Differential Covariance: A New Method to Estimate Functional Connectivity in fMRI

Tiger W. Lin  
wulin@ucsd.edu  
Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, 92092, and Computational Neurobiology Laboratory, Salk Institute for Biological Sciences, La Jolla, CA, 92037, U.S.A.

Yusi Chen  
cyusi@ucsd.edu  
Computational Neurobiology Laboratory, Salk Institute for Biological Sciences, La Jolla, CA, 92037, and Division of Biological Sciences, University of California San Diego, La Jolla, CA, 92092, U.S.A.

Qasim Bukhari  
qbukhari@mit.edu  
McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, U.S.A.

Giri P. Krishnan  
giri.prashanth@gmail.com  
Maxim Bazhenov  
maksimb@ucr.edu  
Department of Medicine, University of California San Diego, La Jolla, CA 92092, U.S.A.

Terrence J. Sejnowski  
sejnowski@salk.edu  
Computational Neurobiology Laboratory, Salk Institute for Biological Sciences, La Jolla, CA, 92037, U.S.A., and Institute for Neural Computation and Division of Biological Sciences, University of California San Diego, La Jolla, CA 92092, U.S.A.

Measuring functional connectivity from fMRI recordings is important in understanding processing in cortical networks. However, because the brain’s connection pattern is complex, currently used methods are prone to producing false functional connections. We introduce differential covariance analysis, a new method that uses derivatives of the signal for estimating functional connectivity. We generated neural activities from dynamical causal modeling and a neural network of Hodgkin-Huxley.
neurons and then converted them to hemodynamic signals using the forward balloon model. The simulated fMRI signals, together with the ground-truth connectivity pattern, were used to benchmark our method with other commonly used methods. Differential covariance achieved better results in complex network simulations. This new method opens an alternative way to estimate functional connectivity.

1 Introduction

Analysis of fMRI recordings to extract functional connectivity aims to characterize a pattern of statistical associations among time series measured from different brain regions (Reid et al., 2019). It has been used quite successfully in identifying brain network interactions at rest and during certain behaviors (Raichle, 2015). Moreover, interesting geometric properties of brain networks can be studied with this method (Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006).

Many statistical methods have been proposed for estimating functional connectivity, including covariance-based methods, lag-based methods (Seth, 2010; Geweke, 1984), and Bayes NET methods (Mumford & Ramsey, 2014). However, without knowing the ground truth, it has been difficult to assess the validity of these methods. To compare these methods, we simulated BOLD signals from network models with known patterns of connectivity (Friston, Harrison, & Penny, 2003). The simulated neural signals were then transformed to BOLD signals with an fMRI forward model using the nonlinear balloon model (Buxton, Wong, & Frank, 1998). In a previous study (Smith et al., 2011), many well-known statistical methods were benchmarked with simulated fMRI data sets and concluded that covariance-based methods performed best.

As Stevenson, Rebesco, Miller, and Körding (2008) reviewed, a fundamental problem for covariance-based methods is that two correlated nodes do not necessarily have a direct physical connection. This could be problematic when one wants to compare the functional connectivity with the anatomical wiring of the brain such as fiber tracts or neural projections. We introduce a new method to reduce these false connections and provide estimates that are closer to ground truth in simulated fMRI data. Our new method is a novel form of functional connectivity with weighted strength that extracts information about the directionality of causal influence. This method, differential covariance, performs better than other covariance-based methods on synthetic data by taking advantage of differential signals.

This article is organized as follows. In section 2 we explain how synthetic BOLD signals were generated and introduce our new method. We also explain how results from different estimators were quantified. In section 3, we compare the performance of our new method with the performance of
other methods. In section 4, we discuss the advantages and generalizability of the differential covariance.

2 Methods

This section describes in detail the methods of generating synthetic BOLD signals, estimating functional connectivity and performance quantification.

2.1 BOLD Signal Simulations.

2.1.1 Dynamic Causal Modeling. We used dynamic causal modeling (DCM) (Friston et al., 2003) and a more realistic thalamocortical model (details are explained in section 2.1.3) to simulate neural signals. Neural signals were then converted to BOLD signals (see Figure 1) using the classical forward balloon model with revised coefficients (Obata et al., 2004). Our simulated BOLD signals resemble ICA processed BOLD signals from real data (Laumann et al., 2015). The detailed DCM method was previously

![Simulated neural signal](image1.png)

![Simulated BOLD signal](image2.png)

![ICA-processed real BOLD signal](image3.png)

Figure 1: Neural signal and BOLD signal example.
explained in Smith et al. (2011). Briefly, the neural signals were generated from the model

\[ \dot{V} = AV + Cu, \]  

(2.1)

where \( V \) are time series of neural signals and \( A \) is the connection matrix. Following the parameters used in the Smith et al. (2011), the diagonal values of \( A \) are assigned to be \(-1\) in our simulations. Its nondiagonal values represent the connection pattern between nodes, which is the functional connectivity we try to estimate. The connection strength is uniformly distributed between 0.4 and 0.6. Figure 3A shows an example connection pattern represented by a 60 \( \times \) 60 matrix. Its first 50 nodes are visible, and they are connected with a sparse pattern. The 10 additional nodes are latent and have a broad connection to the visible nodes. We added the latent inputs to the system because unobserved common inputs exist in real-world problems (Stevenson et al., 2008). Variable \( u \) is the external stimuli. For our simulations, we assume there are no external stimuli to the network. However, a white gaussian noise with 0.1 variance is added to \( u \). \( C \), an identity matrix in this article, indicates that each node has its independent external stimuli.

Once the neural signals were simulated from the neural network model, they were fed into the Balloon model to generate BOLD signals. The model simulates the nonlinear vascular dynamic response due to the neural activities (Obata et al., 2004):

\[ \dot{s} = V - \kappa s - \gamma (f - 1), \]
\[ \dot{f} = s, \]
\[ \tau \dot{v} = f - v^{1/\alpha}, \]
\[ \tau \dot{q} = \frac{f}{E_0} E(f, E_o) - \frac{q}{V} v^{1/\alpha}, \quad E(f, E_o) = 1 - (1 - E_o)^{1/f}, \]
\[ y = V_o(k_1(1 - q) + k_2(1 - \frac{q}{V}) + k_3(1 - v)), \]
\[ k_1 = 4.3 \theta_o E_o TE, \]
\[ k_2 = \epsilon r_o E_o TE, \]
\[ k_3 = 1 - \epsilon, \]  

(2.2)

where the superscript dot stands for first-order derivative with respect to time, \( V \) is the neural signal, \( s \) is the vasodilatory signal, \( f \) is the blood inflow, \( v \) is blood volume, \( q \) is deoxyhemoglobin content, and TE is time of echo in seconds. The BOLD signal \( y \) is a combined signal of the blood volume and the deoxyhemoglobin content. Finally, the BOLD signal \( y \) is sampled at
1 sec. Values and notations of the constants and coefficients are provided in Table 1 in appendix B.

