

Multivariate Granger Causality and Generalized Variance

Adam B. Barrett^{1*}, Lionel Barnett^{2†} and Anil K. Seth^{1§}

¹*Sackler Centre for Consciousness Science, School of Informatics,
University of Sussex, Brighton BN1 9QJ, UK*

²*Centre for Computational Neuroscience and Robotics, School of Informatics,
University of Sussex, Brighton BN1 9QJ, UK*

April 14, 2010

Abstract

Granger causality analysis is a popular method for inference on directed interactions in complex systems of many variables. A shortcoming of the standard framework for Granger causality is that it only allows for examination of interactions between single (univariate) variables within a system, perhaps conditioned on other variables. However, interactions do not necessarily take place between single variables, but may occur among groups, or “ensembles”, of variables. In this study we establish a principled framework for Granger causality in the context of causal interactions among two or more multivariate sets of variables. Building on Geweke’s seminal 1982 work, we offer new justifications for one particular form of multivariate Granger causality based on the generalized variances of residual errors. Taken together, our results support a comprehensive and theoretically consistent extension of Granger causality to the multivariate case. Treated individually, they highlight several specific advantages of the generalized variance measure, which we illustrate using applications in neuroscience as an example. We further show how the measure can be used to define “partial” Granger causality in the multivariate context and we also motivate reformulations of “causal density” and “Granger autonomy”. Our results are directly applicable to experimental data and promise to reveal new types of functional relations in complex systems, neural and otherwise.

PACS numbers: 87.19.L-, 87.10.Mn, 89.75.-k, 87.19.lj

Keywords: Granger causality, causal inference, multivariate statistics, generalized variance

1 Introduction

A key challenge across many domains of science and engineering is to understand the behavior of complex systems in terms of dynamical interactions among their component parts. A common way

*email: Adam.Barrett@sussex.ac.uk

†email: L.C.Barnett@sussex.ac.uk

§email: A.K.Seth@sussex.ac.uk

to address this challenge is by analysis of time series data acquired simultaneously from multiple system components. Increasingly, such analysis aims to draw inferences about *causal* interactions among system variables [1, 2, 3], as a complement to standard assessments of undirected functional connectivity as revealed by coherence, correlation, and the like.

A first step in any dynamical analysis is to identify target variables. Typically, subsequent analysis then assumes that functional (causal) interactions take place among these variables. However, in the general case it may be that explanatorily relevant causal interactions take place among *groups*, or “ensembles”, of variables [4, 5]. It is important to account for this possibility for at least two reasons. First, identification of target variables is usually based on *a priori* system knowledge or technical constraints, which may be incomplete or arbitrary, respectively. Second, even given appropriate target variables, it is possible that relevant interactions may operate at multiple scales within a system, with larger scales involving groups of variables. Consider an example from functional neuroimaging. In a typical fMRI¹ study, the researcher may identify *a priori* several “regions-of-interest” (ROI) in the brain, each represented in the fMRI dataset by multiple voxels, where each voxel is a variable comprising a single time series reflecting changes in the underlying metabolic signal. Assuming that the objective of the study is to assess the causal connectivity among the ROIs, a standard approach is to derive a single time series for each ROI either by averaging or by extracting a principal component [6]; alternatively, repeated pairwise analysis can be performed on each pair of voxels. A more appropriate approach, however, may be to consider causal interactions among the multivariate groups of voxels comprising each ROI. Similar scenarios could be concocted in a very wide range of application areas, including economics, biology, climate science, among others.

In this paper, we describe a principled approach to assessing causal interactions among multivariate groups of variables. Our approach is based on the concept of Granger causality (G-causality) [7, 8], a statistical notion of causality which originated in econometrics but which has since found widespread application in many fields, with a particular concentration in the neurosciences [1, 9]. G-causality is an example of time series inference on stochastic processes and is usually implemented via autoregressive modeling of multivariate time series. The basic idea is simple: one variable (or time series) can be called “causal” to another if the ability to predict the second variable is improved by incorporating information about the first. More precisely, given inter-dependent variables X and Y , it is said that “ Y G-causes X ” if, in a statistically suitable manner, Y assists in predicting the future of X beyond the degree to which X already predicts its own future. It is straightforward to extend G-causality to the conditional case [5], where Y is said to G-cause X , conditional on Z , if Y assists in predicting the future of X beyond the degree to which X and Z together already predict the future of X . Importantly, conditional G-causality is orthogonal to the notion of inferring causality among groups of variables, which is the focus of the present paper and which we term *multivariate* G-causality. In the multivariate case, the above description of G-causality is generalized to interactions among *sets* of interdependent variables \mathbf{X} , \mathbf{Y} , and (in the conditional multivariate case) \mathbf{Z} . The generalization we propose was originally introduced in the field of econometrics by Geweke in 1982 [5], but has since been almost totally overlooked. Indeed a different measure has recently appeared [4]. In the following, we derive several justifications for preferring Geweke’s measure, some of which we examine numerically. We go on to explore a series of implications for the analysis of complex systems in general, with a particular focus on applications in neuroscience.

¹Functional magnetic resonance imaging.

After laying out our conventions in Section 2, in Section 3 we introduce two alternative measures of multivariate G-causality. The formulations differ according to their treatment of the covariance matrices of residuals in the underlying autoregressive models: Geweke’s measure uses the *determinant* of this matrix (the generalized variance), while the other uses the *trace* (the total variance). Section 4 explores several advantageous properties of the determinant formulation as compared to the trace formulation. In brief, the determinant formulation is fully equivalent to transfer entropy [3] under Gaussian assumptions, is invariant under a wider range of variable transformations, is expandable as a sum of standard univariate G-causalities, and admits a satisfactory spectral decomposition. Numerically, we show that Geweke’s measure is just as stable as is the alternative measure based on the total variance. Section 5 extends the determinant formulation to the important case of “partial” G-causality which provides some measure of control with respect to unmeasured latent or exogenous variables. Section 6 extends a previously defined measure of “causal density” [10] which reflects the overall dynamical complexity of causal interactions sustained by a system. In Section 7 we show how multivariate G-causality can enhance a measure of “autonomy” (or “self-causation”) based on G-causality [11], and Section 8 carries the discussion towards the identification of macroscopic variables via the notion of causal independence. Section 9 provides a general discussion and summary of contributions.

2 Notational conventions and preliminaries

We use a mathematical vector/matrix notation in which bold type generally denotes vector quantities and upper-case type denotes matrices or random variables, according to context. All vectors are considered to be *column* vectors. ‘ \oplus ’ denotes *vertical concatenation* of vectors, so that for $\mathbf{x} = (x_1, \dots, x_n)^\top$ and $\mathbf{y} = (y_1, \dots, y_m)^\top$, $\mathbf{x} \oplus \mathbf{y}$ is the vector $(x_1, \dots, x_n, y_1, \dots, y_m)^\top$ of dimension $n + m$, where the symbol ‘ \top ’ denotes the transpose operator. We also write $|\cdot|$ for the determinant and $\text{tr}(\cdot)$ for the trace of a square matrix.

Given jointly distributed multivariate random variables (i.e. random vectors) \mathbf{X}, \mathbf{Y} , we denote by $\Sigma(\mathbf{X})$ the $n \times n$ matrix of covariances $\text{cov}(X_i, X_j)$ and by $\Sigma(\mathbf{X}, \mathbf{Y})$ the $n \times m$ matrix of cross-covariances $\text{cov}(X_i, Y_\alpha)$. We then use $\Sigma(\mathbf{X} | \mathbf{Y})$ to denote the $n \times n$ matrix

$$\Sigma(\mathbf{X} | \mathbf{Y}) \equiv \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y}) \Sigma(\mathbf{Y})^{-1} \Sigma(\mathbf{X}, \mathbf{Y})^\top, \quad (1)$$

defined when $\Sigma(\mathbf{Y})$ is invertible. $\Sigma(\mathbf{X} | \mathbf{Y})$ appears as the covariance matrix of the residuals of a linear regression of \mathbf{X} on \mathbf{Y} (c.f. Eq. (6) below); thus, by analogy with *partial correlation* [12] we term $\Sigma(\mathbf{X} | \mathbf{Y})$ the *partial covariance*² of \mathbf{X} given \mathbf{Y} . Similarly, given another jointly distributed variable \mathbf{Z} , we define the *partial cross-covariance*

$$\Sigma(\mathbf{X}, \mathbf{Y} | \mathbf{Z}) \equiv \Sigma(\mathbf{X}, \mathbf{Y}) - \Sigma(\mathbf{X}, \mathbf{Z}) \Sigma(\mathbf{Z})^{-1} \Sigma(\mathbf{Y}, \mathbf{Z})^\top. \quad (2)$$

The following identity [13] will be useful for deriving certain properties of multivariate G-causality:

$$|\Sigma(\mathbf{X} | \mathbf{Y})| = |\Sigma(\mathbf{X} \oplus \mathbf{Y})| / |\Sigma(\mathbf{Y})|. \quad (3)$$

²This is to be distinguished from the *conditional covariance*, which will in general be a random variable, though later we note that for *Gaussian* variables the notions coincide.

Suppose we have a multivariate stochastic process \mathbf{X}_t in discrete time³ (i.e. the random variables X_{it} are jointly distributed). We use the notation $\mathbf{X}_t^{(p)} \equiv \mathbf{X}_t \oplus \mathbf{X}_{t-1} \oplus \dots \oplus \mathbf{X}_{t-p+1}$ to denote \mathbf{X} itself, along with $p - 1$ lags, so that for each t , $\mathbf{X}_t^{(p)}$ is a random vector of dimension pn . Given the lag p , we also often use the shorthand notation $\mathbf{X}_t^- \equiv \mathbf{X}_{t-1}^{(p)}$ for the lagged variable.

3 Multivariate Granger causality

G-causality analysis is concerned with the comparison of different linear regression models of data. Thus, let us consider the (multivariate) linear regression of one random vector \mathbf{X} , the predictee, on another random vector \mathbf{Y} , the predictor:⁴

$$\mathbf{X} = A \cdot \mathbf{Y} + \boldsymbol{\varepsilon}, \quad (4)$$

where the $n \times m$ matrix A contains the regression coefficients and the random vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ comprises the residuals. The coefficients of this model are uniquely specified by imposing zero correlation between the residuals $\boldsymbol{\varepsilon}$ and the regressors (predictors) \mathbf{Y} . Via the Yule-Walker procedure [1, 13] one obtains

$$A = \Sigma(\mathbf{X}, \mathbf{Y}) \Sigma(\mathbf{Y})^{-1} \quad (5)$$

and finds the covariance matrix of the residuals to be given by

$$\Sigma(\boldsymbol{\varepsilon}) = \Sigma(\mathbf{X} | \mathbf{Y}), \quad (6)$$

with $\Sigma(\mathbf{X} | \mathbf{Y})$ defined as in (1).

Suppose now we have three jointly distributed, stationary⁵ multivariate stochastic processes $\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t$. Then to measure the G-causality from \mathbf{Y} to \mathbf{X} given \mathbf{Z} , one wants to compare the following two multivariate autoregressive (MVAR) models for the processes [8]:

$$\begin{aligned} \mathbf{X}_t &= A \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Z}_{t-1}^{(r)} \right) + \boldsymbol{\varepsilon}_t, \\ \mathbf{X}_t &= A' \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Y}_{t-1}^{(q)} \oplus \mathbf{Z}_{t-1}^{(r)} \right) + \boldsymbol{\varepsilon}'_t. \end{aligned} \quad (7)$$

Thus the predictee variable \mathbf{X} is regressed firstly on the previous p lags of itself plus r lags of the conditioning variable \mathbf{Z} and secondly, in addition, on q lags of the predictor variable \mathbf{Y} (in theory, if not in practice, p, q and r could be infinite).⁶

The standard measure of G-causality used in the literature is defined only for *univariate* predictor and predictee variables Y and X , and is given by the log of the ratio of the residual variances for

³While our analysis may be extended to *continuous* time we focus here on the discrete time case.

