Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling

Anil K. Seth*, Paul Chorley, Lionel C. Barnett

Sackler Centre for Consciousness Science and Department of Informatics,
University of Sussex, Brighton BN1 9QJ, UK
**a.k.seth@sussex.ac.uk (correspondence)*
www.sussex.ac.uk/sackler

Abstract

Granger causality is a method for identifying directed functional connectivity based on time series analysis of precedence and predictability. The method has been applied widely in neuroscience, however its application to functional MRI data has been particularly controversial, largely because of the suspicion that Granger causal inferences might be easily confounded by inter-regional differences in the hemodynamic response function. Here, we show both theoretically and in a range of simulations, that Granger causal inferences are in fact robust to a wide variety of changes in hemodynamic response properties, including notably their time-to-peak. However, when these changes are accompanied by severe downsampling, and/or excessive measurement noise, as is typical for current fMRI data, incorrect inferences can still be drawn. Our results have important implications for the ongoing debate about lag-based analyses of functional connectivity. Our methods, which include detailed spiking neuronal models coupled to biophysically realistic hemodynamic observation models, provide an important 'analysis-agnostic' platform for evaluating functional and effective connectivity methods.

Keywords: Granger causality, functional connectivity, functional MRI, hemodynamic response function, computational modelling

1. Introduction

Granger causality (GC) is a widely used method for identifying directed functional ('causal') connectivity in neural time series data, a key challenge for contemporary neuroscience (Bressler and Seth, 2011; Valdes-Sosa et al., 2011; Bressler and Menon, 2010). Introduced conceptually by Wiener (1956), and operationalized using linear autoregressive modelling of stochastic processes by Granger (1969) and Geweke (1982), GC is based on predictability and precedence in time-series data. The core concept is that a variable **X** is said to 'Granger-cause' a variable **Y** if the past of **X** contains information useful for predicting the future of **Y**, over and above that information already available in the past of **Y** (as well as other conditioning variables **Z**). (Throughout this paper we follow the notational conventions that bold symbols denote vector (multivariate) quantities and upper-case symbols denote either random variables or matrices, according to context.)

Over the last 20 years, GC has emerged as a popular method for analyzing neural time series obtained from many modalities including magneto/encephalography (M/EEG) (Barrett et al., 2012; Cohen and van Gaal, 2012; Gow et al., 2008; Kaminski et al., 2001), functional magnetic resonance imaging (fMRI) (Hwang et al., 2010; Bressler et al., 2008; Wen et al., 2012; Roebroeck et al., 2011b, 2005; Goebel et al., 2003), invasively obtained local-field potentials (LFPs) (Brovelli et al., 2004; Gaillard et al., 2009), and spike train data (Cadotte et al., 2008). In this paper, we focus on the application of GC to fMRI data (GC-fMRI), which is at once the most popular and the most controversial application domain (David et al., 2008; Roebroeck et al., 2011a; Friston, 2009). Although highly promising, GC-fMRI faces a number of challenges which have not yet been adequately addressed. Prominent among these are (i) the fact that the fMRI BOLD signal (as captured by the 'hemodynamic response function', HRF) is an indirect, sluggish, variable (inter-regionally and inter-subjectively) and incompletely understood reflection of the underlying neural activity (Logothetis et al., 2001; Handwerker et al., 2012; Magri et al., 2012), and (ii) the constraint that fMRI protocols involve severe downsampling with sample intervals (repetition times, TRs) typically ranging from 1-3 sec, substantially longer than typical inter-neuron delays. The challenge for GC-fMRI is that these factors may disturb information about precedence and predictability upon which GC analysis depends. In particular, inter-regional variation in the HRF would seem to present a potentially fatal confound to GC analysis: if **X** is causally driving **Y** at the neural level, but if the BOLD response to **X** peaks later than the BOLD response to **Y**, the suspicion is that GC analysis would falsely infer that **Y** is driving **X** (David et al., 2008). Unfortunately, inter-regional variation in HRFs are known to exist as a result of vasculature differences, baseline cerebral bloodflow, pulse or respiration differences, and other factors (Aguirre et al., 1998; Handwerker et al., 2004; Chang et al., 2008; Handwerker et al., 2012).

Here, we examine these issues using a combination of theoretical analysis and simulation modelling. Contrary to prevailing views (David et al., 2008; Friston, 2009; Smith et al., 2011), and extending previous simulation work showing limited robustness to HRF variation (Deshpande et al., 2010; Schippers et al., 2011), we show that GC-fMRI is analytically invariant to convolution of a neural signal by an HRF, even when HRF latencies apparently confound the direction of the underlying neural signal. Our results are based on considering hemodynamic responses as low-pass filters applied to neural signals, and we take advantage of previous work in which we show analytically that GC is invariant under a broad class of filtering operations (Barnett and Seth, 2011). We examine the empirical implications of this theoretical result using both simple vector autoregressive (VAR) models and biophysically detailed models of spiking neuron populations to generate simulated neural activity. To generate simulated BOLD signals we utilize both simple convolutions implementing a low-pass filter (difference-of-gamma approximation (Friston, 2006)) as well as the biophysically detailed extended Balloon-Windkessel (BW) model implementing both neurovascular coupling and blood-flow to BOLD mapping (Buxton et al., 1998; Friston et al., 2000; Friston, 2006). In each case we demonstrate the invariance of GC analysis to hemodynamic convolutions under a wide range of simulated hemodynamic responses including those which confound the underlying neural influence. We further show that, despite these invariance properties, GC-fMRI can still lead to missed and spurious causalities as a result of downsampling and measurement noise (these effects cannot be considered as invertible filters; see Sections 3.4 and 3.5). Effects of hemodynamic filtering on inference of statistical significance of GC values are also discussed.

Taken together, our results mandate refocusing the debate surrounding lag-based functional connectivity methods in fMRI to mitigation of noise and downsampling, as a complementary goal to deconvolution of hemodynamic responses (Roebroeck et al., 2011a). Furthermore, our ‘full’ model incorporating both spiking neurons and the BW hemodynamic model provides a uniquely detailed generative model for testing and comparing connectivity analysis methods in neuroimaging, and by doing so avoids potential circularities induced by using the same model class (e.g., a linear autoregressive model, or the neural model utilized by dynamic causal modelling (DCM, (Friston et al., 2003))) for both generating and analysing simulated data.

2. Materials and Methods

We first describe the generative models for simulated neural activity, both using a simple VAR model and a more detailed spiking neuron platform. We then describe the generation of simulated BOLD signals from the simulated neural activity, again using both simple (difference-of-gamma) and more detailed (Balloon-Windkessel, BW) approaches. There follows a summary of GC and its practical application in the context of these models. Additional methodological details are given in Appendices A and B.

2.1. VAR generative model of neural responses

We first tested GC-fMRI on data from a simple VAR generative model. The model is defined by (Baccalá and Sameshima, 2001):

$$
\begin{aligned}
x_1(t) &= 0.95\sqrt{2}x_1(t-l) - 0.9025x_1(t-2l) + w_1(t)\\
x_2(t) &= 0.5x_1(t-2l) + w_2(t)\\
x_3(t) &= -0.4x_1(t-3l) + w_3(t)\\
x_4(t) &= -0.5x_1(t-2l) + 0.25\sqrt{2}x_4(t-l) + 0.25\sqrt{2}x_5(t-l) + w_4(t)\\
x_5(t) &= -0.25\sqrt{2}x_4(t-l) + 0.25\sqrt{2}x_5(t-l) + w_5(t) \qquad (1)
\end{aligned}
$$

and its connectivity structure is shown in Figure 1A; l determines the neural delay relative to the model update rate. This network provides a nontrivial test for GC analysis, since it includes both reciprocal connectivity (between nodes 4 and 5) and distinctions between direct and indirect connections (e.g., node 5 is directly driven by node 4 only but is also indirectly related to activity in nodes 2 and 3 via a common influence from node 1). Generating dynamics, we assume a timestep of 1 ms and generate 200 sec of data, of which we discard the first 50 sec prior to analysis to amply allow for any numerical 'burn-in' effects. We set $l = 20$ leading to neural delays in the range $[20, 60]$ ms, which represents a plausible neurobiological spread (Schmolesky et al., 1998; Smith et al., 2011; Schippers et al., 2011). To generate simulated BOLD signals the VAR model output is convolved with five HRF kernels generated using the difference-of-gamma approach used within SPM8[1] (see Figure 1B and below). Note that HRF times-to-peak are deliberately confounded with the underlying neural delays. In particular, node 1 (in blue) drives nodes 2, 3, and 4 at the neural level, however its HRF peaks 2-5 *sec* after the HRFs of its targets, a difference considerably

[1] http://www.fil.ion.ucl.ac.uk/spm/software/spm8/

longer than the corresponding neural delays and representing a confound worse than might be expected in neurotypical HRF variance (Handwerker et al., 2004). BOLD signals are also generated by implementing the biomechanically detailed BW model (see below). Figure 1D shows examples of the resulting time-series data that are then submitted for GC analysis. All data were subjected to downsampling, either at high frequency (250 Hz) or at low frequency (0.5 Hz, representing a TR of 2 sec in fMRI). All combinations of downsampling (light, 250 Hz and heavy, 0.5 Hz) and BOLD generation (none, difference-of-gamma, and BW) were tested, giving rise to a matrix of simulated observables including LFP/EEG, a typical fMRI signal ($\mathrm{BOLD}_{0.5}$), as well as fast-sampled BOLD signal (BOLD_{250}). The various data types generated are summarized in Table 1.

	light (250 Hz)	heavy (0.5 Hz)
none	EEG/LFP	DOWN
conv	BOLD_{250}	$\mathrm{BOLD}_{0.5}$
BW	BOLD_{250}	$\mathrm{BOLD}_{0.5}$

Table 1: Combinations of downsampling (columns) and hemodynamic model (rows). Light downsampling of VAR model output (250 Hz) simulates either EEG/LFP data or a fast-sampled BOLD signal (such as in near-infrared spectroscopy). Heavy downsampling (0.5 Hz) simulates typical fMRI signals (TR = 2 sec) or very sparsely sampled LFP/EEG data (DOWN). conv and BW represent different hemodynamic models (difference-of-gamma, BW model respectively).

2.2. Spiking neuronal model

Selected simulations employed a biophysically detailed model of neural activity based on interconnected populations of spiking neurons. Full details of this model are presented in Appendix A; here we give an outline. The model consists of two clusters (X,Y) of spiking neurons (Izhikevich, 2003a) with 6144 neurons in each cluster, and incorporating a total of about 19 million synapses. Synapses are modelled with explicit NMDA, AMPA, GABAa, and GABAb conductances and short-term synaptic plasticity (Dayan and Abbott, 2005). Inputs to cluster X originate only from cluster X, whereas inputs to cluster Y originated from both clusters, specifying a simple functional architecture (X $\rightarrow$ Y) as shown in Figure 2A.

On each instantiation of the model, axonal conductance delays are drawn randomly from the uniform distributions of [1,10] ms for intra-cluster connections and [40,50] ms for inter-cluster connections. Background noise and synaptic weights are implemented within physiological ranges. Simulated

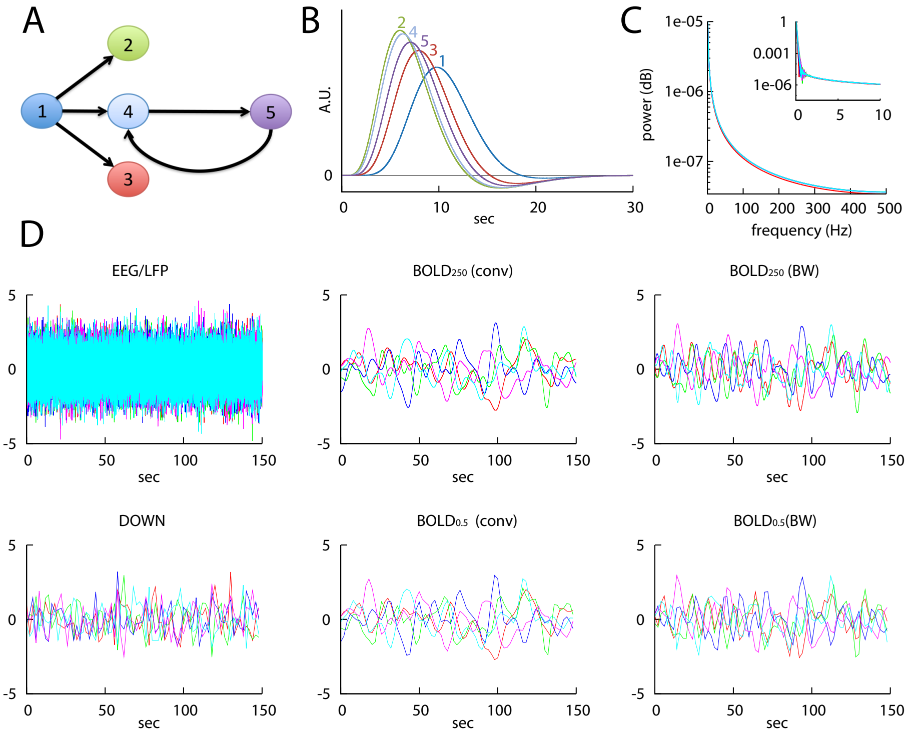

Figure 1: A. The connectivity of the simple VAR generative model; internode ‘neural’ delays are set to 20 ms and the VAR model updates at 1 ms. B. HRF kernels, modelled as the difference between two gamma functions (arbitrary units). The slope is varied in order to achieve HRF shapes that differ in particular with respect to their time-to-peak. Note that time-to-peaks are confounded with the underlying neural delays. C. Frequency-power response of HRF convolutions considered as FIR filters. Inset: detail of frequency range $0-10$ Hz, highlighting steep low frequency roll-off (the ‘smearing’ is due to floating-point inaccuracy). D. Time-series generated by the model. The EEG/LFP time-series is obtained by downsampling the VAR model output to 250 Hz. The BOLD_{250}(conv) timeseries are generated by convolving the VAR output with the corresponding HRF kernel while the BOLD_{250}(BW) timeseries are generated from the VAR output by the Balloon-Windkessel model; both are again downsampled to 250 Hz. The $\text{BOLD}_{0.5}$ time-series apply the same convolution/BW models, but now downsampled to 0.5 Hz, corresponding to a TR of 2 sec. For comparison, the DOWN timeseries downsample the VAR model output to 0.5 Hz but without a BOLD model. Outputs from all 5 nodes are superimposed.

local field potentials (LFPs) for each cluster are obtained at each time-step as the summed AMPA conductances across all neurons within each cluster (Ching et al., 2010). Simulated BOLD responses are obtained by feeding the

LFP signal into either the BW model or the difference-of-gamma approximation (see below). Figure 2B shows example neuronal and LFP output from the model for a period of 2 sec.

