

Published in final edited form as:

*Neuroimage*. 2013 December ; 83: 189–199. doi:10.1016/j.neuroimage.2013.06.056.

## Canonical Granger Causality between Regions of Interest

Syed Ashrafulla, Justin P. Haldar, Anand A. Joshi, and Richard M. Leahy

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA, USA

### Abstract

Estimating and modeling functional connectivity in the brain is a challenging problem with potential applications in the understanding of brain organization and various neurological and neuropsychological conditions. An important objective in connectivity analysis is to determine the connections between regions of interest in the brain. However, traditional functional connectivity analyses have frequently focused on modeling interactions between time series recordings at individual sensors, voxels, or vertices despite the fact that a single region of interest will often include multiple such recordings. In this paper, we present a novel measure of interaction between regions of interest rather than individual signals. The proposed measure, termed canonical Granger causality, combines ideas from canonical correlation and Granger causality analysis to yield a measure that reflects directed causality between two regions of interest. In particular, canonical Granger causality uses optimized linear combinations of signals from each region of interest to enable accurate causality measurements from substantially less data compared to alternative multivariate methods that have previously been proposed for this scenario. The optimized linear combinations are obtained using a variation of a technique developed for optimization on the Steifel manifold. We demonstrate the advantages of canonical Granger causality in comparison to alternative causality measures for a range of different simulated datasets. We also apply the proposed measure to local field potential data recorded in a macaque brain during a visuomotor task. Results demonstrate that canonical Granger causality can be used to identify causal relationships between striate and prestriate cortex in cases where standard Granger causality is unable to identify statistically significant interactions.

### Keywords

connectivity; causality; local field potentials; functional imaging

## 1. Introduction

An important objective in brain research is to understand how information propagates between different regions (Jirsa & McIntosh, 2007). Electrophysiological measurements of brain activity can be useful for achieving this goal, since they provide rich information about the location and temporal dynamics of spontaneous and task-related brain networks. In particular, invasive local field potential (LFP) and electrocorticography (ECoG) measurements as well as noninvasive electroencephalography (EEG) and

---

\*Corresponding Author: leahy@sipi.usc.edu.

magnetoencephalography (MEG) data allow for the modeling of brain connectivity, with wide ranging implications for addressing neuroscientific questions (Astolfi et al., 2007; Bressler et al., 2007; Schoffelen & Gross, 2009), understanding mechanisms of neuropathology (Lin et al., 2009; Wilke et al., 2009), and studying neuropsychological conditions (Hesse et al., 2003). Many connectivity models require assumptions regarding the behavior of the relationship between signals from different regions; for example, autoregressive models try to find areas of the brain whose electrical activity co-varies with past activity in other areas of the brain (Cheung et al., 2010; Hesse et al., 2003).

This paper focuses on models based on Granger causality (GC) (Granger, 1969). Causal models estimate the strength and directionality of signal interactions by analyzing the joint distributions of their time series under appropriate modeling assumptions. Classical methods for causal modeling analyze causality between multiple time series in a pairwise fashion, using bivariate models (Geweke, 1982). However, pairwise analysis is not ideal for functional brain mapping experiments in which multiple time series are available from different sensors, voxels (in a reconstructed source volume), or vertices (from a reconstructed source surface). Usually, such time series can be spatially correlated because of the limited spatial resolution of the mapping techniques or because functional activation is spatially distributed across a larger region of cortex. One approach could be to remove the influence of these confounds via partial measures of causality (Guo et al., 2008), but such partial measures usually require the estimation of a complex parametric model and may be hard to interpret (Eichler, 2006; Ku et al., 2004; Zhou et al., 2009).

Rather than analyzing causality independently between pairs of time series, in some cases it may be more desirable to analyze causality between regions of interest (ROIs) that each include multiple correlated time series (d'Alessandro et al., 2003). Grouping multiple time series together can reduce the number of variables to estimate in a parametric model, can improve the signal-to-noise ratio of the resulting causality measure, and can help in identifying long-range connectivity that might previously have been obscured by larger apparent short-range connectivity that is artificially introduced by crosstalk within the ROI (Bin et al., 2009).

Many methods have been proposed to assess connectivity between ROIs. One popular method is canonical correlation (Hotelling, 1936), which estimates an undirected model of connectivity by maximizing the correlation between weighted linear combinations of signals from two ROIs. In addition to providing a measure of connectivity, canonical correlation also provides an estimate of the relative contribution of each signal from each ROI to the correlation between the ROIs (Deleus & Van Hulle, 2011; Kuylen & Verhallen, 1981). Canonical correlation analysis, however, is not a causal model and provides no information about the direction of information flow between ROIs.

Models of directed interaction between ROIs often use the concept of GC. Multivariate Granger causality (MGC) (Barrett & Seth, 2010) relies on multivariate autoregressive models of the signals between each ROI. While MGC can be accurate when the data records are sufficiently long, MGC involves substantially more parameters to estimate than do bivariate methods and can be prone to overfitting and sensitive to noise. A possible way to

reduce this effect is to use penalized autoregression to promote certain desirable qualities in the estimation of causality such as spatial smoothness or sparse connectivity (Valdés-Sosa et al., 2005). Another approach is Granger canonical correlation analysis (GCCA), (Wu et al., 2011) which aggregates signals in each ROI much like canonical correlation analysis. This approach results in a reduction of the number of parameters to estimate relative to methods like MGC and penalized autoregression, although it does not estimate the amount of causality or the underlying signals responsible for the causal connection (Sato et al., 2010), and can only identify the presence or absence of causality.

In this paper we formulate, develop, and apply a novel directed causal connectivity measure called canonical Granger causality (CGC) that combines the strengths of canonical correlation with the directionality of GC. Our CGC measure combines the strengths and overcomes the disadvantages of GC and MGC by using an optimized weighted linear combination of the time series to parsimoniously represent each ROI with a single time series. This is similar to the idea of canonical correlation, where the signal dimensionality is reduced by considering the weighted sum of signals (Correa et al., 2010). Subsequently, CGC computes standard bivariate GC between the two representative time series.

While CGC and MGC both summarize causal influences between ROIs, CGC is a more parsimonious model and thus is more stable for short timeseries as shown in the following sections. CGC and penalized autoregression both estimate causal models with low complexity, but CGC uses the a priori information to select regions of interest while penalized autoregression simply looks for sparse causal interactions without addressing problems related to crosstalk between signals. Finally, similar to GCCA, CGC also uses weighted sums to represent each region when estimating causality. However, CGC estimates the strength of causality between those regions; results from our simulations will show that CGC can thus better estimate the underlying connectivity between the signals of interest in each region.

A preliminary version of CGC was presented by Ashrafulla et al. (2012). This paper expands substantially upon the results presented in that work, presenting a refined procedure for computing CGC, using extensive simulations to evaluate and characterize the approach relative to methods like MGC and GCCA, and applying the method to identify causality in real LFP data.

This paper is organized as follows. In section 2 we review GC and MGC to establish the groundwork for CGC, and we describe GCCA for comparison. The CGC measure and associated algorithms are presented in section 3. A simulation study is described in section 4 that illustrates the advantages and disadvantages of the proposed approach. In section 5, the proposed measure is applied to real brain data acquired from a macaque performing a visuomotor task (Bressler et al., 1993), where we show that CGC can identify causal interactions between striate and prestriate regions of the occipital lobe. Finally, discussions and conclusions are presented in sections 6 and 7, respectively.

## 2. Review of Granger Causality Measures

### 2.1. Granger Causality (GC)

Let  $x_1$  and  $x_2$  be two time series of length  $N$ . If the past values of  $x_2$  substantially improve the prediction of  $x_1$ , then  $x_2$  is said to “Granger cause”  $x_1$  (Granger, 1969). GC thus attempts to measure the extent to which past values of  $x_2$  can be used to predict the present value of  $x_1$  (Sims, 1972). Mathematically, calculation of GC from  $x_2$  to  $x_1$  considers two different  $P^{th}$  order autoregressive (AR) models given by:

$$x_1[n] = \sum_{p=1}^P b[p] x_1[n-p] + r_1[n] \quad (1)$$

$$\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \sum_{p=1}^P \mathbf{A}[p] \begin{bmatrix} x_1[n-p] \\ x_2[n-p] \end{bmatrix} + \begin{bmatrix} s_1[n-p] \\ s_2[n-p] \end{bmatrix} \quad (2)$$

$n = 1, \dots, N$

where the AR coefficients  $b[p] \in \mathbb{R}$  and  $\mathbf{A}[p] \in \mathbb{R}^{2 \times 2}$ ,  $p = 1, \dots, P$ , are estimated by minimizing the variance or 2-norm of the prediction errors  $r_1[n]$  and  $[s_1[n] \ s_2[n]]^T$  in equations (1) and (2) respectively. The quantities

$$\rho_1^2 = \frac{1}{N-1} \sum_{n=1}^N (r_1[n])^2 \quad \text{and} \quad \sigma_{2 \rightarrow 1}^2 = \frac{1}{n-1} \sum_{n=1}^N (s_1[n])^2 \quad (3)$$

measure the total AR prediction errors under the two different models.  $\rho_1^2$  is a measure of how well the past of  $x_1$  can be used to predict its future values, while  $\sigma_{2 \rightarrow 1}^2$  is a measure of how well the past of both  $x_1$  and  $x_2$  can be used to predict future values of  $x_1$ . GC is computed as the ratio of these two residual errors:

$$G_{x_2 \rightarrow x_1} = \ln \frac{\rho_1^2}{\sigma_{2 \rightarrow 1}^2}. \quad (4)$$

GC takes values between 0 and  $\infty$ , with large values of GC indicating that the past of  $x_2$  contributes significantly to the prediction of  $x_1$ .