⁴Here and in the remainder of this paper we assume, without loss of generality, that all random vectors and random processes have zero mean; thus constant terms are omitted in all linear regressions.

⁵The analysis carries through for the non-stationary case, but for simplicity we assume here that all processes are stationary.

⁶This might be more familiar as *conditional* G-causality, with \mathbf{Z} the conditioning variable. In practice it is the more useful form; for the non-conditional version, \mathbf{Z} may simply be omitted.

the regressions (7). In our notation,⁷

$$\begin{aligned}\mathcal{F}_{Y \rightarrow X | \mathbf{Z}} &\equiv \ln \left(\frac{\text{var}(\varepsilon_t)}{\text{var}(\varepsilon'_t)} \right) \\ &= \ln \left(\frac{\Sigma(\varepsilon_t)}{\Sigma(\varepsilon'_t)} \right) \\ &= \ln \left(\frac{\Sigma(X | \mathbf{X}^- \oplus \mathbf{Z}^-)}{\Sigma(X | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-)} \right),\end{aligned}\tag{8}$$

where the last equality follows from the general formula (6). By stationarity this expression does not depend on time t . Note that the residual variance of the first regression will always be larger than or equal to that of the second, so that $\mathcal{F}_{Y \rightarrow X | \mathbf{Z}} \geq 0$ always. As regards statistical inference, it is known that the corresponding maximum likelihood estimator⁸ $\hat{\mathcal{F}}_{Y \rightarrow X | \mathbf{Z}}$ will have (asymptotically for large samples) a χ^2 -distribution under the null hypothesis $\mathcal{F}_{Y \rightarrow X | \mathbf{Z}} = 0$ [14, 15], and a non-central χ^2 -distribution under the alternative hypothesis $\mathcal{F}_{Y \rightarrow X | \mathbf{Z}} > 0$ [5, 16].

We now consider the case where predictee and predictor variables are no longer constrained to be univariate, i.e. multivariate G-causality. For a multivariate predictor, Eq. (8) above (with Y replaced by the bold-type \mathbf{Y}) is a valid and consistent formula for G-causality. However, for the case of a multivariate predictee there is not yet a standard definition for G-causality. One possibility is to simply use the multivariate mean square error (i.e. total variance, or expected squared length of the multivariate residual), leading to

$$\begin{aligned}\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}}^{tr} &\equiv \ln \left(\frac{\text{tr}(\Sigma(\varepsilon_t))}{\text{tr}(\Sigma(\varepsilon'_t))} \right) \\ &= \ln \left(\frac{\text{tr}(\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Z}^-))}{\text{tr}(\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-))} \right).\end{aligned}\tag{9}$$

We call this the trace version of multivariate G-causality (trvMVGC). As recently noted by Ladrone and colleagues [4] trvMVGC appears to be a natural extension of G-causality to the multivariate case because total variance is a common choice for a measure of goodness-of-fit or prediction error for a multivariate regression. Moreover, the measure is always non-negative, reduces to (8) when the predictee variable is univariate, and the regression matrix coefficients that render the residuals uncorrelated with the regressors also minimize the total variance (this is just the “ordinary least squares” procedure, minimizing mean square error). Nonetheless, an alternative originally proposed by Geweke [5] uses instead the *generalized variance* $|\Sigma(\varepsilon_t)|$, which quantifies the *volume* in which the residuals lie. This leads to the measure

$$\begin{aligned}\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} &\equiv \ln \left(\frac{|\Sigma(\varepsilon_t)|}{|\Sigma(\varepsilon'_t)|} \right) \\ &= \ln \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Z}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-)|} \right).\end{aligned}\tag{10}$$

⁷Note that even though X and Y are univariate, the *lagged* variables \mathbf{X}^- and \mathbf{Y}^- will generally be multivariate (at least if $p, q > 1$); hence they are written in bold type.

⁸We remark that for significance testing of G-causality it is quite common to use the appropriate F -statistic for the regressions (7) rather than $\mathcal{F}_{Y \rightarrow X | \mathbf{Z}}$ itself [8, 17]; the quantities are in any case related by a monotonic transformation.

Like trvMVGC, this measure is always non-negative, reduces to (8) when the predictee variable is univariate, and is consistent with the autoregressive approach inasmuch as the Yule-Walker regression matrix coefficients minimize the generalized variance, $|\Sigma(\epsilon)|$, as well as the total variance, (see Appendix A for a proof). Geweke [5] lists a number of motivations for taking $\mathcal{F}_{Y \rightarrow X|Z}$ as given in Eq. (10) as the natural extension of G-causality to the multivariate case. These include: (i) that the generalized variance version (10) is invariant under (linear) transformation of variables (see Section 4.2); and (ii) that the maximum likelihood estimator of this quantity, $\hat{\mathcal{F}}_{Y \rightarrow X|Z}$, is asymptotically χ^2 -distributed for large samples. In the following section we further justify this choice. Since we advocate the use of Geweke’s measure (10) of multivariate G-causality we abbreviate this simply as MVGC henceforth.

As remarked previously, the expression (10) defines *conditional* MVGC. Geweke [18] gives the following intuitively appealing expression for $\mathcal{F}_{Y \rightarrow X|Z}$ in terms of unconditional MVGCs:

$$\mathcal{F}_{Y \rightarrow X|Z} \equiv \mathcal{F}_{Y \oplus Z \rightarrow X} - \mathcal{F}_{Z \rightarrow X}; \quad (11)$$

that is, the extent to which Y and Z together cause X less the extent that Z on its own causes X . Note that this identity also holds for trvMVGC.

4 Properties of Multivariate Granger causality

In the following subsections we discuss some properties of MVGC and further motivate Geweke’s definition of this measure.

4.1 Gaussian Equivalence with Transfer Entropy

When all variables are Gaussian distributed, the MVGC $\mathcal{F}_{Y \rightarrow X|Z}$ is fully equivalent to the transfer entropy $\mathcal{T}_{Y \rightarrow X|Z}$, an information-theoretic notion of causality [13], with a simple factor of 2 relating the two quantities,

$$\mathcal{F}_{Y \rightarrow X|Z} = 2 \mathcal{T}_{Y \rightarrow X|Z}. \quad (12)$$

Transfer entropy [3, 19] is defined by the difference in entropies

$$\mathcal{T}_{Y \rightarrow X|Z} \equiv H(X | X^- \oplus Z^-) - H(X | X^- \oplus Y^- \oplus Z^-), \quad (13)$$

and quantifies the degree to which knowledge of the past of Y reduces uncertainty in the future of X . The equivalence (12) stems from the entropy of a Gaussian distribution being directly proportional to the logarithm of the determinant of its covariance matrix; and, furthermore, from any conditional entropy involving Gaussian variables being directly proportional to the logarithm of the determinant of the appropriate corresponding *partial* covariance matrix (see [13] for details). Due to the use of the determinant being crucial for this relationship, for trvMVGC the equivalence holds only in the more restricted situation when the predictee variable is univariate.

In addition to motivating MVGC over trvMVGC, the equivalence (12) also provides a justification for the use of linear regression models in measuring causality. Transfer entropy is naturally sensitive to nonlinearities in the data, a property which is rightly seen as desirable for measures of causality and which has motivated the development of several nonlinear extensions to standard G-causality [20, 21]. However, when data are Gaussian, the two linear regressions capture all of the entropy difference that

defines transfer entropy, which implies that non-linear extensions to G-causality are of no additional utility. Indeed for two multivariate Gaussian variables \mathbf{X} and \mathbf{Y} , the partial covariance $\Sigma(\mathbf{X}|\mathbf{Y})$, which is the same quantity as the residual covariance under linear regression, can be simply thought of as the conditional covariance of \mathbf{X} given \mathbf{Y} , because $\text{cov}(\mathbf{X}|\mathbf{Y} = \mathbf{y}) = \Sigma(\mathbf{X}|\mathbf{Y})$ for all \mathbf{y} . Hence, for Gaussian data, linear regression accounts for all the dependence of the regressee on the regressor.

To demonstrate formally that a stationary Gaussian AR process must be *linear*, consider a general stationary multivariate Gaussian process \mathbf{X}_t satisfying

$$\mathbf{X}_t = f(\mathbf{X}_{t-1}^{(p)}) + \boldsymbol{\varepsilon}_t, \quad (14)$$

where $f(\cdot)$ is some sufficiently well-behaved, possibly nonlinear function and the $\boldsymbol{\varepsilon}_t$ are independent of \mathbf{X}_{t-s} for $s = 1, 2, \dots$. For any t then, $\boldsymbol{\varepsilon}_t = \mathbf{X}_t - f(\mathbf{X}_{t-1}^{(p)})$ is independent of $\mathbf{X}_{t-1}^{(p)}$, so that, in particular, for any value $\boldsymbol{\xi}$ taken by $\mathbf{X}_{t-1}^{(p)}$, the conditional expectation

$$\mathbf{E}(\boldsymbol{\varepsilon}_t | \mathbf{X}_{t-1}^{(p)} = \boldsymbol{\xi}) = \mathbf{E}(\mathbf{X}_t | \mathbf{X}_{t-1}^{(p)} = \boldsymbol{\xi}) - f(\boldsymbol{\xi}) \quad (15)$$

does not depend on $\boldsymbol{\xi}$ and nor, by stationarity, on t . But since by assumption \mathbf{X}_t and $\mathbf{X}_{t-1}^{(p)}$ are jointly multivariate Gaussian, by a well-known result $\mathbf{E}(\mathbf{X}_t | \mathbf{X}_{t-1}^{(p)})$ depends linearly on $\boldsymbol{\xi}$, and from (15) it follows that $f(\boldsymbol{\xi})$ must be a linear function of $\boldsymbol{\xi}$.