Note that the clusters in this model represent local neuronal populations and are not intended to explicitly model fMRI units of analysis such as voxels or regions-of-interest. Note also that the motivation for this model is to examine a biophysically detailed generative model of the BOLD signal for which a minimal two-cluster network suffices (as compared to the VAR model described above).

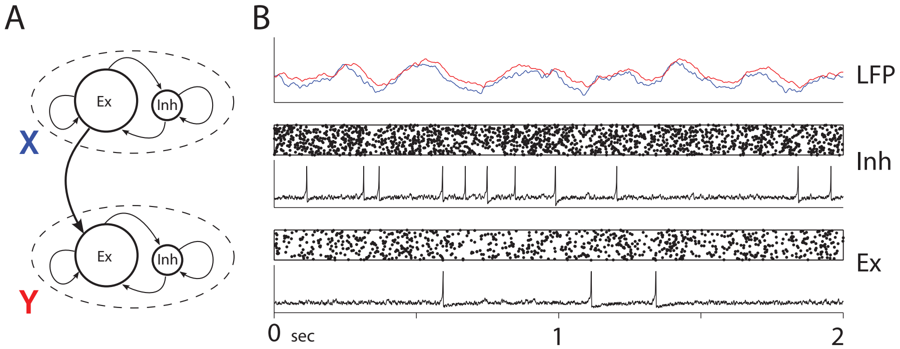

Figure 2: Spiking neuron model. A. Network architecture comprising two clusters of excitatory (Ex) and inhibitory (Inh) neurons. B. Sample output (2 sec) showing simulated LFP signals (top; X in blue, Y in red), and neural responses of Inh (middle) and Ex (bottom) neurons in cluster X. Neural responses are shown as both raster plots of spike timings and representative single-unit membrane potentials.

2.3. Modelling of hemodynamic responses

The BOLD signal arises from a complex interplay among blood volume, blood flow, and increases in oxygen consumption related to neural activity (Roebroeck et al., 2011b; Logothetis, 2008; Buxton, 2012; Buxton et al., 1998). This interplay is reflected by the HRF. A first and simple approach to generating simulated BOLD signals involves a widely used 'canonical' HRF model based on the difference between two gamma functions (Figure 1B). This phenomenological model captures BOLD dynamics very effectively and is easily parameterizable to yield differences in time-to-peak. Here, we use the default parameter settings in SPM8 (which specify a total kernel length of 32 sec) and we modulate the parameter governing the slope of the initial gamma function to generate differences in time-to-peak (Figure 1B).

A second approach involves the biomechanically detailed Balloon-Windkessel (BW) model (Buxton et al., 1998; Friston et al., 2000) which in its extended form (Friston et al., 2000) includes a neurovascular component, linking neural activity to blood flow (rCBF) and a biomechanical hemodynamic component, linking rCBF to the observable BOLD signal. The extended BW model is more biophysically interpretable than the double-gamma approximation but involves a much larger parameter space. Following Friston et al. (2000) we implemented the extended BW model using the default settings provided within SPM8 (Friston (2006); see Table 2). In those simulations involving trial-by-trial jitter of BW model parameters we used standard deviations (also shown in Table 2) again drawn from the prior distributions specified within SPM8 (resting oxygen extraction and resting volume were not jittered). We emphasize that these time-series were generated by directly feeding the simulated neural activity (VAR or spiking model) into the BW model, and not by generating a convolution kernel.

parameter	description	default	jitter
κ	signal decay	0.65	0.0296
γ	autoregulation	0.41	0.0108
τ	transit time	0.98	0.3084
α	outflow exponent	0.32	0.0004
E_0	resting oxygen extraction	0.40	
V_0	resting volume	0.01	

Table 2: SPM8 default BW model parameters and standard deviations of jitter.

The extended BW model requires a neural signal as input. For the VAR model this is trivially the VAR model output, which we take to reflect LFP. For the spiking model several neural signals are available (e.g., LFP, multi-unit spiking activity). For simplicity and ease of comparison we chose to use the full simulated LFP (see Section 2.2). Accumulating evidence suggests that BOLD responses are more tightly coupled to synaptic activity than to spiking activity (Logothetis, 2008), though both may be predictive of BOLD activity in some cases (Rosa et al., 2011). Biophysical mechanisms have been suggested linking presynaptic activity to both LFP signals (via glutamate-induced fluctuations of postsynaptic membrane potentials) and BOLD responses (via triggering extracellular release of vasodilatory agents) (Logothetis, 2008; Rosa et al., 2011; Bonvento et al., 2002). While several studies have associated the BOLD signal preferentially with specific LFP frequencies (notably gamma-band power fluctuations (Logothetis et al., 2001; Murayama et al., 2010; Niessing et al., 2005)), many different LFP frequency bands appear to contribute to BOLD responses in complexly interacting ways (Magri et al., 2012; Scheeringa et al., 2011), and the global LFP response remains a strong predictor of BOLD activity (Magri et al., 2012).

2.4. Granger causality (GC)

Here we describe basic properties of GC relevant for the present study: further details are given in Appendix B. Consider two jointly distributed, possibly multivariate, covariance-stationary stochastic processes $\mathbf{X}_t, \mathbf{Y}_t$(i.e., "variables"), with vector autoregressive (VAR) models for $\mathbf{X}$:

$$\mathbf{X}_t = A_1\mathbf{X}_{t-1} + \ldots + A_p\mathbf{X}_{t-p} + B_1\mathbf{Y}_{t-1} + \ldots + B_p\mathbf{Y}_{t-p} + \varepsilon_t \tag{2}$$

$$\mathbf{X}_t = A'_1\mathbf{X}_{t-1} + \ldots + A'_p\mathbf{X}_{t-p} + \varepsilon'_t \tag{3}$$

where $\varepsilon_t, \varepsilon'_t$ are serially uncorrelated iid residuals (white noise), A_k, B_k, A'_k the VAR coefficients matrices and p the VAR model order. The GC from $\mathbf{Y}$ to $\mathbf{X}$ is then given by:

$$\mathcal{F}_{\mathbf{Y}\to\mathbf{X}} \equiv \ln\left(\frac{\det(\Sigma')}{\det(\Sigma)}\right) \tag{4}$$

where Σ, Σ' are estimators for the residuals covariance matrices $\operatorname{cov}(\varepsilon_t)\,, \operatorname{cov}(\varepsilon'_t)$ respectively. The *generalised variances* $\det(\Sigma)\,, \det(\Sigma')$ quantify the magnitude of the prediction errors of (2) and (3) respectively (Barrett et al., 2010). The quantity $\mathcal{F}_{\mathbf{Y}\to\mathbf{X}}$ may be read as the degree to which the past of $\mathbf{Y}$ improves prediction of $\mathbf{X}$ over and above the degree to which $\mathbf{X}$ is already predicted by its own past. Given a third (possibly multivariate) process $\mathbf{Z}_t$ jointly distributed with $\mathbf{X}_t, \mathbf{Y}_t$, any common causal influence of $\mathbf{Z}$ on $\mathbf{X}$ and $\mathbf{Y}$ may be conditioned out by appending p lags of $\mathbf{Z}_t$ to both regressions (2) and (3). The resulting *conditional GC* (Geweke, 1984) is written $\mathcal{F}_{\mathbf{Y}\to\mathbf{X}\,|\mathbf{Z}}$. A natural spectral decomposition of GC in the frequency domain is also available (Geweke, 1982, 1984), allowing causal effects to be isolated at distinct frequencies or frequency bands (Barnett and Seth, 2011, Section 3.2).

In practice, given empirical time series data, the model order p may be determined by standard techniques such as the Akaike or Bayesian information criteria, or cross-validation (McQuarrie and Tsai, 1998; Edgington, 1995).

2.4.1. Statistical inference

$\mathcal{F}_{\mathbf{Y}\rightarrow\mathbf{X}}$ asymptotically follows a χ^2 distribution, furnishing a simple significance test (see Appendix B). However, for this study we note an important caveat: results presented in Section 3.2 indicate that slow, moving-average (HRF-like) filters may render significance tests unreliable while GC magnitude estimates remained accurate. Importantly, GC magnitudes have meaningful information-theoretic interpretations, regardless of statistical inference: under Gaussian assumptions, GC is entirely equivalent to transfer entropy, an information-theoretic measure of directed (time-asymmetric) information flow between joint processes (Barnett et al., 2009; Schreiber, 2000; Kaiser and Schreiber, 2002), and may thus be considered an absolute informational quantity to be measured in bits. This interpretation justifies foregoing formal significance testing in (most of) our simulations and instead comparing GC magnitudes with the correct (i.e., *a priori* known) causal structure in the data. Practical heuristics for statistical inference are discussed further in Section 4.

2.4.2. Covariance stationarity and model validation

Wide-sense stationarity (i.e., that the mean and autocovariance of the data are constant over time) is a prerequisite for VAR modelling underlying the present GC analysis. Standard tests for stationarity, such as the Augmented Dickey-Fuller (ADF) test, Phillips-Perron test or the KPSS test (Hamilton, 1994) may be used to test a unit-root null hypothesis with drift or deterministic trending. Here we used the ADF and KPSS tests in all simulations; all time-series satisfactorily passed these checks at significance level $\alpha = 0.01$. We also performed a further unit-root check using the *spectral radius* of the (full) VAR model, which reflects how quickly autocorrelation decays in an VAR model, and which must be < 1 for a stationary system (see Appendix B). The spectral radius was calculated for all simulations; in a handful of cases a value of ≥ 1 was obtained due to statistical fluctuations; such simulations were discarded.

Model fit and consistency (i.e., how well the VAR models fit the data) was checked in three ways: first, a Durbin-Watson residuals whiteness test Durbin and Watson (1950) was performed on all data. Second, the adjusted R^2 statistic was calculated to test for the amount of variance accounted for by the (full) VAR model. Finally, the consistency statistic of Ding et al. (2000) was calculated to check for the proportion of the correlation structure accounted for by the VAR model. Results were satisfactory for all time-series.

2.4.3. Practical implementation

GC analysis was implemented using the 'Multivariate Granger causality' (MVGC) Matlab© toolbox[2]. The MVGC toolbox implements pairwise-conditional GC calculation as follows: after stationarity checks, model order is estimated (using the Bayesian Information Criterion, individually for each multivariate time series.) Coefficients for the full regression (2) are estimated in sample via the stable and computationally efficient LWR variant of Morf et al. (1978), along with the resulting residuals covariance matrix. Model validation procedures are performed as described above. The autocovariance sequence for the model is then extracted from the VAR parameters to a suitable number of lags to achieve near-machine precision (see Appendix B.2). Parameters for the reduced regressions (3) are computed from the autocovariance sequence via Whittles's spectral factorisation algorithm (Whittle, 1963). Note that this procedure avoids explicit sample estimation of VAR parameters for the *reduced* models, thus eliminating a further potential source of sampling error. Finally GCs are calculated from the full and reduced residuals covariance matrices as the appropriate log ratio of generalised variances (4).

3. Results

3.1. GC-fMRI using a simple vector autoregressive model

Figure 3 shows the results of fully conditional GC analysis (averaged over 100 instantiations) generated using the simple 5-node VAR model (Figure 1A) and the difference-of-gamma HRF approximation. In these results, and in all subsequent simulation results, model orders p were estimated independently for each instantiation (see Table 3). The middle columns [BOLD$_{250}$ (conf.) and BOLD$_{0.5}$ (conf.) panels] were generated with the HRF kernels confounding the underlying neural delay, as shown in Figure 1B. As expected, GC analysis of the simulated EEG/LFP accurately reveals the underlying structure. Strikingly, GC analysis of the BOLD$_{250}$ data (i.e., following convolution with confounding HRFs) also perfectly recaptures the underlying connectivity. This result overturns the view (David et al., 2008; Friston, 2009) that GC analysis will necessarily be corrupted if hemodynamic latencies confound the underlying neural delays. However, following severe downsampling on top of hemodynamic convolution (BOLD$_{0.5}$),

[2] The MVGC toolbox is currently under development. It is intended to supercede the Granger causal connectivity analysis' (GCCA) toolbox (Seth, 2010).

GC analysis fails to recover the underlying structure. (Indeed, to some extent we see a reversal of causal structure, see Section 4.) Regarding model orders (shown in Table 3); as expected (see Section 3.2 below), convolution with an HRF increases the model order (by about threefold from the 'true' model order of 15), while heavy downsampling reduces the model order to nearly 1.

