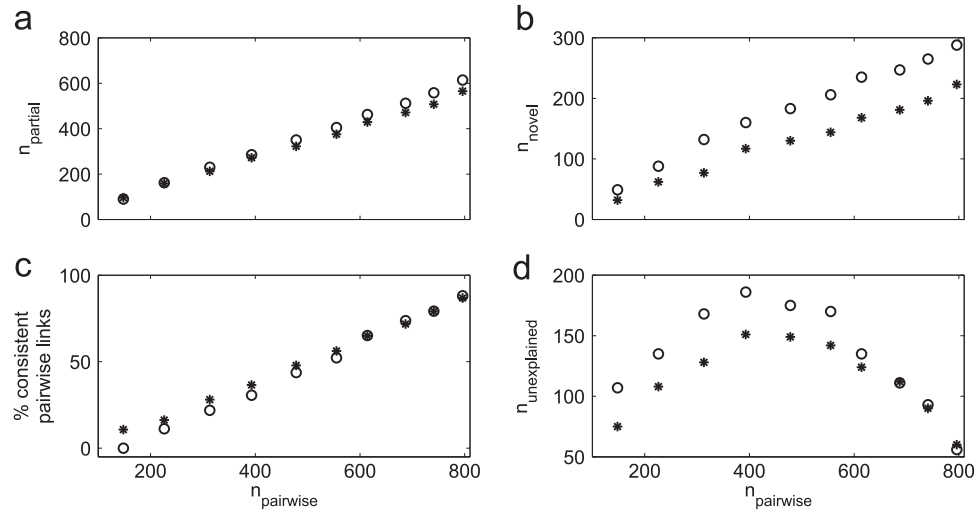


Figure 8. Concerning the genetic application, several quantities are plotted as a function of the number of bivariate causality links exceeding the threshold. (a) The number of retrieved links by PCGC with strategies IB (*) and PB (o). (b) The number of retrieved links by PCGC, with strategies IB (*) and PB (o), which are not present in the bivariate pattern. (c) The percentage of retrieved links, by BVGC, which are consistent with the PCGC with strategies IB (*) and PB (o). (d) The number of retrieved links by BVGC, which are not consistent with PCGC (with strategies IB (*) and PB (o)).

searching for suppressors, several sets of conditioning variables should be used in CGC to explore more possible alternative pathways, especially when there is not *a priori* information on the network structure.

7. Conclusions

In this paper, we considered the inference, from time series data, of the information flow between subsystems of a complex network, which is an important problem in medicine and biology. In particular, we analyzed the effects that synergy and redundancy induce on the Granger causal analysis of time series.

Concerning synergy, we have shown that the search for synergetic contributions in the information flow is equivalent to the search for suppressors, i.e., variables that improve the predictive validity of another variable. Pairwise analysis fails to verify this kind of variable. However, fully multivariate GC solves this problem: conditioning on suppressor variables leads to nonzero GC. In cases when the number of samples is low, we have shown that PCGC is a valuable option, provided that the selection strategy to choose the conditioning variables, succeeds in picking the suppressors. In this paper, we considered two different strategies: choosing the most informative variables for the candidate driver node, or choosing the nodes with the highest pairwise influence to the target. From the several examples analyzed here, we have shown that the first strategy is viable in the presence of redundancy, whereas when the interaction pattern has a tree-like structure, the latter is preferable. However, the issue of selecting variables for PCGC deserves further attention, as it corresponds to the search for suppressor variables and correspondingly of synergetic effects. We have also provided

evidence, for the first time, that synergetic effects are present in an epileptic brain in the preictal condition (just prior to seizure).

We have then shown that CGC approaches do not work well in the presence of redundancy. To handle redundancy, we propose to split the pairwise links into two subsets: those that correspond to indirect connections of the multivariate GC, and those that do not. The links that are not explained, as indirect connections are ascribed to redundancy effects and they are merged to those from CGC to provide the full causality pattern in the system. We have applied this approach to a genetic data set from the HeLa culture, and found that the underlying gene regulatory networks are characterized by both redundancy and synergy, hence these approaches are also promising w.r.t. the reverse engineering of gene regulatory networks.

In conclusion, we observe that the problem of inferring reliable estimates of the causal interactions in real dynamical complex systems, when limited *a priori* information is available, remains a major theoretical challenge. Recently, the most important results in this direction are related to the use of data-driven approaches, such as GC and transfer entropy. In this work, we have shown that in the presence of redundancy and synergy, combining the results from the pairwise and conditioned approaches may lead to more effective analyses.