GC requires the estimation of a modest number ( $P+4P = 5P$ ) of AR coefficients, which makes estimation of GC relatively stable for short time series (low  $N$ ) or more complex models (high  $P$ ). Granger causality does not provide a mechanism for analyzing causality between three or more time series. In addition, when GC is applied pairwise to spatially-correlated signals from multiple areas in the brain, GC is biased towards finding higher causality between spatially-adjacent areas (Wang et al., 2007).

### 2.2. Multivariate Granger Causality (MGC)

MGC is one method to extend GC for analyzing causality between two sets of multiple time series (Barrett & Seth, 2010). Given the vector-valued time series  $\mathbf{y}_1$  (with  $M_1$  values at each time point) and the vector-valued time series  $\mathbf{y}_2$  (with  $M_2$  values at each time point), MGC

attempts to measure the extent to which past values of  $\mathbf{y}_2$  can be used to predict the present value of  $\mathbf{y}_1$ . Analogous to the two AR models used by GC, MGC considers two  $P^{\text{th}}$  order multivariate autoregressive (MVAR) models given by:

$$\mathbf{y}_1[n] = \sum_{p=1}^P \mathbf{B}[p] \mathbf{y}_1[n-p] + \mathbf{r}_1[n] \quad (5)$$

$$\begin{bmatrix} \mathbf{y}_1[n] \\ \mathbf{y}_2[n] \end{bmatrix} = \sum_{p=1}^P \mathbf{A}[p] \begin{bmatrix} \mathbf{y}_1[n-p] \\ \mathbf{y}_2[n-p] \end{bmatrix} + \begin{bmatrix} \mathbf{s}_1[n] \\ \mathbf{s}_2[n] \end{bmatrix} \quad (6)$$

$n = 1, \dots, N$

where  $\mathbf{B}[p] \in \mathbb{R}^{M_1 \times M_1}$  and  $\mathbf{A}[p] \in \mathbb{R}^{(M_1+M_2) \times (M_1+M_2)}$ ,  $p = 1, \dots, P$ , denote the MVAR coefficient matrices which are estimated to minimize the variances of the prediction errors  $\mathbf{r}_1[n]$  and  $[\mathbf{s}_1[n]^T \mathbf{s}_2[n]^T]^T$ . Similar to GC, MGC is defined as the ratio between the sizes of these two prediction errors and is given by:

$$M_{y_2 \rightarrow y_1} = \ln \frac{|\mathbf{P}_1^2|}{|\Sigma_{2 \rightarrow 1}^2|} \quad (2)$$

where  $|\cdot|$  denotes the matrix determinant and

$$\mathbf{P}_1^2 = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{r}_1[n] \mathbf{r}_1^T[n]) \quad (8)$$

$$\Sigma_{2 \rightarrow 1}^2 = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{s}_1[n] \mathbf{s}_1^T[n]). \quad (9)$$

As with GC, MGC takes values between 0 and  $\infty$ , with large values indicating that the past of  $\mathbf{y}_2$  contributes significantly to the prediction of  $\mathbf{y}_1$ . Relative to GC, MGC enables the estimation of causality between two ROIs, with each ROI containing multiple time series.

However, MGC requires the estimation of many more  $(M_1^2 P + (M_1 + M_2)^2 P)$  coefficients, which makes it less stable than GC, particularly when the number of samples is small.

### 2.3. Granger canonical correlation analysis (GCCA)

GCCA (sometimes named cluster Granger causality (Sato et al., 2010)) is an alternate measure of regional causality which is based on lagged correlations. Considering the vector-valued time series  $\mathbf{y}_1$  and  $\mathbf{y}_2$  above, and the order  $P$  which corresponds to the maximum lag in the causal interaction, GCCA is computed as (Wu et al., 2011)

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \underset{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{M_1}, \boldsymbol{\beta} \in \mathbb{R}^{M_2} \\ \|\boldsymbol{\alpha}\|_{\ell_2} = \|\boldsymbol{\beta}\|_{\ell_2} = 1}}{\arg \max} \text{corr}(\boldsymbol{\alpha}^T \mathbf{y}_1[n], \boldsymbol{\beta}^T \mathbf{y}_2[n-P]) \quad (10)$$

$$R_{y_2 \rightarrow y_1} = G_{\beta^T y_2 \rightarrow \alpha^T y_1} \quad (11)$$

where  $\text{corr}(a, b)$  is the correlation between two processes  $a$  and  $b$ , and  $G$  represents the GC defined in (4). GCCA takes values between 0 and 1, with a zero value implying no causality. Since the only parameters to estimate are the weights, GCCA only requires the estimation of  $M_1 + M_2$  parameters. In addition, GCCA can be solved simply via singular value decomposition techniques similar to those used for canonical correlation (Kuylen & Verhallen, 1981). GCCA is able to detect causality relatively well (Sato et al., 2010), but as we will demonstrate in simulation in section 4, GCCA is not effective in determining the strength of causality.

### 3. Canonical Granger Causality (CGC)

This section introduces a new measure, CGC, that combines ideas from canonical correlation analysis and GC analysis. Like MGC, CGC enables the estimation of causality between two ROIs containing multiple time series. However, CGC uses a more parsimonious model than MGC, which makes it less sensitive to noise and less prone to overfitting. Like canonical correlation and GCCA, CGC uses weighted sums of time series to represent each region. However, CGC measures the amount of Granger causality and thus is able to better measure the strength of regional interactions.

Considering the vector-valued time series  $\mathbf{y}_1$  and  $\mathbf{y}_2$  introduced above, we define CGC as:

$$C_{y_2 \rightarrow y_1} = \max_{\substack{\alpha \in \mathbb{R}^{M_1}, \beta \in \mathbb{R}^{M_2} \\ \|\alpha\|_{\ell_2} = \|\beta\|_{\ell_2}}} G_{\beta^T y_2 \rightarrow \alpha^T y_1} \quad (12)$$

where  $G$  represents the Granger causality defined in equation (4). CGC is obtained by computing the standard GC on scalar-valued time series  $\hat{x}_1 = \alpha^T \mathbf{y}_1$  and  $\hat{x}_2 = \beta^T \mathbf{y}_2$  obtained from a linear combination of the original time series from each ROI. The linear combination weights for each ROI,  $\alpha$  and  $\beta$ , are optimized to maximize the GC between  $\hat{x}_1$  and  $\hat{x}_2$ . Note that we have constrained  $\alpha$  and  $\beta$  to have unit norm because GC is invariant to rescaling of the data.

Given the unit-norm constraint, the estimation of  $\alpha$  and  $\beta$  requires estimation of  $M_1 + M_2 - 2$  parameters. Combined with the estimation of  $5P$  AR coefficients for standard GC (cf. section 2.1), CGC requires the estimation of  $M_1 + M_2 - 2 + 5P$  parameters. Note that CGC has substantially fewer parameters to estimate compared to the  $M_1^2 P + (M_1 + M_2)^2 P$  parameters that must be estimated for MGC (cf. section 2.2), and that this difference is particularly pronounced for large values of  $P$ ,  $M_1$ , or  $M_2$ . As we will demonstrate in section 4, this reduction in the number of parameters results in CGC being more stable and accurate in identifying ROI-to-ROI causality than MGC for short data records. However, we note that CGC is an ROI-wise measure that finds the strongest causal interaction between the two ROIs, and hence may not provide direct insight into the individual signal interactions if there are multiple causal connections between ROIs.

Computation of CGC requires optimization with respect to weight vectors  $\mathbf{a}$  and  $\mathbf{b}$ . Due to the unit-norm constraints,  $\mathbf{a}$  lies on the  $(M_1 - 1)$ -sphere  $\mathcal{S}^{M_1-1}$ , and  $\mathbf{b}$  lies on the  $(M_2 - 1)$ -sphere  $\mathcal{S}^{M_2-1}$ . In the appendix we exploit the structure of these constraint sets to introduce a numerical algorithm to optimize these weights, using a variation of a technique developed for optimization on the Steifel manifold (Edelman et al., 1998).

## 4. Simulations

We compared the properties of the CGC, MGC, and GCCA measures using multiple different simulated time series. Performance was evaluated by examining each method's ability to find true causality while avoiding false causality. We tested all measures on simulated time series with varying numbers of samples ( $N$ ), model order ( $P$ ), and noise levels as described below.

### 4.1. Simulation models

We simulated directed networks of 2 ROIs (ROI<sub>1</sub> and ROI<sub>2</sub>) with  $M_1 = M_2 = 4$ . Simulations were constructed from AR models, and included interference and noise (described in section 4.1.2). Simulations were generated for model orders  $P \in \{2, 4, 6, 8, 10\}$ , number of time points  $N \in \{100, 200, 400\}$ , signal-to-interference ratios  $\text{SIR} \in \{1, 5, 25\}$ , and signal-to-noise ratios  $\text{SNR} \in \{1, 5, 25\}$ , with a fixed number of interferers ( $I = 2$ ).

