Journal of Neuroscience Methods 223 (2014) 50–68

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference

Lionel Barnett*, Anil K. Seth

Sackler Centre for Consciousness Science and School of Informatics, University of Sussex, Brighton BN1 9QJ, UK

HIGHLIGHTS

- Matlab© Toolbox for accurate and efficient calculation of Granger causality.
- Calculate Granger causalities (conditional and unconditional) in both time and frequency domains.
- Based on advanced VAR (vector autoregression) theory.
- Avoids separate "full" and "reduced" regressions.

ARTICLE INFO

Article history:
Received 5 September 2013
Received in revised form 16 October 2013
Accepted 26 October 2013

Keywords:
Granger causality
Vector autoregressive modelling
Time series analysis

ABSTRACT

Background: Wiener–Granger causality ("G-causality") is a statistical notion of causality applicable to time series data, whereby cause precedes, and helps predict, effect. It is defined in both time and frequency domains, and allows for the conditioning out of common causal influences. Originally developed in the context of econometric theory, it has since achieved broad application in the neurosciences and beyond. Prediction in the G-causality formalism is based on VAR (vector autoregressive) modelling.
New method: The MVGC Matlab© Toolbox approach to G-causal inference is based on multiple equivalent representations of a VAR model by (i) regression parameters, (ii) the autocovariance sequence and (iii) the cross-power spectral density of the underlying process. It features a variety of algorithms for moving between these representations, enabling selection of the most suitable algorithms with regard to computational efficiency and numerical accuracy.
Results: In this paper we explain the theoretical basis, computational strategy and application to empirical G-causal inference of the MVGC Toolbox. We also show via numerical simulations the advantages of our Toolbox over previous methods in terms of computational accuracy and statistical inference.
Comparison with existing method(s): The standard method of computing G-causality involves estimation of parameters for both a full and a nested (reduced) VAR model. The MVGC approach, by contrast, avoids explicit estimation of the reduced model, thus eliminating a source of estimation error and improving statistical power, and in addition facilitates fast and accurate estimation of the computationally awkward case of conditional G-causality in the frequency domain.
Conclusions: The MVGC Toolbox implements a flexible, powerful and efficient approach to G-causal inference.

1. Introduction

Wiener–Granger causality (G-causality) (Granger, 1969; Geweke, 1982, 1984) is an increasingly popular method for identifying "causal" connectivity in neural time series data (Bressler and Seth, 2011). It can be traced conceptually to Wiener (1956) and was operationalised by Granger in terms of linear autoregressive modelling of stochastic processes (Granger, 1969). G-causality is based on *predictability* and *precedence*. Put simply, a variable X is said to G-cause a variable Y if the past of X contains information that helps predict the future of Y over and above information already in the past of Y. Importantly, G-causality is a measure of *directed functional connectivity* in terms of providing a statistical description of observed responses. In contrast, methods for identifying *effective connectivity* aim to elucidate the "the simplest possible circuit diagram explaining observed responses" (Valdes-Sosa et al., 2011; Aertsen and Preißl, 1991; Friston et al., 2003, 2013).

The MVGC Matlab Toolbox implements numerical routines for calculating multivariate Granger causality (MVGC) from time series

* Corresponding author. Tel.: +44 1273 678101.
E-mail addresses: l.c.barnett@sussex.ac.uk (L. Barnett), a.k.seth@sussex.ac.uk (A.K. Seth).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.10.018

data, both unconditional and conditional, in the time and frequency domains. It supersedes and extends the GCCA (Granger Causal Connectivity Analysis) Toolbox (Seth, 2010). Based on advanced VAR (vector autoregressive) model theory (Hamilton, 1994; Lütkepohl, 2005), it is designed for computational efficiency and accuracy. Improving upon standard approaches to G-causal inference, it avoids separate full and reduced regressions (see Section 2), thus eliminating a common source of statistical inaccuracy and computational inefficiency. It also facilitates estimation of fully multivariate conditional G-causality in the frequency domain.

The MVGC toolbox has been designed with application to empirical neuroscience data in mind. Several issues commonly faced in applying G-causal inference to such data are discussed later in the paper (Section 4). However, the software is not restricted to this application domain. G-causal inference is a very general method which has been productively applied in many areas, though caution is always needed in ensuring that the data satisfy the assumptions underpinning the method (see Section 2.1 and also Section 3.3). Thus, while the MVGC software is designed to be intuitive and straightforward to use, to get the most out of it some understanding of the theoretical basis of G-causal inference, as well as approaches to its numerical computation, is recommended. To facilitate this, this paper presents the conceptual and theoretical basis of G-causality (substantially extending previous presentations), and details the computational implementation of the MVGC toolbox. Although not intended as a user guide—the integrated help system and walk-through demonstration scripts fulfill that purpose—Section 3 provides a helpful overview of the toolbox rationale and design principles, Section 3.1 explains the MVGC computational strategy while Appendices A.1 and A.2 detail the core computational algorithms.

2. G-causality: theory, estimation and inference

Assume two jointly distributed vector-valued stochastic processes ("variables") $\boldsymbol{X} = \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, $\boldsymbol{Y} = \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$. We say that $\boldsymbol{Y}$ does *not* G-cause $\boldsymbol{X}$ if and only if $\boldsymbol{X}$, conditional on its own past, is independent of the past of $\boldsymbol{Y}$; intuitively, past values of $\boldsymbol{Y}$ yield no information about the current value of $\boldsymbol{X}$ beyond information already contained in the past of $\boldsymbol{X}$ itself. If, conversely, the past of $\boldsymbol{Y}$ *does* convey information about the future of $\boldsymbol{X}$ above and beyond all information contained in the past of $\boldsymbol{X}$ then we say that $\boldsymbol{Y}$ G-causes $\boldsymbol{X}$. Much has been made of the fact that this notion of "causality" does not necessarily tally with more conventional (and perhaps more physically intuitive) notions (Pearl, 2009; Valdes-Sosa et al., 2011; Friston, 2011; Roebroeck et al., 2009, 2010) (note, for example, that the processes must be strictly non-deterministic for the above definition even to make sense). We do not engage this debate here; throughout this paper the term "causal" is used exclusively in the Wiener–Granger sense just described.

Testing for Granger (non-)causality requires establishing—in a statistical sense—a conditional (non-)dependency. This may seem to invite an information-theoretic approach, since mutual information is a natural measure of statistical dependence. However in empirical settings a drawback with information-theoretic quantities is the difficulty of estimation in sample and lack of known theoretical distributions for information-theoretic estimators, complicating statistical inference (but see Barnett and Bossomaier, 2013). An alternative is therefore to invoke a model-based or parametric approach for which conditional dependency estimators are more efficient and (preferably) have known sampling distributions, thus facilitating inference. In fact, G-causality is frequently identified with a model-based viewpoint, which may also yield intuitive predictive interpretations of G-causality; for predictive models, the statement "$\boldsymbol{X}$, conditional on its own past, is independent of the past of $\boldsymbol{Y}$" may be reframed as "the past of $\boldsymbol{Y}$ does not help predict $\boldsymbol{X}$ beyond the degree to which $\boldsymbol{X}$ may be predicted by its own past".

The most common operationalisation of G-causality, and the one on which the MVGC Toolbox is based, utilises VAR modelling of time series data. A multivariate time series $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$, where for each time t $\boldsymbol{u}_t$ is a real-valued n-dimensional (column) vector with components $u_{1t}, u_{2t}, \ldots, u_{nt}$, is considered as a realisation of length m of a discrete-time stationary[1] vector stochastic process $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ (the "universe of variables"). A pth order vector autoregressive model for the process—a VAR(p)—takes the form[2]

$$\boldsymbol{U}_t = \sum_{k=1}^{p} A_k \cdot \boldsymbol{U}_{t-k} + \boldsymbol{\varepsilon}_t \tag{1}$$

Here p is the *model order*, which may be infinite. The $n \times n$ real-valued matrices A_k are the *regression coefficients*, and the n-dimensional stochastic process $\boldsymbol{\varepsilon}_t$ the *residuals*, which are independently and identically distributed (iid) and serially uncorrelated; that is, they constitute a white noise process. The parameters of the model are the coefficients A_k and the $n \times n$ *residuals covariance matrix* $\Sigma \equiv \mathrm{cov}(\boldsymbol{\varepsilon}_t)$ which, by stationarity, does not depend on the time t. Interpreted as a *predictive* model, (1) thus models the value of the process at current time t in terms of its past values at times $t-1, \ldots, t-p$. The regression coefficients represent the predictable structure of the data, the residuals the unpredictable.

The determination of a VAR model (1) should not be taken to imply that the time series data modelled by the stochastic process $\boldsymbol{u}_t$ was actually generated by a linear autoregressive scheme. In comparison to effective connectivity techniques like dynamic causal modelling [DCM] (Friston et al., 2003), which make very explicit assumptions about the generative mechanism underlying the observed data (Friston et al., 2013), the VAR models underlying G-causality are "generic", in the sense that they make no assumptions about the mechanism that produced the data, beyond that a VAR model actually exists. Standard theory (Anderson, 1971) yields that a rather general class of covariance-stationary multivariate process—including many nonlinear processes—may be modelled as VARs, albeit of theoretically infinite order (see also the next section). Note that this belies a common misapprehension that VAR-based G-causal analysis is restricted to linear processes. In practice, given empirical time-series data, a finite model order p is selected on theoretical principles, sufficiently high so as to capture predictable variation but not so high as to overfit unpredictable variation in the data (Section 2.4).

2.1. VAR process theory

The MVGC toolbox exploits some advanced aspects of VAR process theory, which we outline here.

For valid G-causality analysis, the VAR coefficients in (1) must be *square summable* and *stable* (Hamilton, 1994; Lütkepohl, 2005; Rozanov, 1967). Square summability requires that $\sum_{k=1}^{p} \|A_k\|^2 < \infty$

[1] G-causal inference is not necessarily limited to stationary processes. Under some circumstances—notably for *multi-trial* data or for near-stationary data that may be "windowed"—the stationarity requirement may be relaxed. This is addressed later on.

[2] Throughout, bold symbols represent vector quantities and normal typeface represents scalar quantities. Capitalised symbols represent random variables or matrices, according to context. Thus we write $\boldsymbol{U}_t$ for the random vector representing the value of the process $\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots$ at time t, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ for a realisation of the process of length m, u_{it} for the ith component of $\boldsymbol{u}_t$, etc. A superscript $\top$ denotes matrix transpose, and a superscript $*$ denotes (complex) matrix conjugate-transpose. $|M|$ denotes the determinant of the square matrix M.

which means intuitively that the coefficients do not "blow up" even for infinite model order p (for finite p square summability is satisfied trivially). Stability means that the VAR coefficients define a *covariance-stationary* process. This is related to the the *characteristic polynomial* for the coefficients sequence A_k, which is:

$$\varphi_A(z) \equiv \left| I - \sum_{k=1}^{p} A_k z^k \right| \tag{2}$$

for the variable z defined in the complex plane $\mathbb{C}$. By standard theory (Lütkepohl, 2005) the coefficients A_k define a stable VAR iff the characteristic polynomial is invertible on the unit disc $|z| \leq 1$ in the complex plane. Defining the *spectral radius* of the VAR as

$$\rho(A) \equiv \max_{\varphi_A(z)=0} \{|z|^{-1}\} \tag{3}$$

an equivalent condition for stability—i.e. that (1) define a covariance-stationary process—is

$$\rho(A) < 1 \tag{4}$$

The *autocovariance sequence* Γ_k for a covariance-stationary (not necessarily VAR) stochastic process $\boldsymbol{u}_t$ is defined as the sequence of $n \times n$ matrices

$$\Gamma_k \equiv \mathrm{cov}(\boldsymbol{u}_t, \boldsymbol{u}_{t-k}) \quad k = \ldots, -2, -1, 0, 1, 2, \ldots \tag{5}$$

Note that by stationarity the Γ_k do not depend on the time t, and that $\Gamma_{-k} = \Gamma_k^\top$ for all lags k. For a VAR process of the form (1), it may be shown [Appendix A.1, (𝔸5)] that in the long run, the autocovariance sequence decays exponentially with lag k, the rate of decay being just the spectral radius $\rho(A)$. Note, importantly, that this implies that if the autocovariance sequence for a process does *not* decay (sub-)exponentially then the process may not be modelled as a VAR (see Section 3.3). This has some importance for analysis of neural data, which we return to in Section 4.2.

For a VAR process (1), the autocovariance sequence is related to the VAR parameters via the *Yule-Walker* equations (Anderson, 1971)

$$\Gamma_k = \sum_{\ell=1}^{p} A_\ell \Gamma_{k-\ell} + \delta_{k0} \Sigma \quad k = \ldots, -2, -1, 0, 1, 2, \ldots \tag{6}$$

Numerically, there are efficient algorithms available for solving (6) for the Γ_k in terms of (A_k, Σ) [Appendix A.1, (𝔸5)] and, conversely, for (A_k, Σ) in terms of the Γ_k [Appendix A.1, (𝔸6)].

The *cross-power spectral density* (CPSD) is defined as the (two-sided) Fourier transform of the autocovariance sequence[3]:

$$S(\lambda) = \sum_{k=-\infty}^{\infty} \Gamma_k e^{-ik\lambda} \quad 0 \leq \lambda \leq 2\pi \tag{7}$$

The $S(\lambda)$ are $n \times n$ Hermitian positive semi-definite matrices. The inverse Fourier transform recovers the autocovariance sequence:

$$\Gamma_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\lambda) e^{i\lambda k} \, \mathrm{d}\lambda \quad k = \ldots, -2, -1, 0, 1, 2, \ldots \tag{8}$$

Numerically, (7) and (8) may be efficiently computed by a (discrete) Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) respectively.

For a VAR process, the CPSD admits a unique *spectral factorisation* (Papoulis and Pillai, 2002)

$$S(\lambda) = H(\lambda) \Sigma H(\lambda)^* \quad 0 \leq \lambda \leq 2\pi \tag{9}$$

where the *transfer function* $H(\lambda)$ is defined as the inverse matrix of the Fourier transform of the regression coefficients:

$$H(\lambda) \equiv \left(I - \sum_{k=1}^{p} A_k e^{-ik\lambda} \right)^{-1} \quad 0 \leq \lambda \leq 2\pi \tag{10}$$

While there is no known closed-form solution of (9) for $H(\lambda)$, Σ in terms of $S(\lambda)$, a classical result (Masani, 1966; Wilson, 1972) states that, given a CPSD $S(\lambda)$ of a VAR process, a unique solution exists[4] and may be computed numerically [Appendix A.1, (𝔸7)]. According to (10) the VAR coefficients A_k may then be recovered from $H(\lambda)$ by a matrix inversion and inverse Fourier transform.

A further technical condition for valid G-causality analysis is that the CPSD be uniformly bounded away from zero almost everywhere (Geweke, 1982, 1984). Explicitly, there must exist a $c > 0$ such that for almost all $0 \leq \lambda \leq 2\pi$

$$c^{-1} I \leqslant S(\lambda) \leqslant cI \tag{11}$$

where for A, B square matrices, $A \leqslant B$ denotes that $B - A$ is positive semidefinite. Importantly, condition (11) guarantees square summability of the regression coefficients (Geweke, 1982, 1984; Rozanov, 1967).

Eqs. (6)–(10) relate the VAR parameters (A_k, Σ), the autocovariance sequence Γ_k and the CPSD $S(\lambda)$. Importantly, these three mathematical objects specify *entirely equivalent* representations for a VAR and any of them may be estimated from empirical time series data. As we shall see (Section 3), the MVGC toolbox makes extensive use of these equivalences to furnish accurate and efficient computational pathways for the calculation of G-causality in the time and frequency domains.

