Chapter 5
Granger Causality: Theory and Applications

Shuixia Guo[1], Christophe Ladroue[2], Jianfeng Feng[1,2]

5.1 Introduction

A question of great interest in systems biology is how to uncover complex network structures from experimental data[1, 3, 18, 38, 55]. With the rapid progress of experimental techniques, a crucial task is to develop methodologies that are both statistically sound and computationally feasible for analysing increasingly large datasets and reliably inferring biological interactions from them [16, 17, 22, 37, 40, 42]. The building block of such enterprise is to being able to detect relations (causal, statistical or functional) between nodes of the network. Over the past two decades, a number of approaches have been developed: information theory ([4]), control theory ([17]) or Bayesian statistics ([35]). Here we will be focusing on another successful alternative approach: Granger causality. In recent Cell papers [7, 12], the authors have come to the conclusion that the ordinary differential equation approach outperforms the other reverse engineering approaches (Bayesian network and information theory) in building causal networks. We have demonstrated that the Granger causality achieves better results than the ordinary differential approach [34].

The basic idea of Granger causality can be traced back to Wiener[47] who conceived the notion that, if the prediction of one time series is improved by incorporating the knowledge of a second time series, then the latter is said to have a causal influence on the first. Granger[23, 24] later formalized Wiener's idea in the context of linear regression models. Specifically, two auto-regressive models are fitted to the first time series – with and without including the second time series – and the improvement of the prediction is measured by the ratio of the variance of the error terms. A ratio larger than one signifies an improvement, hence a causal connection. At worst, the ratio is 1 and signifies causal independence from the second time series to the first. Geweke's decomposition of a vector autoregressive process ([20, 21]) led to a set of causality measures which have a spectral representation and make the interpretation more informative and useful by extending Granger causality to the frequency domain. In this chapter, we aim to present Granger causality and how its original formalism has been extended to address biological and computational issues, as summarized in Fig. 5.1.

[1]Mathematics and Computer Science College, Hunan Normal University, Changsha 410081, P.R.China
[2]Department of Computer Science and Mathematics, Warwick University, Coventry CV4 7AL, UK

Partial Granger Causality In its original conception, Granger Causality is limited to the investigation of pairs of time series. If strong enough, indirect connections produce spurious relations between distant nodes. Conditional Granger causality ([9, 11, 21]) is able to deal with this situation by removing the influence of an external node and thus discarding what would be misleading connections. However, it requires the explicit knowledge of the influencing node; in other words, its applicability largely depends on the ability to measure all relevant variables, which is usually not possible in biological recordings. Both exogenous inputs and endogenous variables can confound accurate causal influences and thus degrade the credibility of the uncovered network structure. In order to eliminate the influences of exogenous inputs and latent variables, and inspired by the definition of partial correlation in statistics, we introduce a new definition of Granger causality: partial Granger causality, which is robust against perturbations due to common unseen variables ([25, 26, 29, 50]).

Frequency Analysis Thanks to Geweke's decomposition of a power spectrum [20, 21], Granger causality can be expressed in the frequency domain [19, 26] and thus provides more information: a spectrum indicating at which frequencies the connection between two time series occurs is produced, instead of a single number. This property is especially useful for studying phenomena in which frequencies play a major role, like electro-physiological recordings (*e.g.* Multi-electrode array in the brain) or rhytmic behaviour like circadian rhythm [15]. The two representations (time and frequency domains) are consistent: integrating the spectrum over all frequencies is equal in practice to the time domain Granger causality.

Complex Granger Causality In complex systems of genes, proteins or neurons, elements often work cooperatively or competitively to achieve a task. If we want to understand biological process in details, it is of substantial importance to study the interactions among groups of nodes. Such group interactions are ubiquitous in biological processes: enzymes act on the production rate of metabolites ([18, 28]), information is passed on from one layer of neurons to the next, transcription factors form complexes which influence gene activity etc.. These interactions will be missed out with traditional Granger causality approaches. To tackle this problem, we extend traditional Granger causality to *complex* Granger causality, defined both in the time and frequency domains ([15, 29]). Furthermore, we validate this approach with real biological data.

Harmonic Granger Causality Partial Granger causality was developed to eliminate the influence of common inputs, be they exogenous and endogenous. But full elimination is only possible if all common external inputs have equal influence on all measured variables, which is generally not realistic to expect in experimental recordings. For example, we know that certain genes are sensitive to local pH or temperature, while others are not. In this situation, we want to know which variables are impacted by environmental inputs, and the extent of this influence. In Sec. 5.5, Granger causality is modified to explicitly include a model of an oscillating external input ([8, 10, 19, 27, 44, 48, 50]). Its influence on each of the observed variables can then be quantified and an accurate network can be built. The method is applied to microarray experiments to study the circadian rythm in a plant.

Granger Causality and Bayesian Networks Bayesian networks are a popular approach for investigating biological systems which has proved successful on many occasions ([35, 39]). A natural question is to know whether one should choose Bayesian network or Granger causality when faced with data. In Sec. 5.6, we present a systematic and computationally intensive comparison between the two methods ([52]). The experiment is done on simulated data, of which the true structure is known and for which we have total control on the parameters. It results that the length of the time series is a crucial factor for the contrasted performances

of the two methods.

Unified Causal Model In contrast to many similar frameworks ([16, 37]), Granger causality does not use the concept of perturbation to define causality. In these frameworks, the states of the system are compared before and after some event (*e.g.* injection of different inhibitors in a cell in[39]). Instead, Granger causality relies on dependence over time to define causality. Unified Causal Model (UCM) is an attempt at including the notions of stimuli and modifying coupling to traditional Granger causality. Sec. 5.7 explains how UCM unifies the seemingly different approaches of Granger causality, phenomenological in nature, and the model-based Dynamic Causal Model (DCM, [17]).

Large Networks While Partial Granger causality is efficient for uncovering small network structures, the method does not scale up for practical reasons: removing the influence of all other variables requires the fit of a linear model so large that it exhausts the number of observables. The system is under-determined and the resulting Granger causality unreliable. To address this issue, we developed an iterative procedure that builds a global network of possibly hundreds of nodes from local investigations. The idea is to gradually prune the network by removing indirect connections, considered one at a time. Applied on simulated networks of 200 nodes, the method shows very good performance. Sec. 5.8 shows its application on a large network of 800 proteins [54].

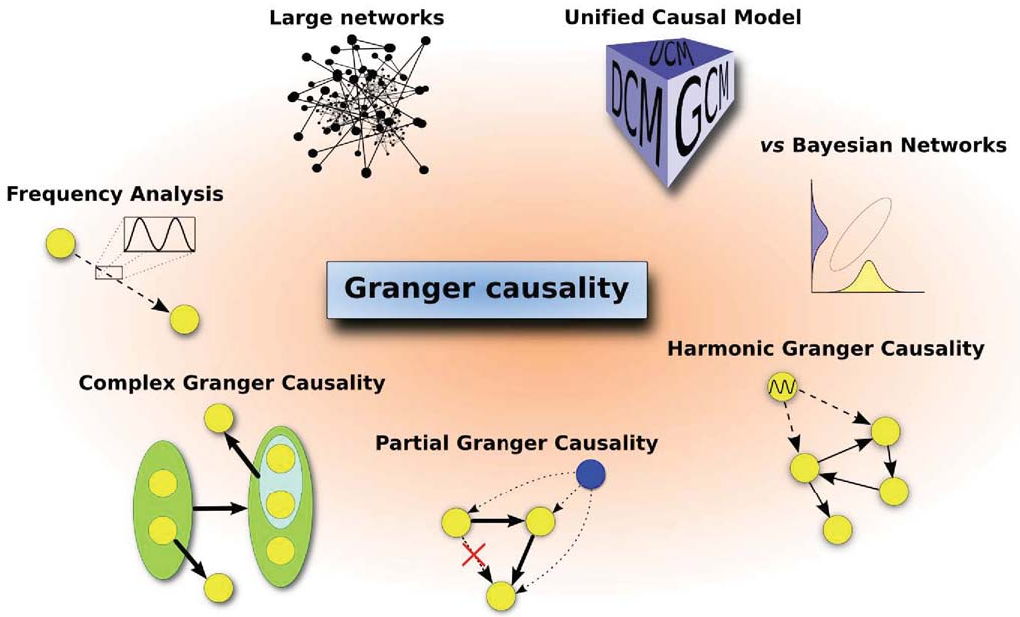

Fig. 5.1 The extensions of traditional Granger causality that will be discussed in this chapter.

5.2 Partial Granger causality

Traditional Granger causality is defined for two time series only. In order to build a network of interactions from a collection of time series, a simple approach is to apply it to all possible pairs of signals. However, indirect links can lead to spurious connections if single connections are strong enough: if $A \rightarrow B$ and $B \rightarrow C$, then very likely $A \rightarrow C$ will be picked up by Granger causality. The end-result is often a very densely connected network. *Conditional* Granger causality ([11, 22]) works by explicitly removing the influence of a third signal and thus avoids producing misleading links.

Critically, the ability of conditional Granger causality to deal with indirect interactions depends on being able to measure all relevant variables in a system. This is not possible most of the times, and both environmental inputs and unmeasured latent variables can confound real causal influences. For example, in experimental data recorded from the inferotemporal (IT) cortex, every measured neuron receives common exogenous inputs from the visual cortex and feedback from the prefrontal cortex. Moreover, even with advanced multi-electrode array techniques, only a tiny subset of interacting neurons in a single area is recorded and there bound to be latent, unobserved variables. In this section, we introduce Partial Granger causality ([25]), an extension of conditional Granger causality which addresses the issue of exogenous and latent variables.

5.2.1 Time domain formulation

For three time series X_t, Y_t and Z_t, define $\overrightarrow{y_t} = (X_t, Y_t, Z_t)$, where X_t and Y_t are 1-dimensional time series, Z_t is a set of time series of dimension m. A general form of an autoregressive model with zero mean and exogenous variable $\overrightarrow{\varepsilon_t^E}$ has the following vector autoregressive representation with the use of the lag operator[1] $\mathscr{L}$:

$$\mathbf{B}(\mathscr{L})\overrightarrow{y_t} = \overrightarrow{\varepsilon_t^E} + \overrightarrow{\varepsilon_t} \tag{5.1}$$

where $\mathbf{B}$ is a polynomial matrix of $\mathscr{L}$, $\mathbf{B}(0) = I_n$, the $n \times n$ identity matrix. The two random vectors $\overrightarrow{\varepsilon^E}$ and $\overrightarrow{\varepsilon}$ are independent. The exogenous variable $\overrightarrow{\varepsilon^E}$ represents the environmental drive and is typically present in any experimental setup.

As already mentioned, the confounding influence of latent variables is possibly even more disruptive than that due to exogenous inputs. To incorporate latent variables, assume that the i^{th} network element receives unmeasured inputs of the form $\sum_{j=1}^{N} x_{ij}(t)/N$, where each x_{ij} is a stationary time series and j is the latent index.