**4.1.1. Simulated Signals of Interest**—Each ROI has one underlying time series involved in information transfer between the ROIs, denoted by  $x_1$  and  $x_2$  respectively. We simulate these signals using bivariate AR processes according to:

$$\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \sum_{p=1}^P \begin{bmatrix} D[p] & A[p] \\ 0 & D[p] \end{bmatrix} \begin{bmatrix} x_1[n-p] \\ x_2[n-p] \end{bmatrix} + \begin{bmatrix} \eta_1[n] \\ \eta_2[n] \end{bmatrix}. \quad (13)$$

Here, the past of  $x_1$  has no effect on  $x_2$  because the corresponding coefficients in the AR model are zero. The only possible causal information transfer is then from ROI<sub>2</sub> to ROI<sub>1</sub>, with the amount of causality modulated by  $A[p]$ ,  $p = 1, \dots, P$ . The process is driven by simulated zero-mean white Gaussian innovations  $\eta_1$  and  $\eta_2$ .

These AR models are generated randomly in such a way that (1) we consistently obtained stable AR models, and (2) the resulting  $P^{\text{th}}$ -order AR signals cannot be accurately modeled by another AR model with order smaller than  $P$ . To test the ability of the three measures to detect causality, we generated two types of processes: those for which there is no causality ( $A[p] = 0$ ,  $p = 1, \dots, P$ ) and those for which there is causality from  $x_2$  to  $x_1$  ( $A[p] \neq 0$  for some  $p = 1, \dots, P$ ). In each case, we generated the coefficients for a given order  $P$  using the following steps:

1. We constructed two sequences:

$$\begin{aligned} d[p] &= \frac{1}{p}, p=1, \dots, P, \\ a[p] &= \text{Uniform}[0.15, 0.3], p=1, \dots, P, \end{aligned}$$



where Uniform  $[j, k]$  is a uniformly-distributed random variable between real numbers  $j$  and  $k$ .

2. We set  $D[p]$  equal to  $0.3 \cdot d[p]$ . For no causality, we set  $A[p] = 0$ ,  $P = 1, \dots, P$ . For causality, we set  $A[p]$  equal to  $d[p] \cdot a[p]$ . Combined with the previous step, this procedure was found to always generate a stable process for our simulation parameters.
3. Using monotonically descending  $d[p]$  coefficients may admit lower-order AR modeling of signals (e.g. an AR model with  $P = 2$  may sufficiently fit signals from an AR process simulated above without reordering for  $P = 4$ ). So, to ensure the signals require order- $P$  AR modeling, we randomly reordered the AR matrices:
  - (a) We let  $(q_1, q_2, \dots, q_P)$  be a random shuffling of the indices  $\{1, \dots, P\}$ .
  - (b) We assigned  $D[q_1] \rightarrow D[1], D[q_2] \rightarrow D[2], \dots, D[q_P] \rightarrow D[P]$ .
  - (c) We applied the same reordering to  $A[p]$ ,  $p = 1, \dots, P$ .

**4.1.2. Observation Model**—Given the underlying signals  $x_1$  and  $x_2$ , we simulated the measured vector-valued time series for each ROI according to

$$\mathbf{y}_r[n] = \mathbf{g}_r x_r[n] + \frac{\boldsymbol{\delta}_r[n]}{\text{SIR}} + \frac{\boldsymbol{\epsilon}_r[n]}{\text{SNR}} \quad r=1, 2, \quad (14)$$

where  $\mathbf{g}_r \in \mathbb{R}^4$  is a vector describing the relative contribution of  $x_r$  to each of the 4 time series in ROI <sub>$r$</sub>  (drawn uniformly from  $\mathcal{S}^4$  in our simulation),  $\boldsymbol{\delta}_r$  represents simulated interference, and  $\boldsymbol{\epsilon}_r$  is simulated measurement noise (white Gaussian with unit variance in our simulation).

The interference signals  $\boldsymbol{\delta}_1$  and  $\boldsymbol{\delta}_2$  are vector-valued time series used to emulate background brain activity that is not associated with causality between the two ROIs. To emulate within-region brain activity, we want the signals to be sums of causally related time series that are independent from signals in other regions of activity. The time series  $\boldsymbol{\delta}_1[n]$  and  $\boldsymbol{\delta}_2[n]$  are each simulated as a sum of randomly generated scalar AR time series  $z_{r,i}[n]$ ,  $i = 1, 2$  according to:

$$\boldsymbol{\delta}_r[n] = \sum_{i=1}^2 \mathbf{h}_{r,i} z_{r,i}[n] \quad r=1, 2, \quad (15)$$

where  $\mathbf{h}_{r,i}$  represent the contribution of each interferer to each time series with its corresponding region, and are drawn uniformly at random from  $\mathcal{S}^4$ . For region  $r \in \{1, 2\}$ , we enforce the desired within-region causality by generating the scalar AR time series  $z_{r,1}$  and  $z_{r,2}$  with autoregression of the same order  $P$  as the source model:

$$\begin{bmatrix} z_{r,1}[n] \\ z_{r,2}[n] \end{bmatrix} = \sum_{p=1}^P \begin{bmatrix} D[p] & B[p] \\ B[p] & D[p] \end{bmatrix} + \begin{bmatrix} \gamma_1[n] \\ \gamma_2[n] \end{bmatrix}. \quad (16)$$

We choose the coefficients  $D[p]$  and  $B[p]$  using the same procedure described in section 4.1.1 for generating  $D[p]$  and  $A[p]$ .



## 4.2. Detecting Causality

We used receiver operating characteristics (ROCs) to evaluate the performance of CGC and MGC. The ROC curve is a plot of the true positive versus false positive rate computed from a test statistic as a function of a decision threshold (Metz, 1978). For the simulation setup in section 4.1, we define a “true positive” as occurring when the causality from ROI<sub>2</sub> to ROI<sub>1</sub> exceeds the decision threshold applied to the Granger measure when  $A[p]$  is nonzero, while a “false positive” occurs when the causality from ROI<sub>2</sub> to ROI<sub>1</sub> exceeds the decision threshold when  $A[p] = 0, p = 1, \dots, P$ .

We performed  $T = 1500$  Monte Carlo simulations, computing for each simulation

- CGC:  $C_{y_2 \rightarrow y_1}$  and  $C_{y_1 \rightarrow y_2}$ ,
- MGC:  $M_{y_2 \rightarrow y_1}$  and  $M_{y_1 \rightarrow y_2}$ , and
- To establish a best case performance limit, we also computed bivariate GC using a linear combination of signals based on the true weights ( $\mathbf{g}_1$  and  $\mathbf{g}_2$ ) used to generate the multivariate data from the bivariate signals of interest:  $G_{\mathbf{g}_2^T y_2 \rightarrow \mathbf{g}_1^T y_1}$  and  $G_{\mathbf{g}_1^T y_1 \rightarrow \mathbf{g}_2^T y_2}$ .

The empirical ROC curves for each measure derived from these simulations are shown in Figure 2 for the simulations where  $N = 400$ , SIR = 5 and SNR = 1. It can be seen from each of the plots that CGC shows a higher true positive rate than MGC for a given false positive rate. Increasing the order  $P$  of the simulated AR processes amplifies the difference between CGC and MGC, due to the increasing difference in the number of parameters between the two measures. Specifically, for  $P = 8$  the number of parameters required for MGC is  $(4^2 \cdot 8 + (4 + 4)^2 \cdot 8) / (4 + 4 + 5 \cdot 8) = 40/3$  times more than the number of parameters required for CGC. Additionally, our simulations show that applying GC to ROI signals formed using perfect prior knowledge of the original weights  $\mathbf{g}_1$  and  $\mathbf{g}_2$  yields better performance than CGC, as expected. We note that with GCCA, we found ROCs and AUCs nearly equal to CGC for all simulations.

In order to further compare the performance of CGC to MGC, we summarize the ROCs by calculating the area under the ROC curve (AUC). The AUC measures the ability of each measure to discern true directions of causality, averaged over all possible choices of the false positive rate. In order to estimate a 95% confidence interval of AUC non-parametrically, we performed 1000 bootstrap simulations (Efron & Tibshirani, 1993).

In Figure 3 the AUC is plotted, along with its bootstrapped confidence intervals, as a function of  $P$  and  $N$ , for a case with SIR = SNR = 5. As  $N$  decreases, the performance of MGC deteriorates substantially, while CGC is still able to reasonably differentiate true causality from false causality. Similar to the previous results, the difference between the performances of CGC and MGC grows even more as  $P$  increases. As before, the performance of MGC and CGC is surpassed by the performance of GC with known weights, though the performance of both MGC and CGC approach that of GC with known weights as  $N$  increases. Additional results in Table 1 of the supplementary material show that as SIR and SNR decrease, the gap in performance of CGC over MGC decreases, but CGC always

has a larger mean AUC; in fact, for most cases CGC also has a 95% AUC confidence interval that has no overlap with the corresponding 95% AUC confidence interval for MGC.

