

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.


For additional information about this publication click this link.

<http://hdl.handle.net/2066/177101>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

ORIGINAL RESEARCH

The impact of hemodynamic variability and signal mixing on the identifiability of effective connectivity structures in BOLD fMRI

Natalia Z. Bielczyk^{1,2}  | Alberto Llera³ | Jan K. Buitelaar^{1,2} | Jeffrey C. Glennon^{1,2} | Christian F. Beckmann^{1,2,3}

¹Donders Institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands

²Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

³Oxford Centre for Functional MRI of the Brain, John Radcliffe Hospital, Oxford, UK

Correspondence

Natalia Z. Bielczyk, Donders Institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands.

Email: natalia.bielczyk@radboudumc.nl

Funding information

NB has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 305697 (OPTIMISTIC), the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 278948 (TACTICS) and European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement no 603016 (MATRICS). CFB gratefully acknowledges funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z]. CFB is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003)

Abstract

Purpose: Multiple computational studies have demonstrated that essentially all current analytical approaches to determine effective connectivity perform poorly when applied to synthetic functional Magnetic Resonance Imaging (fMRI) datasets. In this study, we take a theoretical approach to investigate the potential factors facilitating and hindering effective connectivity research in fMRI.

Materials and Methods: In this work, we perform a simulation study with use of Dynamic Causal Modeling generative model in order to gain new insights on the influence of factors such as the slow hemodynamic response, mixed signals in the network and short time series, on the effective connectivity estimation in fMRI studies.

Results: First, we perform a Linear Discriminant Analysis study and find that not the hemodynamics itself but mixed signals in the neuronal networks are detrimental to the signatures of distinct connectivity patterns. This result suggests that for statistical methods (which do not involve lagged signals), deconvolving the BOLD responses is not necessary, but at the same time, functional parcellation into Regions of Interest (ROIs) is essential. Second, we study the impact of hemodynamic variability on the inference with use of lagged methods. We find that the local hemodynamic variability provide with an upper bound on the success rate of the lagged methods. Furthermore, we demonstrate that upsampling the data to TRs lower than the TRs in state-of-the-art datasets does not influence the performance of the lagged methods.

Conclusions: Factors such as background scale-free noise and hemodynamic variability have a major impact on the performance of methods for effective connectivity research in functional Magnetic Resonance Imaging.

KEYWORDS

brain parcellation, causal discovery, Dynamic Causal Modeling, effective connectivity, functional Magnetic Resonance Imaging, neuronal noise

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Brain and Behavior* published by Wiley Periodicals, Inc.

1 | INTRODUCTION

Studies on the communication in large-scale networks in fMRI were initiated as the *functional* connectivity (FC) research. FC quantifies the strength of communication between brain regions by means of correlation and therefore without specification of directionality (van den Heuvel & Pol, 2010).

Extending network research in fMRI from functional to effective connectivity could provide a substantial advance to the understanding of brain dynamics in health and disease (Bielczyk, Buitelaar, Glennon, & Tiesinga, 2015; Fornito, Zalewsky, & Breakspear, 2015; Friston, 2011; Sporns, 2014). Effective connectivity in fMRI is a complex research problem that involves not only specification of the presence or absence of connections, but also the directionality of the information flow.

There are two main classes of methods for effective connectivity research in fMRI. On one hand, we have structural causal models, which are based on the dependencies between the BOLD time series in different nodes in the networks, without taking time lags into account, for example, Structural Equation Modeling, SEM, (McIntosh & Gonzalez-Lima, 1994), Linear Non-Gaussian Acyclic Models, LiNGAM (Shimizu, Hoyer, Hyvärinen, & Kerminen, 2006) or Bayesian Nets, BNs (Frey & Jojic, 2005). On the other hand, we have state-space models which infer effective connectivity on the basis of the temporal patterns in the dynamics [e.g., Dynamic Causal Modeling, DCM (Friston, Harrison, & Penny, 2003), Granger Causality, GC, (Granger, 1969; Roebroeck, Formisano, & Goebel, 2011; Seth, Barrett, & Barnett, 2015; Solo, 2016), Transfer Entropy, TE (Lizier, Heinzle, Horstmann, Haynes, & Prokopenko, 2011; Vicente, Wibral, Lindner, & Pipa, 2011)]. There is an ongoing debate upon which class of models is better suited for this research problem (Valdes-Sosa, Roebroeck, Daunizeau, & Friston, 2011).

Understanding factors influencing performance of methods for effective connectivity research has been a subject to multiple computational studies. Previous work used the Dynamic Causal Modeling generative model (as the basis for the DCM inference procedure (Friston et al., 2003)) to create benchmark synthetic datasets (Smith et al., 2011). Multiple methods for assessing effective connectivity were tested on this synthetic data, including GC (Roebroeck, Formisano, & Goebel, 2005), Partial Directed Coherence (PDC; Baccalá & Sameshima, 2001), LiNGAM (Shimizu et al., 2006) or TE. In general,

however, the methods tested in the study did not perform much better than chance even though the testing networks were sparse and relatively small.

In this work, we employ a generative DCM forward model as implemented in (Smith et al., 2011). We then perform a simulation study in order to shed more light on the caveats of effective connectivity studies in fMRI datasets. On the basis of these simulations, we propose how certain problems can be overcome by a proper data preprocessing and region definition.

In Section Materials and Methods: The generative model, we introduce the DCM generative model. In Sections Materials and Methods: Impact of the noise and the length of the signal on identifiability of causal structures in fMRI and Materials and Methods: Influence of hemodynamics on the inference of effective connectivity, we describe in detail how we set up networks in order to perform the Linear Discriminant Analysis (LDA) study and to compute lagged cross correlations, respectively. In the Results section, we present the results and in the Discussion, we discuss these results and their practical implications on the effective connectivity research in fMRI.

2 | MATERIALS AND METHODS

2.1 | The generative model

Over the past decade, multiple generative models have been proposed in the context of the DCM (Friston, Kahan, Biswal, & Razi, 2011; Friston et al., 2003; Havlicek et al., 2015; Kiebel, Klöppel, Weiskopf, & Friston, 2007; Li et al., 2011; Marreiros, Kiebel, & Friston, 2008; Seth, Chorley, & Barnett, 2013; Smith et al., 2011; Stephan, Weiskopf, Drysdale, Robinson, & Friston, 2007; Stephan et al., 2008). In this study, we chose the original, single-node per region DCM (Friston et al., 2003; Smith et al., 2011). This model operationalizes the generation of BOLD response from the neuronal networks across two levels: nonobservable neuronal level and the observable hemodynamic level. The latent neuronal dynamics is described by the simple differential relationship:

$$\frac{d\vec{z}(t)}{dt} = A\vec{z}(t - \tau) + C\vec{u}(t) + \vec{\sigma}(t) \quad (1)$$

where $\vec{z}(t)$ denotes the temporary activity across all nodes, $\vec{u}(t)$ denotes binary inputs (trains of on- and off- states in our case), A denotes the

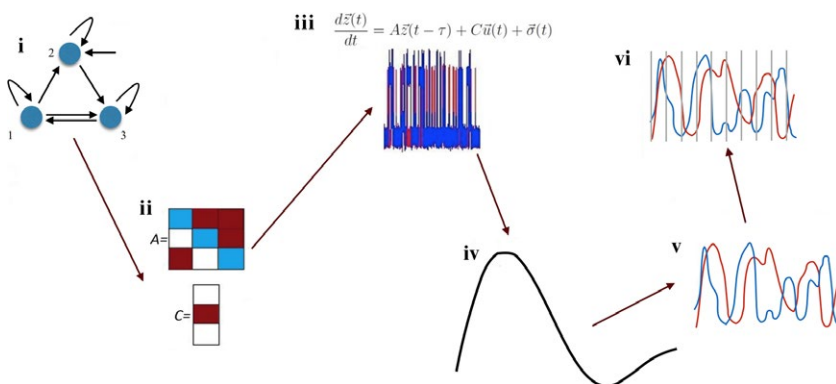


FIGURE 1 The full pipeline for the Dynamic Causal Modeling forward model. The full parameter set for the network (i) includes adjacency matrix (A) and inputs to the nodes (C) (ii). The neuronal dynamics is generated from this network with use of ordinary differential equations (iii). The neuronal time series is then convolved with the hemodynamic response function (iv) to obtain the BOLD response (v), which may be then (vi) subsampled

adjacency matrix of effective connectivity and C denotes connections from (experimental) inputs to the nodes, τ denotes a lag in the neuronal communication, and $\sigma(t)$ denotes the level of stochasticity on the neuronal level. [Correction added on 4 August 2017, after first online publication: in Equations 1 and 2, all delta symbol “ δ ” were changed to letter “ d ”; and both equations have been reformatted to reflect the correct formulas.] In our network setup, the modulatory connectivity does not play a role for the research question, therefore we set all the modulatory connections B from the original DCM model (Friston et al., 2003) to zero. The connectivity (a.k.a. adjacency) matrix A contains self-inhibition in every node (Figure 1i) as originally proposed in (Friston et al., 2003). Additionally, we use small, biologically plausible time lags of 50 ms in the communication between areas throughout all simulations in our study (as also implemented in Smith et al., 2011).