2.2. Unconditional G-causality in the time domain

In its simplest (unconditional, time-domain) form, G-causality is motivated as follows: suppose that $\boldsymbol{U}_t$ is split into two jointly distributed (i.e. inter-dependent) multivariate processes:

$$\boldsymbol{U}_t = \begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{pmatrix} \tag{12}$$

Under a predictive interpretation (*cf.* Section 2), the G-causality from $\boldsymbol{Y}$ to $\boldsymbol{X}$, written $\mathcal{F}_{\boldsymbol{Y} \to \boldsymbol{X}}$, stands to quantify the "degree to which the past of $\boldsymbol{Y}$ helps predict $\boldsymbol{X}$, over and above the degree to which $\boldsymbol{X}$ is already predicted by its own past". In the VAR formulation, this notion is operationalised as follows: the VAR(p) (1) decomposes as

$$\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{pmatrix} = \sum_{k=1}^{p} \begin{pmatrix} A_{xx,k} & A_{xy,k} \\ A_{yx,k} & A_{yy,k} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_{t-k} \\ \boldsymbol{Y}_{t-k} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_{x,t} \\ \boldsymbol{\varepsilon}_{y,t} \end{pmatrix} \tag{13}$$

and the residuals covariance matrix as

$$\Sigma \equiv \mathrm{cov} \begin{pmatrix} \boldsymbol{\varepsilon}_{x,t} \\ \boldsymbol{\varepsilon}_{y,t} \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{14}$$

[3] For simplicity, all spectral quantities are defined on the normalised frequency range $0 \leq \lambda \leq 2\pi$. In an empirical scenario—and indeed in the MVGC implementation—the natural frequency range of definition is $0 \leq \lambda \leq f$ where f is the sampling frequency in Hertz (Hz). In fact, to take advantage of inherent symmetries, the MVGC computes all spectral entities only up to the *Nyqvist frequency* $f/2$.

[4] More precisely, given $S(\lambda)$ there is a unique holomorphic matrix function $\tilde{H}(z)$ on the complex plane with $\tilde{H}(0) = I$, and a unique symmetric positive definite matrix Σ, such that (9) holds for $H(\lambda) \equiv \tilde{H}(e^{-i\lambda})$.

The x-component of the regression (13) is

$$\boldsymbol{X}_t = \sum_{k=1}^{p} A_{xx,k} \cdot \boldsymbol{X}_{t-k} + \sum_{k=1}^{p} A_{xy,k} \cdot \boldsymbol{Y}_{t-k} + \boldsymbol{\varepsilon}_{x,t} \tag{15}$$

from which we see that the dependence of $\boldsymbol{X}$ on the past of $\boldsymbol{Y}$, given its own past, is encapsulated in the coefficients $A_{xy,k}$; in particular, there is *no* conditional dependence of $\boldsymbol{X}$ on the past of $\boldsymbol{Y}$ iff $A_{xy,1} = A_{xy,2} = \cdots = A_{xy,p} = 0$. This leads to consideration of the *reduced* (or *restricted*) regression—as opposed to the *full* regression (15)—given by omitting the "historic" (past) $\boldsymbol{Y}$ dependency:

$$\boldsymbol{X}_t = \sum_{k=1}^{p} A'_{xx,k} \cdot \boldsymbol{X}_{t-k} + \boldsymbol{\varepsilon}'_{x,t} \tag{16}$$

so that now $\boldsymbol{X}$ is predicted by its own past only. Here $A'_{xx,k}$ are the reduced regression coefficients and $\boldsymbol{\varepsilon}'_{x,t}$ the reduced regression residuals, with covariance matrix $\Sigma'_{xx} \equiv \mathrm{cov}(\boldsymbol{\varepsilon}'_{x,t})$. $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}}$, then, stands to quantify the degree to which the full regression (15) represents a "better" model of the data than the restricted regression (16). *Maximum likelihood* (ML) theory (Edwards, 1992) furnishes a natural framework for the analysis of parametric data modelling. In this framework, an appropriate comparative measure for *nested* models of the form (15) and (16) is the *likelihood ratio* statistic or, for its more convenient statistical properties, its logarithm.[5] Technically, it is a test statistic for the null hypothesis of zero causality

$$H_0 : A_{xy,1} = A_{xy,2} = \cdots = A_{xy,p} = 0 \tag{17}$$

in (15). This motivates the *definition* of the G-causality statistic as the appropriate log-likelihood ratio. Now standard ML theory tells us that the likelihood for a realisation of length m of a VAR model of the form (1) is proportional to $|\Sigma|^{-(m-p)/2}$, where $|\Sigma|$, the *generalised variance* of the model (Barrett et al., 2010), is the determinant of the residuals covariance matrix. Thus the (unconditional, time-domain) G-causality from $\boldsymbol{Y}$ to $\boldsymbol{X}$ is defined to be the log-likelihood ratio

$$\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}} \equiv \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|} \tag{18}$$

where $\Sigma_{xx} = \mathrm{cov}(\boldsymbol{\varepsilon}_{x,t})$ and $\Sigma'_{xx} = \mathrm{cov}(\boldsymbol{\varepsilon}'_{x,t})$ are the residuals covariance matrices of the VAR models (15) and (16) respectively. A predictive interpretation of (18) is that the generalised variance of a regression model may be viewed as a measure of model *prediction error*.[6] From this perspective, G-causality (18) thus quantifies the reduction in prediction error when the past of the process $\boldsymbol{Y}$ is included in the explanatory variables of a VAR model for $\boldsymbol{X}$.

An important and frequently misunderstood point is that G-causality is often perceived to lack quantitative meaning beyond its interpretation as a hypothesis-test statistic appropriate only for significance testing (Section 2.5); that is, that its magnitude is essentially meaningless, and in particular should not be compared with other G-causality statistics. However, G-causality magnitudes have a natural interpretation in terms of information-theoretic bits-per-unit-time. This is because, for a large class of joint processes, G-causality and information-theoretic transfer entropy (Schreiber, 2000) are asymptotically equivalent (Barnett and Bossomaier, 2013) [in the Gaussian case the equivalence is exact (Barnett et al., 2009)]. Note that transfer entropy is conceptually isomorphic to G-causality and is often described as a measure of information flow (but see e.g. Lizier and Prokopenko, 2010). Thus, G-causalities may be meaningfully compared, and magnitudes cited, albeit with due caution to statistical significance (Section 2.5).

2.3. Conditional and pairwise-conditional G-causality

The unconditional G-causality statistic introduced above has the undesirable characteristic that if there are joint (possibly historical) dependencies between $\boldsymbol{X}$ and $\boldsymbol{Y}$ and a third set of variables, $\boldsymbol{Z}$ say, then spurious causalities may be reported. Thus, for instance, if there is no direct causal influence $\boldsymbol{Y} \to \boldsymbol{X}$ but there are (possibly lagged) dependencies of $\boldsymbol{X}$ and $\boldsymbol{Y}$ on $\boldsymbol{Z}$ then a spurious $\boldsymbol{Y} \to \boldsymbol{X}$ causality may be reported. These spurious causalities may be eliminated by "conditioning out" the common dependencies – provided they are available in the data. If, however, there are dependencies on unknown (exogenous) or unrecorded (latent) variables, then it will in general be impossible to eliminate entirely their potentially confounding effect on causal inference, although attempts have been made to mitigate their impact [e.g. "partial" G-causality (Guo et al., 2008)[7]].

To illustrate the conditional case, suppose that the universe $\boldsymbol{U}$ of known, recorded variables splits into three jointly distributed (i.e. inter-dependent) multivariate processes

$$\boldsymbol{U}_t = \begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \\ \boldsymbol{Z}_t \end{pmatrix} \tag{19}$$

and we wish to eliminate any joint effect of $\boldsymbol{Z}$ on the inference of the G-causality from $\boldsymbol{Y}$ to $\boldsymbol{X}$. Again the VAR(p) (1) splits analogously to (13) and we may consider the full and reduced regressions

$$\boldsymbol{X}_t = \sum_{k=1}^{p} A_{xx,k} \cdot \boldsymbol{X}_{t-k} + \sum_{k=1}^{p} A_{xy,k} \cdot \boldsymbol{Y}_{t-k} + \sum_{k=1}^{p} A_{xz,k} \cdot \boldsymbol{Z}_{t-k} + \boldsymbol{\varepsilon}_{x,t} \tag{20}$$

$$\boldsymbol{X}_t = \sum_{k=1}^{p} A'_{xx,k} \cdot \boldsymbol{X}_{t-k} + \sum_{k=1}^{p} A'_{xz,k} \cdot \boldsymbol{Z}_{t-k} + \boldsymbol{\varepsilon}'_{x,t} \tag{21}$$

analogous to (15) and (16), but with the conditioning variables $\boldsymbol{Z}_t$ included in *both* regressions. The null hypothesis to be tested is still (17) and the causality $\boldsymbol{Y} \to \boldsymbol{X}$ conditioned on $\boldsymbol{Z}$, which we write $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}$, is again as in (18):

$$\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}} \equiv \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|} \tag{22}$$

but now the inclusion of $\boldsymbol{Z}$ in both regressions accounts for its joint effect. $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}$ may thus be read as "the degree to which the past of $\boldsymbol{Y}$ helps predict $\boldsymbol{X}$, over and above the degree to which $\boldsymbol{X}$ is already predicted by its own past *and the past of* $\boldsymbol{Z}$".

Note that in this (and the previous) section, the source, target and conditioning variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ may themselves be multivariate; that is, they may represent *groups* of variables. It is in this

[5] Alternative test statistics which appear in the G-causality literature are the Wald or Lagrange multiplier estimators (Edwards, 1992). An advantage of the likelihood ratio is that is has pleasant invariance properties under a broad class of transformations (Barrett et al., 2010; Barnett and Seth, 2011) (Section 2.7), a natural spectral decomposition (Geweke, 1982) (Section 2.6) and an information-theoretic interpretation as a transfer entropy (Schreiber, 2000; Kaiser and Schreiber, 2002; Barnett et al., 2009; Barnett and Bossomaier, 2013).

[6] It may seem intuitive to use the mean square error trace(Σ) (the "total error") as a measure of prediction error. However, on the grounds of transformation invariance, frequency decomposition and information-theoretic interpretation, as well as for consistency with the ML formulation, the generalised variance is more appropriate. See Barrett et al. (2010) for a fuller discussion.

[7] Note that there are a number of errors in this paper: firstly, it is stated erroneously that partial G-causality may be negative. Secondly, the sampling distribution given is incorrect. See Roelstraete and Rosseel (2012) and also Barrett et al. (2010).

sense that we use the term "multivariate" G-causality. G-causality is thus able to account for *group* interactions. This is significant, since elements in a multivariate system may function cooperatively or competitively, or interact generally in a more complex fashion than traditional bivariate analysis can accommodate (Ladroue et al., 2009; Barrett et al., 2010).

A case of particular importance is that of *pairwise-conditional* G-causality. Given a universe $\boldsymbol{U}$ of variables comprising n (known, recorded) jointly distributed univariate processes $U_{1t}, \ldots, U_{nt}$, it is frequently of interest to estimate the G-causalities between pairs of variables U_i, U_j, $i \neq j$. The traditional bivariate pairwise G-causalities are the $\mathcal{F}_{U_j \to U_i}$. Since these are prone to spurious effects through joint dependencies as described above, however, it is generally preferable to consider rather the *pairwise-conditional* causalities

$$\mathcal{G}_{i,j}(\boldsymbol{U}) \equiv \mathcal{F}_{U_j \to U_i | \boldsymbol{U}_{[ij]}} \tag{23}$$

where the subscript $[ij]$ denotes omission of the ith and jth variables in the multivariate universe $\boldsymbol{U}$. Thus, when considering the causality $U_j \to U_i$ the joint dependencies of all remaining known variables are conditioned out.[8] The quantities $\mathcal{G}_{ij}(\boldsymbol{U})$, $i \neq j$ may be considered as a weighted directed graph, which we shall sometimes refer to as the *(G-)causal graph*.

2.4. Estimation from time series data

So far we have discussed time series data in terms of abstract stochastic processes; that is, we have assumed that our empirical time series data behaves like a realisation of some VAR process. We now turn to the crucial issues of estimation, under this assumption, of $\mathcal{F}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}$ from[9] a numerical time series $\boldsymbol{u} = \boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$. The first stage is to determine an appropriate *model order* for the regression (1). This may be achieved via standard techniques such as the Akaike or Bayesian information criteria, or cross-validation (McQuarrie and Tsai, 1998). The idea of model order selection is to balance the number of parameters (in the VAR case this is determined by the maximum lag p) so as to achieve the best model fit—perhaps in a ML or error-minimisation sense—while at the same time avoiding overfitting a finite data sequence.

The next stage is to obtain estimates of the model parameters which maximise the likelihood function for the respective VAR models (equivalently, minimise model error). As mentioned (Section 1), the MVGC toolbox obviates the need to estimate the reduced model parameters separately from the data. Here there is a choice of techniques yielding estimates asymptotically equivalent to the ML estimate, notably ordinary least squares (OLS) (Hamilton, 1994) and various multivariate extensions of Durbin recursion (frequently under the banner of "LWR algorithms") (Levinson, 1947; Whittle, 1963; Wiggins and Robinson, 1965; Morf et al., 1978) – see Appendix A.1, ($\mathbb{A}2$). Once we have estimates of all relevant parameters for both the full and reduced models, the G-causality sample estimator $\widehat{\mathcal{F}}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}(\boldsymbol{u})$ is obtained via (22) with the residuals covariance matrices Σ_{xx}, Σ'_{xx} replaced by their respective estimators $\hat{\Sigma}_{xx}(\boldsymbol{u})$, $\hat{\Sigma}'_{xx}(\boldsymbol{u})$. Another route to estimation of causalities in the time domain is via *spectral G-causality* (Section 2.6).

An important but rarely considered issue in G-causality model estimation is that, given that the full process $\boldsymbol{u}_t$ is a VAR (1), it is not clear that the *sub*process $\boldsymbol{X}_t$ will always be a well-defined VAR. In fact condition (11) guarantees that it is.[10] However, even if the full VAR (1) is of finite order, the reduced VAR (16) will generally be of infinite order[11] (similar remarks apply to the reduced regression (21)). Since under the ML formalism the full and reduced model orders should be the same, this implies that the appropriate (finite) empirical model order p should really be estimated for the reduced, rather than the full regression model. Failure to do so has potentially serious implications for G-causal inference (Section 2.5 – see also Section 3.1.2). The MVGC toolbox overcomes this issue in a natural way by only requiring estimation of the full regression (Section 3.1).

2.5. Statistical inference

The next task is to establish the statistical significance of the estimated causality against the null hypothesis (17) of zero causality, and, if desired, confidence intervals for its magnitude. As a likelihood ratio test, standard large-sample theory (Wilks, 1938; Wald, 1943) applies to the time-domain G-causality estimator in both the conditional and unconditional cases. If $\dim(\boldsymbol{X}) = n_x$, $\dim(\boldsymbol{Y}) = n_y$ and $\dim(\boldsymbol{Z}) = n_z$ (with $n_x + n_y + n_z = n$) then the difference in the number of parameters between the full model (20) and the nested reduced model (21) is just $d \equiv pn_xn_y$. Thus under the null hypothesis (17) of zero causality, $(m-p)\widehat{\mathcal{F}}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}(\boldsymbol{u})$, the G-causality estimator scaled by sample size has an asymptotic $\chi^2(d)$ distribution. Under the alternative hypothesis the scaled estimator has an asymptotic noncentral-$\chi^2(d\,;\nu)$ distribution, with noncentrality parameter $\nu = (m-p)\mathcal{F}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}$ equal to the scaled actual causality (which may, for the purpose of constructing confidence intervals, be replaced by its estimator).

In the case of a univariate causal target [i.e. $\dim(\boldsymbol{X}) = n_x = 1$] an alternative asymptotic sampling distribution is available for the R^2-like statistic $\exp(\mathcal{F}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}) - 1$: namely, under the null hypothesis, $\exp(\widehat{\mathcal{F}}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}(\boldsymbol{u})) - 1$ scaled by d_2/d_1 has an asymptotic $F(d_1, d_2)$ distribution where $d_1 = pn_y$ and $d_2 = m - p(n+1)$, and under the alternative hypothesis an asymptotic noncentral-$F(d_1, d_2\,;\nu)$ distribution where again ν is equal to the actual scaled value of the causality. For small samples in particular, the F-distribution may be preferable (it has a fatter tail than the corresponding χ^2 distribution). The MVGC toolbox makes both tests available and defaults to the F-test if the target variable is univariate.

