How to detect the Granger-causal flow direction in the presence of additive noise?

Martin Vinck, Lisanne Huurdeman, Conrado A. Bosman, Pascal Fries, Francesco P. Battaglia, Cyriel M.A. Pennartz, Paul H. Tiesinga

Abstract

Granger-causality metrics have become increasingly popular tools to identify directed interactions between brain areas. However, it is known that additive noise can strongly affect Granger-causality metrics, which can lead to spurious conclusions about neuronal interactions. To solve this problem, previous studies have proposed to detect Granger-causal directionality, i.e. the dominant Granger-causal flow, using either the slope of the coherency (Phase Slope Index; PSI), or by comparing Granger-causality values between original and time-reversed signals (reversed Granger testing). We show that for ensembles of vector autoregressive (VAR) models encompassing bidirectionally coupled sources, these alternative methods do not correctly measure Granger-causal directionality for a substantial fraction of VAR models, even in the absence of noise. We then demonstrate that uncorrelated noise has fundamentally different effects on directed connectivity metrics than linearly mixed noise, where the latter may result as a consequence of electric volume conduction. Uncorrelated noise only weakly affects the detection of Granger-causal directionality, whereas linearly mixed noise causes a large fraction of false positives for standard Granger-causality metrics and PSI, but not for reversed Granger testing. We further show that we can reliably identify cases where linearly mixed noise causes a large fraction of false positives by examining the magnitude of the instantaneous influence coefficient in a structural VAR model. By rejecting cases with strong instantaneous influence, we obtain improved detection of Granger-causal flow between neuronal sources in the presence of additive noise. These techniques are applicable to real data, which we demonstrate using actual area V1 and area V4 LFP data, recorded from the awake monkey performing a visual attention task.

Keywords: Granger-causality; phase slope index; volume conduction; reversed time series; vector autoregressive modelling

The Wiener-Granger definition of causality allows inference of causal relationships between interacting stochastic sources. Causality analysis methods have been applied in many fields, including physics, econometrics, geology, ecology, genetics, physiology and neuroscience (Barnett et al., 2009; Bernasconi and Konig, 1999; Bressler and Seth, 2011; Brovelli et al., 2004; Ding et al., 2006; Geweke, 1982; Granger, 1969; Gregoriou et al., 2009; Hiemstra and Jones, 1994; Hu and Nenov, 2004; Kaufmann and Stern, 1997; Lozano et al., 2009; Marinazzo et al., 2008; Nolte et al., 2008; Rosenblum and Pikovsky, 2001; Salazar et al., 2012; Schreiber, 2000; Smirnov and Mokhov, 2009; Staniek and Lehnertz, 2008; Sugihara et al., 2012). Standard Granger-causality metrics are typically based on linear vector autoregressive (VAR) modeling, with Granger-causality $f_{i \rightarrow j}$ defined by examining $x_i$’s effect on the residual errors in forecasting $x_j(t)$ (Geweke (1982); Granger (1969), eqs. 4-5). In the neurosciences, Granger-causality metrics have become increasingly popular tools to identify functional, frequency-specific directed influences between brain areas (e.g. (Bernasconi and Konig, 1999; Bressler and Seth, 2011; Brovelli et al., 2004; Ding et al., 2006; Friston et al., 2014)). Two recent studies have shown interesting applications of Granger causality to characterize functional interactions in the visual system. Bastos et al. (2014); van Kerkoerle et al. (2014) have shown that gamma frequencies contribute to a feedforward flow of information, whereas alpha and beta frequencies contribute to flow of information in the feedback direction. Interestingly, Bastos et al. (2014) succeeded to reconstruct the visual hierarchy based on anatomical tracing studies on the mere basis of examining the asymmetry of Granger-causality spectra, and showed that this cortical hierarchy was task-dependent.

Granger-causality metrics were originally developed for systems whose measurements are not corrupted by additive noise. It has been shown that they can be strongly affected by both uncorrelated and linearly mixed additive noise (Albo et al., 2004; Friston et al., 2014; Haufe et al., 2012a,b; Nalatore et al., 2007; Newbold, 1978; Nolte et al., 2008; Seth et al., 2013; Sommerlade et al., 2012). Nolte et al. (2008) showed that adding a linear mixture of noise sources (correlated noise) often leads to a misidentification of causal driver and recipient when using the standard Granger-causality metrics proposed by Granger (1969). This is an important issue in the neurosciences, as electric currents spread instantaneously from single brain or noise sources to multiple sensors (“volume conduction”),...
posing a major problem especially for scalp EEG (Electro-encephalography), MEG (Magneto-encephalography), and intracranial LFP (Local Field Potential) signals (Nolte et al., 2004; Nunez and Srinivasan, 2006; Stam et al., 2007; Vinck et al., 2011). This problem can, as far as the estimation of symmetric, undirected connectivity (like coherence, phase locking value) metrics is concerned, be adequately addressed by using metrics that are based on the imaginary component of the cross-spectral density (Nolte et al., 2004; Stam et al., 2007; Vinck et al., 2011). In this paper, we ask whether directed connectivity measures like Granger-causality can be protected against linearly mixed noise as well.

Recently, two directed connectivity measures were introduced to address the volume conduction problem. Nolte et al. (2008) proposed to detect Granger-causal directionality by examining whether fluctuations of one signal precede fluctuations of another signal in time - i.e. temporal precedence - using a measure called phase slope index (PSI). Haufe et al. (2012a,b) proposed to protect Granger-causality metrics against linearly mixed noise by comparing Granger-causality values with those of another signal in time - i.e. temporal precedence - using a model that contains an explicit instantaneous transfer coefficient. This allows us to reject cases where the instantaneous transfer is too large compared to the transfer at other lags. In this paper, we will be concerned with a bivariate signal $x(t)$ described by a bivariate VAR model of order $M$

$$x(t) = \sum_{\tau=1}^{M} A(\tau)x(t-\tau) + U(t),$$

where innovation $U(t)$ - the remaining error after incorporating the predictions from past values of $x(t)$ - has covariance matrix

$$\Sigma \equiv \text{Cov}(U(t), U(t)).$$

We refer to $x_1$ and $x_2$ as the signal sources. The matrices $A(\tau)$ hold the real-valued VAR coefficients. Also $x(t)$ can be represented by the restricted AR model

$$x(t) = \sum_{\tau=1}^{M} F(\tau)x(t-\tau) + V(t)$$

with diagonal coefficient matrix $F(\tau)$ and $\Omega \equiv \text{Cov}(V(t), V(t))$. The standard measure of Granger-causal flow is defined by the

metrics, does not constrain the false positives rate at acceptable levels. In contrast, RGT yields a much smaller fraction of false positives and much better overall performance than standard Granger and PSI, although it still shows failures in a significant fraction of test cases. This paper also aims to advance the theoretical analysis of PSI and RGT; in particular, we use theoretical analysis to identify regimes in which RGT always fails or succeeds.

We find that further performance gains are achievable beyond those obtained by RGT, by indirectly measuring the amount of linearly mixed noise impinging on two measurement sensors. The idea of this approach is quite simple but effective: We can examine the degree to which there is instantaneous (i.e., zero-lag) feedback between time series by using a structural VAR model that contains an explicit instantaneous transfer coefficient. This provides a means to reduce the false positives rate and to improve the overall performance of the analysis in terms of the true and false positives mix.

We apply these techniques to actual LFP data obtained from areas V1 and V4 in the awake monkey performing a visual attention task.

1. Introduction of Granger analysis techniques and VAR model with additive noise

In this section, we define the basic VAR model, the VAR model with additive noise included, the various directed connectivity metrics, and performance measures for the different metrics.

1.1. The bivariate VAR model and a measure of linear Granger feedback

In this paper, we will be concerned with a bivariate signal $x(t)$ described by a bivariate VAR model of order $M$
log-ratio of the variances of the innovation errors (Granger, 1969)

\[ f_{j \rightarrow i} \equiv \ln \left( \frac{\sigma_i^2}{\sigma_j^2} \right), i \neq j. \]  

(4)

The feedback metric \( f_{j \rightarrow i} \) measures the degree to which past values of \( x_i(t) \) improve the prediction of future values of \( x_j(t) \) relative to what can be derived from past values of \( x_j(t) \). We define \( x_1 \) to be Granger-causally dominant over \( x_2 \) if the Granger-causal directionality measure

\[ g \equiv f_{1 \rightarrow 2} - f_{2 \rightarrow 1} > 0. \]  

(5)

In this paper, we study the simplified problem of detecting \( g \), where \( g \) is the sign function, from noisy data, as in Nolte et al. (2008); we are not concerned with the problem of estimating \( f_{2 \rightarrow 1} \) and \( f_{1 \rightarrow 2} \) separately. The problem is stated as providing a measure of \( g \) that optimizes performance in terms of the false and true positives mix. The precise weighting of false positives and true positives depends on the practical application, but typically false positives outweigh true positives in importance by an order of magnitude, which is reflected by the common convention that the false discovery rate (FP/TP) should not exceed a low threshold (e.g. 0.05 or 0.1) (Benjamini, 2010). The problem can therefore be formulated as maximizing a direction detection performance function

\[ U \equiv -G \cdot \text{frac}_{TP} + \text{frac}_{FP} \]  

(6)

where the gain \( G \) is set to 10 in this paper (as in Haufe et al. (2012b)).

1.2. The PSI (Phase Slope Index)

Recently, two procedures were proposed to detect Granger-causal directionality in the presence of additive linearly mixed noise (Haufe et al., 2012a,b; Nolte et al., 2008). PSI is defined

\[ \psi \equiv \int \Im(C'(\omega))d\omega, \]  

with slope of the coherency

\[ C'(\omega) \equiv \lim_{d\omega \rightarrow 0} \frac{C(\omega)C^*(\omega + d\omega)}{d\omega}, \]  

(8)

and where \( \Im[] \) is the imaginary component. In turn, coherency is defined

\[ C(\omega) \equiv \frac{S_{12}(\omega)}{\sqrt{S_{11}(\omega)S_{22}(\omega)}} \]  

(9)

and spectral density matrix (assuming it exists) is defined

\[ S(\omega) \equiv \begin{pmatrix} S_{11}(\omega) & S_{12}(\omega) \\ S_{21}(\omega) & S_{22}(\omega) \end{pmatrix}. \]  

(10)

PSI was motivated by the argument that causal dominance implies temporal precedence (Nolte et al., 2008), with a shift of the peak of the cross-covariance function

\[ s_{12}(\tau) \equiv \text{Cov}(x_1(t), x_2(t + \tau)) \]  

(11)


by a delay of \( \Delta \) corresponding to multiplication of its Fourier Transform, the cross-spectral density \( S_{12}(\omega) \).