Therefore the simulated network becomes a system of delayed differential equations in fact (Bocharov & Rihan, 2000). Furthermore, effective connectivity matrix A must fulfill a few additional conditions, listed in the Appendix *Constraints on the adjacency matrix*.

In this context, the stochastic term $\sigma(t)$ represents neuronal innovations which are not the part of the communication between nodes of the investigated network (Daunizeau, Stephan, & Friston, 2012). It can either represent intrinsic dynamics in the given node (other than inhibition), or input from areas outside the investigated network. Strictly speaking, these innovations are not a “noise” (which would mean stochasticity added to the neuronal time series on the top of the simulated dynamics), but rather a background neuronal dynamics which cannot be explained by the given model. However, for the sake of simplicity we will refer to $\sigma(t)$ as noise in the text below.

In this study, we use two versions of the noise: white and scale-free, pink noise. Pink noise was generated from white noise by applying a Fourier transform, rescaling the spectrum so the power spectral density is proportional to the frequency by factor $1/f$, and subsequently, applying inverse Fourier transform and normalization.

The observational level is given by the classic model for the hemodynamic response, referred to as Balloon–Windkessel model (Buxton, Wong, & Frank, 1998; Friston et al., 2003), is described node-wide, and for every node i it is described by the dynamics of four biophysiological variables as follows:

$$\begin{aligned} \frac{ds_i(t)}{dt} &= z_i(t) - \kappa_i s_i(t) - \gamma_i (f_i(t) - 1) \\ \frac{df_i(t)}{dt} &= s_i(t) \\ \lambda_i \frac{dv_i(t)}{dt} &= f_i(t) - v_i^{\frac{1}{\alpha}}(t) \\ \lambda_i \frac{dq_i(t)}{dt} &= f_i(t) \frac{E(f_i(t), \rho_i)}{\rho_i} - v_i^{\frac{1}{\alpha}-1}(t) q_i(t) \end{aligned} \tag{2}$$

where $s_i(t)$ —vasodilatory signal, $f_i(t)$ —inflow, $v_i(t)$ —blood volume, $q_i(t)$ —deoxyhemoglobin content, $E(f, \rho) = 1 - (1 - \rho)^{1/f}$. The model involves five node-specific constants: κ —rate of signal decay, γ —rate of flow-dependent elimination, λ —hemodynamic transit time, α —Grubb’s exponent, ρ —resting oxygen extraction fraction. Then, the following expression describes the outcome BOLD response:

$$\gamma(t) = V_0(7\rho_i(1 - q_i(t)) + 2(1 - q_i(t)/v_i(t)) + (2\rho_i - 0.2)(1 - v_i(t))) \tag{3}$$

where $V_0 = 0.02$ denotes the resting blood volume fraction. Inputs to the network were simulated as in (Smith et al., 2011): as independent trains of on- and off-states with time resolution of $TR = 5$ ms. The probability of state switches was governed by a Poissonian process of a mean on-state duration of 2.5 s, and a mean off-state duration of 7.5 s.

In order to simulate the natural variability in the human hemodynamics, we sampled the parameters independently for each node, and from the distributions described previously in (Friston et al., 2003). Since this work concerns the effects of the neuronal noise, we did *not* add the thermal noise to the hemodynamic response as it refers to the quality of the scanning.

The simulations of the DCM generative model were performed with a step of 5 ms. At the end of the modeling pipeline, we subsampled the BOLD in order to emulate true restrictions of the fMRI datasets. The TR was one of the parameters in our study, with which we also controlled the length of the signal. In the typical fMRI experiments, the range of TR is 0.7 – 3.0 s, but we made a step beyond this range in order to better understand the influence of the TR on presence of networks signatures in the dynamics, and reduced it down to 0.10 s. The whole pipeline for the data generation with use of the DCM forward model is presented in Figure 1.

2.2 | Impact of the noise and the length of the signal on identifiability of causal structures in fMRI

In order to investigate under what conditions the problem of effective connectivity research becomes ill-posed, we fixed a test network and perturbed the connectivity within this network in order to investigate under what circumstances this perturbation yields detectable effects in the outcome BOLD response. Namely, if the two networks of distinct connectivity patterns yield indistinguishable BOLD response, the effective connectivity problem is ill-posed.

In this work, we restricted ourselves to network design comparable to networks investigated in typical fMRI DCM papers. These studies typically involve comparing between small, literature-informed models of 3–4 nodes. We also chose for simple Directed Acyclic Graphs (DAGs; Thulasiraman & Swamy, 1992) presented in Figure 2. This architecture facilitated following the feed-forward distribution of information throughout the hierarchical connectivity pattern. In Figure 2, the three connectivity patterns proposed in this study are presented. The original network (N_1) contains the projections 1→2, 1→3, 2→3, 3→4. We perturbed this original connectivity pattern in two ways:

1. “flip”: exchanging the connection 2→3 into 3→2 (Figure 2, N_2)
2. “split”: substituting the connection 3→4 into two connections 1→4 and 2→4 (Figure 2, N_3)

Here, we fixed the connectivity strength to 0.15 which refers to connection strengths typically found in the DCM studies (Li et al., 2014; Volza, Eickhoff, Pool, Fink, & Grefkes, 2015).

In this study, we concentrated on the neuronal noise, or “innovations” $z\sigma(t)$: other neuronal activity within a node which is either related to intrinsic dynamics of the brain region represented by that node, or to inflow of activity from other regions lying outside this particular

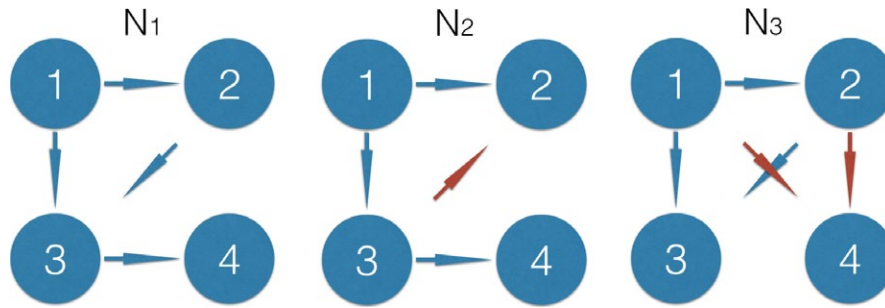


FIGURE 2 The test network and its two perturbations. The original network (N_1) contains projections $1 \rightarrow 2$, $1 \rightarrow 3$, $2 \rightarrow 3$, $3 \rightarrow 4$. The connectivity flip involves exchanging the connection $2 \rightarrow 3$ into $3 \rightarrow 2$ (N_2 , red). The connectivity split involves substituting the connection $3 \rightarrow 4$ into two weaker connections $1 \rightarrow 4$ and $2 \rightarrow 4$ (N_3 , red). This is an extension of the two-node setup presented in Figure 3 to four nodes (inputs and background noise are skipped from the picture for simplicity). The dynamics generated from these networks is presented in Supplementary Material The dynamics for the four-node DAG with perturbations

network. In a previous study by (Smith et al., 2011), this noise was set to very low levels (namely, to the Gaussian white noise of a standard deviation equal to 0.05 of the high input value), whereas in our study, the magnitude of this noise is the variable of interest. We varied this parameter in the simulations between 0.01 and 2 times the value of the high input state (which corresponds to signal-to-noise ratio, SNR = 100 and SNR = 0.50, respectively). Also, as we were interested in quantifying the strength of the effect of mixed signals in the network on the effective connectivity research, we emulated mixed signals with scale-free, pink noise in the neuronal communication $\sigma(t)$, and compared against the usual Gaussian white noise of the same magnitude. Since the DCM generative model was stochastic in our study, we performed 500 instantiations of the network dynamics for each parameter set.

In order to find out under what circumstances one can properly distinguish the original and perturbed networks, we performed a LDA study (Fisher, 1936). In LDA, a classifier supplied with a labeled training dataset, learns a linear combination of features that best separates the data into given classes. That performance can be then validated on a separate testing dataset. In our study, we first compressed the four-node time series into six pairwise Pearson correlations. We chose correlations as features because multiple methods for effective connectivity research utilize correlations between the time series. The classification performance was evaluated by cross validation: for each set of 500 instantiations of the networks, the data were randomly assigned to the training- (400 networks) and the testing set (100 networks), a 100 times. We then computed the mean performance and the standard deviation over all the random assignments.

Then, we investigated the impact of the magnitude and spectral properties of the noise on the neuronal level and the length of the time series, on the classification accuracy.

2.3 | Influence of hemodynamics on the inference of effective connectivity

Second, we studied the particular case of lagged methods for effective connectivity. Lagged methods such as GC, TE and other, new approaches (Hyvärinen, Shimizu, & Hoyer, 2008) assume that there is information preserved in the sequence of the BOLD samples.

Therefore, we were testing the limitations of the lagged methods for effective connectivity research with respect to the variability in the underlying dynamics and the properties of the local hemodynamic responses. With use of the DCM generative model (Section Materials and Methods: The generative model), we simulated the dynamics of a simple, two-node system (Figure 3a). Then, we adapted the lagged cross correlation (as proposed in El-Gohary & McNames, 2007), and proposed a simple quantity Δ based on asymmetry of the lagged cross correlation (Figure 3b), in order to quantify the amount of information preserved on effective connectivity contained in the *sequence* of the samples. Since in our simple model, node 2 received information from node 1 and the neuronal dynamics was delayed by 50 ms, in the absence of hemodynamic variability the BOLD time series in node 2 was delayed with respect to the BOLD time series in node 1. Let us now define correlation between the activity in node 1, $z_1(t)$, and activity in node 2 shifted one sample forward in time, $z_2(t + 1)$, as r_1 (Figure 3b, middle row). Similarly, let us define correlation between the activity in node 1, $z_1(t)$, and activity in node 2 shifted one sample backwards in time, $z_2(t - 1)$, as r_{-1} (Figure 3b, bottom row).

