bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Non-parametric test for connectivity detection in multivariate autoregressive networks and application to multiunit activity data

M Gilson*, A Tauste Campo*, A Thiele, G Deco

January 18, 2017

abstract

Directed connectivity inference has become a cornerstone in neuroscience following the recent progress in imaging techniques to describe anatomical and functional networks. This paper focuses on the detection of existing connections from the observed activity in networks of 50 to 150 nodes with linear feedback in discrete time. Through the variation of multiple network parameters, our numerical results indicate that multivariate autoregressive (MVAR) estimation attains better accuracy than other standard techniques like partial correlations and Granger's causality. Based on these findings, we propose a surrogate-based significance test for connectivity detection that is shown to achieve a good control of false positive rate and to be robust to various network topology. The surrogates are generated using time rolling of the observed time series, which breaks down covariances while preserving variances: this builds a null-hypothesis distribution for each connection, from which the connectivity estimate can be compared. Our results strongly support surrogate-based MVAR estimation as a better alternative to Granger's causality test, as it properly incorporates the network feedback in the estimation model. We apply our method to multiunit activity data recorded from Utah electrode arrays in monkey to examine the detected interactions between channels for a proof of concept.

1 Introduction

In recent years, there has been a growing interest in developing multivariate techniques to infer causal relations among time series. The initial formulation of the problem goes back to the seminal work by Granger in 1960's [11] motivated by the analysis of the pairwise influence between economic time series. A multivariate solution of the problem was provided a decade later by the introduction of Multivariate autoregressive (MVAR) models [2, 17] which allow to estimate causal relationships between nodes in networks based on their observed activity. The MVAR was further combined with spectral analysis to develop the directed transfer entropy function [13, 14], which has been employed to analyze connectivity patterns in neurobiological systems [3, 34]. This inference problem has been regarded in a number of scientific disciplines, gaining central relevance in the field of genetics, in which the inference of gene regulatory networks is a cornerstone. In this field, the growth of inference methods has been specially notorious over the last decade [6, 36] providing techniques that are typically tied to the large-scale and small-sample specificities of gene expression time series [7]. Of particular interest is the development of techniques around the partial correlation measure, which is used to test the conditional independence on multivariate Gaussian data [16]. Although partial correlations "only" estimate undirected dependencies among variables, it can be further combined with directional linear [36] and non-linear methods [23] to infer directed networks.

In neuroscience, this inference problem has been developed over the last decades in conjunction with progress in neuroimaging techniques, in order to study brain connectivity. In this context, we may distinguish two types of estimation procedures: measures relying on an underlying interaction model such as Granger causality and Dynamic Causal Modeling [8] on the one hand; and model-free measures such as

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transfer entropy [27] and directed information [19], which make minimal model assumptions on the other hand. Although model-free approaches have proven useful to describe neural propagation at spike-train level [29, 32], certain assumptions are required when estimating interactions at the neuronal population level, in which broader spatial and temporal scales contribute to shaping the signals. Among the family of dynamic models, the MVAR model stands out for its simplicity and applicability to different problems, in particular in neuroscience for resting-state and task-evoked fMRI data [30, 24, 20] and MEG data [21]. Using the continuous-time version of the MVAR model - namely an Ornstein Uhlenbeck process - we recently proposed an iterative method to fit its parameters, to infer the so-called 'effective' connectivity [10]; this concept simply refers to a dynamic network model whose inputs and connectivity are both estimated, the latter being called effective because it describes the interaction strengths for the considered local node dynamics. Practically, that 'Lyapunov optimization' (LO) method relies on the covariances of the observed activity and aims to reduce the matrix distance between the model and empirical covariances; like MVAR, it uses time shifted covariances to estimate directed interactions.

In this article we perform a comparative study of connectivity estimation techniques applied to a well-known dynamic model, with a view to application to brain network activity. Here we do not aim to estimate the strengths of interactions, but their existence. We use a MVAR process as generative model, since it is canonical to MVAR estimation as well as Granger causality analysis; however, we refer to it as noise-diffusion model to distinguish the generative model and the estimation method. In essence, both Granger causality and MVAR are based on a linear regression and the distinction between them lies in the following: while Grangers causality independently estimates connectivity strengths for each pairwise interaction, MVAR provides a one-shot estimation of pairwise connectivity strengths for the whole network. By comparing these methods, we specifically question the advantage of taking the full network information into account for the estimation. In addition to MVAR and Granger causality, we also infer the network connectivity from the simulated activity using partial correlation analysis and our recently developed iterative LO method [10].

The focus of the present analysis is on networks of about a hundred of nodes for a few thousands of observed time points in their activity, that is, more samples than nodes. We assess the performance of the four considered methods in detecting existing connections at the network level, as well as the computational cost to calculate the estimates. In particular, we test the robustness of the detection for a broad range of network parameters and several topologies. Finally, in order to provide an alternative to Granger's causality analysis in real-data settings, we propose a non-parametric significance test for MVAR estimation that controls the desired false-alarm rate.

2 Methods: Noise-diffusion discrete-time network model and interaction estimation

The activity in a noise-diffusion network (equivalent to a MVAR process of order 1) is described by the following equation:

$$x^{t+1} = Ax^t + \zeta^t \ , \tag{1}$$

where the matrix A describes the interactions between coordinates of the vector $x^t = x_i^t$ with $1 \leq i \leq N$, and ζ^t is Gaussian (possibly colored) noise, whose realizations are independent for successive time steps.

To detect the existence of connections $A_{ij} > 0$, the three methods MVAR, LO and PC rely on the time-shifted covariances of the observed activity variables x^t:

$$\widehat{Q}_{ij}^{\tau} = \frac{1}{T-2} \sum_{1 \leq t \leq T-\tau_{\max}} (x_i^t - \bar{x}_i)(x_j^{t+\tau} - \bar{x}_j) \ , \tag{2}$$

where T denotes the number of successive samples indexed by t, $\tau \in \{0, 1\}$ is the time shift and the mean activity for each node is $\bar{x}_i = \frac{1}{T} \sum_t x_i^t$. Without loss of generality, we assume all variables ζ^t to have zero means, meaning that $\bar{x}_i = 0$ for all nodes i. Moreover, we constrain the study to the single time shift $\tau = 1$, as it gives the most direct link between the connectivity and the covariances, as developed below.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.1 Multivariate autoregressive (MVAR) estimation

Following [17], the Yule-Walker equation gives a consistency equation for the covariance matrices in terms of the connectivity A in the dynamics described by Eq.(1):

$$Q^1 = Q^0 A^\dagger \ , \tag{3}$$

where $\dagger$ denotes the matrix transpose. The matrix Q^0 satisfies

$$Q^0 = A Q^0 A^\dagger + \Sigma \ , \tag{4}$$

where $\Sigma = \langle \zeta \zeta^\dagger \rangle$ is the noise covariance matrix.

The estimation of network interactions relies on evaluating of A from Eq. (3) for the empirical covariance matrices defined as Eq. (2) and calculated for a given time series:

$$A^\dagger = (\widehat{Q}^0)^{-1} \widehat{Q}^1 \ . \tag{5}$$

This least-square estimate for autoregressive models also corresponds to the model with maximum likelihood, given assumptions such as Gaussianity of the observed process. An estimate of the input noise Σ is then given by

$$\Sigma = \widehat{Q}^0 - A \widehat{Q}^0 A^\dagger \ . \tag{6}$$

2.2 Lyapunov optimization to estimate interactions

We also adapt the iterative optimization method in [10] to the present case of noise-diffusion network in discrete time. We optimize all connections by modifying the parameters A and Σ such that the model reproduces the empirical covariances in $\widehat{Q}^0$ and $\widehat{Q}^1$ in Eq. (2). This iterative optimization yields the following updates at each step that are computed from the model error $\Delta Q^0 = (\widehat{Q}^0 - Q^0)$ and $\Delta Q^1 = (\widehat{Q}^1 - Q^1)$:

$$\begin{aligned} \Delta A^\dagger &= \eta_A (\widehat{Q}^0)^{-1} \Big[- \Delta Q^0 A^\dagger + \Delta Q^1 \Big], \\ \Delta \Sigma &= \eta_\Sigma \Big[\Delta Q^0 - A \Delta Q^0 A^\dagger \Big]. \end{aligned} \tag{7}$$

These equations are obtained by differentiating Eqs. (5) and (6) with respect to the parameters A and Σ, respectively; further details for an analogous derivation are provided in [10]. In the present study, the optimization rates are set to $\eta_A = 0.0001$ and $\eta_\Sigma = 0.01$. At each step, conditions on A and Σ can be imposed, e.g., Σ is a diagonal matrix and all its off-diagonal elements are kept equal to 0.