Coherency relates to linear prediction in the context of constructing a noncausal Wiener filter (Wiener, 1949): For the linear model \( \hat{x}_2(t) = \sum_{h=-\infty}^{\infty} h(\tau)x_1(t - \tau) \), the optimal Wiener filter kernel has Fourier Transform \( H(\omega) = \frac{S_{21}(\omega)}{S_{11}(\omega)} \), such that the minimum error \( \epsilon_{\text{min}} = \int_{-\infty}^{\infty} S_{22}(\omega) \left( 1 - |C(\omega)|^2 \right) d\omega \), where the squared coherence \( |C(\omega)|^2 \) plays a similar role as in linear regression analysis. Define the inverse Fourier Transform of the coherency function \( c(\tau) \equiv F^{-1}[C(\omega)] \). If prediction errors tend to be more reduced by the \( \tau > 0 \) than the \( \tau < 0 \) component of \( c(\tau) \), then the coherency \( C(\omega) \) will typically have a positive slope, as follows from the shift/translation property of the Fourier Transform.

Yet, it is not obvious that PSI ‘typically’ detects Granger-causal directionality for bidirectional VAR models, because it ignores the linear feedback between sources and the predictions of the time series by themselves.

1.3. RGT (Reversed Granger Testing)

Another procedure to make a decision about Granger-causal directionality is based on computing the Granger-causal feedback values \( f_{j \rightarrow i} \) for time-reversed signals \( x^r(t) = x(-t) \) (Haufe et al., 2012b). For time-reversed signals, we have spectral density matrix \( S^r(\omega) = S(\omega) \) and cross-covariance matrix \( s^r(\tau) = s(-\tau) \). RGT holds that \( x_1 \) is Granger-causally dominant over \( x_2 \) if the inequalities

\[ g \equiv f_{1 \rightarrow 2} - f_{2 \rightarrow 1} > 0 \]

\[ g^r \equiv f^r_{1 \rightarrow 2} - f^r_{2 \rightarrow 1} > 0 \]  

(12)

both hold, that is if Granger-causal directionality reverses (flips) for the time-reversed signals. Note that in the definition of \( g^r \), the individual feedback terms are already sign-reversed; therefore the signs of \( g \) and \( g^r \) should be equal. No decision about Granger-causal directionality is taken when \( g \neq g^r \) or \( g^r \neq g^r \), that is if Granger-causal directionality does not flip for time-reversed signals. Theoretical motivation for the RGT measure in case of additive linearly mixed noise will be given in Section 4.1.1.

1.4. Additive noise model

Now suppose that additive noise

\[ \epsilon(t) \equiv \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{pmatrix} \equiv K z(t) \]  

(13)

is superimposed onto the signal source \( x(t) \). Here, \( K \) is a real-valued 2x1 linear mixing matrix and \( z(t) \) is an \( S \times 1 \) vector of \( S \) noise sources. We assume that the \( S \) noise sources are uncorrelated with each other, i.e. \( \text{Cov}[z_p(t), z_s(t + \tau)] = 0 \) for all \( s \neq p \) and \( \tau \). We also assume that the noise sources are each uncorrelated with the signal sources \( x(t) \), such that \( \text{Cov}[z_s(t), x_s(t + \tau)] = 0 \) for all \( s \). We refer to the additive noise \( \epsilon(t) \) as uncorrelated (Section 3) if the equality \( \text{Cov}[\epsilon_1(t), \epsilon_2(t + \tau)] = 0 \) holds for all \( \tau \) (e.g. if for \( S = 2 \), \( K \) is diagonal or anti-diagonal) and as linearly
mixed (Section 4) otherwise (e.g. if for $S = 2$, $K$ is not strictly diagonal or anti-diagonal).

The resulting time series with a superposition of additive noise is

$$x^{(e)}(t) \equiv (1 - \gamma)x(t) + \gamma \epsilon(t), \quad (14)$$

where the parameter $\gamma$ determines the signal-to-noise ratio. Throughout this paper, superscript $(e)$ always indicates variables that result from mixing signal and noise. From the linearity of the cross-covariance function and the Fourier Transform, it follows that the spectral density matrix equals

$$S^{(e)}(\omega) = (1 - \gamma)^2 S(\omega) + \gamma^2 S_\epsilon(\omega), \quad (15)$$

and that the cross-covariance function equals

$$s^{(e)}(\tau) = (1 - \gamma)^2 s(\tau) + \gamma^2 s_\epsilon(\tau). \quad (16)$$

That is, the resulting spectral density matrix $S^{(e)}(\omega)$ can be written as a linear mixture of the original spectral density matrix $S(\omega)$ and the spectral density matrix of the noise sources $S_\epsilon(\omega)$; the same holds for the resulting cross-covariance function $s^{(e)}(\tau)$.

### 1.5. Difference between additive noise and innovation

It is important to note that additive noise and innovation exhibit fundamentally different effects. The innovation process $U$ affects future values $x(t + \tau)$ through $A$ (eq. 1), whereas additive noise merely distorts the measurement of $x(t)$ (eq. 14), such that it cannot be subsumed by the innovation $U$ in eq. 1.

### 1.6. Definition of true positives and false positives

Throughout the text, the standard measure of Granger-causal feedback $\text{sgn}(g)$ for the noise-free VAR in eq. 1 ($\gamma = 0$) is taken as a ‘ground-truth’ reference for defining true positives and false positives. We define false negatives, and true and false positives as follows.

(i) If PSI is computed analytically, then we simply examine $\text{sgn}(\psi^{(i)})$, i.e. no statistical testing has to be performed (see Figs 2 and 4). In that case, a true positive for PSI is defined by the equality $\text{sgn}(\psi^{(i)}) = \text{sgn}(g)$; a false positive by the inequality. Let $\frac{\text{frac}_{TP}}{0} \in [0, 1]$ and $\frac{\text{frac}_{FP}}{0} \in [0, 1]$ denote the fractions of false positives and true positives. For PSI, the equality $\frac{\text{frac}_{TP}}{0} + \frac{\text{frac}_{FP}}{0} = 1$ then holds, unless $\gamma = 1$ in eq. 14. Note that with analytical computation, there are no false negatives for PSI, because PSI either yields a value that is positive or negative (since a value of zero will in practice not occur), which means that it either detects the G-causal directionality correctly (true positive) or incorrectly (false positive).

(ii) If PSI is estimated from finite data traces (Fig 6), then statistical testing is performed to test against the null hypothesis that $\psi = 0$. In this case, PSI can yield false negatives, because it can fail to reach statistical significance, and the equality $\frac{\text{frac}_{TP}}{0} + \frac{\text{frac}_{FP}}{0} + \frac{\text{frac}_{FN}}{0} = 1$ holds, where a true positive requires the null hypothesis to be rejected and the equality $\text{sgn}(\psi^{(i)}) = \text{sgn}(g)$ to hold true.

(iii) If RGT is computed analytically, then it has a true positive if the equalities

$$\text{sgn}(g^{(i)}) = \text{sgn}(g),$$

both hold, and a false positive if the inequalities $\text{sgn}(g^{(i)}) \neq \text{sgn}(g)$ and $\text{sgn}(g^{(i)}) \neq \text{sgn}(g)$ both hold; it yields a false negative otherwise.

(iii-a) For RGT, the equality $\frac{\text{frac}_{TP}}{0} + \frac{\text{frac}_{FP}}{0} + \frac{\text{frac}_{FN}}{0} = 1$ holds in case of added noise.

(iii-b) In case of no additive noise ($\gamma = 0$ in eq. 14), the equality $\frac{\text{frac}_{TP}}{0} + \frac{\text{frac}_{FN}}{0} = 1$ holds.

(iv) If RGT is estimated from finite traces (Fig 6), then statistical testing is performed in addition to the requirements in eq. 17.

The false discovery rate (FDR) is defined $\text{FDR} \equiv \frac{\text{frac}_{FP}}{\text{frac}_{TP}}$.

### 2. Case of no additive noise

In this section we consider the case of no additive noise, i.e. the equality $\gamma = 0$ holds, where the noise amplitude parameter $\gamma$ is defined in eq. 14. The case of additive noise will be discussed in Sections 3-4.

#### 2.1. Theoretical considerations for the case of no additive noise

We note that while the parameters of the restricted AR model (eq. 3) are invariant to a time-reversal of $x(t)$, the full VAR model (eq. 1) and associated metrics $f_{i \rightarrow j}$ (eq. 4) are not always reversible.

In Appendix A, we provide a theoretical analysis showing that RGT performs always well in the regime where cross- and auto-correlations have small magnitudes (as confirmed by Fig 1A-B). This case occurs for example for near-Poissonian spike trains from separate brain areas, which have small-magnitude autocorrelations at all non-zero lags, and may have small-magnitude cross-correlations at all lags as well.

In Appendix A, we also perform a theoretical analysis showing that RGT performs always well if the autocorrelation functions of $x_1$ and $x_2$ are similar. This result is important because we can assume that signals from different neocortical areas often have quite similar autocorrelation functions (implying similar power spectra of signals with standardized variance), a result of the relatively homogeneous architecture of the neocortical column. The result does not assume that the time series are optimally described by a linear model, but merely assumes that the autocorrelation functions of $x_1$ and $x_2$ (and power spectra) are relatively similar.

#### 2.2. Simulations for the case of no additive noise: Methods

##### 2.2.1. Methods: Choice of priors

We evaluated the performance of the various metrics analytically, i.e. without using finite-length signals or particular spectral estimation techniques, or making assumptions about the type of probability distribution of the innovation process
and relatively 'di-directional and relatively 'di-flow. (2008) were limited to cases with unidirectional Granger-causal 'prising' or 'nontypical'. Simulations performed by Nolte et al. (2008) were limited to cases with unidirectional Granger-causal flow.

Here we consider classes of priors that encompass both uni-directional and relatively 'difficult' bidirectional VAR models that are more representative of actual cortical interactions. Assuming that variables interact at a delay, we set the covariance matrix to be the identity matrix, i.e. $\Sigma = I$. We examined both a normal prior and a shrinkage prior on the space of VAR models, and evaluate a wide range of parameter settings to generalize our results. The normal prior is defined

$$A_i(\tau) \sim N(0, \sigma^2_A),$$

(18)

where $N$ is the normal (Gaussian) distribution. Larger coefficients $\sigma_A$ correspond to a larger variance (i.e. magnitude) of VAR coefficients and larger Granger feedback (and coherence) values; note that for large $\sigma_A$ many VAR models become unstable. No assumption is made about the distribution (e.g. Gaussian or non-Gaussian) of $U(t)$ beyond its covariance $\Sigma$.

