Chapter 22
Granger Causality: Its Foundation and Applications in Systems Biology

Tian Ge
Fudan University, China

Jianfeng Feng
Fudan University, China & University of Warwick, UK

ABSTRACT

As one of the most successful approaches to uncover complex network structures from experimental data, Granger causality has been widely applied to various reverse engineering problems. This chapter first reviews some current developments of Granger causality and then presents the graphical user interface (GUI) to facilitate the application. To make Granger causality more computationally feasible and satisfy biophysical constraints for dealing with increasingly large dynamical datasets, two attempts are introduced including the combination of Granger causality and Basis Pursuit when faced with non-uniformly sampled data and the unification of Granger causality and the Dynamic Causal Model as a novel Unified Causal Model (UCM) to bring in the notion of stimuli and modifying coupling. Several examples, both from toy models and real experimental data, are included to demonstrate the efficacy and power of the Granger causality approach.

INTRODUCTION

With the rapid progress in the development of experimental techniques, more and more high-throughput datasets measuring temporal behavior of hundreds of or even thousands of proteins or genes are offering rich opportunities for researchers. In order to exploit the full potential of these approaches, we have to be able to convert the resulting data into the most appropriate framework to account for the functioning of the underlying biological system. Over the past two decades, a variety of attempts have been carried out in this field and reverse engineering approaches to uncover network structures in genes, proteins,

DOI: 10.4018/978-1-60960-491-2.ch022

neurons and brain areas are still one of the hottest topics in computational systems biology.

Causality analysis based upon experimental data has become one of the most powerful and valuable tools in discovering connections between different elements in complex biological systems (Cantone et al., 2009; Camacho & Collins, 2009). Comparing approaches including information theory, control theory or Bayesian statistics, here we focus on another successful approach: Granger causality, which is based upon simple ideas and has a concise theory but is even more powerful to capture the nature and dynamics of a biological system. As an example, in one recent comment on a paper in Cell, we have demonstrated that Granger causality outperforms all the other approaches the authors had employed to build causal networks (Zou et al., 2009).

The basic idea of Granger causality can be traced back to Wiener (Wiener, 1956) who put forward the notion that if the prediction of one process can be improved by incorporating the past information of the second process, then the second process causes the first one. Later, Granger followed this point and formalized it in the context of linear regression models (Granger, 1969). Geweke's decomposition of a vector autoregressive process endowed Granger causality with a spectral representation (Geweke, 1982, 1984) and made the interpretation more informative in that interactions in different frequency bands could be clearly figured out instead of only in a single number. Recently, a series of papers based upon its original formalism have been published to make Granger causality suitable to address biological and computational issues in different situations. These useful extensions include partial Granger causality (Guo et al., 2008) which is able to eliminate the influences of exogenous inputs and latent variables; complex Granger causality (Ladroue et al., 2009) which can uncover the interactions among groups of time-series and harmonic Granger causality (Wu et al. 2008) which introduces a model with an oscillating external input and puts special emphasis on environmental effects. These methods can be combined to identify interactions in the time and frequency domains in local and global networks. Furthermore, detailed and intensive comparisons between Granger causality and Bayesian networks have also been carried out (Zou & Feng, 2009). In this chapter, we first apply well established Granger causal analysis approaches to microarray data from *Arabidopsis thaliana* (Arabidopsis) to recover a well-known gene circuit. Our graphical user interface (GUI) is also presented to facilitate the application. These will show the power of Granger causality and its convenient implementation.

In spite of all the successful extensions and applications of Granger causality mentioned above, some limitations still exist which restrict its application on a broader basis. The first issue we encounter, whatever the approach is to be applied to a set of data, is preprocessing. Preprocessing should not be ignored and can sometimes play a critical role in determining final conclusions. A brute-force application of Granger causality could simply result in false or erroneous conclusions. Two general approaches of preprocessing in dealing with temporal data are down-sampling and up-sampling, after filtering out noise or extreme points. We have found that both techniques are useful and are commonly implemented before doing further analysis. The choice of down-sampling or up-sampling depends on the nature of data. In neurophysiology, the original data are usually sampled at a very high frequency, for example, 2 kHz. Even if we fit the data with an autoregressive model with an order of 20, the model only covers a time window of 10 milliseconds, a very short duration. As a result, information in the low frequency band could be lost and the features of slow oscillations such as theta rhythms (4-8 Hz) are difficult to be captured. On the other hand,