CGC does not perform as well as GC with known weights due to the upward bias in CGC: by calculating a maximum over weights, CGC is guaranteed to be higher than GC with known weights. For long time series, the difference between estimated causalities in true and false directions is significantly larger than this bias. However, for short time series, the bias increases and therefore the ROC performance of CGC deteriorates. We do not show results comparing CGC to GCCA, as they showed nearly equal AUC confidence intervals for all simulation parameter choices.

### 4.3. Estimating Causality

In addition to CGC's ability to detect the presence of causality, we also evaluated its ability to quantify the strength of causality. In particular, we analyzed the error of CGC and GCCA with respect to the ground truth of bivariate GC with known weights. We do not analyze MGC as MGC showed much larger bias than CGC or GCCA for most cases. For each simulation, we calculate how well CGC and GCCA match the bivariate causality by taking the squared difference:

$$error_{CGC} = \left( C_{y_2 \rightarrow y_1} - G_{\mathbf{g}_2^T y_2 \rightarrow \mathbf{g}_1^T y_1} \right)^2, \quad (17)$$

$$error_{GCCA} = \left( R_{y_2 \rightarrow y_1} - G_{\mathbf{g}_2^T y_2 \rightarrow \mathbf{g}_1^T y_1} \right)^2. \quad (18)$$

Mean and 95% confidence intervals for these errors are plotted in Figure 4 for multiple  $N$  and  $P$  values, for the case where  $SIR = SNR = 5$ . For low model orders, both CGC and GCCA track the ground truth bivariate GC well; in addition, it can be seen that GCCA has less error than CGC for these cases. For large  $N$ , CGC produces more accurate estimates of causality than GCCA. For small  $N$ , GCCA performs better than CGC in accuracy. However, at small  $N$  it can also be seen that neither estimate is an accurate estimator of the underlying causality. The benefit of GCCA is in fact much smaller than its error from the true causality. Additional results shown in the supplementary material in Table 2 and Table 3 indicate that errors rates for both CGC and GCCA is relatively unaffected by  $SIR$ , but that CGC is generally more accurate than GCCA as  $SNR$  decreases.

## 5. Real Experimental Data

We also applied CGC and MGC to estimate causality between brain regions using data collected during a visuomotor task performed by a macaque. This data is a prototypical example of a case with real brain data where we are interested in ROIs rather than individual channels (Goldberg et al., 2002; Turi et al., 2012). The data for this task was made available by Dr. Steven Bressler (Florida Atlantic University, Boca Raton, FL, USA) and consisted of local field potential (LFP) recordings as explained below.

## 5.1. Experimental Setup

A macaque was implanted with transcortical bipolar electrodes to record LFPs at multiple brain locations on one hemisphere, including multiple recordings in the striate and prestriate. The macaque was also trained to perform a visuomotor decision making task (Bressler et al., 1993). The macaque initiated each trial of the experiment by depressing a lever, after which a visual stimulus appeared for 100ms. The visual cue would inform the monkey to either let go of the lever (“Go”) or keep the lever down (“NoGo”).

Voltages were recorded at a 200Hz sampling rate. Defining the lever’s initial descent to be at time  $t = 0$ , we focus on a time interval up to 300ms after stimulus ( $t = 300\text{ms}$ ). After pooling together multiple runs, we have 480 trials for each condition “Go” (corresponding to a stimulus image of 4 dots in the shape of a diamond) and “NoGo” (corresponding to a stimulus image of 4 dots in the shape of a line).

## 5.2. Preprocessing

In order to examine causality in visual cortex, we computed time-varying estimates of causality over the period 0-300ms, using the windowing parameters described by Ding et al. (2000). The 6 electrodes located in the visual cortex naturally form two anatomical ROIs:  $S_1$ ,  $S_2$  and  $S_3$  form a striate ROI and IT, P1 and P2 form a prestriate ROI, as illustrated in Figure 5. These ROIs and their connectivity have been associated with visual pattern recognition (Liang et al., 2000; Mishkin et al., 1983). The ability to cluster electrodes into anatomically constrained ROIs in this manner is typical of multielectrode invasive recordings.

The focus of our analysis was on alpha and beta band activity because of previous work that studied these bands with this data (Bressler et al., 1999; Chen et al., 2006); therefore, we applied a 14th-order elliptical filter to the data with a passband of 8–30 Hz and a stopband attenuation of 60 dB for frequencies below 5 Hz and above 40 Hz. Such filtering may lead to spurious causality (Florin et al., 2010) so for each autoregressive model estimated below we used a model selection procedure using the Akaike Information Criterion to determine the order of the autoregression model. Such a procedure, under the constraint that the same filter is used for every signal, will reduce the amount by which filtering affects causality estimates (Barnett & Seth, 2011). We also subtracted the ensemble mean from the signals to remove a cause of 2<sup>nd</sup> order non-stationarity, and scaled each time series to have unit power.

The CGC and MGC measures were computed between the striate and prestriate ROIs for each time point using a sliding time-window of duration 55ms, yielding time series of causality values. We chose an intermediate window length to heuristically balance between the benefits of short windows (capturing short-time causality, avoiding non-stationarity) and the benefits of long windows (reducing the effects of trial-to-trial ERP variability (Wang et al., 2008), decreasing the variance in causality estimation). Separate causality measures were obtained for the trials corresponding to the “Go” response and the trials corresponding to the “NoGo” response. Bootstrapping with 1000 resamples was used to compute confidence intervals on all estimates.

### 5.3. Striate-Prestriate Causality

The resulting CGC and MGC measures for each condition are plotted as a function of time in Figure 6. Both measures indicate that the peak causality from striate to prestriate occurs in both tasks at 57.5ms. However, the difference between the two tasks is more pronounced for CGC than MGC; in fact, for CGC there is a statistically significant ( $p = 0.02$ ) difference in causality between the two tasks when correcting for multiple comparisons across time through the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). The task-related difference in causality may be attributed to the increased complexity in the stimulus image of a diamond as opposed to the stimulus image of a line (Bressler et al., 1999).

In addition, both measures indicate an increase in causality from prestriate to striate specific to the “Go” task at 137.5ms. Again, the difference between “Go” and “NoGo” at this time point is quite pronounced for CGC. There is a statistically significant ( $p = 0.01$ ) difference between the causality between these two conditions for CGC when correcting for multiple comparisons as before, while MGC does not demonstrate such statistically significant differences at this level between the two conditions at any point during the experiment. Note that signal propagation from the prestriate to striate has been observed in previous literature on visuomotor processing when visual stimulation leads to a motor action response (Lock et al., 2003). The differences in connectivity as observed using CGC for “Go” and “NoGo” can be explained as a result of this signal interaction since “Go” involves visual and visuomotor networks, while “NoGo” involves only visual networks.

### 5.4. Electrode Weights

The computation of CGC also yields linear combination weights and the signals of interest for each ROI. We analyzed the weights at the two time points, 57.5ms and 137.5ms, corresponding to peak differences in causality between “Go” and “NoGo.” Results are only shown for the “Go” task, which demonstrated causality at both time points. In Figure 7 we plot the weight vectors estimated in CGC in each direction of causality. At each sensor, we also plotted the confidence interval using the previous bootstrap procedure.

At the first time of large causality (57.5ms), the signals from  $S_3$  for the striate and  $P_2$  for the prestriate contribute substantially to the linear combination used in CGC, as indicated by the large magnitude of the weights at those sensor locations. Later at 137.5ms, only  $P_2$  contributes substantially to the linear combination for the prestriate, while both  $S_1$  and  $S_3$  contribute for the striate. The presence of  $S_3$  and  $P_2$  is consistent with previous analysis of this data (Liang et al., 2000). However, in the causality of interest at 137.5ms, we found that there was a significant increase in the weight for  $S_1$  with respect to the first time of large causality. Using a two-sample t-test, we found that there was a significant difference ( $p = 0.05$ ) in the weight of  $S_1$  between the two timepoints and directions of causality. The presence of  $S_1$  is not found with bivariate GC (Liang et al., 2000) and therefore needs to be further investigated.

## 6. Discussion

We propose a new measure, CGC, that models directed connectivity between two ROIs by estimating optimal weighted sums of the signals from each ROI. The resulting scalar-valued weighted sum can be viewed as a regional signal representing the ROI. In each ROI, the signals of interest are the ones that participate in a causal interaction between ROIs, with the weights representing the spatial topography of these signals.

With simulated and LFP data, we find CGC to be more robust to sample size and model order than an alternative method for calculating regional causality (MGC). In simulation, CGC showed improved causality detection performance compared to MGC as shown by the difference in AUC and TPR between causality measures. In LFP data, CGC showed short-time task differences in causality between the striate and prestriate; such statistically significant differences were not observed with MGC. This striate-prestriate causality has previously been associated with tasks involving visual pattern recognition (Liang et al., 2000; Mishkin et al., 1983) and visuomotor processing where the visual stimulus causes a motor action (Lock et al., 2003). CGC is more able to discriminate between causal and non-causal interactions for shorter sample sizes and higher autoregressive model orders partially because CGC simplifies the model of regional causality by taking a weighted sum of the signals in each region.