The shrinkage prior is a popular prior in macro-economic Bayesian estimation and is defined

$$A_i(\tau) \sim N(0, \theta^2 \lambda^2/\tau^2),$$

(19)

with the coupling parameter $\theta = 1$ for $i = j$ and $0 \leq \theta \leq 1$ otherwise (Litterman, 1986). In this case the parameter $\lambda$ (comparable to $\sigma_A$ in eq. 18) controls the variance of the VAR coefficients, and $\theta$ the coupling strength. The shrinkage prior assumes that more recent lags carry more predictive power than distant lags (shrinkage $1/\tau^2$), reflecting the decay of synaptic potentials in the brain, i.e. the finite influence of an impulse in one neuronal group on another neuronal group. It also assumes that a time series is a better forecaster of itself than of other time series ($\theta \in [0, 1]$), reflecting the architecture of the neocortex in the sense that the majority of neuronal connections is local (Buzsáki, 2006), such that a local group of neurons generating one LFP/EEG/MEG signal is expected to have a much greater causal influence on itself than on a distant group of neurons that generates another LFP/EEG/MEG signal. Furthermore, it encompasses both unidirectional and bidirectional VAR models, which both occur in the nervous system. Probabilities of false and true positives were estimated by Monte Carlo sampling an ensemble of stable VAR models from the prior $P(A(\tau))$, and computing the various Granger-causality metrics analytically as indicated in the following subsections. We note that it would not be possible to evaluate a flat (noninformative) prior on $\mathbf{A}$, because it would be an improper prior having infinite support that does not allow for Monte Carlo sampling.

Because most VAR models become unstable for large variance of the VAR coefficients $\sigma_A$, we iteratively updated draws by replacing the maximum squared element of $\mathbf{A}$ with a new random variate from $N(0, \sigma^2_A)$, until the VAR model became stable. For the shrinkage prior we directly rejected the small fraction of unstable VAR models.

2.2.2. Computation of the PSI

PSI values were obtained by computing the spectral density matrix $S(\omega)$ analytically and using the broad-band frequency spectrum, sampled at $10^3$ (without loss of generalization) frequencies. Starting from the VAR model in eq. 1 the spectral density matrix is obtained by using the stochastic spectral (Cramer) representation for the wide-sense stationary time series

$$x_j(t) = \int_0^\pi e^{i\omega t} dZ_j(\omega)$$

(20)

(Granger, 1969; Percival and Walden, 1993) (in analogy to the Fourier Transform of a deterministic signal) such that

$$\begin{bmatrix}
\rho_{11}(\omega) & \rho_{12}(\omega) \\
\rho_{21}(\omega) & \rho_{22}(\omega)
\end{bmatrix} = \begin{bmatrix}
\rho_{11}(\omega) & \rho_{12}(\omega) \\
\rho_{21}(\omega) & \rho_{22}(\omega)
\end{bmatrix}$$

(21)

where $a_{ij}$ is defined in terms of the Discrete Fourier Transform of the VAR coefficients

$$a_{ij}(\omega) \equiv \delta_{ij} - \sum_{\tau=1}^{M} A_{ij}(\tau) e^{-i\omega \tau},$$

(22)

and where the $dZ$’s are the spectral representations of $x(t)$ and $U(t)$. The spectral density matrix is now given as

$$S(\omega) \equiv \begin{bmatrix}
\rho_{11}(\omega) & \rho_{12}(\omega) \\
\rho_{21}(\omega) & \rho_{22}(\omega)
\end{bmatrix} = \begin{bmatrix}
\rho_{11}(\omega) & \rho_{12}(\omega) \\
\rho_{21}(\omega) & \rho_{22}(\omega)
\end{bmatrix}$$

(23)

where the ‘transfer function’ $H(\omega) \equiv a(\omega)^{-1}$ and $\ast$ is the adjoint operator (Hermitian conjugate). Thus, the spectral density matrix can be analytically computed starting from the coefficients of a given VAR model. We then computed PSI, using eq. 7, sampling at $10^3$ frequencies (results were not affected by increasing the number of frequencies, as the spectral functions are relatively smooth with the model orders used, resulting in only few peaks in the coherence spectra, see Fig 1D).

Note that using an iterative spectral factorization algorithm, the VAR model can also be obtained from $S(\omega)$ in return (Dhamala et al., 2008; Wilson, 1972). The spectral Granger values are now defined (Dhamala et al., 2008; Geweke, 1982)

$$F_{j \rightarrow i}(\omega) \equiv \frac{S_{ii}}{S_{ii} - \left(\Sigma_{jj} - \frac{\sigma^2_j}{\tau^2}\right) |H_{ji}(\omega)|^2}.$$  

(24)

2.2.3. Computation of RGT

For the $g^{rev}$ metric, i.e. the Granger-causal dominance for time-reversed signals, we computed the VAR coefficients $A^{rev}(\tau)$ for the time-reversed signals $x(-t)$ by (i) analytically solving the Yule-Walker equations

$$s(\tau) = \sum_{m=1}^{M} A(m)s(\tau - m) + \delta_{\tau0}\Sigma$$

(25)
for the cross-covariance function \( s(\tau) \) given specified \( A(m) \) and \( \Sigma \), (ii) using the equality \( s^{rev}(\tau) = s(\tau) \), and (iii) analytically solving the Yule-Walker equations for \( A^{rev} \), where

\[
s^{rev}(\tau) = \sum_{m=1}^{M} A^{rev}(m)s^{rev}(\tau - m) + \delta_{\tau 0}\Sigma^{rev}
\]

We then used eq. 4 and 5 to compute \( g^{rev} \). Spectral matrix factorization (Dhamala et al., 2008) of the spectral density matrix \( S(\omega) \) yielded quantitatively nearly identical results (data not shown).

### 2.3. Simulations for the case of no additive noise: Results

Fig 1A-B shows the true and false positives ratios as a function of the parameters \( \sigma_{x} \) and \( \lambda \) that control the variance (i.e., magnitude) of the VAR coefficients, and the model order \( M \). The fraction of false negatives is not shown for RGT as for RGT, \( 1 -frac_{\text{TP}}{\text{FP}} = frac_{\text{FP}}{\text{FN}} \) in case of no additive noise. For both normal and shrinkage priors, PSI and RGT failed to identify the Granger-causally dominant variable for a large fraction of simulated VAR models (Fig 1A-B). The fraction of failures \( [1 - frac_{\text{TP}}{\text{FP}}] \) increased as a function of the magnitude of VAR coefficients (\( \sigma_{x} \) and \( \lambda \)). Because, in case of no noise, RGT does not yield false positives but only true positives and false negatives, it yielded much better performance than PSI in terms of a gain function \( U = \text{frac}_{\text{TP}}{\text{FP}} - 10\text{frac}_{\text{FN}}{\text{FP}} \), except for the case of a unidirectional prior (see Supplementary Information). Although PSI is a directed connectivity metric and a valid measure of the temporal ordering of signals, we conclude that it does not constitute a valid measure of Granger-causality in general, because the FDR can reach 30-40% for many priors. Thus, for PSI to be a valid measure of Granger-causality, one needs to assume that the interaction between sources is nearly unidirectional.

RGT performed especially well in the regime of the normal prior with small variance of the VAR coefficients. The good performance of RGT in this regime is conform our analytical results presented in Appendix A. This regime yielded realistic peak coherences (0.1-0.4) for brain signals (Fig 1C), however it also yielded signals that are too white (i.e., relatively flat power spectrum) for the colored brain EEG/MEG signals (Bosman et al., 2012; Brovelli et al., 2004; Buschman and Miller, 2007; Gregoriou et al., 2009; Saalmann et al., 2012; Salazar et al., 2012) (although not for e.g. near-Poissonian spike trains). The most difficult regime for RGT that yielded both colored signals and realistic peak coherences for MEG and EEG brain signals was the case of the shrinkage prior with \( \lambda \) large (i.e., large magnitude of VAR coefficients) and \( \theta \) small (i.e., weak coupling). In practice, we expect the performance of RGT to be better however: In our simulations, the two variables \( x_{1} \) and \( x_{2} \) could have widely different autocorrelation functions (and power spectra), while RGT performs better for variables having similar autocorrelation functions (Appendix A).

### 3. Case of uncorrelated noise

#### 3.1. Theoretical considerations for the case of uncorrelated noise

The case of predominantly uncorrelated noise occurs for example for spike trains, and distant current source densities or bipolarly derived intracranial signals (Mitzdorf, 1985). It remains unknown whether the dominant Granger-causal flow can be reliably detected in case of uncorrelated noise, as Nolte et al. (2008) evaluated the case of correlated noise only. In case of uncorrelated noise the equality \( s_{x,\epsilon}(\tau) = 0 \) holds for all \( \tau \), thus uncorrelated noise affects \( S_{\epsilon}^{rev}(\omega) \) and \( S_{\epsilon}^{rev}(\tau) \) only for \( i = j \), as follows from eqs. 15-16. Hence, it affects PSI only weakly and via \( C(\omega) \)'s denominator (see Fig 2A,B). In Appendix A we show that if the magnitudes of cross- and autocorrelations are relatively small, then \( \text{sgn}(g^{rev}(\tau)) \) tends to be unaffected by uncorrelated noise (see Fig 2A,B) even though \( f_{j,i}^{rev}(\tau) \) and \( f_{j,i}^{rev}(\epsilon) \) typically decrease at the rate of the signal-to-noise ratio (Appendix A). The behavior of the various measures in other regimes will be explored through simulations in this Section.

#### 3.2. Simulations for the case of uncorrelated noise: Methods

We evaluated normal and shrinkage priors by drawing an ensemble of VAR models from these priors. When the VAR coefficients for \( x(t) \) were drawn randomly from a normal (or shrinkage) prior, we also drew the VAR coefficients for the noise sources \( e \) from a normal (or shrinkage) prior, except that \( A_{n}(\tau) \), the coefficients for the noise time series, was set to be diagonal for all \( \tau \). Values of \( f_{j,i}^{rev}(\tau) \) and \( f_{j,i}^{rev}(\epsilon) \) were obtained using the following algorithm:

(i) We analytically solve the Yule-Walker equations (eq. 25) yielding expressions for \( s(\tau) \) and \( s_{\epsilon}(\tau) \), which are combined into \( s^{rev}(\tau) = (1 - \gamma)^2 s(\tau) + \gamma^2 s_{\epsilon}(\tau) \) (eq. 16).

(ii) We obtain an expression for \( A^{rev}(\epsilon) \), the coefficients for the signal corrupted by noise, where

\[
x^{rev}(\epsilon)(t) = \sum_{\tau=1}^{M} A^{rev}(\epsilon)(\tau)x^{rev}(\epsilon)(t + \tau) + U^{rev}(\epsilon)(t),
\]

by solving the Yule-Walker equations again.

(iii) From the expressions of \( A^{rev}(\epsilon)(\tau) \) and \( U^{rev}(\epsilon)(t) \) we then obtain the expression for the VAR coefficients of the time-reversed signal, \( A^{rev}(\epsilon)(\tau) \) (see Section 2.2.3).

(iv) Because of added noise, \( A^{rev}(\epsilon)(\tau) \) will typically be non-zero for \( \tau > M \), but decays at steep exponential rate. We therefore computed the order \( M^{rev}(\epsilon) \) by increasing \( M^{rev}(\epsilon) \) until all \( A^{rev}(\epsilon)(M^{rev}(\epsilon) + M + 1), \ldots, A^{rev}(\epsilon)(M^{rev}(\epsilon)) \) coefficients computed using step (iii) had absolute values <10^{-4} lower than the respective coefficients in \( A^{rev}(\epsilon)(1), \ldots, A^{rev}(\epsilon)(M) \) (where \( M \) is the original order of \( x \)), that is we set \( M^{rev}(\epsilon) \) such that the VAR coefficients had decreases by a factor of 10^{-4}. The same procedure was used for \( A^{rev}(\epsilon)(\tau) \).