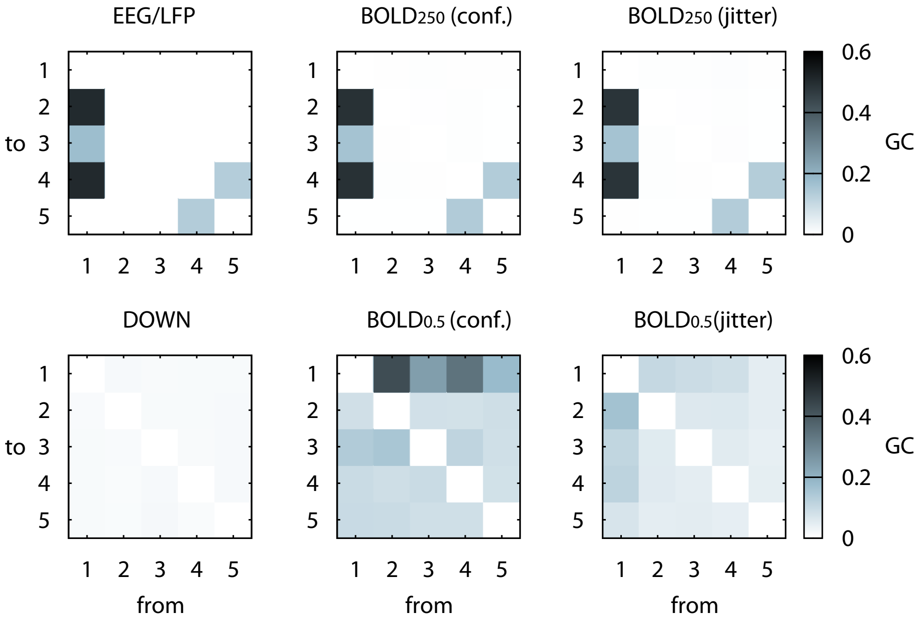

Figure 3: Conditional GC analysis of data generated from the simple VAR model (150 sec simulation time). Each panel shows the mean GC, over 100 separate simulation runs, between each pair of nodes. EEG/LFP: VAR model output downsampled to 250 Hz. DOWN: VAR output downsampled to 0.5 Hz. BOLD$_{250}$ (conf.): VAR output convolved with confounding HRFs (Figure 1B) and downsampled to 250 Hz. BOLD$_{0.5}$ (conf.): VAR output convolved with confounding HRFs and downsampled to 0.5 Hz. BOLD$_{250}$ (jitter): VAR output convolved with jittered HRFs and downsampled to 250 Hz. BOLD$_{0.5}$ (jitter): VAR output convolved with jittered HRFs and downsampled to 0.5 Hz.

We repeated an additional 100 simulations of the VAR model but, on each simulation, drawing the HRF parameters for each node from a random distribution incurring about 20% variation around equal prior values (see Section 2.3). The right-hand column BOLD$_{250}$ (jitter) and BOLD$_{0.5}$ (jitter) in Figure 3 shows that (conditional) GC inferences remain robust given simulation-by-simulation jitter of HRF shapes. In the context of this model

	mean	std
EEG/LFP	15.00	0.00
BOLD_{250} (conf.)	48.60	1.22
BOLD_{250} (jitter)	41.47	2.87
DOWN	1.00	0.00
$\text{BOLD}_{0.5}$ (conf.)	3.30	0.50
$\text{BOLD}_{0.5}$ (jitter)	2.10	0.30

Table 3: Empirical model orders (p) estimated over 100 instantiations of each data type, corresponding to the VAR-type simulations described in Figure 3. Means and standard deviations are given.

this jitter could reflect either variance across individuals or variance within individuals across trials. Model orders are again shown in Table 3 and are consistent with those reported in the previous (non-jittered) analysis. Note that Figure 3 (and subsequent GC plots) report average GC magnitudes without explicit statistical significance testing. The reason for this is that hemodynamic filtering can cause standard significance tests to perform unreliably, while GC magnitudes nonetheless retain meaningful interpretation in terms of information flow ((Barnett et al., 2009); see Section 2 and Section 3.2 below).

3.1.1. Simple VAR model with BW hemodynamics

Figure 4 shows conditional GC analysis of data generated by the BW model. The BW model takes as input the VAR model output which we assume to represent a broadband LFP signal. Results are averaged over 100 separate simulations. For each simulation BW parameters are jittered randomly (and independently for each variable) around priors drawn from the DCM component of SPM8 (see Section 2). As in Figure 3, GC analysis of the simulated BOLD_{250} response accurately recovers the underlying causal structure whereas heavily downsampled data (DOWN and $\text{BOLD}_{0.5}$) fail to detect any causal structure. This result shows that the empirical invariance of GC-fMRI to hemodynamic convolution is not dependent on using a filter-based approximation but generalizes to a biophysically detailed neuronal-to-BOLD mapping.

3.2. Invariance of GC to HRF convolution: Analytical results

To examine the theoretical basis for the invariance properties illustrated above, we next analyse the impact of hemodynamic convolution from the perspective of recent results establishing the invariance properties of GC

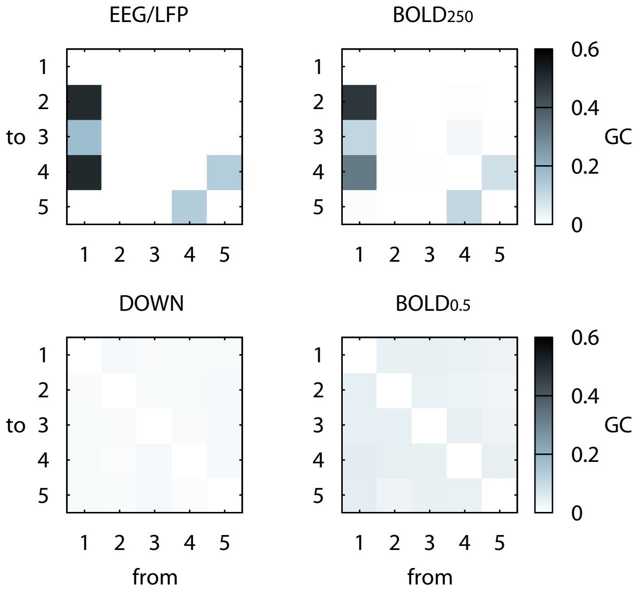

Figure 4: Conditional GC analysis of data generated from the simple VAR model. Details are as for Figure 3 except here the BOLD time-series are generated from the VAR output by the BW HRF model, rather than the HRF double-gamma approximation.

under stable invertible filtering (Barnett and Seth, 2011). (As a reminder, in this paper we follow the notational conventions that bold symbols denote vector (multivariate) quantities and upper-case symbols denote either random variables or matrices, according to context.)

3.2.1. Filter invariance of GC

Barnett and Seth (2011) showed that GC is invariant under a broad class of stable, invertible multivariate digital filters. We briefly review the relevant results. A multivariate digital filter is specified by a rational transfer function $G(z) = P(z)^{-1}Q(z)$, where $Q(z) = \sum_{k=0}^{r} Q_k z^k$ and $P(z) = \sum_{l=0}^{s} P_l z^l$ are $n \times n$ square matrix holomorphic functions (r, s may be infinite), normalized so that $P(0) = I$; z is a complex scalar variable. We indicate filter-transformed quantities by a tilde, so that for a multivariate time series $\mathbf{u}_t$ the filter action may be represented as:

$$\sum_{l=0}^{s} P_l \cdot \tilde{\mathbf{u}}_{t-l} = \sum_{k=0}^{r} Q_k \cdot \mathbf{u}_{t-k} \tag{5}$$

The filter is of FIR (finite impulse response) type iff $P(z) \equiv I$; otherwise it is of IIR (infinite impulse response) type. The filter is *stable* if $\det(P(z)) \neq 0$ on the unit disk $|z| \leq 1$ (i.e., all poles of $G(z)$ lie outside the unit circle (Antoniou, 1993)), and *invertible* if the matrix $Q(0)$ is invertible. Intuitively, a filter is stable if an impulse does not "blow up" (a FIR filter is always stable). Invertibility guarantees that an inverse filter exists; this precludes, for example, a pure delay such as $\tilde{\mathbf{u}}_t = \mathbf{u}_{t-1}$ which, though stable, is not invertible.

Given covariance-stationary processes $\mathbf{X}_t, \mathbf{Y}_t$ as before, the filtered process $\tilde{\mathbf{X}}_t$ may also be modelled as a full VAR regressed on its own past and that of $\tilde{\mathbf{Y}}$ as in (2), and as a reduced VAR regressed just on its own past as in (3). Importantly, the model order for the filter-transformed regressions will generally be higher (in theory infinite) than that of the original regressions. We remark that there is an argument that filtering might be better studied in the context of VARMA (vector autoregressive moving-average) than VAR modelling since, as alluded to in Barnett and Seth (2011), finite-order VARMA—but not VAR—models are preserved under finite-order digital filtering. The class of finite-order VARMA models is in addition closed under subsampling (Bergstrom, 1984; Solo, 2007) as well under the application of certain types of additive noise (Solo, 2007). In this study, however, we confine our attention to VAR processes since VAR modelling is by far the commonest operationalisation of GC analysis in neuroscience, and because of the substantial practical difficulties of GC analysis for VARMA models. For example, for VARMA models, the null (non-causality) condition is nonlinear as opposed to the linear condition [Appendix B, eq. B.1] for VAR models, and there are awkward computational issues regarding system identification and maximum-likelihood parameter estimation (Lütkepohl, 2005). In Barnett and Seth (2011) we show, under the stability and invertibility assumptions on $G(z)$, and provided that the cross-filter term $G_{xy}(z) \equiv 0$, that the VAR coefficients and residuals covariances transform in such a way that GC is left invariant: $\mathcal{F}_{\tilde{\mathbf{Y}} \to \tilde{\mathbf{X}}} = \mathcal{F}_{\mathbf{Y} \to \mathbf{X}}$. If a conditioning variable $\mathbf{Z}_t$ is included, the conditional GC $\mathcal{F}_{\mathbf{Y} \to \mathbf{X} \,|\, \mathbf{Z}}$ is similarly shown to be invariant [in this case the requisite condition is $G_{xy}(z) = G_{xz}(z) = G_{zy}(z) \equiv 0$]. We note that invertibility is a *sufficient* but not a *necessary* condition for invariance; see the minimal worked example below. An important caveat is that, despite theoretical invariance, filtering may still impact adversely on statistical inference of GC, due mostly to the increase in empirical model order; in particular, increased incidences of Type I errors (false positives) are to be expected. We discuss this issue in more detail below.

3.2.2. GC Filter invariance and HRF convolutions

We now examine GC filter-invariance with regard to the HRF convolutions. Firstly, note that a convolution may be considered as a FIR (and hence stable) filter. Although different HRF convolutions may apply to different neural variables, there are no cross-terms so that—provided they are invertible filters—the theoretical invariance applies. Secondly, although Barnett and Seth (2011) address the effects of filtering on accuracy of estimation and statistical inference, the filters analysed empirically there are (a) the same for each variable and (b) of a rather different type to the HRF filters considered here. In particular, Barnett and Seth (2011) examine standard high/lowpass and notch filters as might be applied in a standard signal processing context (Antoniou, 1993); the double-gamma HRF convolutions, on the other hand, are more akin to slow, moving-average filters.

Here we demonstrate the effects of HRF-like filtering through analysis of the minimal two-variable stationary VAR

$$\begin{aligned} X_t &= cY_{t-\ell} + \varepsilon_t \\ Y_t &= \qquad\quad\, + \eta_t \end{aligned} \tag{6}$$

where ε_t, η_t are uncorrelated white noise terms, so that the residuals covariance matrix is the identity matrix I. The constant c mediates the strength of causality $Y \to X$, with causal lag $\ell \geq 1$. In Appendix C we calculate the GC from Y to X as

$$\mathcal{F}_{Y\to X} = \log\left(1 + c^2\right) \tag{7}$$

(*cf.* Barnett and Seth (2011), Section 4). It is clear that GC in the opposite direction is zero: $\mathcal{F}_{X\to Y} = 0$.

We then apply convolutions to X, Y corresponding to a bivariate FIR filter with transfer function

$$G(z) = \begin{bmatrix} g(z) & 0 \\ 0 & z^k h(z) \end{bmatrix} \tag{8}$$

with $g(z), h(z)$ invertible univariate transfer functions and $k \geq 0$ an integer. The convolution applied to the predictor variable Y is thus an invertible filter followed by a pure k-lag onset delay for $k > 0$, in which case the filter is not invertible and invariance does not necessarily apply. While for the most part we do not consider explicit onset delays a few remarks are worthwhile; additional details are given in Appendix C (see also Section 4). For $k < \ell$ (convolution onset delay is shorter than causal lag; this includes the case $k = 0$ when the bivariate filter (8) remains invertible), the convolutions effectively reduce the causal lag by k, so that causalities are unchanged. If $k = \ell$ (convolution onset delay is equal to causal lag), the original causality $\mathcal{F}_{Y \to X}$ is destroyed by the convolution and no causality is introduced in the reverse direction. If $k > \ell$ (convolution onset delay is longer than causal lag), causality in the convolved process is effectively reversed.

As mentioned, the HRF convolutions may be viewed as slow, moving-average filters (with consequently strong lowpass characteristics), which are also likely to feature differential times-to-peak (*cf.* Figure 1B,C). To model these features, we applied binomially weighted moving-average convolutions of the form (8) with transfer functions $g(z) = [\frac{1}{2}(1+z)]^2$ and $h(z) = [\frac{1}{2}(1+z)]^3$ to realizations of length 1000 time steps of the minimal VAR (6) with causal lag $\ell = 1$ and onset delay $k = 0$, so that causalities are theoretically unchanged by filtering. The causal constant c was set to ≈ 2.5277, giving the theoretical G-causality $\mathcal{F}_{Y \to X} = 2$ according to (7). An empirical model order of 40 was estimated by the Bayesian Information Criterion. Despite the theoretically infinite model order of the convolved process, the model VAR coefficients as reflected in the estimated VAR transfer function (Appendix C) are well approximated at this model order (Figure 5). Over 1000 trial runs, the filtered G-causality $\mathcal{F}_{\tilde{Y} \to \tilde{X}}$ was estimated with a mean of 1.9469 ± 0.0653 standard deviations, very close to the theoretical value of 2. The theoretically zero $\mathcal{F}_{\tilde{X} \to \tilde{Y}}$ was estimated with a mean of 0.1413 ± 0.0168 standard deviations. Consistent with the theoretical invariance, the causal structure is thus well reflected empirically.