4.2 Invariance under transformation of variables

The partial covariance $\Sigma(\mathbf{X}|\mathbf{Y})$ transforms in a simple way under linear transformation of variables. If T and U are respective matrices for linear transformations on \mathbf{X} and \mathbf{Y} then we have that

$$\Sigma(T \cdot \mathbf{X} | U \cdot \mathbf{Y}) \equiv T \Sigma(\mathbf{X} | \mathbf{Y}) T^\top. \quad (16)$$

Using this formula, and the properties of the determinant and trace operators, we can find the respective groups of linear transformations under which MVGC and trvMVGC are invariant. For MVGC, we find that the most general transformation that $\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}}$ is invariant under is given by

$$\begin{aligned} \mathbf{X} &\rightarrow T_{xx} \cdot \mathbf{X}, \\ \mathbf{Y} &\rightarrow T_{yx} \cdot \mathbf{X} + T_{yy} \cdot \mathbf{Y} + T_{yz} \cdot \mathbf{Z}, \\ \mathbf{Z} &\rightarrow T_{zx} \cdot \mathbf{X} + T_{zz} \cdot \mathbf{Z}, \end{aligned} \quad (17)$$

where the matrices T_{xx} , T_{yy} and T_{zz} on the diagonal are non-singular. All these symmetries are desirable properties for a causality measure. There ought to be invariance under redefinition of the individual variables within each of \mathbf{X} , \mathbf{Y} and \mathbf{Z} , (i.e. under the diagonal components T_{xx} , T_{yy} and T_{zz} of Eq. (17)), because MVGC is designed to measure causality between unified wholes rather than between arbitrarily defined constituent elements. The “off-diagonal” components T_{yx} , T_{yz} and T_{zx} are also intuitive. Adding components of \mathbf{Z} or \mathbf{X} to the predictor \mathbf{Y} should not change the value of MVGC, because MVGC is designed to measure the ability of \mathbf{Y} at predicting \mathbf{X} *over and above* \mathbf{Z} and \mathbf{X} . Similarly, adding components of \mathbf{X} onto \mathbf{Z} should not make a difference because the predictee \mathbf{X} could already be thought of as a conditional variable before transformation.

trvMVGC has an invariance under a similar group of transformations but with one significant restriction, namely that the matrix T_{xx} must be *conformal* (angle-preserving), that is T_{xx} must satisfy $T_{xx}T_{xx}^\top = cI$ for some constant c . This difference can have practical consequences. The broader invariance of MVGC (under *all* linear transformations T_{xx}) means that this measure, but not trvMVGC, is insensitive to certain common inaccuracies of data collection, namely those in which variables within a given set \mathbf{X} are contaminated by contributions from other variables (see Discussion). To put this point another way, if one wishes to infer MVGC between *hidden* variables by analyzing MVGC between observed variables, these two quantities are actually the same if the relationship between hidden and observed variables is linear and can be written in the form given in Eq. (17). One may also wish to measure the MVGC from the independent components of the predictor to the independent components of the predictee. Again, the invariance properties of MVGC mean that one does not need to explicitly find these independent components; one can simply compute MVGC between observed components. These observations indicate that MVGC takes into account correlation between variables in a principled way. We see this explicitly in Section 4.3.

The restriction $T_{xx}T_{xx}^\top = cI$ for trvMVGC further implies that an uneven *rescaling* of the components of the predictee variable may change the value of $\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}}^{tr}$. This too has practical implications, namely that trvMVGC but not MVGC can be affected by magnitude differences in the components of \mathbf{X} , perhaps resulting from these components reflecting underlying mechanisms that are differently amplified or differentially accessible to the measuring equipment, a common situation in many neuroscience contexts (see Discussion). This sensitivity is undesirable because causal connectivity should be based on the information content of signals (c.f. Section 4.1), and not on their respective magnitudes.

It is worth noting that for transfer entropy the symmetry group can be extended to include all non-singular (not necessarily linear) transformations of the predictee variable, since the entropies are invariant under such transformations.⁹ Since G-causality is essentially a linear version of transfer entropy, the former should at least be invariant under the linear subgroup of transformations.

4.3 Expansion of Multivariate Granger Causality

MVGC is expandable as a sum of G-causalities over all combinations of *univariate* predictor and predictee variables contained within the multivariate composites. The existence of this expansion depends on the fact that determinants are decomposable into products, and that logarithms of products are decomposable into sums of logarithms. No such decomposition exists for the logarithm of a trace, and so there is no obvious way of expanding trvMVGC into combinations of univariate components.

The expansion of MVGC is not entirely straightforward because different terms in the sum involve conditioning on the past and present of different subsets of variables. However each predictor/predictee combination appears precisely once in the sum, and each term can be explained intuitively. The general formula may be written as

$$\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} = \sum_{i=1}^n \sum_{\alpha=1}^m \mathcal{F}_{Y_\alpha \rightarrow X_i | \mathbf{Z} \oplus \mathbf{X} \oplus \mathbf{Y}_1 \oplus \dots \oplus \mathbf{Y}_{\alpha-1} \oplus X_1^0 \oplus \dots \oplus X_{i-1}^0}, \quad (18)$$

⁹If the predictee variable has a continuous (multivariate) distribution, we note that the Jacobian determinants in the standard change-of-variables formula for entropy calculation cancel out.

where the superscript ‘⁰’ indicates conditioning on the present (in addition to the past) of the corresponding variables. Thus, in the term for causality from Y_α to X_i one conditions on (i) the past of the entire multivariate conditional variable \mathbf{Z} , (ii) the past of the entire multivariate predictee variable \mathbf{X} , (iii) the past of all predictor variables Y_β with $\beta < \alpha$ and (iv) the present of all predictee variables X_j with $j < i$. The derivation of the expansion (18) is given in Appendix B.

For the case of a multivariate predictor and a univariate predictee we have

$$\mathcal{F}_{\mathbf{Y} \rightarrow X} = \mathcal{F}_{Y_1 \rightarrow X} + \mathcal{F}_{Y_2 \rightarrow X | Y_1} + \mathcal{F}_{Y_3 \rightarrow X | Y_1 \oplus Y_2} + \cdots + \mathcal{F}_{Y_m \rightarrow X | Y_1 \oplus Y_2 \oplus \dots \oplus Y_{m-1}} . \quad (19)$$

This formula is consistent with the intuitive idea that the total degree to which the multivariate \mathbf{Y} helps predict the univariate X is: the degree to which Y_1 predicts X , plus the degree to which Y_2 helps predict X over and above the information already present in Y_1 , and so on.

For the case of a multivariate predictee and a univariate predictor we have

$$\mathcal{F}_{Y \rightarrow \mathbf{X}} = \mathcal{F}_{Y \rightarrow X_1 | \mathbf{X}} + \mathcal{F}_{Y \rightarrow X_2 | \mathbf{X} \oplus X_1^0} + \mathcal{F}_{Y \rightarrow X_3 | \mathbf{X} \oplus X_1^0 \oplus X_2^0} + \cdots + \mathcal{F}_{Y \rightarrow X_n | \mathbf{X} \oplus X_1^0 \oplus X_2^0 \oplus \dots \oplus X_{n-1}^0} . \quad (20)$$

This formula supports the intuition that the total degree to which the univariate Y helps predict the multivariate \mathbf{X} is: the degree to which the past of Y helps predict the current value of X_1 over and above the degree to which the past of the whole of \mathbf{X} predicts the current value of X_1 , plus the degree to which the past of Y helps predict the current value of X_2 over and above the degree to which the past of the whole of \mathbf{X} and the current value of X_1 predicts the current value of X_2 , and so on.

We remark on two implications of the expansion of MVGC. First, Ladroue and colleagues suggested that use of generalized residual variance for causal inference on high-dimensional data might suffer from problems of numerical stability. However, the expansion of MVGC into low-dimensional, univariate G-causalities suggests that there should be no problem (see Section 4.3.1 for numerical evidence of this). Second, the expansion (18) indicates that MVGC controls for, to some extent, the influence of unmeasured latent/exogenous variables (see also Section 5). By conditioning on the present of certain appropriate predictee variables for each term of the expansion, only the effects of each predictor on independent components of the predictees enter the equation. This property stems from the fact that the determinant of the residual covariance matrix reflects not just residual variances, but also the extent to which these residual variances are independent of each other. This is another advantage of the MVGC measure over trvMVGC, which does not depend on residual correlations.

4.3.1 Stability of Multivariate Granger Causality

We tested numerically our claim (Section 4.3) that MVGC should not be less stable than trvMVGC. We studied MVAR(1) processes whose dynamics are given by

$$\mathbf{X}_t = A \cdot \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t , \quad (21)$$

where \mathbf{X} contains 8 variables, the sum of each row of A (i.e. total afferent to each element) is 0.5, all components in a given row of A are equal and positive, and each component of $\boldsymbol{\varepsilon}_t$ is an independent Gaussian random variable of mean 0 and variance 1. We generated 30 random “connectivity” matrices (or systems) A_i , ($i = 1, \dots, 30$), each with an average of 2 non-zero components per row. For each A_i

we obtained 10 sets of 3000 (post equilibrium) data points via Eq. (21). For each set, we computed the MVGC across each bipartition of the system corresponding to A_i . We then calculated, for each bipartition, the standard deviation of the MVGC across the 10 data sets and (excluding bipartitions with standard deviation less than 0.01) the corresponding coefficient of variation (CoV, standard deviation divided by mean). This procedure allowed us to obtain, for each A_i , a maximum CoV. Figure 1(a) shows that the maximum CoV is generally very small and never large, confirming the stability of MVGC.

To compare the stability of MVGC with that of trvMVGC, for each A_i and for each bipartition we divided the CoV for MVGC by the CoV for trvMVGC. Figure 1(b) shows the distribution of the average of this ratio across all bipartitions. The clustering of this distribution at ≈ 1 , with no outliers, confirms that MVGC and trvMVGC have similar stability properties, at least in the systems we have simulated.

To generalize these results we next used a genetic algorithm (GA) [22, 23] to see if we could find a network for which MVGC becomes unstable. The GA was initialized using a population composed of the 30 random systems A_i described above. We ran the GA for 130 generations. In each generation, we computed the fitness of each system as the maximum CoV of MVGC. Systems were selected to proceed to subsequent generations using stochastic rank-based selection. Mutations enabled the adding of new non-zero components to A_i , the removal of existing non-zero components, or the swapping of components, followed by renormalization of each row to sum to 0.5 again; two mutations were applied per system. After 130 generations (sufficient for fitness to asymptote) the average fitness (i.e. maximum CoV) in the population was ≈ 0.25 , and the maximum was 0.39, which is still a low value. For the A_i that gave this highest value, we compared the CoV obtained using MVGC with that obtained using trvMVGC following the procedure described above. The average ratio (across all bipartitions) was ≈ 1.00 , (maximum value 1.12), indicating that MVGC and trvMVGC had similar stability properties even for systems optimized to be unstable with respect to MVGC. Further, we examined some A_i for which the sums of the rows differed (i.e. having heterogeneous afferent connectivity); these systems had similar stability properties to those described above. Finally, stability properties were unaffected when computations were based on 1000 (rather than 3000) data points.

Taken together, these simulation results confirm that MVGC is numerically stable, and is not appreciably different from trvMVGC in terms of stability properties.

4.4 Spectral decomposition

In this section we review the spectral decomposition of G-causality [5, 1]. For simplicity we limit ourselves to the unconditional case, although the procedure may be readily extended to the conditional case (as described in e.g. Refs. [18, 1, 24]). We assume multivariate predictor and predictee variables, and show that MVGC but not trvMVGC has a satisfactory spectral decomposition.