Using spectral matrix factorization (Wilson, 1972) of the analytically computed spectral density matrix \( S^{rev}(\omega) \) yielded nearly identical results to those obtained with the algorithm presented above, but it was much slower to compute.
PSI was directly computed from the expression for the spectral density matrix in eq. 15, sampling at $10^3$ frequencies (as above, results were not affected by increasing the number of frequencies).

3.3. Simulations for the case of uncorrelated noise: Results

As predicted theoretically in Section 3.1, we found that PSI was minimally affected by uncorrelated noise, although, as shown in Section 2, it yielded many false positives without any additive noise. The standard measure of Granger-causal directionality $\text{sgn}(g^{\text{rev}})$ was only weakly affected for small $\sigma_A$, and only moderately for large $\sigma_A$ and $\gamma$ (Fig 2A-B). For the shrinkage prior, we also observed weak effects on all measures. Thus, while the individual Granger-causal feedback measures $f_2^{\text{rev}}$ and $f_2^{\text{forward}}$ are typically strongly affected by noise, the standard measure of Granger-causal directionality $\text{sgn}(g^{\text{rev}})$ is quite robust to uncorrelated noise. Overall, RGT yielded the smallest fraction of (<0.1) false positives, at the expense of a reduction in true positives. Thus, our results suggest that both RGT and standard Granger-causality are viable options for the regime of uncorrelated noise; which measure to choose depends on assumptions on the signal-to-noise ratio and the relative weighting of true and false positives.

We verified that our results held true across different model orders $M$ and for a unidirectional prior (See the Supplementary Information).

4. Case of linearly mixed noise

4.1. Theoretical considerations for the case of linearly mixed noise

4.1.1. PSI

Linearly mixed noise is common in the neurosciences (the ‘volume conduction’ problem), where electric currents from single neural or noise sources spread instantaneously to multiple electro-magnetic measurement sensors (Nolte et al., 2004; Stam et al., 2007; Vinck et al., 2011). For the signal component that is due to linearly mixed noise, the equality $s_\gamma^{\text{rev}}(\tau) = s_\gamma^{\text{forward}}(-\tau)$ holds for all lags $\tau$, i.e. the cross-covariance function of the noise sources is strictly symmetric. This implies that if there are only noise sources ($\gamma = 1$ in eq. 14), then the imaginary component of the spectral density matrix vanishes, i.e., the equation $\Im(S^{\text{rev}}(\omega)) = 0$ holds for all $\omega$ (Nolte et al., 2004), implying that the PSI for the noise sources equals zero, i.e. $\psi^{\text{rev}} = 0$. It was therefore argued that PSI should have reduced noise sensitivity in comparison to the standard measure of Granger-causal directionality $\text{sgn}(g^{\text{rev}})$ (Nolte et al., 2008). While this holds true for the extreme case of $\gamma = 1$, it is not clear whether it also holds true for values of $\gamma$ close to one. PSI can, by construction, be affected by adding linearly mixed noise. First note that the real component of the coherency, $\Re(C(\omega))$, does play a role in the PSI metric, as the imaginary component of the slope of the coherency $\Im[C(\omega)]$ cannot be rewritten to

$$\lim_{d\omega \rightarrow 0} \frac{\Im(C(\omega))\Im(C(\omega + d\omega))}{d\omega}. \tag{28}$$

Second, note that adding a symmetric cross-covariance function $s_\gamma^{\text{rev}}(\tau)$ to an asymmetric function $s(\tau)$, the effect of adding linearly mixed noise (eq. 16), can flip asymmetries in the opposite direction, because the cross-covariance function $s(\tau)$ can attain both negative and positive values. That is, the inequality $|s_{12}^{\text{rev}}(\tau)| > |s_{12}^{\text{forward}}(\tau)|$ does not imply the inequality $|s_{12}^{\text{rev}}(\tau)| < |s_{12}^{\text{forward}}(-\tau)|$ (Fig 3). For example, suppose that $s_{12}(\tau) = u(\tau)e^{-\tau} \cos(\omega t)$ with $u(\tau)$ the Heaviside step function. In this case, all the energy in the cross-covariance function is concentrated at positive delays. If now $s_{12}(\tau) = -u(\tau)e^{-\tau} \cos(\omega t)$, then $s_{12}^{\text{rev}}(\tau) = 0$ for $\tau \geq 0$ and $s_{12}^{\text{rev}}(\tau) = -e^{-\tau} \cos(\omega t)$ for $\tau < 0$. Suddenly, the energy in the cross-covariance function is concentrated at negative delays.

These considerations reveal that the effect of additive noise is fundamentally different for the case of linearly mixed noise as it affects the cross-covariance function $f_{12}^{\text{rev}}(\tau)$ in a lag-dependent manner, as opposed to uncorrelated noise, and can flip asymmetries around $\tau = 0$ (see Fig 3), which would then appear as a switch of causal flow.

4.1.2. RGT

Good performance of RGT is predicted in the limit of dominant linearly mixed noise, as we will now prove from the spectral definition of Granger-causality (Dhamala et al., 2008; Geweke, 1982). If linearly mixed noise is dominant (i.e. $\gamma$ high), then the normalized magnitude of the imaginary component of the cross-spectral density vanishes, i.e.

$$\frac{\Im[S_{12}]}{\Re[S_{12}]} \rightarrow 0. \tag{29}$$

Define $J \equiv H^{\text{rev}}$. Now take the spectral matrix factorization

$$S^{\text{rev}}(\omega) = J(\omega)\Sigma^{\text{rev}} J^*(\omega). \tag{30}$$

As $\gamma \rightarrow 1$, $H^{\text{rev}}(\omega) \rightarrow H(\omega)$. Hence, $F_{j\rightarrow i}^{\text{rev}}(\omega) \rightarrow F_{j\rightarrow i}^{\text{forward}}(\omega)$. Since $F_{j\rightarrow i}$ and $f_{j\rightarrow i}$ are Fourier pairs (Geweke, 1982), it follows that $f_{j\rightarrow i}^{\text{rev}}(\omega) \rightarrow f_{j\rightarrow i}^{\text{forward}}(\omega)$. Thus, it follows that $[f_{j\rightarrow i}^{\text{rev}} - f_{j\rightarrow i}^{\text{forward}}] \rightarrow 0$, and hence $\text{sgn}(g^{\text{rev}}) \rightarrow \text{sgn}(g)$, as the relative magnitude of linearly mixed noise increases, i.e. $\gamma \rightarrow 1$. This theoretical result is confirmed by our simulations presented in Fig 4. On the other hand, although the PSI converges to zero, i.e. $\psi^{\text{rev}} \rightarrow 0$, as the relative magnitude of linearly mixed noise increases, i.e. $\gamma \rightarrow 1$. However, $\text{sgn}(g^{\text{rev}})$ does not necessarily converge to $\text{sgn}(g)$ for $\gamma$ close to 1, which is confirmed by our simulations presented in Fig 4.

These theoretical considerations also reveal cases where RGT yields false negatives in the absence of noise: If the imaginary component of $\Sigma f$ vanishes (i.e. zero-phase or 180 degree phase interaction), then RGT will yield a false negative; thus, similar to connectivity metrics based on the imaginary component of the cross-spectral density (Nolte et al., 2004; Stam et al., 2007; Vinck et al., 2011), RGT is esp. sensitive in detecting interactions occurring at a phase delay away from 0 or 180 degrees.

4.2. Simulations for the case of linearly mixed noise: Methods

Previous simulations performed by Nolte et al. (2008) were performed for unidirectional VAR models with finite-length re-
alizations. The question remains whether the observed lack of false positives for PSI (maximum $\approx 5\%$) (Nolte et al., 2008) was due to limited sampling and the use of a unidirectional prior. We set the linear mixing matrix $K \sim N(0, I)$ and the number of noise sources $S = 2$ (where $K$ and $S$ occur in eq. 13), like Nolte et al. (2008). We computed RGT, Granger and PSI values as outlined in Section 3.2.

We also performed simulations for the finite sampling regime. We randomly generated finite length realizations according to eq. 1 with Gaussian innovation. For PSI, the spectral density matrix $S$ was estimated parametrically by first estimating the VAR coefficients (using the BSMART toolbox, Cui et al. (2008)) and then using eq. 23. PSI was then computed using sampling at $10^3$ frequencies (as above, results were not affected by increasing the number of frequencies). The model order of the VAR parameters was determined empirically using the Bayesian information criterion (BIC) (Seth, 2010). The BIC criterion was computed in iterative fashion: first finding the optimal model order between 1 and 16; if the model order was equal to 16, then the optimal order between 16 and 32 was searched, etc. To compute PSI nonparametrically, we estimated $S$ using Bartlett-averaging with a Hann taper and 400 datapoint windows, as in Nolte et al. (2008).

4.3. Simulations with analytical approximations: Results

We first show the results from simulations where RGT, Granger and PSI were computed analytically from the given VAR models of signal and noise sources. Our simulations confirm Nolte et al. (2008)'s observation that the standard Granger-causal directionality measure $\text{sng}(g^{(s)})$ fails to identify the Granger-causally dominant variable when linearly mixed noise dominates (high $\gamma$), yielding poor control of the FDR.

A surprising result of this paper is that the PSI also exhibited a large number of false positives for high $\gamma$, with poor control of the false discovery rate. In contrast, RGT yielded a much smaller fraction of false positives (maximum $\approx 0.1-0.15$ across $\gamma$'s (Fig 4A-B). The RGT shows non-monotonous behavior, because at the extremes of $\gamma = 0$ and $\gamma = 1$, there are no false positives, however for intermediate values of linear mixing, there will be acceptances that are false.

Evaluation of the performance function $U = -10 \frac{\text{frac}_U}{\text{frac}_T} + \frac{\text{frac}_U}{\text{frac}_T}$ indicates that RGT has much better asymptotic performance $U$ than both PSI and standard Granger measure $\text{sng}(g^{(s)})$ (Fig 5). RGT is the only decision procedure yielding a performance $U$ around 0, while Granger measure $\text{sng}(g^{(s)})$ and PSI had a performance $U$ far below zero. These findings contradict the previous conclusion that PSI outperforms RGT and standard Granger-causality (Haufe et al., 2012b; Nolte et al., 2008) and reveal that RGT is the most conservative control for linearly mixed noise.

We verified that our results held true across different model orders $M$ and for a unidirectional prior (See the Supplementary Information).