2.1.2 Benchmark Data Set from Previous Work. Based on DCM, using different connection patterns, a benchmark data set of 28 simulations was previously created (Smith et al., 2011). We slightly modified the simulation process to evaluate the performance of our method and previous methods over this popular benchmark data set.

First, the original binary external stimulus was replaced by a random gaussian input because our method cannot deal with sharp state transitions. However, our method is able to give a correct estimation in a network with gradual state transition, as in the case of the thalamocortical model simulations (see section 2.1.3).

Second, we reduced the sampling interval (TR) from 3 seconds to 1 second. This is a reasonable modification because an increasing number of fMRI studies (such as the Human Connectome Project) are adopting a densely sampled protocol with TR less than or equal to 1 second. Third, because the new method needs more data to provide a good estimation, simulation duration was increased from 600 seconds to 6000 seconds for most of the simulations. Fourth, the thermal noise added to the BOLD signal was changed from 1% to 0.1%. All other procedures in the original paper (Smith et al., 2011) were followed. Twenty-eight different connection patterns were first simulated using DCM and passed through the forward balloon model, as shown in equation 2.2.

2.1.3 Thalamocortical Model. To further validate our method, a more realistic Hodgkin-Huxley-based ionic neural network model was used in this article to generate neural signals, and the neural signals were then passed through the forward balloon model (see equation 2.2) to generate the BOLD signals. The thalamocortical model used in this study was based on several previous studies (Bazhenov, Timofeev, Steriade, & Sejnowski, 2002; Chen, Chauvette, Skorheim, Timofeev, & Bazhenov, 2012; Bonjean et al., 2011). It was originally used to model spindle and slow wave activity. The thalamocortical model was structured as a one-dimensional, multilayer array of cells. It consisted of 50 cortical pyramidal (PY) neurons, 10 cortical inhibitory (IN) neurons, 10 thalamic relay (TC) neurons, and 10 reticular (RE) neurons. The connection between the 50 PY neurons is shown in Figure 6A. For the rest of the connection types, a neuron connects to all target neurons within a predefined radius, as described in Bazhenov et al. (2002). The details of this model can be found in our previous publication (Lin, Das, Krishnan, Bazhenov, & Sejnowski, 2017).

For each simulation, we simulated the network for 4000 seconds with 600 seconds at each sleep stage (Awake → Stage 2 → Slow Wave → REM → Stage 2) and 200 seconds of transition period between stages. The simulated neural signals were transferred to BOLD signals and sampled at 10 Hz.


2.2 New Method for Functional Connectivity Estimation. In our new method, we first build a backward model to reconstruct the neural signals from the BOLD signals. Then we apply our differential covariance method to the reconstructed neural signals to estimate the functional connectivity.

2.2.1 Backward Model. To derive a backward model, we first linearized the high-order terms in the balloon model (see equation 2.2) according to Khalidov, Fadili, Lazeyras, Van De Ville, and Unser (2011). To be more specific, let \( \{x_1, x_2, x_3, x_4\} = \{s, 1 - f, 1 - v, 1 - q\} \), and linearize around the resting point \( \{x_1, x_2, x_3, x_4\} = \{0, 0, 0, 0\} \). Then we have

\[
\begin{align*}
\dot{x}_1 &= V - \kappa x_1 + \gamma x_2, \\
\dot{x}_2 &= -x_1, \\
\dot{x}_3 &= \frac{1}{\tau} x_2 - \frac{1}{\tau \alpha} x_3, \\
\dot{x}_4 &= cx_2 - \frac{1 - \alpha}{\alpha \tau} x_3 - \frac{1}{\tau} x_4, \\
c &= \frac{1}{\tau} + \frac{1 - E_o}{E_o \tau} \ln(1 - E_o), \\
y &= V_o((k_1 + k_2)x_4 + (k_3 - k_2)x_3).
\end{align*}
\]

After rearrangement, we have

\[
q_1 \frac{dV}{dt} + q_0 V = p_4 \frac{dy^4}{dt^4} + p_3 \frac{dy^3}{dt^3} + p_2 \frac{dy^2}{dt^2} + p_1 \frac{dy}{dt} + p_0 y,
\]

where constants and coefficients are defined in appendix C. Equation 2.4 shows that if we compute the high-order derivatives of \( y \), we can reconstruct the original neuronal signal \( V \) as a linear combination of its first-order derivative and itself. We denote this reconstructed neuronal signal as \( z \);

\[
z = q_1 \frac{dV}{dt} + q_0 V.
\]

2.2.2 Differential Covariance. In our previous work (Lin et al., 2017), we demonstrated that when derivative signals are used, the differential covariance method can reduce the false functional connections in a simulated network of neuronal activities. Here we briefly explain this method.

First, we assume that using the backward model above, we have reconstructed the neural signals \( z(t) \) of some fMRI nodes. \( z(t) \) is an \( N \times M \) matrix, where \( N \) is the number of nodes recorded in the network and \( M \) is the number of time points during the simulation. We compute the derivative of each time series by calculating \( dz(t) = \frac{z(t+1) - z(t-1)}{2\Delta t} \). Then the covariance between \( z(t) \) and \( dz(t) \) is computed and denoted as \( \Delta_C \), which is a \( N \times N \) matrix defined as the

\[
\Delta_C = \text{cov}(z(t), dz(t)).
\]
\[
\Delta_{C_{i,j}} = \text{cov}(dz_i(t), z_j(t)),
\]

(2.6)

where \( z_j(t) \) is the reconstructed neural signal of node \( j \), \( dz_i(t) \) is the derivative neural signal of node \( i \), and \( \text{cov} \) is the sample covariance function for two time series.

Note that different from our previous work (Lin et al., 2017), the differential covariance is computed using the reconstructed neural signal, \( z = q_1 \frac{dV}{dt} + q_0 V \), instead of the neural signal \( V \), which is not recoverable using the balloon model.

2.2.3 Partial Differential Covariance. As Stevenson et al. (2008) previously noted, one problem of the covariance method is the propagation of covariance. Here we designed a customized partial covariance algorithm to reduce this type of error in our differential covariance method. We use \( \Delta_{P_{i,j}} \) to denote the partial covariance between \( dz_i(t) \) and \( z_j(t) \).

Let \( Z \) be a set of all nodes except \( i \) and \( j \): \( Z = \{1, 2, \ldots, i-1, i+1, \ldots, j-1, j+1, \ldots, N\} \). Using the derivation from section A.1.2, we have

\[
\Delta_{P_{i,j}} = \Delta_{C_{i,j}} - \text{COV}_{j,Z} \cdot \text{COV}^{-1}_{Z,Z} \cdot \Delta_{C_{i,Z}},
\]

(2.7)

where \( \Delta_{C_{i,j}} \) and \( \Delta_{C_{i,Z}} \) were computed from the previous section and \( \text{COV} \) is the covariance matrix of the nodes.