According to the Wold representation, any stationary variable $\xi(t)$ can be expressed as the summation of the form $\sum_k \psi_k \varepsilon(t-k)$, and we have

$$x_{ij}(t) = \sum_{k=1} \psi_{ij,k} \varepsilon_{ij}^L (t-k).$$

where ε_{ij}^L is the latent variable of indices (i, j). Therefore

[1] A lag operator $\mathscr{L}$ is such that $\mathscr{L}X_t = X_{t-1}$. Thus, applying the operator k times yields: $\mathscr{L}^k X_t = X_{t-k}$. The auto-regressive model $X_t + A_1 X_{t-1} + A_2 X_{t-2} + \ldots = \varepsilon_t$ can be represented as $(Id + A_1\mathscr{L} + A_2\mathscr{L}^2 + \ldots)X_t = \varepsilon_t$.

$$\begin{aligned}\sum_{j=1}^{N} x_{ij}(t)/N &= \sum_{j=1}^{N}\sum_{k=1} \psi_{ij,k}\varepsilon_{ij}^{L}(t-k)/N \\ &= \sum_{k=1} \bar{\psi}_{i,k}\varepsilon_{i}^{L}(t-k)\end{aligned}$$

where $\psi_{\cdot}$ are constants. In other words, each node i receives a latent input ε_i^L which depends on its history.

So the model Eq. (5.1) becomes

$$\mathbf{B}(\mathscr{L})\overrightarrow{y_t} = \overrightarrow{\varepsilon_t^E} + \overrightarrow{\varepsilon_t} + \mathbf{B}^{(1)}(\mathscr{L})\overrightarrow{\varepsilon_t^L} \tag{5.2}$$

where the random vectors $(\overrightarrow{\varepsilon_t^E}, \overrightarrow{\varepsilon_t^L})$ and $\overrightarrow{\varepsilon_t}$ are independent and $\mathbf{B}^{(1)}(\mathscr{L})$ is another polynomial matrix of $\mathscr{L}$ of appropriate size.

Now consider two time series X_t and Z_t which admit a joint autoregressive representation of the form

$$\begin{cases} X_t = \sum\limits_{i=1}^{\infty} a_{1i}X_{t-i} + \sum\limits_{i=1}^{\infty} c_{1i}Z_{t-i} + \overrightarrow{\varepsilon_{1t}} + \overrightarrow{\varepsilon_{1t}^E} + \overrightarrow{B_1(\mathscr{L})\varepsilon_{1t}^L} \\ Z_t = \sum\limits_{i=1}^{\infty} b_{1i}Z_{t-i} + \sum\limits_{i=1}^{\infty} d_{1i}X_{t-i} + \overrightarrow{\varepsilon_{2t}} + \overrightarrow{\varepsilon_{2t}^E} + \overrightarrow{B_2(\mathscr{L})\varepsilon_{2t}^L} \end{cases} \tag{5.3}$$

For simplicity of notation, let us define

$$u_i(t) = \overrightarrow{\varepsilon_{it}} + \overrightarrow{\varepsilon_{it}^E} + \overrightarrow{B_i(\mathscr{L})\varepsilon_{it}^L}$$

The noise covariance matrix for the model can be represented as

$$S = \begin{bmatrix} \mathrm{var}(u_{1t}) & \mathrm{cov}(u_{1t},u_{2t}) \\ \mathrm{cov}(u_{2t},u_{1t}) & \mathrm{var}(u_{2t}) \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xz} \\ S_{zx} & S_{zz} \end{bmatrix}$$

In the same fashion, the vector autoregressive representation for a system involving three variables X_t, Y_t and Z_t can be written as follows:

$$\begin{cases} X_t = \sum\limits_{i=1}^{\infty} a_{2i}X_{t-i} + \sum\limits_{i=1}^{\infty} b_{2i}Y_{t-i} + \sum\limits_{i=1}^{\infty} c_{2i}Z_{t-i} + \overrightarrow{\varepsilon_{3t}} + \overrightarrow{\varepsilon_{3t}^E} + \overrightarrow{B_3(\mathscr{L})\varepsilon_{3t}^L} \\ Y_t = \sum\limits_{i=1}^{\infty} d_{2i}X_{t-i} + \sum\limits_{i=1}^{\infty} e_{2i}Y_{t-i} + \sum\limits_{i=1}^{\infty} f_{2i}Z_{t-i} + \overrightarrow{\varepsilon_{4t}} + \overrightarrow{\varepsilon_{4t}^E} + \overrightarrow{B_4(\mathscr{L})\varepsilon_{4t}^L} \\ Z_t = \sum\limits_{i=1}^{\infty} g_{2i}X_{t-i} + \sum\limits_{i=1}^{\infty} h_{2i}Y_{t-i} + \sum\limits_{i=1}^{\infty} k_{2i}Z_{t-i} + \overrightarrow{\varepsilon_{5t}} + \overrightarrow{\varepsilon_{5t}^E} + \overrightarrow{B_5(\mathscr{L})\varepsilon_{5t}^L} \end{cases} \tag{5.4}$$

The noise covariance matrix for the model can be represented as

$$\Sigma = \begin{bmatrix} \mathrm{var}(u_{3t}) & \mathrm{cov}(u_{3t},u_{4t}) & \mathrm{cov}(u_{3t},u_{5t}) \\ \mathrm{cov}(u_{4t},u_{3t}) & \mathrm{var}(u_{4t}) & \mathrm{cov}(u_{4t},u_{5t}) \\ \mathrm{cov}(u_{5t},u_{3t}) & \mathrm{cov}(u_{5t},u_{4t}) & \mathrm{var}(u_{5t}) \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

In order to consider the influence from Y to X while controlling for the effect of the exogenous input, we consider the variance of u_{1t} when we eliminate the influence of u_{2t}:

$$\mathrm{cov}(u_{1t},u_{1t})-\mathrm{cov}(u_{1t},u_{2t})\mathrm{cov}(u_{2t},u_{2t})^{-1}\mathrm{cov}(u_{2t},u_{1t})=S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}$$

Similarly, the variance of u_{3t} while eliminating the influence of u_{5t} equals to

$$\mathrm{cov}(u_{3t},u_{3t})-\mathrm{cov}(u_{3t},u_{5t})\mathrm{cov}(u_{5t},u_{5t})^{-1}\mathrm{cov}(u_{5t},u_{3t})=\Sigma_{xx}-\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$$

The value of $S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}$, a scalar, measures the accuracy of the autoregressive prediction of X based on its previous values conditioned on Z and eliminating the influence of the latent variables, whereas the value of $\Sigma_{xx}-\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$, also a scalar, represents the accuracy of predicting present value of X based on the previous history of both X and Y, conditioned on Z and eliminating the influence of latent variables. Granger causality defines the causality from one process to another by comparing the improvement in prediction when the first process is taken into account. Similarly we define this causal influence by

$$F=\ln\left(\frac{|S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}|}{|\Sigma_{xx}-\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}|}\right) \tag{5.5}$$

We call F *partial* Granger causality. Note that conditional Granger causality is defined by $F=\ln\left(\frac{|S_{xx}|}{|\Sigma_{xx}|}\right)$. The essential difference between them is that with conditional Granger causality, the effect of latent variables remains present both in the denominator $|\Sigma_{xx}|$ and in the numerator $|S_{xx}|$. In contrast, partial Granger causality uses the conditional variance in both the denominator $|\Sigma_{xx}-\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}|$ and numerator $|S_{xx}-S_{xz}S_{zz}^{-1}S_{zx}|$. As a result, the effects of the latent and exogenous variables are both taken into account.

5.2.2 Numerical example

Example 1 We simulated a 5-node oscillatory network structurally connected with different delays. In order to illustrate the robustness of partial Granger causality, we add exogenous inputs and latent variables to the model.

$$\begin{cases}
x_1(t) = 0.95\sqrt{2}x_1(t-1)-0.9025x_1(t-2)+\varepsilon_1(t)+a_1\varepsilon_6(t)\\
\qquad\quad +b_1\varepsilon_7(t-1)+c_1\varepsilon_7(t-2)\\
x_2(t) = 0.5x_1(t-2)+\varepsilon_2(t)+a_2\varepsilon_6(t)+b_2\varepsilon_7(t-1)+c_2\varepsilon_7(t-2)\\
x_3(t) = -0.4x_1(t-3)+\varepsilon_3(t)+a_3\varepsilon_6(t)+b_3\varepsilon_7(t-1)+c_3\varepsilon_7(t-2)\\
x_4(t) = -0.5x_1(t-2)+0.25\sqrt{2}x_4(t-1)+0.25\sqrt{2}x_5(t-1)+\varepsilon_4(t)\\
\qquad\quad +a_4\varepsilon_6(t)+b_4\varepsilon_7(t-1)+c_4\varepsilon_7(t-2)\\
x_5(t) = -0.25\sqrt{2}x_4(t-1)+0.25\sqrt{2}x_5(t-1)+\varepsilon_5(t)+a_5\varepsilon_6(t)\\
\qquad\quad +b_5\varepsilon_7(t-1)+c_5\varepsilon_7(t-2)
\end{cases}$$

where $\varepsilon_i(t), i=1,2,\cdots,7$ are zero-mean uncorrelated processes with identical variances. $a_i\varepsilon_6$ is the exogenous input and the term $b_i\varepsilon_7(t-1)+c_i\varepsilon_7(t-2)$ represents the influence of latent variables. From the model (depicted in Fig. 18.1 (**A**)), one can see that $x_1(t)$ is a direct source to $x_2(t), x_3(t)$, and $x_4(t)$, $x_4(t)$ and $x_5(t)$

share a feedback loop. There is no direct connection between $x_1(t)$ and $x_5(t)$. We perform a simulation of this system with $a_i \sim U[0,1], b_i = 2, c_i = 5, i = 1, \cdots, 5$ to generate a data set of 2000 data points with a sample rate of 200 Hz.

A bootstrap is used to calculate 95% confidence intervals. Fig. 18.1(**B**, upper panel) shows the values for both partial Granger causality (F_1) and conditional Granger causality (F_2) when applied to the simulated data. Partial Granger causality outperforms the conditional Granger causality.

The values of the conditional Granger causality are all very small due to the dominating nuisance effect of latent variables and common inputs, while the partial Granger causality reveals the correct structure.