Since G-causality is non-negative, $\widehat{\mathcal{F}}_{\boldsymbol{Y} \to \boldsymbol{X}|\boldsymbol{Z}}(\boldsymbol{u})$ will generally be positively biased, and under the alternative hypothesis the noncentral asymptotic distributions will be more accurate the closer the actual causality is to zero (Wald, 1943; Geweke, 1982). Some techniques for dealing with bias are discussed in Section 3.4.

For very small samples the theoretical asymptotic distributions may not be sufficiently accurate and nonparametric data-derived empirical sampling distributions may be preferable for statistical inference. For this purpose the MVGC toolbox supplies *permutation test* (Anderson and Robinson, 2001) and *non-parametric bootstrap* (Efron, 1982) routines for significance testing and computation of confidence intervals respectively; the toolbox also features routines suitable for simulation of surrogate time series data [Appendix A.1, ($\mathbb{A}3$)].

Finally, if multiple causalities are to be simultaneously considered, in view of the multiple null hypotheses we should take into account family-wise error rates or false discovery rates (Hochberg

[8] Confusingly, the term "multivariate" G-causality is sometimes applied in the literature to this case.

[9] This section and the following applies equally to the *un*conditional case (Section 2.2); we need simply take $\boldsymbol{Z}$ to be empty; i.e. of dimension 0.

[10] As noted by Geweke (Geweke, 1982), since the full CPSD $S(\lambda)$ is assumed to satisfy condition (11), so too does the CPSD $S_{xx}(\lambda)$ of the process $\boldsymbol{X}_t$ alone, from which it follows that the coefficients of the autoregression of $\boldsymbol{X}_t$ are square summable, and the reduced VAR is thus well-defined. See Geweke (1982, 1984) for full details.

[11] In fact if $\boldsymbol{U}_t$ is a finite-order VAR, then a subprocess $\boldsymbol{X}_t$ will in general be a finite-order VARMA process.

and Tamhane, 1987); again, this functionality is available in the MVGC toolbox.

2.6. G-causality in the frequency domain

A powerful feature of G-causality is that it may be decomposed in a natural way by frequency (Geweke, 1982, 1984). The resulting *spectral G-causality* integrates to the time-domain causality previously introduced, which may thus be considered an average over all frequencies of the spectral causality. The decomposition in the unconditional case is derived as follows: a split of $\boldsymbol{U}$ into sub-processes $\boldsymbol{X}$, $\boldsymbol{Y}$ as in (12) induces a decomposition

$$S(\lambda) = \begin{pmatrix} S_{xx}(\lambda) & S_{xy}(\lambda) \\ S_{yx}(\lambda) & S_{yy}(\lambda) \end{pmatrix} \tag{24}$$

of the cross-power spectral density (Section 2.1) and a similar decomposition for the transfer function $H(\lambda)$. Now $S_{xx}(\lambda)$ is just the CPSD of the sub-process $\boldsymbol{X}$, and from (9) we may derive

$$\begin{aligned} S_{xx}(\lambda) &= H_{xx}(\lambda)\Sigma_{xx}H_{xx}(\lambda)^* + 2\Re\mathfrak{e}\{H_{xx}(\lambda)\Sigma_{xy}H_{xy}(\lambda)^*\} \\ &\quad + H_{xy}(\lambda)\Sigma_{yy}H_{xy}(\lambda)^* \end{aligned} \tag{25}$$

Following Geweke (1982), we note that in the special case that $\Sigma_{xy} \equiv 0$, which may always be effected by a linear transformation of variables leaving $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}}$ invariant (Barrett et al., 2010), (25) takes the simpler form

$$S_{xx}(\lambda) = H_{xx}(\lambda)\Sigma_{xx}H_{xx}(\lambda)^* + H_{xy}(\lambda)\Sigma_{yy}H_{xy}(\lambda)^* \tag{26}$$

whereby the CPSD of $\boldsymbol{X}$ splits into an "intrinsic" term and a "causal" term. This motivates definition of the (unconditional) *spectral G-causality* from $\boldsymbol{Y}$ to $\boldsymbol{X}$ as

$$\boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda) \equiv \ln\left(\frac{|S_{xx}(\lambda)|}{|S_{xx}(\lambda) - H_{xy}(\lambda)\Sigma_{yy}H_{xy}(\lambda)^*|}\right) \tag{27}$$

or, in terms of untransformed variables (i.e. where $\Sigma_{xy} \neq 0$),

$$\boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda) \equiv \ln\left(\frac{|S_{xx}(\lambda)|}{|S_{xx}(\lambda) - H_{xy}(\lambda)\Sigma_{y|x}H_{xy}(\lambda)^*|}\right) \tag{28}$$

where the *partial covariance matrix* $\Sigma_{y|x}$ is defined by

$$\Sigma_{y|x} \equiv \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \tag{29}$$

Geweke then establishes the fundamental spectral decomposition of G-causality in the unconditional case[12]

$$\frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda)\mathrm{d}\lambda = \mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}} \tag{30}$$

so that time domain causality may be thought of as the average over all frequencies of spectral causality.

Sample estimation of $\boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda)$ admits at least two possible approaches. A straightforward procedure is to estimate parameters A_k, Σ for the VAR(p) (1) as before (after selecting a suitable model order). The transfer function may then by calculated by (fast) Fourier transform of the estimated regression coefficients according to (10), the CPSD from (9) and the partial covariance matrix from (29). An estimate for the unconditional spectral G-causality is then obtained by plugging the appropriate estimates into (28). Note that no reduced regression is required. An alternative procedure (Dhamala et al., 2008a,b) is to estimate the CPSD $S(\lambda)$ directly from the data – many standard techniques are available to this end. Then, as mentioned previously (Section 2.1), $H(\lambda)$ and Σ may be computed numerically and an estimate for the spectral G-causality is again obtained from (28). The MVGC toolbox facilitates both techniques; we recommend the former due to improved numerical accuracy and computational efficiency (Section 3.1 and Appendix A.1). As regards statistical inference, in contrast to the time-domain case there are (to our knowledge) no known (asymptotic) distributions for the sample distribution of $\hat{\boldsymbol{f}}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda)$ (see Geweke, 1984 for a fuller discussion on this issue) and nonparametric techniques are best deployed for significance testing and derivation of confidence intervals[13].

The conditional case is less straightforward. Suppose that $\boldsymbol{U}$ splits into sub-processes $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ as per (19). We then consider the reduced regression [*cf.* (21)]

$$\begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Z}_t \end{pmatrix} = \sum_{k=1}^{p} \begin{pmatrix} A'_{xx,k} & A'_{xz,k} \\ A'_{zx,k} & A_{zz,k} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_{t-k} \\ \boldsymbol{Z}_{t-k} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_t^\dagger \\ \boldsymbol{Z}_t^\dagger \end{pmatrix} \tag{31}$$

to define the residuals $\boldsymbol{X}^\dagger$, $\boldsymbol{Z}^\dagger$ as new variables. In Geweke (1984) the identity

$$\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}} \equiv \mathcal{F}_{\boldsymbol{Y}\oplus\boldsymbol{Z}^\dagger\to\boldsymbol{X}^\dagger} \tag{32}$$

is established, where $\boldsymbol{Y} \oplus \boldsymbol{Z}_t^\dagger \equiv \begin{pmatrix} \boldsymbol{Y}_t \\ \boldsymbol{Z}_t^\dagger \end{pmatrix}$, which expresses the conditional time-domain G-causality $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}$ as an *un*conditional causality in terms of the new variables $\boldsymbol{X}^\dagger$, $\boldsymbol{Y} \oplus \boldsymbol{Z}^\dagger$. Accordingly, the spectral causality in the unconditional case is defined as

$$\boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}(\lambda) \equiv \boldsymbol{f}_{\boldsymbol{Y}\oplus\boldsymbol{Z}^\dagger\to\boldsymbol{X}^\dagger}(\lambda) \tag{33}$$

and the spectral decomposition

$$\frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{f}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}(\lambda)\,\mathrm{d}\lambda = \mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}} \tag{34}$$

again obtains.

To estimate $\hat{\boldsymbol{f}}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}(\lambda)$ in sample, a separate reduced regression—which, as has been pointed out (Chen et al., 2006a), may lead to substantial inaccuracies[14]—may be avoided as follows: after estimation of VAR parameters for (1) (or spectral estimation), it may be shown [Appendix A.1, (𝔸11) and (𝔸12)] that the transformation of variables $\boldsymbol{X}$, $\boldsymbol{Z} \to \boldsymbol{X}^\dagger$, $\boldsymbol{Z}^\dagger$ defined by (31) induces a transformation of the autocovariance sequence and CPSD which may be effected computationally. Then the conditional spectral causality may be calculated as for the unconditional case via (33). Again, little is known of the sampling distribution of $\boldsymbol{f}_{\boldsymbol{Y}} \to \boldsymbol{X}|\boldsymbol{Z}(\lambda)$ so that nonparametric techniques must be used for statistical inference.

As in the time-domain case, pairwise-conditional spectral causalities may be calculated as

$$g_{ij}(\boldsymbol{U};\lambda) \equiv \boldsymbol{f}_{U_j\to U_i|\boldsymbol{U}_{[ij]}}(\lambda) \tag{35}$$

2.7. Filter invariance of G-causality

We have previously shown that multivariate G-causality, conditional and unconditional and in both time and frequency domains,

[12] In fact equality in (30) holds strictly when the condition $|A_{yy}(\lambda) - \Sigma_{yx}\Sigma_{xx}^{-1}A_{xy}(\lambda)| \neq 0$ is satisfied on $0 < \lambda \leq 2\pi$; otherwise it should be replaced by $\leq$. In practice, according to Geweke (1982), the equality condition is "almost always" satisfied.

[13] Our unpublished simulation studies suggest that, in fact, the sampling distribution for $\hat{\boldsymbol{f}}_{\boldsymbol{Y}\to\boldsymbol{X}}(\lambda)$ at any fixed frequency λ is actually the same as for the time-domain causality $\widehat{\mathcal{F}}_{\boldsymbol{Y}\to\boldsymbol{X}}$, although we do not yet have a theoretical justification for this claim. This appears also to hold in the conditional case.

[14] The "block decomposition" technique proposed in Chen et al. (2006a) appears, according to our unpublished simulations, to yield inaccurate results. In particular, the integral equality (34) generally fails even for simple test data in large samples. It is possible that the analytic derivation in Chen et al. (2006a) may be flawed.

is theoretically invariant under the application of (almost) arbitrary stable, invertible[15] digital filters (Antoniou, 1993; Barnett and Seth, 2011). Empirically, however, filtering of stationary time series data may severely compromise statistical inference of causalities estimated in sample. This is due primarily to an increase in empirical VAR model order following filtering. An important implication is that G-causality restricted to particular frequency bands cannot be measured by first filtering the data within the desired bands. While the computed values may change following filtering, this likely reflects not the desired "band-limited" G-causality but rather spurious values arising from inaccuracies in model fitting and statistical inference. The solution proposed in Barnett and Seth (2011) is to avoid filtering (of *stationary* data – see below) altogether. Then if spectral causalities are required, values outside the frequency range of interest may simply be ignored, while appropriate time-domain causalities may be obtained by averaging spectral causalities (28) and (33) over the frequency range $\mathcal{B}$ of prior interest [*cf.* (30) and (34)] to obtain *band-limited G-causality*:

$$\mathcal{F}_{\mathbf{Y}\to\mathbf{X}}(\mathcal{B}) \equiv \frac{1}{\mu(\mathcal{B})}\int_{\mathcal{B}} \boldsymbol{f}_{\mathbf{Y}\to\mathbf{X}}(\lambda)\,\mathrm{d}\lambda \tag{36}$$

$$\mathcal{F}_{\mathbf{Y}\to\mathbf{X}|\mathbf{Z}}(\mathcal{B}) \equiv \frac{1}{\mu(\mathcal{B})}\int_{\mathcal{B}} \boldsymbol{f}_{\mathbf{Y}\to\mathbf{X}|\mathbf{Z}}(\lambda)\,\mathrm{d}\lambda \tag{37}$$

where $\mu(\mathcal{B}) \equiv \int_{\mathcal{B}} \mathrm{d}\lambda$ is the measure (length) of $\mathcal{B}$. Once again, nonparametric methods must be used for statistical inference of band-limited G-causality.

In practice, some filtering may still be appropriate in cases where the data are not stationary to begin with, in which case G-causal analysis is likely anyway to fail or deliver spurious results (Section 3.3). In these cases filtering may be a valid and useful tool to improve stationarity; e.g. notch-filtering of line-noise in electronically recorded time series data (Barnett and Seth, 2011), or finite differencing to eliminate drift (Seth, 2010). In the latter case, filters of the form $\tilde{u}_t = u_t - u_{t-1}$ may be applied iteratively until (approximate) stationarity is achieved; but note that differencing constitutes a *non-invertible* filter, so that stationary data should *not* be differenced. An important application of filter invariance in neuroscience (Section 4) is to G-causal analysis of fMRI (functional Magnetic Resonance Imaging) data, where serious confounds have been previously suspected; this is discussed further in Section 4.3.

3. MVGC Toolbox design

Central to the design of the MVGC toolbox is the equivalence (Section 2.1) of the VAR parameters (A_k, Σ), the autocovariance sequence Γ_k and the cross-power spectral density $S(\lambda)$ as representations for a VAR process. The MVGC toolbox exploits these equivalences to provide numerically accurate algorithms for moving flexibly between the alternative VAR representations, thus furnishing computationally efficient pathways for calculation of G-causalities, conditional and unconditional, in the time and frequency domains. A schematic of computational pathways implemented in the toolbox is given in Fig. 1.

3.1. Computational strategy

The basic operating principle of estimation of G-causalities via the MVGC toolbox is that (after determination of a suitable model order) a VAR (1) is fitted to the time series data just once ($\mathbb{A}2$) and all subsequent calculations are based on the estimated model parameters $\hat{A}_k$, $\hat{\Sigma}$. This approach is theoretically consistent with the fundamental assumption of G-causal estimation, that the time series data is a realisation of a stationary VAR (1). In principle, any of the equivalent VAR representations might be chosen for initial estimation—e.g. via sample autocovariance ($\mathbb{A}1 \to \mathbb{A}6$) or spectral estimation ($\mathbb{A}4 \to \mathbb{A}7$)—but empirically direct estimation ($\mathbb{A}2$) proves to be the most stable, accurate and computationally efficient; nonetheless the MVGC toolbox allows implementation of alternative methods.