2.2.4 Sparse Latent Regularization. Finally, we applied the sparse latent regularization method to our estimation (Chandrasekaran, Sanghavi, Parrilo, & Willsky, 2011; Yatsenko et al., 2015). As explained in section A.4, during recording, there are observable nodes and latent nodes in a network. If the connections between the observable nodes are sparse and the number of latent nodes is small, this method can separate the covariance matrix into these two parts, and one can regard the sparse matrix as the intrinsic connections between the observable nodes.

Here we define \( \Delta_S \) as the sparse intrinsic connections from the observed nodes and \( L \) as the residual covariance part introduced from the latent inputs. Then, by solving

\[
\arg\min_{\Delta_S} \| \Delta_S \|_1 + \alpha \cdot \text{tr}(L)
\]

(2.8)

under the constraint that

\[
\Delta_P = \Delta_S + L,
\]

(2.9)

we retrieve the intrinsic connections between the nodes measured, where \( \| \|_1 \) is the element-wise L1-norm of a matrix and \( \text{tr}(\cdot) \) is the trace of a matrix. \( \alpha \) is the penalty ratio between the L1-norm of \( \Delta_S \) and the trace of \( L \). It was
Figure 2: The three types of false connections in the covariance-based methods. Solid lines indicate the physical wiring between neurons, and the black circles at the end indicate the synaptic contacts (i.e., the direction of the connections). The dotted lines are the false connections introduced by the covariance-based methods. (A) Confounder errors, which are due to two neurons receiving the same synaptic inputs. (B) Chain errors, which are due to the propagation of correlation. (C) Latent confounder errors, which are due to unobserved common inputs.

set to $\frac{1}{\sqrt{N}}$ for all our estimations. $\Delta_P$ is the partial differential covariance computed from section 2.2.3.

2.3 Performance Quantification. The performance of each method is judged by four quantified values. The first three values indicate the method’s abilities to reduce the three types of false connections (see Figure 2). The last one indicates the method’s ability to correctly estimate the true positive connections against all possible interference.

We define $G$ as the ground-truth connectivity matrix, where

$$G_{i,j} = \begin{cases} 
1, & \text{if node } i \text{ projects to node } j \text{ with excitatory input} \\
-1, & \text{if node } i \text{ projects to node } j \text{ with inhibitory input} \\
0, & \text{otherwise}
\end{cases} \quad (2.10)$$

Then we can use a three-dimensional tensor to represent the false connections caused by common inputs. For example, if node $j$ and node $k$ receive common input from node $i$,

$$M_{i,j,k} = \begin{cases} 
1, & \text{iff } G_{i,j} = 1 \text{ and } G_{i,k} = 1 \\
0, & \text{otherwise}
\end{cases}. \quad (2.11)$$

Therefore, we can compute a mask that labels all the false connections due to the confounder structure:

$$\text{Mask}^{(1)}_{j,k} = \sum_{i \in \{\text{observed nodes}\}} M_{i,j,k}. \quad (2.12)$$
For the chain structure (e.g., node $i$ projects to node $k$, then node $k$ projects to node $j$), the mask is defined as

$$\text{Mask}^{(2)}_{i,j} = \sum_{k \in \{\text{observed nodes}\}} G_{i,k} G_{k,j}, \quad \text{or} \quad \text{Mask}^{(2)} = G \cdot G. \quad (2.13)$$

For the false connections caused by latent confounder structure, the mask is defined as

$$\text{Mask}^{(3)}_{i,k} = \sum_{i \in \{\text{unobserved nodes}\}} M_{i,j,k}. \quad (2.14)$$

Finally, $\text{Mask}^{(4)}$ as the general false connection mask is defined as

$$\text{Mask}^{(4)}_{i,j} = \begin{cases} 1, & \text{if } G_{i,j} = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (2.15)$$

Given a certain mask $\text{Mask}^{(k)}$ ($k = 1, 2, 3, 4$) and ground-truth $G$, the true positive set ($\text{TP}_k$) and the false positive set ($\text{FP}_k$) are defined as

$$(i, j) \in \text{TP}_k, \text{if } G_{i,j} \neq 0 \text{ and } \text{Mask}^{(k)}_{i,j} = 0,$$

$$(i, j) \in \text{FP}_k, \text{if } \text{Mask}^{(k)}_{i,j} \neq 0 \text{ and } G_{i,j} = 0. \quad (2.16)$$

To quantify the performance of a functional connectivity estimation $\text{FC}$, we follow the convention of Smith et al. (2011) and use the c-sensitivity index defined in the paper. C-sensitivity is defined as the percentage of connections in $\text{TP}$ that have an FC value above the 95% quantile of the FC values in $\text{FP}$,

$$\text{c-sensitivity}_k = \frac{|\{(i, j) \in \text{TP}_k : |\text{FC}_{i,j}| > Q_{95\%}(\text{FP}_k)\}|}{|\{(i, j) \in \text{TP}_k\}|}, \quad (2.17)$$

where $(i, j)$ represents the connection from node $i$ to node $j$, $|\text{FC}_{i,j}|$ is the absolute value of the functional connectivity from node $i$ to node $j$, $Q_{95\%}(\text{FP}_k)$ represents the 95% quantile of the FC values in $\text{FP}_k$, and $|\{}|$ is the number of elements in a given set.

3 Results

3.1 False Connections in Covariance-Based Methods. It is a well-known problem that the commonly used covariance-based methods produce systematic false connections (Stevenson et al., 2008). Shown in Figure 3A is the ground truth of the connection in our DCM model (nodes 1...
Figure 3: DCM model’s connection ground truth and connectivity estimation from covariance-based methods. (A) Ground-truth connection matrix. Nodes 1 to 50 are visible nodes. Nodes 51 to 60 are latent nodes. (B) Estimation from the covariance method. (C) Estimation from the precision matrix method. (D) Estimation from the sparsely regularized precision matrix (ICOV) method. (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.
to 50 are the observable nodes). Figure 3B shows the results from the covariance method, Figure 3C is the precision matrix, and Figure 3D is the ICOV (sparsely regularized precision matrix). Compared to the ground truth, all of these methods produce extra false connections. Here we briefly review the cause of these three types of false connections. The detailed mechanisms have been well reviewed previously (Stevenson et al., 2008).

We define the diagonal strip of connections in the ground truth (Figure 3A) as the +6 to +10 diagonal lines, because they are 6 to 10 steps away from the diagonal line of the matrix.

Shown in Figure 2A, confounder errors are produced because two nodes receive the same input from another node. The same input that passes into the two nodes generates positive correlation between the two nodes. However, there is no physiological connection between these two nodes. For example, as shown in Figure 3A, node 1 projects to nodes 6 and 7 ($A_{16}, A_{17} > 0$), but nodes 6 and 7 are not connected. However, because they share the same input from node 1, in Figure 3B, there are positive connections on coordinates (6,7) and (7,6). Because this connection pattern is also applied to other nodes (node 2 projects to nodes 7 and 8, node 3 projects to nodes 8 and so forth), there are false connections on the ±1 to ±4 diagonal lines of Figure 3B.