Therefore, the BOLD time series in node 2 shifted one sample forward in time $z_2(t + 1)$ (r_1 , middle row) correlated with the BOLD time series $z_1(t)$ *higher* than the BOLD time series in node 2 shifted one sample backwards in time $z_2(t - 1)$ (r_{-1} , bottom row). On the basis of this expected difference, we proposed the variable $\Delta = r_1 - r_{-1}$, and we expected this quantity to be positive for the connection $1 \rightarrow 2$. Now, since in the absence of hemodynamic variability between the upstream and the downstream node, we expected $r_1 > r_{-1}$, the value of Δ for the connection $1 \rightarrow 2$ should be positive. We then introduced the variability in the hemodynamic lags and investigated whether the positive sign of Δ still holds. We also investigated how the TR influences the performance of effective connectivity research with use of Δ , and whether or not further improving the TR to levels lower than 0.70 s as implemented in the state-of-the-art HCP data (van Essen et al., 2013) would improve the performance of the lagged based methods. Therefore, we compared the results for TR = 0.70 s with TR = 0.10 s.

Since in this part of the study, we were only focused on the variability in the hemodynamic lags as a confound to the effective connectivity

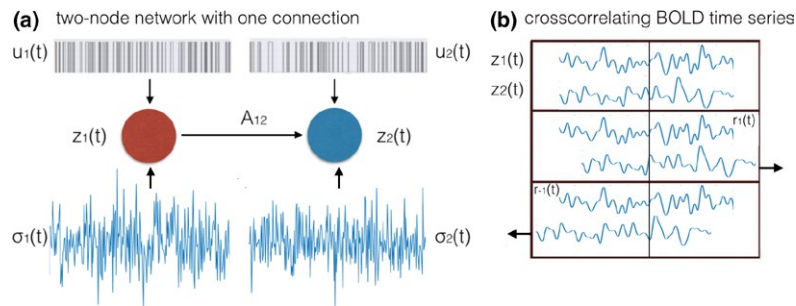


FIGURE 3 Defining Δ for the two-node noisy system. (a) The upstream node $[z_1(t)]$ is sending information to the downstream node $[z_2(t)]$ through a single connection of weight A_{12} . Both regions received a binary signal $u_i(t)$ and neuronal noise $\sigma_i(t)$. (b) Computing cross correlation between the two time series. Correlation between two time series can be computed without a lag (upper panel), with time series generated in the downstream region shifted one sample ahead in time (r_1 , middle panel), or backwards in time (r_{-1} , lower panel)

research, we have set the level of neuronal noise to low magnitudes in both nodes (standard deviation, STD = 0.01, white Gaussian noise), we have set the connectivity strength to a very high value of $w = 0.9$ and we performed a single but long simulation ($T = 3, 600$ s) of the neuronal dynamics of this two-node system. Then, we convolved the neuronal time series in the upstream node $[z_1(t)]$ with a fixed BOLD response (hemodynamic parameters at the mean of the distributions given in (Friston et al., 2003), which gives a hemodynamic lag of 3.14 s. We used 60,000 different BOLD responses to convolve the downstream neuronal time series $z_2(t)$ with. We then derived the Δ value for all pairs of BOLD time series while assuming TR = 0.70 s (which refers to the state-of-the-art Human Connectome Project datasets), and a very high time resolution of TR = 0.10 s for comparison.

3 | RESULTS

3.1 | Impact of the noise, the hemodynamics and the length of the signal on the presence of the network signatures in the BOLD

The results of the classification study with LDA are presented in Figure 4a,b. In panel a, we present the results for the flipped connection: the neuronal time series, BOLD and BOLD subsampled with two different TRs (3.0 s and 10.0 s), for the white noise (gray) and the pink noise (red) case. We present the classification accuracy in the y-axis against the SNR in the x-axis. In panel b, we present the results for the split connection, accordingly. In panel c, we present an example demonstrating the dependence of the LDA results on the TR and the number of samples in the time series. In panel d, we present the precision of the six pairwise correlations in the function of the TR and the number of samples.

First, we observe that the split of connection in the network is easier to detect than the flip. This result is intuitive because perturbing the network with a flip involves direct manipulation of one pairwise correlation (pair 2-3), whereas perturbing the network with a split involves direct manipulation of three pairwise correlations (pairs 1-4, 2-3, and 3-4).

Second, the dropout in the accuracy of LDA classifier in a function of SNR is similar for the neuronal time series and for the full length

BOLD time series. This result suggests that, while using the methods of effective connectivity that are based on correlations between signals, deconvolving BOLD into the neuronal time series might not be necessary to perform the inference. The hemodynamic response can even be beneficial to effective connectivity research as it removes noise-related higher frequencies while leaving experiment-related lower frequencies intact.

However, this effect does not hold for the scale-free pink noise: the hemodynamic convolution can no longer filter out the noise which has high power in the slow frequency range. Importantly, subsampling BOLD time series from very low time step used in the simulations of the generative model (5 ms, which gives a total of 120,000 samples for the 10 min of simulation) into the low TR = 3.0 s resolution (which gives only 200 samples) does not effectively blur the network signatures contained in the pairwise correlations, since LDA performs classification with almost identical accuracy for the BOLD time series subsampled with TR = 3.0 s as for the full-length BOLD time series obtained from convolving the fast neuronal time series (Figure 4, middle columns). This result can be explained by plotting the success rate of LDA classification for fixed set of parameters: the full BOLD time series, pink noise of SNR = 0.5, split, in the function of the TR (Figure 4c). In this case, TR = 3.0 s is the critical value of TR which gives performance comparable to the full BOLD time series. This TR relates to the total of 200 samples. For the TRs longer than 3.0 s, however, the performance of LDA gradually drops. This can be explained by two factors. First, the precision in (all six) pairwise correlations over 500 instantiations of the network in a function of TR (Figure 4d)—which is in fact the function of the length of the time series. All the six precision functions are constant until the TR reaches a critical value of TR = 3.0 s, and above that value, the precision in pairwise correlations drops. Second, TRs longer than 3.0 s give a sampling rate below the Nyquist frequency for the hemodynamic response, which also influences the amount of information contained in the time series. To sum up, given the small network of four nodes with connectivity weights in the range of 0.15, the BOLD time series of about 200 samples should be long enough to estimate pairwise correlations between the nodes, and as these are the features fed into the classifier, it is long enough to distinguish between different connectivity patterns.