2.3 Partial correlations (PC)

The theory for PC has been developed for variables x_i sampled from a multivariate Gaussian distribution with given covariance matrix Q, which is estimated from the data using Eq. (2) with $\tau = 0$. The partial correlations are calculate from the inverse matrix of the empirical estimate $\widehat{Q}^0$

$$P_{ij} = \frac{(\widehat{Q}^0)^{-1}_{ij}}{\sqrt{(\widehat{Q}^0)^{-1}_{ii} (\widehat{Q}^0)^{-1}_{jj}}}, \tag{8}$$

The matrix element P_{ij} gives an estimate of direct interactions between x_i and x_j.

Provided the underlying Gaussian process is stationary, parametric tests on P_{ij} estimate the probability of the existence of an interaction between the nodes i and j, conditioned on the remaining observed nodes. The null hypothesis is constructed to account for possible indirect connections between i and j through neighboring nodes.

For PC to estimate adequately the network interactions, it requires that $(Q^0)^{-1}$ is linearly related to a solution for A - say with a factor a - in Eq. (4), for given Q^0 and Σ. From Eq. (3), this can only occur if Q^1 is a scaled version of the identity matrix, i.e., with time-shifted cross-covariance terms are small compared to the diagonal ones. This constraint makes it unlikely for PC to describe the interactions in general for a model with arbitrary A and Σ that generates a covariance matrix Q^0.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.4 Granger causality analysis

In contrast to the previous methods relying on covariances, Granger causality analysis is usually defined using time series. In a network, two flavors of Granger causality exist: 'pairwise' or multivariate ('full'). To detect the existence of connection $j \to i$, we consider the comparisons of the residuals from the following linear regressions (LR)

$$\begin{aligned} \mathrm{GR2}(x_j \to x_i) &= \ln\left[\frac{\mathrm{LR}\big(x_i^{2\le t\le T}|x_{i,j}^{1\le t\le T-1}\big)}{\mathrm{LR}\big(x_i^{2\le t\le T}|x_i^{1\le t\le T-1}\big)}\right]; \\ \mathrm{GRf}(x_j \to x_i) &= \ln\left[\frac{\mathrm{LR}\big(x_i^{2\le t\le T}|x_{1,\cdots,N}^{1\le t\le T-1}\big)}{\mathrm{LR}\big(x_i^{2\le t\le T}|x_{1,\cdots,j-1,j+1,\cdots,N}^{1\le t\le T-1}\big)}\right]. \end{aligned} \tag{9}$$

In the LR notation above (similar to conditional probabilities), the left argument (future) is regressed against the right argument (past); the superscript t indicates the considered range for the history used in the regressions; the subscripts indicate the nodes involved. In comparison, the MVAR estimation corresponds to $\mathrm{LR}\big(x_{1,\cdots,N}^{2\le t\le T}|x_{1,\cdots,N}^{1\le t\le T-1}\big)$; note that if we were to compare it to $\mathrm{LR}\big(x_{1,\cdots,N}^{2\le t\le T}|x_{1,\cdots,j-1,j+1,\cdots,N}^{1\le t\le T-1}\big)$, we would consider the influence of j on the rest of the network, without isolating a single connection.

In the case of pairwise Granger, which has a univariate causal target x_i, the recommended parametric test to infer the significance of the interaction relies on the F-statistics for small number of samples: the null hypothesis of no interaction for $\mathrm{GR2}(x_j \to x_i)$ corresponds to [4, with $m = T$, $p = 1$, $n_x = 1$ and $n_y = 2$ using their notation]

$$\frac{\mathrm{LR}\big(x_i^{2\le t\le T}|x_{i,j}^{1\le t\le T-1}\big) - \mathrm{LR}\big(x_i^{2\le t\le T}|x_i^{1\le t\le T-1}\big)}{\mathrm{LR}\big(x_i^{2\le t\le T}|x_i^{1\le t\le T-1}\big)} = [\exp(GR2_{ij}) - 1] > \frac{\phi(\epsilon, 1, T-3)}{T-3} \tag{10}$$

with ϵ the desired sensitivity and ϕ the inverse survival function of the F-distribution [35].

2.5 Receiver operative characteristic (ROC) curve to estimate the performance; shuffling method

From the off-diagonal matrix elements of a given connectivity estimate, we can define a threshold for which larger values are detected to be connections (undirected for PC, directed for others). For each threshold value, misses and false alarms are counted knowing the original A used to generate the observed network activity. By varying the threshold across the entire distribution, we obtain a ROC curve with the rate of false alarms on the x-axis and correct detection on the y-axis. The area under the curve - ranging from 0 to 1 - gives a single "absolute" value for the performance of the connectivity estimation avoiding the issue of setting a single threshold. Random detection corresponds to 0.5.

2.6 Experimental setup and processing of electrode measurements to extract MUAe activity

All procedures were carried out in accordance with the European Communities Council Directive RL 2010/63/EC, the US National Institutes of Health Guidelines for the Care and Use of Animals for Experimental Procedures, and the UK Animals Scientific Procedures Act. Two male macaque monkeys (5 - 14 years of age) were used in the experiment; only the data for the first one is used here. A surgical operation was performed under sterile conditions, in which a custom-made head post (Peek, Tecapeek) was embedded into a dental acrylic head stage. Details of surgical procedures and post-operative care have been published previously [33]. During the surgery microelectrode chronic Utah arrays (5*5 grids), attached to a CerePort base (Blackrock Microsystems) were implanted into V1. Electrodes were 1 mm in length in line with procedures described in [31]. Stimulus presentation was controlled using CORTEX software (Laboratory of Neuropsychology, NIMH, http://dally.nimh.nih.gov/index.html) on a computer with an Intel Core i3-540 processor. Stimuli were displayed at a viewing distance of 0.54 m, on a 25" Sony

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Trinitron CRT monitor with a resolution of 1280 by 1024 pixels, yielding a resolution of 31.5 pixels/degree of visual angle (dva). The monitor refresh rate was 85 Hz for monkey 1, and 75 Hz for monkey 2. A gamma correction was used to linearize the monitor output, and the gratings had 50% contrast. Monkeys performed a passive viewing task where they fixated centrally while stationary sinusoidal grating of either horizontal or vertical orientation and 2 cycle per degree spatial frequency, were presented in a location that covered all receptive fields recorded from the 25 electrode tips. Stimuli were presented 500 ms after fixation onset for 150 ms. Raw data were acquired at a sampling frequency of 32556 Hz using a 64-channel Digital Lynx 16SX Data Acquisition System (Neuralynx, Inc.). Following each recording session, the raw data were processed offline using commercial (Neuralynx, Inc.). Signals were extracted using Cheetah 5 Data Acquisition Software, with bandpass filtering set to allow for spike extraction (600-9000 Hz) and saved at 16-bit resolution.

In the present study, we focus on the period starting 200 ms before and finishing 200 ms after the stimulus onset, for 4 conditions (vertical gradings with pre/post cue in the receptor/opposite field) that will not be compared in details. The electrode recordings is firstly down-sampled from 32556 Hz to 1 kHz. A high-pass filter above 375 Hz is then applied - 3rd-order Butterworth filter at 0.75 of the Nyquist frequency [35] - followed by a smoothing of 4 ms to extract the envelope of the resulting signal, by retaining the 250 time points of 1000-ms period surrounding the stimulus onset.

3 Benchmark of detection performance for synthetic data

The workflow of our estimation procedure is schematically represented in Fig.1. We first consider a noise-diffusion network with discrete time as defined by Eq. (1) with the underlying connectivity matrix A (Fig. 1A, left). Then we generate its activity during a period T and evaluate the empirical covariances $\widehat{Q}$ using Eq. (2) with time shifts 0 and 1 respectively (Fig. 1A, middle). Covariance matrices are used to estimate the initial connectivity matrix A for partial correlation (PC), the multivariate autoregressive method (MVAR) and the Lyapunov optimization (LO) method while bivariate (GR2) and multivariate (GRf) Granger causality estimates are directly estimated from the simultaneous time series (Fig. 1A, right). For each method, we compare the estimated values against the original connectivity values (Fig. 1B, left). To infer absent connections, we apply continuous thresholds across the range of estimated values. Specifically, for any given threshold, we obtain a pair of false positive and negative rates associated with the right-tail distribution area of estimation values across absent connections and the left-tail distribution area of estimation values across present connections in A, respectively (Fig. 1B, middle). Finally, we calculate the resulting receiver operating characteristic (ROC) curve for the considered network simulation (Fig. 1B, right).