We found in simulation that CGC was a better estimate of the underlying bivariate causality than GCCA whenever there was a large number of samples, high model order, and/or low SNR. This is the result of the differences between the way the two measures estimate their weight vectors: CGC is maximizing Granger causality while GCCA is only maximizing lagged cross-correlation at order  $P$ . As a result, GCCA degrades in performance relative to CGC as model order increases, because GCCA only analyzes correlations of lag equal to the model order while CGC summarizes the correlations of lags from 0 up to the model order. One consequence of our simulation results is that GCCA may have a difficult task in differentiating between two unequally strong causal interactions, while CGC differentiates between those unequally strong interactions due to the subsequent change in the underlying Granger causality. For low model order, both CGC and GCCA have low error in estimating the underlying bivariate causality. In those cases, GCCA shows less error for most of the simulations, but differences were overshadowed by the inaccuracy of CGC and GCCA to estimate GC as shown in Figure 4.

CGC provides, in addition to causality, the weights used to estimate the signals in each region related to that causality. In simulations, we saw that for larger sample size ( $N = 400$ ) and smaller model order ( $P = 2, 4, 6$ ), CGC estimated causality with similar accuracy to the causality using the correct (known) weights. In the causal network in the macaque, high weights were observed at  $S_3$  and  $P_2$ , consistent with previous analysis of this data. However, in the portion of the response related to the visuomotor network, the sensor  $S_1$  also had significant weight which is inconsistent with earlier findings (Liang et al., 2000).

CGC depends on predefined ROIs. In the macaque data, the LFP recordings include two sets of electrodes grouped in two well-defined anatomical ROIs: the striate and prestriate.

However, anatomical ROIs are not required for CGC; we can use decompositions such as principal components (PCA) or associations such as correlation to cluster recordings into distinct functional ROIs. Whether the ROIs are defined anatomically or functionally, CGC is able to more robustly detect causal interactions between ROIs without the need to specify which specific subregions are involved. Moreover, the weight vectors obtained during model fitting provide an estimate of the relative contribution of these subareas to the causal interaction.

In some cases, it might be of interest to apply CGC to ROIs that each contain a large number of signals. However, CGC should be used with caution in these settings. In particular, CGC is just like other canonical correlation-type methods, and requires that the number of time samples be much larger than the number of signals in each ROI to avoid degeneracies. If this condition is not satisfied, then CGC will spuriously estimate infinite causality in both directions. In our simulations, we used 8 signals, much smaller than the minimum of 100 samples simulated, to avoid this issue. In addition, the computational complexity of CGC is linear in the number of signals in each ROI, as shown in section 3. Hence, CGC may require a lot of runtime to find the optimal weights when measuring regional causality between ROIs.

With LFP signals we have shown CGC's ability to find inter-regional connectivity as well as signals of interest in a region. We can extend this work to other functional brain modalities such as electrocorticography (ECoG), electroencephalography (EEG) and magnetoencephalography (MEG). While the LFP signals are directly related to local neuronal populations, the sensors in ECoG, EEG and MEG are sensitive to much larger areas of neuronal activity. If two ROIs are widely separated then CGC should produce meaningful results for EEG and MEG data. However, in cases where there is significant crosstalk or linear mixing between ROIs, the causality measures can be compromised (Hui et al., 2010). In that case, it may be possible to modify the CGC approach to first remove linear mixing between ROIs (Brookes et al., 2011; Hipp et al., 2012; Soto et al., 2010) before computing causality.

CGC could also potentially be applied to reduce the effects of linear mixing in other modalities with poor spatial resolution like cortical current density mapping. Cortical current density map ROIs typically contain many vertices, so a dimensionality reduction step would need to be used before applying CGC. With appropriate preprocessing and data reduction, then, regional measures such as CGC would potentially model dynamic causality on the brain by incorporating the distributed nature of cortical current density maps through weighted sums.

Currently, since we do not know the distribution of CGC under the assumption of no causality, we have used resampling procedures such as permutation testing or bootstraps to test for significant CGC. However, it should be noted that canonical correlation, through its reformulation as a maximum eigenvalue problem, has an associated parametric null distribution that can be derived using random matrix theory (Mardia et al., 1980). It might be possible to use similar techniques to derive a null distribution for CGC, though our attempts to do this have not yet been successful.



The CGC model presented here can be extended to the frequency domain methods of finding causality (Geweke, 1982; Kaminski et al., 2001). By computing causality at different frequencies, we may be able to discover significant causality over restricted frequency bands when there is no significant overall causality. In many cases, it has been found that causal interactions are localized in frequency (Lin et al., 2009; Wilke et al., 2009; Wu et al., 2008); however, frequency measures are difficult for all but the simplest of cases as it is hard to find a frequency domain measure that is both interpretable and can be summarized easily by the original measure (Chicharro, 2012). Another extension of CGC would allow more than one interaction between the two ROIs by computing two more sets of weight vectors; hence a measure of causality could be determined that lies somewhere between CGC and MGC. This extension may better approximate causal relations in the brain, where it may be expected that multiple signals in one ROI, none uniquely determined by any channel in that ROI, are interacting with multiple signals in a second ROI.

In this paper, CGC is proposed as a bivariate model: the only signals being analyzed are those in the two ROIs. If there are other signals that may be driving or influencing signals from the two ROIs, we can attempt to remove the influences of these other signals by simple linear regression (Guo et al., 2008), as was done in partial GCCA (Wu et al., 2011) and partial MGC (Barrett & Seth, 2010). However, removing the effects of all signals not in the two ROIs can increase the variance in causality estimation. Often, when computing partial causality, it is thus prudent to use some data reduction (e.g. principal components) on all signals not in the two ROIs before calculating partial causality (Zhu et al., 2009). A similar data reduction approach may have to be incorporated into a partial CGC formulation.

In addition, partial causality measures suffer from the artifact of redundancy (Angelini et al., 2010). Redundancy is the presence of the same signal of interest in multiple recordings. CGC is different from partial causality measures, in that it responds to redundancy in a different way. Redundancy refers to the presence of the same signal of interest in multiple signals, and leads to spuriously low causality in methods like partial GCCA and partial MGC when the shared signal is removed during the regression process. CGC does not regress away redundant signals of interest and will not suffer from spuriously low causality if there is shared information between signals from the same ROI. However, CGC can suffer from spuriously high causality if the same signal of interest is present in both ROIs.

## 7. Conclusion

The analysis of brain imaging data often focuses in anatomically or functionally defined ROIs, rather than discrete cortical locations. However, most existing methods of causal functional connectivity analysis from electrophysiological data focus on interactions between individual pairs of signals or multivariate models of interaction between all electrodes. This paper presents a novel measure, CGC, that models the causal interaction between ROIs by borrowing ideas from canonical correlation and Granger causality. With this new measure, causality between ROIs is modeled through weighted sums of the signals in the individual ROIs. We also presented a numerical method to determine the weights and corresponding causality from one ROI to another. In simulation, we demonstrated that CGC can effectively differentiate both the presence and strength of causality for complex model



order, moderate noise levels and low data sizes. In an application to macaque LFP data from a visuomotor task, we demonstrated CGC's ability to determine dynamic differences in causality in the occipital lobe corresponding to the type of visual stimulus and visuomotor response.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Dr. Steven Bressler for providing the macaque data. This work supported by NIH under grants R01 EB009048, R01 EB000473, T32 EB00438, and NSF grant BCS-1028389.

## Appendix A: Numerical Implementation

To simplify notation in the sequel, we define the vector  $\mathbf{w}$  from the set  $\mathcal{S}$  as:

$$\mathbf{w} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \in \mathcal{S} \triangleq \mathcal{S}^{M_1-1} \times \mathcal{S}^{M_2-1}. \quad (\text{A.1})$$

We assume that we have the ability to compute the unconstrained gradient (i.e., ignoring the unit-norm constraints) of the CGC measure in (12),  $G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1}$ , with respect to  $\mathbf{w}$ :

$$\nabla_{\mathbf{w}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} = \begin{bmatrix} \nabla_{\boldsymbol{\alpha}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} \\ \nabla_{\boldsymbol{\beta}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} \end{bmatrix} \quad (\text{A.2})$$

where the subscripts indicate the components of the gradient corresponding to the weight vectors  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . In our implementation, this gradient computation is performed numerically using an adaptive finite-difference method, with the AR coefficients for GC estimated using a modified version of Burg's algorithm as described by de Waele & Broersen (2003).

Due to the properties of  $\mathcal{S}$ , it can be shown (Absil et al., 2008) that the constrained gradient of  $G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1}$  with respect to  $\mathcal{S}$  is a projection of the unconstrained gradient:

$$\nabla_{\mathcal{S}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} = \begin{bmatrix} (\mathbf{I} - \boldsymbol{\alpha} \boldsymbol{\alpha}^T) \nabla_{\boldsymbol{\alpha}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} \\ (\mathbf{I} - \boldsymbol{\beta} \boldsymbol{\beta}^T) \nabla_{\boldsymbol{\beta}} G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1} \end{bmatrix} \quad (\text{A.3})$$

We use a nonlinear conjugate gradient algorithm with Polak-Ribière updates (Edelman et al., 1998) to maximize  $G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1}$ . The algorithm requires the specification of two termination criterion parameters ( $\epsilon$  for the weights and  $\delta$  for the value of  $G_{\boldsymbol{\beta}^T \mathbf{y}_2 \rightarrow \boldsymbol{\alpha}^T \mathbf{y}_1}$ ), a

conjugate gradient restart parameter  $Q$ , and an initial point  $\mathbf{w}_0 = \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix} \in \mathcal{S}$  (which we choose randomly).