Shown in Figure 2B, the chain errors are due to the propagation of the covariance. Because node $A$ connects to node $B$ and node $B$ connects to node $C$, the covariance method presents covariance between nodes $A$ and $C$, which do not have a physical connection. This phenomenon is shown in Figure 3B as the extra diagonal strips.

Shown in Figure 2C, the latent confounder errors are produced due to the common input passed into two nodes. However, in this case, they are from the unobserved latent inputs. For this DCM model, it is due to the inputs from the 10 latent nodes (nodes 51 to 60), as shown in Figure 3A. Because the latent nodes have broad connections to the observed nodes, they introduce a square box shape covariance pattern into Figure 3B. And this type of false connection remains in the precision matrix and ICOV estimations (see Figures 3C and 3D) as the yellow color false connections on the bottom-left corner and the up-right corner.

3.2 Estimation from Differential Covariance Method. In this section, we explain the estimation results from the differential covariance method. Comparing the ground-truth connections in Figure 4A with our final estimation in Figure 4D, we see that our method essentially transformed the connections in the ground truth into a map of sources and sinks in a network. An excitatory connection, $i \rightarrow j$, in our estimations has negative value for $\Delta S_{ij}$ and positive value for $\Delta S_{ji}$, which means the current is coming out of the source $i$ and goes into the sink $j$. We note that there is another ambiguous case: an inhibitory connection $j \rightarrow i$, which produces the same...
Figure 4: Differential covariance analysis of the DCM model. The color in panels B, C, D, F, G, and H indicates the direction of the connections. For element $A_{ij}$, warm color indicates $i$ is the sink, $j$ is the source (i.e., $i \leftarrow j$), and cool color indicates $j$ is the sink and $i$ is the source (i.e., $i \rightarrow j$). (A) Ground-truth connection matrix. (B) Estimation from the differential covariance method ($\Delta C$). (C) Estimation from the partial differential covariance ($\Delta P$). (D) Estimation from the sparse + latent regularized partial differential covariance ($\Delta S$). (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.
Differential Covariance

Figure 5: Performance quantification (c-sensitivity) of different methods with respect to their abilities to reduce the three types of false connections and their abilities to estimate the true positive connections using the dynamic causal modeling. The first three results (BOLD Cov, BOLD Precision, and BOLD ICOV) are from covariance-based methods (covariance matrix, precision matrix, and sparsely regularized precision matrix) directly applied to simulated BOLD signals. The next three results (Back Cov, Back Precision, and Back ICOV) are from covariance-based methods applied to backward model reconstructed neural signals. The last three results ($\Delta_C$, $\Delta_P$, and $\Delta_S$) are from differential covariance methods applied to backward model reconstructed neural signals.

results in our estimations. Our method cannot differentiate these two cases; instead, they indicate sources and sinks in a network.

3.2.1 The Differential Covariance Method Reduces Confounder Errors. Comparing Figures 3B and 4B, we see that confounder errors on $\pm 1$ to $\pm 4$ diagonal lines of Figure 3B are reduced in Figure 4B. This is quantified in Figure 5A, where the differential covariance method has a higher c-sensitivity score given confounder errors as the false positive set.

3.2.2 The Partial Differential Covariance Reduces Chain Errors. Second, as we can see in Figure 4B, there are propagation errors. By applying the partial
covariance method, we regress out the interfering terms. Thus, there are fewer chain errors. Each method’s performance for reducing chain errors is quantified in Figure 5B.

3.2.3 The Sparse + Latent Regularization Reduces Latent Confounder Errors. Third, we use the sparse latent regularization to remove the covariance introduced by the latent variables. As already noted, when the observable nodes’ connections are sparse and the number of latent inputs is small, covariance introduced by the latent inputs can be separated. As shown in Figure 4D, the external covariance in the background of Figure 4C is reduced, while the true positive connections and the directionality of the connection are maintained. Each method’s ability to reduce this type of error is quantified in Figure 5C, and each method’s overall performance at reducing all errors is quantified in Figure 5D.

3.3 Thalamocortical Model’s Results. In addition to the DCM simulations, we also benchmarked various methods with a more realistic Hodgkin-Huxley-based model.

The covariance-based methods are applied to both the raw BOLD signals and the reconstructed neural signals. The differential covariance-based methods are applied to both the raw BOLD signals and the reconstructed neural signals as well. Compared to the ground truth in Figure 6A, covariance-based estimations failed to show the true connections. Also, confounder errors are noticeable on the ±1 diagonal line of Figures 6C and 6D. We show later that applying the covariance-based methods to the reconstructed neural signals does provide better results; however, covariance-based methods still suffer from the three types of false connections.

However, by applying differential covariance to the reconstructed neural signals, we see in Figure 7B that $\Delta C$ provides directionality information of the connection. However, $\Delta C$ estimation is contaminated by connections introduced from the propagation of the signals and the latent inputs. Confounder errors are greatly reduced. In Figure 7C, the partial covariance reduces chain errors and produces $\Delta P$. Finally, in Figure 7D, the sparse + latent regularization is applied to reduce latent confounder errors.

In Figure 8, each estimator’s performance in reducing different types of false positive connections is quantified. As explained above, while covariance-based methods perform well at chain errors and latent confounder errors, a differential covariance method performs better at reducing confounder errors. Thus, it achieves better performance overall.

3.4 The Backward Model and the Differential Covariance Are Both Needed to Produce Correct Estimations. We have explained our newly developed backward model and the differential covariance method. Now we show that both steps are necessary to produce good estimation.
Figure 6: Connectivity estimation from covariance-based methods of synthetic BOLD signals from thalamocortical model. (A) Ground-truth connection of the PY neurons in the thalamocortical model. (B) Estimation from the covariance method. (C) Estimation from the precision matrix method. (D) Estimation from the sparsely regularized precision matrix (ICOV) method. (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.

The purpose of the backward model is to restore the dynamic form of neural signals, such that it is suitable for the differential covariance estimation. As shown in Figure 9, if we directly apply the differential covariance method to the raw BOLD signals, the estimator’s performance is poor.
Figure 7: Differential covariance analysis of the thalamocortical model’s reconstructed neural signals. The color in panels B, C, D, F, G, and H indicates the direction of the connections. For element $A_{ij}$, warm color indicates $i$ is the sink, $j$ is the source (i.e., $i \leftarrow j$), and cool color indicates $j$ is the sink, $i$ is the source (i.e., $i \rightarrow j$). (A) Ground-truth connection matrix. (B) Estimation from the differential covariance method. (C) Estimation from the partial differential covariance method. (D) Estimation from the sparse + latent regularized partial differential covariance method. (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.
Figure 8: Performance quantification (c-sensitivity) of different methods with respect to their abilities to reduce the three types of false connections and their abilities to estimate the true positive connections using the thalamocortical model. The first three results are from covariance-based methods directly applied to simulated BOLD signals. The next three results are from covariance-based methods applied to backward model reconstructed neural signals. The last three results are from differential covariance methods applied to backward model reconstructed neural signals.