3.1 Estimation power improves when considering the whole network

We assess the performance of the studied methods across randomly connected graphs, which are simulated with different sizes ($N = 50$ to 150 nodes), connectivity strengths and number of time samples; the input noise corresponds to independent ζ in Eq. (1), meaning that the corresponding covariance matrix Σ in Fig. 1A is diagonal. For the case of the partial correlation we consider non-directional networks generated with symmetric connectivity matrices A. For each method, we analyze the tradeoff between the accuracy and the computational cost. The accuracy is firstly measured through the area under the ROC curve and the cost by the average computational time per simulation for each network configuration. Fig. 2A displays this ROC-based accuracy as a function of the number of time observations represented by box plots for 500 randomly connected networks. As theoretically predicted in Section 2, the partial correlation in general fails to provide a good estimation of the underlying connectivity in the discretized noise diffusion model. When considering many samples ($\sim 10^4$), all methods except PC can achieve high-accuracy estimates. However, for small sample sets, the MVAR method is shown to exhibit superior performance than Granger causality ones: as measured by the Mann-Whittney test, $p < 10^{-45}$, $p < 10^{-18}$ and $p < 10^{-5}$ for the three values of observed samples, respectively. The LO and MVAR performance are similar, $p > 0.2$ for the three same cases. The computation time required to perform each estimate is given in Fig. 2B for different

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

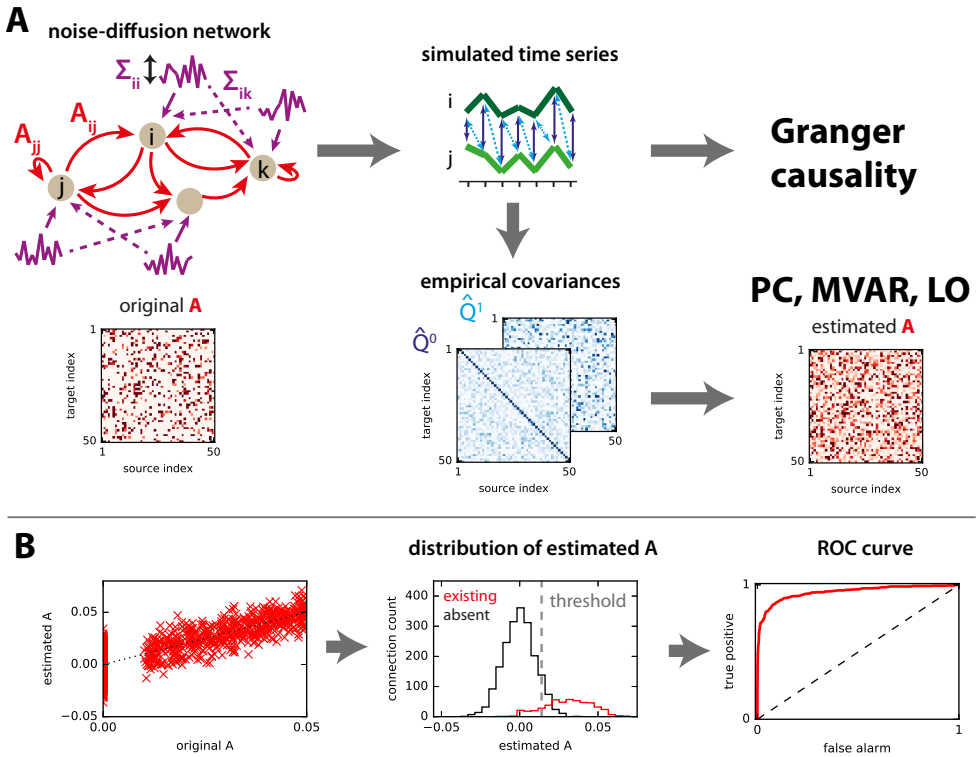

Figure 1: **Network model, calculation of empirical covariances with sampling and connectivity estimation. A:** (Left) For a network with randomly generated directed connectivity A and input variances Σ, its activity is simulated using Eq. (1). (Middle) The empirical covariance matrices $\widehat{Q}^0$ and $\widehat{Q}^1$ are calculated from the same time series using the formula in Eq. (2) for $\tau = 0$ and 1. (Right) Inference of connectivity: Granger causality analysis is directly applied to the generated simultaneous time series to estimate the existence of connections A_{ij} from node j to i. In parallel, the empirical covariance matrices are used to estimate the undirected connectivity using partial correlation (PC) on the one hand, and directed connectivity using multivariate autoregression (MVAR) and Lyapunov optimization (LO) on the other hand. **B:** (Left) For each generated network and estimation method, we compare the estimated against the original weights $A_{i,j}$ for all pairs. (Middle) The distribution of estimation values across present (red) and absent connections (black) in the original connectivity matrix A. The gray dashed line stands for a given threshold value in the estimates domain. (Right) By moving the threshold value across the whole distributions domain, we obtain pairs of true positive and false-alarm rates that are used to construct a ROC curve.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simulation times. This plot primarily shows that MVAR are orders of magnitude below other methods. GRf and LO are more than one order of magnitude longer to compute than GR2. Yet, LO scales better than GRf with the number of observed samples. Among the remaining methods, GR2 attained a longer computation time than MVAR.

For the generative model considered here, partial correlation is shown to be a mismatched estimation method and Granger causality methods should be avoided for their poor accuracy (and long computation time for 'full' Granger). In contrast, MVAR attains the best performance among the studied methods for both accuracy and computation time.

3.2 Influence of network parameters on ROC-based detection

Next, we examine the robustness of these estimates are against controlled network parameters. To do so, we plot the ROC-based performances in Fig. 2A as a function of the network size N, the network density, the minimum weight over the whole network and the mean sum of incoming weights per node; note that the weights in the 500 networks considered in Fig. 2 are uniformly distributed between a minimum and a maximum, the inputs are white noise (i.e., without cross-correlation). For illustration purpose, the 500 networks are grouped in quartiles for each network parameter in Fig. 2C. Not surprisingly, the estimation accuracy of all methods decreased as a function of the network size N and density, and increased as a function of the minimum connectivity weight and the mean incoming weight per node. Interestingly, in challenging configurations (e.g., minimum weight close to 0), MVAR and LO methods consistently show a superior performance by a larger gap compared to Granger causality methods. These findings further corroborate the use of MVAR as an accurate and robust method to estimate connections in noise diffusion models over random network topologies.

3.3 A robust non-parametric significance test for MVAR

We have thus far examined the performance of different estimation methods based on ROC, considering the whole range of thresholds to discriminate between existing and absent connections in the simulated networks. However, in real data settings, the decision on the existence of a connection is typically based on comparing the value of the connection estimate with a given statistical threshold; for Granger causality, such tests have been developed, for example, based on the F statistics [4]. Equivalently, it is sufficient to know how the values of the estimates for absent connections are distributed, in order to select a desired false-alarm (or false-positive) rate. In this section, we focus on the MVAR method and develop a significance test that provides a null-hypothesis distribution for absent connections. We also discuss possible variations of the method.

Our approach relies on the fact that covariances reflect the underlying connectivity: we thus construct the surrogate estimates by repeatedly performing random circular shifts of the observed time series independently for all nodes and computing the corresponding covariance matrices, from which we evaluate the surrogate connectivity (see Fig. 3A). In the surrogate covariance matrices, this procedure destroys the off-diagonal covariances for both $\widehat{Q}^0$ and $\widehat{Q}^\tau$, while preserving the autocovariances on the diagonal. The core result underlying our surrogate approach is illustrated in Fig. 3B: the distribution of surrogate estimates (thick black line) is compared against the distribution of existing (red) and absent (blue) connections in a simulated random network model, underlying the match between the blue and black curve. Then, we consider two manners of pooling the off-diagonal elements of the generated surrogate matrices to obtain the null-hypothesis distribution. The black histogram in Fig. 3C is obtained from 100 surrogates and lumps together the $N(N-1)$ off-diagonal elements of the matrix. In comparison, the gray distribution corresponds to the diagonal elements and does not match the blue curve. From the black distribution of surrogate value, we can thus perform a detection test with a threshold defined by a percentage of the right tail that corresponds to the desired false-alarm rate (here 5%), as illustrated by the same black curve and the gray dashed line in Fig. 3C; we refer to this as the *global* test since the black distribution includes all matrix elements except for the diagonal. If, instead, we consider for each connection only the 100 corresponding matrix elements in the 100 surrogates, we obtain the *local* version of the previous distribution, from which a similar detection can be performed. Now we examine which of the two tests