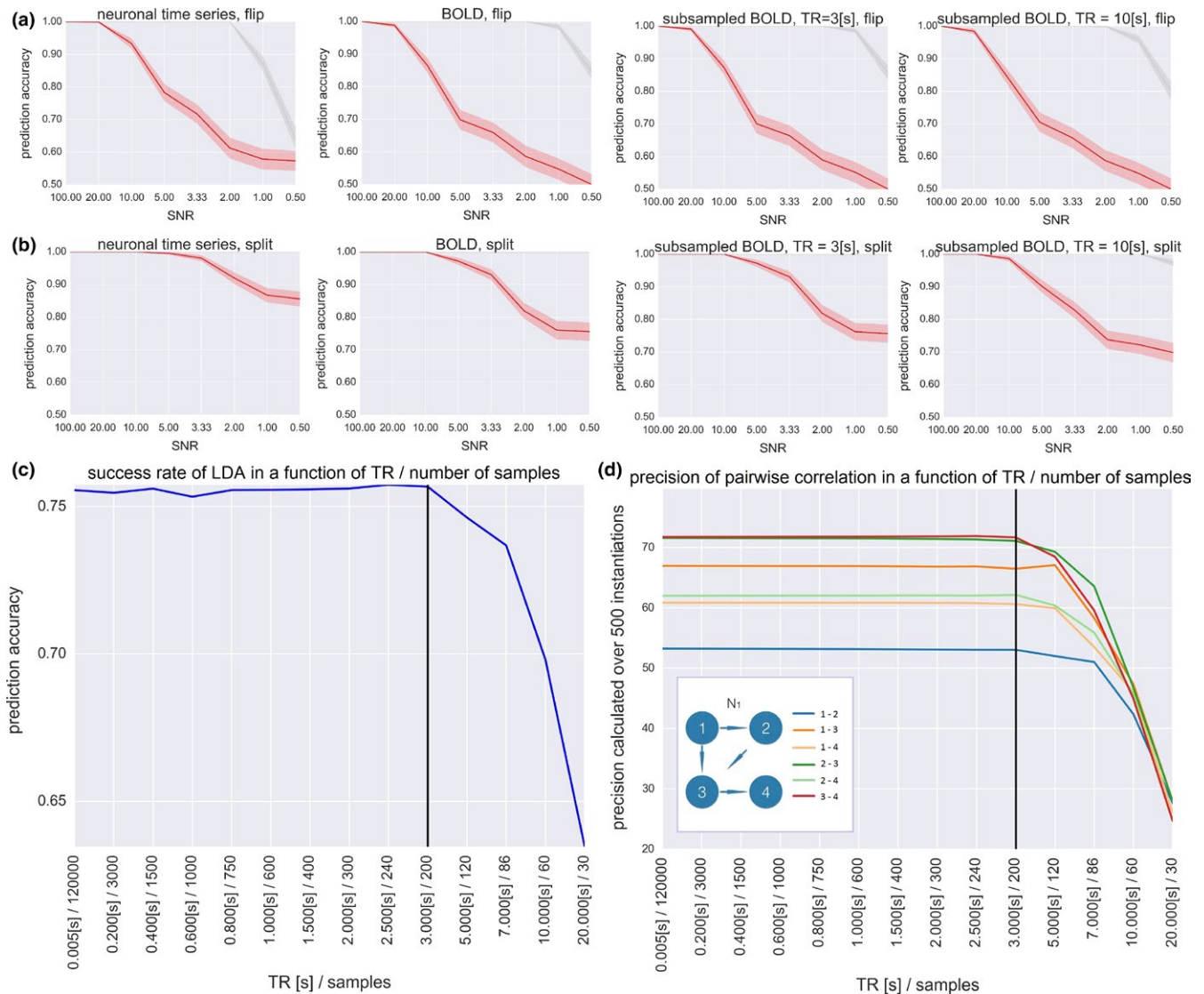


FIGURE 4 Results of the classification study with the Linear Discriminant Analysis. (a) Neuronal time series, BOLD and BOLD subsampled with two different TRs for the flipped connection. Gray line: white Gaussian noise in the neuronal communication. Red line: pink noise in the neuronal communication. Shaded error bars denote the standard deviation of the Linear Discriminant Analysis (LDA) results over 500 simulations. (b) The same for the split connection. (c) An example demonstrating the dependence of the LDA results on the TR and the number of samples, for fixed combination of parameters: full BOLD time series, pink noise of SNR = 0.5, split connection. TR = 3.0 s is the critical value of TR which gives performance comparable to the full BOLD time series. (d) Precision of the six pairwise correlations in the function of the TR and the number of samples, for fixed combination of parameters: the original network connectivity pattern N_1 , the full length BOLD time series, pink noise of SNR = 0.5. The precision in correlation decreases along with decreasing length of the time series

3.2 | Influence of hemodynamics on effective connectivity research

In Figure 5, we demonstrate the results of testing Δ on the two-node system with varying hemodynamic lags in the downstream region. Hemodynamic lags are operationalized as the time to peak of the hemodynamic response.

In Figure 5a, the histogram of Δ values over 500 instantiations of the neuronal dynamics of a two-node network with one connection (Figure 3) is presented. Indeed, all values are positive. In Figure 5b, values of Δ for all TRs between 50 ms and 3 s and over 500 instantiations of the neuronal dynamics are presented. As expected, both the

mean value and precision of Δ drop toward zero along with growing TR (black line). At the point of TR = 0.70 s (green line), 90% of Δ values are still positive (magenta curve). This means that in the absence of any hemodynamic lags and at the TR = 0.70 s, Δ would have empirical accuracy rate of 90%.

In Figure 5c, the results for the analysis on the hemodynamic level are presented: Δ is expressed as the function of the difference between hemodynamic lags in the upstream and the downstream region (referred to as a “relative lag”). The plot presents scatter plot over 60,000 different convolutions with hemodynamic parameters independently sampled from the distributions given in (Friston et al., 2003). Since the hemodynamic lag in the downstream node can be either higher or lower

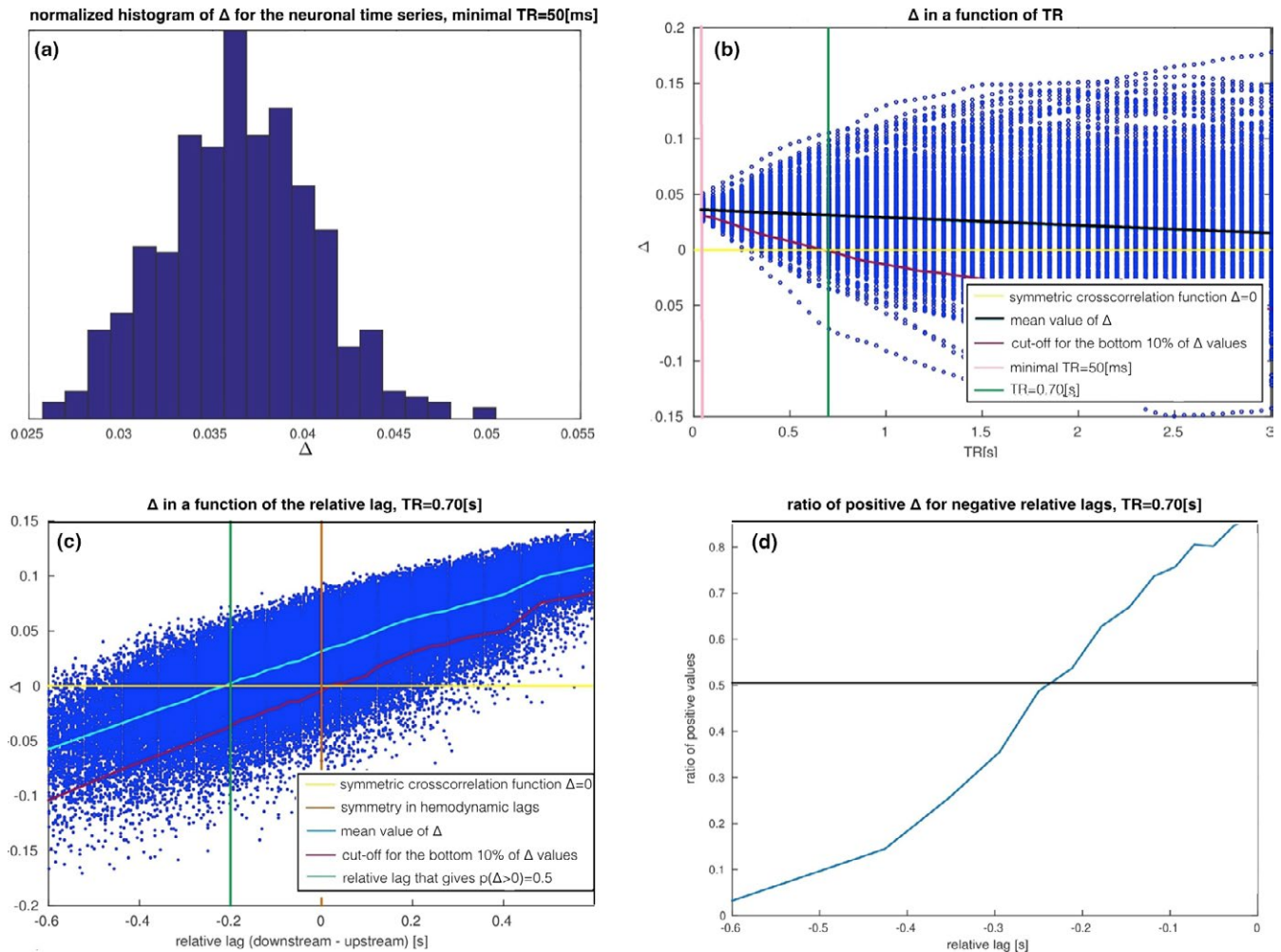


FIGURE 5 Impact of the relative hemodynamic lags computed as a difference between time to peak of the hemodynamic response in the upstream and downstream region on Δ . (a) The histogram of Δ values over 500 instantiations of the neuronal dynamics, for the optimal TR = 50 ms matching the true delay in the neuronal communication, gives only positive values as expected. (b) Δ over 500 instantiations of the neuronal dynamics and in a function of the TR. At TR = 0.70 s, the percentage of positive Δ values drops down to 90%, which is still high over the chance level. (c) Δ in the function of the *difference* between hemodynamic lags in the upstream and the downstream region (“relative lag”). (d) The relationship between the percentage of positive Δ values and the relative lags

than the reference hemodynamic lag in the upstream node, the relative lags have positive and negative values (equal lags are marked with orange line). We divided the whole set of relative lags into 30 intervals, and computed the mean for relative lags lying within each interval (Figure 5c, light blue curve). For each of the intervals, we also computed the cut-off for the bottom 10% of all the values (magenta curve).

First, for positive relative lags Δ becomes highly indicative for the connection as the vast majority of the Δ values are positive. This is because in this case, next to the neuronal dynamics in the downstream region following the dynamics in the upstream region, the slower hemodynamics further increases the lag in the downstream BOLD time series. The upper 90% of the Δ values (magenta curve) at the relative lag of zero (orange line) is positive (yellow line), which corresponds with the previous results for the underlying neuronal dynamics (Figure 5b).

The negative relative lags on the other hand make this statistic worse: the lower the value of the relative lag, the higher the percentage

of Δ values that become negative, the higher the chance of inferring wrong directionality of the link between the two nodes. Figure 5d demonstrates this relationship. For the critical value of the relative lag of around -0.22 s, the percentage of positive Δ values drops below 50%.