4.4. Simulations with finite samples: Results

For the finite sampling regime (Fig 6A-D), we find that with increasing number of observations, true positives and false positives fractions converged to the asymptotic values shown in Fig 4. PSI converged at a much slower rate when computed with nonparametric (Fourier) than with parametric spectral estimation of the spectral density matrix (Fig 6C-D). Thus, the discrepancy with earlier conclusions (Nolte et al., 2008) can be explained by considering only finite samples ($M = 5; n = 6 \cdot 10^3$), a unidirectional prior, and nonparametric spectral estimation of PSI - all of these factors suppress the fraction of false positives (Haufe et al., 2012a; Nolte et al., 2008).

5. Criterion on instantaneous influence

5.1. Definition of the instantaneous influence strength measure

Our results indicate that RGT strongly improves overall performance relative to standard Granger-causality techniques and PSI. Nevertheless, it can still yield quite a significant fraction of false positives for intermediate levels of linearly mixed noise, such that the performance measure $U = -10 \frac{\text{frac}_U}{\text{frac}_T} \frac{\text{frac}_U}{\text{frac}_T}$ still reaches values of about -1 for noise amplitudes $\gamma$ around 0.5 to 0.7 (see Fig 9). These are structural failures that do not vanish with increasing durations of datasets. We therefore asked whether further gains in overall performance are achievable.

PSI is based on the notion that linearly mixed noise adds a symmetric component to the cross-covariance function $s_{12}(\tau)$ (like symmetric synchronization metrics as in Nolte et al. (2004); Stam et al. (2007); Vinck et al. (2011)), and is designed to suppress this component for estimation of directed connectivity, but fails to do so (Fig 4). However, there may be other features of the VAR model that are diagnostic of linear mixing and that can be used to further improve overall performance, as we will show in what follows. One consequence of adding linearly mixed noise is that it gives rise to strong instantaneous causality. This does not occur in the absence of linearly mixed noise, assuming delayed interactions. Thus, if linearly mixed noise dominates, it is to be expected that instantaneous influence is relatively large in comparison to the feedback at other lags in magnitude.

To formalize this notion, our strategy will be to rewrite the standard (reduced-form) VAR model to a structural VAR (SVAR) model that has a zero-lag coefficient for instantaneous influence. The magnitude of this coefficient can then be compared with the magnitude of VAR coefficients at other delays. This yields a diagnostic criterion for the relative magnitude of linearly mixed noise that can be used to further reduce the false positive rate and false discovery rate. The strength of instantaneous influence can be compared with the strength of feedback at other lags by rewriting the reduced-form VAR model (eq. 1) to a structural VAR model

$$B(0) \mathbf{x}(t) = \sum_{\tau=1}^{M} B(\tau) \mathbf{x}(t-\tau) + \mathbf{w}, \quad (31)$$

with diagonal matrix $\text{Cov} (\mathbf{w}, \mathbf{w}) = \Lambda$. The difference between the reduced form VAR (eq. 1) and the structural VAR is that the structural VAR contains a VAR coefficient $B(0)$ that accounts for the instantaneous influence from $\mathbf{x}(t)$ to $\mathbf{x}(t)$, such that the
covariance matrix of the innovation $\Lambda = \text{Cov}(\mathbf{W}, \mathbf{W})$ can remain diagonal. If the covariance matrix $\Sigma$ of the innovation in the reduced form VAR (eq. 1) is diagonal, then the SVAR equates the reduced form VAR. Using the SVAR representation of the data allows the noise correlations in $\Sigma$ in the reduced form VAR to be put on the same scale as the VAR coefficients $A(\tau)$.

Because the SVAR has one more degree of freedom than the reduced-form VAR, $B(0)$ cannot be uniquely determined from the reduced form VAR coefficients in eq. 1 without additional constraints. Thus, we need to estimate the instantaneous influence matrix $B(0)$ separately for the prediction of $x_1$ and $x_2$ (recursive causal ordering) (Geweke, 1982; Sims, 1980). For predicting $x_2$,

$$B(0) = B(12)(0) = \begin{pmatrix} 1 & 0 \\ b_{12} & 1 \end{pmatrix}$$

(32)

yielding $b_{12} = -\Sigma_{21}/\Sigma_{11}$. For predicting $x_1$,

$$B(0) = B(21)(0) = \begin{pmatrix} 1 & b_{21} \\ 0 & 1 \end{pmatrix}$$

(33)

yielding $b_{21} = -\Sigma_{21}/\Sigma_{22}$. Note that $B(0)$ does not have a unique solution without the triangular constraint. We then obtain a diagonal covariance matrix as the Cholesky decomposition $\Lambda = B(0) \Sigma B(0)^T$, and $B(\tau) = B(0) A(\tau)$, where $A(\tau)$ is the VAR coefficient matrix for the reduced-form VAR (eq. 1). We define a measure of instantaneous influence strength (IIS) as

$$I \equiv \frac{|b_{12}| + |b_{21}|}{\max_{\tau \geq 0} \left| B_{21}(\tau) \right| + \max_{\tau \geq 0} \left| B_{12}(\tau) \right|}$$

(34)

The IIS measure compares the magnitude of the instantaneous VAR coefficient to the maximum magnitude of the VAR coefficients at other lags. Because the magnitude of VAR coefficients is affected by a scaling of the signals ($\Sigma$), signals should first be Z-transformed in practice. If $I > 1/2$, we say that the instantaneous influence exceeds the maximum magnitude of feedback at other lags; the value of 1/2 follows from the fact that in eq. 34 we derive the SVAR(0) value for each signal by assuming that it absorbs all the instantaneous causal influence, while we can only assume that each channel absorbs 1/2 of that influence.

5.2. Use of the instantaneous influence strength measure as a rejection criterion or diagnostic tool

Fig 7 shows the probability of obtaining a certain IIS value as a function of $\gamma$. Low IIS values correspond to high probability to low values of $\gamma$, whereas high values of IIS correspond with high probability to high values of $\gamma$. This reveals that the IIS is strongly indicative of the signal-to-noise ratio, i.e. the amount of linearly mixed noise added.

This finding suggests that we can use the IIS as a rejection criterion in order to improve overall performance, only taking a decision about $\text{sgn}(g)$ if $I < \beta$, where $\beta \geq 0$ is some threshold. That is, for RGT, a true positive would correspond to $\text{sgn}(g)^{\text{rev}} = \text{sgn}(g)$, $\text{sgn}(g)^{\text{rev}} = \text{sgn}(g)$ and $I < \beta$, while a false positive would correspond to $\text{sgn}(g)^{\text{rev}} \neq \text{sgn}(g)$, reveals that this rejection procedure leads to an increase in detection performance (Fig 9).

Thus, the effect of linearly mixed noise can be strongly mitigated by not taking a decision about Granger-causal directionality for datasets in which the instantaneous influence coefficient in a structural VAR model is too large. Different choices of $\beta$ than $\beta = 1/2$ may be possible, however. To inquire this, we evaluated the performance gain $U$ for various levels of the threshold $\beta$. Evaluation of the gain function $U$ reveals that across the various priors there exists a different value of $\beta$ that is optimal (Fig 10). For both standard Granger and RGT, average performance is close to maximum around $\beta = 1/2$, although lower thresholds for standard Granger would yield comparable gain.

Another use of the instantaneous influence measure IIS is as a diagnostic feature for the relative magnitude of linearly mixed noise superimposed on our data, or as a confidence weight to emphasize the contribution of particular trials, channel-combinations, sessions or subjects. For example, a value of $I = 0.1$ would indicate a high signal-to-noise ratio and a very good control of the false discovery rate even for the standard measure of Granger-causal directionality $\text{sgn}(g)^{\text{rev}}$ (Fig 11). A value of $I = 10$ would indicate that linearly mixed noise is extremely dominant and that the FDR is unacceptable for all Granger-causality measures, as the signal-to-noise ratio is likely small.

For values of $\beta \approx 1/2$, the FDR for RGT is smaller than 0.15, at least for the priors examined. For standard Granger, the FDR lies around 0.2-0.25, which may be higher than acceptable. To constrain the FDR for standard Granger below 0.1, one would have to require a threshold of $\beta \approx 0.1$. Thus, for RGT, a threshold of $\beta \approx 1/2$ works reasonably well both in terms of FDR and average performance, while for standard Granger, one should choose a lower threshold than $\beta = 1/2$ in order to constrain the FDR. Thus, the following decision strategy could be taken: if $I$ is small enough ($I < 0.1$), indicating low noise levels, one may use standard Granger because it avoids the relatively large fraction of false negatives that RGT brings in case of low noise, although average detection performance is comparable to RGT; if $I$ indicates medium noise levels (say $0.1 < I < 0.5$), one can use RGT as it still yields an acceptable FDR for all $I$ in this interval; if $I$ exceeds 0.5, indicating high noise levels, it is to be advised to reject any decision about causal directionality.
6. Application to experimental data

In order to demonstrate that the discussed techniques also work for experimental data we applied them to simultaneous LFP recordings from area V1 and area V4 that were made from ECoG grids in an awake monkey performing a spatial visual attention task (Bosman et al., 2012; Rubehn et al., 2009). LFP data was bipolarly referenced, as in Bosman et al. (2012). Here we analyze the epoch of visual stimulation, in which two gratifying stimuli were presented simultaneously, and the monkey was cued to attend to one of the two gratifying stimuli (Bosman et al., 2012). For one of the two monkeys (monkey P), we analyzed bivariate causality between area V1 sites having receptive fields that overlap with the receptive fields of the area V4 sites. We used the same sites as selected in Bosman et al. (2012). Bosman et al. (2012) reported that in the gamma-frequency band (40-90 Hz), the dominant Granger-causal flow between V1 and V4 is ‘bottom-up’, i.e., runs from area V1 to area V4, likely reflecting the feedforward projection from superficial V1 layers, in which gamma power is known to be particularly strong (Buffalo et al., 2011), to the superficial/granular V4 layers.

Our analysis strongly suggests that this finding is unlikely to be explained by the influence of uncorrelated or linearly mixed noise. We find that Granger feedback values in the gamma range reliably reverse when time-reversing the signals (Fig 12). For the original signals, the dominant Granger-causal drive runs from V1 to V4, while for time-reversed signals, the dominant Granger-causal drive runs from V4 to V1. We then applied our SVAR measure of instantaneous influence strength, which yielded low values (mean±SEM = 0.1926±0.0083), indicating that the relative contribution of linearly mixed noise to our data was quite small. For all priors analyzed, these values of I in combination with RGT yield FDR ratios smaller than 0.1. In sum, we conclude that the finding that gamma coherence between V1 and V4 primarily serves as a substrate for feedforward coherence (Bastos et al., 2014; Bosman et al., 2012) is unlikely to be an effect of linearly mixed noise or uncorrelated noise and likely reflects a true bottom up drive.

7. Discussion

7.1. Overview of results

In this study, we examined the performance of various directed connectivity measures, and developed a measure of the relative magnitude of linearly mixed noise that can be used to identify VAR models with either a high or low expected false discovery rate. We considered the following cases: detecting the dominant Granger-causal flow for bidirectional VAR models in the presence of (1) no noise, (2) additive uncorrelated noise, or (3) additive linearly mixed noise. The latter case models the effect of volume conduction of single sources to multiple sensors in EEG and MEG data. It is known that standard Granger-causality fails in the presence of substantial amounts of linearly mixed noise (Haufe et al., 2012a; Nolte et al., 2008). We examined the phase slope index (PSI), which quantifies the slope of the coherency, and reversed Granger testing (RGT), which demands that Granger-causal directionality flips for time-reversed signals. We evaluated these techniques using analytical calculations for a wide range of priors on VAR models, including normal priors and shrinkage priors, and for a wide range of parameters.