The performance quantified using c-sensitivity was shown in Figure 8. The backward model is a better procedure (see appendix E) for reconstructing a neural signal compared to blind deconvolution (Wu et al., 2013), which is commonly used to deconvolute resting-state fMRI signals. In addition, the backward model is robust to parameter jittering because the same set of backward parameters could still reconstruct the signals simulated using forward balloon model with a wide range of parameters (see appendix F).

The differential covariance method is also needed to reduce the false connections. In Figure 10, estimations from the covariance-based methods applied to the reconstructed neural signals failed to reduce the false connections mentioned. Confounder errors that are on the ±1 and ±2 diagonal lines are apparent in Figures 10C and 10D. This comparison can also be seen from the quantifications on Figures 5 and 8. For both
Figure 9: Differential covariance method applied directly to simulated BOLD signals without applying the backward model. The color in panels B, C, D, E, F, G, and H indicates direction of the connections. For element $A_{ij}$, warm color indicates $i$ is the sink, $j$ is the source (i.e. $i \leftarrow j$), and cool color indicates $j$ is the sink, $i$ is the source (i.e. $i \rightarrow j$). (A) Ground-truth connection matrix. (B) Estimation from the differential covariance method. (C) Estimation from the partial differential covariance method. (D) Estimation from the sparse + latent regularized partial differential covariance method. (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.
Figure 10: Covariance-based methods applied to backward model reconstructed signals. (A) Ground-truth connection matrix. (B) Estimation from the covariance method. (C) Estimation from the precision matrix method. (D) Estimation from the sparsely regularized precision matrix (ICOV) method. (E) Zoom-in of panel C. (F) Zoom-in of panel D. (E) Zoom-in of panel A. (F) Zoom-in of panel B. (G) Zoom-in of panel C. (H) Zoom-in of panel D.
Figure 11: (A) Performances (c-sensitivity) of different methods on the thalamocortical model simulations. (B) Average performances across the 28 data sets from a previous study (Smith et al., 2011). The Bayes net column refers to the average performance of all Bayes net methods mentioned in section A.2. The method’s performance on individual data sets is provided in section A.5.

simulations, even using the same backward-processed signals, the differential covariance method performs better at reducing confounder errors than the covariance-based methods. While the covariance-based methods perform well at reducing chain errors and latent confounder errors, the differential covariance method’s performance is better overall. Therefore, the differential covariance method is needed to make better estimations after the backward reconstruction step.

3.5 Benchmark Using Previous Study’s Data Set. To show that our new method is generalizable outside of the two example simulations just given, we benchmarked various methods using a previously generated data set containing 28 BOLD signal simulations (Smith et al., 2011). Top performers mentioned in the previous study (Smith et al., 2011) (ICOV, Precision matrix, Bayes net, and Patel’s κ) are included as benchmarks. Shown in Figure 11A, for the thalamocortical ionic model, we see that the $\Delta S$ outperforms others. This is mainly because of the method’s better performance at reducing confounder errors.

For the 28 simulations from the previous study (Smith et al., 2011), the differential covariance method’s average performance is comparable to other top methods identified previously (Figure 11B). Performance on individual data sets is provided in section A.5. Because most of these 28 simulations only have five nodes and the connection matrix is not sparse, we
dropped the sparse + latent regularization step in our method, except for simulation 4, the only one that has a sparse connection matrix (50 nodes, about 70 connections).

4 Discussion

4.1 Main Results. One goal of this article was to reconstruct neural signals from BOLD signals. We used synthetic BOLD data from detailed neural models to benchmark a novel differential covariance method to estimate the underlying functional connectivity. Although the relationship between fMRI and neuronal activity is a well-studied topic (for reviews, see Heeger and Ress, 2002 and Logothetis, 2002), we found that a custom-designed backward model was needed for our differential covariance method to efficiently and validly recover the underlying ground truth (Figure 9). This new method not only reduced estimation errors made by previous methods but also provided information about the directionality of the connections.

For the DCM data set from a previous study (Smith et al., 2011), differential covariances had similar performance to the previously benchmarked best methods (covariance-based methods). Each step of the differential covariance method is specialized at reducing one type of false connection. However, since the presence of these three types of false connections are minimal in this data set our new method does not outperform the covariance-based methods. Further, because the size of the network is small, the connectivity matrix is not sparse, so we cannot take advantage of the sparse + latent regularization step in our method. We believe this will not be an issue when applying differential covariance method to real-world problems because the brain network has a much higher dimensionality and is sparser.

4.2 Caveats and Future Directions. We made several key assumptions that should be further tested and tuned in the future. Our approach depends on the neural signals reconstructed from the fMRI transfer function. Even though the fMRI transfer function we used (the balloon model) is based on the fundamental vesicular dynamics, whether high-quality neural signals can be reconstructed from the noisy real fMRI instruments should be further studied. The backward reconstruction process also had lower noise tolerance (see appendix D). In addition, the sparse and low-rank assumptions were critical to the cases we simulated. These assumptions should be validated in applications to fMRI recordings.

The next step is to evaluate the performance of differential covariance analysis on fMRI recordings and compare the results with other methods for measuring functional connectivity. Preliminary results from analyzing fMRI resting state data from humans and mice with this new method are promising and will be reported elsewhere. Estimating functional connectivity is of great importance in understanding the dynamical patterns of
brain activity and differential covariance analysis of fMRI recordings offers
an new way to identify causal interactions between brain region.

Appendix A: Previous Methods

In this appendix we review several top-performing methods benchmarked
in Smith et al. (2011).

A.1 Covariance-Based Methods. The well-used covariance-based
methods are used as a benchmark in this article because a prior work (Smith
et al., 2011) has shown its superior performance by applying to simulated
BOLD signal.

A.1.1 Covariance Method. The covariance matrix is defined as

\[
\text{COV}_{x,y} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y),
\]  

(A.1)

where \( x \) and \( y \) are two variables, \( \mu_x \) and \( \mu_y \) are their population mean.

A.1.2 Precision Matrix. The precision matrix is the inverse of the covari-
ance matrix:

\[
P = \text{COV}^{-1}.
\]  

(A.2)

It can be considered as one kind of partial correlation. Here we briefly re-
view this derivation because we use it to develop our new method. The
derivation here is based on and adapted from Cox and Wermuth (1996).

We begin by considering a pair of variables \((x, y)\) and remove the corre-
lation in them introduced from a control variable \( z \).