To conclude, we consider the impact of the TR on these results. In Figure 6a, we demonstrate a reproduction of Figure 5c but with a low TR = 0.10 s. Despite much faster temporal sampling, the results of the lag differences are almost identical, and the main difference being the range of absolute values of Δ (on the y axis). Therefore, the difference is quantitative rather than qualitative. This effectively means that at TR = 0.70 s, the hemodynamics is captured sufficiently well so that lowering TR further does not add any extra information that could be utilized with lagged methods for effective connectivity represented by Δ . This effect is explained in Figure 6b, where an exemplary cross-correlation function between the BOLD response between the upstream and the downstream region is presented (blue curve). Both

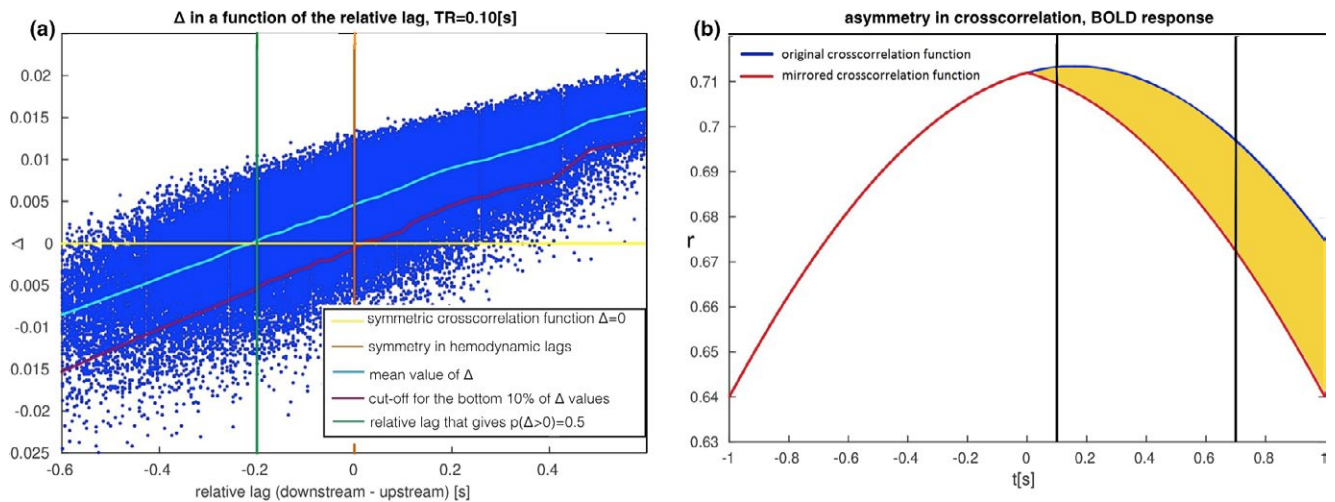


FIGURE 6 Impact of TR on Δ . (a) A reproduction of Figure 5c but with a small TR = 0.10 s. The absolute values of Δ change into narrower range than for TR = 0.70 s, but the sign of Δ stays the same. (b) Dependence of Δ on low TRs. The difference between cross-correlation function (blue curve) and its version mirrored around zero (red curve) for $t = TR$ equals Δ (yellow area). As we decrease TR from 0.70 s to 0.10 s (black lines), we can observe that TR has impact on the absolute value, but not on the sign of Δ

time series are convolved with the identical, canonical hemodynamic response. The cross-correlation peaks at 50 ms and then declines as expected. We mirrored the cross-correlation function around zero (red curve), and Δ can be computed as the difference between the original value of cross-correlation and its mirror counterpart (yellow area). TRs of 0.10 and 0.70 s are marked with vertical black lines. In this low range, the TR has an impact on the *absolute value* of Δ , but not on its sign.

4 | DISCUSSION

As a summary, our findings suggest that

1. In effective connectivity research with use of methods based on correlations, deconvolution of BOLD time series into neuronal time series might not be necessary. As opposed to the distribution-based methods, in lagged methods the variability in the local hemodynamic response can inverse the direction of a connection, therefore deconvolution of BOLD time series into the neuronal time series is recommended
2. Scale-free, pink noise induced by misparcellation is detrimental to the signatures of distinct connectivity patterns in the BOLD time series, therefore functional parcellation into ROIs (e.g., by Bayesian clustering or hierarchical ICA) is recommended
3. Decreasing the TR from 0.7 s (as is implemented in HCP datasets (van Essen et al., 2013), would not improve the performance of the lagged methods

Our LDA prediction study sheds the light on the factors influencing effective connectivity research in fMRI. One remark is that LDA does not indicate *how* to establish the causal relations between the nodes in the network: it just uses the distinct features in the data and therefore,

determines when it is theoretically possible. In our study, we chose pairwise correlations as features because Pearson correlation is nonparametric and in multiple methods for effective connectivity research in fMRI, it serves as the basis for the effective connectivity research (Hyvärinen & Smith, 2013; Patel, Bowman, & Rilling, 2006). Having that in mind, we compared the impact of the presence of slow hemodynamic response, short time series and mixed signals on the LDA results.

The results of our LDA study suggest that the *fixed* hemodynamics is not detrimental to the network signatures based on pairwise correlations: the LDA classifier distinguishes networks as easily on the basis on the neuronal time series as on the basis of the BOLD. Moreover, the hemodynamic response can work as a denoiser for the white noise (Figure 4a,b, gray lines). This result is intuitive once we take the spectrum of the white noise into account: a big portion of the power is carried within the high-frequency range which is easily denoised with the hemodynamic response. This result demonstrates that, in some cases (namely, when the signal has low frequency, and at the same time a good portion of the power of the noise is in the high-frequency range, and the hemodynamic response is fixed), the BOLD response is not necessarily a detrimental factor to the effective connectivity research. Therefore, according to our results, deconvolving the BOLD time series is not necessary while using methods for effective connectivity based on correlations.

We also obtain an intuitive result linking to the success rate of the LDA classifier to the precision in features estimation. Due to our results, at certain length of the BOLD time series, the precision in estimating pairwise correlations saturates, therefore further increasing the length of the time series does not improve the classification accuracy any further. This is a different characteristic from the relationship between the length of the time series and the accuracy of *signal detection*, where increasing the length of the time series tends to improve the performance, as it was found in a theoretical study by Murphy, Bodurka, & Bandettini (2007).

Furthermore, unlike slow hemodynamics and short time series, mixing signals represented in our simulations by the transition from white to scale-free (pink) noise, hinders the effective connectivity research in fMRI. For this reason, studying scale-free noise and its role for the inference with the DCM—as opposed to the Wiener process typically implemented in the stochastic DCM (Daunizeau et al., 2012)—is worth considering. Furthermore, the effect of mixed signals can to some extent be controlled. There are two sources of signal mixing which partly can be addressed by the researcher during data preprocessing and before applying any method for effective connectivity research to the fMRI data. The first source is associated with the background neuronal activity in the networks. In order to control for this possible confound effect, one can *partial out* the signal coming from outside the chosen pair of ROIs as an initial step to the effective connectivity study. The second source of the scale-free noise is a possible misparcellation into arbitrary brain areas during the preprocessing pipeline. From the perspective of network analysis, mixing signals of various frequencies is equivalent to inducing a pink noise in the underlying neuronal dynamics (Bak, Tang, & Wiesenfeld, 1987). Therefore, one should pursue efforts aiming at functional (as opposed to anatomical) segregation into ROIs. Most of the recent open-source fMRI datasets such as Human Connectome Project, ADHD-200 or ABIDE, already support the functional parcellation of the data (Craddock, James, Holtzheimer, Hu, & Mayberg, 2012; Glasser et al., 2016; Rosenberg et al., 2016). Multiple functional parcellations are available in the field of fMRI (Bellec et al., 2006; Bellec, Rosa-Neto, Lyttelton, Benali, & Evans, 2010; Blumensath et al., 2013; Chen et al., 2013; Craddock et al., 2012; Eickhoff et al., 2011; Flandin et al., 2002; Glasser et al., 2016; Golland, Golland, & Malach, 2007; Janssen, Jylänki, Kessels, & van Gerven, 2015; Janssen, Jylänki, & van Gerven, 2016; Kahnt, Chang, Park, Heinzle, & Haynes, 2012; Lashkari et al., 2012; Lashkari, Vul, Kanwisher, & Golland, 2010; Michel et al., 2012; Orban et al., 2014; Thirion et al., 2006; Tucholka et al., 2008; van den Heuvel, Mandl, & Pol, 2008; Yeo et al., 2011), and the issue of optimal functional parcellation is broadly discussed in the field (Stanley et al., 2013). In particular, in cognitive paradigms, the ROIs can be built in a data-driven way and on the basis of the patterns of activation only (task localizers, Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Heinzle, Wenzel, & Haynes, 2012).

Second, once we chose ROIs, the effect of mixing signals can be further reduced by a proper signal extraction from ROIs. One possibility is to consider only the first eigenvariate within the anatomical ROIs as proposed by Sato et al. (2010) (and as is implemented in the original version of the DCM inference procedure (Friston et al., 2003) instead of *averaging* activity over the voxels within an ROI).

Finally, our results suggest that once the effect of mixed signals is under control (so that only white noise is present in the neuronal dynamics), the signatures of distinct connectivity patterns are present in the BOLD time series even in a high noise regime. This may encourage further endeavors in the search for new markers of directed connections between brain regions from fMRI data.