## Initialization

Let  $\mathbf{g}_0 = \nabla_{\mathbf{w}_0}^{\mathcal{S}} G_{\beta^T y_2 \rightarrow \alpha^T y_1}$  be the initial gradient and  $\mathbf{d}_0 = \begin{bmatrix} \mathbf{d}_0^\alpha \\ \mathbf{d}_0^\beta \end{bmatrix} = \mathbf{g}_0$  be the initial ascent direction, where we have split  $\mathbf{d}_0$  into two components of length  $M_1$  and  $M_2$  corresponding to the weight vectors  $\alpha$  and  $\beta$ .

## Main Loop

For iteration  $k = 1, \dots, Q$ , we optimize over successive search paths in  $\mathcal{S}$ :

### Search Path

Form the one-dimensional curve

$$\mathbf{w}(t) = \begin{bmatrix} \alpha_{k-1} \cos(\|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t) + \frac{\mathbf{d}_{k-1}^\alpha}{\|\mathbf{d}_{k-1}^\alpha\|_{\ell_2}} \sin(\|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t) \\ \beta_{k-1} \cos(\|\mathbf{d}_{k-1}^\beta\|_{\ell_2} t) + \frac{\mathbf{d}_{k-1}^\beta}{\|\mathbf{d}_{k-1}^\beta\|_{\ell_2}} \sin(\|\mathbf{d}_{k-1}^\beta\|_{\ell_2} t) \end{bmatrix} \quad (\text{A.4})$$

The normalization of the two components of the gradient is required to allow the entire path to lie in  $\mathcal{S}$ . Geometrically, the path represents the traversal of great circles around the two spheres in  $\mathcal{S}$ , starting at  $\mathbf{w}_{k-1}$  and moving along the direction  $\mathbf{d}_{k-1}$ .

### Search

Find the maximal value of  $G_{\beta^T y_2 \rightarrow \alpha^T y_1}$  over the search path using the golden section search (Kiefer, 1953), with the location of the maximum given by  $\hat{t}$ .

### Update

1. Update  $\mathbf{w}_k = \mathbf{w}(\hat{t})$ .
2. Update  $\mathbf{g}_k = \nabla_{\mathbf{w}_k}^{\mathcal{S}} G_{\beta^T y_2 \rightarrow \alpha^T y_1}$ .
3. Define the step size  $\gamma_k$  (Polak & Ribière, 1969) using the modified gradient  $\bar{\mathbf{g}}_{k-1}$ ,

$$\gamma_k = (\mathbf{g}_k - \bar{\mathbf{g}}_{k-1})^T \mathbf{g}_{k-1} \quad (\text{A.5})$$

$$\bar{\mathbf{g}}_{k-1} = \begin{bmatrix} \mathbf{I} - \alpha_{k-1} \alpha_{k-1}^T & 0 \\ 0 & \mathbf{I} - \beta_{k-1} \beta_{k-1}^T \end{bmatrix} \mathbf{g}_{k-1} \quad (\text{A.6})$$

4. Update the search direction  $\mathbf{d}_k$  using the transported previous search direction  $\bar{\mathbf{d}}_{k-1}$ :

$$\mathbf{d}_k = \mathbf{g}_k + \gamma_k \bar{\mathbf{d}}_{k-1} \quad (\text{A.7})$$

$$\bar{\mathbf{d}}_{k-1} = \begin{bmatrix} -\alpha_{k-1} \sin \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t \right) + \frac{\mathbf{d}_{k-1}^\alpha}{\|\mathbf{d}_{k-1}^\alpha\|_{\ell_2}} \cos \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t \right) \\ -\beta_{k-1} \sin \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t \right) + \frac{\mathbf{d}_{k-1}^\beta}{\|\mathbf{d}_{k-1}^\beta\|_{\ell_2}} \cos \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t \right) \end{bmatrix} \quad (\text{A.8})$$

### Termination

Exit the process if

$$\|\mathbf{w}_k - \mathbf{w}_{k-1}\|_{\ell_2} < \varepsilon \quad (\text{A.9})$$

or

$$G_{\beta_k^T y_2 \rightarrow \alpha_k^T y_1} - G_{\beta_{k-1}^T y_2 \rightarrow \alpha_{k-1}^T y_1} < \delta. \quad (\text{A.10})$$

### Repeat

Go to iteration  $k + 1$

### Reset

If we reach the last iteration in the loop, so that  $k = Q$ , and neither of the termination criterion are satisfied, set  $k = 0$ ,  $\alpha_0 = \alpha_Q$ , and  $\beta_0 = \beta_Q$ , and repeat the algorithm from the beginning.

By using this modified version of the nonlinear conjugate gradient method designed for the constraint set  $\mathcal{S}$ , the algorithm converges faster than the standard conjugate Sgradient descent method (Edelman et al., 1998). We observed that setting  $\varepsilon = \delta = 10^{-6}$  and  $Q = M_1 + M_2 - 2$  allowed the optimization to converge. We also observed a substantial decrease in algorithm runtime by modifying the search path in each iteration from a one-dimensional curve (dependent on  $t$ ) to a two-dimensional curve dependent on  $t_1$  and  $t_2$ :

$$\mathbf{w}(t_1, t_2) = \begin{bmatrix} \alpha_{k-1} \cos \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t_1 \right) + \frac{\mathbf{d}_{k-1}^\alpha}{\|\mathbf{d}_{k-1}^\alpha\|_{\ell_2}} \sin \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t_1 \right) \\ \beta_{k-1} \cos \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t_2 \right) + \frac{\mathbf{d}_{k-1}^\beta}{\|\mathbf{d}_{k-1}^\beta\|_{\ell_2}} \sin \left( \|\mathbf{d}_{k-1}^\alpha\|_{\ell_2} t_2 \right) \end{bmatrix} \quad (\text{A.11})$$

and then optimizing over this curve first in  $t_1$  and then in  $t_2$ , in each case using a golden section search.

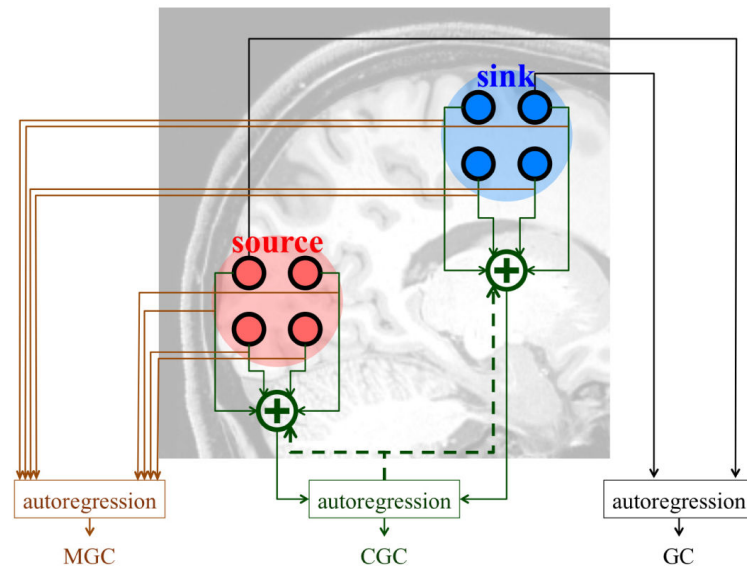
Note that the cost function for CGC is non-convex because reversing the sign of either set of weights does not affect the variances in the Granger causality computation. This sign ambiguity leads to equivalent solutions for the optimal weights in CGC. However, in simulation we found that this optimization algorithm consistently converges to the same maximum found via a multiresolution grid search.