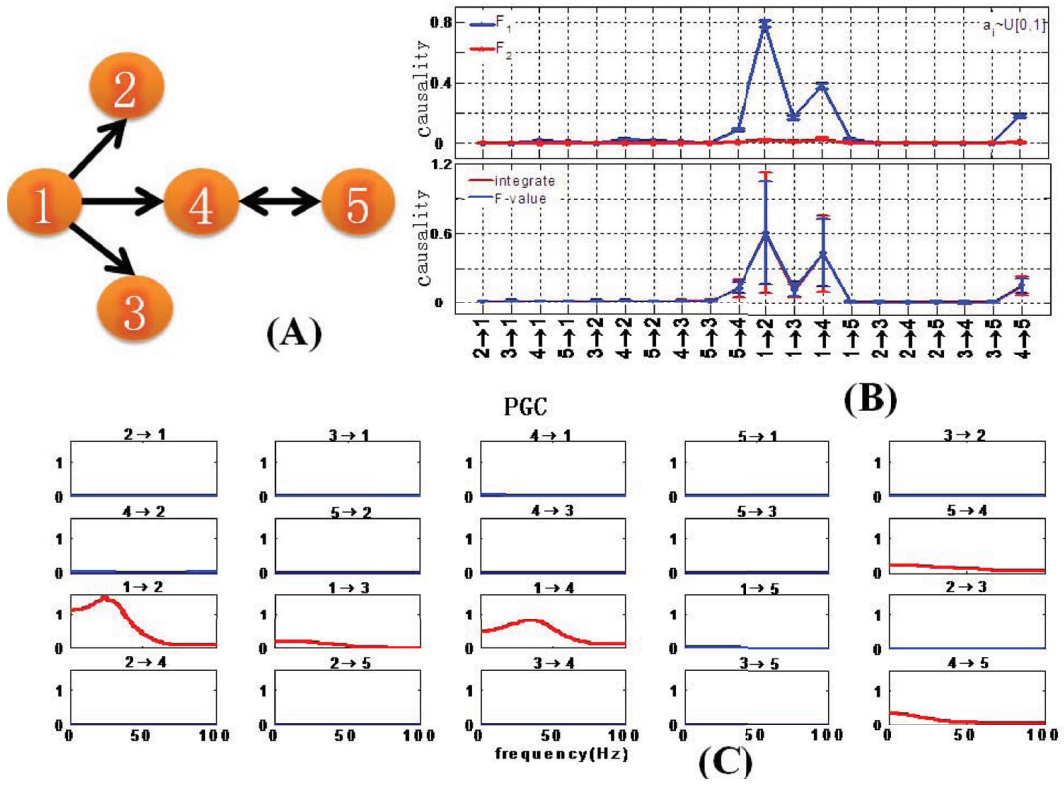

Fig. 5.2 Granger Causality applied to the system in example 1. (**A**) The true network structure. (**B**)(upper panel) Comparison of the partial Granger causality F_1 and the conditional Granger causality F_2 . F_2 fails to pick up any true connections while the inferred links from F_1 is consistent with the correct structure(**A**). (bottom panel) Comparison of the partial Granger causality in the time domain (blue line) and frequency domain (red line, the integral of the frequency domain formulation in the interval $[-\pi, \pi]$. (**C**) Results of the frequency domain decomposition of all 20 pairs of signals.

5.3 Frequency analysis

Granger causality summarizes the influence of one time series on another with a single non-negative number. It is possible to extract more information about their connection by going to the frequency domain. Thanks to Geweke's decomposition [20, 21], the power spectrum of the target signal can be written as a sum of easily intepretable quantities, leading to a natural definition of Granger Causality in the frequency domain. Instead of a single number, a whole spectrum expliciting at which frequencies the signals interact is obtained. We present here a similar decomposition for the more recent partial Granger causality.

To derive the spectral decomposition of the time domain partial Granger causality, we first multiply the matrix

$$P_1 = \begin{pmatrix} 1 & -S_{xz}S_{zz}^{-1} \\ \mathbf{0} & I_m \end{pmatrix} \tag{5.6}$$

to both sides of Eq. (5.3). The normalized equations are represented as:

$$\begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} X_t^* \\ Z_t^* \end{pmatrix} \tag{5.7}$$

with $D_{11}(0) = 1, D_{22}(0) = I_m, D_{21}(0) = \mathbf{0}, \mathrm{cov}(X_t^*, Z_t^*) = 0$, we note that $\mathrm{var}(X_t^*) = S_{xx} - S_{xz}S_{zz}^{-1}S_{zx}, \mathrm{var}(Z_t^*) = S_{zz}$. For Eq. (5.4), we also multiply the matrix

$$P = P_3 \cdot P_2 \tag{5.8}$$

where

$$P_2 = \begin{pmatrix} 1 & \mathbf{0} & -\Sigma_{xz}\Sigma_{zz}^{-1} \\ \mathbf{0} & 1 & -\Sigma_{yz}\Sigma_{zz}^{-1} \\ \mathbf{0} & \mathbf{0} & I_m \end{pmatrix} \tag{5.9}$$

and

$$P_3 = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} \\ -(\Sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy})(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_m \end{pmatrix} \tag{5.10}$$

to both sides of Eq.(5.4). The normalized equation of Eq. (5.4) becomes

$$\begin{pmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix} \tag{5.11}$$

where $\varepsilon_{xt}, \varepsilon_{yt}, \varepsilon_{zt}$ are independent, with variances $\hat{\Sigma}_{xx}, \hat{\Sigma}_{yy}$ and $\hat{\Sigma}_{zz}$:

$$\begin{cases} \hat{\Sigma}_{zz} = \Sigma_{zz} \\ \hat{\Sigma}_{xx} = \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} \\ \hat{\Sigma}_{yy} = \Sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy} - \dfrac{(\Sigma_{yx} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zx})(\Sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy})}{(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})} \end{cases}$$

After Fourier transforming Eq. (5.7) and Eq. (5.11), we can rewrite these two equations as the following expression:

$$\begin{pmatrix} X(\lambda) \\ Z(\lambda) \end{pmatrix} = \begin{pmatrix} G_{xx}(\lambda)\ G_{xz}(\lambda) \\ G_{zx}(\lambda)\ G_{zz}(\lambda) \end{pmatrix} \begin{pmatrix} X^*(\lambda) \\ Z^*(\lambda) \end{pmatrix} \tag{5.12}$$

and

$$\begin{pmatrix} X(\lambda) \\ Y(\lambda) \\ Z(\lambda) \end{pmatrix} = \begin{pmatrix} H_{xx}(\lambda)\ H_{xy}(\lambda)\ H_{xz}(\lambda) \\ H_{yx}(\lambda)\ H_{yy}(\lambda)\ H_{yz}(\lambda) \\ H_{zx}(\lambda)\ H_{zy}(\lambda)\ H_{zz}(\lambda) \end{pmatrix} \begin{pmatrix} E_x(\lambda) \\ E_y(\lambda) \\ E_z(\lambda) \end{pmatrix} \tag{5.13}$$

Note that $X(\lambda)$ and $Z(\lambda)$ from Eq. (5.12) are identical with that from Eq. (5.13), we thus have

$$\begin{aligned} \begin{pmatrix} X^*(\lambda) \\ Y(\lambda) \\ Z^*(\lambda) \end{pmatrix} &= \begin{pmatrix} G_{xx}(\lambda)\ 0\ G_{xz}(\lambda) \\ 0\quad 1\quad 0 \\ G_{zx}(\lambda)\ 0\ G_{zz}(\lambda) \end{pmatrix}^{-1} \begin{pmatrix} H_{xx}(\lambda)\ H_{xy}(\lambda)\ H_{xz}(\lambda) \\ H_{yx}(\lambda)\ H_{yy}(\lambda)\ H_{yz}(\lambda) \\ H_{zx}(\lambda)\ H_{zy}(\lambda)\ H_{zz}(\lambda) \end{pmatrix} \begin{pmatrix} E_x(\lambda) \\ E_y(\lambda) \\ E_z(\lambda) \end{pmatrix} \\ &= \begin{pmatrix} Q_{xx}(\lambda)\ Q_{xy}(\lambda)\ Q_{xz}(\lambda) \\ Q_{yx}(\lambda)\ Q_{yy}(\lambda)\ Q_{yz}(\lambda) \\ Q_{zx}(\lambda)\ Q_{zy}(\lambda)\ Q_{zz}(\lambda) \end{pmatrix} \begin{pmatrix} E_x(\lambda) \\ E_y(\lambda) \\ E_z(\lambda) \end{pmatrix} \end{aligned} \tag{5.14}$$

where $\mathbf{Q}(\lambda) = \mathbf{G}^{-1}(\lambda)\mathbf{H}(\lambda)$. Now the power spectrum of X^* is

$$S_{x^*x^*}(\lambda) = Q_{xx}(\lambda)\hat{\Sigma}_{xx}Q'_{xx}(\lambda) + Q_{xy}(\lambda)\hat{\Sigma}_{yy}Q'_{xy}(\lambda) + Q_{xz}(\lambda)\hat{\Sigma}_{zz}Q'_{xz}(\lambda) \tag{5.15}$$

where $'$ denotes the conjugate transpose. Note that $\hat{\Sigma}_{xx} = \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$; the first term of $S_{x^*x^*}$ can be thought of as the intrinsic power eliminating exogenous inputs and latent variables, while the remaining two terms as the combined causal influence from Y mediated by Z. This interpretation leads immediately to the definition

$$f_{Y\to X|Z}(\lambda) = \ln \frac{|S_{x^*x^*}(\lambda)|}{|Q_{xx}(\lambda)\hat{\Sigma}_{xx}Q'_{xx}(\lambda)|} \tag{5.16}$$

Note that according to Eq. (5.7), the variance of X^* equals to $S_{xx} - S_{xz}S_{zz}^{-1}S_{zx}$. By the Kolmogrov formula [21] for spectral decompositions and under some mild conditions, the Granger causality in the frequency domain and in the time domain measures satisfy

$$F_{Y\to X|Z} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{Y\to X|Z}(\lambda)d\lambda \tag{5.17}$$

Example 2

We apply the frequency analysis to the data presented in example 1. Fig. 18.1(**B**, bottom panel) presents a comparison between the time domain partial Granger causality and the integrated frequency domain partial Granger causality (the summation over all frequencies). As expected from Kolmogorov formula, the decomposition in the frequency domain fits very well with the partial Granger causality in the time domain. Fig. 18.1 (**C**) shows the spectra for the partial Granger causality for all 20 pairs of signals. There are direct causal links from 1 to 2,3 and 4, and a feedback between 5 and 4. Most importantly they are consistent with the results in the time domain.

This approach has been used with success on data where one would expect frequency to be important, *e.g.* electrophysiological experiments, but also on microarray or protein data ([26, 54]). Sec. 5.5.4 presents an example on gene expression changes during circadian rhythm in a plant.

5.4 Group interaction: complex Granger causality

So far, we have only considered Granger causality between two individual signals. This can be limiting for the study of biological systems, where cooperative and competitive actions are a frequent occurence. For example, one would like to study the flow of information between brain regions rather than between individual neurons, or to elucidate transcription factor complexes (an AND-like combination of proteins) in the cell. In this section, we present *complex* Granger causality [29], a measure of causality between collections of time series. It can be considered the natural extension of partial Granger Causality to the multi-dimensional case. However, we'll see that cross-interaction between sources can now have an influence on the strength of the connection to a target group.