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

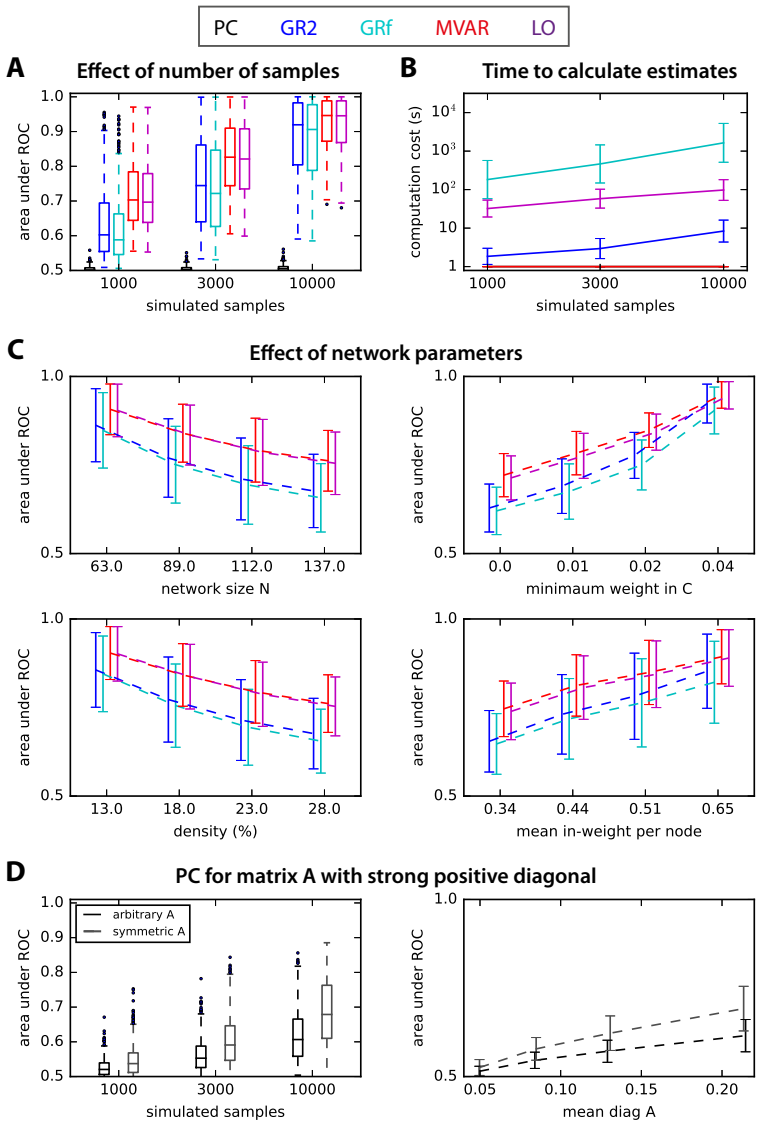

Figure 2: **ROC accuracy and computation time. A:** Area under ROC for estimated A obtain from Granger (pairwise for GR2 and with full network for GRf), PC, MVAR and LO. The x-axis indicates three sample size for the observed network activity. The boxplots indicate the median (horizontal bar in the middle of the box) as well as the upper and lower tail spreads (standard deviation) over 500 simulated networks of various sizes and connectivity strengths. **B:** Computation time needed to obtain the Granger estimates for all connections, as compared to PC and MVAR estimates. The graph show the increase of the computation time as a function of the simulated time. **C:** Effect of network parameters on ROC-based detection performance. Influence of network size N, connectivity density, sum of recurrent connectivity strengths, minimum weight in A, mean noise on the diagonal of Σ and mean off-diagonal noise in Σ on the ROC-based accuracy in Fig. 2A. In each graph, the network configurations have been grouped in quartiles according to the parameter plotted on the x-axis, and the corresponding group mean and standard deviations are indicated; the curves are displaced horizontally to favor the legibility. **D:** ROC-based performance of PC for arbitrary (black) and symmetric (gray) original matrices A, whose diagonal elements are all set to positive values. The left panel displays the performance as a function of the number of simulated samples as in A; the right panel as a function of the mean value of the diagonal elements.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

leads to a better performance, considering the trade-off that local distribution may be tailored to their corresponding connections, but at the expense of losing sample size (distinct thresholds in Fig. 3C). Following the same network example, we assess in Fig. 3D the resulting true-positive and false-alarm rates attained by both local and global tests (circles and triangles, respectively; 50 surrogates each) and place them on a ROC curve for thresholds determined by right-tail percentages ranging from 1 to 5%. Fig. 3D shows some variability in the resulting false-alarm rates, but both methods satisfactorily coincide with the desired rate and the ROC curve.

This small difference is further illustrated in Fig. 3E, where the actual false-alarm rates attained by both methods (with varying number of surrogates) over 500 randomly generated networks are plotted against right-tail percentages of their associated null distributions. The control of false-alarm rates hardly shows any difference across the various numbers of surrogates and test types (local versus global), demonstrating the robustness of our method. Following, we fix a given false-alarm rate (4%) (e.g., for the same 500 random networks) and evaluate the true negative (miss) rate of both methods depending on the actual weight strength: in Fig. 3F, connections are grouped in terciles for each network configuration. In both panels, connections with larger weights are easier to detect, but only mildly when compared to the trend with, e.g., the minimum weight in the network in Fig. 2C. Interestingly, the detection performance for local test improves with the number of surrogates and seems to stabilize above 200 (right panel), unlike the global test for which 25 surrogates give the same miss rate as 400 (left panel).

From Fig. 3F, we conclude that the local test is preferable to the global test provided more than 100 surrogates are generated. However, the computational cost increases with the number of surrogates, as illustrated in Fig. 4A by the red curves: for 1000 to 10000 observed samples, generating 50 surrogates takes as much time as the parametric significance test from pairwise Granger analysis (GR2 in dark blue) for all connections in the network. On the other hand, generating 400 surrogates is still much faster than the full Granger analysis (GRf in cyan).

Finally, we also compare the parametric significance F-test for pairwise Granger [4] with our non-parametric method in Fig. 4B. In both Fig. 3 and 4, the novelty compared to Fig. 2 is that networks are simulated with cross-correlated inputs (i.e pink noise, indicated by the purple dashed arrows in Fig. 1A). Importantly, the false-rate alarm is rather stable for our method, whereas the parametric tests for Granger's causality based on residual noise worsen their performance as the input cross-correlation increases.

3.4 Significance test applied to LO estimates

We also apply the same surrogate-based test to the LO estimate to assess the viability of the generated null distribution and compare it against the non-parametric MVAR. Due to the large computation cost for LO, we focus on the global test with 25 surrogates (see Fig. 4A where the purple curve has similar order of magnitude as GRf in cyan) and the same 500 randomly generated networks with various sizes as in Fig. 3. In these simulated networks, the LO method hardly overestimates the false-alarm rate more than the MVAR, but attains slightly better true positive detection for a relative small number of observed samples (1000 samples in Fig. 5A). This means that the null distribution obtained by time rolling is also valid for LO, albeit a bit less precise. In addition, the influence of global network parameters such as the network minimum weight is similar for both MVAR and LO in Fig. 5B, which also agrees with the ROC-based performance in Fig. 2C. Last, the similarity between both estimates, measured by the Pearson and Spearman correlations, is large and the match becomes quasi perfect when the number of samples increases. Together with (Fig. 5C), these findings support the use of surrogate-based LO in experimental conditions where costly computations are manageable (purple curve in Fig. 4A). The advantages of LO as compared to MVAR, namely the possibility to impose topological constraints on the connectivity as was discussed in [10], will not be discussed here; we simply want to demonstrate the accuracy of LO and compare it to MVAR.

3.5 Influence of network topology

In this part, we test and compare the robustness of global and local surrogate-based detection tests to specific connections and topological configurations. Here, the number of observed samples is 3000, the