First, we define the covariance matrix as

\[
\text{COV}_{xyz} = \begin{bmatrix}
\sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\
\sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\
\sigma_{zx} & \sigma_{zy} & \sigma_{zz}
\end{bmatrix}.
\]  

(A.3)

By solving the linear regression problem

\[
w_x = \arg \min_{w} E(x - w \ast z)^2,
\]

\[
w_y = \arg \min_{w} E(y - w \ast z)^2,
\]  

(A.4)
we have
\[
\begin{align*}
w_x &= \sigma_{xz}\sigma_{zz}^{-1}, \\
w_y &= \sigma_{yz}\sigma_{zz}^{-1},
\end{align*}
\]
(A.5)
then we define the residual of \(x, y\) as
\[
\begin{align*}
r_x &= x - w_x \ast z \\
r_y &= y - w_y \ast z.
\end{align*}
\]
(A.6)
Therefore, the covariance of \(r_x, r_y\) is
\[
\text{COV}_{r_x, r_y} = \sigma_{xy} - \sigma_{xz} \ast \sigma_{zz}^{-1} \ast \sigma_{yz}.
\]
(A.7)

On the other hand, if we define the precision matrix as,
\[
P_{xyz} = \begin{bmatrix}
p_{xx} & p_{xy} & p_{xz} \\
p_{yx} & p_{yy} & p_{yz} \\
p_{zx} & p_{zy} & p_{zz}
\end{bmatrix},
\]
(A.8)
then, using Cramer’s rule, we have
\[
p_{xy} = -\frac{\begin{vmatrix}
\sigma_{xy} & \sigma_{xz} \\
\sigma_{yx} & \sigma_{zz}
\end{vmatrix}}{|\text{COV}_{xyz}|}.
\]
(A.9)
Therefore,
\[
p_{xy} = \frac{-\sigma_{zz}}{|\text{COV}_{xyz}|}(\sigma_{xy} - \sigma_{xz} \ast \sigma_{zz}^{-1} \ast \sigma_{yz}).
\]
(A.10)
So \(p_{xy}\) and \(\text{COV}_{r_x, r_y}\) differ by a ratio of \(\frac{-\sigma_{zz}}{|\text{COV}_{xyz}|}\).

A.1.3 Sparsely Regularized Precision Matrix. If the precision matrix is expected to be sparse, one way to efficiently and accurately estimate it, especially when the dimension is large, is to regularize the estimation using the Lasso method (Hsieh, Sustik, Dhillon, & Ravikumar, 2014; Hsieh, Sustik, Dhillon, Ravikumar, & Poldrack, 2013; Friedman, Hastie, & Tibshirani, 2008; Banerjee, Ghaoui, d’Aspremont, & Natsoulis, 2006),
\[
\arg\min_{S \succ 0}\{-\log \det(S) + tr(SC) + \lambda \ast \|S\|_1},
\]
(A.11)
where \(C\) is the sample covariance matrix.
We used the implementation (https://www.cs.ubc.ca/~schmidtm/Software/L1precision.html) mentioned in the previous work (Smith et al., 2011) to benchmark our new method.

A.2 Bayes-Net Method. A variety of the Bayes net estimation methods are implemented in the Tetrad toolbox (http://www.phil.cmu.edu/tetrad/). Following the convention of the prior work (Smith et al., 2011), we tested CCD, CPC, FCI, PC, and GES. PC (“Peter and Clark”) is a causal inference algorithm used to estimate the causal relationship in a directed acyclic graph (DAG; Meek, 1995). CPC (Conservative PC) was based on the PC algorithm, but produces fewer false connections (Ramsey, Zhang, & Spirtes, 2012). GES (Greedy Equivalence Search) is another implementation based on the PC algorithm, but it uses a score system to estimate the causal connections (Chickering, 2003; Ramsey et al., 2010). The FCI (Fast Causal Inference) algorithm is a different approach, which takes into account the presence of latent confounder and selection bias (Zhang, 2008). CCD (Cyclic Causal Discovery), unlike the other algorithms mentioned, allows cycles in the network for estimation (Richardson & Spirtes, 1996). The code used in this article is based on the Tetrad Toolbox and further adapted by Smith et al. (2011) for fMRI data.

A.3 Patel’s κ Method. Patel’s κ measurement assesses the relationship between pairs of distinct brain regions by comparing expected joint and marginal probabilities of elevated activity of voxel pairs through a Bayesian paradigm (Patel, Bowman, & Rilling, 2006). For this article we used the implementation generously provided by Smith et al. (2011).

A.4 Sparse Latent Regularization. Prior studies (Banerjee et al., 2006; Friedman et al., 2008) have shown that regularizations can provide better estimation if the ground-truth connection matrix has a known structure (e.g., sparse). For all data tested in this article, the sparse latent regularization (Yatsenko et al., 2015) worked best. For a fair comparison, we applied the sparse latent regularization to both the precision matrix method and our differential covariance method.

In the original sparse latent regularization method, people made the assumption that a larger precision matrix $S$ is the joint distribution of the $p$ observed neurons and $d$ latent neurons (Yatsenko et al., 2015), that is,

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

where $S_{11}$ corresponds to the observable neurons. If we can measure only the observable neurons, the partial correlation computed from the observed neural signals is
\[ C_{ob}^{-1} = S_{ob} = S_{11} - S_{12} \cdot S_{22}^{-1} \cdot S_{21} \]  
(A.12)

because the latent latent neurons as shown in equation A.12 introduce correlations into the measurable system. We denote this correlation introduced from the latent inputs as

\[ L = S_{12} \cdot S_{22}^{-1} \cdot S_{21}. \]  
(A.13)

If we can make the assumption that the connections between the visible neurons are sparse, that is, \( S_{11} \) is sparse, and the number of latent neurons is much smaller than the number of visible neurons, \( d \ll p \). Then prior work (Chandrasekaran et al., 2011) has shown that if \( S_{ob} \) is known, \( S_{11} \) is sparse enough and \( L \)'s rank is low enough (within the bound defined in Chandrasekaran et al., 2011), the solution of

\[ S_{11} - L = S_{ob} \]  
(A.14)

is uniquely defined and can be solved by the following convex optimization problem,

\[
\arg\min_{S_{11}, L} \| S_{11} \|_1 + \alpha \cdot tr(L),
\]

(A.15)

under the constraint that

\[ S_{ob} = S_{11} - L. \]  
(A.16)

Here, \( \| \cdot \|_1 \) is the L1-norm of a matrix, and \( tr() \) is the trace of a matrix. \( \alpha \) is the penalty ratio between the L1-norm of \( S_{11} \) and the trace of \( L \) and is set to \( 1/\sqrt{N} \) for all our estimations.

This method is used to regularize precision matrix. For our differential covariance estimation, we need to make small changes to the derivation. Note that if we assume that the neural signals of the latent neurons are known and let \( l \) be the indexes of these latent neurons, then from our section 2.2.3,

\[ \Delta S_{i,j} = \Delta p_{i,j} - \text{COV}_{j,l} \cdot \text{COV}_{l,l}^{-1} \cdot \Delta_{C_{i,j}}^T \]  
(A.17)

removes the \( V_{\text{latent}} \) terms.