To test for statistical significance p-values were calculated based on the theoretical χ^2 asymptotic null distribution (see Section 2.4) and tested at a significance level of $\alpha = 0.01$, with a Bonferroni correction for the dual null hypotheses on causalities in the $Y \to X$ and $X \to Y$ directions. Causality in the $Y \to X$ direction was always correctly reported as significant; however, causality in the $X \to Y$ direction was frequently also reported as significant; i.e. there was a high incidence of false positives. We repeated the experiment, but this time with the *same* convolution $(1+z)^3$ applied to both time series. $\mathcal{F}_{\tilde{Y} \to \tilde{X}}$ was estimated with a mean of 2.0429 ± 0.0669, and $\mathcal{F}_{\tilde{X} \to \tilde{Y}}$ with a mean of 0.0494 ± 0.0106. The causal structure was thus again correctly reflected, but now significances were correctly reported in both directions. We repeated these experiments using permutation testing (Barnett and Seth, 2011) instead of the theoretical χ^2 asymptotic null distribution; results were unchanged indicating that the reported effects on statistical significance testing were not due to failure of the theoretical distribution.

In summary, the above results indicate that while slow moving-average (i.e., HRF-like) filtering may lead to unreliable significance test results when

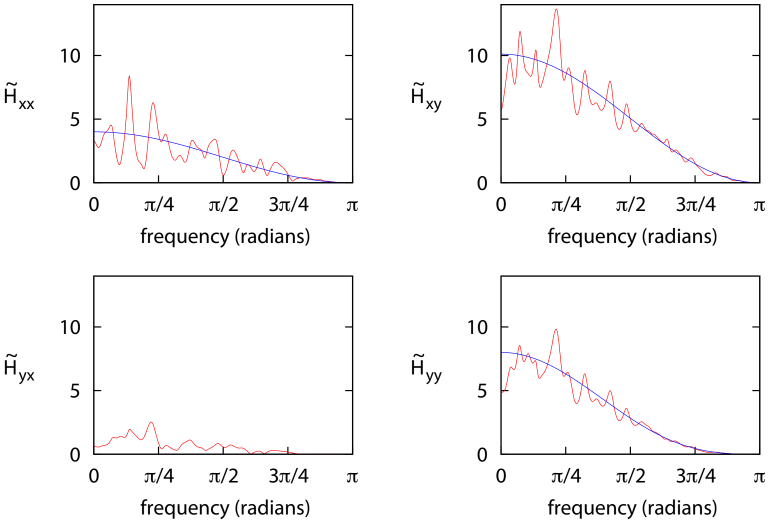

Figure 5: Modulus of VAR transfer function (see Appendix C for a typical 'minimal VAR' convolved time series. Theoretical values are plotted in blue, estimated values in red.

different kernels are applied to different variables, the actual values of estimated GC nevertheless correctly reflect causal structure. Practical heuristics for statistical inference for GC-fMRI are discussed further in Section 4.

3.3. GC-fMRI using a spiking neuronal model

In developing analysis methods it is important to rely neither on assumptions necessary for analytical approaches, nor on simulation models leveraging the same framework as the analysis method itself (nor, of course, on generative models that are simply inappropriate). One way to address this need is to utilize more biologically realistic simulations of neural dynamics (Valdes-Sosa et al., 2011). Following this approach, we next analyzed a model based on large populations of spiking neurons organized into two clusters (Figure 2; Section 2.2). We generated 10 sets of 260 sec of data from this model, discarding the first 60 sec of each run to allow for simulation burn-in. We used these 10 sets to generate data for 100 separate trials of GC-fMRI. For each trial one of the 10 data sets was randomly selected, and simulated LFP and BOLD responses were generated (BOLD responses were generated using the BW model with trial-to-trial jitter of BW parameters affecting time-to-peak, see Section 2.3). Again, the first 50 sec prior to analysis were discarded to allow for HRF 'burn-in'. Resulting GC values were averaged across the 100 trials. Empirical model orders are shown in Table 4 and are consistent with those obtained from the simple VAR generative model (Table 3). Figure 6 shows that, as with the simple VAR model, GC analysis of both the LFP and BOLD responses, under light downsampling, reliably identifies the underlying causal structure. However, heavy downsampling of both the LFP and the BOLD signal leads to the causal structure being missed. This result demonstrates that the invariance of GC to hemodynamic responses is retained even when using biophysically detailed models of both the BOLD signal and the underlying neural activity, escaping the circularity of using a VAR model for both data generation and analysis.

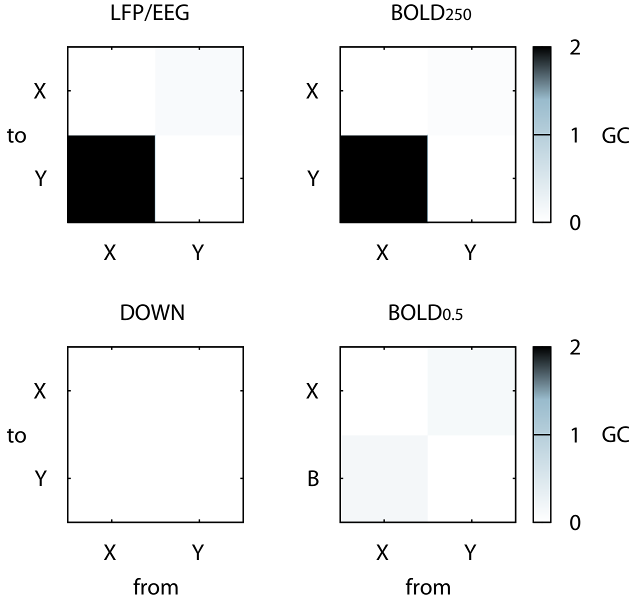

Figure 6: GC analysis of data generated from the spiking neuronal model (150 sec simulation time). Each panel shows the mean GC, over 100 separate simulation runs, between the two clusters. LFP/EEG: model output downsampled to 250 Hz. DOWN: model output downsampled to 0.5 Hz. BOLD_{250}: generated from model output by BW model and downsampled to 250 Hz. $\mathrm{BOLD}_{0.5}$: generated from model output by BW model and downsampled to 0.5 Hz.

3.4. Impact of downsampling

The above results demonstrate an invariance of GC to hemodynamic responses, but not to downsampling. In theory, this is to be expected since nei-

	mean	std
EEG/LFP	16.40	0.49
BOLD_{250}	23.73	0.53
DOWN	1.00	0.00
$\text{BOLD}_{0.5}$	1.97	0.59

Table 4: Empirical model orders (p) estimated over 100 instantiations of each data type, corresponding to the full spiking model simulations described in Figure 6. Means and standard deviations are given.

ther downsampling nor measurement noise can be characterized as invertible filtering operations, and so will not benefit from the invariance properties described above. In Solo (2007) it is shown that the *existence* of a Granger-causal effect is in fact preserved under downsampling [see also Breitung and Swanson (2002)] and also under certain types of additive noise (Section 3.5); however, in Solo (2007) the definition of "strong" Granger causality is more restrictive than the conventional ("weak") Granger causality and, moreover, the results there do not inform about the effect on magnitude or statistical inference of GC under downsampling and/or measurement noise. Therefore, to better understand the interaction between hemodynamic convolution and downsampling in practice for standard GC, we investigated sensitivity of GC-fMRI to different levels of downsampling using both the VAR model and the spiking neuron model.

3.4.1. VAR model

Figure 7 shows conditional GC analysis of time-series generated by the VAR model, convolved with confounding HRF kernels (difference-of-gamma approximation, as in Figure 1B) and downsampled at a range of frequencies. Again, results are averages over 100 trials. We see that GC for the simulated BOLD signal (right column) degrades slightly more rapidly than GC for unconvolved data, shown for comparison (left column). At low (fMRI-comparable) sample frequencies there is evidence that the confounding HRFs lead to a reversal in the direction of GC.

3.4.2. Spiking neuron model

Figure 8 shows GC analysis of time-series generated by the spiking neuron model, with BOLD signal generated from the simulated LFP by the BW model, and downsampled at a range of frequencies. Results are averages over 100 trials randomly selected from the 10 simulated LFP data sets. In this case GC for both the simulated LFP and BOLD signals degrades quite rapidly. It is of interest that the 'difference of influence' term ($\mathcal{F}_{\mathbf{Y}\to\mathbf{X}} - \mathcal{F}_{\mathbf{X}\to\mathbf{Y}}$) (Roebroeck et al., 2005) applied to the simulated BOLD signal would perform better on this data than GC applied to the simulated LFP data, however this can be attributed to the fact that over 100 separate trials the effects of jittered HRF latencies are averaged out.

3.4.3. Interaction of downsampling and hemodynamic confound

Figure 9 shows GC analysis of time series generated by the spiking neuron model, with BOLD signal generated from the simulated LFP by HRF difference-of-gamma convolutions with differential confounding delays-to-peak around a reference delay-to-peak of 4 sec applied to both X and Y outputs. The BOLD signal is then downsampled at a range of frequencies. The confounding delays vary from $[0, 1000]$ ms; note that the X → Y causal delay in the LFP is $[40, 50]$ ms. Results are again averages over 100 trials randomly selected from the 10 simulated LFP data sets. At high (EEG) sample frequencies, causal structure estimation is robust to confounding delay on the HRF time-to-peak. However, as the downsample frequency is decreased, the impact of the confounding delays becomes more prominent until at low (fMRI) frequencies estimated GCs are already destroyed with no HRF confound, and are reversed with strong HRF confound.

3.5. Impact of measurement noise

The final set of simulations investigated the impact of measurement noise on GC-fMRI, again for both the simple VAR generative model and the spiking neuron model. In all experiments white noise was additively combined with the raw or BOLD model output at a series of signal-to-noise ratios (SNRs), and then downsampled. As mentioned (Section 3.4), measurement noise cannot be considered as an invertible filtering operation and so the theoretical invariance (Section 3.2) does not apply here.

Figure 10 illustrates the degrading effect of measurement noise on the raw (unconvolved) VAR output downsampled to 250 Hz. (Conditional) GC can be seen to degrade slowly with increasing measurement noise. However, GCs for the BOLD timeseries (not shown) as derived from the VAR output by either HRF convolutions or the BW model, were destroyed, even at a (high) downsample frequency of 250 Hz and a SNR of 10 (i.e. noise level = 0.1 of signal standard deviation). Similarly, downsampling the raw output (i.e. without convolution) at a (typical fMRI) frequency of 0.5 Hz also destroyed all GCs. These findings also held for the spiking neuron model when measurement noise was added in the same way, as shown in Figure 11. Summarizing these findings, even small amounts of noise in combination with either BOLD simulation and/or downsampling have a drastic impact on GC estimation.

4. Discussion

In this paper, we first illustrated the invariance of GC-fMRI under HRF variability using a simple multivariate vector autoregressive (VAR) model for generating the underlying 'neural' signal. We then provided the analytical foundation for this result by invoking a previous finding showing invariance of GC under a broad class of digital filters (Barnett and Seth, 2011). We next established the robustness of the result by analyzing a biophysically detailed model implementing spiking neurons and a BW neurovascular-hemodynamic model. We then used both this detailed model, and the comparatively simple VAR model, to characterize systematically the effects of downsampling and measurement on noise on GC-fMRI.

4.1. Summary of findings

The application of GC to fMRI data has been frequently challenged on the basis that inter-regional differences in hemodynamic response latency may readily confound directional interactions in the underlying neural dynamics, leading to false inferences (David et al., 2008; Friston, 2009; Smith et al., 2011). This point has been made particularly strenuously in the context of an ongoing discussion comparing the relative merits of GC and alternative model-based methods (e.g., dynamic causal modelling, DCM, (Friston et al., 2003)) for identifying directed interactions from fMRI data. On the other hand, a series of simulations have revealed some degree of robustness of GC when applied to fMRI signals (Roebroeck et al., 2005; Deshpande et al., 2010; Schippers et al., 2011); however the extent and theoretical foundation of this resilience has so far remained unclear. Here, we have addressed these issues by a rigorous combination of theory and simulation modelling at multiple levels of biophysical detail including (i) analytically solvable VAR models, (ii) VAR models with nontrivial topologies and realistic hemodynamics, and (iii) a detailed 'analysis agnostic' simulation model involving spiking neuron populations, explicit synaptic conductances underlying LFP generation, and biomechanically detailed hemodynamics.

Theoretically we have established that GC is invariant to hemodynamic convolution (excluding confounding onset delays). This result is based on recognizing that differences in hemodynamic time-to-peak do not reflect differential 'buffering' (i.e., onset delays) of the underlying neuronal signal but rather reflect different convolution kernels within the framework of low-pass filtering. We have built on previous results showing invariance properties of GC under invertible filtering (Barnett and Seth, 2011) to accommodate the case of HRF-like filters, with different kernels applied to different variables. This theoretical result was confirmed in simulation, using both VAR models and a biophysically detailed spiking neuronal model to generate the underlying 'neural' signal. The VAR model implemented a nontrivial 5 node topology including reciprocal, direct, and indirect connections (Figure 1A) whereas the spiking model traded increased biophysical realism for a simple 2 node functional architecture (Figure 2). In both models, simulated BOLD responses were generated using both a standard difference-of-gamma approximation and the extended BW model (Friston et al., 2000). In some simulations HRFs were chosen deliberately to confound the underlying neural influences in terms of time-to-peak (Figure 1B) while in others HRF parameters were jittered on a trial-to-trial basis within biophysically plausible ranges. In all cases we observed a striking resilience of GC to variable hemodynamic filtering (e.g., Figures 3,4,5). However, our simulations also revealed that hemodynamic filtering when confounded with neural influences did corrupt GC inferences if combined with severe downsampling (Figures 3,4,7) and/or even low levels of measurement noise (Figures 10,11), accounting for similar effects observed in previous simulation studies (Deshpande et al., 2010; Schippers et al., 2011) and in empirical data (David et al., 2008). As expected, convolution with an HRF substantially increases the empirical model order, while heavy downsampling substantially reduces the model order. Our results therefore do not mandate the routine application of GC to fMRI. They do, however, transform what had previously been considered a problem in principle (i.e., the idea that hemodynamic response variation necessarily confounds GC inferences) to a problem in practice (i.e., given sufficiently fast sampling and low measurement noise, GC-fMRI will be invariant to such variation).