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

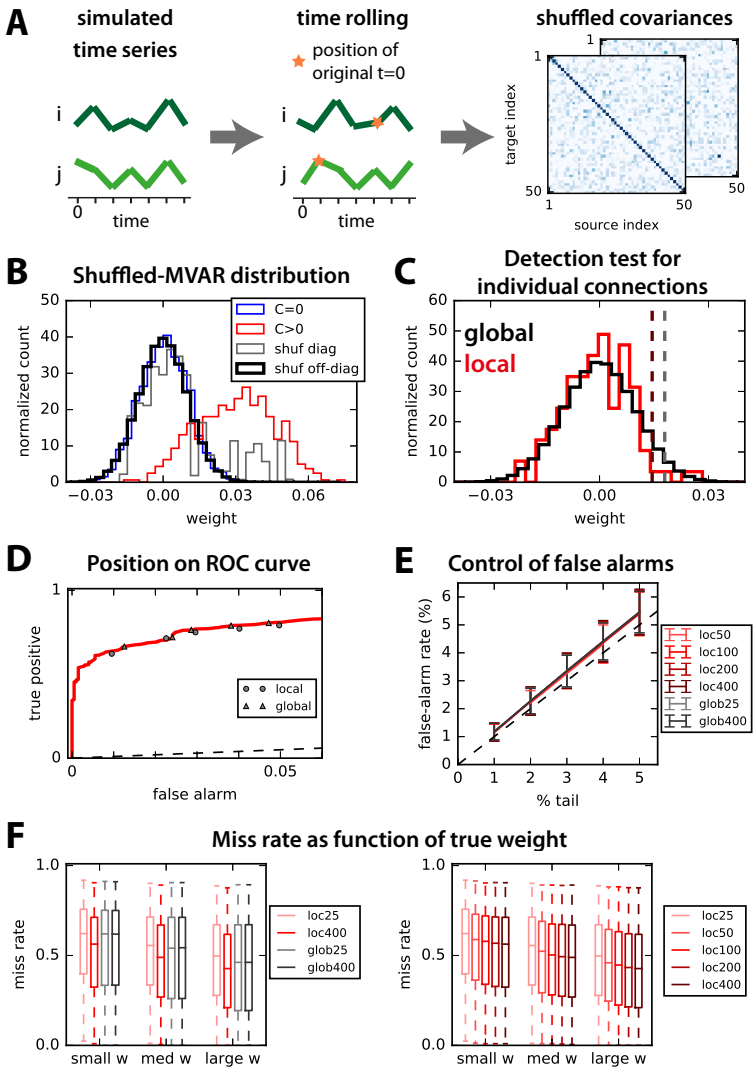

Figure 3: **Non-parametric test to assess significance based on time-rolling surrogates. A:** Schematic illustration of time rolling applied on the observed time series to generate the surrogate covariances. **B:** Pooled distributions of estimated weights for the existing (in red) and absent (blue) connections. The thick black curve indicates the distribution of connections over 100 surrogates (except self-connections on the matrix diagonals), which closely matches the blue distribution. In contrast, the gray curve indicate the same distribution for diagonal elements of the surrogates, which does not match the blue curve. **C:** For a given connection, we compare two methods: 'local' in red where the null distribution corresponds to the surrogates corresponding to that connection as matrix element; 'global' in black where the null distribution is the pooled distribution for all off-diagonal elements of the surrogates (same as in B). The dark red and gray dashed lines indicate the detection thresholds corresponding to the 5% tail for those two options. **D:** Example of performance of the two non-parametric methods for 100 surrogates on the ROC curve for a desired false-alarm rate ranging from 1 to 5%: comparison between the local (circles) and global (triangles) method. **E:** Comparison of the desired (% of the tail of null distribution) and actual rate of false alarms for the two non-parametric methods when varying the the number of surrogates. Error bars indicate one standard deviation for 500 random networks; importantly, inputs for these networks have cross-correlation, unlike in Fig. 2. **F:** Influence of the strength of original weight on the detection performance for a desired false-alarm rate set to 4%. In both panels, lighter colors indicate smaller numbers of surrogates, in red for the local test and gray for the global test.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

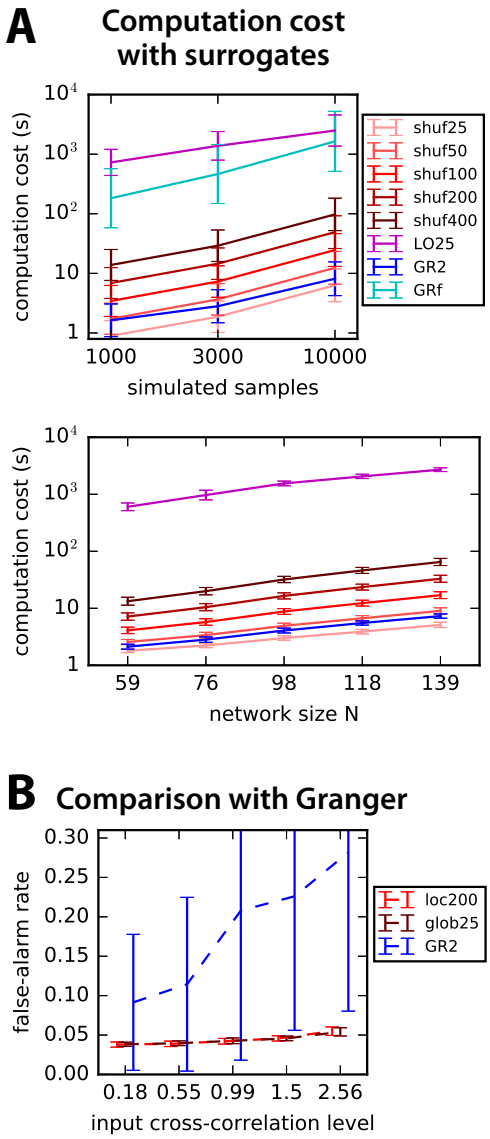

Figure 4: **Comparison of our surrogate method with pairwise Granger causality analysis. A:** Computation cost for the evaluation of the shuffling-based surrogates as a function of the number of simulated samples (top) and the network size N (bottom). The red curves with various intensities correspond to MVAR with 25 to 400 surrogates (same as Fig. 3E and F); the blue and cyan curves to pairwise and full Granger analysis; the purple curve to LO with 25 surrogates. Error bars in all panels indicate one standard deviation over the 500 random networks as in Fig. 3E. **B:** Comparison between the parametric test for pairwise Granger (blue curve) and our non-parametric methods (red for local and gray for global). The x-axis indicates the strength of input correlations (i.e., pink noise) in the simulated network. The desired false-alarm rate is set to 4% as in Fig. 3F. The number of time samples is 3000 here.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

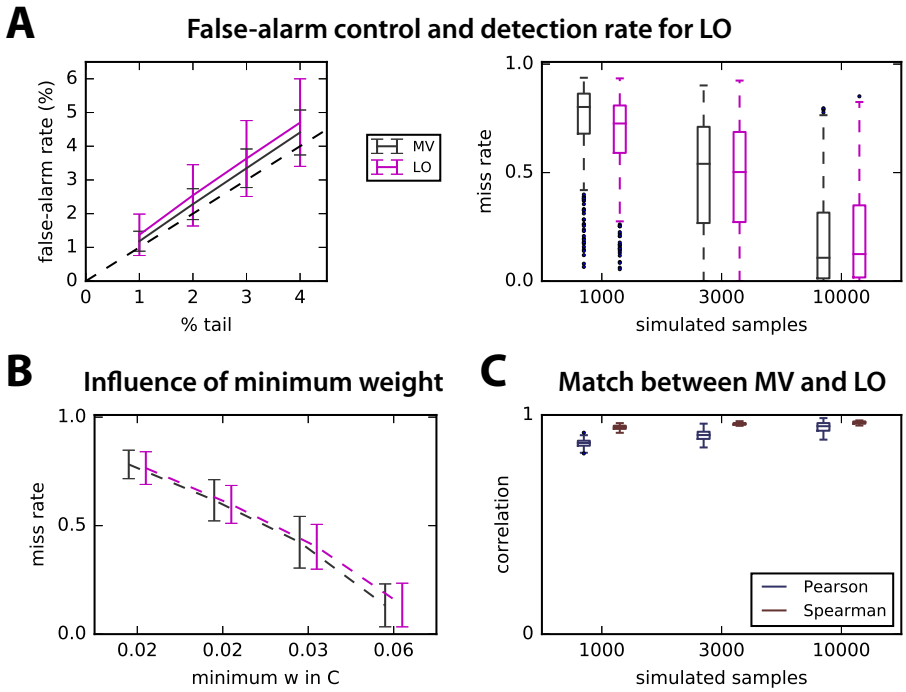

Figure 5: **Global significant test for LO. A:** Comparison of the control of the rate of false alarm and detection miss between MVAR (gray) and LO (purple) for the global test with 25 surrogates and the 500 random networks in Fig. 3 (the error bars indicate one standard deviation). **B:** Influence of the minimum weight in the original A on the miss rate (similar to Fig. 2C for the ROC-based performance). **C:** Similarity between the MVAR and LO estimates as measured by the Pearson and Spearman correlation coefficients.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

number of surrogates is 100 for the local test and 50 for the global test and the results correspond 500 networks of each type; in all cases, the simulated networks vary in size N, level of input cross-correlation, connectivity density and distribution of recurrent weights. To start with, we compare the miss rate for unidirectional, reciprocal and self connections in the random networks of Fig. 3. Fig. 6A shows that the miss rate is similar in unidirectional and reciprocal connections with the local test, which performs slightly better than the global test (as in Fig. 3F). In contrast, self connections cannot be detected with a local approximation due to the method construction. The intuition is that, because time-rolling surrogates preserve the autocovariances, they fail to build a proper null-distribution for self-connections; however, the global null-distribution appears suitable for their detection.