In case of no noise, we find that PSI produces many errors, yielding many false positives. RGT also produces a large number of false negatives, but does not produce false positives in case of no noise. Thus, Granger-causal inferences obtained from time delays (as measured by e.g. PSI) should be interpreted cautiously. In fact, the dominant Granger-causal flow can be reliably inferred from time delays only by making the assumption that the interaction between sources is unidirectional. If this assumption can be made, then sgn(g) is probably known already. It should be emphasized that inference of temporal precedence is a meaningful enterprise by itself; PSI is a valid measure of temporal precedence that has the important advantage of not being affected by uncorrelated noise.

We find that uncorrelated noise has only weak to moderate effects on the standard metric of G-causal directionality $g = f_{ij} - f_{ji}$. If the SNR can be assumed to be larger than 1, then $\text{sgn}(g)$ will typically yield a reliable measure of Granger-causal directionality with well-controlled FDR. We have also shown that uncorrelated noise has negligible effects in the limiting case that signals are close to white and coherence values are low, even though both $f_{ij}$ and $f_{ji}$ can be strongly decreased.

Compared to uncorrelated noise, linearly mixed noise exerts much stronger effects on metrics of G-causal dominance. Our simulations show that linearly mixed noise strongly affects the PSI, contrary to previous results, which were obtained using nonparametric spectral estimation and relatively short data traces (Nolte et al., 2008). This discrepancy is principally explained by the finding that the false positives fraction is strongly underestimated when using short data traces. We find that using data traces of similar length as in Nolte et al. (2008) but with parametric estimation of the PSI already yields a much higher false positives fraction. Furthermore, the false positives fraction is higher for bidirectionally coupled sources than for unidirectionally coupled sources. RGT has a compromised true positives fraction relative to standard Granger measures (Granger, 1969), but yields a much smaller fraction of false positives, resulting in a higher overall performance. In fact, our theoretical analysis shows that the fraction of false positives always (i.e., independent of the specific VAR model) converges to zero as the relative contribution of linearly mixed noise increases.

We then developed a means to further suppress the false positive fraction and increase overall performance for all the measures studied. This was achieved by examining the relative magnitude of instantaneous transfer, by comparing the zero-lag coefficient with the nonzero-lag coefficients in a structural VAR model. Exclusion of cases where the zero-lag coefficient has a larger magnitude than the other coefficients yields a strong increase in overall performance gain.

Large values of the instantaneous influence strength measure $I$ can also be indicative of the presence of a third neuronal driving source: While the axonal outputs from one brain area cannot be transmitted to another brain area without a synaptic delay, a third area can simultaneously drive the activity of two
distant brain areas, which can cause instantaneous influence. Thus, high values of \( I \) may either indicate the abundance of linearly mixed noise, or the presence of third driving sources, which can both lead to false conclusions about the Granger-causal flow between brain areas. Note that the effect of a third driving source is fundamentally different than the effect of additive noise, because the effect of a third driving source can be explicitly modelled using a multivariate VAR, whereas additive noise cannot, since noise at time \( t \) merely affects the observations at time \( t \), but not at future time-points. Our practical recommendation is to combine the use of RGT and the measure \( I \) (with a threshold of 1/2), as it leads to a substantial increase in performance gain, with a strong reduction in the fraction of false positives. Furthermore, in practice, if independent replicates (trials sessions, EEG channels, subjects) are available within same experiment, then it may add insight to explore the relationship between the measure \( I \) and the asymmetry of Granger values. One can also use the measure \( 1/I \) effectively as a continuous confidence weight, emphasizing the contribution of trails/sessions/subjects/channel-combinations with a relatively small amount of instantaneous influence (possibly in combination with a hard inclusion criterion).

Note that computation of IIS require sufficiently fine sampling. If too coarse sampling is used, then strong interactions at short lags can give rise to a high IIS value. Given that synaptic delays in the nervous system are on the order of a few to tens of ms, it is recommended to adjust the sampling frequency based on prior knowledge about the delays of interaction between the two sources under consideration.

We then applied these techniques to actual neuronal data (Bosman et al., 2012), where we studied Granger-causal flow between area V1 and area V4. As in Bosman et al. (2012), we find that the dominant Granger-causal flow in the gamma-frequency band (40-90 Hz) is bottom-up, i.e. runs from V1 to V4, consistent with area V1 sitting lower in the cortical hierarchy than area V4 (Markov et al., 2014). Application of RGT and the instantaneous influence strength measure indicates that this finding is unlikely to be explained by the contribution of either uncorrelated or correlated noise. Our analysis shows that conclusions drawn based on coherence and Granger analysis of this dataset (Bastos et al., 2014; Bosman et al., 2012; Brunet et al., 2014) are unlikely to be affected by linearly mixed noise.

7.2. **Comparison to previous work**

Compared to previous work (Haufe et al., 2012a; Nolte et al., 2008), we made several advances that we believe lead to a substantial reinterpretation and reevaluation of existing directed connectivity measures (RGT, PSI and standard Granger-causality).

Firstly, we evaluated all directed connectivity measures using analytical computations, without generating finite data traces and not making any assumptions about the distribution of the innovation noise beyond its covariance matrix. We supplemented and validated this analysis using computations based on finite data traces. Analytical computations have - despite their complexity - the advantage of not underestimating the false and true positives ratios, and of showing the measures’ behavior when all statistical quantities can be optimally estimated from the data. The lack of false positives observed for PSI by Nolte et al. (2008) can to a large degree be explained by Nolte et al. (2008) using only finite and relatively short data traces. This leads to the paradoxical conclusion that PSI performs better for short than for very long data traces. Indeed, in terms of the performance function \( U = -10 \cdot \text{frac}_p + \text{frac}_p \), PSI’s performance deteriorates as we increase the sample size (Fig 6). However, it can also be seen from Fig 6 that parametric spectral estimation already leads to much higher FP ratios, and that the FP ratio for a given sample size is also strongly dependent on the prior used to generate VAR models. This suggests that repeatedly computing PSI over small data segments (bootstrapping) does not provide a viable solution to the noise problem.

Our finding that RGT strongly outperforms PSI in the analytical regime seems somewhat surprising given earlier conclusions that PSI performs slightly better than RGT (Haufe et al., 2012a,b). Importantly, a solid analytical proof was lacking showing the behavior of PSI and RGT for low signal-to-noise ratios. Here, we gave an analytical proof that the false positive ratio for RGT indeed approaches zero as the signal-to-noise ratio approaches zero, while we gave counterexamples showing that this is not the case for PSI, except in the limiting case that there is only correlated noise and no signal.

Secondly, compared to previous work, we did not only evaluate the case of correlated noise, but also the case of uncorrelated noise. We find that uncorrelated noise only weakly affects standard Granger-causality metrics when it comes to detecting the dominant Granger-causal flow, even though individual Granger-causal flow values (from source 1 to 2, and from 2 to 1) can be strongly affected. Thus, the conclusion that Granger-causality metrics are strongly affected by correlated noise (Nolte et al., 2008) should not be generalized to all forms of noise, and should not be generalized to all data types. In-tracranial measurements such as LFPs, especially when bipolearly derived, current source densities and spike trains should be much less affected by correlated noise than scalp EEG and MEG.

Thirdly, compared to previous work, we evaluated a much broader range of priors for generating ensembles of VAR models. Nolte et al. (2008) only evaluated unidirectional priors (i.e., one signal is the driver and one signal is the recipient), which leads to a severe underestimation of FP ratios for low signal-to-noise ratios. Here, we used several types of priors (normal and shrinkage) that include both unidirectional and bidirectional VAR models, and evaluated a wide range of parameter settings to generalize our results (difference variances and model orders). We note that the set of priors used in this paper is not exhaustive and that other, e.g. sparse priors on the space of VAR models, have been considered previously (Valdés-Sosa et al., 2005). The use of a wide range of bidirectional priors is critical as the behavior of RGT and PSI is not easily predictable even in the case of no noise. In fact, we show that RGT can attain false negative ratios up to 40% in case of no noise. The consequence is that RGT may systematically fail to detect certain neuronal interactions, and that the use of RGT should be avoided if the signal-to-noise ratio can be assumed to be high.
Yet, a more positive message follows from our theoretical result that RGT performs especially well if the autocorrelation functions of two time series are relatively similar, in other words if the time series have similar power spectra after z-scoring (i.e., standardizing the variance) the signals.

Fourthly, to make further progress on the problem of linearly mixed noise, we took a fundamentally different approach than Nolte et al. (2008). Following the important and influential result that undirected connectivity measures (like coherence or phase locking value) can be protected against linearly mixed noise by considering only the imaginary component of the cross-spectral density (Nolte et al., 2004; Stam et al., 2007; Vinck et al., 2011), the fundamental idea of Nolte et al. (2008) was to devise a directed connectivity measure that is insensitive to symmetric features of the cross-covariance function, as linearly mixed noise adds a strictly symmetric component to the cross-covariance function. RGT was proposed based on the same rationale (Haufe et al., 2012a), and (Haufe et al., 2012a) concluded that its performance is similar to PSI. However we have shown that it is much more effective than PSI in suppressing the false positives ratio, and that it has a better theoretical rationale than PSI. Here, we develop the idea to explicitly model the instantaneous transfer between two time series, by including a zero-lag transfer coefficient in the VAR model (a structural VAR model). We note that, in neuroscience context, instantaneous interactions have been modelled by several previous papers (Faes and Nollo, 2010; Gates et al., 2010). The novelty of our approach lies in an explicit quantification of the magnitude of the zero-lag transfer coefficients relative to the magnitude of transfer coefficients at other lags, and using this as a diagnostic tool of linearly mixed noise. Compared to the VAR model, the ‘regular’ VAR model does not contain a coefficient to quantify instantaneous transfer between signals, but merely absorbs the instantaneous transfer into the covariance matrix of the innovation process. We have shown that rejecting SVAR models with a large-magnitude zero-lag coefficient leads to a large improvement in the overall performance for all directed connectivity measures.

7.3. Combination with other techniques and extension to multivariate case

We would like to emphasize that in the present paper we have attempted to solve the common noise problem by checking for volume conduction / linearly mixed noise, instead of reducing the scale of the problem with a model, spatial filtering, or techniques like independent component analysis. Connectivity measures can be directly computed on the ‘raw’ sensor level, which is a common approach for EEG or LFP signals (e.g. Salazar et al., 2012; van Kerkoerle et al., 2014). However, prior to computing connectivity measures, it is often desirable to first apply some form of spatial filtering or source estimation. For example, bipolar referencing, current source densities (CSDs) or Laplacians can be used to remove the influence of a common reference and mitigate the influence of distal sources, thereby improving signal-to-noise ratio (e.g. Bollimunta et al., 2008; Bosman et al., 2012; Nunez and Srinivasan, 2006; van Kerkoerle et al., 2014)). Noise reduction through spatial filtering becomes especially important when multiple distant noise sources that are out of phase are (non-linearly) mixed (Schoffelen and Gross, 2014; Sirota et al., 2008; Vinck et al., 2011).