References

- [1] Bashan A, Bartsch R P, Kantelhardt J W, Havlin S and Ivanov P Ch 2012 *Nat. Commun.* **3** 702
- [2] Barabasi A L 2002 *Linked: The New Science of Networks* (Cambridge, MA: Perseus Publishing)
- [3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D-U 2006 *Phys. Rep.* **424** 175
- [4] Abbott L F and van Vreeswijk C 1993 *Phys. Rev. E* **48** 1483
- [5] Gardner T S, Bernardo D, Lorenz D and Collins J J 2003 *Science* **301** 102
- [6] Tucker C L, Gera J F and Uetz P 2001 *Trends Cell Biol.* **11** 102
- [7] Jeong H, Tombor B, Albert R, Oltvai Z N and Barabasi A L 2000 *Nature* **407** 651
- [8] Guimer R and Nunes Amaral L A 2005 *Nature* **433** 895
- [9] de la Fuente I M, Cortes J M, Perez-Pinilla M B, Ruiz-Rodriguez V and Veguillas J 2011 *PLoS One* **6** e27224
- [10] Pereda E, Quiroga R and Bhattacharya J 2005 *Prog. Neurobiol.* **77** 1
- [11] Wibral M, Vicente R and Lizier J T (ed) 2014 *Directed Information Measures in Neuroscience* (Berlin: Springer)
- [12] Sameshima K and Baccala L A (ed) 2014 *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis* (Boca Raton, FL: CRC Press)
- [13] Granger C W J 1969 *Econometrica* **37** 424
- [14] Bressler S L and Seth A K 2011 *Neuroimage* **58** 323
- [15] Schreiber T 2000 *Phys. Rev. Lett.* **85** 461
- [16] Barnett L, Barrett A B and Seth A K 2009 *Phys. Rev. Lett.* **103** 238701
- [17] Hlavackova-Schindler K 2011 *Appl. Math. Sci.* **5** 3637
- [18] Vicente R, Wibral M, Lindner M and Pipa G 2011 *J. Comput. Neurosci.* **30** 45
- [19] Zhou D, Zhang Y, Xiao Y and Cai D 2014 *New J. Phys.* **16** 043016
- [20] Bartsch R P and Ivanov P Ch 2014 *Commun. Comput. Inf. Sci.* **438** 270–87
- [21] Bartsch R P, Schumann A Y, Kantelhardt J W, Penzele T and Ivanov P Ch 2012 *Proc. Natl Acad. Sci.* **109** 10181–6
- [22] Ge T, Cui Y, Lin W, Kurths J and Liu C 2012 *New J. Phys.* **14** 083028
- [23] Geweke J F 1984 *J. Am. Stat. Assoc.* **79** 907–15
- [24] Marinazzo D, Pellicoro M and Stramaglia S 2012 *Comput. Math. Methods Med.* **2012** 303601
- [25] Wu G, Liao W, Chen H, Stramaglia S and Marinazzo D 2013 *Brain Connectivity* **3** 294–301

- [25] Angelini L *et al* 2010 *Phys. Rev. E* **81** 037201
- [26] Griffith V and Koch C 2014 Quantifying synergistic mutual information *Guided Self-Organization: Inception* vol 9 ed M Prokopenko (Berlin: Springer) pp 159–90
- [27] Harder M, Salge C and Polani D 2013 *Phys. Rev. E* **87** 012130
- [28] Lizier J T, Flecker B and Williams P L 2013 *IEEE Symposium on Artificial Life (ALIFE)* pp 43–52
- [29] Yang S and Gu J 2004 *J. Zhejiang Univ. Sci.* **5** 1382
- [30] Schneideman E, Bialek W and Berry M J 2003 *J. Neurosci.* **23** 11539
- [31] Gat I and Tishby N 1998 *NIPS* (Boston, MA: MIT Press) pp 111–7
- [32] Barrett A B, Barnett L and Seth A K 2010 *Phys. Rev. E* **81** 041907
- [33] Zhou Z, Chen Y, Ding M, Wright P, Lu Z and Liu Y 2009 *Hum. Brain Mapp.* **30** 2197
- [34] Marinazzo D, Liao W, Pellicoro M and Stramaglia S 2010 *Phys. Lett. A* **374** 4040
- [35] Conger A J 1974 *Educ. Psychol. Meas.* **34** 35–46
- [36] Thompson F T and Levine D U 1997 *Mult. Linear Regression Viewpoints* **24** 11–13
- [37] Stramaglia S, Wu G, Pellicoro M and Marinazzo D 2012 *Phys. Rev. E* **86** 066211
- [38] Chicharro D and Ledberg A 2012 *Phys. Rev. E* **86** 041901
- [39] Hlavackova-Schindler K, Palus M, Vejmelka M and Bhattacharya J 2007 *Phys. Rep.* **441** 1
- [40] Marinazzo D, Pellicoro M and Stramaglia S 2008 *Phys. Rev. E* **77** 056215
- [41] Vapnik V 1995 *The Nature of Statistical Learning Theory* (New York: Springer)
- [42] Shawe-Taylor J and Cristianini N 2004 *Kernel Methods For Pattern Analysis* (London: Cambridge University Press)
- [43] Marinazzo D, Pellicoro M and Stramaglia S 2008 *Phys. Rev. Lett.* **100** 144103
- [44] Ancona N and Stramaglia S 2006 *Neural Comput.* **18** 749
- [45] <http://math.bu.edu/people/kolaczyk/datasets.html>, accessed May 2012
- [46] Kramer M A, Kolaczyk E D and Kirsch H E 2008 *Epilepsy Res.* **79** 173
- [47] Masters J R 2002 *Nat. Rev. Cancer* **2** 315
- [48] Fujita A *et al* 2007 *BMC Syst. Biol.* **1** 1
- [49] Whitfield M L *et al* 2002 *Mol. Biol. Cell* **13** 1977
- [50] Wang K *et al* 2009 *Nat. Biotechnol.* **27** 829837