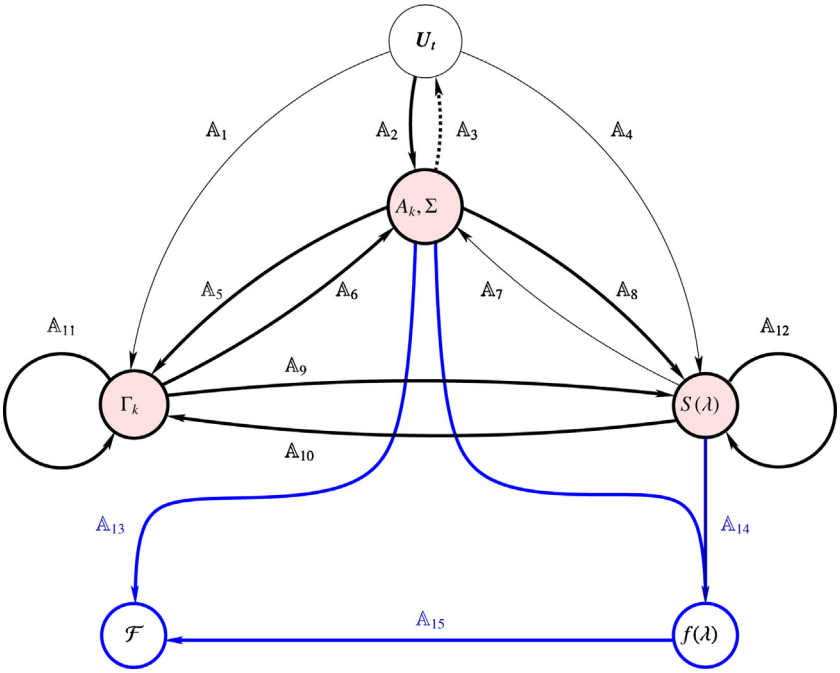
Having acquired the VAR model, (conditional) G-causality in the time (22) and frequency (33) domains involves estimation of VAR parameters for the reduced regressions (21) and (31) respectively. There is no simple algorithm (to our knowledge) to compute reduced regression VAR parameters directly from full regression parameters. Indeed, the standard approach [e.g. as implemented in Seth (2010)] is to perform the reduced regressions explicitly from the data ($\mathbb{A}2$); however this leads to inaccuracies that may be particularly serious in the spectral case (see Section 2.6 and Chen et al., 2006a). Fortunately, numerical algorithms are available for obtaining reduced regression parameters from the autocovariance sequence ($\mathbb{A}7$) via solution of the Yule-Walker equations (6), or from the CPSD ($\mathbb{A}7$) via spectral factorisation (9); the autocovariance sequence and CPSD may themselves be computed from the (full) VAR parameters by "reverse" Yule-Walker solution ($\mathbb{A}5$) and straightforward spectral calculation (Fourier transform) ($\mathbb{A}8$) respectively. The recommended pathway for the MVGC Toolbox is to use the autocovariance sequence. This is on the grounds that, as revealed by empirical testing over a broad range of VAR parameters, system size and model orders, Whittle's algorithm ($\mathbb{A}8$) (Whittle, 1963) for Yule-Walker solution is faster, more stable and more accurate than Wilson's spectral factorisation algorithm ($\mathbb{A}7$) (Wilson, 1972); it also has additional useful properties (see below). The toolbox thus uses the autocovariance sequence as the dominant representation for all G-causality routines ($\mathbb{A}13$, $\mathbb{A}14$) although, again, the alternative spectral pathways may be implemented if desired. Note that a useful "reality check" for the accuracy of G-causality computations is that time-domain causalities should integrate with reasonable accuracy to their spectral counterparts (34) ($\mathbb{A}15$).

3.1.1. Reduced model order, spectral resolution and autocovariance lags

As noted previously (Section 2.4), even if the model order for the full regression (20) is finite (and in practice a finite full model order will, of course, need to be selected), the model order for the reduced regressions (21) and (31) will be in theory infinite. If, as described above, reduced VAR parameters are calculated from the autocovariance sequence (or CPSD) how do we choose a suitable *finite* model order?

First, we note that spectral factorisation (9) specifies the transfer function $H(\lambda)$ at frequencies λ for which the CPSD $S(\lambda)$ is specified. Now the VAR coefficients A_k are calculated by matrix inversion of the transfer function followed by Fourier transform. Since the Fast Fourier Transform (FFT) (Cooley and Tukey, 1965) will be used, this implies that to calculate VAR coefficients from the CPSD, the CPSD must be specified at some set of frequencies $\lambda_1, \ldots \lambda_q$ evenly spaced on $[0, f]$, where f is the sampling frequency—indeed, Wilson's algorithm ($\mathbb{A}7$) assumes this, and approximates the $H(\lambda_k)$ accordingly. Then the A_k will naturally be derived for $k = 1, \ldots, q$; i.e. the *spectral resolution* q will, effectively, be the model order for the reduced regression. Unfortunately, for the spectral factorisation approach there is no obvious way to choose q. Furthermore (in contrast to Whittle's algorithm – see below), given a CPSD $S(\lambda)$ for a stable VAR process evaluated, for a given spectral resolution q at evenly

[15] In Barnett and Seth (2011) it is erroneously stated that to guarantee invariance a filter need only be *causal*—i.e. not contain a pure lag. In fact full invertibility (minimum phase filtering) is required to ensure that the condition (11) is not violated for the filtered process. We are grateful to Victor Solo (private communication) for pointing this out.



$\mathbb{A}1$	Sample autocovariance estimation
$\mathbb{A}2$	VAR parameter estimation
$\mathbb{A}3$	VAR simulation (for testing)
$\mathbb{A}4$	Sample spectral estimation
$\mathbb{A}5$	Yule-Walker reverse solution
$\mathbb{A}6$	Yule-Walker solution
$\mathbb{A}7$	VAR spectral factorisation
$\mathbb{A}8$	VAR spectral calculation
$\mathbb{A}9$	Fourier transform of autocovariance sequence
$\mathbb{A}10$	Inverse Fourier transform of CPSD
$\mathbb{A}11$	Autocovariance transform for reduced regression
$\mathbb{A}12$	Spectral transform for reduced regression
$\mathbb{A}13$	Time-domain G-causality calculation
$\mathbb{A}14$	Spectral (unconditional) G-causality calculation
$\mathbb{A}15$	Integration of spectral G-causality

U_t	time series
A_k	VAR coefficients
Σ	Residuals covariance matrix
Γ_k	Autocovariance sequence
$S(\lambda)$	CPSD
$\mathcal{F}$	G-causality (time domain)
$f(\lambda)$	G-causality (frequency domain)

Fig. 1. Schematic of computational pathways for the MVGC Toolbox. The "inner triangle" (shaded circles) represents the equivalence between the VAR representations outlined in Section 2.1. Bold arrows represent recommended (useful and computationally efficient) pathways, while blue arrows represent actual MVGC calculation.

spaced λ_k, there is no guarantee that VAR parameters derived by Wilson's algorithm will actually be stable. These problems do not afflict the autocovariance approach, described next.

For the autocovariance route, there is a natural answer for the choice of q, based on specific properties of VAR processes and Whittle's Yule-Walker solution algorithm ($\mathbb{A}6$). Given an autocovariance sequence Γ_k for $k = 0, 1, \ldots q$ lags, Whittle's algorithm calculates A_k for $k = 1, \ldots q$. As noted in Section 2.1, for a stationary VAR process the autocovariance sequence Γ_k decays exponentially; thus for large enough k, Γ_k will become numerically insignificant. This implies that we only need calculate the Γ_k for $k \leq$ some q, where the truncation decision is based on a specified numerical tolerance (which might be determined, e.g. by machine precision). Noting that the exponential decay factor is just the spectral radius ρ—easily calculated from the (assumed known) full VAR parameters—the MVGC Toolbox reverse Yule-Walker ($\mathbb{A}5$) implementation core/var_to_autocov calculates q so that $\rho^q <$ a given tolerance, which defaults to 10^{-8} (empirically this yields good accuracy).

Since the CPSD is calculated by FFT from the autocovariance sequence ($\mathbb{A}9$), q also becomes the spectral resolution for any subsequent spectral calculations. Thus a principled value for model order for the reduced regression is obtained. A further pleasant feature of Whittle's algorithm is that given the finite autocovariance sub-sequence Γ_k of a stable VAR model up to any given number of lags q, the derived VAR parameters are—in contrast to Wilson's algorithm—guaranteed to be stable. We emphasise that the *full* model order p as derived by standard model order selection criteria, and the *reduced* model order q as derived from autocovariance decay, are effectively independent; as implemented by the MVGC Toolbox, the latter depends only on the spectral radius ρ of the fitted VAR model.

3.1.2. Computational efficiency and accuracy

The MVGC Matlab implementation (Appendix A) has been designed with great regard to efficiency. All algorithms are wherever possible completely vectorised and designed to exploit Matlab's computational strengths, including the use, where appropriate, of key methods which invoke machine-optimised, multi-threaded libraries, such as linear solvers (LAPACK) and Fast Fourier Transform (FFTW).[16] Detailed benchmarking is beyond the scope of this article, but in testing we have not in general found

[16] It is difficult, in lieu of published information, to deduce the computational complexity and scaling of Matlab's core algorithms, particularly regarding the impact of potential parallelisation.

performance to be a major issue, even with large, highly multivariate datasets.

The most performance-critical algorithm in the MVGC workflow under normal usage scenarios (Section 3.2) is likely to be the Yule-Walker solution algorithm (𝔸6, core/autocov_to_var). The performance of this algorithm depends critically on the number of autocovariance lags deemed necessary for given numerical accuracy – which, in turn, depends entirely on the spectral radius of the estimated VAR (see previous Section). Thus performance in practice is closely tied to the actual dataset being analysed. An implication is that, if the spectral radius ρ is close to 1—i.e. the VAR estimate is "near unstable"—then the single-regression approach of the MVGC may well be more computationaly intensive than the traditional dual-regression method ("*GCCA mode*", in the toolbox parlance – see Section 3.3) and it may be necessary to limit the number of autocovariance lags.[17] This may be achieved either explicitly (via the `acmaxlags` parameter) or by adjusting the decay tolerance (via the `acdectol` parameter) in the routine 𝔸6, core/autocov_to_var, effectively trading-off performance for numerical accuracy (but note that statistical inference may be compromised – see below). In another usage scenario—smaller spectral radius coupled with very long time-series lengths—the MVGC approach may actually be more efficient, since here VAR estimation (𝔸2, core/tsdata_to_var) may become a computational bottleneck, and a single regression will thus be preferable.

Crucially however, in arguably the most important usage case—multivariate conditional G-causality in the frequency domain (Section 2.6)—performance is not the issue at stake. This calculation is especially relevant in neuroscience applications (Section 4), where results are required (a) to exclude common cross-variable influences by conditioning and (b) to be restricted to specific frequency bands. The traditional approach here has been to pre-filter the data, ostensibly to restrict causal estimates to frequencies of prior interest; however, as already described (Section 2.7), this approach is fatally flawed, insofar as it not only fails to restrict G-causality as desired, but in addition seriously compromises statistical inference. Thus it becomes essential to compute conditional spectral causalities where, as has been established (Chen et al., 2006a), the traditional dual-regression approach is known to produce spurious results, including negative causal estimates. The MVGC single-regression method handles this essential case correctly (and here too, the performance/accuracy trade-off described above is available).

Regarding accuracy, as previously mentioned (Section 2.4) the traditional dual-regression method fails to take account of the theoretically infinite model order of the reduced regression, with adverse effects on statistical inference. We illustrate this with a comparative analysis of the MVGC single-regression method vs. the traditional dual-regression method for the minimal VAR(1)

$$\begin{aligned} X_t &= aX_{t-1} + cY_{t-1} + \varepsilon_{x,t} \\ Y_t &= \quad\quad bY_{t-1} + \varepsilon_{y,t} \end{aligned} \tag{38}$$

Here the residuals $\varepsilon_{x,t}$, $\varepsilon_{y,t}$ are normally distributed, uncorrelated and unit-variance white noise, a, b represent decay parameters (stability requires that $|a| < 1$, $|b| < 1$) and c controls the strength of G-causality from $Y \to X$. G-causalities for (38) may be fully solved analytically [Appendix B; see also Barnett and Seth (2011)]. Evidently $\mathcal{F}_{X\to Y} \equiv 0$ and in Appendix B we calculate expressions for $\mathcal{F}^{(\infty)}_{Y\to X}$, G-causality as approximated by the MVGC single-regression method where infinite lags are assumed for the reduced regression, and for $\mathcal{F}^{(1)}_{Y\to X}$, G-causality as approximated by the traditional dual-regression method, with model order 1 (i.e. 1 autoregression lag) for both full and reduced regressions.

We simulated the process (38) 10, 000 times with time series lengths of 100 time steps, for $a = 0.8$, $b = 0.9$ and varying causal coefficient c, calculating sample distributions for G-causality estimators for both the statistically significant causality $Y \to X$ and the statistically non-significant ("null") causality $X \to Y$. Results are displayed in Fig. 2. We see [Fig. 2(a)] that, due to finite-sample effects, $\widehat{\mathcal{F}}^{(\infty)}_{Y\to X}$ slightly underestimates the true causality $\mathcal{F}^{(\infty)}_{Y\to X}$ on average, while $\widehat{\mathcal{F}}^{(1)}_{Y\to X}$ slightly underestimates the 1-lag causality $\mathcal{F}^{(1)}_{Y\to X}$ – which itself substantially *over*estimates the true causality. The latter may be explained by the fact that the model order 1 reduced regression fails to take into account the full explanatory power of the history of the process X_t on itself. We also see [Fig. 2(b)] that the dispersion (as measured by standard deviation) of the sample distribution is significantly greater for $\widehat{\mathcal{F}}^{(1)}_{Y\to X}$; this may be explained by the additional sampling error incurred by the extra regression involved. In the non-causal $X \to Y$ direction, we see that both the mean [Fig. 2(c)] and standard deviation [Fig. 2(d)] are significantly greater for $\widehat{\mathcal{F}}^{(1)}_{X\to Y}$ than for $\widehat{\mathcal{F}}^{(\infty)}_{X\to Y}$; while the latter decays with increasing causal strength c, the former remains approximately constant.

Next we investigated the robustness of G-causal inference to *measurement noise* under the two estimation techniques: we considered the process

$$\begin{aligned} \tilde{X}_t &= X_t + \nu\eta_{x,t} \\ \tilde{Y}_t &= Y_t + \nu\eta_{y,t} \end{aligned} \tag{39}$$

where $\eta_{x,t}$, $\eta_{y,t}$ are normally distributed unit-variance white noise terms uncorrelated with the X_t, Y_t, representing additive measurement noise of intensity ν. The addition of noise is known to degrade the ability to detect G-causalities (Solo, 2007).[18] Again we simulated the process (39) 10, 000 times with time series lengths of 100 time steps, for $a = 0.8$, $b = 0.9$, $c = 1$ and varying noise intensity ν. Results are displayed in Fig. 3. Again we see that mean estimated causalities as well as dispersion are higher for the dual-regression estimates than the single-regression estimates, most severely in the null (non-causal) $X \to Y$ direction. Here we also calculated Type II [false negative, Fig. 3(c)] and Type I [false positive, Fig. 3(f)] error rates, using the asymptotic null F-distribution (Section 2.5), at a significance level of $\alpha = 0.05$. We see that while the impact of dual regression on Type II errors is negligible, the impact on Type I errors is quite severe in comparison with the single-regression MVGC method.

3.2. MVGC work-flow

A typical work-flow for calculation of (conditional) G-causalities from empirical time series data, as exemplified in the demo/mvgc_demo.m toolbox demonstration script, is illustrated as follows:

1. **Model order estimation**: For each model order p up to a chosen maximum, fit a VAR model to the full "universe of data" time series $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ (𝔸2) and calculate the likelihood $L \propto |\Sigma^{-(m-p)/2}|$.

[17] This is mitigated by the observation that model order estimates based on autocovariance decay for Whittle's algorithm may be pessimistically large; in general, the A_k tend to become negligibly small (and the residuals covariance matrix Σ converges) for $k \ll$ number of autocovariance lags.

[18] Solo's "weak linear GC" corresponds to our G-causality, as opposed to his "strong linear GC", which corresponds to what has been independently termed "partial Granger causality" (see Section 2.3).