A more advanced form of spatial filtering is source reconstruction or inverse modelling, e.g. using minimum-norm estimates or minimum-variance beamforming. Source reconstruction can be used for EEG/MEG signals prior to computing connectivity measures (Babiloni et al., 2005; Ding et al., 2007; Hui et al., 2010; Schoffelen and Gross, 2009, 2014), although spurious correlations can still be abundant after application of source reconstruction (Schoffelen and Gross, 2009). Future simulation studies and theoretical work are needed to assess the performance of applying the diagnostic techniques discussed in the present paper (RGT, IIS) after source reconstruction / inverse modelling in comparison to either source reconstruction with standard Granger-causality or applying RGT and ISS without source reconstruction.

An alternative approach to remove uncorrelated or linearly mixed noise is a model-based approach in which one directly attempts to model the separate contributions of the signal and noise sources (Cheung et al., 2010; Friston et al., 2014; Nalatore et al., 2007). Nalatore et al. (2007) used a state-space Kalman-filter approach in which an expectation-maximization algorithm was developed to mitigate the effect of uncorrelated, Gaussian white noise. This algorithm may be employed to remove uncorrelated noise sources prior to the application of standard Granger causality. Although the assumption of Gaussianity is sensible from the perspective of Central Limit Theorem, a difficulty with this approach is that uncorrelated noise can not always be modelled as a Gaussian white source. In this present paper, we have shown that estimation of G-causal directionality is not strongly affected by uncorrelated noise sources, even if they are not white or Gaussian white.

For the common noise problem, it has been shown that estimating VAR parameters for cortical sources after first applying source reconstruction can lead to biased VAR estimates (Cheung et al., 2010). To address this problem, Cheung et al. (2010); Cheung and Van Veen (2011) used a state-space modelling approach in which both the VAR and the observation equation, which links sources to sensors through the lead field matrices, were considered simultaneously. Future studies should examine whether applying RGT or IIS on the VAR estimates obtained using such state-space estimation yields further improvements in detecting G-causal directionality.

Various connectivity measures have been proposed to estimate coherence or phase synchronization based on the imaginary component of the cross-spectral density, whose expected value is not affected by adding linearly mixed, uncorrelated noise sources (Nolte et al., 2004). These include the imaginary component of the coherence (Nolte et al., 2004) the phase lag index (PLI) (Stam et al., 2007) and the weighted phase lag index (WPLI) (Vinck et al., 2011). While these measures were developed with the idea to measure the magnitude of coherency or strength of phase synchronization, these measures could in principal also be used as diagnostic measures of the amount of volume conduction. However, we believe that the discussed
techniques (RGT, IIS) are to be preferred for the following two reasons: 1) If the imaginary component of the coherency (ICC) or PLI/WPLI is significantly greater than zero, it is still possible that there is a strong linear mixing component that gives rise to spurious G-causal directionality estimates, especially if the availability of a large sample size permits the detection of small coupling values. Hence, some threshold would be required on the strength of e.g. ICC/PLI/WPLI; it is not clear how to choose this threshold. 2) An advantage of IIS is that, in case of no noise, it can identify G-causal links for all VAR systems (Figure 8), while ICC/PLI/WPLI perform poorly when interactions occur around a 0 or 180 degrees phase relationship.

Finally, we would like to point out that in this paper we have considered only the simplified bivariate VAR case. Principally, to interpret G-causality in terms of causality one needs to take a multivariate approach in which third sources are also modelled (Granger, 1969). It is always possible that a G-causal relationship between two signals can be explained by a third source that is not included in the VAR model. In practice, multivariate approaches give rise to estimation problems due to the increased number of model parameters. Our conclusion that the fraction of false positives for RGT approaches zero in the limiting case of $\gamma \to 1$ (only linear mixing) extends to multivariate settings, because also in that case the cross-spectral density matrix is real-valued and not affected by a reversal of the time-series. An extension of the SVAR approach used here to the multivariate case is called for. Because the nuisance from third sources, which may give rise to instantaneous interactions, can be removed using multivariate VAR models, we expect that the IIS approach benefits from a multivariate model. One possible extension of IIS to the multivariate case is to model the instantaneous influences for only one pair of sources at a time, and have the instantaneous influences between the other sources absorbed by the covariance matrix $W$ in eq. 31. The exact performance of IIS and RGT in the multivariate case needs to be examined using future simulations.

7.4. Outlook

Granger-causality has become an increasingly popular technique to study directed interactions between neuronal populations in spike train, LFP, EEG or MEG signals. Yet, Granger-causality metrics can be strongly affected by the addition of linearly mixed noise (e.g. caused by volume conduction). Here, we show that two effective, complementary control analyses can be performed to avoid spurious conclusions: 1) computing Granger-causality metrics on time-reversed signals and ensuring that the directionality of Granger-causal flow reverses; 2) rewriting the reduced form VAR model to a structural VAR model containing an instantaneous influence coefficient, and ensuring that the magnitude of that instantaneous influence coefficient is relatively small. We show that, for a broad range of priors, a combination of these techniques yields a small fraction of false positives. In the limit of having only or no linearly mixed noise, these techniques give a hard performance guarantee that no false positives emerge. However, it needs to be emphasized that false positives can, in practice, still emerge for intermediate levels of linearly mixed noise, and that for those cases no universal performance guarantees can at present be given outside the realm of the studied VAR priors. Future work should generalize this conclusion by studying a broader range of VAR priors. However, it remains an open question whether hard, universal performance bounds on the expected fractions of false and true positives can be obtained for data contaminated by linearly mixed noise.

Acknowledgements

This work was supported by EU FP7- ICT grant 270108 (Goal Leaders; to CMP), a Rubicon NWO grant (to MV) and EU FP-7 grant 284801 (Enlightenment; to FPB). We thank Fernando Lopes da Silva and Steven Bressler for useful comments on early versions of this manuscript. MV designed the study. MV and LH performed simulations and analyzed data. CAB and PF contributed experimental data. All authors contributed to interpretation of results. MV drafted manuscript. All authors edited and revised manuscript.

Appendix A

Theoretical analysis of validity reversed Granger testing

We decompose $f_{j\rightarrow i}$ in terms of auto- and cross-correlations, revealing the limiting case where reversed Granger testing performs well. By dividing $\Omega_i$ and $\Sigma_j$ by the overall signal variance (i.e., using the scale invariance property of G-causality), we can rewrite $f_{j\rightarrow i} \equiv \ln \frac{D_i}{D_j}$ as

$$f_{j\rightarrow i} = -\ln \left( \frac{1 - R^2_{i\rightarrow j}}{1 - R^2_{j\rightarrow i}} \right),$$

(36)

By treating the VAR model as a multiple regression model, it can be seen that the coefficient of determination $R^2$ (explained variance) (Pierce, 1979) equals

$$R^2_{i\rightarrow j} = u_{i\rightarrow j}^{-1} [D_{j\rightarrow i}]^{-1} u_{j\rightarrow i}^T,$$

(37)

where

$$u_{i\rightarrow j} \equiv (\rho_{ij}(1), \ldots, \rho_{ij}(M), \rho_{ji}(1), \ldots, \rho_{ji}(M)),$$

(38)

with cross-correlation function

$$\rho_{ij}(\tau) \equiv \frac{s_{ij}(\tau)}{\sqrt{s_{ii}(0)s_{jj}(0)}}.$$  

(39)

(Subscript $i, j \rightarrow i$ indicates that $x_i$ and $x_j$ are used to predict $x_i$). The symmetric block matrix

$$D_{i\rightarrow j} \equiv \begin{pmatrix} D_{ii} & D_{ij} \\ D_{ji} & D_{jj} \end{pmatrix}$$

(40)

holds the correlations among the predictors, where $D_{ii(km)} \equiv \rho_{ii}(k-m)$, $D_{12(km)} \equiv \rho_{12}(m-k)$. Define

$$[D_{i\rightarrow j}]^{-1} \equiv \begin{pmatrix} E_{ii} & E_{i1} \\ E_{12} & E_{jj} \end{pmatrix}$$

(41)
Note that $E$ is symmetric. For the restricted model $v_{i\rightarrow j} = (\rho_{i1}(1), \cdots, \rho_{iM}(M))$, and $R^2_{i\rightarrow j} = v_{i\rightarrow j}[D_{i\rightarrow j}]^{-1}v_{i\rightarrow j}^T$.

We now show the behavior of reversed Granger testing in two limiting cases.

1) Suppose that the cross- and autocorrelations are small in magnitude, that is $\forall \tau, |\rho_{ij}(\tau)| \leq \mu << 1$ (note that $\rho_{ii}(0) = 1$). Then, $E_{ij} \approx 1 + O(\mu^2)$ for $i \neq j$ and $E_{ii} \approx (D_{i\rightarrow i})_{ii} + O(\mu^2)$ for $i \neq j$. Taylor expansion of $\ln(\tau,\mu)$ around $\mu = 1$ (which is justified because of the low Granger values) yields

$$f_{j\rightarrow i} \approx \sum_{\tau=1}^{M} \rho_{12}^2(\tau) + O(\mu^3) \approx f_{j\rightarrow i}^{rev}.$$  \hspace{2cm} (42)

Thus, for small auto- and cross-correlations, Granger values approximately equal the energy of the left or right-hand side of the cross-correlation function. Indeed, for small $A(\tau)$’s, reversed G-causality testing invariably identifies $sgn(g)$, and tends to fail for a larger % of VAR models when the squares of $A(\tau)$’s are larger (corresponding to larger squared correlations) (Fig 1A-B, main text).

2) If the signals have similar autocorrelation functions, i.e., if $D_{i1} \approx D_{22}$, then it follows directly from eq. 36 that $f_{j\rightarrow i} = f_{j\rightarrow i}^{rev}$.

**Theoretical analysis of influence uncorrelated noise**

Now suppose that uncorrelated noise $n(t)$ is added only to $x_1(t)$. Then,

$$\rho_{11}^{(e)}(\tau) = \frac{s_{11}(\tau) + \rho_{10}(\tau)\sigma_n^2}{\sigma_1^2 + \sigma_0^2}.$$  \hspace{2cm} (43)

Suppose that, in addition, $\rho_{10}(\tau)$ is small (i.e., nearly white noise). Then, $\rho_{11}^{(e)}(\tau) \approx \rho_{11}(\tau)$ for $\tau \neq 0$ with $\theta = (\sigma_1^2 + \sigma_0^2)/\sigma_1^2$.