Consider the stationary MVAR

$$\mathbf{X}_t = A \cdot \mathbf{X}_{t-1}^{(p)} + \boldsymbol{\varepsilon}_t = \sum_{k=1}^p A_k \cdot \mathbf{X}_{t-k} + \boldsymbol{\varepsilon}_t. \quad (22)$$

We may write this as

$$A(L) \cdot \mathbf{X}_t = \boldsymbol{\varepsilon}_t, \quad (23)$$

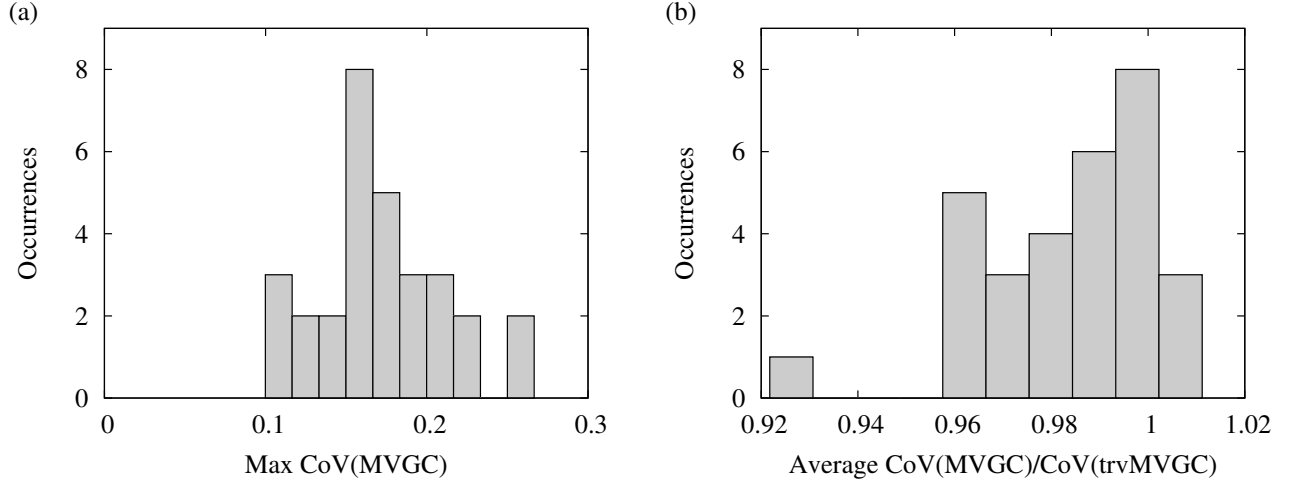


Figure 1: Stability of MVGC. (a) Histogram of the maximum CoV of MVGC, observed over 10 trials of 3000 time-steps, for each of 30 different systems, as described in Section 4.3.1. (b) Histogram of the average ratio between the CoV of MVGC and the CoV of trvMVGC, for each of the 30 systems. MVGC is numerically stable (a) and is not appreciably different from trvMVGC in terms of stability properties (b).

where L denotes the (single time step) lag operator, and

$$A(L) \equiv I - \sum_{k=1}^p A_k L^k. \quad (24)$$

Eq. (23) may be solved as

$$\mathbf{X}_t = H(L) \cdot \boldsymbol{\varepsilon}_t, \quad (25)$$

where $H(L) \equiv A(L)^{-1}$. Transforming into the frequency domain via the discrete-time Fourier transform $\mathbf{X}(\lambda) = \sum_{t=-\infty}^{\infty} \mathbf{X}_t e^{-i\lambda t}$ yields $A(\lambda) \cdot \mathbf{X}(\lambda) = \boldsymbol{\varepsilon}(\lambda)$ (replace L by $e^{-i\lambda}$), so that

$$\mathbf{X}(\lambda) = H(\lambda) \cdot \boldsymbol{\varepsilon}(\lambda), \quad (26)$$

where $H(\lambda) \equiv A(\lambda)^{-1}$ is the *transfer matrix*. The (power) *spectral density* of \mathbf{X} is then given by

$$S(\lambda) = H(\lambda) \Sigma(\boldsymbol{\varepsilon}) H^*(\lambda). \quad (27)$$

From a standard result [25], since $H(L)$ is a square matrix lag operator with the identity matrix as leading term, we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |H(\lambda) H^*(\lambda)| d\lambda = 0, \quad (28)$$

provided that all roots of the characteristic polynomial $|A(L)|$ lie outside the unit circle, which is a necessary condition for the existence of the stationary process (22). From (27) we may then derive the relation [26]

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |S(\lambda)| d\lambda = \ln |\Sigma(\boldsymbol{\varepsilon})|. \quad (29)$$

Consider now the stationary MVAR

$$\mathbf{X}_t \oplus \mathbf{Y}_t = A \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Y}_{t-1}^{(q)} \right) + \boldsymbol{\varepsilon}_{x,t} \oplus \boldsymbol{\varepsilon}_{y,t} \quad (30)$$

with coefficients matrix

$$A \equiv \begin{pmatrix} A_{xx} & A_{xy} \\ A_{yx} & A_{yy} \end{pmatrix} \quad (31)$$

and residuals covariance matrix

$$\Sigma(\boldsymbol{\varepsilon}_x \oplus \boldsymbol{\varepsilon}_y) \equiv \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}. \quad (32)$$

Let us split the corresponding transfer matrix $H(\lambda)$ as

$$H(\lambda) \equiv A(\lambda)^{-1} = \begin{pmatrix} H_{xx}(\lambda) & H_{xy}(\lambda) \\ H_{yx}(\lambda) & H_{yy}(\lambda) \end{pmatrix} \quad (33)$$

and the spectral density as

$$S(\lambda) = \begin{pmatrix} S_{xx}(\lambda) & S_{xy}(\lambda) \\ S_{yx}(\lambda) & S_{yy}(\lambda) \end{pmatrix}. \quad (34)$$

Then $S_{xx}(\lambda)$ is just the spectral density of \mathbf{X} , which from (27) is given by

$$S_{xx}(\lambda) = H_{xx}(\lambda)\Sigma_{xx}H_{xx}^*(\lambda) + 2\Re\{H_{xx}(\lambda)\Sigma_{xy}H_{xy}^*(\lambda)\} + H_{xy}(\lambda)\Sigma_{yy}H_{xy}^*(\lambda). \quad (35)$$

The idea is that we wish to decompose this expression into a part reflecting the effect of \mathbf{X} itself and a part reflecting the causal influence of \mathbf{Y} . The problem is that, due to the presence of the “cross” term, $S_{xx}(\lambda)$ does not split cleanly into an \mathbf{X} and a \mathbf{Y} part. Geweke [5] addresses this issue by introducing the transformation

$$\mathbf{X} \oplus \mathbf{Y} \rightarrow U \cdot (\mathbf{X} \oplus \mathbf{Y}), \quad (36)$$

where

$$U \equiv \begin{pmatrix} I & 0 \\ -\Sigma_{yx}\Sigma_{xx}^{-1} & I \end{pmatrix}. \quad (37)$$

Note that this transformation leaves the G-causality $\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X}}$ invariant (c.f. Section 4.2) and, for the transformed regression, we have $\Sigma_{xy} \equiv 0$; that is, the residuals $\boldsymbol{\varepsilon}_x, \boldsymbol{\varepsilon}_y$ are uncorrelated. Thus, assuming the transformation (37) has been pre-applied, Eq. (35) becomes

$$S_{xx}(\lambda) = H_{xx}(\lambda)\Sigma_{xx}H_{xx}^*(\lambda) + H_{xy}(\lambda)\Sigma_{yy}H_{xy}^*(\lambda), \quad (38)$$

whereby the spectral density of \mathbf{X} splits into an “intrinsic” part and a “causal” part. The spectral G-causality of $\mathbf{Y} \rightarrow \mathbf{X}$ at frequency λ is now defined to be

$$f_{\mathbf{Y} \rightarrow \mathbf{X}}(\lambda) \equiv \ln \left(\frac{|S_{xx}(\lambda)|}{|H_{xx}(\lambda)\Sigma_{xx}H_{xx}^*(\lambda)|} \right) \quad (39)$$

or, in terms of the *untransformed* variables,

$$f_{\mathbf{Y} \rightarrow \mathbf{X}}(\lambda) \equiv \ln \left(\frac{|S_{xx}(\lambda)|}{|S_{xx}(\lambda) - H_{xy}(\lambda)\Sigma_{y|x}H_{xy}^*(\lambda)|} \right), \quad (40)$$

with $S_{xx}(\lambda)$ as in (35) and $\Sigma_{y|x} \equiv \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.

Geweke (Ref. [5], Theorem 2) then establishes the fundamental motivating relationship between frequency and time domain G-causality:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y \rightarrow X}(\lambda) d\lambda = \mathcal{F}_{Y \rightarrow X}, \quad (41)$$

provided that all roots of $|A_{yy}(L)|$ lie outside the unit circle.¹⁰ The proof of this relation relies crucially on the result (28) which, we note, involves the *determinant* of the transfer matrix. Thus if the trace, rather than the determinant, were to be used in the definition (39) for $f_{Y \rightarrow X}(\lambda)$ then we could not expect to obtain a relation corresponding to (41), since (i) the trace of the spectral density in Eq. (27) does not factorize, (ii) there is no trace analogue to Eq. (28), and thus (iii) no analogue to Eq. (29). This would seem to preclude a satisfactory spectral decomposition for the trace version of G-causality. Similar remarks apply to conditional G-causality in the spectral domain.

In Ref. [4], however, it is conjectured that a trace analogue of Eq. (41) does indeed hold. To test this conjecture we performed the following experiment: we simulated 1000 MVAR(1) processes of the form

$$\mathbf{X}_t \oplus Y_t = A \cdot (\mathbf{X}_{t-1} \oplus Y_{t-1}) + \boldsymbol{\varepsilon}_{x,t} \oplus \varepsilon_{y,t}, \quad (42)$$

where \mathbf{X} has dimension 2 and Y dimension 1. Residuals $\boldsymbol{\varepsilon}_{x,t}, \varepsilon_{y,t}$ were completely uncorrelated, with unit variance (i.e. $\Sigma(\boldsymbol{\varepsilon}_{x,t} \oplus \varepsilon_{y,t})$ was the 3×3 identity matrix) so that, in particular, the Geweke transformation (37) was unnecessary. For each trial the 3×3 coefficients matrix A was chosen at random with elements uniform on $[-\frac{1}{2}, \frac{1}{2}]$, and the process (42) simulated for 10^6 stationary time steps (the occasional unstable process was rejected). Time domain causalities $\mathcal{F}_{Y \rightarrow X}$, $\mathcal{F}_{Y \rightarrow X}^{tr}$ and frequency domain causalities $f_{Y \rightarrow X}(\lambda)$, $f_{Y \rightarrow X}^{tr}(\lambda)$ were calculated in sample using $p = 10$ lags. (As noted previously,¹⁰ equality in (41) is only assured in the limit of infinite lags; 10 lags was found empirically to achieve good accuracy without overfitting the data.) Relative errors of integrated spectral MVGC with respect to time-domain MVGC, expressed as a percentage, were defined as

$$\begin{aligned} E_{\%} &\equiv 100 \times \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y \rightarrow X}(\lambda) d\lambda - \mathcal{F}_{Y \rightarrow X}}{\mathcal{F}_{Y \rightarrow X}}, \\ E_{\%}^{tr} &\equiv 100 \times \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y \rightarrow X}^{tr}(\lambda) d\lambda - \mathcal{F}_{Y \rightarrow X}^{tr}}{\mathcal{F}_{Y \rightarrow X}^{tr}}, \end{aligned} \quad (43)$$

for MVGC and trvMVGC respectively. (The integrals were computed by standard numerical quadrature.) Results, displayed in Table 1, confirm to good accuracy the theoretical prediction of Eq. (41) for MVGC (the small negative bias on $E_{\%}$ is due to the finite number of lags), while for trvMVGC relative errors are several orders of magnitude larger and furthermore were not decreased by choosing longer stationary sequences and/or more lags. The full distribution of relative errors is also displayed as a histogram in Fig. 2.

¹⁰A subtlety to note is that even if the MVAR (30) has a finite number of lags $p, q < \infty$, the exact *restricted* regression of \mathbf{X} on its own past will generally require an infinite number of lags [5]. Thus in theory, for exact equality in (41), an infinite number of lags is required to calculate the term $\Sigma(\mathbf{X} | \mathbf{X}^-)$ which appears in $\mathcal{F}_{Y \rightarrow X}$ (using a finite number of lags will generally result in an overestimate of $\mathcal{F}_{Y \rightarrow X}$, since residual errors will be larger than for the exact regression). As applied to empirical data, it is in any case good practice to choose “sufficient” lags for all regressions so as to model the data adequately without overfitting [27, 28].

error	mean	std. dev.	abs. mean
$E_{\%}$	-0.0004	0.0005	0.0005
$E_{\%}^{tr}$	-0.0488	10.5995	8.1799

Table 1: Comparison of relative errors of integrated spectral MVGC and trvMVGC with respect to time domain MVGC and trvMVGC, for a random sample of MVAR(1) processes. Top row shows MVGC, bottom row shows trvMVGC. See text for details. Figures in the “abs. mean” column are the means of the absolute values $|E_{\%}|$ and $|E_{\%}^{tr}|$.