## References

- Absil, P-A.; Mahony, R.; Sepulchre, R. Optimization Methods on Matrix Manifolds. Princeton University Press; Princeton, NJ: 2008.
- Angelini L, de Tommaso M, Marinazzo D, Nitti L, Pellicoro M, Stramaglia S. Redundant variables and Granger causality. *Phys Rev E*. 2010; 81:037201.
- Ashrafulla, S.; Haldar, JP.; Joshi, AA.; Leahy, RM. Proc IEEE ISBI. IEEE; Barcelona, Spain: 2012. Canonical Granger causality applied to functional brain data; p. 1751-1754.
- Astolfi L, Cincotti F, Mattia D, Marciani MG, Baccalá LA, de Vico Fallani F, Salinari S, Ursino M, Zavaglia M, Ding L, Edgar JC, Miller GA, He B, Babiloni F. Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *HBM*. 2007; 28:143–57.
- Barnett L, Seth AK. Behaviour of Granger causality under filtering: theoretical invariance and practical application. *J Neuro Methods*. 2011; 201:404–19.
- Barrett AB, Seth AK. Multivariate Granger causality and generalized variance. *Phys Rev E*. 2010; 81:1–14.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Society B*. 1995; 57:289–300.
- Bin G, Gao X, Yan Z, Hong B, Gao S. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *J Neural Engr*. 2009; 6:1–6.
- Bressler SL, Coppola R, Nakamura R. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature*. 1993; 366:153–6. [PubMed: 8232553]
- Bressler SL, Ding M, Yang W. Investigation of cooperative cortical dynamics by multivariate autoregressive modeling of event-related local field potentials. *Neurocomputing*. 1999; 26-27:625–631.
- Bressler SL, Richter CG, Chen Y, Ding M. Cortical functional network organization from autoregressive modeling of local field potential oscillations. *Statistics in Medicine*. 2007; 26:3875–85. [PubMed: 17551946]
- Brookes MJ, Woolrich M, Luckhoo H, Price D, Hale JR, Stephenson MC, Barnes GR, Smith SM, Morris PG. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *PNAS*. 2011; 108:16783–8. [PubMed: 21930901]
- Chen Y, Bressler SL, Ding M. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J Neuro Methods*. 2006; 150:228–37.
- Cheung BLP, Riedner BA, Tononi G, van Veen BD. Estimation of cortical connectivity from EEG using state-space models. *IEEE Trans Biomed Engr*. 2010; 57:2122–34.
- Chicharro D. On the spectral formulation of Granger causality. *Biol Cyber*. 2012:331–347.
- Correa NM, Adali T, Li Y-O, Calhoun VD. Canonical correlation analysis for data fusion and group inferences: Examining applications of medical imaging data. *IEEE Sig Proc*. 2010; 27:39–50.
- d'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echaz J, Litt B. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Trans Biomed Engr*. 2003; 50:603–615.
- Deleus F, Van Hulle MM. Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis. *J Neuro Methods*. 2011; 197:143–57.
- Ding M, Bressler SL, 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 Cyber*. 2000; 83:35–45.
- Edelman A, Arias TA, Smith ST. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*. 1998; 20:303–353.
- Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. Chapman and Hall; Boca Raton, FL, USA: 1993.
- Eichler M. On the evaluation of information flow in multivariate systems by the directed transfer function. *Biol Cyber*. 2006; 94:469–82.
- Florin E, Gross J, Pfeifer J, Fink GR, Timmermann L. The effect of filtering on Granger causality based multivariate causality measures. *NeuroImage*. 2010; 50:577–88. [PubMed: 20026279]

- Geweke J. Measurement of linear dependence and feedback between multiple time series. *J Amer Stat Assoc.* 1982; 77:304–313.
- Goldberg JA, Boraud T, Maraton S, Haber SN, Vaadia E, Bergman H. Enhanced synchrony among primary motor cortex neurons in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine primate model of Parkinson's disease. *J Neurosci.* 2002; 22:4639–53. [PubMed: 12040070]
- Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica.* 1969; 37:424.
- Guo S, Seth AK, Kendrick KM, Zhou C, Feng J. Partial Granger causality—eliminating exogenous inputs and latent variables. *J Neuro Methods.* 2008; 172:79–93.
- Hesse W, Moller E, Arnold M, Schack B. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *J Neuro Methods.* 2003; 124:27–44.
- Hipp JF, Hawellek DJ, Corbetta M, Siegel M, Engel AK. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuro.* 2012; 15:884–90.
- Hotelling H. Relations between two sets of variables. *Biometrika.* 1936; 28:321–377.
- Hui HB, Pantazis D, Bressler SL, Leahy RM. Identifying true cortical interactions in MEG using the nulling beamformer. *NeuroImage.* 2010; 49:3161–74. [PubMed: 19896541]
- Jirsa V.; McIntosh, AR. *Handbook of Brain Connectivity.* Springer; 2007.
- Kaminski MJ, Ding M, Truccolo WA, Bressler SL. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cyber.* 2001; 85:145–157.
- Kiefer JC. Sequential minimax search for a maximum. *Proc AMS.* 1953; 4:502–506.
- Ku R, Kaminski MJ, Blinowska KJ. Determination of EEG activity propagation: pair-wise versus multichannel estimate. *IEEE Trans Biomed Engr.* 2004; 51:1501–10.
- Kuylen AAA, Verhallen TMM. The use of canonical analysis. *J Econ Psych.* 1981; 1:217–237.
- Liang H, Ding M, Nakamura R, Bressler SL. Causal influences in primate cerebral cortex during visual pattern discrimination. *Neuroreport.* 2000; 11:2875–80. [PubMed: 11006957]
- Lin F-H, Hara K, Solo V, Vangel M, Belliveau JW, Stueben SM, Härmäläinen MS. Dynamic Granger-Geweke causality modeling with application to interictal spike propagation. *HBM.* 2009; 30:1877–86.
- Lock TM, Baizer JS, Bender DB. Distribution of corticotectal cells in macaque. *Experimental Brain Research.* 2003; 151:455–70. [PubMed: 12851806]
- Mardia, KV.; Kent, JT.; M, BJ. *Multivariate Analysis.* 1st ed. Academic Press; San Diego, CA: 1980.
- Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine.* 1978; 8:283–298. [PubMed: 112681]
- Mishkin M, Ungerleider LG, Macko KA. Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences.* 1983; 6:414–417.
- Polak E, Ribière G. Note sur la convergence de méthodes de directions conjuguées. *Revue.* 1969; 3:35–43.
- Sato JR, Fujita A, Cardoso EF, Thomaz CE, Brammer MJ, Amaro E. Analyzing the connectivity between regions of interest: an approach based on cluster Granger causality for fMRI data analysis. *NeuroImage.* 2010; 52:1444–55. [PubMed: 20472076]
- Schoffelen J-M, Gross J. Source connectivity analysis with MEG and EEG. *HBM.* 2009; 30:1857–65.
- Sims C. Money, income and causality. *Amer Eco Rev.* 1972; 62:540–52.
- Soto, JLP.; Pantazis, D.; Jerbi, K.; Baillet, S.; Leahy, RM. *Proc IEEE ISBI.* IEEE; Rotterdam: 2010. Canonical correlation analysis applied to functional connectivity in MEG; p. 113-116.
- Turi G, Gotthardt S, Singer W, Vuong TA, Munk M, Wibral M. Quantifying additive evoked contributions to the event-related potential. *NeuroImage.* 2012; 59:2607–24. [PubMed: 21982933]
- Valdés-Sosa PA, Sánchez-Bornot JM, Lage-Castellanos A, Vega-Hernández M, Bosch-Bayard J, Melie-García L, Canales-Rodriguez E. Estimating brain functional connectivity with sparse multivariate autoregression. *Phil Trans B.* 2005; 360:969–81.
- de Waele S, Broersen PMT. Order selection for vector autoregressive models. *IEEE Trans Sig Proc.* 2003; 51:427–433.

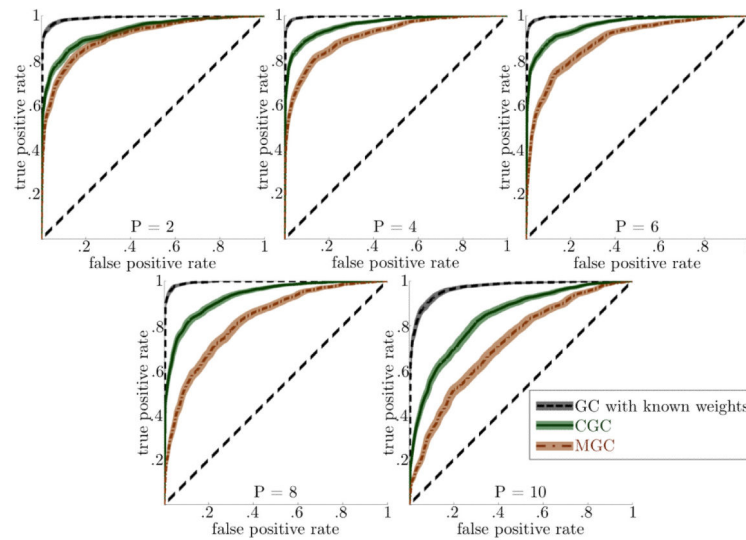
- Wang X, Chen Y, Bressler SL, Ding M. Granger causality between multiple interdependent neurobiological time series: blockwise versus pairwise methods. *Int J Neural Sys.* 2007; 17:71–8.
- Wang X, Chen Y, Ding M. Estimating Granger causality after stimulus onset: a cautionary note. *NeuroImage.* 2008; 41:767–76. [PubMed: 18455441]
- Wilke CT, van Drongelen W, Kohrman MH, He B. Identification of epileptogenic foci from causal analysis of ECoG interictal spike activity. *Clin Neuro.* 2009; 120:1449–56.
- Wu G, Chen F, Kang D, Zhang X, Marinazzo D, Chen H. Multiscale causal connectivity analysis by canonical correlation: theory and application to epileptic brain. *IEEE Trans Biomed Engr.* 2011; 58:3088–96.
- Wu J, Liu X, Feng J. Detecting causality between different frequencies. *J Neuro Methods.* 2008; 167:367–75.
- Zhou Z, Chen Y, Ding M, Wright P, Lu Z, Liu Y. Analyzing brain networks with PCA and conditional Granger causality. *HBM.* 2009; 30:2197–206.
- Zhu, X.; Ji, L.; Jin, X.; Yi, J. Proc 9th ICEMI. IEEE; 2009. Fitting and reconstruction of three-dimensional curve based on orthogonal curvature; p. 4–323–4–328.