In the second part of our study, we utilize lagged cross correlation in order to determine under what conditions the effective

connectivity-related information is preserved in the sequence of the samples and could be retrieved with the lagged methods. The main difference between our implementation of the hemodynamic variability and the previous studies is that we concentrate on the *time to peak* of the hemodynamic response as the representation of the hemodynamic lag, as opposed to the *time to onset* of the hemodynamic response as in (Smith et al., 2011). If the only difference between two hemodynamic responses was the time to onset, then a lag of 50 ms in the underlying neuronal communication would imply that the time to onset in the hemodynamic response in the upstream region 50 ms later than the time to onset in the hemodynamic response in the downstream region would automatically flip the sign of Δ and therefore also the outcome of the inference. However, this definition of the hemodynamic lags contains an assumption that hemodynamic response is equal to zero for the initial period of time. Such response profiles, however, are not biophysically plausible as in reality, the hemodynamic response has a positive derivative already at the start (Figure 1C, see Section Materials and Methods: The generative model for equations). In our study we employed the classic Balloon-Windkessel model (Buxton et al., 1998; Friston et al., 2003) in order to generate a natural distribution of the hemodynamic responses derived from neurophysiological experiments, and we operationalized hemodynamic lags as the *time to peak* instead. We found that within certain range of the relative hemodynamic lags, the asymmetry in the lagged cross correlation carries information about effective connectivity that can be derived from two BOLD time series.

In our setup, we use a fixed neuronal delay of 50 ms, whereas in another computational study on the influence of hemodynamic response on GC by (Witt & Meyerand, 2009), this variable was a parameter of interest. In our study, neuronal lags higher than 50 ms would increase the asymmetry of the Δ in a function of relative lags on behalf of the positive relative lags. However, we motivate this small neuronal lag by experimental evidence suggesting that the axonal transmission, as the slowest phase of the neuronal transmission, involves time delays in the range of dozens of milliseconds (Sabatini & Regehr, 1999).

In many applications, lagged methods for effective connectivity in fMRI are applied to the deconvolved BOLD time series, with an example of GC (David et al., 2008; Goodyear et al., 2016; Hutcherson et al., 2015; Ryali, Supekar, Chen, & Menon, 2011; Ryali et al., 2016; Sathian, Deshpande, & Stilla, 2013; Wheelock et al., 2014). However, as demonstrated in Figure 5b, the natural variability in the neuronal dynamics results with an upper bound on the accuracy of the lagged based methods: even assuming a perfect deconvolution (which is never the case in practice), which would allow for perfect retrieval of the neuronal time series from the BOLD time series, for the TR = 0.70 s, the accuracy rate of Δ would be <100% (for the parameter space we are exploring in this study, this accuracy would be on the level of 90%). This result suggests that it might be worth considering to perform the GC analysis voxel-wise, and to average the resulting GC over the whole ROI instead of computing the GC ROI-wise as is often done in the GC studies (Chen et al., 2017; Deshpande, Hu, Stilla, & Sathian, 2008; Goodyear et al., 2016; Regner et al., 2016; Yang et al., 2017). After voxel-wise application of GC, the result can be averaged over

all voxel-wise GC scores between the two regions. This voxel-wise application of GC can be performed in multiple ways: the GC score can be averaged with LASSO regularization (Tang, Bressler, Sylvester, Shulman, & Corbetta, 2012), through a multivoxel pattern-based causality mapping as proposed by Kim, Kim, Ahmad, & Park (2013), or through hierarchical clustering as proposed by Deshpande, LaConte, James, Peltier, & Hu (2009). This strategic twist is already being used in some studies (Katwal, Gore, Gatenby, & Rogers, 2013; Zhao et al., 2016), however, it is not the most common approach in the field yet [voxel-wise modeling is, however, increasingly popular for finding activation patterns in cognition from fMRI data (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016)]. Applying GC voxel-wise has two major advantages. First, computing the final value of GC as a mean value over GC values derived from a large number of voxels should reduce the natural inaccuracy of the lagged methods when applied to the stochastic neuronal dynamics (Figure 5b). Second, it neutralizes the bias coming from possible inaccuracies in the blind deconvolution algorithms.

We also discuss the influence of the local distribution of the hemodynamic lags present in the investigated networks on the performance of the lagged methods for effective connectivity (Figure 5c,d). This part of the analysis refers to the studies in which the lagged methods are applied to the *unconvolved* BOLD, which is often a practice in GC research (Chen et al., 2017; Regner et al., 2016; Zhao et al., 2016). In such case, the utility of lagged methods in fMRI research depends on the variability in the hemodynamic lags. According to our results, under the assumption that the lag in the neuronal communication is in the range of 50 ms, any inference with lagged methods is possible only if the mean hemodynamic lag between all the voxels within the upstream region are no more than 200 ms higher than the mean hemodynamic lag between all the voxels within the downstream region.

In the work by Handwerker, Ollinger, & D'Esposito (2004), it was found that the variability in the hemodynamic lags *between* subjects is higher than the variability *within* subjects. Furthermore, a study by Menon (2012) has demonstrated a large variation in hemodynamics depending on the vessel sizes in the voxels. Even vasculature changes across cortical layers have been shown to cause around 400 ms differences in hemodynamic time-to-peak (Silva & Koretsky, 2002).

This experimentally determined local variability in the hemodynamic lags is higher than the "safe range" found in our study (200 ms). Therefore, further neurophysiological studies on the variability in hemodynamic response across the human brain are necessary to advance our ability to characterize causal interactions from BOLD fMRI data, especially with use of the lagged methods such as GC.

Lastly, our results from this part of the study demonstrate that upsampling the BOLD time series to as little as $TR = 0.10$ s would *not* significantly improve the ability to retrieve the directionality of a connection with use of the lagged methods as the results differ from the results obtained for $TR = 0.70$ s only quantitatively, but not qualitatively. Our conclusions are concordant with recent theoretical considerations by Solo (2016), but differ from conclusions derived by Seth et al. (2013), Barnett & Seth (2017) and Lin et al. (2014), who conclude that GC estimation improves with upsampling. The design of our study is also different than the previous studies. The novelty

in our study is that we reduced the effective connectivity research with lagged methods to the simplest possible case. First, we quantify the amount of asymmetry in the lagged cross-correlation function between two time series, which is a model-free and probably the simplest lagged method for effective connectivity research (and is precise at picking on the time lag of the interaction, as demonstrated in Supplementary Material *Comparison between lagged cross correlation and Granger Causality*, Figure S1A). Second, we reduce the setup to the simplest, two-node neural mass model with a single connection, and with neuronal dynamics modeled with ordinary differential equations, as in the standard DCM generative model (Seth et al., 2013), by comparison, use much more complex design, involving a few nodes with more complex, spiking neuronal dynamics). Therefore, we believe that our results give insights into how, in general, a lagged method for rendering effective connectivity can be affected by the neuronal and hemodynamic variability, as well as by the TR.

One last remark is that in this study, we performed a simulation in order to approach the research questions posed in this article, as done in e.g. in Smith et al. (2011). One reason for this choice is that still, little is known about the effective connectivity structures in the human brain, as they are by definition dynamic and context dependent. Multiple ongoing projects aim at establishing effective connectivity in the brain on the basis of TMS and EEG (Esser, 2008) procedures, therefore extending our investigations to experimental fMRI data might be feasible in the future.

Altogether, our results suggest that neither the slow hemodynamics, a relatively short time series of a few hundred samples, nor TR as high as 0.7 s can affect the effective connectivity research in fMRI to a large extent. On the other hand, a proper region definition can facilitate the inference, therefore a good choice of the preprocessing pipeline improves chances for success in the effective connectivity study in fMRI.

ACKNOWLEDGMENTS

We thank Andre Marquand, Erik van Oort, and the rest of the SIN group at the Donders Institute for Brain, Cognition and Behavior, as well as Luca Ambrogioni, Jacob Billings and Foteini Protopapa, for the discussion and advice.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. JG has acted as a consultant to Boehringer Ingelheim in the last 3 years, but is not an employee or shareholder of this company.

AUTHOR CONTRIBUTIONS

NB, AL, and CB designed the study. NB conducted the simulations of the forward model. NB drafted the work. AL, CB, JB, and JG critically revised the work.