5.4.1 Time Domain Formulation

Consider three multiple stationary time series $\overrightarrow{X_t}$, $\overrightarrow{Y_t}$ and $\overrightarrow{Z_t}$ with k,l and m dimensions respectively. We first consider the relationship from $\overrightarrow{Y_t}$ to $\overrightarrow{X_t}$ conditionned on $\overrightarrow{Z_t}$. The joint autoregressive representation for $\overrightarrow{X_t}$ and $\overrightarrow{Z_t}$ can be written as

$$\begin{cases} \overrightarrow{X_t} = \sum_{i=1}^{\infty} a_{1i}\overrightarrow{X}_{t-i} + \sum_{i=1}^{\infty} c_{1i}\overrightarrow{Z}_{t-i} + \overrightarrow{\varepsilon_{1t}} \\ \overrightarrow{Z_t} = \sum_{i=1}^{\infty} b_{1i}\overrightarrow{Z}_{t-i} + \sum_{i=1}^{\infty} d_{1i}\overrightarrow{X}_{t-i} + \overrightarrow{\varepsilon_{2t}} \end{cases} \tag{5.18}$$

where $\overrightarrow{\varepsilon_{.t}}$ are vectors representing exogenous and endogenous inputs and noise. The noise covariance matrix for the system can be represented as

$$S = \begin{pmatrix} \mathrm{var}(\overrightarrow{\varepsilon_{1t}}) & \mathrm{cov}(\overrightarrow{\varepsilon_{1t}}, \overrightarrow{\varepsilon_{2t}}) \\ \mathrm{cov}(\overrightarrow{\varepsilon_{2t}}, \overrightarrow{\varepsilon_{1t}}) & \mathrm{var}(\overrightarrow{\varepsilon_{2t}}) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xz} \\ S_{zx} & S_{zz} \end{pmatrix}$$

where var and cov represent variance and co-variance respectively. In the same manner, the vector autoregressive representation for the system involving the three time series $\overrightarrow{X_t}$, $\overrightarrow{Y_t}$ and $\overrightarrow{Z_t}$ can be written in the following way:

$$\begin{cases} \overrightarrow{X_t} = \sum_{i=1}^{\infty} a_{2i}\overrightarrow{X}_{t-i} + \sum_{i=1}^{\infty} b_{2i}\overrightarrow{Y}_{t-i} + \sum_{i=1}^{\infty} c_{2i}\overrightarrow{Z}_{t-i} + \overrightarrow{\varepsilon_{3t}} \\ \overrightarrow{Y_t} = \sum_{i=1}^{\infty} d_{2i}\overrightarrow{X}_{t-i} + \sum_{i=1}^{\infty} e_{2i}\overrightarrow{Y}_{t-i} + \sum_{i=1}^{\infty} f_{2i}\overrightarrow{Z}_{t-i} + \overrightarrow{\varepsilon_{4t}} \\ \overrightarrow{Z_t} = \sum_{i=1}^{\infty} g_{2i}\overrightarrow{X}_{t-i} + \sum_{i=1}^{\infty} h_{2i}\overrightarrow{Y}_{t-i} + \sum_{i=1}^{\infty} k_{2i}\overrightarrow{Z}_{t-i} + \overrightarrow{\varepsilon_{5t}} \end{cases} \tag{5.19}$$

The noise covariance matrix for the above system can be represented as

$$\Sigma = \begin{pmatrix} \mathrm{var}(\overrightarrow{\varepsilon_{3t}}) & \mathrm{cov}(\overrightarrow{\varepsilon_{3t}}, \overrightarrow{\varepsilon_{4t}}) & \mathrm{cov}(\overrightarrow{\varepsilon_{3t}}, \overrightarrow{\varepsilon_{5t}}) \\ \mathrm{cov}(\overrightarrow{\varepsilon_{4t}}, \overrightarrow{\varepsilon_{3t}}) & \mathrm{var}(\overrightarrow{\varepsilon_{4t}}) & \mathrm{cov}(\overrightarrow{\varepsilon_{4t}}, \overrightarrow{\varepsilon_{5t}}) \\ \mathrm{cov}(\overrightarrow{\varepsilon_{5t}}, \overrightarrow{\varepsilon_{3t}}) & \mathrm{cov}(\overrightarrow{\varepsilon_{5t}}, \overrightarrow{\varepsilon_{4t}}) & \mathrm{var}(\overrightarrow{\varepsilon_{5t}}) \end{pmatrix} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

The conditional variance $S_{xx} - S_{xz}S_{zz}^{-1}S_{zx}$ measures the accuracy of the autoregressive prediction of $\overrightarrow{X}$ based on its previous values conditioned on $\overrightarrow{Z}$ whereas the conditional variance $\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$ measures the accuracy of the autoregressive prediction of $\overrightarrow{X}$ based on its previous values of both $\overrightarrow{X}$ and $\overrightarrow{Y}$ conditioned on $\overrightarrow{Z}$. Note that now both $S_{xx} - S_{xz}S_{zz}^{-1}S_{zx}$ and $\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$ are matrices and not scalars. We denote $T_{x|z}$ and $T_{xy|z}$ their respective traces, which we use to compare their relative size. Following the original concept of Granger causality, we define the partial complex Granger causality from group $\overrightarrow{Y}$ to group $\overrightarrow{X}$ conditioned on group $\overrightarrow{Z}$ to be

$$F_{\overrightarrow{Y} \to \overrightarrow{X} \,|\, \overrightarrow{Z}} = \ln\left(\frac{T_{x|z}}{T_{xy|z}}\right) \tag{5.20}$$

If Y and X are one-dimensional, the definition reduces to that of partial Granger causality. Partial Complex causality has the same property of removing the influence of explicit and unseen variables from the connection from Y to X.

5.4.2 Frequency Domain Formulation

The derivation of the frequency domain formulation follows the steps seen in Sec. 5.3, with the minor difference that the trace replaces the absolute value.

We first normalize Eq. (5.18) by multiplying the matrix

$$P_1 = \begin{pmatrix} I_k & -S_{xz}S_{zz}^{-1} \\ \mathbf{0} & I_m \end{pmatrix} \tag{5.21}$$

to both sides of it. The normalized equations are represented as:

$$\begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} \overrightarrow{X_t} \\ \overrightarrow{Z_t} \end{pmatrix} = \begin{pmatrix} \overrightarrow{X_t^*} \\ \overrightarrow{Z_t^*} \end{pmatrix} \tag{5.22}$$

where $\mathrm{var}(\overrightarrow{X_t^*}) = S_{xx} - S_{xz}S_{zz}^{-1}S_{zx}, \mathrm{var}(\overrightarrow{Z_t^*}) = S_{zz}$. For Eq. (5.19), we also multiply the matrix

$$P = P_3 \cdot P_2 \tag{5.23}$$

where

$$P_2 = \begin{pmatrix} I_k & \mathbf{0} & -\Sigma_{xz}\Sigma_{zz}^{-1} \\ \mathbf{0} & I_l & -\Sigma_{yz}\Sigma_{zz}^{-1} \\ \mathbf{0} & \mathbf{0} & I_m \end{pmatrix} \tag{5.24}$$

and

$$P_3 = \begin{pmatrix} I_k & \mathbf{0} & \mathbf{0} \\ -(\Sigma_{yx} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zx})(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1} & I_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_m \end{pmatrix} \tag{5.25}$$

to both sides of Eq.(5.19). The normalized equation of Eq. (5.19) becomes

$$\begin{pmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{pmatrix} \begin{pmatrix} \overrightarrow{X_t} \\ \overrightarrow{Y_t} \\ \overrightarrow{Z_t} \end{pmatrix} = \begin{pmatrix} \overrightarrow{\varepsilon_{xt}} \\ \overrightarrow{\varepsilon_{yt}} \\ \overrightarrow{\varepsilon_{zt}} \end{pmatrix} \tag{5.26}$$

where $\overrightarrow{\varepsilon_{xt}}, \overrightarrow{\varepsilon_{yt}}, \overrightarrow{\varepsilon_{zt}}$ are independent, and their variances being $\hat{\Sigma}_{xx}, \hat{\Sigma}_{yy}$ and $\hat{\Sigma}_{zz}$ with

$$\begin{cases} \hat{\Sigma}_{zz} = \Sigma_{zz} \\ \hat{\Sigma}_{xx} = \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} \\ \hat{\Sigma}_{yy} = \Sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy} - \dfrac{(\Sigma_{yx} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zx})(\Sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy})}{(\Sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy})} \end{cases}$$

As shown in Sec. 5.3, we can obtain the power spectrum of $\overrightarrow{X_t^*}$

$$S_{x^*x^*}(\omega) = Q_{xx}(\omega)\hat{\Sigma}_{xx}Q_{xx}^*(\omega) + Q_{xy}(\omega)\hat{\Sigma}_{yy}Q_{xy}^*(\omega) + Q_{xz}(\omega)\hat{\Sigma}_{zz}Q_{xz}^*(\omega) \tag{5.27}$$

Considering the traces of both sides of Eq. 5.27, we have:

$$\mathrm{tr}(S_{x^*x^*}(\omega)) = \mathrm{tr}(Q_{xx}(\omega)\hat{\Sigma}_{xx}Q_{xx}^*(\omega)) + \mathrm{tr}(Q_{xy}(\omega)\hat{\Sigma}_{yy}Q_{xy}^*(\omega)) + \mathrm{tr}(Q_{xz}(\omega)\hat{\Sigma}_{zz}Q_{xz}^*(\omega)) \tag{5.28}$$

As before, we can think of the first term as the intrinsic power while eliminating exogenous inputs and latent variables, and the remaining two terms as the combined influence from $\overrightarrow{Y}$ mediated by $\overrightarrow{Z}$.

This interpretation leads immediately to the definition

$$f_{\overrightarrow{Y}\to\overrightarrow{X}|\overrightarrow{Z}}(\omega) = \ln \frac{\mathrm{tr}(S_{x^*x^*}(\omega))}{\mathrm{tr}(Q_{xx}(\omega)\hat{\Sigma}_{xx}Q_{xx}^*(\omega))} \tag{5.29}$$

Note that according to Eq. (5.20), the variance of X^* equals to $\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$. By the Kolmogrov formula [21] for spectral decompositions and under the same mild conditions, the Granger causality in the frequency domain and in the time domain measures satisfy

$$F_{\vec{Y}\to\vec{X}|\vec{Z}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{\vec{Y}\to\vec{X}|\vec{Z}}(\omega)d\omega \tag{5.30}$$

5.4.3 *Effect of correlation between sources*

The complex Granger causality between a group and a target signal can be affected by the source signals' cross-correlations. Let us consider a model where $y_i, i = 1,2,\cdots,N$ are identical random processes. The Granger causality from $(y_i(t), i = 1,\cdots,N)$ to their weighted sum $y(t) := a\sum_{i=1}^{N} y_i(t) + \varepsilon_t$ is $\log(1+a^2N(1+\rho(N-1)))$ where ρ is the correlation coefficient between y_i's and ε_t is normaly distributed. Fig. 5.3 illustrates how the complex interaction depends on the correlation. If the original signals are not correlated (black dashed line), taken as group they have increasingly higher interaction with y with the number of units. But this interaction is always higher the more positively cross-correlated they are. Conversely, negative cross-correlation reduces the interaction, all the way down to zero even though the target signal y is made up of each of these signals by construction. Collaborative activity enhances the interaction, but antagonistic activity reduces or even suppresses the interaction.

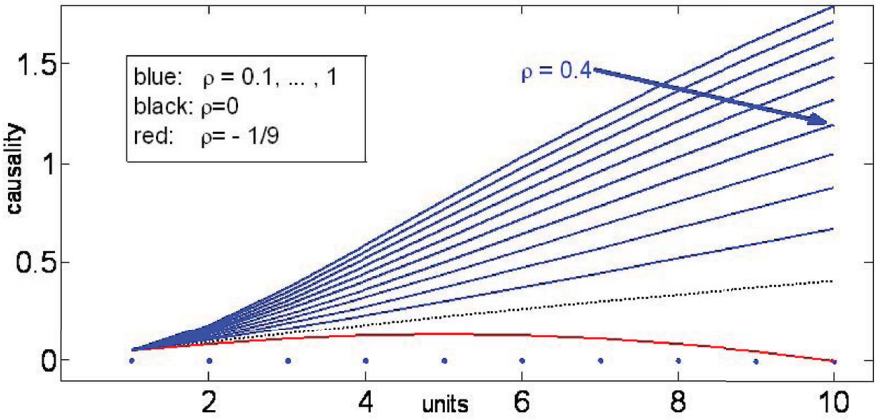

Fig. 5.3 The role of correlation in the complex interaction. The Granger causality vs. units (N) for different cross-correlation coefficients with a=0.022.(ρ = 1/9 is the smallest possible value for the cross-correlations of 10 units.)