Even if \( l \) is unknown,

\[ \text{COV}_{j,l} \cdot \text{COV}_{l,l}^{-1} \cdot \Delta_{C_{i,j}}^T \]
Figure 12: The first column is method performance (c-sensitivity) on our more realistic thalamocortical ionic model. The second column is method average performance across the 28 data sets from a previous study Smith et al. (2011). Columns 3 to 30 are method performances for each of the 28 data sets.

is low rank, because it is bounded by the dimensionality of $COV_{I,I}$, which is $d$. And $\Delta \mathbf{S}$ is the internal connections between the visible neurons, which should be a sparse matrix. Therefore, letting

$$S_{ob} = \Delta P,$$
$$S_{11} = \Delta \mathbf{S},$$
$$L = -COV_{j,l} \cdot COV_{I,I}^{-1} \cdot \Delta C_{c,l}^{T}, \quad (A.18)$$

we can use the original sparse + latent method to solve for $\Delta \mathbf{S}$. In this article, we used the inexact robust PCA algorithm (http://perception.csl.illinois.edu/matrix-rank/sample_code.html) to solve this problem (Lin, Liu, & Su, 2011).

**A.5 Methods’ Performance for Individual Data Sets.** In Figure 12, the performance of the different methods mentioned above is quantified using c-sensitivity on individual data sets simulated in Smith et al. (2011). For comparison, the methods’ performance on thalamocortical model and their average performance across all 28 data sets are shown in the first two columns.
Table 1: Balloon Model Parameters.

| Parameter | Description                  | Value       |
|-----------|------------------------------|-------------|
| $\kappa$  | Rate of signal decay         | 0.65 per s |
| $\gamma$ | Rate of flow-dependent elimination | 0.41 per s |
| $\tau$   | Hemodynamic transit time      | 0.98 per s |
| $\alpha$ | Grubbs exponent              | 0.32        |
| $E_o$     | Resting oxygen extraction fraction | 0.5        |
| $\epsilon$ | Ratio of intra- and extravascular signal | 0.5 |
| $\vartheta_o$ | Frequency offset at the outer surface of the magnetized vessel for fully deoxygenated blood | 40.3 per s |
| TE        | Time of echo                 | 40 ms       |

Table 2: Backward Reconstruction Parameters.

| Parameter | Description                  | Value       |
|-----------|------------------------------|-------------|
| $\kappa$  | Rate of signal decay         | 0.65 per s |
| $\gamma$ | Rate of flow-dependent elimination | 0.41 per s |
| $\tau$   | Hemodynamic transit time      | 0.98 per s |
| $\alpha$ | Grubbs exponent              | 0.32        |
| $E_o$     | Resting oxygen extraction fraction | 0.5        |
| $\epsilon$ | Ratio of intra- and extravascular signal | 0.5 |
| $\vartheta_o$ | Frequency offset at the outer surface of the magnetized vessel for fully deoxygenated blood | 40.3 per s |
| TE        | Time of echo                 | 40 ms       |

Appendix B: Balloon Model Parameters

The values and notations for parameters involved in the forward balloon model (see equation 2.2) are shown in Table 1. The values of the coefficients derived from these parameters are evaluated in equation B.1.

\[
\begin{align*}
  k_1 &= 4.3 \vartheta_o E_o TE = 3.47, \\
  k_2 &= \epsilon r_o E_o TE = 0.25, \\
  k_3 &= 1 - \epsilon = 0.5. \\
\end{align*}
\]

Appendix C: Backward Model Parameters

The values and notations for parameters involved in the backward reconstruction model (see equation 2.3) are shown in Table 2. The values of the coefficients (equation 2.4) derived from these parameters are evaluated in
equation C.1.

\[ k_1 = 4.3 \theta_o E_o TE = 3.47, \]
\[ k_2 = \epsilon \rho_o E_o TE = 0.25, \]
\[ k_3 = 1 - \epsilon = 0.5, \]
\[ p_4 = \tau = 0.98, \]
\[ p_3 = 1 + \frac{1}{\alpha} + \tau \kappa = 4.76, \]
\[ p_2 = \frac{1}{\tau \alpha} + \tau \gamma + \left(1 + \frac{1}{\alpha}\right) \kappa = 6.27, \]
\[ p_1 = \left(1 + \frac{1}{\alpha}\right) \gamma + \frac{\kappa}{\tau \alpha} = 3.75, \]
\[ p_0 = \frac{\gamma}{\tau \alpha} = 1.30, \]
\[ c = \frac{1}{\tau} + \frac{1 - E_o}{E_o \tau} \ln(1 - E_o) = 0.31, \]
\[ q_1 = V_o \left[-(k_1 + k_2) \tau c - (k_3 - k_2)\right] = -0.03, \]
\[ q_0 = V_o \left[-(k_1 + k_2) \frac{(\tau c - 1 + \alpha)}{\alpha \tau} - \frac{(k_3 - k_2)}{\tau}\right] = 0.08. \quad \text{(C.1)} \]

**Appendix D: Noise Tolerance**

When using the numerical differentiation method to compute the reconstructed signal in equation 2.4, we noticed that the method’s noise tolerance depends on the sampling frequency. For example, if white noise \( \mathcal{N}(0, \sigma^2) \) was introduced into BOLD signal \( y \) sampled at 10 Hz, the noise was magnified by 10 times for \( \frac{dy}{dt} = \frac{y(t+1) - y(t-1)}{2dt} \) because noise variance in the nominator is \( 2\sigma^2 \) and \( dt = 0.1 \). The observation was confirmed in Figure 13 where differential covariance analysis was applied to both a noisy signal (right panel) and a clean signal (left panel) after backward reconstruction. While we could still identify the overall connection structure from the noisy signal sampled at 1 Hz, the connection pattern was totally destroyed for the noisy signal sampled at 19 Hz.

**Appendix E: Benchmark with Blind Deconvolution**

Blind deconvolution aims to reduce the confounding effects of the hemodynamic response function (HRF) on resting-state fMRI signals. For
Figure 13: Noise tolerance of the method. BOLD signals were simulated from the thalamocortical model. (A) Left panel: $\Delta S$ estimated from the clean signal sampled at 1 Hz. Right panel: $\Delta S$ estimated from the noisy (white noise with variance equal 1% of the signal level) signal sampled as 1 Hz. (B) Left panel: $\Delta S$ estimated from clean signal sampled as 10 Hz, Right panel: $\Delta S$ estimated from noisy (1%) signal sampled at 10 Hz.

For task-related fMRI, neural population dynamics can be captured by modeling signal dynamics with explicit exogenous inputs, which is absent during resting-state fMRI recordings. Blind deconvolution considers resting-state fMRI as spontaneous event related and then estimates region-specific HRF and performs deconvolution on the signal. For details, see Wu et al. (2013).