Now we consider more elaborate network topologies than the random connectivity (Erdös-Rényi) considered so far, namely modular and hierarchical networks. In Fig. 6B, we simulate 500 modular networks with two groups (green and blue) linked by hubs (red, about 5 to 15% or the nodes). We find a similar miss rate in hub-group and intra-group connections using local and global surrogates. In Fig. 6C we simulate hierarchical networks, where connections are grouped into connections linking center and intermediate nodes, connections linking intermediate nodes and leaves, and self connections. This network type is much sparser than the two types in A and B, yielding a much better detection performance (miss rate < 0.1 for most network configurations in Fig. 6C). Further, it exhibits more homogenous nodes which allows for the global approximation to more suitable than the local approximation in this topology. Again, self-connections in this topology suffer from a improper null-hypothesis distribution, causing the local test to fail dramatically for them. As can be seen in Fig. 6D, the control of rate of false alarms attained by local and global approaches in the three studied network topologies is very similar across network topologies and approximation types.

Finally, we consider a network with both positive and negative interactions (from 5 to 50% of existing connections) and perform the test by defining a threshold on both tails of the null distributions. As can be seen in Fig. 6E, the excitatory/inhibitory nature of the connections affect neither the false-alarm nor the miss rate.

4 Application to experimental data

4.1 Multiunit activity data obtained from Utah electrode array in monkey

Now we consider data recorded from a monkey performing a visual task, where the stimulus corresponds to vertical gratings covering all recorded V1 receptive fields from the Utah arrays (see Methods for details). We aim to provide a proof of concept for the connectivity analysis for this type of data, so as to complement the more classical analysis based on the activity of individual channels; therefore we do not focus on comparing the 4 stimulus conditions or other details.

The multiunit activity envelope (MUAe) is obtained as described in Methods. In Fig. 7A, the resulting MUAe is represented for two out of the 26 channels (red and purple) for two trials in the top and middle panels, 400 ms before and 600 ms after the stimulus onset. The average MUAe over 200 trials for these two channels is displayed in the bottom panel, exhibiting a peak immediately after the stimulus. Among the 26 channels, about a third show a large increase in activity after the stimulus onset as compared to before (namely, a post-stimulus mean activity larger by more than three standard deviations compared to the pre-stimulus activity); almost all channels show a moderate increase of one standard deviation.

To investigate further the temporal information conveyed by MUAe, we calculate the pairwise covariances between channels, after centering the MUAe activity individually for each trial. We firstly verify that the model can be applied to these data, by examining the MUAe autocovariances in Fig. 7B, which exhibit a profile corresponding to an exponential decay up to two time shifts (i.e., 8 ms for the downsampling every 4 ms), namely a straight line in the log plot. This suits an autoregressive model with large positive values on the diagonal of the connectivity matrix A, which generates this type of profiles.

The two connectivity matrices for the 26 channels estimated using the MVAR method before and after the stimulus are illustrated in Fig. 7C for condition 1: we find larger off-diagonal values for the period after the stimulus than before. This is actually true for all conditions, as indicated by the more

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

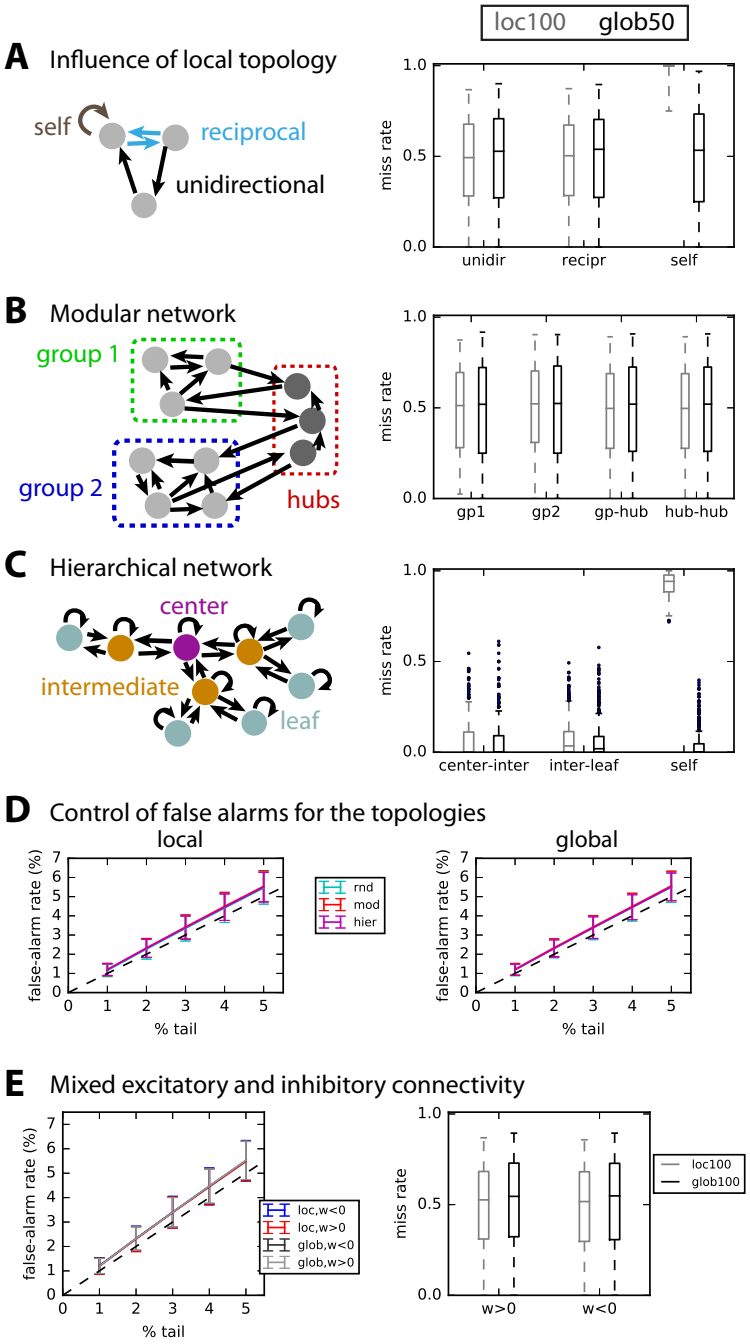

Figure 6: **Robustness to network topology. A:** Detection performance for unidirectional, reciprocal and self connections in 500 the randomly connected networks used so far in Figs. 3 and 5. **B:** Detection performance for modular topology schematically represented in the left: each of the 500 networks comprises of two groups connected by hubs. The connections are separated depending on whether the connect group nodes or hubs, as indicated by the diagram on the left. **C:** Similar to B with a hierarchical topology, where connections are grouped in 3 subsets: from center to intermediate nodes; from intermediate nodes to leaves; self connections. The results concern 500 such networks, which all have very low density. **D:** Control of false-alarm rate for the local (left) and global (right) significance tests and the three network topologies; the plot is similar to Fig. 3E. **E:** Control of false-alarm rate and miss rate for networks with both excitatory and inhibitory connections.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spread distributions in red as compared to gray in Fig. 7D. The channels appear to be coordinated at the considered time scale of 4 ms and their collective interaction scheme is affected by the stimulus presentation.

4.2 Significance test for real data: interactions related to stimulus presentation

We then use the local and global tests based on 1000 time-rolling surrogates to retain only significant interactions from the estimates in Fig. 7C: this leaves a few interactions for the pre-stimulus period in Fig. 8A (left panel), 11 out of 650, which is of the order of the desired false-alarm rate set to 1% (namely, the extreme 0.5% of each tail). In contrast, many more post-stimulus interactions survive the significance tests in the right panel: interestingly, many of these connections are unidirectional, as indicated by the red pixels in Fig. 8B, to be compared to the green pixels indicating reciprocal connections of the same sign. The summary of symmetric versus asymmetric interactions for the 4 conditions is given in Fig. 8C for both local and global tests for the post-stimulus period. Varying the threshold on the tail of the null-distributions, we see that the number of detected interactions is close to the desired false-alarm rate for the pre-stimulus period in Fig. 8D (dark red and black curves, respectively). In contrast, post-stimulus interactions are more numerous for both local and global tests (light red and grey); moreover, the difference between post- and pre-stimulus interactions remains roughly stable for the considered range of false-alarm rate, especially for the local test. The latter detects about 5% significant interactions (corresponding to 6% for a desired false-alarm rate of 1% with the local test), that is, $26 \times 25 \times 0.05 \sim 32$ interactions in the network of 26 nodes. In contrast, the global test sets a single threshold for all connections, disregarding disparities across channels and leaving only the strongest interactions in absolute values. Therefore we expect the local test to work better for real data.