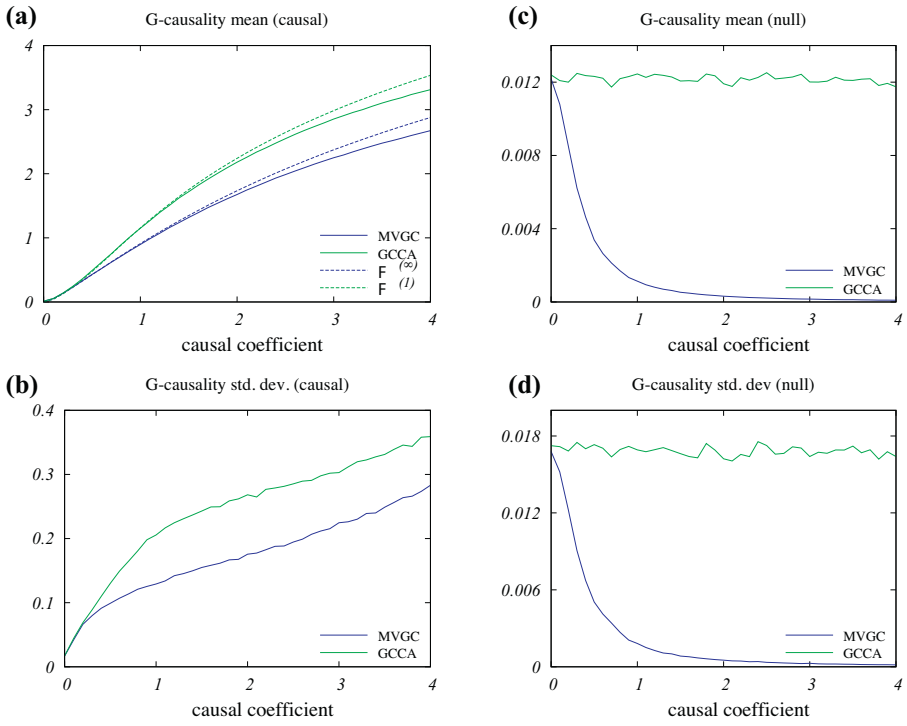


Fig. 2. MVGC (single-regression) vs. GCCA (dual-regression) accuracy for the minimal causal VAR(1) (38), with $a = 0.8$, $b = 0.9$ and varying causal coefficient c (x-axis). Sample statistics estimated over 10,000 runs of 100 time steps each. (a) sample mean significant G-causality $\widehat{\mathcal{F}}_{Y\to X}$, as estimated by MVGC and GCCA methods. $\mathcal{F}^{(\infty)}$ plots the theoretical (infinite lags) G-causality as estimated by the MVGC method, while $\mathcal{F}^{(1)}$ plots the theoretical (1-lag) causality as estimated by the GCCA method; (b) standard deviation of significant G-causality; (c) sample mean null G-causality $\widehat{\mathcal{F}}_{X\to Y}$; (d) standard deviation of null G-causality. See text for details.

Use this to calculate the chosen model order criterion (AIC or BIC). Select the best model order p according to the criterion.

2. **VAR model estimation**: Estimate the corresponding VAR model parameters (A_k, Σ) for the selected model order (𝔸2) and check that the spectral radius (3) is <1 (other statistical tests on the VAR parameters and residuals may be performed at this stage).
3. **Autocovariance calculation**: Calculate the autocovariance sequence Γ_k from the VAR parameters (𝔸5), to a suitable number of lags (as described above). This involves "reverse solution" of the Yule-Walker equations (6).
4. **Time domain**: For each conditional causality $\mathcal{F}_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}$ required [possibly for the entire causal graph (23)]:
   (a) Calculate the VAR parameters for both the full and reduced regressions (20) and (21) from the autocovariance sequence Γ_k (𝔸6). This involves solution of the Yule-Walker equations (6).
   (b) Calculate the time-domain conditional G-causality according to (22) (𝔸13).
   (c) Test the resulting causalities for significance at a given level using the analytical sampling distribution (Section 2.5), taking care to adjust for multiple hypotheses, and construct confidence intervals if desired.
5. **Frequency domain**: For each conditional spectral causality $\boldsymbol{f}_{\boldsymbol{Y}}\to\boldsymbol{X}|\boldsymbol{Z}(\lambda)$ required [possibly for all pairwise-conditional spectral causalities]:
   (a) Transform the autocovariance sequence for $\boldsymbol{X}$, $\boldsymbol{Z}$ to the autocovariance sequence for $\boldsymbol{X}_t^\dagger, \boldsymbol{Z}_t^\dagger$ (𝔸11); then, as per (33), calculate the unconditional spectral causality $\boldsymbol{f}_{\boldsymbol{Y}\oplus\boldsymbol{Z}^\dagger\to\boldsymbol{X}^\dagger}(\lambda)$.
   (b) Calculate the VAR parameters for the full regression (15) from the transformed autocovariance sequence (𝔸6).
   (c) Calculate the transfer function $H(\lambda)$ by Fourier transformation of the regression coefficients (10), and then the CPSD $S(\lambda)$ according to (9) (𝔸8).
   (d) Calculate the partial residuals covariance (29) and then the conditional spectral causality according to (28) (𝔸14).
   (e) Test the spectral causalities for significance at a given level by permutation test (Section 2.5), and construct confidence intervals, if desired, by non-parametric bootstrap.
   - If both time and frequency domain causalities are required, we can alternatively calculate the time-domain causalities by integrating their spectral counterparts (34) (𝔸15).

3.3. Potential problems and some solutions

Users of the GCCA (Granger Causal Connectivity Analysis) Toolbox (Seth, 2010), which the MVGC Toolbox supersedes, may occasionally find that time series data which appeared to yield satisfactory results using the GCCA software trigger errors or warnings when analysed via the MVGC Toolbox. This will typically be because the MVGC Toolbox is more stringent in its requirements and performs more thorough error-checking. There may be situations (see below) where the more robust GCCA approach can be useful. The MVGC toolbox may optionally be deployed in "GCCA mode" (see demo/mvgc_demo_GCCA), which takes the more traditional approach of separate full and reduced regressions (calculation of conditional spectral causalities will not be available). Generally, however, data which yield apparently reasonable results in GCCA mode but fail in standard MVGC mode should be treated with caution.

Likely reasons for reported problems with time series data are (i) colinearity, (ii) non-stationarity, (iii) long-term memory, (iv) strong moving average component and (v) heteroscedasticity. We briefly discuss each in turn.

(i) Colinearity occurs when there are linear relationships between variables (i.e. between individual time series) in multivariate time series data. In this case there is an ambiguity in the VAR

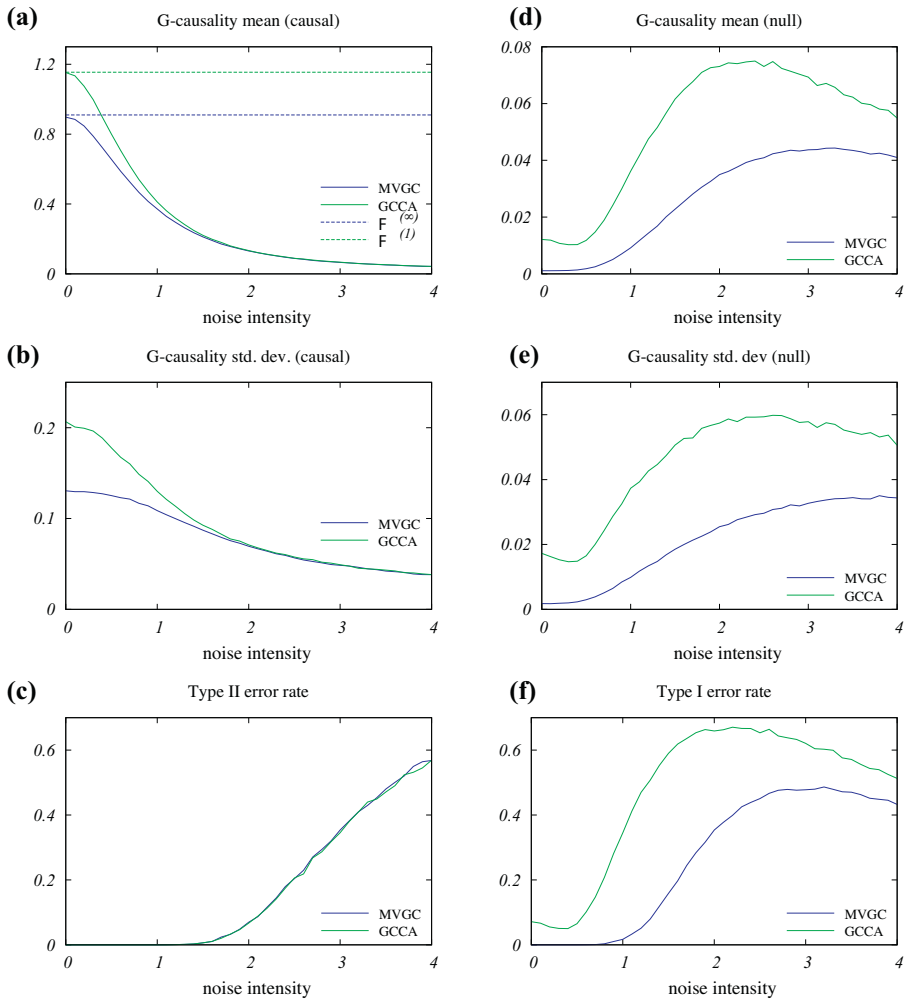


Fig. 3. MVGC (single-regression) vs. GCCA (dual-regression) accuracy for the minimal causal VAR(1) with additive noise (39), with $a = 0.8$, $b = 0.9$, $c = 1$ and varying noise intensity ν (x-axis). Sample statistics estimated over 10,000 runs of 100 time steps each. (a) Sample mean significant G-causality $\widehat{\mathcal{F}}_{Y\to X}$, as estimated by MVGC and GCCA methods. $\mathcal{F}^{(\infty)}$ plots the theoretical (infinite lags) G-causality as estimated by the MVGC method, while $\mathcal{F}^{(1)}$ plots the theoretical (1-lag) causality as estimated by the GCCA method; (b) standard deviation of significant G-causality; Type II error rate (proportion of false negatives); (d) sample mean null G-causality $\widehat{\mathcal{F}}_{X\to Y}$; (e) standard deviation of null G-causality; (f) Type I error rate (proportion of false positives). Error rates are calculated using the asymptotic null F-distribution (Section 2.5) at a significance level of $\alpha = 0.05$. See text for details.

representation of the data. Colinearity (or near-colinearity) will most likely be detected in the VAR estimation stage ($\mathbb{A}2$) and reported as "rank-deficient" or "ill-conditioned" regressions by the core/tsdata_to_var routine; this should be tested for by the caller (see demo/mvgc_demo). One solution is to eliminate linear dependencies, possibly via a Principal Component Analysis (PCA) (Jolliffe, 2002) or factor model approach (Flamm et al., 2013), or a signal separation technique such as Independent Component Analysis (ICA) (Hyvärinen et al., 2001).

(ii) The principal stationarity check performed by the MVGC Toolbox (explicitly by the routine utils/var_specrad or, more usually, as part of the standard workflow [Section 3.2, step 3] by core/var_to_autocov) is that the spectral radius of the estimated full VAR model (1) is less than one (4). If this condition is not satisfied then analysis cannot proceed. Note that the routine core/var_to_autocov performs exhaustive error checking and also produces useful diagnostics: the utility utils/var_info reports all errors, warnings and diagnostics generated by core/var_to_autocov (see e.g. demo/mvgc_demo).

Non-stationary data may be dealt with in several ways. Stationarity can sometimes be achieved by standard techniques such as de-trending. As mentioned earlier (Section 2.7), pre-filtering (e.g. notch filtering of electrical line noise or finite differencing) may also be useful, albeit at the potential cost of undermining model fitting and statistical inference. For *unit root* processes (processes exhibiting random walk-like non-stationarities) (Hamilton, 1994), a more principled route to G-causal analysis is via *co-integration* (Granger, 1981; Engle and Granger, 1987). The MVGC Toolbox is not currently able to deal directly with co-integrated processes.

An alternative approach is to divide the data into (overlapping or non-overlapping) windows, on the logic that shorter windows are more likely to be approximately stationary. This implies a tradeoff between likelihood of stationarity (shorter is better) and accuracy of model fit (longer is better). An advantage of windowing—assuming there is sufficient data—is that *time-varying* G-causality can be analysed (Ding et al., 2000). This is particularly useful given a large number of temporally aligned trials: then so-called "vertical regression" can be implemented using extremely short windows. This method depends on an assumption (which requires justification) that each trial is an independent realisation of the same underlying stochastic generative process; see the demo/mvgc_demo_nonstationary script for an example.

(iii) The spectral radius condition may also fail if the time series, although stationary, has *long-term memory*—that is the autocorrelation does not decay exponentially. This may arise in particular for iEEG/LFP data (Section 4.2). In this case, the data is fundamentally unsuited to VAR modelling (*cf.* Section 2.1), which may silently yield spurious results. The MVGC Toolbox is not currently able to deal with such data. Models for long-term memory processes do exist, e.g. VARFIMA (Vector Autoregressive Fractionally Integrated Moving Average) models (Granger and Joyeux, 1980), but G-causal inference for such models is difficult and the theory underdeveloped. Pre-analysis to determine the presence of long-term memory may be performed via checks on sample autocovariance (core/var_to_autocov) and/or CPSD (core/var_to_cpsd), where long-term memory typically manifests itself as power-law behaviour.

(iv) A problem may arise when the data, although stationary, contain a strong (and in particular a "slow") *moving average* component, violating the condition (11). In this case, the routine core/var_to_autocov may report warnings or errors (specifically, it may fail to solve the associated 1-lag problem – see Appendix A.1, $\mathbb{A}5$) and it may become necessary to use the more robust GCCA mode described above. This may in particular be the case for fMRI BOLD data (Section 4.3). A recent study (Seth et al., 2013) (see also Section 4.3) indicates that empirically, even if square summability [or the condition (11)] is violated, G-causal analysis may still, under some circumstances, yield meaningful causal magnitudes and directionality. Clearly, however, if the coefficients of a VAR are not square summable, then in an empirical setting they cannot be well-approximated at finite model order, so that the estimated VAR will inevitably be misspecified. This may result in diminished accuracy of causal estimates and compromised statistical inference (Section 2.5).

In rare cases it is possible that, as regards the reverse Yule-Walker calculation ($\mathbb{A}5$) (Whittle's algorithm), the reduced model order q, as determined by autocovariance decay, may turn out to be too small for the VAR coefficients A_k to decay sufficiently, preventing the residuals covariance matrix Σ from converging. This is unlikely to occur given a valid VAR model and a reasonably small choice of autocovariance decay tolerance.

(v) Finally, under standard VAR model scenarios, it is usually assumed that the *variance* of the residual terms does not depend on the actual values of the process. If this condition is violated, the process is said to be *heteroscedastic*. Note that heteroscedasticity does not in itself violate stationarity. However, while a stationary heteroscedastic process might well be modellable in theory as an (infinite order) VAR, such a model will not be parsimonious and statistical inference is likely to suffer; indeed, it is well-known that heteroscedasticity can invalidate standard statistical significance tests, and may confound G-causal inference (Luo et al., 2011). While there is extensive research (mainly in the econometrics literature) into GARCH (Generalised AutoRegressive Conditional Heteroscedastic) models (Silvennoinen and Teräsvirta, 2009), which autoregress residuals variances on their own history and/or the history of the process itself, G-causal analysis of such models is somewhat fragmented. In the neurosciences (Section 4) the literature on detection and functional analysis of heteroscedasticity (Ozaki, 2012) is limited.

Future iterations of the MVGC will address some or all of the above challenges. These iterations will take advantage of our recent work (Barnett and Bossomaier, 2013) establishing a very general equivalence of G-causality and transfer entropy for a broad class of predictive models (e.g. VARMA, VARFIMA, GARCH, co-integration) within a maximum likelihood framework, enabling the derivation of analytic (asymptotic) sampling distributions for statistical inference.

3.4. Debiasing of G-causality magnitudes

G-causality is by definition >0 and hence any empirical estimation is subject to bias. This is not important when assessing statistical significance but may be important if the objective is to accurately quantify the magnitude of a G-causality interaction in terms of bits (leveraging the asymptotic equivalence with transfer entropy (Barnett et al., 2009; Barnett and Bossomaier, 2013)). To ensure accurate magnitude estimation, *debiasing* is recommended. As illustrated in Barrett et al. (2012), a useful approach to debiasing is to generate surrogate distributions by dividing the data into windows and (many times) randomly rearranging these windows independently for each variable. The mean G-causality across this surrogate data set should give a good estimate of the bias, which can then be subtracted from the sample estimate obtained from the non-shuffled data.

4. Application to neuroscience time series data

Although G-causality is an entirely general method for analysis of time series data, the MVGC toolbox has been developed with application to neuroscience data in mind. Methodological development in this application domain is rapidly advancing and a full review is beyond the present scope (see, for example, Deshpande and Hu, 2012; Friston et al., 2013; Bressler and Seth, 2011; Ding et al., 2006). This section summarises some of the main issues involved in application of G-causality (as implemented by the MVGC toolbox) to some of the more common varieties of neuroscience time-series data.