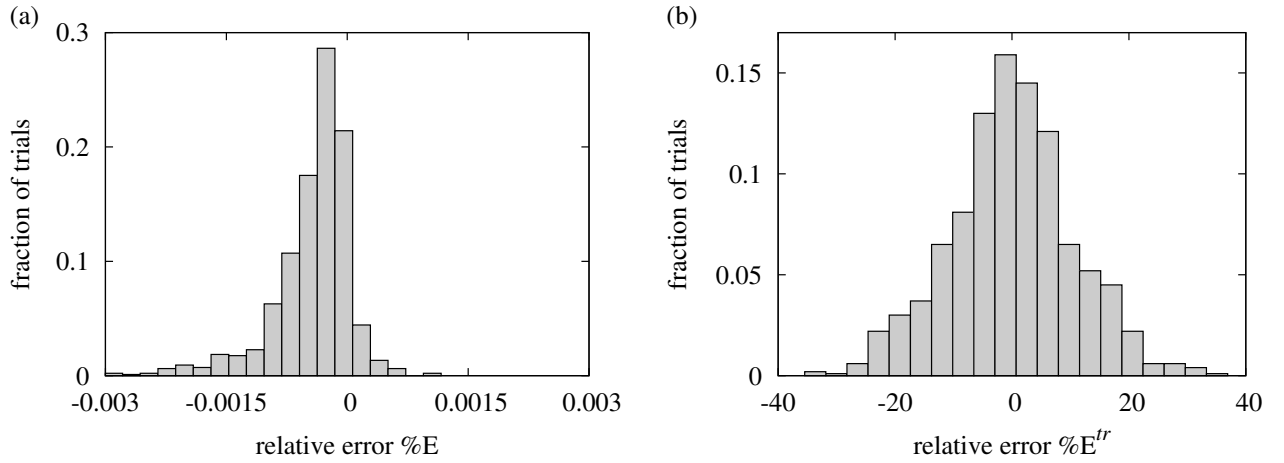


Figure 2: Distribution of relative errors of integrated spectral multivariate G-causality with respect to the time domain for (a) MVGC (b) trvMVGC, for a random sample of MVAR(1) processes.

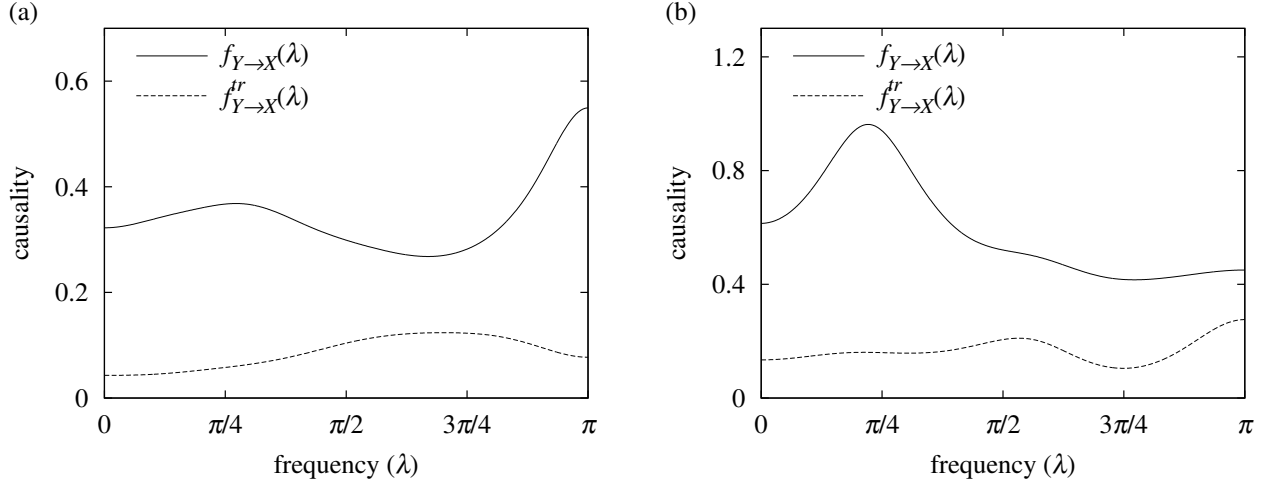


Figure 3: Comparison of MVGC and trvMVGC in the frequency domain: spectral MVGC and trvMVGC plotted against frequency for (a) a typical MVAR(3) process with $\dim(\mathbf{X}) = 2$, $\dim(Y) = 1$ and (b) a typical MVAR(5) process with $\dim(\mathbf{X}) = 3$, $\dim(\mathbf{Y}) = 2$.

We also repeated the experiment with higher order MVAR(p) processes, higher dimensional predictee and predictor variables and correlated residuals ε_x . In all cases, results confirmed the accuracy of (41) for MVGC and yielded large relative errors for trvMVGC. We remark that *qualitative* differences (i.e. aside from differences of scale) between spectral MVGC and trvMVGC could be substantial (Fig. 3). These differences, furthermore, appeared in general to be exaggerated by the presence of residual correlations; this is consonant with the sensitivity of MVGC as contrasted with the lack of sensitivity of trvMVGC to residual correlations (see Sections 4.3 and 5).

It is straightforward to show that $f_{\mathbf{Y} \rightarrow \mathbf{X}}(\lambda)$ is invariant under the same group of linear transformations (17) as $\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X}}$; again, $f_{\mathbf{Y} \rightarrow \mathbf{X}}^{tr}(\lambda)$ will in general be invariant only under the restricted group with T_{xx} conformal; this extends to the conditional case.

5 Multivariate partial Granger causality

Recently, a *partial G-causality* measure has been introduced [29] which exploits a parallel with the concept of *partial coherence* [30] in order to control for latent/exogenous influences on standard G-causality. Partial G-causality modifies the standard G-causality measure by including terms based on residual correlations between the predictee variable and the conditional variables. Consider, in addition to the regressions (7), the following regressions of the *conditioning* variable \mathbf{Z}_t :

$$\begin{aligned} \mathbf{Z}_t &= B \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Z}_{t-1}^{(r)} \right) + \boldsymbol{\eta}_t, \\ \mathbf{Z}_t &= B' \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Y}_{t-1}^{(q)} \oplus \mathbf{Z}_{t-1}^{(r)} \right) + \boldsymbol{\eta}'_t. \end{aligned} \tag{44}$$

Here the roles of the predictee and conditioning variables are reversed. Then for univariate predictor and predictee the partial G-causality of Y on X given \mathbf{Z} is defined by conditioning the respective

residual covariances for the regressions of X on the corresponding residuals for the regressions of Z :

$$\mathcal{F}_{Y \rightarrow X|Z}^P \equiv \ln \left(\frac{\Sigma(\varepsilon_t | \eta_t)}{\Sigma(\varepsilon'_t | \eta'_t)} \right) . \quad (45)$$

This extends naturally to the fully multivariate case (c.f. Eq. (10)), and we define partial MVGC (pMVGC) as

$$\mathcal{F}_{Y \rightarrow X|Z}^P \equiv \ln \left(\frac{|\Sigma(\varepsilon_t | \eta_t)|}{|\Sigma(\varepsilon'_t | \eta'_t)|} \right) \quad (46)$$

$$= \ln \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Z}^- \oplus \mathbf{Z})|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^- \oplus \mathbf{Z})|} \right) \quad (47)$$

where the RHS (47) follows from the identity (64) derived in Appendix C, (with $\mathbf{W} \equiv \mathbf{X}^- \oplus \mathbf{Z}^-$ and $\mathbf{W} \equiv \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-$ for the numerator and denominator terms respectively). Comparing with (10) we see thus that pMVGC differs from MVGC in the inclusion of the *present* of the conditioning variable Z in the respective regressions. Seen in this form, it is clear that, as is the case for MVGC, pMVGC is always non-negative.¹¹ One could alternatively express pMVGC as (non-partial) MVGC conditioned on a “forward lagged” version of Z : defining $\tilde{Z}_t \equiv Z_{t+1}$ we have $Z_t \oplus Z_{t-1}^{(r)} \equiv \tilde{Z}_{t-1}^{(r+1)}$, or $\tilde{Z}^- = Z \oplus Z^-$ (note the additional lag on \tilde{Z}^-), so that, from Eq. (47),

$$\mathcal{F}_{Y \rightarrow X|Z}^P = \mathcal{F}_{Y \rightarrow X|\tilde{Z}} . \quad (48)$$

As noted in Section 4.3, (non-partial) MVGC to some extent already controls for the influence of latent/exogenous variables because the generalized variance is sensitive to residual correlations. However, pMVGC takes into account even more correlations with the explicit aim of controlling for latent/exogenous influences. pMVGC may therefore be preferable when such influences are expected to be (a) strong and (b) relatively uniform in their influence on the measured system. Indeed, pMVGC (and the original measure of partial G-causality) can only be effective in compensating for latent/exogenous variables that affect *all* modeled variables (i.e. predictee, predictor and conditioning) to a roughly equal degree [29].

It is interesting to note that pMVGC may be expressed in terms of non-partial MVGCs as

$$\mathcal{F}_{Y \rightarrow X|Z}^P = \mathcal{F}_{Y \rightarrow Z \oplus X} - \mathcal{F}_{Y \rightarrow Z|X} . \quad (49)$$

by straightforward application of Eq. (3). As expected, (49) includes a term with a mandatory multivariate predictee, since it is only in this case that residual correlation can make a difference. It is interesting that Z appears as a *predictee* variable; this might be understood as pMVGC using the conditioning variable Z as a “proxy” by which to assess the influence of latent or exogenous variables.

A “trace” version of pMVGC may be defined analogously to (46). Again by Eq. (64) of Appendix C, the identity corresponding to (47) will hold, as will the trace analogue of (48). However, the

¹¹In [29] it is stated that partial G-causality may in some circumstances be *negative*; the justification for this is unclear.

analogue of (49) will *not* hold in general, since the traces of the partial covariance matrices will in general not factorize appropriately.¹²

From (48) it is straightforward to derive a spectral decomposition $f_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}}^P(\lambda)$ for pMVGC, which will integrate correctly to the time-domain pMVGC $\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}}^P$. Again, a spectral decomposition for the corresponding trace version is likely to be problematic, insofar as it will fail in general to integrate correctly to the time-domain value (c.f. Section 4.4).

6 Causal density

A straightforward application of MVGC is to measures of *causal density*, the overall level of causal interactivity sustained by a multivariate system \mathbf{X} . A previous measure of causal density [22] has been defined as the average of all pairwise (and hence univariate) G-causalities between system elements, conditioned on the remaining system elements.¹³

$$\text{cd}(\mathbf{X}) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{F}_{X_i \rightarrow X_j | \mathbf{X}_{[ij]}} \quad (50)$$

where $\mathbf{X}_{[ij]}$ denotes the subsystem of \mathbf{X} with variables X_i and X_j omitted, and n is the total number of variables. Causal density provides a useful measure of the dynamical “complexity” of a system inasmuch as elements that are completely independent will have zero causal density, as will elements that are completely integrated in their dynamics. Exemplifying standard intuitions about complexity [31], high causal density will only be achieved when elements behave somewhat differently from each other, in order to contribute novel potential predictive information, and at the same time are globally integrated, so that the potential predictive information is in fact useful [32, 33].