Finally, we check the relationship between the strengths of significant interactions and the increase of average MUAe observed in Fig. 7A. In Fig. 8E, the plotted dots correspond to the sum of incoming (left panel) and outgoing (right panel) significant interactions for each node in the pre- and post-stimulus periods (gray and red, respectively). Note that the plotted dots at the extreme right of both panels correspond to channel 12, which has a high MUAe activity that does not significantly increase with the stimulus presentation. The summed interaction values positively correlate with the mean MUAe only for the outgoing connections: $p \sim 10^{-3}$ with a coefficient of 0.32 before the stimulus and $p < 10^{-4}$ with 0.41 after; in contrast, incoming connections have $p > 0.01$ for both periods. This suggests a gain increase of the neuronal populations due to the stimulus, which may drive propagating activity through their anatomical connections and induce effective interactions between them. If the detected interactions were a result of a noisy detection test, there would likely be no difference between incoming and outgoing interactions. Our results thus suggest that the estimated structure is not the mere consequence of observation noise.

5 Discussion

5.1 Beyond Granger causality tests: properly integrating the network feedback in the connectivity detection

Granger causality (both GR2 and GRf) lead to individual tests for each connection, whereas our method provides simultaneous detections for all connections at once. In addition to the computational cost (especially larger for GRf), the conceptual difference of our approach is that it does regress the activity of the *whole* network with its past, a time step earlier. This is to our understanding the key reason why our network-oriented method outperforms classical Granger causality tests: it takes into account network effects due to the feedback in a proper manner, which appears crucial in networks with input cross-correlations (Fig. 4B). In a sense, our method can be seen as an "unscented" extension of Granger causality for the multivariate case. The MVAR/LO estimation methods output a unique network estimate for which a multidimensional null-hypothesis is built. Our time-rolling procedure used to generate the surrogates can be encompassed in the family of contained randomization methods [28, 26]. Specifically,

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

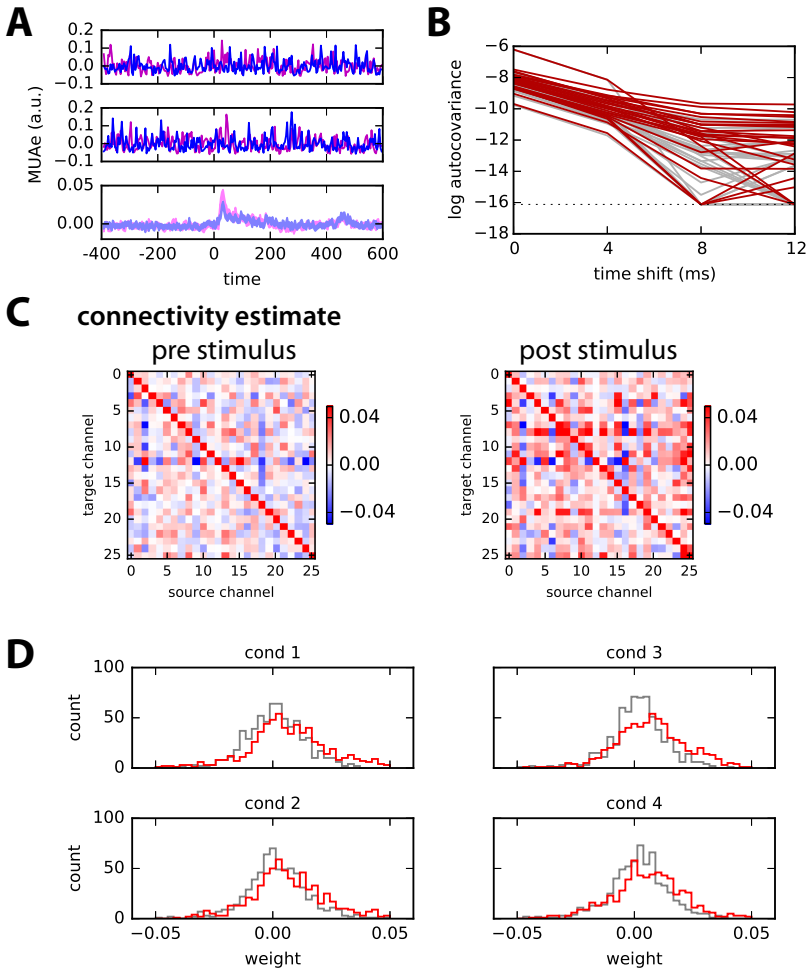

Figure 7: **Application to multiunit activity (MUAe) data. A:** Example of two trials (top and middle panels) of multiunit activity envelope (MUAe) for two channels of recordings using Utah electrode array in the primary visual cortex of a monkey (in arbitrary units; see text for further details). The bottom panel represents the average over 200 trials (with standard-error mean for the thickness of the curve). The stimulus is presented at time point 100 (actually 400 ms, since the smoothing corresponds to a smoothing window of 4 ms). **B:** Autocovariances of MUAe signals for time shifts up to 12 ms averaged over 200 trials plotted with a log y-axis: comparison of signals before (gray) and after (red) the stimulus presentation. **C:** MVAR estimates of the connectivity between the 26 channels for the MUAe activity 200 ms before and after stimulus presentation (i.e., 50 time points each), averaged over 200 trials. The scaling has been optimized to enhance the legibility off-diagonal elements. **D:** Comparison of the distribution of connectivity weights (off-diagonal elements in D) for the 4 conditions. Gray and red indicate before and after the stimulus, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

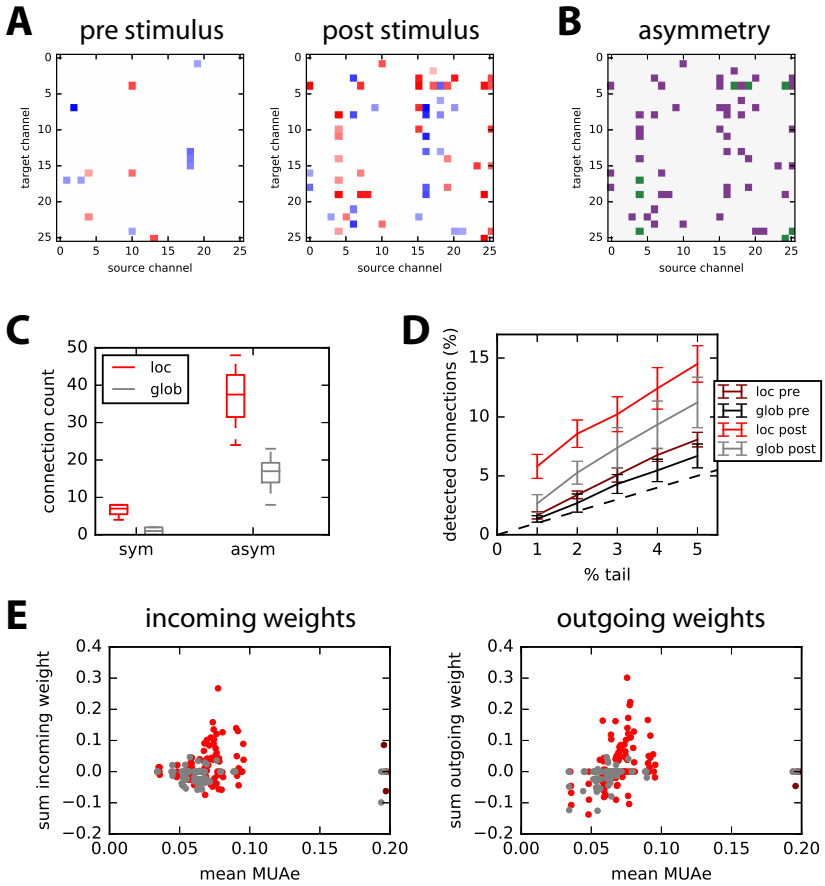

Figure 8: **Detection of significant connections. A:** Examples of significant connections with the time-rolling surrogate method; left and right panels correspond to pre- and post-stimulus periods, respectively. The p-value corresponds to the upper and lower 0.5% tails of 1000 surrogates (local test). Many more connections are found for post- than pre-stimulus period. **B:** Matrix of post-stimulus interactions where green pixels indicate symmetric and purple asymmetric significant connections; other are left blank. **C:** Comparison of number of asymmetric connections versus symmetric connections for the local (red) and global (gray) tests over the 4 conditions. **D:** Ratios of detected connections for pre- and post-stimulus periods with the desired false-alarm rate equal to 1 to 5% corresponding to both local and global tests. **E:** Sum of incoming (left) and outgoing (right) significant weights for each channel (dots) for the pre- and post-stimulus period in gray and red, respectively, plotted against the mean MUAe for the 4 conditions.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

it preserves the autocovariances while destroying the cross-covariances (off-diagonal elements in the covariances matrices with and without time shift); note that the resulting border effect is negligible here as we observe more than 1000 time samples. The number of observed samples of the network activity determines the level of spurious covariances, which in turn induce the width of the null-distribution for the off-diagonal elements used to determine the significance of the links. We found that using individually the null-distribution corresponding to each connection (local test) gives slightly better results than lumping together all off-diagonal elements of all surrogates (global test) provided sufficiently many surrogates are generated, which requires more computation time. Further studies should thoroughly compare the proposed method against additional methods preserving autocorrelation functions - e.g. phase randomization [28], trial permutation [18] - in synthetic and real data to elucidate the potentially distinct information brought by each method.