Given that fMRI involves both hemodynamic filtering and downsampling we carefully examined their interaction. Corroborating previous studies (Roebroeck et al., 2005; Deshpande et al., 2010; Schippers et al., 2011) we found that when HRF latencies were confounded with neural latencies under severe downsampling, GC inferences tended to follow the hemodynamics rather than the neural mechanisms (e.g., compare the top and bottom panels of the middle column in Figure 3). Parametrically, as downsampling becomes more severe the impact of confounding HRFs is increased (Figure 9). Interestingly, HRF filtering can allow some detection of causal interactions even if the downsampling frequency is too low to capture these interactions in the neural data itself (Figure 8); however the validity of these inferences will of course depend on the HRFs not confounding the neural latencies. Notably, downsampling and measurement noise degrade GC inferences even in the absence of HRF filtering (see Figures 7,8 for downsampling; Figures 10,11 for noise).

Hemodynamic filtering impacted the reliability of statistical significance testing, generating an increased preponderance of false positives (Type 1 errors). This effect, observed previously (Roebroeck et al., 2005), is likely due to the increased difficulty in adequate VAR model fitting given increases in model order following filtering (Barnett and Seth, 2011). We quantified these effects using slow moving-average filters applied to a simple simulation for which GC values are analytically available (Section 3.2). Even though false positive rates were inflated, estimated GC magnitudes closely matched their true theoretical values. This result is significant because GC magnitudes have a meaningful interpretation in terms of information flow measured in bits (Barnett et al., 2009), regardless of statistical inference. Interestingly, we observed that statistical significance testing was largely restored when the *same* convolution/filter was applied to each variable. Although we cannot fully account for this observation at present, we suspect that it will again depend on issues of empirical model fitting. In practice, statistical inference may be performed on distributions of GC magnitudes between experimental conditions, using nonparametric tests such as the Wilcoxon rank sum [see (Barrett et al., 2012) for an application to EEG data]. This approach will determine whether the GC in one condition is significantly different from that in another, but not whether a particular GC interaction is significant in itself. In the latter case, the fact that GC magnitudes remain accurately estimated warrants a simple heuristic of setting an arbitrary threshold to distinguish between weak (or 'insignificant') and strong (or 'significant') interactions.

Summarizing these results, GC is fully invariant to HRF filtering in theory (excluding confounding onset delays) and strikingly invariant in practice, given sufficiently fast sampling and low measurement noise. While even low noise levels can disrupt GC inferences, the impact of downsampling depends on a complex interaction between hemodynamic and neural latencies, and in the absence of confounds HRF filtering can ameliorate the impact of downsampling. Although filtering can induce additional false positives during statistical testing, GC magnitudes interpreted information-theoretically remain accurately estimated in sample.

4.2. Comparison with previous simulations

Following the initial studies of Roebroeck et al. (2005), a number of more recent simulation studies have addressed GC-fMRI with varying outcomes. In an influential comparative analysis, Smith et al. (2011) found that lag-based methods (such as GC) performed poorly in reconstructing network topology as compared to alternative methods based on higher-order statistical moments. However, their analysis was performed using the DCM forward model to generate simulated data, which is not appropriate for subsequent GC analysis; in particular, neural dynamics were simulated by a (non-neuronal) system of continuous differential equations modulated by binary impulse functions which is very different from the stationary stochastic data generated by an VAR model or by our detailed spiking model of LFPs.

Deshpande et al. (2010) examined GC performance under systematic variation of TR, measurement noise, strength of causal influence, and underlying neural delay. They used experimentally obtained LFP traces as their neural signals and simulated neural interactions via simple time-shifts of a single LFP trace. While this strategy guarantees the biophysical plausibility of the LFP signal it represents a highly oversimplified model of neural interactions. Also, these authors used a difference-of-gamma approach to HRF generation and did not consider the biophysically detailed BW model. Nonetheless their conclusions were in broad agreement with those presented here inasmuch as GC-fMRI performs better under low noise and fast sampling conditions. Schippers et al. (2011) focused on group-level GC-fMRI as opposed to single 'subject' analyses conducted in this paper and in Deshpande et al. (2010). They used a VAR model to generate 'neural' data and difference-of-gamma HRF convolutions fitted to the various HRF shapes measured in Handwerker et al. (2004). They considered only bivariate situations and assumed independent HRF variations between subjects, similar to our 'jittered' HRF models (see e.g., Figure 3). Their simulations indicated reasonable sensitivity of GC-fMRI at the group level (i.e., when averaging across HRF variations), though in a way highly dependent on the underlying neural delay. Our results significantly extend these previous simulations by (i) establishing a theoretical basis for invariance of GC-fMRI to HRF variation; (ii) validating this invariance in models spanning a wide spectrum of biophysical detail, and (iii) testing systematically the impact of downsampling, TR, neural delay, HRF confound, and measurement noise in relevant subsets of these simulations.

4.3. Agnostic simulation models

A key feature of our study has been to combine biophysically realistic neurovascular and hemodynamic models (i.e., the extended BW model) with similarly detailed models of underlying neural dynamics and the resulting LFPs (i.e., the spiking model). The full model contained >12,000 spiking neurons and ∼19 million synapses, with explicit synaptic conductances, short-term synaptic plasticity, and realistic neural delays. Simulated LFPs from this model were fed into the extended BW model including both neurovascular and hemodynamic components, generating a simulated BOLD response. One motivation for this additional detail is to validate our results on time-series data generated by more biologically realistic simulations of neural dynamics (Valdes-Sosa et al., 2011). An important benefit of this approach is that it escapes the potential circularity of using similar model classes to both generate data and perform subsequent analyses. To date, most simulation studies of GC have used a VAR generative model (e.g., (Roebroeck et al., 2005; Schippers et al., 2011) but see (Deshpande et al., 2010)); similarly, simulation studies of DCM have tended to use the relatively abstract DCM neuronal model (e.g. (Friston et al., 2003; Stephan et al., 2008), though see (Marreiros et al., 2008)). These approaches, while convenient, may lead to biases in favour of the corresponding analysis method. Here we have sought to avoid any such biases by using 'agnostic' simulation models which make no assumptions about subsequently applied analysis methods. In future studies it will be interesting to adapt a single generative model, such the spiking model, to generate multiple datasets suitable for GC and DCM respectively, enabling a principled comparison among these different approaches.

Alternative simulations approaches exploring a middle ground between full spiking models and VAR approximations may also be useful. For example, network simulations based on Wilson-Cowan mean-field approximations (Wilson and Cowan, 1972) could allow more complex networks to be studied by relaxing computational constraints. However, the Wilson-Cowan approach is more appropriate for modelling of stable states and global transitions under slowly changing inputs, which we do not consider here (but which forms an interesting topic for future research). Moreover, the spiking neural model allows us to verify that fast (correlated) fluctuations in simulated neuronal time-series do not affect the GC analysis.

We made one simplifying assumption in the combination of the spiking and extended BW models, which was to use the full LFP signal as the neuronal input to the BW model. Although this is consistent with an observed strong association between global LFP and BOLD responses (Magri et al., 2012), the relation between neuronal activity and BOLD remains incompletely understood and is undoubtedly complex (Logothetis, 2008; Rosa et al., 2011; Magri et al., 2012; Buxton, 2012). Future studies could examine more complicated neurovascular models (Rosa et al., 2011) or band-limited LFP inputs, as well as more complex topologies and interactions between changing neural-level parameters and GC analysis of the resulting LFP and BOLD signals. An important caveat is that simulation data, no matter how detailed, can never be taken to replace empirical data. Analysis of empirical data will always face additional challenges such as stationarity and absence of a ground-truth comparison; these issues are discussed further in Section 4.6 below.

4.4. HRF variation in fMRI

Measurement and modelling of HRFs remains a challenging issue not only for connectivity analyses but in fMRI generally (Handwerker et al., 2012; Buxton, 2012), since all analysis methods including the ubiquitous general linear model, as well as DCM, make assumptions about HRF shape and variability. In a seminal study, Handwerker et al. (2004) were able to estimate HRF shapes (and hence variation) in primary sensorimotor cortex in 20 subjects performing a button-press task. However, more general characterization of HRFs globally has remained infeasible in the absence of corresponding knowledge of 'ground truth' neuronal responses (David et al., 2008). In particular, when measuring BOLD signals alone it is impossible to disentangle neural delays from HRF effects (Handwerker et al., 2004), in the absence of prior assumptions (e.g., those embedded in DCM). Our simulations utilize HRF parameters drawn either from widely-used prior distributions (Friston et al., 2003) or which explicitly impose confounds worse than might be expected on the basis of measured HRF variation (Handwerker et al., 2004). Importantly, the simulated neural delays (20-50 ms) lie well within plausible physiological ranges (Schmolesky et al., 1998; Smith et al., 2011; Schippers et al., 2011). Together, these factors imply that our simulations are testing situations likely worse than encountered in practice. As described above we also use both double-gamma approximations and the extended BW model in order to ensure that our results do not depend on a filter-based approximation to the BOLD response; both models can be readily parameterized to generate variation in hemodynamic time-to-peak.

HRFs vary along multiple dimensions, potentially including onset delay as well other factors leading to differential time-to-peak (Handwerker et al., 2004, 2012; Menon, 2012). Our results confirm that confounds in explicit onset delays do, as expected, affect GC inference. This raises the question of how prevalent onset delay variations are in fMRI responses. Most descriptions of HRF variation emphasize time-to-peak, width, and post-undershoot shapes (Handwerker et al., 2012), and indeed the detailed BW model has no explicit onset delay term (Friston et al., 2000). While onset delay variations have been observed (Menon et al., 1998; Menon, 2012) it is difficult in practice to discriminate between onset delay and other parameters affecting time-to-peak. From a theoretical perspective, the key distinction is between a completely flat initial segment of a convolution kernel (a non-invertible filter for which GC is not invariant) and a very shallow slope (for which the invariance applies). Moreover, differential onset delays may in fact reflect underlying neural delays and processing rather than variation in neurovascular or hemodynamic properties. This is the idea underlying fMRI-based 'mental chronometry', which is supported by observed trial-to-trial correlations between onset delay and behavioral responses such as reaction time (Menon et al., 1998; Menon, 2012) and indeed the detailed BW model. Under this interpretation, onset delay variation will help, not hinder, GC analysis.

4.5. Functional, effective, and causal connectivity

Connectivity analysis methods (as applied to time series data) are commonly divided into two classes: functional and effective (Friston et al., 1993; Friston, 2011). Standardly, functional connectivity describes statistical dependencies among observed responses and is associated with undirected quantities such as correlation and synchrony. By contrast, effective connectivity is best conceived as 'the simplest possible circuit diagram explaining observed responses' (Aertsen and Preißl, 1991) and is explicitly a claim about underlying mechanisms, and especially their modulation by experimental condition (Friston, 2011). Effective connectivity has its roots in structural equation modelling and is currently best expressed in the context of neuroimaging via the framework of Bayesian model selection in DCM (Friston et al., 2003). GC, in virtue of supporting causal inference, has often been classified as effective connectivity (Deshpande et al., 2010; Schippers et al., 2011; Roebroeck et al., 2005). However GC is perhaps better thought of as *directed* functional connectivity, which we may also call 'causal connectivity' (Bressler and Seth, 2011): GC patterns are clearly dependent on underlying causal mechanisms but should not be taken to be identical with them. The mislabelling of GC as effective connectivity is one possible reason for some of the confusion surrounding the relative merits of GC and DCM (Bressler and Seth, 2011; Roebroeck et al., 2011b; Valdes-Sosa et al., 2011). We emphasize that the methods are fundamentally different and complementary, with causal connectivity permitting a more exploratory description of dynamical interactions, recognizing that (i) a single mechanism may support multiple dynamics and (ii) in neural systems dynamical patterns and their underlying mechanisms may mutually specify each other via plasticity (Bressler and Seth, 2011).

A related distinction between GC and DCM is often made on the basis of data-driven/exploratory (GC) versus model-driven/confirmatory (DCM) approaches. This distinction is however not sharp and rather reflects endpoints of a spectrum, with the middle ground becoming increasingly populated (Roebroeck et al., 2011b; Bressler and Seth, 2011; Ryali et al., 2011). Advances in Bayesian model selection are allowing more extensive repertoires of causal models to be compared, potentially moving DCM in the direction of data-driven approaches (Penny et al., 2010). On the other hand, even standard GC has a model-driven element inasmuch as observed responses are assumed to be well represented by an VAR model (related approaches such as transfer entropy avoid this assumption (Schreiber, 2000)). Recent developments of GC have combined VAR models with 'observation' equations linking VAR dynamics with observed responses in the framework of state-space modelling, providing a compromise between standard GC and DCM (Ryali et al., 2011); see also Smith et al. (2010).

4.6. Implications for fMRI functional connectivity analysis

Our results show that GC analysis of fMRI signals is possible *in principle* given sufficiently fast sampling and low measurement noise. The finding that hemodynamic convolution *per se* does not inevitably confound GC analysis opens new avenues for advancing data-driven functional connectivity analyses. Specifically, methods which can reduce (or model) measurement noise and also increase sampling rates have the potential to allow robust GC inference on fMRI BOLD signals. Fortunately, progress is already being made in both directions. New methods are being developed for mitigation of measurement noise using novel preprocessing (Rasmussen et al., 2012) and post-processing (Nalatore et al., 2009) algorithms. High field MR scanners allow faster sampling, and novel imaging sequences may radically increase sampling rates: for example a recent multiplexed echo-planar imaging sequence is reaching TRs in the range of 200 ms without loss of coverage (Feinberg and Yacoub, 2012; Feinberg et al., 2010). Other developing neuroimaging methods, notably near-infrared spectroscopy, already allow arbitrarily fast sampling of BOLD responses (Im et al., 2010) albeit with presently low SNRs. While statistical inference remains a problem even for fast-sampled BOLD signals, the accurate estimation of GC magnitudes indicates that reliable inference can be made between distributions of GC values [as in (Barrett et al., 2012)], and that thresholds discriminating weak from strong GC values will also have heuristic value (see Section 4.1).