Using MVGC, various extensions to (50) can be suggested, based on the various possible interactions between multivariate predictors, predictees and conditional variables. These extensions may provide a more principled measure of complexity by analyzing a target system at multiple scales. First we define the causal density from size k to size r , $\text{cd}_{k \rightarrow r}(\mathbf{X})$, as the average MVGC from a subset of size k to a subset of size r , conditioned on the rest of the system:

$$\text{cd}_{k \rightarrow r}(\mathbf{X}) = \frac{1}{n_{k,r}} \sum_{i=1}^{n_{k,r}} \mathcal{F}_{\mathbf{V}_i^k \rightarrow \mathbf{U}_i^r | \mathbf{W}_i^{n-k-r}}, \quad (51)$$

where $\mathbf{X} = \mathbf{V}_i^k \cup \mathbf{U}_i^r \cup \mathbf{W}_i^{n-k-r}$ denotes the i^{th} of the $n_{k,r} \equiv \binom{n}{k} \binom{n-k}{r}$ distinct tripartitions of \mathbf{X} into disjoint sub-systems of respective sizes k , r and $(n-k-r)$. Then using this, one could define the *bipartition causal density* (bcd) as the average of $\text{cd}_{k \rightarrow (n-k)}(\mathbf{X})$ over predictor size k ,

$$\text{bcd}(\mathbf{X}) = \frac{1}{n-1} \sum_{k=1}^{n-1} \text{cd}_{k \rightarrow (n-k)}(\mathbf{X}). \quad (52)$$

¹²In [4], under the section headed “Partial Complex Granger causality”, the quantity developed appears to be (the trace version of) what is conventionally referred to as *conditional* G-causality, rather than partial G-causality as introduced in [29] and referenced in this section.

¹³This is the “weighted” version of causal density. An unweighted and [0,1] bounded alternative can be defined as the fraction of all pairwise conditional causalities that are statistically significant at a given significance level.

Interestingly, this quantity is closely related to the popular Tononi-Sporns-Edelman “neural complexity” measure [34] which averages (contemporaneous) *mutual information* across bipartitions; (we are currently exploring this in work in preparation). It could also be interesting to compare causal density at different scales of predictor plus predictee size; thus we define

$$\text{cd}_s(\mathbf{X}) \equiv \frac{1}{s-1} \sum_{k=1}^{s-1} \text{cd}_{k \rightarrow (s-k)}(\mathbf{X}) . \quad (53)$$

Then the original causal density measure of Eq. (50) is just cd_2 and bcd is cd_n . The average of this over all scales can be used to define a complete *tripartition causal density* (tcd):

$$\text{tcd}(\mathbf{X}) \equiv \frac{1}{n-1} \sum_{s=2}^n \text{cd}_s(\mathbf{X}) . \quad (54)$$

A comparison of the properties of all versions of causal density, as well as related complexity measures, is in progress. We remark that it is straightforward to define spectral versions of these causal density measures.

7 Autonomy in complex systems

G-causality has recently been adapted to provide an operational measure of “autonomy” in complex systems [11]. A variable \mathbf{X} can be said to be “G-autonomous” with respect to a (multivariate) set of external variables \mathbf{Z} if its own past states help predict its future states over and above predictions based on \mathbf{Z} . This definition rests on the intuition of autonomy as “self determination” or “self causation”. We can formalize this notion along the lines of MVGC as follows. Consider the regressions

$$\begin{aligned} \mathbf{X}_t &= A \cdot \mathbf{Z}_{t-1}^{(r)} + \boldsymbol{\varepsilon}_t , \\ \mathbf{X}_t &= A' \cdot \left(\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Z}_{t-1}^{(r)} \right) + \boldsymbol{\varepsilon}'_t , \end{aligned} \quad (55)$$

which differ from Eqs. (7) primarily because the predictee variable \mathbf{X} is *not* regressed on itself in one of the equations. The G-autonomy of \mathbf{X} is then given by

$$\mathcal{A}_{\mathbf{X}|\mathbf{Z}} = \ln \left(\frac{|\Sigma(\boldsymbol{\varepsilon}_t)|}{|\Sigma(\boldsymbol{\varepsilon}'_t)|} \right) . \quad (56)$$

The extension of G-autonomy to the multivariate case is important because it accommodates situations in which groups of elements may be jointly autonomous (self-determining, self-causing), even though the activity of individual elements within the group may be adequately predicted by combinations of activities of other elements in the group. Univariate formulations of G-autonomy [11] would fail in these cases. Consider as a trivial example an element X_1 which is G-autonomous with respect to a background \mathbf{Z} . If X_1 is now duplicated by the element X_2 it will no longer appear as G-autonomous within the multivariate system $X_1 \oplus X_2 \oplus \mathbf{Z}$. However, the multivariate variable $X_1 \oplus X_2$ will be (jointly) G-autonomous with respect to \mathbf{Z} .

As discussed in [11] G-autonomy also provides the basis for a notion of “G-emergence” as applied to the relation between *macroscopic* variables “emerging” from the activity of *microscopic* constituents. G-emergence operationalizes the intuition that a macro-level variable is emergent to the extent that it is simultaneously *autonomous from* and *dependent upon* its micro-level constituents [11, 35]. Extension of G-emergence to the multivariate case using MVGC is straightforward, allowing consideration of multivariate micro- and macro-variables.

8 Macroscopic variables and causal independence

Given the ability to assess multivariate causal interactions, a second challenge arises: the identification of relevant groupings of variables into multivariate ensembles. One approach to this challenge adopts the perspective of statistical mechanics on the emergence of novel macroscopic variables, given a microscopic description of a system [36, 37]. Here, we suggest that MVGC may furnish a useful method for macro-variable identification in this context. Let us assume that \mathbf{Z}_t represents a set of microscopic variables defining a complex (possibly stochastic) dynamical system, and $\mathbf{X}_t \equiv f(\mathbf{Z}_t)$ a set of macroscopic variables functionally (possibly deterministically) dependent on the microscopic variables. There is then a sense in which \mathbf{X} represents a “parsimonious” high-level description of the system, to the extent that it predicts its own dynamical evolution without recourse to the low level of description of the system represented by \mathbf{Z} ; that is, to the extent that \mathbf{X} exhibits strong *causal independence* with respect to \mathbf{Z} . In this view, $\mathcal{F}_{\mathbf{Z} \rightarrow \mathbf{X}}$ furnishes a natural measure of the *lack* of this causal independence, which might then be used to *identify* parsimonious macroscopic variables by minimizing $\mathcal{F}_{\mathbf{Z} \rightarrow f(\mathbf{Z})}$ over candidate functions $f(\cdot)$. The multivariate formulation MVGC would appear to be significant in this context for reasons similar to the G-autonomy case. Specifically, it may be that a set of macroscopic variables \mathbf{X} may *jointly* have high causal independence with respect to the microscopic variables \mathbf{Z} , while the component variables X_i may individually have lower causal independence.

The notions of G-autonomy, G-emergence, and causal independence are distinct but related. In short, G-autonomy measures “self-causation”, causal independence measures the *absence* of useful predictive information between microscopic and macroscopic descriptions of a system, and G-emergence measures a combination of macro-level autonomy and micro-to-macro causal *dependence*. It is possible, and is left as an objective of future work, that all three measures could be applied usefully to systems that avail multiple levels of descriptions, (i) to identify relevant groupings of observables at each level, (ii) to decompose causal interactions within each level, and finally (iii) to quantitatively characterize inter-level relationships.

9 Discussion

We have described and motivated a measure of multivariate causal interaction that is a natural extension of the standard G-causality measure. The measure, originally introduced by Geweke [5] but almost totally overlooked since, uses the generalized variance (the determinant of the residual covariance matrix) and we have termed it *multivariate G-causality* (MVGC). It contrasts with another recent proposal [4] for addressing the same problem which uses instead the total variance (the trace of the residual covariance matrix). In this paper, we have presented several theoretical justifications,

augmented by numerical modeling, for preferring MVGC over the trace version, which we summarize below. We have also extended MVGC to address novel challenges in the analysis of complex dynamical systems, including quantitative characterization of “causal density”, “autonomy”, and the identification of novel macroscopic variables via causal independence.

9.1 Importance of multivariate causal analysis

In many analyses of complex systems, particularly in neuroscience and biology, there may be no simple or principled relationship between observed variables and explanatorily relevant collections, or ensembles, of these variables. In the Introduction we already remarked on fMRI, where explanatorily relevant ROIs are each composed of multiple observables (voxels) which are arbitrarily demarcated with respect to underlying neural mechanisms. Other non-invasive neuroimaging methods share similar varieties of arbitrariness: both electroencephalography (EEG) and magnetoencephalography (MEG) provide signals which are complex convolutions of underlying neural sources. In these and similar cases, multivariate causal analysis, and MVGC in particular, can be used to aggregate univariate observables into meaningful multivariate (ensemble) variables. It bears emphasizing that MVGC is fundamentally different from conditional G-causality [38], which assesses the causal connectivity between two univariate variables, conditioned on a set of other variables.

Even when it is possible to measure directly the activity of variables of interest, it is still important to consider multivariate interactions. Continuing with the neuroscience example, it may be that multiple ROIs act *jointly* to influence other ROIs, or cognitive and/or behavioral outputs. In single cell recordings this point is even more pressing: since Hebb [39] it has been increasingly appreciated that neurons act as ensembles, rather than singly, in the adaptive function of the brain [40]. MVGC is well suited to disclosing causal relationships among these ensembles as a window onto underlying principles of brain operation.

Of course, the application of MVGC is not limited to neuroscience. Multivariate interactions are likely to be important in a very broad range of application areas. For example, genetic, metabolic, and transcriptional regulatory networks may be usefully decomposed into multivariate ensembles influencing other such ensembles [4]. Indeed, multivariate interactions may be important in any system, natural or artificial, which can be described in terms of multiple simultaneously acquired time series.

9.2 Generalized variance vs total variance

A different approach to multivariate causal analysis was recently proposed by Ladroue and colleagues [4]. This involved a measure (which we call *trvMVGC*) based on the trace of the residual covariance matrix (the total variance), rather than the determinant (the generalized variance). Geweke [5] provided the original justifications for the determinant form, but did not explicitly discuss the trace form. As noted in Section 3 of Ref. [5], Geweke’s motivations included (i) MVGC is invariant under (linear) transformations of variables, and (ii) the maximum likelihood estimator of MVGC is asymptotically χ^2 -distributed for large samples; (there is no standard test statistic for *trvMVGC*). In this paper we have substantially enhanced this list, in each case comparing MVGC explicitly with *trvMVGC*. In summary: (iii) MVGC is fully equivalent to transfer entropy under Gaussian assumptions, whereas for *trvMVGC* this equivalence only holds for the univariate case; (iv) MVGC is invariant under *all* (non-singular) linear transformations of the predictee variable, while *trvMVGC* is invariant only

under conformal linear transformations (see below); (v) only MVGC is expandable as a sum of univariate G-causalities; (vi) MVGC but not trvMVGC admits a satisfactory spectral decomposition, inasmuch as it guarantees a consistent relationship with the corresponding time-domain formulation; (vii) only MVGC depends on residual correlations, and through these accommodates in a natural way the influence of exogenous or latent variables, and (viii) the partial version of MVGC, pMVGC is decomposable in terms of non-partial MVGCs, but this is not true in general for trvMVGC.