REFERENCES

- Bacalá, L. A., & Sameshima, K. (2001). Partial Directed Coherence: a new concept in neural structure determination. *Biological Cybernetics*, *84*(6), 463–474.
- Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: an explanation of $1/f$ noise. *Physical Review Letters*, *59*(4), 381–384. <https://doi.org/10.1103/PhysRevLett.59.381>
- Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68. <https://doi.org/10.1016/j.jneumeth.2013.10.018>
- Barnett, L., & Seth, A. (2017). Detectability of Granger causality for subsampled continuous-time neurophysiological processes. *Journal of Neuroscience Methods*, *275*, 93–121. <https://doi.org/10.1016/j.jneumeth.2016.10.016>
- Bellec, P., Perlbarg, V., Jbabdi, S., Pélérini-Issac, M., Anton, J. L., Doyon, J., & Benali, H. (2006). Identification of large-scale networks in the brain using fMRI. *NeuroImage*, *29*(4), 1231–1243. <https://doi.org/10.1016/j.neuroimage.2005.08.044>
- Bellec, P., Rosa-Neto, P., Lyttelton, O. C., Benali, H., & Evans, A. C. (2010). Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *NeuroImage*, *51*(3), 1126–1139. <https://doi.org/10.1016/j.neuroimage.2010.02.082>
- Bielczyk, N. Z., Buitelaar, J. K., Glennon, J. C., & Tiesinga, P. H. E. (2015). Circuit to construct mapping: a mathematical tool for assisting the diagnosis and treatment in major depressive disorder. *Frontiers in Psychiatry*, *6*(29), 2015. <https://doi.org/10.3389/fpsy.2015.00029>
- Blumensath, T., Jbabdi, S., Glasser, M. F., Essen, D. C. V., Ugurbil, K., Behrens, T. E., & Smith, S. M. (2013). Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *NeuroImage*, *76*, 313. <https://doi.org/10.1016/j.neuroimage.2013.03.024>
- Bocharov, G. A., & Rihan, F. A. (2000). Numerical modelling in biosciences using delay differential equations. *Journal of Computational and Applied Mathematics*, *125*(1-2), 183–199. [https://doi.org/10.1016/S0377-0427\(00\)00468-4](https://doi.org/10.1016/S0377-0427(00)00468-4)
- Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. *Magnetic Resonance in Medicine*, *39*(6), 855–864. <https://doi.org/10.1002/mrm.1910390602>
- Chen, H., Li, K., Zhu, D., Jiang, X., Yuan, Y., Lv, P., ... Liu, T. (2013). Inferring group-wise consistent multimodal brain networks via multi-view spectral clustering. *IEEE Transactions in Medical Imaging*, *32*(9), 1576–1586. <https://doi.org/10.1109/TMI.2013.2259248>
- Chen, Y. C., Xia, W., Chen, H., Feng, Y., Xu, J. J., Gu, J. P., ... Yin, X. (2017). Tinnitus distress is linked to enhanced resting-state functional connectivity from the limbic system to the auditory cortex. *Human Brain Mapping*, *38*(5), 2384–2397. <https://doi.org/10.1002/hbm.23525>
- Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, *33*, 1914–1928. <https://doi.org/10.1002/hbm.21333>
- Daunizeau, J., Stephan, K. E., & Friston, K. (2012). Stochastic dynamic causal modelling of fMRI data: Should we care about neural noise? *NeuroImage*, *62*(1), 464–481. <https://doi.org/10.1016/2012.04.061>
- David, O., Guillemain, I., Sallet, S., Rey, S., Deransart, C., Segebarth, C., & Depaulis, A. (2008). Identifying neural drivers with functional mri: An electrophysiological validation. *PLoS Biology*, *6*(12), e315. <https://doi.org/10.1371/journal.pbio.0060315>
- Deshpande, G., Hu, X., Stilla, R., & Sathian, K. (2008). Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *NeuroImage*, *40*(4), 1807. <https://doi.org/10.1016/j.neuroimage.2008.01.044>
- Deshpande, G., LaConte, S., James, G. A., Peltier, S., & Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, *30*(4), 1361–1373. <https://doi.org/10.1002/hbm.20606>
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *NeuroImage*, *57*(3), 938. <https://doi.org/10.1016/j.neuroimage.2011.05.021>
- El-Gohary, M., & McNames, J. (2007). Establishing causality with whitened cross-correlation analysis. *IEEE Transactions on Biomedical Engineering*, *54*(12), 2214–2222.
- van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T., Yacoub, E., Ugurbil, K., ... WU-Minn HCP Consortium. (2013). The Human Connectome Project: a data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. <https://doi.org/10.1016/j.neuroimage.2012.02.018>
- Esser, S. K. (2008). A modeling TMS-EEG approach to effective connectivity. <https://doi.org/grantome.com/grant/nih/f31-ns055553-03>
- Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. <https://doi.org/10.1152/jn.00032.2010>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., & Poline, J. B. (2002). Improved detection sensitivity in functional MRI data using a brain parcelling technique. *Medical Image Computing and Computer-Assisted Intervention*, *2488*, 467–474.
- Fornito, A., Zalewsky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*, 159–172. <https://doi.org/doi:10.1038/nrn3901>
- Frey, B. J., & Jojic, N. (2005). A comparison of algorithms for inference and learning in probabilistic Graphical Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(9), 1392–1416. <https://doi.org/10.1109/TPAMI.2005.169>
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13–36. <https://doi.org/10.1089/brain.2011.0008>
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic Causal Modeling. *NeuroImage*, *19*(4), 1273–1302. [https://doi.org/10.1016/S1053-8119\(03\)00202-7](https://doi.org/10.1016/S1053-8119(03)00202-7)
- Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2011). A DCM for resting state fMRI. *NeuroImage*, *94*, 396–4. <https://doi.org/10.1016/j.neuroimage.2013.12.009>
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. <https://doi.org/10.1038/nature18933>
- Golland, P., Golland, Y., & Malach, R. (2007). Detection of spatial activation patterns as unsupervised segmentation of fMRI data. *Medical Image Computing and Computer-Assisted Intervention*, *10*(1), 110–118. <https://doi.org/10.1007/978-3-540-75757-3>
- Goodyear, K., Parasuraman, R., Chernyak, S., Madhavan, P., Deshpande, G., & Krueger, F. (2016). Advice taking from humans and machines: An fMRI and effective connectivity study. *Frontiers in Human Neuroscience*, *4*(10), 542. <https://doi.org/10.3389/fnhum.2016.00542>
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. <https://doi.org/10.2307/1912791>
- Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. <https://doi.org/10.1016/j.neuroimage.2003.11.029>
- Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., & Uludag, K. (2015). Physiologically informed Dynamic Causal Modeling of fMRI data. *NeuroImage*, *122*(355–72), 2015. <https://doi.org/doi:10.1016/j.NeuroImage.07.078>
- Heinzle, J., Wenzel, M. A., & Haynes, J.-D. (2012). Visuomotor functional network topology predicts upcoming tasks. *Journal of Neuroscience*,