5.5 Harmonic Granger causality

With partial Granger causality, the external influence is assumed present but not known in explicit form. However, there are situations in which a model of an external factor is available, for example the oscillating amount of sunlight received by an organism. In this section, we consider the inclusion of an harmonic oscillator in the formulation of Granger causality. This new term represents a periodically-changing envi-

ronmental input (*e.g.* sunlight, pH, temperature) and since its contribution to the prediction is explicit, it is possible to elucidate its Granger influence on each of the observed time series.

This extension should be amenable to a more general class of models: any function can be written as a sum of periodic functions through Fourier analysis ([27, 44]). Moreover, it should be possible to define it for more than two time series and use it like partial Granger causality. Both aspects are the subject of further study.

5.5.1 Time domain formulation

As we have seen in the previous sections, the core idea of Granger analysis is to measure how much the predicition of a target time series has improved with the knowledge of a possible source. The innovation of harmonic Granger causality is to have two possible sources of influences – another observed time series and the oscillator – which leads to two quantities of interest.

Consider a target time series X, a possible source Y and the periodic environmental factor E. As usual, we use an autoregressive model to predict X but this time we include different combinations of contributions:

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{1i} X_{t-i} + C_{1x}\cos(2\pi f_{1x} t + \phi_x) + \varepsilon_{1t} \\ X_t = \sum_{i=1}^{\infty} a_{2i} X_{t-i} + \sum_{i=1}^{\infty} b_{2i} Y_{t-i} + C_{2x}\cos(2\pi f_{2x} t + \phi_x) + \varepsilon_{3t} \\ X_t = \sum_{i=1}^{\infty} a_{3i} X_{t-i} + \sum_{i=1}^{\infty} b_{3i} Y_{t-i} + \varepsilon_{5t} \end{cases} \tag{5.31}$$

We denote Σ_i the variance of each error term ε_{it}. The only difference between the first and second equation is the inclusion of Y in the model. Therefore, we can naturally define the harmonic Granger causality from Y to X as:

$$F_{Y\to X} = \ln \frac{\Sigma_1}{\Sigma_3}$$

Similarly, the only difference between the second and the third equation is the inclusion of the oscillator. We define the harmonic Granger causality from E to X as:

$$F_{E\to X} = \ln \frac{\Sigma_5}{\Sigma_3}$$

These newly-defined quantities behave in the same manner as before: if Y (resp. E) does not contribute to X, $b_{2i} = 0$ and $\varepsilon_{3t} = \varepsilon_{1t}$, which implies $F_{Y\to X} = 0$. Otherwise, $\Sigma_1 > \Sigma_3$, since the fit has more degrees of freedom, and $F_{Y\to X} > 0$.

5.5.2 Two remarks about harmonic Granger causality

1. The addition of the harmonic term in the causality analysis is motivated by the fact that signals from experimental data are often periodic. A closer model would include more than one oscillator, to take into

account a larger number of background periodic influences. This Fourier-like method could lead to a more accurate picture of the network structure.

2. Harmonic Granger causality requires the fitting of three autoregressive models, whose parameters have to be estimated. When the model does not feature an harmonic term, a usual least-square fitting or classical more sophisticated techniques are sufficient and can easily be implemented [11, 33, 36]. When the autoregressive model includes an harmonic term, a two-step procedure can improve parameter estimation. First, we identify (manually or automatically) the dominating oscillation present in X by looking at its spectrum after Fourier transformation. The dominating frequency is filtered out to produce a new $\tilde{X}$ lacking this frequency, which is then fitted using usual methods [27, 44].

5.5.3 Frequency domain formulation

Harmonic Granger causality has an equivalent formulation in the frequency domain, whereby we can obtain the causality spectra showing the frequencies at which the influence of one node is exerted on another. Deriving the frequency domain formulation requires some mathematical manipulations in order to decompose the spectrum into clearly separated expressions.

To begin with, we rewrite the following auto-regressive model

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{2i} X_{t-i} + \sum_{i=1}^{\infty} b_{2i} Y_{t-i} + C_{2x}\cos(2\pi f_{2x} t + \phi_x) + \varepsilon_{3t} \\ Y_t = \sum_{i=1}^{\infty} c_{2i} Y_{t-i} + \sum_{i=1}^{\infty} d_{2i} Y_{t-i} + C_{2y}\cos(2\pi f_{2y} t + \phi_y) + \varepsilon_{4t} \end{cases} \tag{5.32}$$

in terms of the lag operator

$$\begin{pmatrix} D_{11}(L)\ D_{12}(L) \\ D_{21}(L)\ D_{22}(L) \end{pmatrix} \begin{pmatrix} X_t + O_t^x \\ Y_t + O_t^y \end{pmatrix} = \begin{pmatrix} \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix} \tag{5.33}$$

where O_t^x and O_t^y represent the oscillator added to X and Y respectively. We first normalize Eq. (5.33) by multiplying the matrix

$$P = \begin{pmatrix} 1 & 0 \\ -\frac{\Sigma_{34}}{\Sigma_3} & 1 \end{pmatrix} \tag{5.34}$$

to both sides of it, where $\Sigma_{34} = \mathrm{cov}(\varepsilon_{3t}, \varepsilon_{4t}), \Sigma_3 = \mathrm{var}(\varepsilon_{3t})$, The normalized equations are represented as:

$$\begin{pmatrix} \tilde{D}_{11}(L)\ \tilde{D}_{12}(L) \\ \tilde{D}_{21}(L)\ \tilde{D}_{22}(L) \end{pmatrix} \begin{pmatrix} X_t + O_t^x \\ Y_t + O_t^y \end{pmatrix} = \begin{pmatrix} \tilde{\varepsilon}_{3t} \\ \tilde{\varepsilon}_{4t} \end{pmatrix} \tag{5.35}$$

where $\tilde{\varepsilon}_{3t}$ and $\tilde{\varepsilon}_{4t}$ are now independent. Fourier-transforming both sides of Eq. (5.35) leads to

$$\begin{bmatrix} X(\omega) + O_x(\omega) \\ Y(\omega) + O_y(\omega) \end{bmatrix} = \begin{bmatrix} H_{xx}(\omega)\ H_{xy}(\omega) \\ H_{yx}(\omega)\ H_{yy}(\omega) \end{bmatrix} \begin{bmatrix} E_x(\omega) \\ E_y(\omega) \end{bmatrix} \tag{5.36}$$

where the transfer function is $H(\omega) = \tilde{\mathbf{D}}^{-1}(\omega)$. We can also rewrite Eq. (5.36) in the following expression:

$$\begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix} = \begin{bmatrix} H_{xx}(\omega)\ H_{xy}(\omega) \\ H_{yx}(\omega)\ H_{yy}(\omega) \end{bmatrix} \begin{bmatrix} E_x(\omega) \\ E_y(\omega) \end{bmatrix} + \begin{bmatrix} \tilde{O}_x(\omega) \\ \tilde{O}_y(\omega) \end{bmatrix} \tag{5.37}$$

where

$$\begin{bmatrix} \tilde{O}_x(\omega) \\ \tilde{O}_y(\omega) \end{bmatrix} = - \begin{bmatrix} O_x(\omega) \\ O_y(\omega) \end{bmatrix}$$

It can now be seen that X is defined as follows:

$$X(\omega) = H_{xx}(\omega)E_x(\omega) + H_{xy}(\omega)E_y(\omega) + \tilde{O}_x(\omega) \tag{5.38}$$

To obtain the frequency decomposition of the time domain causality, we look at the auto-spectrum of X_t

$$\begin{aligned} S_{xx}(\omega) &= X(\omega)X^*(\omega) \\ &= H_{xx}E_xE_x^*H_{xx}^* + \tilde{O}_x\tilde{O}_x^* + H_{xx}E_x\tilde{O}_x^* + \tilde{O}_xE_x^*H_{xx}^* \\ &\quad + H_{xy}E_y\tilde{O}_x^* + H_{xy}E_yE_y^*H_{xy}^* + \tilde{O}_xE_y^*H_{xy}^* \\ &= S_1 + S_2 \end{aligned} \tag{5.39}$$

$S_1 = H_{xx}E_xE_x^*H_{xx}^* + \tilde{O}_x\tilde{O}_x^* + H_{xx}E_x\tilde{O}_x^* + \tilde{O}_xE_x^*H_{xx}^*$, viewed as intrinsic part, involves only the variance of ε_{3t}, which is the noise term from the model for X_t. $S_2 = H_{xy}E_y\tilde{O}_x^* + H_{xy}E_yE_y^*H_{xy}^* + \tilde{O}_xE_y^*H_{xy}^*$, viewed as the causal part, involves only the variance of ε_{4t}, which is the noise term from the model for Y_t.

Note that if the harmonic term O_x is not present, there are only two terms in the expression for S_{xx}; it is consistent with the frequency decomposition of the time domain pairwise causality [11].

Finally, we can define the causal influence from Y_t to X_t at frequency ω as

$$f_{Y\to X}(\omega) = \ln\left(\frac{S_{xx}(\omega)}{S_1(\omega)}\right) \tag{5.40}$$

5.5.4 A Circadian circuit

Next we show a biological example to further confirm our partial complex Granger causality approach and harmonic Granger causality approach. We collected microarray data of Arabidopsis leaves of 32,448 genes that were observed over 11 days. The plants are grown in laboratory conditions, where they are subjected to 12 hours of artificial daylight followed by 12 hours of no light representing night time. Gene microarray data is collected at regular intervals (twice a day) throughout the experiment, so the data length is 22. A circadian circuit has been reported in the literature [13, 30, 31]. The circuit comprises of 8 genes: PRR5,PRR7, PRR9, ELF4, LHY, CCA1, TOC1 and GI. The time domain trace of the expression of these genes is shown in Fig. 5.5(**A**). Each of the genes with the exception of GI exhibits highly oscillatory behaviour with a time period of one day. This periodicity is attributed to the presence of incident sunlight during the day time and its absence during the night.