All the above factors suggest that MVGC should be preferred to trvMVGC. Taken individually they may differ in their significance but taken together they emphasize that MVGC, but not trvMVGC, provides a comprehensive and *theoretically consistent* extension of standard G-causality to the multivariate case. While this consistency is the most important reason to prefer MVGC to trvMVGC, let us consider further three of the individual properties. First, the equivalence with transfer entropy is important because it justifies the use of linear modeling for multivariate causal analysis, at least where Gaussian assumptions are reasonable. Second, the broader range of invariance is important because it means that MVGC is robust to a wider range of common inaccuracies during data collection, in particular those in which univariate variables are contaminated by contributions from other variables and in which different components of multivariate ensembles are differently scaled by measurement constraints. It is likely that this additional robustness will have significant practical importance in many experimental applications, for example in EEG and MEG where individual sensors detect signals from multiple neural sources and may differentially amplify these sources according to their distance from the sensors and their alignment with the cortical surface. Finally, the lack of a satisfactory spectral version of trvMVGC, which we establish both theoretically and numerically (Section 4.4 and Figures 2 and 3), implies that frequency-domain results obtained using trvMVGC are unreliable, both in their magnitude and in their spectral profile.

Ladroue *et al.* [4] note Geweke’s form (i.e. MVGC) and suggest trvMVGC is preferable in view of possible numerical instabilities attending the computation of determinants for high-dimensional data. However the existence of an expansion of MVGC in terms of univariate G-causalities (18) seems to counter this claim, since the univariate causalities would not be expected to be unstable. Numerical simulations (Section 4.3.1 and Figure 1) confirm our view.

9.3 Quantities derived from MVGC

In the second part of the paper we used MVGC to derive several novel measures that have the potential to shed substantial new light on complex system dynamics.

First, MVGC leads immediately to a series of redefinitions of our previous “causal density” measure [22], which aims to capture the dynamical complexity of a system’s dynamics in terms of coexisting integration and differentiation. Extension to the multivariate case allows causal density to be evaluated at multiple levels of description thus furnishing a more principled measure of dynamical complexity. Causal density has been suggested as a measure of neural dynamics that captures certain aspects of consciousness [32]. It has been shown¹⁴ to increase in response to perceived stimuli as compared to non-perceived stimuli in a visual masking task [41], and it captures the complex dynamics of small-world networks more effectively than does a prominent competing measure, neural complexity [33]. Multivariate causal density has the potential to further strengthen and generalize these contributions.

¹⁴In approximation.

Second, MVGC can be used to generalize the concept of G-autonomy, which operationalizes the notion of autonomy as “self causation” [11]. Multivariate G-autonomy is a significant enhancement because it deals with the case in which a group of variables may be jointly autonomous even though, individually, no variable is autonomous. Our results therefore pave the way to informative application of this measure to complex systems.

Third, MVGC can be helpful in considering relations between microscopic and macroscopic levels of description of a system. One approach is to consider how *causally independent* a macroscopic variable is, with respect to its set of constituent micro-variables. We have suggested that this notion can be used to identify parsimonious macro-variables by maximizing causal independence over a space of functions relating micro- and macro-variables. Alternatively, the concept of G-emergence operationalizes the idea that an emergent macro-variable is both *autonomous from* and *causally dependent* on its underlying micro-level constituents. Unlike the “causal independence” view, G-emergence may be better suited to characterizing the degree of emergence as opposed to identifying prospective macro-variables; G-emergence also explicitly measures micro-to-macro causal dependence rather than assuming that it is present.

Finally, the concepts of redundancy and synergy amongst variables have been recently introduced, via the use of a variant of the trvMVGC measure [42]. These quantities aim at detecting functionally relevant partitions of a system by grouping variables according to their summed causal influences. Because of the advantages of MVGC over trvMVGC, we suggest it may be useful to redefine redundancy and synergy in terms of MVGC.

9.4 Summary

Models of complex systems typically contain large numbers of variables. Having a measure for directed interactions between groups of variables, as opposed to just single variables, provides a useful tool for the analysis of such systems. We have demonstrated that MVGC is such a measure, and we have provided a series of justifications, theoretical and numerical, to prefer it over a related measure, trvMVGC. Like all measures of directed interaction based on G-causality, MVGC can be measured for freely collected data, without perturbing or providing inputs to the system. Finally, in contrast to alternative approaches such as structural equation modeling [43] or dynamic causal modeling [2], MVGC can be applied with very little prior knowledge of the system under consideration.

Acknowledgements

AKS is supported by EPSRC Leadership Fellowship EP/G007543/1, which also supports the work of ABB. Support is also gratefully acknowledged from the Dr. Mortimer and Theresa Sackler Foundation.

Appendix

A Minimizing the determinant of the residuals covariance matrix

We wish to show that minimizing the determinant $|\Sigma(\epsilon)|$, where $\epsilon = \mathbf{X} - \mathbf{A} \cdot \mathbf{Y}$ as specified in (4), leads to the same values (5) for the regression coefficients \mathbf{A} . We thus solve for \mathbf{A} in the simultaneous

equations

$$\frac{\partial |\Sigma(\boldsymbol{\varepsilon})|}{\partial A_{i\alpha}} = 0, \quad (57)$$

where i runs from $1 \dots n$, α from $1 \dots m$ and $\Sigma(\boldsymbol{\varepsilon})$ is given by

$$\Sigma(\boldsymbol{\varepsilon}) = \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y}) A^\top - A \Sigma(\mathbf{X}, \mathbf{Y})^\top + A \Sigma(\mathbf{Y}) A^\top. \quad (58)$$

We use the formula for an invertible square matrix B

$$\frac{\partial |B|}{\partial B_{jk}} = |B| (B^{-1})_{kj}. \quad (59)$$

Assuming $\Sigma(\boldsymbol{\varepsilon})$ invertible and setting $W \equiv |\Sigma(\boldsymbol{\varepsilon})| \Sigma(\boldsymbol{\varepsilon})^{-1}$ we have

$$\begin{aligned} \frac{\partial |\Sigma(\boldsymbol{\varepsilon})|}{\partial A_{i\alpha}} &= \sum_{j,k} \frac{\partial |\Sigma(\boldsymbol{\varepsilon})|}{\partial \Sigma(\boldsymbol{\varepsilon})_{jk}} \frac{\partial \Sigma(\boldsymbol{\varepsilon})_{jk}}{\partial A_{i\alpha}} \\ &= \sum_{j,k} W_{kj} \frac{\partial \Sigma(\boldsymbol{\varepsilon})_{jk}}{\partial A_{i\alpha}} \quad \text{from (59)} \\ &= \sum_{j,k} W_{kj} \frac{\partial}{\partial A_{i\alpha}} [\Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y}) A^\top - A \Sigma(\mathbf{X}, \mathbf{Y})^\top + A \Sigma(\mathbf{Y}) A^\top]_{jk} \quad \text{from (58)} \\ &= \sum_{j,k} W_{kj} \frac{\partial}{\partial A_{i\alpha}} \left[- \sum_{\beta} \Sigma(\mathbf{X}, \mathbf{Y})_{j\beta} A_{k\beta} - \sum_{\beta} \Sigma(\mathbf{X}, \mathbf{Y})_{k\beta} A_{j\beta} + \sum_{\beta, \gamma} \Sigma(\mathbf{Y})_{\beta\gamma} A_{j\beta} A_{k\gamma} \right] \\ &= \sum_{j,k} W_{kj} \left[- \sum_{\beta} \Sigma(\mathbf{X}, \mathbf{Y})_{j\beta} \delta_{ik} \delta_{\alpha\beta} - \sum_{\beta} \Sigma(\mathbf{X}, \mathbf{Y})_{k\beta} \delta_{ij} \delta_{\alpha\beta} + \sum_{\beta, \gamma} \Sigma(\mathbf{Y})_{\beta\gamma} (A_{j\beta} \delta_{ik} \delta_{\alpha\gamma} + A_{k\gamma} \delta_{ij} \delta_{\alpha\beta}) \right] \\ &= - \sum_j W_{ij} \Sigma(\mathbf{X}, \mathbf{Y})_{j\alpha} - \sum_k W_{ki} \Sigma(\mathbf{X}, \mathbf{Y})_{k\alpha} + \sum_{\beta, j} W_{ij} \Sigma(\mathbf{Y})_{\beta\alpha} A_{j\beta} + \sum_{\gamma, k} W_{ki} \Sigma(\mathbf{Y})_{\alpha\gamma} A_{k\gamma} \\ &= 2 \{W [A \Sigma(\mathbf{Y}) - \Sigma(\mathbf{X}, \mathbf{Y})]\}_{i\alpha} \end{aligned}$$

after gathering terms and simplifying, and Eq. (5) follows.

B Proof of expansion of multivariate Granger causality

Here we prove Eq. (18). We consider the case of there being no conditional third variable, since the extension to this case is trivial. We first expand in terms of predictor variables according to

$$\begin{aligned}
\mathcal{F}_{Y \rightarrow X} &= \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^-)|} \right) \\
&= \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^-)| \cdot |\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^-)| \cdot |\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \mathbf{Y}_2^-)| \cdots |\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \dots \mathbf{Y}_{m-1}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^-)| \cdot |\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \mathbf{Y}_2^-)| \cdots |\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \dots \mathbf{Y}_m^-)|} \right) \\
&= \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^-)|} \right) + \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \mathbf{Y}_2^-)|} \right) + \cdots \\
&\quad + \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \dots \mathbf{Y}_{m-1}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus \dots \mathbf{Y}_m^-)|} \right) \\
&= \mathcal{F}_{Y_1 \rightarrow X} + \mathcal{F}_{Y_2 \rightarrow X | Y_1} + \mathcal{F}_{Y_3 \rightarrow X | Y_1 \oplus Y_2} + \cdots + \mathcal{F}_{Y_m \rightarrow X | Y_1 \oplus Y_2 \oplus \dots \oplus Y_{m-1}} .
\end{aligned} \tag{60}$$

To expand in terms of predictees we use the expansion

$$|\Sigma(\mathbf{X} | \mathbf{W})| = \Sigma(X_1) \Sigma(X_2 | \mathbf{W} \oplus X_1) \Sigma(X_3 | \mathbf{W} \oplus X_1 \oplus X_2) \cdots \Sigma(X_n | \mathbf{W} \oplus X_1 \oplus \dots \oplus X_{n-1}) , \tag{61}$$

which follows from repeated application of Eq. (3). We obtain

$$\begin{aligned}
\mathcal{F}_{Y_1 \rightarrow X} &= \log \left(\frac{|\Sigma(\mathbf{X} | \mathbf{X}^-)|}{|\Sigma(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}_1^-)|} \right) \\
&= \log \left(\frac{\Sigma(X_1 | \mathbf{X}^-) \Sigma(X_2 | \mathbf{X}^- \oplus X_1) \cdots \Sigma(X_n | \mathbf{X}^- \oplus X_1 \oplus X_2 \oplus \dots \oplus X_{n-1})}{\Sigma(X_1 | \mathbf{X}^- \oplus \mathbf{Y}_1^-) \Sigma(X_2 | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus X_1) \cdots \Sigma(X_n | \mathbf{X}^- \oplus \mathbf{Y}_1^- \oplus X_1 \oplus X_2 \oplus \dots \oplus X_{n-1})} \right) \\
&= \mathcal{F}_{Y_1 \rightarrow X_1 | X} + \mathcal{F}_{Y_1 \rightarrow X_2 | X \oplus X_1^0} + \mathcal{F}_{Y_1 \rightarrow X_3 | X \oplus X_1^0 \oplus X_2^0} + \cdots + \mathcal{F}_{Y_1 \rightarrow X_n | X \oplus X_1^0 \oplus X_2^0 \oplus \dots \oplus X_{n-1}^0} ,
\end{aligned} \tag{62}$$

and similar for the other components of the sum in Eq. (60), from which the result follows.