Here we performed blind deconvolution on BOLD signals simulated from the thalamocortical model and applied differential covariance analysis on the deconvoluted signal (see Figure 14). The performance of blind deconvolution is not as good as the backward reconstruction process.

Appendix F: Robustness of the Backward Model

There is considerable uncertainty regarding $E_o$ and $\epsilon$ (Stephan, Weiskopf, Drysdale, Robinson, & Friston, 2007). To evaluate whether the backward model is applicable to different parameter values of $E_o$ and $\epsilon$, we randomly sampled $E_o$ and $\epsilon$ from the following distribution (see equation F.1) and simulated multiple sets of BOLD signals. Backward reconstruction with
Figure 14: Differential covariance analysis applied to blind deconvoluted BOLD signal simulated from the thalamocortical model.

Figure 15: C-sensitivity of $\Delta_5$ with different forward balloon model parameters.
$E_o = 0.5$ and $\epsilon = 0.5$ followed by differential covariance analysis was applied to the signals and the performance of $\Delta_s$ was quantified by c-sensitivity (see Figure 15). The c-sensitivity of $\Delta_s$ remains to be high with respect to a wide range of forward balloon model parameters:

$$E_o \sim U(0.2, 0.8); \quad \epsilon \sim 0.5 \ast \exp(N(0, 0.5^{0.5})). \quad (F.1)$$

Acknowledgments

This research was supported by the Office of Naval Research (N00014-16-1-2829) and NIH/NIBIB (R01EB026899).

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1), 63–72.

Banerjee, O., Ghaoui, L. E., d’Aspremont, A., & Natsoulis, G. (2006). Convex optimization techniques for fitting sparse gaussian graphical models. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 89–96). New York: ACM.

Bazhenov, M., Timofeev, I., Steriade, M., & Sejnowski, T. J. (2002). Model of thalamocortical slow-wave sleep oscillations and transitions to activated states. *Journal of Neuroscience*, 22(19), 8691–8704.

Bonjean, M., Baker, T., Lemieux, M., Timofeev, I., Sejnowski, T., & Bazhenov, M. (2011). Corticothalamic feedback controls sleep spindle duration in vivo. *Journal of Neuroscience*, 31(25), 9124–9134.

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.

Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., & Willsky, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2), 572–596.

Chen, J.-Y., Chauvette, S., Skorheim, S., Timofeev, I., & Bazhenov, M. (2012). Interneuron-mediated inhibition synchronizes neuronal activity during slow oscillation. *Journal of Physiology*, 590(16), 3987–4010.

Chickering, D. M. (2003). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3, 507–554.

Cox, D. R., & Wermuth, N. (1996). *Multivariate dependencies: Models, analysis and interpretation*. Boca Raton, FL: CRC Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302.

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association*, 79(388), 907–915.
Heeger, D. J., & Ress, D. (2002). What does FMRI tell us about neuronal activity? Nature Reviews Neuroscience, 3(2), 142–151.

Hsieh, C.-J., Sustik, M. A., Dhillon, I. S., & Ravikumar, P. (2014). QUIC: Quadratic approximation for sparse inverse covariance estimation. Journal of Machine Learning Research, 15(1), 2911–2947.

Hsieh, C.-J., Sustik, M. A., Dhillon, I. S., Ravikumar, P. K., & Poldrack, R. (2013). BIG & QUIC: Sparse inverse covariance estimation for a million variables. In C. J. C. Burgess, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), Advances in neural information processing systems, 26 (pp. 3165–3173). Red Hook, NY: Curran.

Khalidov, I., Fadili, J., Lazeyras, F., Van De Ville, D., & Unser, M. (2011). Activelets: Wavelets for sparse representation of hemodynamic responses. Signal Processing, 91(12), 2810–2821.

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., ... Petersen, S. E. (2015). Functional system and areal organization of a highly sampled individual human brain. Neuron, 87(3), 657–670.

Lin, T. W., Das, A., Krishnan, G. P., Bazhenov, M., & Sejnowski, T. J. (2017). Differential correlation: A new method to estimate sparse connectivity from neural recordings. Neural Computation, 29(10), 2581–2632.

Lin, Z., Liu, R., & Su, Z. (2011). Linearized alternating direction method with adaptive penalty for low-rank representation. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), Advances in neural information processing systems, 24 (pp. 612–620). Red Hook, NY: Curran.

Logothetis, N. K. (2002). The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 357(1424), 1003–1037.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (pp. 403–410). San Mateo, CA: Morgan Kaufmann.

Mumford, J. A., & Ramsey, J. D. (2014). Bayesian networks for FMRI: A primer. NeuroImage, 86, 573–582.

Obata, T., Liu, T. T., Miller, K. L., Luh, W.-M., Wong, E. C., Frank, L. R., & Buxton, R. B. (2004). Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: Application of the balloon model to the interpretation of BOLD transients. NeuroImage, 21(1), 144–153.

Patel, R. S., Bowman, F. D., & Rilling, J. K. (2006). A Bayesian approach to determining connectivity of the human brain. Human Brain Mapping, 27(3), 267–276.

Raichle, M. E. (2015). The brain’s default mode network. Annual Review of Neuroscience, 38, 433–447.

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., & Glymour, C. (2010). Six problems for causal inference from FMRI. NeuroImage, 49(2), 1545–1558.

Ramsey, J., Zhang, J., & Spirtes, P. L. (2012). Adjacency-faithfulness and conservative causal inference. arXiv:1206.6843.

Reid, A. T., Headley, D. B., Mill, R. D., Sanchez-Romero, R., Uddin, L. Q., Marinazzo, D., ... Cole, M. W. (2019). Advancing functional connectivity research from association to causation. Nature Neuroscience, 1(10).
Richardson, T., & Spirtes, P. (1996). *Automated discovery of linear feedback models*. https://www.cmu.edu/dietrich/philosophy/docs/tech-reports/75_Richardson.pdf

Seth, A. K. (2010). A Matlab toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods, 186*(2), 262–273.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., . . . Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage, 54*(2), 875–891.

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.

Stevenson, I. H., Rebesco, J. M., Miller, L. E., & Körding, K. P. (2008). Inferring functional connections between neurons. *Current Opinion in Neurobiology, 18*(6), 582–588.

Wu, G.-R., Liao, W., Stramaglia, S., Ding, J.-R., Chen, H., & Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain networks from resting state FMRI data. *Medical Image Analysis, 17*(3), 365–374.

Yatsenko, D., Josić, K., Ecker, A. S., Froudarakis, E., Cotton, R. J., & Tolias, A. S. (2015). Improved estimation and interpretation of correlations in neural circuits. *PLOS Computational Biology, 2*, 199–207.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence, 172*(16), 1873–1896.

Received December 17, 2019; accepted July 1, 2020.