4.1. Application to surface EEG and MEG data

M/EEG data is well suited for analysis by G-causality in virtue of the high time resolution, fast sampling, and its typically stochastic nature. For analysis of *steady-state* M/EEG the primary consideration is to ensure that the time-series are covariance stationary. As mentioned (Section 3.3), non-stationary data can be treated in a variety of ways which may achieve stationarity. First, time-series can be differenced (Section 2.7) which is useful for removing long-term trends or drifts; however the interpretation of any subsequent analysis may be altered (linear or piecewise de-trending provide a simpler and less problematic alternative that may be effective). Second, the data can be windowed (Section 3.3). Finally, minimal pre-filtering may be applied to help achieve stationarity (see Section 2.7), for example by removing electrical line noise by notch filtering at 50 (or 60) Hz (Barnett and Seth, 2011).

For *evoked* data, as exemplified by event-related potentials (ERPs), nonstationarity is likely to be a common issue. Given a large number of trials, "vertical regression" as outlined in Section 3.3 can be implemented using extremely short windows. An alternative (complementary) approach is to subtract the ensemble average ERP from each trial, which—assuming low inter-trial ERP variation—should result in stationary residual time series reflecting the induced response (Chen et al., 2006). However, this method (i) risks "throwing out the baby with the bathwater" by dispensing with the ERP itself, and (ii) will fail if there is any substantial inter-trial ERP variability, which is common (Wang et al., 2008).

Volume conduction in EEG may lead to excessive colinearity (Section 3.3) among EEG time series in sensor space (this is less of a problem for MEG). Some useful approaches here are application of a surface Laplacian transform, which has the effect of spatially decorrelating the data (e.g. Cohen and van Gaal, 2013),

use of factor models (Flamm et al., 2013) or projection of sensor-level VAR model coefficients onto the locations of neural sources (Michalareas et al., 2013). The validity of application to *source-localised* EEG data depends on the method of source-localisation and is beyond the present scope (see Barrett et al., 2012 for an illustrative example). Finally, as emphasised earlier, spectral G-causality of M/EEG data should be measured by the band-limited approach (Section 2.7) and not by prefiltering into the desired band-pass and then applying time-domain G-causality.

4.2. Application to intracranial EEG and LFP data

Intracranially recorded EEG (iEEG) and local field potential (LFP) data also provide data well suited for G-causality analysis. Unlike surface M/EEG this data is usually spatially highly precise while sharing the advantages of high temporal resolution and sampling rate. However, perhaps counter-intuitively, LPF and iEEG data can sometimes appear to be "too clean" inasmuch as the stochastic variance on which VAR modelling depends is overshadowed by non-stationary deflections and fluctuations which may have both neural and non-neural (e.g. due to electrode movement) origins, and which may reflect long-term memory effects (Section 2.1) which are difficult to accommodate within a VAR framework. As well as the techniques described above (for M/EEG), one additional possibility for iEEG/LFP is to transform the data using a bipolar montage/reference which may reduce the impact of common sources and drifts, emphasising the residual stochastic activity. However, a fully adequate treatment of inherently nonstationary (and especially long-term memory) data is likely to require substantially adapted methods, e.g. VARFIMA modelling (Section 3.3). Despite these caveats, G-causality analysis of stationary iEEG/LFP data is likely to be highly informative; see Gaillard et al. (2009) for an illuminating example.

4.3. Application to fMRI BOLD data

Application of G-causality to fMRI BOLD data has been highly controversial for apparently good reasons; e.g. David et al. (2008), Valdes-Sosa et al. (2011). First, the BOLD signal (as captured by the *hemodynamic response function*, HRF) is an indirect, sluggish, and variable (inter-regionally and inter-subjectively), transformation of underlying neural activity (Handwerker et al., 2012). Second, typical fMRI protocols involve severe downsampling with sample intervals (repetition times, TRs) normally ranging from 1 to 3 s, substantially longer than typical inter-neuron delays. Inter-regional HRF variation has been argued to be particularly devastating for G-causality analysis: if X is causally driving Y at the neural level, but if the BOLD response to X peaks later than the BOLD response to Y, the suspicion is that G-causality analysis would falsely infer that Y is driving X (David et al., 2008). Perhaps surprisingly, this is not the case. In fact, G-causality of fMRI BOLD data is robust to a wide variety of changes in HRF properties, including notably their time-to-peak (Seth et al., 2013). This is because the HRF is effectively a slow, moving average filter, to which G-causality is in principle invariant. (Intuitively, this turns on recognising that a convolution is not the same as a delay. We have recently provided a detailed theoretical proof together with a range of simulations which fully establish this result[19] (Seth et al., 2013). Notably, these simulations include detailed population-based spiking neuron generative models coupled to the biomechanically realistic Balloon–Windkessel model of hemodynamic responses, eliminating concerns that the invariance properties described above depend on simplified VAR-based generative models of neuronal responses.

Unfortunately, the severe downsampling imposed by fMRI, together with measurement noise, still undermine G-causal inference of BOLD data in many applications. When confounding HRFs are combined with these other factors, false G-causality inferences are indeed likely. This precludes naive application of G-causality to fMRI data generally. While technological developments supporting ultra-low TRs (Feinberg and Yacoub, 2012) and de-noising (Rasmussen et al., 2012) promise to alleviate these problems, in the absence of these developments a conservative methodology is recommended. Useful strategies include (i) using as short a TR as possible by compromising on coverage; (ii) examining *changes* in G-causality between experimental conditions, rather than attempting to identify "ground truth" G-causality patterns; (iii) correlating changes in G-causality magnitude with behavioural variables such as reaction times across trials (or trial blocks) (Wen et al., 2012), and (iv) computing the so-called "difference of influence" term (Roebroeck et al., 2005) which may provide some robustness to HRF variation. Alternative promising approaches include estimation of state-space models which jointly parameterize functional connectivity and hemodynamic responses (Ryali et al., 2011) or blind deconvolution of the HRF to retrieve the underlying neuronal processes (Havlicek et al., 2011; Wu et al., 2013). In general, G-causality of fMRI BOLD data should be treated with caution and may best be interpreted as exploratory (Friston et al., 2013). See Seth et al. (2013) for further discussion of this important application domain.

4.4. Application to spiking (i.e. point process) data

Point process data obtained from direct neural recordings represents the other end of the spatiotemporal scale from fMRI, but carries its own challenges with respect to G-causality analysis. The main challenge is that point process spike train data is not suitable for modelling by linear VAR models. A theoretically principled approach is to replace the VAR modelling step with fitting of full and reduced point process models, within a framework of maximum likelihood estimation (Okatan et al., 2005; Truccolo et al., 2005; Kim et al., 2011; Gerhard et al., 2013). The MVGC toolbox does not support this in its current version. A simpler (although arguably less principled) approach is to obtain an estimate of the point process spectrum of spiking neural data; G-causal analysis may then proceed via spectral factorisation (Section 2.1), effectively treating the estimated spectrum as if it derived from a VAR process (Nedungadi et al., 2009). A "quick and dirty" alternative is to convolve each spike train with a Gaussian or half-Gaussian (causal) kernel, choosing the kernel width by some appropriate function of the mean inter-spike interval – see e.g. Cadotte et al. (2008), Kispersky et al. (2011). Assuming the resulting time series data satisfy stationarity tests, G-causality can then be applied. Due caution is needed in any interpretation of the results since this, again, is not a theoretically principled solution.

5. Conclusions

The MVGC toolbox provides a comprehensive set of Matlab routines for implementing G-causality analysis in the time and frequency domains and in both conditional and unconditional cases. It optimises computational efficiency, numerical accuracy and statistical inference by leveraging multiple equivalent representations

[19] The proof turns on the identification of the HRF convolution as an invertible filter and the consequent filter-invariance of G-causality (Section 2.7). While there is good evidence that the HRF convolution is causal (Handwerker et al., 2004), whether it is fully invertible (i.e. minimum-phase) for typical HRF parameters (Handwerker et al., 2004) requires further research.

of a VAR model by regression parameters, the autocovariance sequence, and the cross-power spectral density of the underlying process. By this approach, it is able to compute G-causality quantities by a "one shot" regression, without requiring separate "reduced" regressions which reduce statistical power and model estimation accuracy; the awkward case of conditional spectral G-causality is also fluently catered for. The toolbox provides extensive functionality for statistical significance testing, model order estimation, and error checking. It seamlessly integrates into the Matlab environment, offering multiple demonstration scripts and extensive documentation throughout.

As with any advanced statistical method, G-causality analysis via the MVGC toolbox should be implemented with care and with a good understanding of the underlying statistical principles and practical constraints, as described in this paper and elsewhere. This being the case, we hope that the MVGC toolbox will help address a major challenge for current neuroscience, namely the deciphering of the functional organisation of neural systems. We also anticipate its fruitful application in other contexts involving complex dynamic systems, both within biology and beyond.

Acknowledgements

We are grateful to Adam Barrett for helpful discussions and comments, and for testing pre-release versions of the MVGC toolbox. For financial support we are grateful to the EPSRC (G/700543/1) and to the Dr. Mortimer and Dame Theresa Sackler Foundation, which supports the work of the Sackler Centre for Consciousness Science.

Software availability: The MVGC toolbox may be downloaded from www.sussex.ac.uk/sackler. The software is freely available for non-profit academic usage under the GNU General Public License (GPL), version 3 or optionally any later version; see www.gnu.org/licenses.

Appendix A. MVGC algorithms

A.1. Core algorithms

We now review the algorithms 𝔸1–𝔸15 implementing the MVGC toolbox key computational pathways (Fig. 1). All multivariate time series $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ in what follows are n-variable of length m and assumed already *de-meaned*; i.e. the sample mean

$$\overline{\boldsymbol{u}} \equiv \frac{1}{m}\sum_{t=1}^{m}\boldsymbol{u}_t \tag{A.1}$$

has been subtracted from each sample $\boldsymbol{u}_t$ (see stats/demean). Although all MVGC routines that reference time series data generally accept/return *multi-trial* data, here for clarity we assume single-trial data. A model order p is assumed to have been selected.

𝔸1 Sample autocovariance estimation

The sample autocorrelation sequence [*cf.* (5)] is estimated as

$$\hat{\Gamma}_k(\boldsymbol{u}) = \frac{1}{m-k-1}\sum_{t=k+1}^{m}\boldsymbol{u}_t \cdot \boldsymbol{u}_{t-k}^{\top} \quad k = 0, 1, 2, \ldots \tag{A.2}$$

The factor of $1/(m-k-1)$ rather than $1/(m-k)$ is used to obtain an unbiased estimate. We do not recommend this as a computational pathway since (auto)covariance estimates can be unacceptably biased by error in estimation of the mean (A.1). An experimental routine implementing a potentially more accurate algorithm due to Shkolnisky et al. (Shkolnisky et al., 2008) is included, but we nonetheless do not recommend this pathway. Another drawback is that it is difficult to ascertain how many lags k of autocovariance will be appropriate to avoid overfitting while adequately modelling variation in the data.

Key routines: core/tsdata_to_autocov, experimental/tsdata_to_autocov_debias

𝔸2 VAR parameter estimation

The MVGC toolbox supplies two algorithms for fitting parameters (A_k, Σ) for a VAR model (1). Both yield VAR parameter estimates that are asymptotically equivalent to the corresponding ML estimate.

The first method is a standard OLS, which computes a least-squares estimate for the sample estimators $\hat{A}_k$. Given a sample time series $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ and estimated regression coefficient matrices $\hat{A}_1, \ldots, \hat{A}_p$, the residual errors for the regression are

$$\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{u}_t - \sum_{k=1}^{p}\hat{A}_k\boldsymbol{u}_{t-k} \quad t = p+1, \ldots, m \tag{A.3}$$

In an OLS, the $\hat{A}_k$ are chosen so as to minimise the mean squared error $E^2 = (1/(m-p))\sum_{t=p+1}^{m}\|\hat{\boldsymbol{\varepsilon}}_t\|^2$ where $\|\cdot\|$ denotes the L^2 (Euclidean) vector norm. In the MVGC toolbox the OLS is implemented by the Matlab "`/`" (`mrdivide`) operator, which solves the overdetermined linear system $\sum_{k=1}^{p}\hat{A}_k\boldsymbol{u}_{t-k} = \boldsymbol{u}_t, \quad t = p+1, \ldots, m$ in the least-squares sense via QR decomposition. An estimate for the residuals covariance matrix is then obtained as the unbiased sample covariance of the residual errors [*cf.* Eq. (A.2)]:

$$\hat{\Sigma} = \frac{1}{m-p-1}\sum_{t=p+1}^{m}\hat{\boldsymbol{\varepsilon}}_t \cdot \hat{\boldsymbol{\varepsilon}}_t^{\top} \tag{A.4}$$

The second supported method is a variant of the so called "LWR algorithm"—a term generally referring to a class of multivariate extensions to Durbin recursion (Levinson, 1947; Durbin, 1960; Whittle, 1963; Wiggins and Robinson, 1965)—due to Morf et al. (Morf et al., 1978). The Morf variant estimates regression coefficient matrices $\hat{A}_k$ recursively for $k = 1, 2, \ldots$ from the time series data $\boldsymbol{u}_t$. It is widely acknowledged to be very stable, and has the advantage that not only VAR coefficients but also estimates $\hat{\Sigma}$ for the residuals covariance matrix (and hence the VAR maximum likelihood estimator) are computed recursively. It is thus extremely efficient for use in likelihood-based model selection criteria such as the AIC or BIC (McQuarrie and Tsai, 1998), in comparison with OLS estimation, which needs to be recomputed for each model order. In practice, for a single estimate, the Morf algorithm may be faster or slower than an equivalent OLS, depending on number of variables, length of time series and model order. We refer the reader to Morf et al. (1978) for details of the algorithm.

Key routines: core/tsdata_to_var, core/tsdata_to_infocrit

𝔸3 VAR simulation

The routine core/var_to_tsdata returns simulated multi-trial, multivariate VAR(p) test data $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ according to (1) with normally distributed iid residuals $\boldsymbol{\varepsilon}_t$, for given VAR parameters $(A_1, \ldots, A_p, \Sigma)$. Initial samples will generally contain non-stationary transients, and may thus be truncated. Truncation may be performed automatically; a sufficient number of initial samples to approximate stationarity is calculated according to the estimated number of time steps for autocovariance to decay to near zero (𝔸5).

Key routine: core/var_to_tsdata

𝔸4 Spectral estimation

Although we do not recommend spectral estimation as a starting point for MVGC estimation (Sections 3.1 and 3.1.1 – although see Dhamala et al., 2008a) examination of cross-power spectra may be a useful preliminary step in time series data analysis.

The routine core/tsdata_to_cpsd implements two different methods of estimating the cross-power spectral density from time series data (further algorithms may be added in future). The first algorithm uses Welch's method (Welch, 1967; Oppenheim et al., 1998) which splits the data into overlapping "windows", computes modified periodograms of the overlapping segments, and averages the resulting periodograms to produce the power spectral density estimates. It is implemented via the Matlab Signal Processing Toolbox functions `pwelch` and `cpsd`. Window length and overlap parameters may be specified by the user. The second algorithm uses a multi-taper method (Percival and Walden, 1993) adapted from the Chronux neural data analysis package (Mitra and Bokil, 2008). In addition to window length and overlap parameters, a time bandwidth parameter and the number of tapers may also be set by the user. Discrete prolate spheroidal (Slepian) sequences for the multi-taper method are calculated using the function `dpss` from the Matlab Signal Processing Toolbox.

Key routine: core/tsdata_to_cpsd

𝔸5 Yule-Walker reverse solution

Given VAR(p) parameters ($A_1, \ldots, A_p, \Sigma$), we wish to solve the Yule-Walker equations (6) for the autocovariance sequence Γ_k. Firstly, we note that if the Γ_k are known for $k = 0, 1, \ldots, p-1$, then since (6) expresses Γ_k in terms of $\Gamma_{k-1}, \ldots, \Gamma_{k-p}$, the Γ_k may be calculated recursively for $k \geq p$. Thus if we can solve (6) up to $k = p-1$ then we can calculate Γ_k up to arbitrary lags.