C Partial covariance of residuals for two variables jointly dependent on a third

Given the regressions

$$\begin{aligned}
\mathbf{X} &= A \cdot \mathbf{W} + \boldsymbol{\varepsilon} , \\
\mathbf{Z} &= B \cdot \mathbf{W} + \boldsymbol{\eta} ,
\end{aligned} \tag{63}$$

where the regression coefficients A, B are derived from an ordinary least squares, Yule-Walker or equivalent procedure, we show that

$$\Sigma(\boldsymbol{\varepsilon} | \boldsymbol{\eta}) = \Sigma(\mathbf{X} | \mathbf{Z} \oplus \mathbf{W}) , \tag{64}$$

assuming that all (partial) covariance matrices which appear below are invertible. We have

$$\begin{aligned}\Sigma(\boldsymbol{\varepsilon}) &= \Sigma(\mathbf{X} | \mathbf{W}) , \\ \Sigma(\boldsymbol{\eta}) &= \Sigma(\mathbf{Z} | \mathbf{W}) , \\ \Sigma(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) &= \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) .\end{aligned}\tag{65}$$

Thus we may calculate that

$$\Sigma(\boldsymbol{\varepsilon} | \boldsymbol{\eta}) = \Sigma(\mathbf{X} | \mathbf{W}) - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{X} | \mathbf{W}) .\tag{66}$$

Using the block matrix inversion formula for $\Sigma(\mathbf{Z} \oplus \mathbf{W})$, we may also calculate that

$$\Sigma(\mathbf{X} | \mathbf{Z} \oplus \mathbf{W}) = \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} \Sigma(\mathbf{W}, \mathbf{X}) .\tag{67}$$

Now expanding the $\Sigma(\mathbf{X} | \mathbf{W}) \equiv \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W}, \mathbf{X})$ term in (66), we find using (67) that (64) is equivalent to

$$\begin{aligned}\Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W}, \mathbf{X}) + \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{X} | \mathbf{W}) \\ = \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{X}) + \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} \Sigma(\mathbf{W}, \mathbf{X}) .\end{aligned}$$

Or, rearranging and factorizing,

$$\begin{aligned}\left[\Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} - \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} \right] \Sigma(\mathbf{W}, \mathbf{X}) \\ = \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} [\Sigma(\mathbf{Z}, \mathbf{X}) - \Sigma(\mathbf{Z}, \mathbf{X} | \mathbf{W})] .\end{aligned}\tag{68}$$

Now the term in square brackets on the RHS of (68) simplifies to $\Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W}, \mathbf{X})$ so that, factoring out $\Sigma(\mathbf{W}, \mathbf{X})$, (68) is equivalent to

$$\begin{aligned}\left[\Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} - \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \right] \\ \times \Sigma(\mathbf{W}, \mathbf{X}) = 0 .\end{aligned}\tag{69}$$

We now show that the term in the square brackets in (69) is zero; i.e. that

$$\Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} - \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} = 0 ,\tag{70}$$

thus proving (64). Rearranging and factoring out $\Sigma(\mathbf{W})^{-1}$, (70) becomes

$$\left[\Sigma(\mathbf{X}, \mathbf{W}) - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \right] \Sigma(\mathbf{W})^{-1} = \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) \Sigma(\mathbf{W} | \mathbf{Z})^{-1} ,$$

or, multiplying through on the right by $\Sigma(\mathbf{W} | \mathbf{Z})$,

$$\left[\Sigma(\mathbf{X}, \mathbf{W}) - \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \right] \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W} | \mathbf{Z}) = \Sigma(\mathbf{X}, \mathbf{W} | \mathbf{Z}) .$$

Expanding $\Sigma(\mathbf{W} | \mathbf{Z})$, factorizing and rearranging again, we get

$$\begin{aligned} & \left[\Sigma(\mathbf{X}, \mathbf{Z}) - \Sigma(\mathbf{X}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W}, \mathbf{Z}) \right] \Sigma(\mathbf{Z})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \\ &= \Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W} | \mathbf{Z}) , \end{aligned}$$

or, since the term in square brackets on the LHS is just $\Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W})$,

$$\Sigma(\mathbf{X}, \mathbf{Z} | \mathbf{W}) \left[\Sigma(\mathbf{Z})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) - \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W} | \mathbf{Z}) \right] .$$

We now show that, again, the term in square brackets is zero; i.e. that

$$\Sigma(\mathbf{Z})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) = \Sigma(\mathbf{Z} | \mathbf{W})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W} | \mathbf{Z}) . \quad (71)$$

Multiplying through on the left by $\Sigma(\mathbf{Z} | \mathbf{W})$, (71) is equivalent to

$$\Sigma(\mathbf{Z} | \mathbf{W}) \Sigma(\mathbf{Z})^{-1} \Sigma(\mathbf{Z}, \mathbf{W}) = \Sigma(\mathbf{Z}, \mathbf{W}) \Sigma(\mathbf{W})^{-1} \Sigma(\mathbf{W} | \mathbf{Z}) ,$$

which follows immediately on expanding $\Sigma(\mathbf{Z} | \mathbf{W})$ and $\Sigma(\mathbf{W} | \mathbf{Z})$, thus establishing (64).

References

- [1] M. Ding, Y. Chen and S. Bressler. Granger causality: Basic theory and application to neuroscience. In S. Schelter, M. Winterhalder, and J. Timmer, editors, *Handbook of Time Series Analysis*, 438–460. Wiley, Weinheim, 2006.
- [2] K. Friston, L. Harrison and W. Penny. Dynamic causal modeling. *Neuroimage*, 19(4):1273–302, 2003.
- [3] T. Schreiber. Measuring information transfer. *Phys Rev Lett*, 85(2):461–4, 2000.
- [4] C. Ladroue, S. Guo, K. Kendrick and J. Feng. Beyond element-wise interactions: Identifying complex interactions in biological processes. *PLoS One*, 4:e6899–e6899, 2009.
- [5] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc*, 77(378):304–313, 1982.
- [6] Z. Zhou, Y. Chen, M. Ding, P. Wright, Z. Lu, and Y. Liu. Analyzing brain networks with PCA and conditional Granger causality. *Hum Brain Mapp*, 30:2197–2206, 2009.
- [7] N. Wiener. The theory of prediction. In E. F. Beckenbach, editor, *Modern Mathematics for Engineers*. McGraw Hill, New York, NY, 1956.
- [8] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
- [9] S. L. Bressler and A. K. Seth. Wiener-Granger causality: A well established methodology. *Neuroimage*, x:xx–xx, 2010.

- [10] A. K. Seth. Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*, 1(1):50–63, 2009.
- [11] A. K. Seth. Measuring autonomy and emergence via Granger causality. *Artificial Life*, 16(2), 2009.
- [12] M. G. Kendall and A. Stuart. *The advanced theory of statistics, volume 2: “Inference and Relationship”*. Griffin, London, 1979.
- [13] L. Barnett, A. B. Barrett and A. K. Seth. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett*, 103:238701, 2009.
- [14] C. W. J. Granger. Economic processes involving feedback. *Inform Control*, 6:28–48, 1963.
- [15] P. Whittle. The analysis of multiple stationary time series. *J Royal Stat Soc B*, 15(1):125–139, 1953.
- [16] A. Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *T Am Math Soc*, 54(3):426–482, 1943.
- [17] A. K. Seth. A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Meth*, 186:262–273, 2010.
- [18] J. Geweke. Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc*, 79(388):907–915, 1984.
- [19] A. Kaiser and T. Schreiber. Information transfer in continuous processes. *Physica D*, 166:43–62, 2002.
- [20] Y. Chen, G. Rangarajan, J. Feng and M. Ding. Analyzing multiple nonlinear time series with extended Granger causality. *Phys Lett A*, 324:26–35, 2004.
- [21] D. Marinazzo, M. Pellicoro, and S. Stramaglia. Kernel method for nonlinear Granger causality. *Phys Rev Lett*, 100(14):144103, 2008.
- [22] A. K. Seth. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems*, 16:35–54, 2005.
- [23] O. Sporns, G. Tononi and G.M. Edelman. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*, 10:127–141, 2000.
- [24] Y. Chen, S. L. Bressler and M. Ding. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J Neurosci Meth*, 150:228–237, 2006.
- [25] Y. A. Rozanov. *Stationary Random Processes*. Holden-Day, San Francisco, CA, 1967.
- [26] J. Doob. *Stochastic Processes*. John Wiley, New York, NY, 1953.

- [27] H. Akaike. A new look at the statistical model identification. *IEEE Trans Autom Control*, 19:716-723, 1974.
- [28] G. Schwartz. Estimating the dimension of a model. *The Annals of Statistics*, 5(2):461-464, 1978.
- [29] S. Guo, A. K. Seth, K. Kendrick, C. Zhou and J. Feng. Partial Granger causality: Eliminating exogenous inputs and latent variables. *J Neurosci Meth*, 172:79-93, 2008.
- [30] L. A. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern*, 84:463-474, 2001.
- [31] O. Sporns. Complexity. *Scholarpedia*, 2(10):1623, 2007.
- [32] A. K. Seth, E. Izhikevich, G. N. Reeke and G. M. Edelman. Theories and measures of consciousness: An extended framework. *P Natl Acad Sci, USA*, 103(28):10799-10804, 2006.
- [33] M. Shanahan. Dynamical complexity in small-world networks of spiking neurons. *Phys Rev E Stat Nonlin Soft Matter Phys*, 78:041924-041924, 2008.
- [34] G. Tononi, O. Sporns and G. M. Edelman. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *P Natl Acad Sci, USA*, 91:5033-5037, 1994.
- [35] M. A. Bedau. Weak emergence. *Philosophical Perspectives*, 11:375-399, 1997.
- [36] C. R. Shalizi and C. Moore. What is a macrostate: Subjective observations and objective dynamics. <http://arxiv.org/abs/cond-mat/0303625>, 2003.
- [37] C. R. Shalizi, R. Haslinger, J-B. Rouquier, K. L. Klinkner and C. Moore. Automatic filters for the detection of coherent structure in spatiotemporal systems. *Phys Rev E Stat Nonlin Soft Matter Phys*, 73:036104-036104, 2006.
- [38] A. K. Seth. Granger causality. *Scholarpedia*, 2(7):1667, 2007.
- [39] D. O. Hebb. *The organization of behavior*. Wiley, New York, NY, 1949.
- [40] K. D. Harris. Neural signatures of cell assembly organization. *Nat Rev Neurosci*, 6:399-407, 2005.
- [41] R. Gaillard, S. Dehaene, C. Adam, S. Clémenceau, D. Hasboun, M. Baulac, L. Cohen and L. Naccache. Converging intracranial markers of conscious access. *PLoS Biol*, 7:e61-e61, 2009.
- [42] L. Angelini, M. de Tommaso, D. Marinazzo, L. Nitti, M. Pellicoro and S. Stramaglia. Redundant variables and Granger causality. *Phys Rev E*, in press, 2010.
- [43] R. B. Kline. *Principles and practice of structural equation modeling*. Guilford Press, New York, NY, 2005.