**Figure 1.**

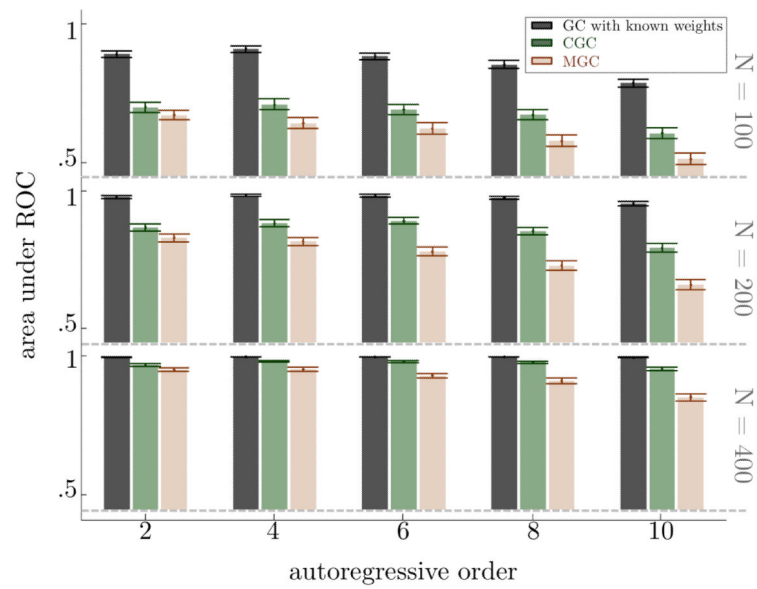
We want to estimate the causality from the source region to the sink region via the time series recordings (represented by dark circles) from each region. GC analyzes the causality between pairs of recordings using univariate and bivariate autoregressive models. MGC fits much larger parametric autoregressive models to collections of time series. CGC parsimoniously represents each ROI using a single signal formed through an optimized weighted linear combination of signals from the ROI. CGC then applies standard GC analysis to these parsimonious representations.





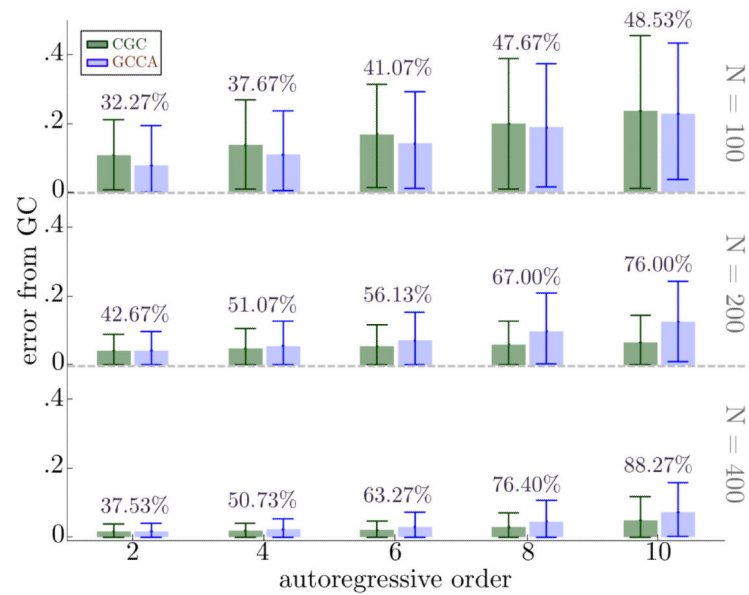
**Figure 2.**

CGC has a higher true positive rate for a given false positive rate than MGC. The dark lines are the mean ROCs, with the shadows indicating the 95% confidence intervals after bootstrapping. Knowing the true weights is better than CGC because CGC has an upward bias in the condition of no causality due to its maximization procedure, but such weights are not known in most practical applications.



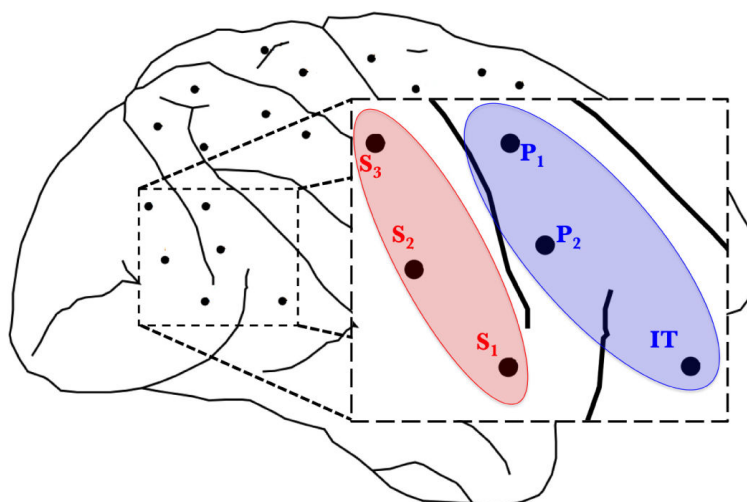
**Figure 3.**

CGC can better differentiate between causality and non-causality over MGC, and that difference is amplified as the autoregressive order increases. For each method, length of timeseries, and autoregressive order, the height of the bar corresponds to the mean AUC while the error bars describe the 95% confidence interval of the AUC.

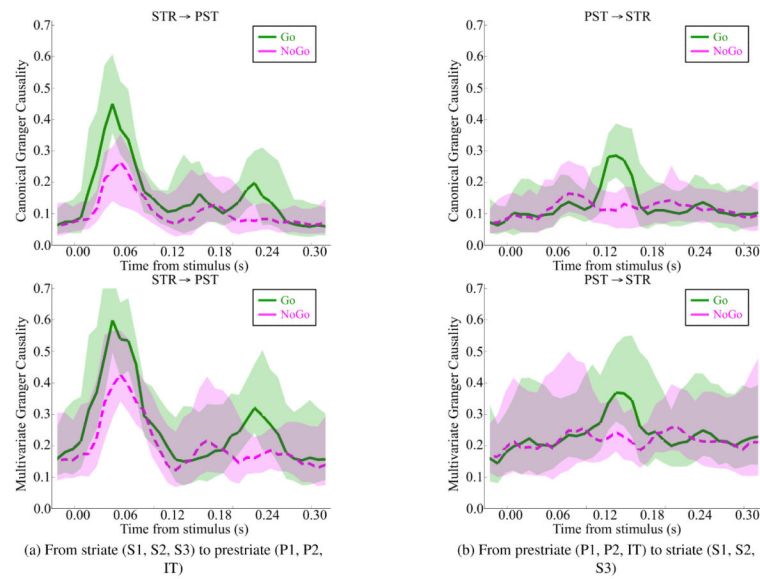


**Figure 4.**

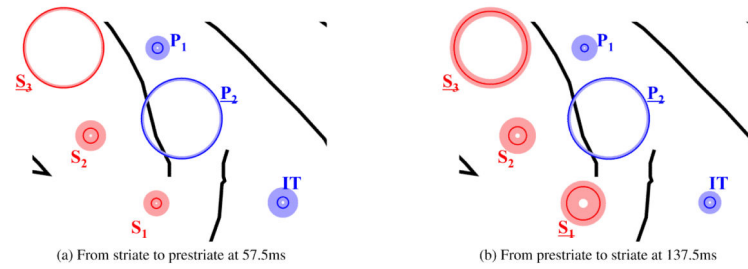
As the autoregressive order increases, for moderate timeseries lengths, CGC becomes much more adept at estimating the underlying true bivariate causality than GCCA. For each choice of method, length of timeseries  $N$ , and autoregressive order  $R$ , the bar is the mean error over simulations and the error bars describe the 95% confidence interval of error. For short timeseries, CGC seems to perform worse, but the mean value and variance of errors show that this drawback is minimal because both CGC and GCCA are equally inadequate at estimating the strength of the true causality.



**Figure 5.** Location of the 6 LFP channels in the monkey recorded during the experiment. We analyzed striate-prestriate causal interactions related to the stimulus: a visual cue to perform a motor task.

**Figure 6.**

Both CGC (top row) and MGC (bottom row) find causality at 57.5ms from striate to prestriate (left column), but CGC finds a statistically significant difference in causality between the two conditions. In addition, both CGC and MGC find some causality at 137.5ms from prestriate to striate (right column) under the “Go” condition, though only CGC finds a statistically significant difference in causality between the two conditions. This prestriate to striate causal interaction may be part of visuomotor processing, as “Go” requires visuomotor interaction while “NoGo” only requires visual processing.



**Figure 7.**

Most of the causality in both timepoints of interest are driven by the relationship between two recording sites:  $S_3$  and  $P_2$ . The main plot shows the relative magnitude of the weight estimated for each channel; the larger the circle radius, the larger the magnitude of the weight for that channel. The dark line shows the mean magnitude of the weight over the full set of trials, while the shaded region shows the bootstrapped confidence interval. The analysis also indicates that  $S_1$  has a higher weight in causality from prestriate to striate at 137.5ms than from striate to prestriate at 57.5ms.