Previous discussions of GC-fMRI have recommended additional strategies for coping with HRF variability. In particular, Roebroeck et al. (Roebroeck et al., 2005) have recommended (i) comparing the difference in GC between experimental conditions as opposed to attempting to establish the 'ground truth' connectivity for a given condition, and (ii) analyzing the 'difference of influence' term ($\mathcal{F}_{\mathbf{Y}\to\mathbf{X}} - \mathcal{F}_{\mathbf{X}\to\mathbf{Y}}$) rather than each direction separately. The first strategy is motivated by the notion that HRFs are less likely to vary between experimental conditions than between brain regions or subjects, and the second by the observation of false positives in GC-fMRI as a result of HRF smoothing; however a problem with the difference-of-influence strategy is that it is constitutively unable to detect bidirectional causal interactions. In this paper we have deliberately avoided these strategies in order to better understand the properties of GC analysis of fMRI signals *per se*. It is likely that additional research exploiting these strategies may further enhance the potential for GC-fMRI, for example by accommodating any differences in explicit onset delays, should such differences exist. Future work should also consider the influence of different experimental designs on GC-fMRI, with particular attention to event-related designs which may challenge assumptions of covariance stationarity needed for GC analysis. Finally, there may be important opportunities in further examining the potential for examining GC-fMRI in the context of VARMA modelling (Sections 3.2, 3.4, 3.5 and (Roebroeck et al., 2011b)) in light of the properties of these models with respect to filtering, downsampling and measurement noise. We hope that future work might reinforce and clarify the results presented here in that regard.

Acknowledgments

We gratefully acknowledge support from the EPSRC (Leadership Fellowship EP/G007543/1 to AKS, Ph.D. studentship to PC) and the Dr. Mortimer and Theresa Sackler Foundation (AKS and LB) via the Sackler Centre for Consciousness Science. We thank Adam Barrett for helpful discussions, and four reviewers for their insightful comments.

Appendix A. Spiking neuron model

The model consists of two clusters (X,Y) with 6144 neurons in each cluster (Figure 2). Neurons are modelled using Izhikevich's phenomenological description (Izhikevich, 2003b) of cortical pyramidal (excitatory) and basket (inhibitory) cell types, with explicit axonal conductance delays. In both clusters, the ratio of inhibitory inter-neurons is set at 2%. Each neuron receives 1536 afferent connections, with input to each neuron calculated via explicit NMDA, AMPA, GABAa, and GABAb conductances (Dayan and Abbott, 2005) and incorporating short-term synaptic plasticity (Zucker and Regehr, 2002).

The model architecture reflects two locally recurrent cortical regions, interconnected by a single feed-forward (X→Y) pathway. Neurons in cluster X receive excitatory input only from cluster X, while neurons in Y receive excitatory input equally from X and Y. Inhibitory synaptic interactions are local within each cluster, reflecting local inhibition in cortex (Mountcastle, 1998). Axonal conductance delays are drawn from the uniform distributions of [1,10] ms for intra-cluster connections and [40,50] ms for inter-cluster connections.

Simulations were performed on combined CPU/GPU hardware (Intel Xeon / nVidia GeForce GTX) using C/CUDA.

Appendix A.1. Generation of neural activity

The discrete-time formulation of Izhikevich (2003b)'s model is defined by a pair of difference equations with discrete after-spike reset:

$$v' = 0.04v^2 + 5v + 140 - u + I + \xi \tag{A.1}$$

$$u' = a(bv - u) \tag{A.2}$$

$$\text{if } v \geq 30, \text{ then} \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{A.3}$$

where v represents the membrane potential (mV), u (dimensionless) governs the relaxation dynamics of the model neuron and $' = \frac{d}{dt}$. Following Izhikevich (2003b), regular-spiking excitatory pyramidal neurons are modelled using $a = 0.02$, $b = 0.2$, $c = -65$ and $d = 8$, and fast-spiking inhibitory inter-neurons are modelled using $a = 0.1$, $b = 0.2$, $c = -65$ and $d = 2$. This model captures several important features of neuronal dynamics not found in standard 'integrate-and-fire' models, such as spike-frequency adaptation and dynamic refractoriness.

Appendix A.2. Synaptic dynamics

Synaptic input to each neuron (I) is calculated from explicit conductances (g) for each post-synaptic neuron, as

$$\begin{aligned} I_j = {} & g_{\text{AMPA}}(v-0) \\ & + g_{\text{NMDA}} \frac{[(v+80)/60]^2}{1+[(v+80)/60]^2}(v-0) \\ & + g_{\text{GABAa}}(v-70) \\ & + g_{\text{GABAb}}(v-90) \end{aligned} \tag{A.4}$$

in which separate state variables (g_{AMPA}, g_{NMDA}, g_{GABAa} and g_{GABAb}) are integrated for each of the 4 receptor types. Spikes arriving at neuron j from neuron i step-increase the corresponding conductance variables by an amount proportional to both the synaptic weight ($\omega_{ij} \in [0,1]$) and the state of the synapse with respect to short-term plasticity (R and w, see below). Specifically, spikes arriving from excitatory pre-synaptic neurons cause conductance variables associated with glutamatergic receptors (g_{AMPA}, g_{NMDA}) to be augmented, whereas spikes from inhibitory inter-neurons augment GABAergic conductances (g_{GABAa} and g_{GABAb}) at the post-synaptic neuron. Omitting receptor-type subscripts we have

$$g'_j = \frac{-g_j}{\tau} + \sum_{i=0}^{S} \delta(t-t_i^*)\Omega_{ij} \tag{A.5}$$

where

$$\Omega = \omega R w \tag{A.6}$$

Here, t_i^* is the arrival time of last pre-synaptic spike of neuron i, δ is the Dirac delta function and S is the number of afferent synapses per neuron. Conductances for each receptor type therefore decay exponentially according to a receptor-specific time constant τ. For NMDA and GABAb type receptors we set $\tau = 0.1\,\text{ms}$ (implementing 'slow' synapses) and for AMPA and GABAa type receptors we set $\tau = 0.01\,\text{ms}$ ('fast' synapses) (Dayan and Abbott, 2005).

External (e.g. background) synaptic input is calculated for each neuron by a discrete random process at each time-step, such that ξ (Eq. A.1) follows the uniform distribution

$$\xi \sim U(-6.5, 6.5) \quad (\xi \in \mathbb{R}) \tag{A.7}$$

which is sufficient to cause neurons to fire irregular spike trains at 1–5 Hz without external stimulation (*cf.* Softky and Koch (1993)).

Appendix A.3. Short-term synaptic plasticity

Short-term synaptic plasticity was implemented to facilitate stable network dynamics (Abbott, 1997; Zucker and Regehr, 2002). Following Markram et al. (1998), each spike arrival updates synapse-specific variables governing facilitation (w) and depression (R) at the corresponding synapse. For facilitation, w is step-increased by $U(1-w)$ for each pre-synaptic spike, otherwise decaying by a rate governed by parameter F:

$$w' = \frac{(U - w)}{F} + \delta(t - t^*)U(1 - w) \tag{A.8}$$

For depression, R is step-increased by Rw for each pre-synaptic spike and decays to 1 at a rate governed by parameter D:

$$R' = \frac{(1 - R)}{D} - \delta(t - t^*)Rw \tag{A.9}$$

We used parameter values $U = 0.5$, $F = 1000$, $D = 800$ for excitatory synapses and $U = 0.2$, $F = 20$, $D = 700$ for inhibitory synapses, reflecting STP as observed in cortex (Izhikevich, 2004).

Appendix A.4. Generation of simulated LFPs

Simulated LFP time-series are obtained by assuming a strong correlation with dendritic AMPA currents (Ching et al., 2010). At each time-step an LFP value for each cluster ($V_{\mathrm{X/Y}}$) is computed according to:

$$V_{\mathrm{X/Y}}(t) = \frac{1}{NS} \sum_{j=0}^{N} \sum_{i=0}^{S_E} g_{ij}(t) \tag{A.10}$$

where N is the number of neurons in each cluster (X,Y), S_E is the number of afferent excitatory synapses per neuron and $g = g_{\mathrm{AMPA}}$.

Appendix B. Granger causality: statistical inference and stationarity

Appendix B.1. Statistical inference

Considering the VAR models (2) and (3), GC may be considered as a test statistic for the null hypothesis

$$H_0 : \; B_1 = \ldots = B_p = 0 \tag{B.1}$$

of zero GC influence. That is, $\mathbf{Y}$ Granger-causes $\mathbf{X}$ iff the 'full' regression (2) provides a significantly better model of $\mathbf{X}_t$ than the 'reduced' regression (3) at a given significance level. Standard theory (Hamilton, 1994) provides that, since the model (3) is nested in the model (2), the appropriate test statistic for H_0 is the corresponding likelihood ratio, which (at least under Gaussian assumptions on the distribution of residuals) is precisely the GC quantity $\mathcal{F}_{\mathbf{Y}\to\mathbf{X}}$ of (4), where model parameters are taken to be maximum likelihood estimators. Importantly, maximum likelihood estimates for the full and reduced regression parameters are known to be asymptotically equivalent to those obtained by an ordinary least squares (OLS) or equivalent procedure, allowing easy computation (Hamilton, 1994).

Classical asymptotic theory (Wilks, 1938; Wald, 1943) yields that under the null hypothesis H_0, $n\,\mathcal{F}_{\mathbf{Y}\to\mathbf{X}\,|\mathbf{Z}}$, where n is the number of observations, is asymptotically χ^2 distributed, with degrees of freedom given by the difference in the number of parameters between the models (2) and (3), furnishing a simple significance test for the GC statistic. (If $\mathbf{X}$ is univariate, an alternative test is available: the R^2-like statistic $\exp(\mathcal{F}_{\mathbf{Y}\to\mathbf{X}\,|\mathbf{Z}}) - 1$, scaled appropriately, has an asymptotic F-distribution under the null hypothesis (Hamilton, 1994).) If multiple GCs are calculated for a multivariate system, significance test results should also be adjusted for multiple hypotheses. Standard approaches include the Bonferroni, Sidak or various "false discovery rate" procedures (Hochberg and Tamhane, 1987).

Appendix B.2. Stationarity, autocovariance, and spectral radius

Given a multivariate system $\mathbf{U}_t$ comprising n component variables $U_{i,t}$ (the "universe" of variables), the GC between pairs of variables (U_j, U_i), conditioned on the remaining variables in the system, is given by $\mathcal{F}_{U_j\to U_i\,|\mathbf{U}_{[ij]}}$, where $\mathbf{U}_{[ij]}$ denotes omission of the variables U_i, U_j. The full regressions (2) for all such pairwise causalities is equivalent to the regression

$$\mathbf{U}_t = A_1\mathbf{U}_{t-1} + \ldots + A_p\mathbf{U}_{t-p} + \varepsilon_t \tag{B.2}$$

Having estimated coefficients A_k, the autocovariance sequence of the model $\Gamma_k \equiv \operatorname{cov}(\mathbf{U}_t, \mathbf{U}_{t-k})$ can be obtained (this involves solution of a discrete Lyapunov equation (Bartels and Stewart, 1972) and "reverse solution" of the associated Yule-Walker equations (Hamilton, 1994)). As described in Section 2.4.3, knowing Γ_k allows the parameters of the reduced regressions (2) to be extracted directly, without additional model fitting.

Regarding stationarity, (B.2) represents a wide-sense stationary VAR model iff all roots of the *characteristic equation* $\det\bigl(A(z^{-1})\bigr) = 0$, where

$A(z) = I - \sum_{k=1}^{p} A_k z^k$, lie within the unit disc $|z| < 1$ in the complex z-plane (Hamilton, 1994). The spectral radius of the model is defined as the largest modulus of all roots of the characteristic equation; it governs how quickly autocorrelation decays in the VAR model, and must be < 1 for a stationary system.

Appendix C. Analysis of the minimal 2-variable VAR under convolution with differential onset delay

To calculate the (unconditional) Granger causality $\mathcal{F}_{Y \to X}$ for the minimal two variable VAR (6) we use the following theory (Doob, 1953): for a VAR of the form (B.2) we define the pth order square matrix coefficients polynomial $A(z) \equiv I - \sum_{k=1}^{p} A_k z^k$. The *cross-power spectral density* (CPSD) at frequency ω of a stationary multivariate stochastic process $\mathbf{U}_t$ is defined as $S(z) \equiv \sum_{k=-\infty}^{\infty} \Gamma_k z^k$, where $z = e^{-i\omega}$ (i.e. z lies on the unit circle $|z| = 1$) and $\Gamma_k \equiv \mathrm{cov}(\mathbf{U}_t, \mathbf{U}_{t-k})$ is the autocovariance at k lags. A standard result then states that for a covariance-stationary VAR process of the form (B.2), the CPSD factorises uniquely as

$$S(z) = H(z) \Xi H^*(z) \tag{C.1}$$

where the *transfer function* $H(z) \equiv A(z)^{-1}$ (matrix inverse) has a unique extention to a holomorphic function on the unit disc $|z| \le 1$ with $H(0) = I$, $\Xi = \mathrm{cov}(\varepsilon_t)$ is the residuals covariance matrix and '*' denotes matrix conjugate transpose. Although efficient algorithms are available to perform the *spectral factorisation* (C.1) numerically (Whittle, 1963; Wilson, 1972), no generally applicable analytic algorithm is known. However, if the factorisation is known, it enables extraction of the residuals covariance matrix Ξ from $S(z)$. If $\mathbf{U}_t$ decomposes into jointly-continuous processes $\mathbf{X}_t, \mathbf{Y}_t$ then the spectral factorisation formula (C.1) applies in particular to the component $S_{xx}(z)$ of the CPSD of $\mathbf{U}_t$—which is just the CPSD of $\mathbf{X}_t$—allowing calculation of the residuals covariance matrix Σ' for the reduced VAR (3). Then $\mathcal{F}_{\mathbf{Y} \to \mathbf{X}}$ may be easily obtained from (4) (note that the residuals covariance matrix Σ for the full VAR (2) is just the component Ξ_{xx} of the residuals covariance matrix of $\mathbf{U}_t$).