Consider now the case $p = 1$. Setting $A \equiv A_1$, the first two Yule-Walker equations are $\Gamma_0 = A\Gamma_1{}^\top + \Sigma$ and $\Gamma_1 = A\Gamma_0$, leading to (note that Γ_0 is a symmetric matrix)

$$\Gamma_0 = A\Gamma_0 A^\top + \Sigma \tag{A.5}$$

This is a discrete-time Lyapunov equation (Bartels and Stewart, 1972) for Γ_0, for which efficient numerical solvers are available. The Matlab Control System Toolbox function `dlyap` solves equations of the form (A.5); if the Control System Toolbox is not available, then the MVGC toolbox function utils/dlyap_aitr implements an efficient iterative solver. Thus we may calculate Γ_0 for a VAR(1). We now use a standard trick (Lütkepohl, 2005) to express a VAR(p) as a VAR(1) on a new set of variables. Given the VAR(p) (1), we obtain the VAR(1)

$$\boldsymbol{U}_t^p = \boldsymbol{A}^p \boldsymbol{U}_{t-1}^p + \boldsymbol{\varepsilon}_t^p \tag{A.6}$$

where

$$\boldsymbol{U}_t^p \equiv \begin{pmatrix} \boldsymbol{U}_t \\ \boldsymbol{U}_{t-1} \\ \vdots \\ \boldsymbol{U}_{t-p+1} \end{pmatrix} \tag{A.7}$$

$$\boldsymbol{A}^p \equiv \begin{pmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{pmatrix} \tag{A.8}$$

$$\boldsymbol{\varepsilon}_t^p \equiv \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{A.9}$$

and the residuals covariance matrix is

$$\boldsymbol{\Sigma}^p \equiv \mathrm{cov}(\boldsymbol{\varepsilon}_t^p) = \begin{pmatrix} \Sigma & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} \tag{A.10}$$

We find that the covariance matrix of $\boldsymbol{U}_t^p$ is given by

$$\boldsymbol{\Gamma}_0^p = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \ldots & \Gamma_{p-1} \\ \Gamma_1{}^\top & \Gamma_0 & \ldots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p-1}{}^\top & \Gamma_{p-2}{}^\top & \ldots & \Gamma_0 \end{pmatrix} \tag{A.11}$$

so that solving the discrete-time Lyapunov equation $\Gamma_0^p = \mathbf{A}^p \boldsymbol{\Gamma}_0^p \mathbf{A}^{p\top} + \boldsymbol{\Sigma}^p$ for Γ_0^p yields $\Gamma_1, \ldots, \Gamma_{p-1}$ and the Γ_k may be calculated recursively up to arbitrary lags as described above.

The autocovariance for a VAR decays exponentially with lag (Section 2.1); more precisely, the Yule-Walker equations for the VAR(1) (A.6) yield

$$\boldsymbol{\Gamma}_k^p = (\boldsymbol{A}^p)^k \boldsymbol{\Gamma}_0^p \quad k = 1, 2, \ldots \tag{A.12}$$

Now it is easy to show that the spectral radius $\rho(A)$ of the original VAR(p) is just the largest modulus of the eigenvalues of $\boldsymbol{A}^p$, so that $\|\Gamma_k^p\|$ decays with rate $\rho(A)$ where $\|\cdot\|$ is any consistent matrix norm. In the MVGC Toolbox routine core/var_to_autocov, the autocovariance sequence is calculated up to a maximum of κ lags such that $\rho(A)^\kappa$ is smaller than a given specified numerical tolerance (which defaults to 10^{-8}) beyond which it is presumed that the Γ_k will be negligibly small. Thus the maximum number of lags required will depend on the spectral radius of the VAR, and may potentially become very large as the VAR approaches the unstable regime $\rho(A) \to 1$.

Key routine: core/var_to_autocov

𝔸6 Yule-Walker solution

Given an autocovariance sequence Γ_k the Yule-Walker equations (6) may be solved for VAR parameters ($A_1, \ldots, A_q, \Sigma$) under the assumption that Γ_k is the autocovariance sequence of a VAR(q). The equations for $k = 1, \ldots, q$ may be written as $\begin{pmatrix} \Gamma_1 & \ldots & \Gamma_q \end{pmatrix} = \begin{pmatrix} A_1 & \ldots & A_q \end{pmatrix} \cdot \boldsymbol{\Gamma}_0^q$ so that

$$\begin{pmatrix} A_1 & \ldots & A_q \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \ldots & \Gamma_q \end{pmatrix} \cdot (\boldsymbol{\Gamma}_0^q)^{-1} \tag{A.13}$$

and the $k = 0$ equation then gives $\Sigma = \Gamma_0 - \begin{pmatrix} A_1 & \ldots & A_q \end{pmatrix} \cdot \begin{pmatrix} \Gamma_1 & \ldots & \Gamma_q \end{pmatrix}^\top$. In fact, if the Γ_k are estimated in sample (*cf.* 𝔸1), then (A.13) is just the OLS solution for VAR(q) coefficients (*cf.* 𝔸2).

However, the solution (A.13) has the practical disadvantage that for n variables it requires inversion of an $nq \times nq$ matrix, which may be computationally prohibitive if q is large (*cf.* 𝔸5). Furthermore, if the *true* VAR model is stable and of order p, the A_k computed from (A.13) are not guaranteed to be stable, even if $q = p$. In practice, the true model order will not in general be known exactly; indeed, in

the MVGC approach, the Γ_k will generally be calculated as in (𝔸5) up to q large enough to ensure negligible autocovariance for $k > q$.

The MVGC Toolbox function core/autocov_to_var uses instead an LWR algorithm due to Whittle (Whittle, 1963) to solve (6) recursively for VAR parameters ($A_1, \ldots, A_q, \Sigma$). Importantly, Whittle's algorithm, unlike the OLS solution (A.13), guarantees that if the true VAR is stable of order p, then the calculated model ($A_1, \ldots, A_q, \Sigma$) is also stable, even if not of the correct order; i.e. if $q \neq p$. For n variables the algorithm requires $2q$ separate $n \times n$ matrix inversions, and is thus likely to be computationally tractable up to far higher model orders. The reader is referred to Whittle (1963) for details.

Key routine: core/autocov_to_var

𝔸7 VAR spectral factorisation

Although as mentioned previously (Section 2.1) there is no known closed-form solution of (9) for $H(\lambda)$, Σ in terms of $S(\lambda)$, it is known that a unique solution to the *spectral factorisation* problem exists. The routine core/cpsd_to_var deploys an iterative algorithm due to Wilson (Wilson, 1972) to achieve this numerically. The implementation is based on code kindly provided by G. Rangarajan (see also Dhamala et al., 2008a,b). The VAR coefficients A_k may then be recovered from $H(\lambda)$ by a matrix inversion and inverse Fourier transform; the utility function utils/trfun2var performs this calculation.

We remark that this is not a recommended MVGC computational pathway, since our tests indicate that, numerically, over a wide range of scenarios, spectral factorisation may be more efficiently and accurately calculated by transforming to the time domain (i.e. to the autocovariance sequence (𝔸10), and using Whittle's LWR algorithm (𝔸6).

Key routine: core/cpsd_to_var

𝔸8 VAR spectral calculation

The function core/var_to_cpsd implements the relations (9) and (10) directly to compute a VAR CPSD $S(\lambda)$ from VAR coefficients (A_k, Σ); a fast Fourier transform (FFT) is used to calculate the transfer function $H(\lambda)$ from the VAR coefficients.

Key routine: core/var_to_cpsd

𝔸9, 𝔸10 Transforming between the autocovariance sequence and cross-power spectral density

Calculating the CPSD from the autocovariance sequence entails the Fourier transform (7), implemented in the toolbox as an FFT. In terms of normalised frequencies, i.e. with period 2π, the (discrete) FFT of a sequence G_k approximates

$$\mathcal{F}[G](\lambda) = \sum_{k=0}^{\infty} G_k e^{-i\lambda k} \quad 0 \le \lambda \le 2\pi \tag{A.14}$$

From (7) we thus find:

$$S(\lambda) = \mathcal{F}[\Gamma](\lambda) + \mathcal{F}[\Gamma](\lambda)^* - \Gamma_0 \quad 0 \le \lambda \le 2\pi \tag{A.15}$$

Transforming in the other direction entails the inverse Fourier transform (8), implemented in the toolbox as an inverse fast Fourier transform (IFFT). Since the (discrete) IFFT of a function $P(\lambda)$ approximates

$$\mathcal{F}^{-1}[P]_k = \frac{1}{2\pi} \int_0^{2\pi} P(\lambda) e^{i\lambda k} \, \mathrm{d}\lambda \quad k = 0, 1, 2, \ldots \tag{A.16}$$

we want to represent (8) as an integral over $[0, 2\pi]$, rather than over $[-\pi, \pi]$. We may calculate that

$$\Gamma_k = (-1)^k \mathcal{F}^{-1}[P]_k \quad k = 0, 1, 2, \ldots \tag{A.17}$$

with

$$P(\lambda) = S(\lambda - \pi) = \begin{cases} \overline{S(\pi - \lambda)} & 0 \le \lambda \le \pi \\ S(\lambda - \pi) & \pi < \lambda \le 2\pi \end{cases} \tag{A.18}$$

in terms of $S(\lambda)$ on $[0, \pi]$, which is what we require since the MVGC toolbox represents all spectral quantities up to the Nyqvist frequency, which corresponds to $[0, \pi]$ for normalised frequencies (*cf.* Section 2.1).

Key routines: core/autocov_to_cpsd, core/cpsd_to_autocov

𝔸11, 𝔸12 Autocovariance and spectral transforms for reduced regression

Let Γ_k be the autocovariance sequence and $S(\lambda)$ the CPSD for a universe $\boldsymbol{U}_t$ of variables. We assume that we may take $\Gamma_k = 0$ for $k > q$ (i.e. we take q large enough that the autocovariance decays to some tolerance, as in 𝔸5). Suppose that $\boldsymbol{U}_t$ splits as $\boldsymbol{U}_t = \begin{pmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{pmatrix}$ and we have a reduced regression

$$\boldsymbol{X}_t = \sum_{k=1}^{q} B_k \boldsymbol{X}_{t-k} + \boldsymbol{X}_t^{\dagger} \tag{A.19}$$

with (white noise) residuals $\boldsymbol{X}_t^{\dagger}$. Note that we we may restrict the reduced regression to q lags, since from the Yule-Walker equations (6) it may be seen that $\Gamma_k = 0$ for $k > q$ implies that $B_k = 0$ for $k > q$. Considering the residuals $\boldsymbol{X}_t^{\dagger}$ as new variables, the problem is to calculate the autocovariance sequence $\underset{k}{\Gamma^{\dagger}}$ and CPSD $S^{\dagger}(\lambda)$ for the transformed universe $\boldsymbol{U}_t^{\dagger} = \begin{pmatrix} \boldsymbol{X}_t^{\dagger} \\ \boldsymbol{Y}_t \end{pmatrix}$. Now since the $\boldsymbol{X}_t^{\dagger}$ are residuals of a regression they are white noise (i.e. iid and serially uncorrelated), so that $\underset{xx,k}{\Gamma^{\dagger}} = \delta_{k0} \underset{xx}{\Sigma^{\dagger}}$, where $\underset{xx}{\Sigma^{\dagger}} = \mathrm{cov}(\boldsymbol{X}_t^{\dagger})$. We also have $\underset{yy,k}{\Gamma^{\dagger}} = \Gamma_{yy,k}$ and from (A.19) it follows that the $\Gamma_{xy}^{\dagger}$ and $\Gamma_{yx}^{\dagger}$ are derived by convolving corresponding Γ terms with B:

$$\Gamma_k^{\dagger} = \begin{pmatrix} \delta_{k0} \underset{xx}{\Sigma^{\dagger}} & \Gamma_{xy,k} - \sum_{\ell=1}^{q} B_\ell \Gamma_{xy,k-\ell} \\ \Gamma_{yx,k} - \sum_{\ell=1}^{q} \Gamma_{yx,k+\ell} B_\ell{}^{\top} & \Gamma_{yy,k} \end{pmatrix} \tag{A.20}$$

for $k = 0, 1, 2, \ldots$. We thus need to calculate $\underset{xy,k}{\Gamma^{\dagger}}$ to $k = 2q$ lags (although in practice it seems that q lags is sufficient) and $\underset{yx,k}{\Gamma^{\dagger}}$ to q lags. In the spectral domain, again since $\boldsymbol{X}^{\dagger}$ is white noise, the CPSD of $\boldsymbol{X}^{\dagger}$ is just the flat spectrum $\underset{xx}{S^{\dagger}}(\lambda) = \underset{xx}{\Sigma^{\dagger}}$, and we also have $\underset{yy}{S^{\dagger}}(\lambda) = S_{yy}(\lambda)$. From (A.20), and the Convolution Theorem (Oppenheim et al., 1998), we have:

$$S^{\dagger}(\lambda) = \begin{pmatrix} \underset{xx}{\Sigma^{\dagger}} & B(\lambda) S_{xy}(\lambda) \\ S_{yx}(\lambda) B(\lambda)^* & S_{yy}(\lambda) \end{pmatrix} \tag{A.21}$$

where $B(\lambda) \equiv I - \sum_{k=1}^{q} B_k e^{-ik\lambda}$ is the Fourier transform of the reduced regression coefficients.

The principal application of the transformations is to the reduced regression (31) in the calculation of conditional spectral MVGCs. Although the recommended pathway for this calculation

is via the autocovariance sequence, in fact the spectral transformation (A.21) is more efficient in practice (especially if q is large), since fast Fourier transformation may be deployed; in fact the default method in the routine core/autocov_xform is to convert to the CPSD (𝔸11), apply (A.21) and then convert back to the autocovariance sequence (𝔸12). The function core/cpsd_xform is also implemented for completeness.

Key routines: core/autocov_xform, core/cpsd_xform

A.2. G-causality algorithms

𝔸13 Time-domain G-causality calculation

Time domain causalities are calculated from the autocovariance sequence Γ_k—most likely obtained via core/var_to_autocov (𝔸5) from estimated VAR parameters—by the routine gc/autocov_to_mvgc according to (18) in the unconditional and (22) in the conditional case. The routine core/autocov_to_var (𝔸6) is used to calculate the residuals covariance matrices for both the full and reduced regressions. The routine gc/autocov_to_pwcgc calculates similarly the pairwise-conditional causalities (23).

Key routines: gc/autocov_to_mvgc, gc/autocov_to_pwcgc

𝔸14 Frequency-domain (unconditional) G-causality calculation

Spectral causalities are calculated from the autocovariance sequence Γ_k—again most likely obtained via core/var_to_autocov (𝔸5) from estimated VAR parameters—by the routines gc/autocov_to_smvgc or gc/autocov_to_spwcgc for the pairwise-conditional case. In the unconditional case (28), VAR parameters (A_k, Σ) for the full regression are first calculated by core/autocov_to_var (𝔸6), and the CPSD $S(\lambda)$ and transfer function $H(\lambda)$ are then calculated by core/var_to_cpsd (𝔸8).

In the conditional case, VAR parameters are firstly obtained from the autocovariance sequence Γ_k for the reduced regression (31) by core/autocov_to_var (𝔸6), and are used to transform the autocovariance sequence by core/autocov_xform (𝔸11). The conditional spectral causalities are then calculated from the transformed autocovariance sequence $\Gamma_k^\dagger$ as *un*conditional spectral causalities according to (33). There is in fact a simplification due to the fact that [*cf.* 𝔸12, Eq. (A.21)] the power spectrum $S_{xx}^\dagger(\lambda)$ for the transformed target variable is flat and will thus already have been calculated as the covariance matrix $\Sigma_{xx}^\dagger$.

Key routines: gc/autocov_to_smvgc, gc/autocov_to_spwcgc

𝔸15 Integration of spectral G-causality

The routine gc/smvgc_to_mvgc implements the spectral MVGC integrals (30) and (34) and the band-limited variants (36) and (37); see Sections 2.6 and 2.7 respectively. Numerical integration is performed by a simple trapezoidal rule quadrature.