The problem of multiple comparison is intrinsic to brain connectivity detection as the number of testable connections across brain regions is massive [25]. In this context, different approaches have been developed to control the familywise error rate in the weak sense. For instance, many studies on neuroimage data [9, 22] or electrophysiology [15] have resorted to procedures that control the false discovery rate (FDR) [5], namely, the expected number of falsely declared connections among the total number of detections. These methods make decisions on single connections relying on the entire sequence of p-values computed for each connection and yield substantial statistical power gains over more conservative methods such as Šidák-Bonferroni [1]. With the ever growing application of graph theory to brain connectivity, new methods have been proposed that exploit the clustered structure of the the declared connections [37, 12] to propose cluster-based statistical tests [18] that attain similar performance to FDR methods. The present work can therefore be understood as a primary step before performing any or several multiple-correction procedures. By defining an accurate null model of inexistent connections, p-value estimates per connection are improved and cluster-based surrogate distributions can be better approximated, which is expected to empower the overall control of false positive rates in network connectivity analysis.

During the initial comparison of MVAR with other methods, we also found that partial correlation do not perform well for the discrete-time generative model when the connectivity matrix is arbitrary (Fig. 2A). Nevertheless, the performance improves when the connectivity matrix A has strong positive elements on its diagonal (Fig. 2D), which corresponds to autocovariance profiles with exponential decay (Fig. 7B and C for an example with real data).

5.2 Applications to real electrophysiological and neuroimaging data

Here we have applied the method to MUAe recorded from macaque area V1. Our point was to verify the applicability of our network connectivity analysis to this type of data rather than addressing biological questions, which is beyond the scope of the current manuscript. Looking at individual trials Fig. 7A, it is rather surprising that the variable MUAe activity convey temporal information due to causal interactions that can be detected. This variability is therefore not an absolute limitation to temporal coordination, although the latter only becomes apparent over multiple trial repetitions.

In a previous paper [10], we applied a continuous-time version of the network model, namely a multivariate Ornstein-Uhlenbeck process, to fMRI data and demonstrated that it is well adapted to this type of data. No significance test was used there, since the point was to estimate the strengths of interactions between brain regions using the LO method and not detect interactions; rather we used topological constraints inferred from diffusion-weighted magnetic resonance imaging or diffusion tensor imaging. We nevertheless expect the surrogate-based detection test to be viable for that related model, as it was demonstrated for the discrete-time version without topological constraints here.

More generally, we target adequate preprocessing of multivariate time series - activity aggregation over hundreds of voxels for fMRI and MUAe for electrode recordings - such that the autocovariance profiles match the exponential decay of the dynamic model with linear feedback that underlies the connectivity analysis: provided it resembles an exponential decay over a few time shifts as in Fig. 7B, we expect our approach to be applicable. In theory, stationarity of the time series remains a critical issue as we need sufficiently many observed samples to obtain precise covariances from which we evaluate the connectivity estimates. On the other hand, despite the non-stationary nature of the MUAe signals illustrated in Fig. 7A,

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

we obtain significant interactions that preferentially arise for channels exhibiting an increase of MUAe, but are not merely a consequence of it: the method provides information about the directionality of the interactions in Fig. 8B to E.

Acknowledgments

MG acknowledges funding from the Marie Sklodowska-Curie Action (grant H2020-MSCA-656547). MG and GD were supported by the Human Brain Project (grant FP7-FET-ICT-604102 and H2020-720270 HBP SGA1). GD and ATC were supported by the European Research Council Advanced Grant DYSTRUCTURE 588 (Grant 295129) The authors are grateful to Robert Castelo for constructive discussions.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

[1] H. Abdi. The bonferonni and šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, 3:103–107, 2007.

[2] T. Amemiya. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica: Journal of the Econometric Society*, pages 999–1012, 1974.

[3] F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P. Rossini, L. Ding, Y. Ni, J. Cheng, K. Christine, J. Sweeney, and B. He. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution eeg and fmri data by directed transfer function. *Neuroimage*, 24(1):118–131, 2005.

[4] L. Barnett and A. Seth. The mvgc multivariate granger causality toolbox: A new approach to granger-causal inference. *J Neurosci Methods*, 223:50–68, 2014.

[5] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[6] N. Bing and I. Hoeschele. Genetical genomics analysis of a yeast segregant population for transcription network inference. *Genetics*, 170(2):533–542, 2005.

[7] R. Castelo and A. Roverato. A robust procedure for Gaussian graphical model search from microarray data with p larger than n. *The Journal of Machine Learning Research*, 7:2621–2650, 2006.

[8] K. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

[9] C. Genovese, N. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

[10] M. Gilson, R. Moreno-Bote, A. Ponce-Alvarez, P. Ritter, and G. Deco. Estimation of directed effective connectivity from fmri functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput Biol*, 12(3):e1004762, 2016.

[11] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[12] C. Han, S. Yoo, S. Seo, D. Na, and J.-K. Seong. Cluster-based statistics for brain connectivity in correlation with behavioral measures. *PLoS one*, 8(8):e72332, 2013.

[13] M. Kaminski and K. Blinowska. A new method of the description of the information flow in the brain structures. *Biological cybernetics*, 65(3):203–210, 1991.

[14] M. Kamiński, M. Ding, W. Truccolo, and S. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.

[15] A. Lage-Castellanos, E. Martínez-Montes, J. Hernández-Cabrera, and L. Galán. False discovery rate and permutation test: an evaluation in erp data analysis. *Statistics in medicine*, 29(1):63–74, 2010.

[16] S. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

[17] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[18] E. Maris and R. Oostenveld. Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, 164(1):177–190, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/100669; this version posted January 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[19] J. Massey. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, pages 303–305. Citeseer, 1990.

[20] A. Messé, D. Rudrauf, H. Benali, and G. Marrelec. Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. *PLoS Comput Biol*, 10:e1003530, 2014.

[21] G. Michalareas, J. Schoffelen, G. Paterson, and J. Gross. Investigating causality between interacting brain areas with multivariate autoregressive models of meg sensor data. *Human Brain Mapping*, 34(4):890–913, 2013.

[22] T. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, 12(5):419–446, 2003.

[23] A. Reverter and E. Chan. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*, 24(21):2491–2497, 2008.

[24] B. Rogers, S. Katwal, V. Morgan, C. Asplund, and J. Gore. Functional mri and multivariate autoregressive models. *Magnetic Resonance Imaging*, 28(8):1058–1065, 2016/11/05 2010.

[25] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[26] T. Schreiber. Constrained randomization of time series data. *Physical Review Letters*, 80(10):2105, 1998.

[27] T. Schreiber. Measuring information transfer. *Phys Rev Let*, 85(2):461, 2000.

[28] T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77(4):635, 1996.

[29] K. So, A. Koralek, K. Ganguly, M. Gastpar, and J. Carmena. Assessing functional connectivity of neural ensembles using directed information. *J of Neu Eng*, 9(2):026004, 2012.

[30] A. Storkey, E. Simonotto, H. Whalley, S. Lawrie, L. Murray, and D. Mcgonigle. Learning structural equation models for fmri. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1329–1336. MIT Press, 2007.

[31] H. Supèr and P. Roelfsema. Chronic multiunit recordings in behaving animals: advantages and limitations. *Prog Brain Res*, 147:263–282, 2005.

[32] A. Tauste Campo, M. Martinez-Garcia, V. Nácher, R. Luna, R. Romo, and G. Deco. Task-driven intra-and interarea communications in primate cerebral cortex. *Proc Natl Acad Sci U.S.A.*, 112(15):4761–4766, 2015.

[33] A. Thiele, L. Delicato, M. Roberts, and M. Gieselmann. A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. *J Neurosci Methods*, 158:207–211, 2006.

[34] C. Wilke, L. Ding, and B. He. Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function. *IEEE Trans Biomed Eng*, 55(11):2557–2564, 2008.

[35] www.scipy.org. the scientific python library.

[36] Y. Yuan, C. Li, and O. Windram. Directed Partial Correlation: Inferring Large-Scale Gene Regulatory Network through Induced Topology Disruptions. *PLoS ONE*, 6(4), 2011.

[37] A. Zalesky, A. Fornito, and E. Bullmore. Network-based statistic: identifying differences in brain networks. *Neuroimage*, 53(4):1197–1207, 2010.