The coefficients polynomial and transfer function for (6) are

$$A(z) = \begin{bmatrix} 1 & -cz^{\ell} \\ 0 & 1 \end{bmatrix} \quad H(z) = \begin{bmatrix} 1 & cz^{\ell} \\ 0 & 1 \end{bmatrix} \tag{C.2}$$

so that by (C.1) the CPSD is given by

$$S(z) \equiv H(z)H^*(z) = \begin{bmatrix} 1+c^2 & cz^\ell \\ c\bar{z}^\ell & 1 \end{bmatrix} \tag{C.3}$$

From this we see that $S_{xx}(z) = 1 + c^2$, which is constant, so that X_t is just white noise with variance $1 + c^2$. This gives immediately the value $\mathcal{F}_{Y\to X} = \log\left(1 + c^2\right)$ of (7). Note that this value does not depend on the causal lag. It is trivial to show that $\mathcal{F}_{X\to Y} = 0$.

We now apply convolutions of the form (8); without loss of generality we take $g(0) = h(0) = 1$. The transformed CPSD matrix is then ((Barnett and Seth, 2011, Section 3))

$$\tilde{S}(z) \equiv G(z)S(z)G^*(z) = \begin{bmatrix} (1+c^2)g(z)g(\bar{z}) & c\bar{z}^{k-\ell}g(z)h(\bar{z}) \\ cz^{k-\ell}g(\bar{z})h(z) & h(z)h(\bar{z}) \end{bmatrix} \tag{C.4}$$

The x and y autospectra factorise trivially as $\tilde{S}_{xx}(z) = g(z)\cdot(1+c^2)\cdot g(\bar{z})$ and $\tilde{S}_{yy}(z) = h(z)\cdot 1\cdot h(\bar{z})$. Thus the residuals variances of the *reduced* regressions for the transformed VAR are unchanged: $\tilde{\Xi}'_{xx} = 1+c^2$, $\tilde{\Xi}'_{yy} = 1$. To calculate the residuals variance of the full regression we need to perform a spectral factorisation $\tilde{S}(z) = \tilde{H}(z)\tilde{\Xi}\tilde{H}^*(z)$ as in (C.1), for a transfer function $\tilde{H}(z)$ holomorphic around zero with $\tilde{H}(0) = I$ and a positive-definite covariance matrix $\tilde{\Xi}$. While as mentioned there is no known general recipe for obtaining a spectral factorisation analytically, in this simple example factorisations are not difficult to find. There are three cases:

Case I: $k < \ell$ - convolution onset delay is shorter than causal lag (this includes the case $k = 0$ when the filter is actually invertible). Then

$$\tilde{H}(z) = \begin{bmatrix} g(z) & cz^{\ell-k}g(z) \\ 0 & h(z) \end{bmatrix} \quad \tilde{\Xi} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{C.5}$$

Comparing with (C.2), we see that the convolutions effectively reduce the causal lag by k, so that causalities are unchanged: $\mathcal{F}_{\tilde{Y}\to\tilde{X}} = \mathcal{F}_{Y\to X}$ and $\mathcal{F}_{\tilde{X}\to\tilde{Y}} = 0$.

Case II: $k = \ell$ - convolution onset delay is equal to causal lag. Here the factorisation is

$$\tilde{H}(z) = \begin{bmatrix} g(z) & 0 \\ 0 & h(z) \end{bmatrix} \quad \tilde{\Xi} = \begin{bmatrix} 1+c^2 & c \\ c & 1 \end{bmatrix} \tag{C.6}$$

which yields $\mathcal{F}_{\tilde{Y}\to\tilde{X}} = \mathcal{F}_{\tilde{X}\to\tilde{Y}} = 0$, so that the original causality $\mathcal{F}_{Y\to X}$ is destroyed by the convolution and no causality is introduced in the reverse direction.

Case III: $k > \ell$ - convolution onset delay is longer than causal lag. We have

$$\tilde{H}(z) = \begin{bmatrix} g(z) & 0 \\ c(1+c^2)^{-1}z^{k-\ell}h(z) & h(z) \end{bmatrix} \quad \tilde{\Xi} = \begin{bmatrix} 1+c^2 & 0 \\ 0 & (1+c^2)^{-1} \end{bmatrix} \tag{C.7}$$

This gives $\mathcal{F}_{\tilde{Y}\to\tilde{X}} = 0$ and $\mathcal{F}_{\tilde{X}\to\tilde{Y}} = \mathcal{F}_{Y\to X}$, so that causality in the transformed VAR is effectively reversed.

References

Abbott, L.F., 1997. Synaptic Depression and Cortical Gain Control. Science 275, 221–224.

Aertsen, A., Preißl, H., 1991. Dynamics of activity and connectivity in physiological neuronal networks, in: Schuster, H. (Ed.), Nonlinear Dynamics and Neuronal Networks. VCH Publishers Inc., New York, pp. 281–302.

Aguirre, G.K., Zarahn, E., D'esposito, M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360–369.

Antoniou, A., 1993. Digital Filters: Analysis, Design, and Applications. McGraw-Hill, New York, NY.

Baccalá, L.A., Sameshima, K., 2001. Partial directed coherence: a new concept in neural structure determination. Biol Cybern 84, 463–474.

Barnett, L., Barrett, A.B., Seth, A.K., 2009. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Lett. 103, 238701.

Barnett, L., Seth, A.K., 2011. Behaviour of Granger causality under filtering: Theoretical invariance and practical application. J. Neurosci. Methods 201, 404–419.

Barrett, A.B., Barnett, L., Seth, A.K., 2010. Multivariate Granger causality and generalized variance. Phys. Rev. E 81, 41907.

Barrett, A.B., Murphy, M., Bruno, M.A., Noirhomme, Q., Boly, M., Laureys, S., Seth, A.K., 2012. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. PLoS One 7, e29072–e29072.

Bartels, R.H., Stewart, G., 1972. Solution of the equation AX + XB = C. Comm. A.C.M. 15, 820–826.

Bergstrom, A., 1984. Continuous time stochastic models and issues of aggregation over time, in: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics. Elsevier. volume 2. chapter 20, 1 edition. pp. 1145–1212.

Bonvento, G., Sibson, N., Pellerin, L., 2002. Does glutamate image your thoughts? Trends Neurosci 25, 359–364.

Breitung, J., Swanson, N.R., 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. J. Time Series Anal. 23, 651–665.

Bressler, S., Seth, A., 2011. Wiener-granger causality: A well established methodology. Neuroimage 58, 323–329.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: Emerging methods and principles. Trends Cogn Sci 14, 277–290.

Bressler, S.L., Tang, W., Sylvester, C.M., Shulman, G.L., Corbetta, M., 2008. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. Journal of Neuroscience 28, 10056–61.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S., 2004. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proceedings of the National Academy of Sciences, USA 101, 9849–9854.

Buxton, R.B., 2012. Dynamic models of BOLD contrast. Neuroimage 62, 953-61.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39, 855–864.

Cadotte, A.J., DeMarse, T.B., He, P., Ding, M., 2008. Causal measures of structure and plasticity in simulated and living neural networks. PLoS One 3, e3355–e3355.

Chang, C., Thomason, M.E., Glover, G.H., 2008. Mapping and correction of vascular hemodynamic latency in the BOLD signal. Neuroimage 43, 90–102.

Ching, S., Cimenser, A., Purdon, P.L., Brown, E.N., Kopell, N.J., 2010. Thalamocortical model for a propofol-induced alpha-rhythm associated with loss of consciousness. Proc Natl Acad Sci U S A 107, 22665–22670.

Cohen, M.X., van Gaal, S., 2012. Dynamic Interactions between Large-Scale Brain Networks Predict Behavioral Adaptation after Perceptual Errors. Cereb Cortex .

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol 6, 2683–2697.

Dayan, P., Abbott, L., 2005. Theoretical Neuroscience: Computational And Mathematical Modeling of Neural Systems. Computational Neuroscience, MIT Press.

Deshpande, G., Sathian, K., Hu, X., 2010. Effect of hemodynamic variability on Granger causality analysis of fMRI. Neuroimage 52, 884–896.

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

Doob, J., 1953. Stochastic Processes. John Wiley, New York.

Durbin, J., Watson, G.S., 1950. Testing for serial correlation in least squares regression. I. Biometrika 37, 409–428.

Edgington, E.S., 1995. Randomization Tests. Marcel Dekker, New York.

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5, e15710–e15710.

Feinberg, D.A., Yacoub, E., 2012. The rapid development of high speed, resolution and precision in fMRI. Neuroimage 62, 720-725.

Friston, K., Frith, C.D., Liddle, P.F, Frackowiak, R.S., 1993. Functional connectivity: the prinicipal-component analysis of large PET data sets. J Cereb Blood Flow Metab 13,1 5-14.

Friston, K., 2006. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press.

Friston, K., 2009. Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol 7, e33–e33.

Friston, K.J., 2011. Functional and effective connectivity: A review. Brain Connectivity 1, 13–36.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. Neuroimage 12, 466–477.

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

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc. 77, 304–313.

Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. J. Am. Stat. Assoc. 79, 907–915.

Goebel, R., Roebroeck, A., Kim, D.S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging 21, 1251–1261.

Gow, D.W., Segawa, J.A., Ahlfors, S.P., Lin, F.H., 2008. Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. Neuroimage 43, 614–623.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Hamilton, J.D., 1994. Time series analysis. Princeton University Press, Princeton, NJ.

Handwerker, D.A., Gonzalez-Castillo, J., D'Esposito, M., Bandettini, P.A., 2012. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. Neuroimage 62, 1017-23.

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651.

Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. John Wiley, New York.

Hwang, K., Velanova, K., Luna, B., 2010. Strengthening of top-down frontal cognitive control networks underlying the development of inhibitory control: a functional magnetic resonance imaging effective connectivity study. J Neurosci 30, 15535–15545.

Im, C.H., Jung, Y.J., Lee, S., Koh, D., Kim, D.W., Kim, B.M., 2010. Estimation of directional coupling between cortical areas using Near-Infrared Spectroscopy (NIRS). Opt Express 18, 5730–5739.

Izhikevich, E., 2003a. Simple model of spiking neurons. IEEE Transactions on Neural Networks 14, 1569–1572.

Izhikevich, E.M., 2003b. Simple model of spiking neurons. IEEE Transactions on Neural Networks 14, 1569–72.

Izhikevich, E.M., 2004. Which model to use for cortical spiking neurons? IEEE Transactions on Neural Networks 15, 1063–70.

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

Kaminski, M., Ding, M., Truccolo, W.A., Bressler, S.L., 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biological Cybernetics 85, 145–157.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin.

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., Logothetis, N.K., 2012. The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. J Neurosci 32, 1395–1407.

Markram, H., Wang, Y., Tsodyks, M., 1998. Differential signaling via the same axon of neocortical pyramidal neurons. Proceedings of the National Academy of Sciences of the United States of America 95, 5323–8.

Marreiros, A.C., Kiebel, S.J., Friston, K.J., 2008. Dynamic causal modelling for fMRI: a two-state model. Neuroimage 39, 269–278.

McQuarrie, A.D.R., Tsai, C.L., 1998. Regression and Time Series Model Selection. World Scientific Publishing, Singapore.

Menon, R.S., 2012. Mental chronometry. Neuroimage 62, 1068-71.

Menon, R.S., Luknowsky, D.C., Gati, J.S., 1998. Mental chronometry using latency-resolved functional MRI. Proc Natl Acad Sci U S A 95, 10902–10907.

Morf, M., Viera, A., Lee, D.T.L., Kailath, T., 1978. Recursive multichannel maximum entropy spectral estimation. IEEE Trans. Geosci. Elec. 16, 85 –94.

Mountcastle, V.B., 1998. Perceptual Neuroscience: The Cerebral Cortex. Harvard University Press, Cambridge, MA.

Murayama, Y., Biessmann, F., Meinecke, F.C., Müller, K.R., Augath, M., Oeltermann, A., Logothetis, N.K., 2010. Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. Magn Reson Imaging 28, 1095–1103.

Nalatore, H., Ding, M., Rangarajan, G., 2009. Denoising neural data with state-space smoothing: Method and application. J Neurosci Methods 179, 131–141.

Niessing, J., Ebisch, B., Schmidt, K.E., Niessing, M., Singer, W., Galuske, R.A.W., 2005. Hemodynamic signals correlate tightly with synchronized gamma oscillations. Science 309, 948–951.

Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P., 2010. Comparing families of dynamic causal models. PLoS Comput Biol 6, e1000709–e1000709.

Rasmussen, P.M., Abrahamsen, T.J., Madsen, K.H., Hansen, L.K., 2012. Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation. Neuroimage 60, 1807–1818.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using granger causality and fmri. NeuroImage 25, 230–242.

Roebroeck, A., Formisano, E., Goebel, R., 2011a. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. Neuroimage 58, 296-302.

Roebroeck, A., Seth, A., Valdes-Sosa, P., 2011b. Causal time series analysis of functional magnetic resonance imaging data. Journal of Machine Learning Research 12, 65–94.