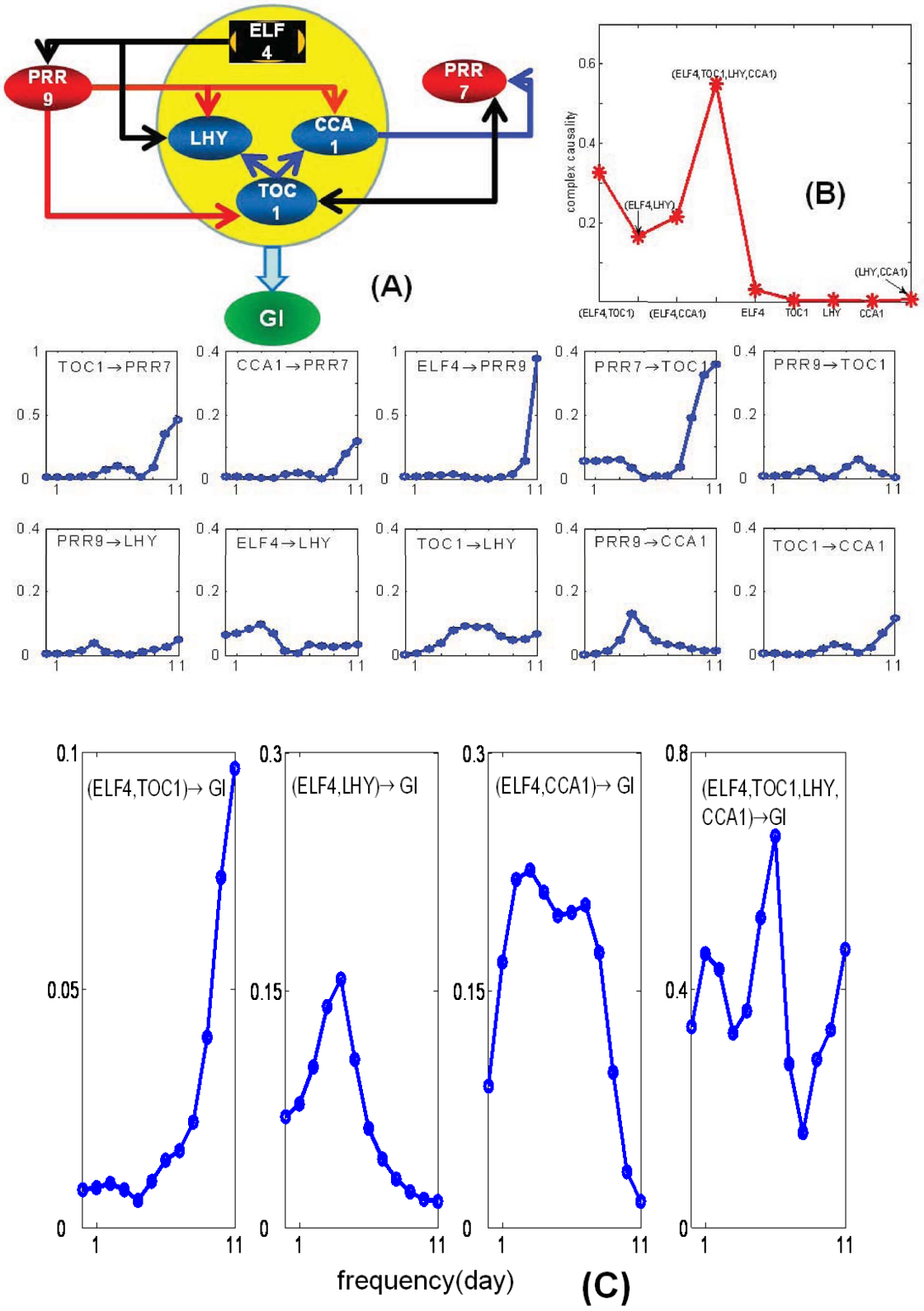

Fig. 5.4 Network of a circadian circuit in plant Arabidopsis leaf. (**A**). The gene circuit obtained in terms of Partial Granger causality, GI hasn't any interactions with other six genes when the relationships between single genes are considered. However, ELF4, TOC1, LHY, CCA1 together exhibit a significant interaction with GI. (**B**). Complex interactions between different group of genes and GI. (**C**) Gene interactions in the frequency domain.

By calculating node-to-node partial Granger causality between single node, we find the gene circuit as plotted in Fig. 18.3(**A**). GI is an isolated gene in our structure, without having any interactions with other six genes. In fact, this also coincides with the experimental findings. On page 4 [30], it is mentioned that *The GI single mutant had a relatively weak phenotype, whereas our assays of the triple GI; LHY; CCA1 mutant demonstrate GI's importance*. We thus turn our attention to partial complex Granger causality. Fig. 18.3(**B**) tells us that all single genes ELF4, TOC1, LHY, CCA1 and (LHY, CCA1) have very little influence on GI. However, ELF4, TOC1, LHY and CCA1 together exhibit a significant interaction with GI, which is in agreement with the experimental finding. We then analyse the interactions in the frequency domain, the detailed results are show in Fig. 18.3(**C**). No surprisingly, most of the interactions show a periodic behavior by exhibiting a peak at 11 day period.

Due to the strong oscillatory behaviour of the data shown in Fig. 5.5(**A**), in the next step, we use harmonic Granger causality to analyse it. The task regarding this data set is twofold, firstly we wish to identify which of the genes are driven by the external oscillation. And secondly, wish to determine how the genes are connected to form the network governing flowering of the plant[31, 41, 43]. The method to determine environmental input and network connectivity is as follows. There are 56 pairwise combinations possible with eight genes; for each of these 56 gene pairs the parameters of three candidate models as Eqs. 5.31 meaning with and without the oscillation term and with and without the causal term respectively are calculated. From the error term of each model we can infer both the presence of an external environmental driver and the possibility of a connection between the pair of genes. The final results are shown in Fig. 5.5(**B**). The plot reveals that four of the genes in this network receive external inputs: PRR9, CCA1, TOC1 and ELF4. The first two of these four genes agree with Ueda's network [45, 46]. The structure of Ueda's network is also very close to the structure of our network, both showing a high level of connectivity.

Here we use two different approaches: partial complex Granger causality and harmonic Granger causality to analyze the same data, the two structures shown in Fig. 18.3(**A**) and Fig. 5.5(**B**) are also different. The reason is that the network in Fig. 18.3(**A**) is inferred from partial Granger causality while network Fig. 5.5(**B**) is inferred from pairwise Granger causality. More reliable network would be constructed by combining both of these approaches, which is our further topic to be studied.

5.6 A comparative study between Granger causality and Bayesian network

A popular approach for building a causal network is the use of Bayesian networks ([6, 34, 35, 39]). Bayesian networks are based on the concept of conditional probability and are part of a class of probabilistic graphical models. While having sound theoretical foundations, they suffer some limitations, amongst which: a same set of probability distributions can have multiple graphical representations and all networks are directed acyclic graphs, which precludes feedback loops. The latter can be addressed by Dynamic Bayesian Networks ([51]), which consider how the data changes over time. The usual approach for the estimation of a Bayesian network is to decide on a scoring function (the likelihood of observing the data given a network structure) and search for the best candidate in the space of possible graphs. Given the size of this space (exponential in the number of nodes), one usually relies on sampling methods like Markov-Chain Monte Carlo ([5])), a long and computer intensive process.

In this section, we compare the performances of Granger causality and Bayesian Network on a simulated dataset of known structure in order to investigate the relative merits of the two methods. As a benchmark, we use a modified version of the system from example 1 where all exogenous inputs and latent variables have been removed:

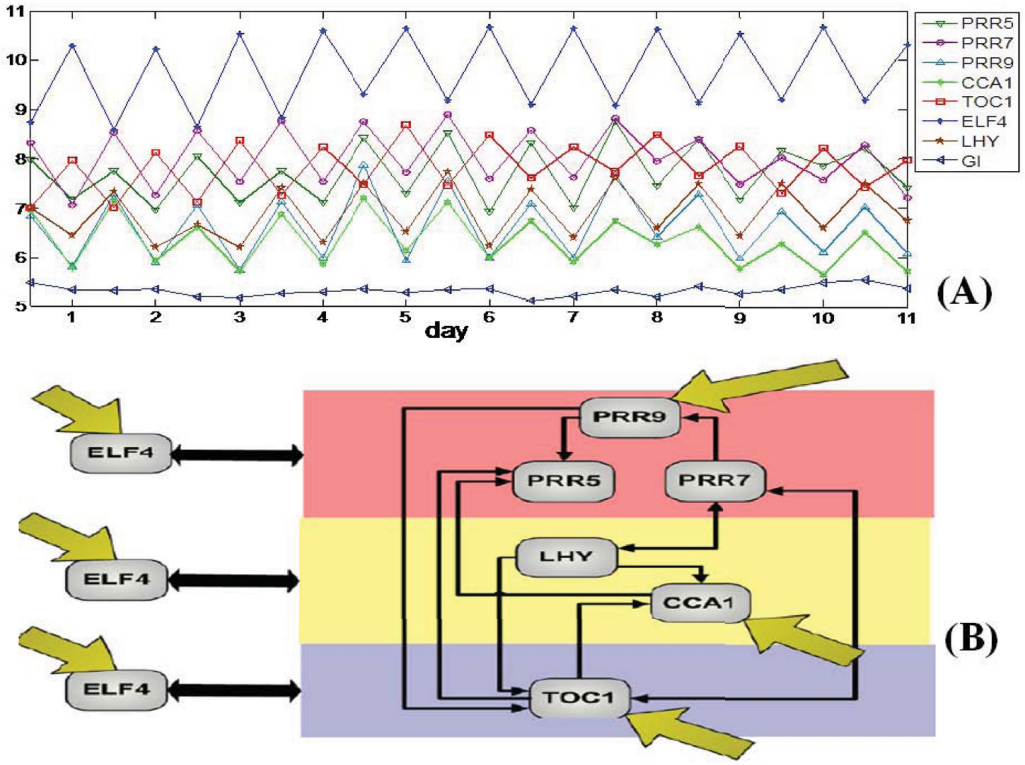

Fig. 5.5 (**A**) Time domain traces of gene expression of eight genes. (**B**) Network of a circadian circuit in plant Arabidopsis leaf using harmonic Granger causality approach. Four genes including PRR9, CCA1, TOC1 and ELF4 receive external inputs.

Example 3

$$\begin{cases} x_1(t) = 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + \varepsilon_1(t) \\ x_2(t) = 0.5x_1(t-2) + \varepsilon_2(t) \\ x_3(t) = -0.4x_1(t-3) + \varepsilon_3(t) \\ x_4(t) = -0.5x_1(t-2) + 0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \varepsilon_4(t) \\ x_5(t) = -0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + \varepsilon_5(t) \end{cases}$$

The experiment goes as follows: applied to the same data, both Granger causality and a dynamic Bayesian network produce a tentative network. The operation is done 100 times and a final network from each approach is built with edges that appear at least 95% of the time. The sample size (or number of time points) varies from 1000 down to 20, in order to investigate its impact on accuracy . Fig. 5.6 shows the how the performances of the two methods, with Bayesian networks on the right hand side and Granger Causality on the left hand side.

From this experiment, we find that both approaches can reveal correct network structures for the data with a large sample size (1000 here). As one would expect, the accuracy decreases with the size of the

data, with more and more links gone missing but the two methods perform identically. However, at $n = 20$, Granger causality lacks all correct links, while the Bayesian network still features true positives ($X_1 \to X_4$ and $X_1 \to X_2$). The reason for the poorer performance of Granger causality is that at small sample size the linear fit becomes less constrained, which makes the Granger coefficient unstable.

In conclusion, both Granger causality and Bayesian Networks are sensitive to sample size[52]. With long enough time series, they are as accurate as each other and the choice for one or the other should be done on other parameters, like computational requirements (light for Granger causality, heavy for Bayesian networks) or inclusion of *prior* knowledge (natural in a Bayesian setting).