At the same time,

$$\rho_{12}^{(e)}(\tau) = \rho_{12}(\tau)/(\sigma_1^2 + \sigma_0^2)^{1/2}\sigma_2$$  \hspace{2cm} (44)

depends relative to $\rho_{12}$ by only $1/\sqrt{\theta}$, independent of the noise’s color. As for $f_{1\rightarrow 2}$, note that $\rho_{22}^{(e)}(\tau) = \rho_{22}(\tau)$, while $\rho_{22}^{(e)}(\tau)$ decreases at rate $1/\sqrt{\theta}$.

The impact of uncorrelated noise on Granger values can be analyzed in the limiting case that the magnitudes of the cross- and autocorrelations are small, that is $\forall \tau, |\rho_{ij}(\tau)| \leq \mu << 1$. It follows from eq. 42 that $sgn(g)$ (and $f_{j\rightarrow i}/f_{j\rightarrow i}$) tends to be unaffected by noise (see main text Fig 2A,B), even though $f_{j\rightarrow i}$ and $f_{j\rightarrow i}$ decrease at rate $1/\theta$.


Figure 1: Phase slope index (PSI) and reversed Granger testing (RGT) often fail to detect Granger-causal directionality in case of no additive noise, i.e. $\gamma = 0$ in eq. 14. (A) We randomly generated stable VAR models according to a normal prior $A_{ij}(\tau) \sim N(0, \sigma^2_A)$ and determined the dominant Granger-causal directionality. We then took a decision about the dominant Granger-causal directionality using either PSI or RGT. Fractions (relative to the total number of simulated VAR models) of true positives (left) and false positives (right) as a function of VAR model order $M$ and $\sigma_A$, separate for PSI and RGT. Various lines represent different levels of $\sigma_A (0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.6)$, the standard deviation of the VAR coefficients, with fraction of true positives decreasing as a function of $\sigma_A$. Thicker lines with higher transparency correspond to lower values of $\sigma_A$. (B) Same as (A), but now for shrinkage prior $A_{ij}(\tau) \sim N(0, \theta^2 \lambda^2 / \tau^2)$, with the coupling parameter $\theta = 1$ for $i = j$ and $0 \leq \theta \leq 1$ otherwise. The shrinkage prior assumes that recent lags carry more information than distant lags, and that a signal is a better predictor of itself than of other signals. The parameter $\lambda$ controls the variance of the VAR coefficients. (A-B) Note that the fraction of false negatives equals 1 minus the fractions of false and true positives. (C-D) Normal prior: mean peak coherence and number of peaks as a function of standard deviation of VAR coefficients $\sigma_A$ and model order $M$. The coherence is an increasing function of $\sigma_A$ and $M$. The number of peaks in the coherence spectrum increases linearly as a function of model order $M$. (E) Shrinkage prior: mean peak coherence as a function of $\lambda$ and $\theta$ (line definitions as in (B)). The coherence is an increasing function of $\lambda$ and $\theta$. 

16
Figure 2: Uncorrelated noise has very weak effects on PSI and weak to moderate effects on other Granger-causality metrics. (A) We randomly generated stable VAR models of order $M = 4$ according to a normal prior $A_{ij}(τ) \sim N(0, σ^2_A)$ and determined the dominant Granger-causal directionality. Two uncorrelated noise sources were superimposed on the signal sources and also generated according to a stable VAR model that was again randomly generated by a normal prior $A_{ij}(τ) \sim N(0, σ^2_A)$. From the noisy signal, we then detected the Granger-causal directionality using either phase slope index (PSI), reversed Granger testing (RGT) or standard Granger-causality. Fractions (relative to the total number of simulated VAR models) of true positives (left) and false positives (right) as a function of noise amplitude $γ$ and $σ_A$. For small $σ_A$, noise has no effects on the calculated metrics. For large $σ_A$, noise has moderate effects. (B) Same as (A), but now for the shrinkage prior $A_{ij}(τ) \sim N(0, θ^2/τ^2)$, with the coupling parameter $θ = 1$ for $i = j$ and $θ = 0.3$ otherwise. The parameter $λ$ controls the variance of the VAR coefficients. (A-B) See the Supplementary Information for generalization to higher model order and unidirectional prior.
Figure 3: Measures that quantify the asymmetry of the cross-covariance function, like the phase slope index (PSI), are not necessarily robust against the addition of linearly mixed noise. Observed cross-covariance $s_{12}^{\epsilon}$ (B) is sum of (asymmetric) signal and (symmetric) linearly mixed noise cross-covariance (A). Here, $s_{1}^{\epsilon}$ causes the asymmetry in $s_{12}$ to flip around $\tau = 0$. 
Figure 4: Linearly mixed (i.e., correlated noise) has strong effects on the false positives fractions of standard Granger-causality and phase slope index (PSI), but only weak to moderate effects on the false positives fraction of reversed Granger testing (RGT). (A) We randomly generated stable VAR models of order $M = 4$ according to a normal prior $A_{ij}(\tau) \sim N(0, \sigma_A^2)$. A linear mixture of two uncorrelated noise sources was superimposed on the signal sources. Both noise sources were generated according to a stable VAR model that was randomly generated by a normal prior $A_{ij}(\tau) \sim N(0, \sigma_A^2)$. Fractions (relative to the total number of simulated VAR models) of true positives (left) and false positives (right) as a function of noise amplitude $\gamma$. (B) Same as (A), but now for the shrinkage prior $A_{ij}(\tau) \sim N(0, \theta^2, \lambda^2/\tau^2)$, with the coupling parameter $\theta = 1$ for $i = j$ and $\theta = 0.3$ otherwise. The parameter $\lambda$ controls the variance of the VAR coefficients. (A-B) See the Supplementary Information for generalization to higher model order and unidirectional prior.
Figure 5: RGT has better overall performance than PSI and Granger. (A-C) Gain $-10F(\gamma) + T(\gamma)$, with $\gamma$ the noise level, and $F$ and $T$ the fractions of false positives and true positives, respectively (Normal prior, $\sigma_A = 0.01$). (A-C) correspond to different variance of the VAR coefficients ($\sigma_A = 0.01, 0.1, 0.3$). (D-F) Same as (A-C), but now for shrinkage prior, with (A-C) corresponding to various levels of the variance of VAR coefficients ($\lambda = 0.01, 0.1, 0.3$).
Figure 6: All metrics converge to analytically computed values with increasing number of observations. (A-B) We randomly generated stable VAR models of order $M = 4$ according to a normal prior $A_i(\tau) \sim N(0, \sigma^2_A)$, in this case $\sigma_A = 0.1$. A linear mixture of two uncorrelated noise sources was superimposed on the signal sources. Both noise sources were generated according to a VAR model that was randomly generated by a normal prior $A_i(\tau) \sim N(0, \sigma^2_A)$. We then generated data traces of finite length with normally distributed innovation noise. Jackknifing was used to estimate significance, using $|\mu/\sigma| = 2$ as cutoff. For PSI, the spectral density matrix was estimated parametrically via the VAR fits. (B) Same as (A), but now using $\sigma_A = 0.3$. (C-D) Same as (A-B), but now also estimating the spectral density matrix nonparametrically using Discrete Fourier Transforms of 400 datapoint segments multiplied with Hann windows, and Bartlett-averaging.
Figure 7: The instantaneous influence strength measure $I$ provides information about the relative magnitude of linearly mixed noise. (A-C) Shown the cumulative density function as a function of the instantaneous influence strength measure $I$, for various levels of noise amplitude $\gamma$. Green-to-black color gradient corresponds to increasing levels of $\gamma = 0, 0.1, \ldots, 0.9$. The variable $I$ measures the magnitude of the instantaneous influence compared to the feedback at other lags. (A) to (C): variance of normal prior on VAR coefficients $\sigma^2 = 0.01, 0.1, 0.3$. (D-F) As (A-C), but now shown the probability density function. (A-F) show that low values of $I$ tend to correspond to a high signal-to-noise ratio, and that high values of $I$ correspond to a small signal-to-noise ratio.
Figure 8: Using an inclusion criterion on the instantaneous influence strength $I$ (include only $I < \beta$ and reject $I > \beta$) can strongly reduce the fraction of false positives. (A-B) As in Fig 4, we show the fraction of true and false positives as a function of the noise amplitude $\gamma$. The difference with 4 is that in this figure, we take a decision about the Granger-causal directionality only if the instantaneous influence strength $I < 1/2$ (eq. 34). Thus, for $I > 1/2$, we always obtain false negatives. Note that the sum of true and false positives and false negatives equals 1. (C) The fractions of false positives (filled bars) and true positives (open bars) across all levels of noise amplitude $\gamma$, separate for all VAR models, VAR models for which $I > 1/2$, and VAR models for which $I < 1/2$ (Normal prior, $\sigma_A = 0.1$). The mixture of true positives and false positives is better for $I < 1/2$ than for $I > 1/2$. (D-F) Same as (A-C), but now for shrinkage prior.
Figure 9: Using an inclusion criterion on the instantaneous influence strength $I < 1/2$ can strongly increase overall performance in terms of false and true positives, for RGT, PSI and Granger. (A-C) Gain $-10F(\gamma) + T(\gamma)$, with $\gamma$ the noise level, and $F$ and $T$ the fractions of false positives and true positives, respectively (Normal prior, $\sigma_A = 0.01$). Dashed: including only VAR models for which $I < 1/2$. Solid: all VAR models. (A-C) correspond to different variance of the VAR coefficients ($\sigma_A = 0.01, 0.1, 0.3$). (D-F) Same as (A-C), but now for shrinkage prior, with (A-C) corresponding to various levels of the variance of VAR coefficients ($\lambda = 0.01, 0.1, 0.3$).
Figure 10: There exists an optimal rejection criterion on the instantaneous influence strength. (A) False discovery rate (total number of false positives divided by total number true positives) as a function of log_{10}(\beta), with all VAR models for which I < \beta included and VAR models for which I > \beta rejected (yielding false positives). (B-C) Performance gain \(-10 \sum_{\gamma=1}^{\log_{10}(I)} F(\gamma) + \sum_{\gamma=1}^{I} I(\gamma)\) for Granger (B) and RGT (C), as a function of \beta (only I < \beta included). Yellow (B) and blue (C) lines indicate average across tested priors. (D-F) Same as (A-C), but now for shrinkage prior.
Figure 11: The false discovery rate is an increasing function of the instantaneous influence strength $I$. (A) Mean false discovery rate total number of false positives divided by total number true positives) for different levels of $I$, for the various normal priors. (B) Same, but now for shrinkage prior.
Figure 12: Application to neuronal data. (A) Spectral Granger feedback values between V1 LFP sites and V4 LFP sites as a function of frequency. Shadings represent standard error of the mean across channel-combinations. (B) Original Granger asymmetry values (defined as the difference between V1 $\rightarrow$ V4 flow minus V4 $\rightarrow$ V1 flow) vs. Granger asymmetry values for time-reversed signals (defined as the difference between V4 $\rightarrow$ V1 flow minus V1 $\rightarrow$ V4 flow).