Rosa, M.J., Kilner, J.M., Penny, W.D., 2011. Bayesian comparison of neurovascular coupling models using EEG-fMRI. PLoS Comput Biol 7, e1002070–e1002070.

Ryali, S., Supekar, K., Chen, T., Menon, V., 2011. Multivariate dynamical systems models for estimating causal interactions in fMRI. Neuroimage 54, 807–823.

Scheeringa, R., Fries, P., Petersson, K.M., Oostenveld, R., Grothe, I., Norris, D.G., Hagoort, P., Bastiaansen, M.C.M., 2011. Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron 69, 572–583.

Schippers, M.B., Renken, R., Keysers, C., 2011. The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. Neuroimage 57, 22–36.

Schmolesky, M.T., Wang, Y., Hanes, D.P., Thompson, K.G., Leutgeb, S., Schall, J.D., Leventhal, A.G., 1998. Signal timing across the macaque visual system. J Neurophysiol 79, 3272–3278.

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

Seth, A.K., 2010. A MATLAB toolbox for Granger causal connectivity analysis. J Neurosci Methods 186, 262–273.

Smith, J.F., Pillai, A., Chen, K., Horwitz, B., 2010. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. Neuroimage 52, 1027–1040.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for FMRI. Neuroimage 54, 875–891.

Softky, W.R., Koch, C., 1993. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. The Journal of Neuroscience 13, 334–50.

Solo, V., 2007. On causality I: Sampling and noise, in: Proceedings of the 46th IEEE Conference on Decision and Control, pp. 3634–3639.

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E.M., Breakspear, M., Friston, K.J., 2008. Nonlinear dynamic causal models for fMRI. Neuroimage 42, 649–662.

Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: Influence, causality and biophysical modeling. Neuroimage 58, 339–361.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. T. Am. Math. Soc. 54, 426–482.

Wen, X., Yao, L., Liu, Y., Ding, M., 2012. Causal interactions in attention networks predict behavioral performance. J Neurosci 32, 1284–1292.

Whittle, P., 1963. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 50, 129–134.

Wiener, N., 1956. The theory of prediction, in: Beckenbach, E.F. (Ed.), Modern Mathematics for Engineers. McGraw Hill, New York, pp. 165–190.

Wilks, S.S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 6, 60–62.

Wilson, G.T., 1972. The factorization of matricial spectral densities. SIAM J. Appl. Math. 23, 420–426.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12, 1–24.

Zucker, R.S., Regehr, W.G., 2002. Short-term synaptic plasticity. Annual Review of Physiology 64, 355–405.

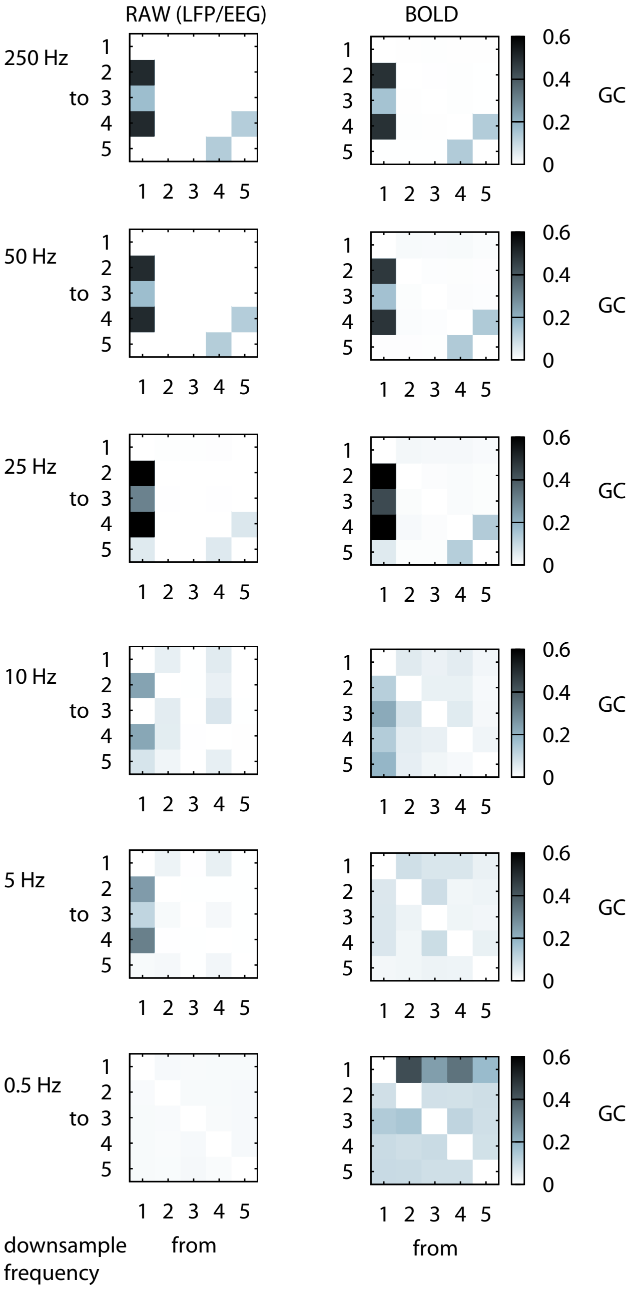

Figure 7: Conditional GC analysis of data generated from the simple VAR model. RAW: VAR model output. BOLD: VAR output convolved with confounding HRFs. Downsampled at 250, 50, 25, 10, 5 and 0.5 Hz.

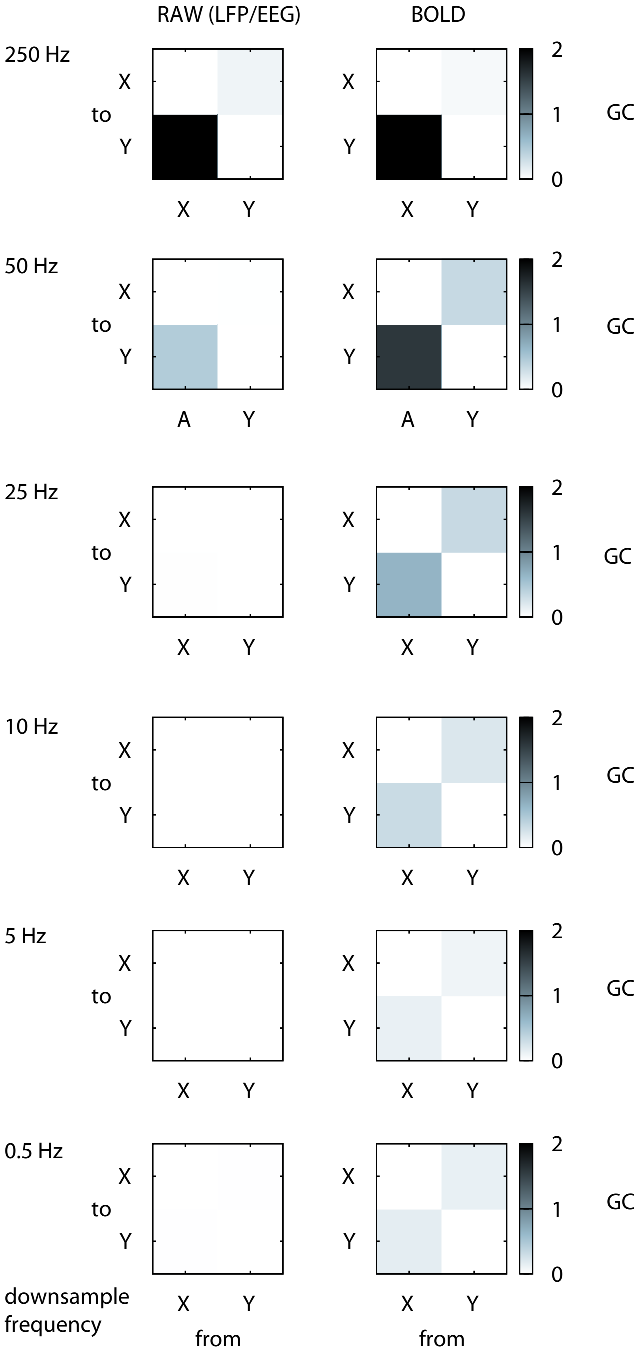

Figure 8: Conditional GC analysis of data generated from the spiking neuronal model. RAW: LFP output. BOLD: generated from LFP output by BW model. Downsampled at 250, 50, 20, 10, 5 and 0.5 Hz.

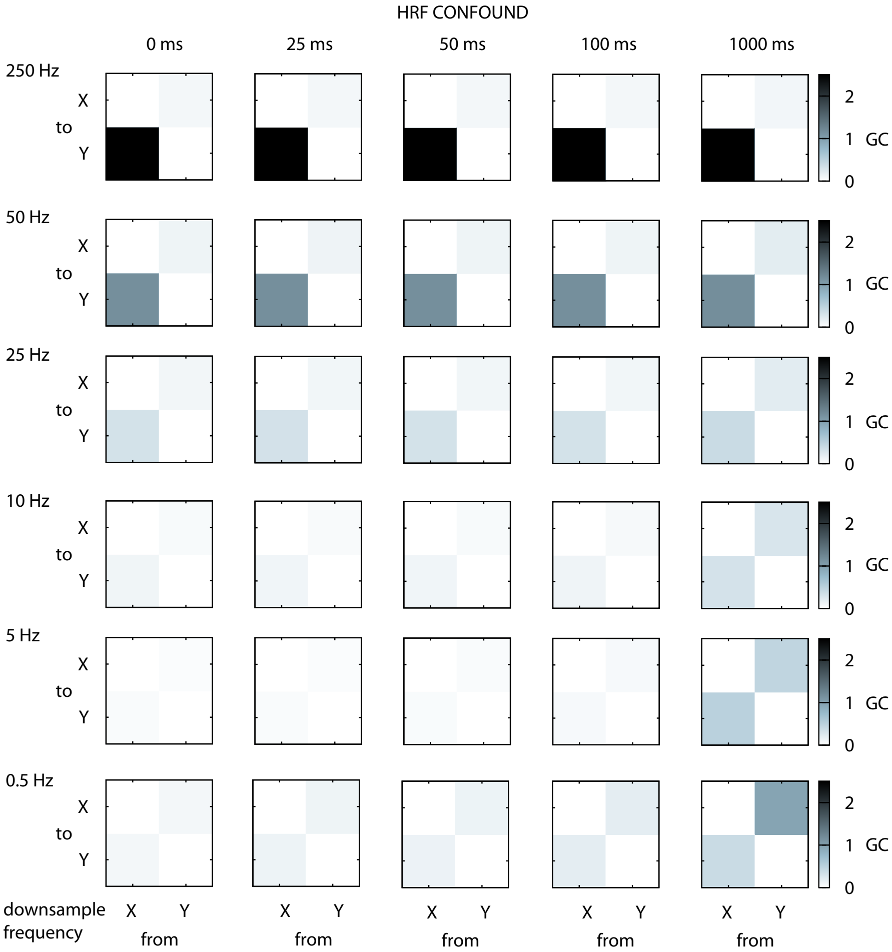

Figure 9: GC analysis of data generated from the spiking neuronal model. LFP output is downsampled at 250, 50, 20, 10, 5 and 0.5 Hz and convolved with confounding HRF convolutions with differential delays-to-peak (HRFDEL) of 0, 25, 50, 100 and 1000 ms around a reference delay-to-peak of 4 secs. Note that the X → Y causal delay in the LFP is 20 ms.

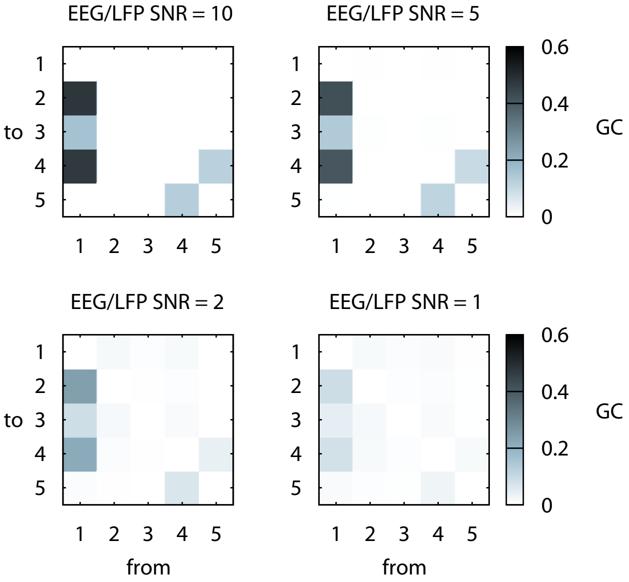

Figure 10: Conditional GC analysis of data generated from the simple VAR model. EEG/LFP: Data generated from the VAR model downsampled at 250 Hz. Gaussian white noise with given signal-to-noise ratio (SNR) then applied pre-downsampling.

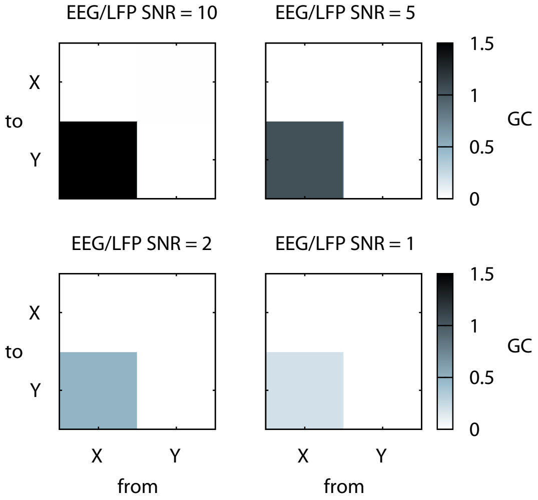

Figure 11: GC analysis of data generated from the spiking neuronal model. EEG/LFP: LFP output downsampled at 250 Hz. Gaussian white noise with given signal-to-noise ratio (SNR) then applied pre-downsampling.