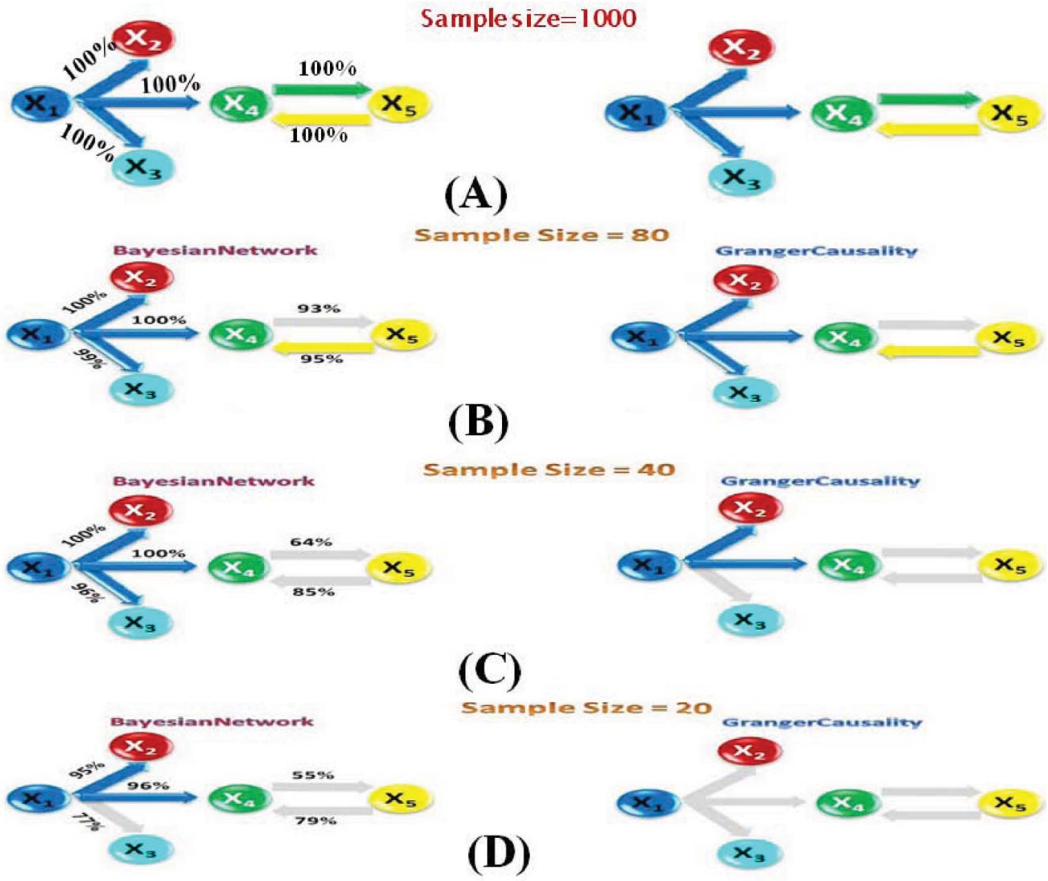

Fig. 5.6 Networks inferred with Bayesian Network and Granger causality for various sample sizes. Grey edges indicate undetected causalities (false negatives). For each sample size n, we simulated a data set of 100 realizations of n time points. High-confidence arcs, appearing in at least 95% of the 100 networks are shown. (**A**) The sample size is 1000. (**B**) The sample size is 80. (**C**) The sample size is 40. (**D**) The sample size is 20.

5.7 Unified Causal Model (UCM)

A successful approach for the study of brain region communication is the Dynamic Causal Model (DCM, [16, 17]). DCM uses a model of how the data were generated, equating the observed variables to the result of a convolution of a linear combination of latent variables. In a functional magnetic neuroimaging context, the convolution captures the varying delays on the blood flow (haemodynamics response). Moreover, the model includes the effect of a deterministic input – for example a stimulus – on the signal dynamics. A causal inference can be drawn from fitting the model (and other competing models) and estimating the likelihood of observing the data. Based on these two properties, DCM seems radically different from the Granger Causality Model (GCM). GCM does not impose a model but is purely based on the data statistics. And GCM does not natively incorporate the notion of input – although some extensions do, as seen in this chapter, *e.g.* Harmonic Granger causality, Sec. 5.5.

The Unified Causal Model is an attempt at reconciling the two approaches, whose mathematical formulations are actually quite close. We first recall the definition of DCM:

$$\begin{aligned} \frac{dX}{dt} &= (A + u_t B)X_t + u_t C \\ Y_t &= g(X) + \varepsilon_t \end{aligned} \tag{5.41}$$

where X represents brain region states and are unobserved. Matrices A,B and C model which interactions take place and to what extent and a deterministic input is represented by u_t. The function g is the convolution capturing the physiological effect of the blood flow. The observations Y_t are used to estimate the validity of the model. The traditional Granger causality has to be modified in such a way that it includes the deterministic input in a very similar fashion. We write the two modified auto-regressive models as:

$$\begin{aligned} X_t &= \textstyle\sum_{i=1}(a_{1i} + b_{1i}u_{t-i})X_{t-i} + c_1 v_{t-1} + \varepsilon_{1t} \\ X_t &= \textstyle\sum_{i=1}(a_{2i} + b_{2i}u_{t-i})X_{t-i} + c_2 v_{t-1} + \sum_{i=1}(d_{2i} + e_{2i}u_{t-i})Y_{t-i} + \varepsilon_{2t} \end{aligned} \tag{5.42}$$

where both X_t and Y_t are two observed signals. As usual, we write X_t first in terms of its past, then adding the knowledge of Y_t. The definition of the Granger causality is the measure of the prediction improvement:

$$F_{Y\to X} = \ln \frac{\mathrm{var}(\varepsilon_1)}{\mathrm{var}(\varepsilon_2)} \tag{5.43}$$

Thus modified, Granger causality is able to use known input signals in a way similar to that of DCM. It is also possible to derive a frequency domain formulation of the Granger causality. This method has been applied on local field potential recordings in sheep's brain and demonstrated learning induced changes in inter- and intra-hemispheric connectivity [19].

5.8 Large Networks

So far, the focus has been put on relatively small networks, with 20 nodes at most. The reason for this is purely numerical: with a limited number of time points, as is the case with experimental data, the linear auto-regressive model fit rapidely becomes under-constrained when the number of signals increases and the

solution ceases to be unique. The resulting Granger causality, defined as a ratio, is unstable and unreliable. For example, a partial Granger causality between two signals, conditioned on n others requires the estimation of $(2+n)^2 p$ parameters, where p is the order (maximum time delay) of the model; for a dataset of 10 signals and the minimum order $p = 1$, one needs at least 100 time points in order to calculate the Granger causality between pairs of nodes conditioned on the rest.

To bypass this difficulty, we proposed a procedure for building large networks of hundreds of nodes using partial Granger causality. The rationale is as follows: if the usual Granger causality from $Y \to X$ is large but significantly decreases when conditioned on a third signal Z ($F_{Y\to X|Z}$), then the connection $Y \to X$ is only indirect and should be discarded. We use this principle to find the direct ancestors (signals acting on a target X with no intermediate) of each nodes. At step 0, we search for all signals Y such that $F_{Y\to X}$ is large. We call Ω_0 this collection of candidate ancestors. At step 1, we filter this set further with keeping the signals $Y \in \Omega_0$ such that $F_{Y\to X|Z}$ is still large for all $Z \in \Omega_0$. We call Ω_1 this new set and carry on the procedure by conditioning on groups of 2, then 3 etc. signals until such an operation is not possible (the size of Ω_i decreases at each iteration). The result is a list of direct ancestors for each node, which we aggregate to produce the global network.

We test the validity of this approach on simulated data. We built an Erdös-Rényi random graph with $N = 200$ nodes and $M = N \ln N = 1060$ edges. We generate N time series with an auto-regressive model such that they follow the random network's structure: the transition matrix A is build from the transpose of network's adajency matrix by replacing non-zero entries by a random value. The matrix A is then scaled to have a maximum eigenvalue less than 1, in order to make the system stable. Each time series is 200 time-points long and normal noise of unit variance is added throughout.

Fig. 5.7(A) shows the resulting receiver operating characteristic (ROC) curve [14], that is the graph obtained by plotting the false positive rate against the true positive rate. A random guess is represented by the dashed line. The method shows a maximum true positive rate just over 0.5, which is not very high. However, the false positive rate is always very low: the method misses many ancestors but its guesses are rarely wrong. This is crucial for biological applications: it meancs that the results can be used in further experiments, for example by indicating which protein/gene to manipulate. Fig. 5.7(B) shows how the true positive rate varies with respect to the strength of the connection (*i.e.* the associated weight in the transition matrix). Weak connections are more likely missed out.

The procedure is easy to parallelize (each node can be processed separately) and easy to implement. It has been applied on a dataset of 812 proteomic time series ([54]) to produce a large and complete network.

5.9 Summary

In this chapter, we introduced several important extensions of Granger causality that have been devised for tackling issues specific to biological phenomena. We started with *Partial Granger causality*, which is not only able to remove the influence of other observables but also to reduce the influence of external, unseen inputs. This is of considerable importance for biological data where typically only a small portion of the quantities of interest is actually observed. We showed that this extension (and the others) has its counterpart in the *frequency domain*, so that instead of summarizing the strength of the connection in one single number as is the case in the time domain, we can obtain a spectrum indicating at which frequency the interaction takes place. Partial Granger causality has then been further extended to *Complex Granger causality*. Complex Granger causality captures the effect of group action, a frequent occurence in biology where the whole is often more significant than the parts. *Harmonic Granger causality* used an explicit form of possible external

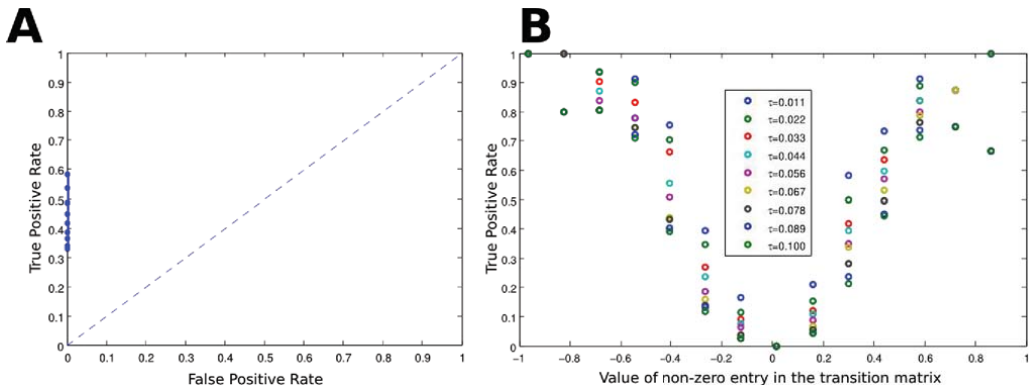

Fig. 5.7 Perfomance of the network building procedure. (A) Receiver Operator Characteristic curve. (B) Sensitivity of the procedure to the connection strength.

influence for modeling periodic inputs like sunlight. Since it is one of the most popular approaches for reverse-engineering biological systems, *Bayesian networks* has been systematically compared with the much simpler Granger causality and it was shown that they both give similar results provided that the number of time points is large enough. It was also shown that another popular approach for studying brain imaging data in particular called Dynamic Causal Model, and Granger causality both can be seen as a particular case of the more general *Unified Causal Model*. And finally, we presented a new method for building *large networks* of hundreds of nodes with limited data, an standing problem in systems biology.