Key routines: gc/smvgc_to_mvgc

Appendix B. Exact solution of a minimal causal VAR(1)

We consider the VAR(1) of (38):

$$\begin{aligned} X_t &= aX_{t-1} + cY_{t-1} + \varepsilon_{x,t} \\ Y_t &= \quad\quad bY_{t-1} + \varepsilon_{y,t} \end{aligned} \tag{B.1}$$

where $|a| < 1$, $|b| < 1$ and the residuals $\varepsilon_{x,t}$, $\varepsilon_{y,t}$ are unit-variance uncorrelated white noise. The coefficients matrix of (B.1) is

$$A = \begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \tag{B.2}$$

and we may easily calculate the transfer matrix as

$$H(z) = \frac{1}{(1-az)(1-bz)} \begin{pmatrix} 1-bz & cz \\ 0 & 1-az \end{pmatrix} \tag{B.3}$$

with $z = e^{-i\lambda}$. The CPSD of the joint (X_t, Y_t) process (B.1) may thus be calculated from (9) as

$$S(z) = \frac{1}{|1-az|^2|1-bz|^2} \begin{pmatrix} |1-bz|^2 + c^2 & cz(1-a\bar{z}) \\ c\bar{z}(1-az) & |1-az|^2 \end{pmatrix} \tag{B.4}$$

for $|z| = 1$ in the complex plane (*cf.* Barnett and Seth, 2011, Section 4). By inspection, the x-component $S_{xx}(z)$ of the CPSD may be factorised explicitly in the form

$$S_{xx}(z) = \frac{|1-bz|^2 + c^2}{|1-az|^2|1-bz|^2} = \sigma^2 h(z) h(z)^* \tag{B.5}$$

where the transfer function is of the form

$$h(z) = \frac{1-rz}{(1-az)(1-bz)} \tag{B.6}$$

and σ^2 is just the residuals variance of the process X_t considered as a VAR(∞). Note that this implies that, although the joint process (X_t, Y_t) is VAR(1), the process X_t alone is actually VARMA(2, 1) (*cf.* Section 2.4). In order to satisfy (B.5) and (B.6), the unknowns σ^2 and r must satisfy

$$\sigma^2|1-rz|^2 = |1-bz|^2 + c^2 \tag{B.7}$$

for all $z = e^{-i\lambda}$ on the unit circle in the complex plane, which requires

$$\sigma^2(1+r^2) = \Delta \equiv 1 + b^2 + c^2 \tag{B.8}$$

$$\sigma^2 r = b \tag{B.9}$$

We may then solve for σ^2:

$$\sigma^2 = \frac{1}{2}\left(\Delta + \sqrt{\Delta^2 - 4b^2}\right) \tag{B.10}$$

From (18) [note that by assumption $\Sigma_{xx} = \mathrm{var}\varepsilon_{x,t} \equiv 1$, and that $\Sigma'_{xx} = \sigma^2$] the (infinite-lag) G-causality in the $Y \to X$ direction is

$$\mathcal{F}_{Y\to X}^{(\infty)} = \ln \sigma^2 = \ln\left[\frac{1}{2}\left(\Delta + \sqrt{\Delta^2 - 4b^2}\right)\right] \tag{B.11}$$

[*cf.* Barnett and Seth (2011), Eq. (45)].

To calculate the 1-lag G-causality, we consider the 1-lag reduced regression

$$X_t = a' X_{t-1} + \varepsilon'_t \tag{B.12}$$

which we need to solve for the residuals variance $\sigma'^2 \equiv \mathrm{var}(\varepsilon'_t)$ in the least-squares (or ML) sense. The OLS solution is given by the partial covariance (Barnett et al., 2009)

$$\sigma'^2 = var(X_t) - cov(X_t X_{t-1})^2 var(X_{t-1})^{-1} = \frac{\gamma_0^2 - \gamma_1^2}{\gamma_0} \tag{B.13}$$

where $\gamma_k \equiv \mathrm{cov}(X_t X_{t-k})$ is the autocovariance sequence for X_t. Now from the Yule-Walker equations (6) for the joint process (X_t, Y_t) with $k = 0, 1$ some straightforward algebra yields

$$\gamma_0 = \frac{1 + (1+ab)\delta}{1-a^2} \tag{B.14}$$

$$\gamma_1 = \frac{a + (a+b)\delta}{1-a^2} \tag{B.15}$$

with

$$\delta \equiv \frac{c^2}{(1-ab)(1-b^2)} \tag{B.16}$$

leading to

$$\mathcal{F}^{(1)}_{Y\to X} = \ln \sigma'^2 = \ln\left[\frac{1+2\delta+(1-b^2)\delta^2}{1+(1+ab)\delta}\right] \tag{B.17}$$

References

Aertsen A, Preißl H. Dynamics of activity and connectivity in physiological neuronal networks. In: Schuster H, editor. Nonlinear dynamics and neuronal networks. New York: VCH Publishers Inc.; 1991. p. 281–302.

Anderson MJ, Robinson J. Permutation tests for linear models. Aust New Zeal J Stat 2001;43:75–88.

Anderson TW. The statistical analysis of time series. New York: Wiley; 1971.

Antoniou A. Digital filters: analysis, design, and applications. New York, NY: McGraw-Hill; 1993.

Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett 2009;103:238701.

Barnett L, Bossomaier T. Transfer entropy as a log-likelihood ratio. Phys Rev Lett 2013;109:138105.

Barnett L, Seth AK. Behaviour of Granger causality under filtering: theoretical invariance and practical application. J Neurosci Methods 2011;201:404–19.

Barrett AB, Barnett L, Seth AK. Multivariate Granger causality and generalized variance. Phys Rev E 2010;81:41907.

Barrett AB, Murphy M, Bruno MA, Noirhomme Q, Boly M, Laureys S, et al. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. PLoS ONE 2012;7:e29072.

Bartels RH, Stewart G. Solution of the equation $AX+XB=C$. Commun ACM 1972;15:820–6.

Bressler S, Seth A. Wiener–Granger causality: a well established methodology. Neuroimage 2011;58:323–9.

Cadotte AJ, DeMarse TB, He P, Ding M. Causal measures of structure and plasticity in simulated and living neural networks. PLoS ONE 2008;3:e3355.

Chen Y, Bressler SL, Ding M. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. J Neurosci Methods 2006a;150:228–37.

Chen Y, Bressler SL, Knuth KH, Truccolo WA, Ding M. Stochastic modeling of neurobiological time series: power, coherence, Granger causality, and separation of evoked responses from ongoing activity. Chaos 2006;16, 026113–026113.

Cohen MX, van Gaal S. Dynamic interactions between large-scale brain networks predict behavioral adaptation after perceptual errors. Cereb Cortex 2013;23:1061–72.

Cooley JW, Tukey JW. An algorithm for the machine computation of the complex fourier series. Math Comput 1965;19:297–301.

David O, Guillemain I, Saillet S, Reyt S, Deransart C, Segebarth C, et al. Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol 2008;6:2683–97.

Deshpande G, Hu X. Investigating effective brain connectivity from FMRI data: past findings and current issues with reference to granger causality analysis. Brain Connect 2012;2:235–45.

Dhamala M, Rangarajan G, Ding M. Estimating Granger causality from Fourier and wavelet transforms of time series data. Phys Rev Lett 2008a;100:018701.

Dhamala M, Rangarajan G, Ding M. Analyzing information flow in brain networks with nonparametric Granger causality. Neuroimage 2008b;41:354–62.

Ding M, Bressler S, Yang W, Liang H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol Cybern 2000;83:35–45.

Ding M, Chen Y, Bressler S. Granger causality: basic theory and application to neuroscience. In: Schelter S, Winterhalder M, Timmer J, editors. Handbook of time series analysis. Wienheim: Wiley; 2006. p. 438–60.

Durbin J. The fitting of time series models. Rev Inst Int Stat 1960;28:233–44.

Edwards AWF. Likelihood (expanded edition). Baltimore: Johns Hopkins University Press; 1992.

Efron B. The jackknife, the bootstrap, and other resampling plans. In: Society of Industrial and Applied Mathematics CBMS-NSF Monographs; 1982. p. 38.

Engle F, Granger CWJ. Co-integration and error correction: representation, estimation and testing. Econometrica 1987;55:251–76.

Feinberg DA, Yacoub E. The rapid development of high speed, resolution and precision in fMRI. Neuroimage 2012.

Flamm C, Graef A, Pirker S, Baumgartner C, Deistler M. Influence analysis for high-dimensional time series with an application to epileptic seizure onset zone detection. J Neurosci Methods 2013;214:80–90.

Friston K, Moran R, Seth AK. Analysing connectivity with Granger causality and dynamic causal modelling. Curr Opin Neurobiol 2013;23:172–8.

Friston KJ. Functional and effective connectivity: a review. Brain Connect 2011;1:13–36.

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage 2003;19:1273–302.

Gaillard R, Dehaene S, Adam C, Clémenceau S, Hasboun D, Baulac M, et al. Converging intracranial markers of conscious access. PLoS Biol 2009;7, e61–e61.

Gerhard F, Kispersky T, Gutierrez GJ, Marder E, Kramer M, Eden U. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. PLoS Comput Biol 2013;9:e1003138.

Geweke J. Measurement of linear dependence and feedback between multiple time series. J Am Stat Assoc 1982;77:304–13.

Geweke J. Measures of conditional linear dependence and feedback between time series. J Am Stat Assoc 1984;79:907–15.

Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 1969;37:424–38.

Granger CWJ. Some properties of time series data and their use in econometric model specification. J Econometrics 1981;16:121–30.

Granger CWJ, Joyeux R. An introduction to long-memory time series models and fractional differencing. J Time Series Anal 1980;1:15–30.

Guo S, Seth A, Kendrick K, Zhou C, Feng J. Partial granger causality: eliminating exogenous inputs and latent variables. J Neurosci Methods 2008;172:79–93.

Hamilton JD. Time series analysis. Princeton, NJ: Princeton University Press; 1994.

Handwerker DA, Gonzalez-Castillo J, D'Esposito M, Bandettini PA. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. Neuroimage 2012.

Handwerker DA, Ollinger JM, D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 2004;21:1639–51.

Havlicek M, Friston KJ, Jan J, Brazdil M, Calhoun VD. Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. Neuroimage 2011;56:2109–28.

Hochberg Y, Tamhane AC. Multiple comparison procedures. New York: John Wiley; 1987.

Hyvärinen A, Karhunen J, Oja E. Independent component analysis. New York: John Wiley; 2001.

Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer-Verlag; 2002.

Kaiser A, Schreiber T. Information transfer in continuous processes. Physica D 2002;166:43–62.

Kim S, Putrino D, Ghosh S, Brown EN. A Granger causality measure for point process models of ensemble neural spiking activity. PLoS Comput Biol 2011;7:e1001110.

Kispersky T, Gutierrez GJ, Marder E. Functional connectivity in a rhythmic inhibitory circuit using granger causality. Neural Syst Circ 2011:1.

Ladroue C, Guo S, Kendrick K, Feng J. Beyond element-wise interactions: identifying complex interactions in biological processes. PLoS ONE 2009;4:e6899.

Levinson N. The Wiener RMS (root-mean-square) error criterion in filter design and prediction. J Math Phys 1947;25:261–78.

Lizier JT, Prokopenko M. Differentiating information transfer and causal effect. Eur Phys J B 2010;73:605–15.

Luo Q, Ge T, Feng J. Granger causality with signal-dependent noise. Neuroimage 2011;57:1422–9.

Lütkepohl H. New introduction to multiple time series analysis. Berlin: Springer-Verlag; 2005.

Masani P. Recent trends in multivariate prediction theory. In: Krishnaiah PR, editor. Multivariate analysis. New York: Academic Press; 1966. p. 51–382.

McQuarrie ADR, Tsai CL. Regression and time series model selection. Singapore: World Scientific Publishing; 1998.

Michalareas G, Schoffelen JM, Paterson G, Gross J. Investigating causality between interacting brain areas with multivariate autoregressive models of MEG sensor data. Hum Brain Mapp 2013;34:890–913.

Mitra P, Bokil H. Observed brain dynamics. New York: Oxford University Press; 2008.

Morf M, Viera A, Lee DTL, Kailath T. Recursive multichannel maximum entropy spectral estimation. IEEE Trans Geosci Electron 1978;16:85–94, http://dx.doi.org/10.1109/TGE.1978.294569.

Nedungadi AG, Rangarajan G, Jain N, Ding M. Analyzing multiple spike trains with nonparametric Granger causality. J Comput Neurosci 2009;27:55–64.

Okatan M, Wilson MA, Brown EN. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. Neural Comput 2005;17:1927–61.

Oppenheim AV, Shafer RW, Buck JR. Discrete-time signal processing. 2nd ed New Jersey: Prentice Hall; 1998.

Ozaki T. Time series modeling of neuroscience data. Boca Raton, FL: CRC Press; 2012.

Papoulis A, Pillai SU. Random variables and stochastic processes. New York: McGraw-Hill; 2002.

Pearl J. Causality: models, reasoning and inference. 2nd ed. Princeton, New York: Cambridge University Press; 2009.

Percival DB, Walden AT. Spectral analysis for physical applications: multitaper and conventional univariate techniques. Cambridge, UK: Cambridge University Press; 1993.

Rasmussen PM, Abrahamsen TJ, Madsen KH, Hansen LK. Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation. Neuroimage 2012;60:1807–18.

Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using granger causality and fMRI. Neuroimage 2005;25:230–42.

Roebroeck A, Formisano E, Goebel R. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. Neuroimage 2009.

Roebroeck A, Seth A, Valdes-Sosa P. Causal time series analysis of functional magnetic resonance imaging data. J Mach Learn Res 2010;12:65–94.

Roelstraete B, Rosseel Y. Does partial Granger causality really eliminate the influence of exogenous inputs and latent variables? J Neurosci Methods 2012;206:73–7.

Rozanov YA. Stationary random processes. San Francisco: Holden-Day; 1967.

Ryali S, Supekar K, Chen T, Menon V. Multivariate dynamical systems models for estimating causal interactions in fMRI. Neuroimage 2011;54:807–23.

Schreiber T. Measuring information transfer. Phys Rev Lett 2000;85:461–4.

Seth AK. A MATLAB toolbox for Granger causal connectivity analysis. J Neurosci Methods 2010;186:262–73.
Seth AK, Chorley P, Barnett L. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. Neuroimage 2013;65:540–55.
Shkolnisky Y, Sigworth FJ, Singer A. A note on estimating autocovariance from short-time observations; 2008, Unpublished, Available from: http://www.journalogy.org/Publication/5978390/a-note-on-estimating-autocovariance-from-short-time-observations
Silvennoinen A, Teräsvirta T. Multivariate GARCH models. In: Mikosch T, Kreiss JP, Davis RA, Andersen TG, editors. Handbook of financial time series. Berlin: Springer-Verlag; 2009. p. 201–29.
Solo V. On causality I: sampling and noise. In: IEEE Proceedings of the 46th Conference on Decision and Control; 2007. p. 3634–9.
Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol 2005;93:1074–89.
Valdes-Sosa PA, Roebroeck A, Daunizeau J, Friston K. Effective connectivity: influence, causality and biophysical modeling. Neuroimage 2011;58:339–61.
Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 1943;54:426–82.
Wang X, Chen Y, Ding M. Estimating Granger causality after stimulus onset: a cautionary note. Neuroimage 2008;41:767–76.
Welch PD. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 1967;15:70–3.
Wen X, Yao L, Liu Y, Ding M. Causal interactions in attention networks predict behavioral performance. J Neurosci 2012;32:1284–92.
Whittle P. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 1963;50:129–34.
Wiener N. The theory of prediction. In: Beckenbach EF, editor. Modern mathematics for engineers. New York: McGraw Hill; 1956. p. 165–90.
Wiggins RA, Robinson EA. Recursive solution of the multichannel filtering problem. J Geophys Res 1965;70:1885–91.
Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 1938;6:60–2.
Wilson GT. The factorization of matricial spectral densities. SIAM J Appl Math 1972;23:420–6.
Wu GR, Liao W, Stramaglia S, Ding JR, Chen H, Marinazzo D. A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. Med Image Anal 2013;17:365–74.