- 32(29), 9960–9968. <https://doi.org/10.1523/JNEUROSCI.1604-12.2012>
- van den Heuvel, M., Mandl, R., & Pol, R. H. (2008). Normalized cut group clustering of resting-state fMRI data. *PLoS ONE*, 3(4), e2001. <https://doi.org/10.1016/j.neuroimage.2008.08.010>.
- van den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519–534. <https://doi.org/10.1016/j.euroneuro.2010.03.008>
- Hutcherson, N. L., Sreenivasan, K. R., Deshpande, G., Reid, M. A., Hadley, J., White, D. M., ... Lahti, A. C. (2015). Effective connectivity during episodic memory retrieval in schizophrenia participants before and after antipsychotic medication. *Human Brain Mapping*, 36(4), 1442–1457. <https://doi.org/10.1002/hbm.22714>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532, 453–458. <https://doi.org/10.1038/nature17637>
- Hyvärinen, A., Shimizu, S., & Hoyer, P. O. (2008). Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-gaussianity. ICML '08 Proceedings of the 25th international conference on Machine learning, 424–431. <https://doi.org/10.1145/1390156.1390210>
- Hyvärinen, A., & Smith, S. (2013). Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14(1), 111–152.
- Janssen, R. J., Jylänki, P., Kessels, R. P., & van Gerven, M. A. (2015). Probabilistic model-based functional parcellation reveals a robust, fine-grained subdivision of the striatum. *NeuroImage*, 105(3), 811–815. <https://doi.org/10.1016/j.neuroimage.2015.06.084>
- Janssen, R. J., Jylänki, P., & van Gerven, M. A. J. (2016). Let's not waste time: Using temporal information in clustered activity estimation with spatial adjacency restrictions (caesar) for parcellating fMRI data. *PLoS ONE*, 11(12), e0164703. <https://doi.org/10.1371/journal.pone.0164703>
- Kahnt, T., Chang, L. J., Park, S. Q., Heinze, J., & Haynes, J.-D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *Journal of Neuroscience*, 32, 6240–6250. <https://doi.org/10.1523/JNEUROSCI.0257-12.2012>
- Katwal, S. B., Gore, J. C., Gatenby, J. C., & Rogers, B. P. (2013). Measuring relative timings of brain activities using fMRI. *NeuroImage*, 66, 436–438. <https://doi.org/10.1016/j.neuroimage.2012.10.052>
- Kiebel, S. J., Kloppe, S., Weiskopf, N., & Friston, K. J. (2007). Dynamic Causal Modeling: a generative model of slice timing in fMRI. *NeuroImage*, 34(4), 1487–1496. <https://doi.org/10.1016/j.neuroimage.2006.10.026>
- Kim, E., Kim, D.-S., Ahmad, F., & Park, H. (2013). Pattern-based Granger causality mapping in fMRI. *Brain Connectivity*, 3(6), 569–577. <https://doi.org/10.1089/brain.2013.0148>
- Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., & Golland, P. (2012). Search for patterns of functional specificity in the brain: a nonparametric hierarchical Bayesian model for group fMRI data. *NeuroImage*, 59(2), 1348–1368. <https://doi.org/10.1016/j.neuroimage.2011.08.031>
- Lashkari, D., Vul, E., Kanwisher, N., & Golland, P. (2010). Discovering structure in the space of fMRI selectivity profiles. *NeuroImage*, 50(3), 1085–1098. <https://doi.org/10.1016/j.neuroimage.2009.12.106>
- Li, B., Friston, K. J., Liu, J., Liu, Y., Zhang, G., Cao, F., ... Hu, D. (2014). Impaired frontal-basal ganglia connectivity in adolescents with internet addiction. *Scientific Reports*, 4, 5027.
- Li, B., Piriz, J., Mirrione, M., Chung, C., Proulx, C. D., Schulz, D., ... Malinow, R. (2011). Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. *Nature*, 470(7335), 535–539. <https://doi.org/10.1038/nature09742>
- Lin, F. H., Ahveninen, J., Raij, T., Witzel, T., Chu, Y. H., Jääskeläinen, I. P., ... Belliveau, J. W. (2014). Increasing fMRI sampling rate improves Granger causality estimates. *PLoS ONE*, 9(6), e100319. <https://doi.org/10.1371/journal.pone.0100319>
- Lizier, J. T., Heinze, J., Horstmann, A., Haynes, J. D., & Prokopenko, M. (2011). Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of Computational Neuroscience*, 30(1), 85–107. <https://doi.org/10.1007/s10827-010-0271-2>
- Marreiros, A. C., Kiebel, S. J., & Friston, K. J. (2008). Dynamic causal modelling for fMRI: a two-state model. *NeuroImage*, 39(1), 269. <https://doi.org/10.1016/j.neuroimage.2007.08.019>
- McIntosh, A., & Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2, 2–22.
- Menon, R. S. (2012). Mental chronometry. *NeuroImage*, 62(2), 1068–1071. <https://doi.org/10.1016/j.neuroimage.2011.01.011>
- Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., & Thirion, B. (2012). A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, 45(6), 2041–2049. <https://doi.org/10.1016/j.patcog.2011.04.006>
- Murphy, K., Bodurka, J., & Bandettini, P. A. (2007). How long to scan? the relationship between fMRI temporal signal to noise and necessary scan duration. *NeuroImage*, 34(2), 565–574. <https://doi.org/10.1016/j.neuroimage.2006.09.032>
- Orban, P., Doyon, J., Petrides, M., Mennes, M., Hoge, R., & Bellec, P. (2014). The richness of task-evoked hemodynamic responses defines a pseudohierarchy of functionally meaningful brain networks. *Cerebral Cortex*, 25(9), 2658–2669. <https://doi.org/10.1093/cercor/bhu064>
- Patel, R., Bowman, F. D., & Rilling, J. (2006). A Bayesian approach to determining connectivity of the human brain. *Human Brain Mapping*, 27(3), 267–276. <https://doi.org/10.1002/hbm.20182>
- Regner, M. F., Saenz, N., Maharajh, K., Yamamoto, D. J., Mohl, B., Wylie, K., ... Tanabe, J. (2016). Top-down network effective connectivity in abstinent substance dependent individuals. *PLoS ONE*, 11(10), e0164818. <https://doi.org/10.1371/journal.pone.0164818>
- Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, 25(1), 230–242. <https://doi.org/10.1016/j.neuroimage.2004.11.017>
- Roebroeck, A., Formisano, E., & Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*, 58(2), 296–302. <https://doi.org/10.1016/j.neuroimage.2009.09.036>
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, 19, 165–171. <https://doi.org/10.1038/nn.4179>
- Ryali, S., Shih, Y. Y., Chen, T., Kochalka, J., Albaugh, D., Fang, Z., ... Menon, V. (2016). Combining optogenetic stimulation and fMRI to validate a multivariate dynamical systems model for estimating causal brain interactions. *NeuroImage*, 132, 398–405. <https://doi.org/10.1016/j.neuroimage.2016.02.067>
- Ryali, S., Supekar, K., Chen, T., & Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fMRI. *NeuroImage*, 54(2), 807–823. <https://doi.org/10.1016/j.neuroimage.2010.09.052>
- Sabatini, B. L., & Regehr, W. G. (1999). Timing of synaptic transmission. *Annual Reviews in Physiology*, 61, 521–542. <https://doi.org/10.1146/annurev.physiol.61.1.521>
- Sathian, K., Deshpande, G., & Stilla, R. (2013). Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. *Journal of Neuroscience*, 33(12), 5387–5398. <https://doi.org/10.1523/JNEUROSCI.3482-12.2013>
- Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., Brammer, M. J., & Amaro, E. (2010). Analyzing the connectivity between regions of interest: an approach based on cluster Granger causality for fMRI data analysis. *NeuroImage*, 52(4), 1444–1455. <https://doi.org/10.1016/j.neuroimage.2010.05.022>

- Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neurosci*, 35(8), 3293–3297. <https://doi.org/10.1523/jneurosci.4399-14.2015>.
- Seth, A. K., Chorley, P., & Barnett, L. C. (2013). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage*, 65, 540–555. <https://doi.org/10.1016/j.neuroimage.2012.09.049>
- Shimizu, S., Hoyer, P. O., Hyvärinen, A., & Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, 2003–2030.
- Silva, A. C., & Koretsky, A. P. (2002). Laminar specificity of functional MRI onset times during somatosensory stimulation in rat. *Proceedings of the National Academy of Sciences*, 99(23), 15182–15187. <https://doi.org/10.1073/pnas.222561899>
- Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., ... Woolrich, M. (2011). *Network modelling methods for fMRI*. *NeuroImage*, 54(2), 875–891. <https://doi.org/10.1016/j.NeuroImage.08.063>
- Solo, V. (2016). State-space analysis of Granger-geweke causality measures with application to fMRI. *Neural Computation*, 28(5), 914–949. https://doi.org/10.1162/NECO_a_00828
- Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, 17(5), 652–660. <https://doi.org/10.1038/nn.3690>
- Stanley, M. L., Moussa, M. N., Paolini, B. M., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2013). Defining nodes in complex brain networks. *Frontiers in Computational Neuroscience*, 7, 169. <https://doi.org/10.3389/fncom.2013.00169>
- Stephan, K. E., Kasper, L., Harrison, L., Daunizeau, J., van den Ouden, H. E. M., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, 42(2), 649–662. <https://doi.org/10.1016/j.neuroimage.2008.04.262>
- Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., & Friston, K. J. (2007). Comparing hemodynamic models with DCM. *NeuroImage*, 38(3), 387–401. <https://doi.org/10.1016/j.neuroimage.2007.07.040>
- Tang, W., Bressler, S. L., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2012). Measuring Granger causality between cortical regions from voxelwise fMRI BOLD signals with lasso. *PLoS Computational Biology*, 8(5), e1002513. <https://doi.org/10.1371/journal.pcbi.1002513>
- Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., & Poline, J.-P. (2006). Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. *Human Brain Mapping*, 27, 678–693. <https://doi.org/10.1002/hbm.20210>
- Thulasiraman, K., & Swamy, M. N. S. (1992). *Directed acyclic graphs. Graphs: theory and algorithms*. New York, NY: John Wiley and Sons.
- Tucholka, A., Thirion, B., Perrot, M., Pinel, P., Mangin, J. F., & Poline, J. B. (2008). Probabilistic anatomic-functional parcellation of the cortex: how many regions? *Medical Image Computing and Computer-Assisted Intervention*, 11(2), 399–406. https://doi.org/10.1007/978-3-540-85990-1_48
- Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., & Friston, K. (2011). Effective connectivity: influence, causality and biophysical modeling. *NeuroImage*, 58(2), 339–361. <https://doi.org/10.1016/j.neuroimage.2011.03.058>
- Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1), 45–67. <https://doi.org/10.1007/s10827-010-0262-3>
- Volza, L. J., Eickhoff, S. B., Pool, E.-M., Fink, G. R., & Grefkes, C. (2015). Differential modulation of motor network connectivity during movements of the upper and lower limbs. *NeuroImage*, 119, 44–53. <https://doi.org/10.1016/j.neuroimage.2015.05.101>
- Whelock, M. D., Sreenivasan, K. R., Wood, K. H., Hoef, L. W. V., Deshpande, G., & Knight, D. C. (2014). Threat-related learning relies on distinct dorsal prefrontal cortex network connectivity. *NeuroImage*, 102(2), 904–912. <https://doi.org/10.1016/j.neuroimage.2014.08.005>
- Witt, T., & Meyerand, M. E. (2009). The effects of computational method, data modeling, and TR on effective connectivity results. *Brain Imaging Behavior*, 3(2), 220–231. <https://doi.org/10.1007/s11682-009-9064-5>
- Yang, H., Wang, C., Zhang, Y., Xia, L., Feng, Z., Li, D., ... Wang, J. (2017). Disrupted causal connectivity anchored in the posterior cingulate cortex in amnesic mild cognitive impairment. *Frontiers in Neurology*, 8(10). <https://doi.org/10.3389/fneur.2017.00010>
- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165. <https://doi.org/10.1152/jn.00338.2011>
- Zhao, Z., Wang, X., Fan, M., Yin, D., Sun, L., Jia, J., ... Gong, J. (2016). Altered effective connectivity of the primary motor cortex in stroke: A resting-state fMRI study with Granger causality analysis. *PLoS ONE*, 11(11), e0166210. <https://doi.org/10.1371/journal.pone.0166210>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Bielczyk NZ, Llera A, Buitelaar JK, Glennon JC, Beckmann CF. The impact of hemodynamic variability and signal mixing on the identifiability of effective connectivity structures in BOLD fMRI. *Brain Behav*. 2017;7:e00777. <https://doi.org/10.1002/brb3.777>