Acknowledgments

This work is supported by grants from EPRSC (UK, CARMEN EP/E002331/1) and EU grant (BION) and NSFC (China).

Appendix: Estimating the error covariance matrix

The main quantity of interest for calculating Granger causality is the covariance matrix of the error term ε in the d-dimensional auto-regressive (AR) model

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon(t)$$

The coefficients of the matrices A_i need to be estimated from the data. Typically, they are found by minimizing the variance of the error between the prediciton X_t and the observation at the same time t. Morf's procedure [33] provides a fast and robust way of estimating A_i and the covariance matrix of ε via a recursive algorithm.

Unless prior knowledge informs us about the likely order p of the AR model (for example by knowing at which time-scale one should expect interactions to take place), it also has to be estimated from the data. The goodness of fit alone is not sufficient for selecting the optimal order: adding a new order implies adding d^2 more unknowns to the system, which considerably improves the fitting simply by increasing the degrees of freedom. The Akaike Information Criterion (AIC, [2]) provides a trade-off between the model complexity (a function of p) and the fit. For an AR model of dimension d, of order p and with T observations, this quantity can be written as:

$$\mathtt{AIC}(p) = 2d^2p + d(T-p)\log\left(2\pi \sum_{\substack{t=p+1\ldots T \\ i=1\ldots d}} \frac{\varepsilon_i^2(t)}{d(T-p)}\right)$$

The optimal order p is then defined as the one that minimizes the AIC.

A point estimate of the Granger causality is often not sufficient for making a strong conclusion about the data and a confidence interval is required. In some cases ([25] for Partial Granger Causality), it is possible to derive the confidence interval in closed form. In general, it is estimated via a bootstrap procedure ([32]) – given the optimal AR model, an ensemble of signals are generated which each produce a value for the Granger causality. The final estimation of the Granger causality and its confidence interval are defined as the average and the standard error of this set of estimates.

References

1. R. Aebersold, L.E. Hood, and J.D. Watts. Equipping scientists for the new biology. *Nature Biotechnology*, **18**(4): 359, 2000.
2. H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **19**(6):716–723, December 1974.
3. U. Alon. Biological networks: the tinkerer as an engineer. *Science*, **301**(5641): 1866–1867, 2003.
4. D. Anastassiou. Computational analysis of the synergy among multiple interacting genes. *Mol Syst Biol*, **3**: 83, February 2007.
5. C. Andrieu, N. de Freitas, A. Doucet, and M.I. Jordan. An introduction to mcmc for machine learning. *Machine Learning*, **V50**(1):5–43, January 2003.
6. W.L. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, **2**: 159, 1994.
7. I. Cantone, L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo and M.P. Cosma. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, **137**(1): 172–181, 2009.
8. B. Chance, R.W. Estabrook and A. Ghosh. Damped sinusoidal oscillations of cytoplasmic reduced pyridine nucleotide in yeast cells. *Proceedings of the National Academy of Sciences*, **51**(6): 1244–1251, 1964.
9. Y. Chen, S.L. Bressler and M. Ding. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *Journal of neuroscience methods*, **150**(2): 228–237, 2006.
10. J.J. Chrobak and G. Buzsaki. Gamma oscillations in the entorhinal cortex of the freely behaving rat. *Journal of Neuroscience*, **18**: 388–398, 1998.
11. M. Ding, Y. Chen and S.L. Bressler. Granger causality: Basic theory and application to neuroscience. In J. Timmer, B. Schelter, M. Winterhalder, editor. *Handbook of Time Series Analysis*, pages 451–474. Wiley-VCH Verlage, 2006.
12. D.M. Camacho and J.J. Collins. Systems biology strikes gold. *Cell*, **137**(1): 24–26, 2009.
13. M.R. Doyle, S.J. Davis, R.M. Bastow, H.G. McWatters, L. Kozma-Bognár, F. Nagy, A.J. Milla and R.M. Amasino. The elf4 gene controls circadian rhythms and flowering time in arabidopsis thaliana. *Nature*, **1419**: 74–77, 2002.
14. T. Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, **27**(8): 861–874, June 2006.
15. J.F. Feng, D.Y. Yi, R. Krishna, S.X. Guo and V. Buchanan-Wollaston. Listen to genes: Dealing with microarray data in the frequency domain. *PLoS ONE*, **4**(4): e5098+, April 2009.
16. K.J. Friston, L. Harrison and W. Penny. Dynamic causal modelling. *NeuroImage*, **19**(4): 1273–1302, August 2003.

17. K. Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol*, **7**(2):e1000033+, February 2009.
18. T.S. Gardner, D. di Bernardo, D. Lorenz and J.J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**(5629): 102–5, 2003.
19. T. Ge, K.M. Kendrick and J.F. Feng. A unified dynamic and granger causal model approach demonstrates brain hemispheric differences during face recognition learning. *PLoS Comp. Biol.*, submitted, 2009.
20. J.F. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, **77**(378): 304–313, 1982.
21. J.F. Geweke. Measures of conditional linear-dependence and feedback between time series. *Journal of the American Statistical Association*,**79**(388): 907–915, 1984.
22. B. Gourévitch, R.L. Bouquin-Jeannès and G. Faucon. Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biology Cybern.*, **95**(4): 349–369, 2006.
23. C. Granger. Investigating causal relations by econometric models and cross- spectral methods. *Econometrica*, **37**: 424–438, 1969.
24. C. Granger. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, **2**: 329–352, 1980.
25. S. Guo, A.K. Seth, K.M. Kendrick, C. Zhou and J.F. Feng. Partial Granger causality–eliminating exogenous inputs and latent variables. *Journal of neuroscience methods*, **172**(1): 79, 2008.
26. S. Guo, J. Wu, M. Ding, and J.F. Feng. Uncovering interactions in the frequency domain. *PLoS Computational Biology*, **4**(5): e1000087, 2008.
27. S He. Estimation of the mixed AR and hidden periodic model. *Acta Mathematicae Applicatae Sinica-English Series*, **13**(2): 196–208, 1997.
28. E. Klipp, R. Herwig, A. Kowald, C. Wierling and H. Lehrach. Systems biology in practice: concepts, implementation and application, 2005.
29. C. Ladroue, S.X. Guo, K. Kendrick and J.F. Feng. Beyond element-wise interactions: Identifying complex interactions in biological processes. *PLoS ONE*, **4**(9): e6899, 2009.
30. J.C. Locke, L. Kozma-Bognar, P.D. Gould, B. Feher, E. Kevei, F. Nagy, M.S. Turner, A. Hall and A.J. Millar. Experimental validation of a predicted feedback loop in the multi-oscillator clock of arabidopsis thaliana. *Molecular System Biology*, **2**:59, 2006.
31. H.G. McWatters, E. Kolmos, A. Hall, M.R. Doyle, R.M. Amasino, P. Gyula, F. Nagy, A.J. Millar and S.J. Davis. ELF4 is required for oscillatory properties of the circadian clock. *Plant physiology*, **144**(1): 391, 2007.
32. D.S. Moore. *The Basic Practice of Statistics*. W.H.Freeman & Co Ltd, 2003.
33. M. Morf, A. Vieira, D.T.L. Lee and T. Kailath. Recursive multichannel maximum entropy spectral estimation. *Geoscience Electronics, IEEE Transactions on*, **16**(2): 85–94, 1978.
34. S. Mukherjee and T.P. Speed. Network inference using informative priors. *Proceedings of the National Academy of Sciences*, **105**(38): 14313–14318, 2008.
35. C.J. Needham, J.R. Bradford, A.J. Bulpitt and D.R. Westhead. A primer on learning in Bayesian networks for computational biology. *PLoS Comput Biol*, **3**(8): e129, 2007.
36. A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.*, **27**(1):27–57, March 2001.
37. J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.
38. J. Quackenbush. Computational analysis of microarray data. *Nature Reviews Genetics*, **2**(6):418–427, 2001.
39. K. Sachs, O. Perez, D. Pe'er, D.A. Lauffenburger, and G.P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, **308**(5721): 523–529, 2005.
40. M. Schelter, B. an Winterhalderm and J. Timmer. *Handbook of time series analysis: recent theoretical developments and applications*. Wiley-VCH, Weinheim, 2006.
41. T.F. Schultz and S.A. Kay. Circadian clocks in daily and seasonal control of development. *Science*, **301**(5631): 326–328, 2003.
42. T.P. Speed. *Statistical analysis of gene expression microarray data*. CRC Press, 2003.
43. A.N. Stepanova and J.M. Alonso. Arabidopsis ethylene signaling pathway. *Science*, **276**: 1872–1874, 2005.
44. G.C. Tiao and M.R. Grupe. Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, **67**(2): 365–373, 1980.
45. H.R. Ueda. Systems biology flowering in the plant clock field. *Molecular System Biology*, **2**: 60, 2006.
46. H.R. Ueda, W.B. Chen, A. Adachi, H. Wakamatsu, S. Hayashi, T. Takasugi, M. Nagano, K. Nakahama, Y. Suzuki, S. Sugano, M. Iino, Y. Shigeyoshi and S. Hashimoto. A transcription factor response element for gene expression during circadian night. *Nature*, **418**(6897): 534–539, 2002.
47. N. Wiener. The theory of prediction. *Modern mathematics for engineers, Series*, **1**: 125–139, 1956.

48. J.H. Wu, K. Kendrick and J.F. Feng. Detecting correlation changes in electrophysiological data. *Journal of Neuroscience Methods*, **161**(1): 155–65, 2007.
49. J.H. Wu, X.G. Liu and J.F. Feng. Detecting causality between different frequencies. *Journal of Neuroscience Methods*, **167**(2): 367–75, 2008.
50. J.H. Wu, J.L. Sinfield and J.F. Feng. Analysis of biological rhythms using a harmonic causal method. *BMC Systems Biology*, submitted, 2009.
51. J. Yu, A.V. Smith, P.P. Wang, and A.J. Hartemink. Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, **20**(18):3594–3603, December 2004.
52. C.L. Zou and J.F. Feng. Granger causality vs. dynamic bayesian network inference: a comparative study. *BMC Bioinformatics*, **10**(1):122, 2009.
53. C.L. Zou, K.M. Kendrick and J.F. Feng. The fourth way: Granger causality is better than the three other reverse-engineering approaches. *Cell*,http://www.cell.com/comments/S0092-8674(09)00156-1, 2009.
54. C.L. Zou, C. Ladroue, S.X. Guo and J.F. Feng. Identifying interactions in the time and frequency domains in local and global networks. *Mol. Sys. Biol.*, 2009.
55. M. Zylka, L. Shearman, J. Levine, X. Jin, D. Weaver and S. Reppert. Molecular analysis of mammalian *timeless*. *Neuron*, **21**(5): 1115–1122, November 1998.