Inference of multi-scale connectivity in the brain

Inaugural-Dissertation

zur Erlangung der Doktorwürde der Fakultät für Biologie

der Albert-Ludwigs-Universität Freiburg im Breisgau

vorgelegt von

Jonathan Schiefer

geboren in Herrenberg

Freiburg im Breisgau

April, 2018

Dekanin der Fakultät für Biologie:	Prof. Dr. Bettina Warscheid
Promotionsausschausvorsitzender:	Prof. Dr. Andreas Hiltbrunner
Referent:	Prof. Dr. Stefan Rotter
Koreferent:	Prof. Dr. Dr. h.c. Jürgen Hennig
Drittprüfer:	Prof. Dr. Carsten Mehring
Datum der mündlichen Prüfung:	11.07.2018

Contents

Abstract v

Zusammenfassung vii

Acknowledgements ix

Declaration xi

1 Introduction 1

2 From Correlation to Causation: Estimation of Effective Connectivity from Continuous Brain Signals based on Zero-Lag Covariance 11

3 Sparse Estimation of Resting-State Effective Connectivity from fMRI Cross-Spectra 29

4 Estimating connectivity from simulated ECoG signals 49

5 Discussion 63

Bibliography 78

Abstract

The brain consists of neurons which are interconnected by synapses. The connectivity of the networks formed by the neurons and synapses is a key feature for the function and dysfunction of the brain. In humans, studying the connectivity comes with various challenges. Thus, the access to connectivity in humans is limited. The aim of this dissertation is to introduce a novel method to estimate effective connectivity of neural populations from continuous recordings of the activity of these populations which overcomes limitations of existing methods. This method estimates the connectivity of neural populations based on the covariance of the measured activity. The key mechanism for doing so is a L^1-minimization via a gradient descent on the manifold of unitary matrices. The fact that the gradient of a matrix on the unitary manifold is skew-hermitian and with this in the corresponding Lie-Algebra, is exploited in the update step to project the gradient back on the manifold via the exponential map.

As presented in this thesis, this method works reliably for sparse networks with more than 40 nodes and a sufficient network interaction. The method can be applied on zero-lag covariance matrices, hence there is no restriction on the sampling rate of the measurement. Although based on a linear interaction model, the method also achieves reasonable results for networks which interact non-linearly. A comparison with structural measures based on fMRI data shows better agreement than state-of-the-art methods. Also, the method is robust against noise, unobserved nodes and variable hemodynamics of BOLD signals.

In this thesis, a novel method for the estimation of effective connectivity from covariances of neural activity is presented. The features of this method make it applicable on a broad range of data-types, including data based on the BOLD effect or electro physiologic data such as ECoG. Applications of fast fMRI data show plausible and coherent results.

Zusammenfassung

Das Gehirn besteht aus Neuronen welche durch Synapsen verbunden sind. Die Konnektivität, die von Neuronen durch Synapsen geformt wird, ist ein Haupt-Charakteristikum für die Funktion und Funktionsstörung des Gehirns. Um die Konnektivität im menschlichen Gehirn untersuchen zu können, müssen mehrere Herausforderungen bewältigt werden. Dadurch ist der Zugang zur Konnektivität im menschlichen Gehirn limitiert. Ziel dieser Dissertation ist es, eine neue Methode zur Schätzung von effektiver Konnektivität neuronaler Populationen aus kontinuierlichen Signalen dieser Populationen vorzustellen, welche die Einschränkungen existierender Methoden überwindet. Die Methode schätzt die Konnektivität von neuronalen Populationen auf Basis der Kovarianzen der gemessenen Aktivität. Der zugrunde liegende Mechanismus ist eine L^1-Minimierung die durch ein Gradientenverfahren auf der Mannigfaltigkeit der unitären Matrizen berechnet wird. Dabei wird ausgenutzt, dass der Gradient einer Matrix der unitären Mannigfaltigkeit schiefhermitesch ist und durch die Exponentialfunktion auf die unitäre Mannigfaltigkeit zurück projiziert werden kann.

Diese Arbeit zeigt, dass diese Methode für spärliche Netzwerke mit mehr als 40 Knoten und einer ausreichenden Netzwerk-Interaktion zuverlässig funktioniert. Die Methode kann auf die Kovarianzen, die ohne zeitliche Verschiebung berechnet werden, angewendet werden und hat dadurch keine Einschränkungen in der Anwendbarkeit bezüglich der zeitlichen Abtastrate der Messung. Obwohl die Methode auf einem lineare Interaktions-Modell basiert, können damit auch gute Ergebnisse für Netzwerke mit nicht-linearer Interaktion erzielt werden. Vergleiche mit Maßen von struktureller Konnektivität basierend auf fMRI-Daten zeigen eine bessere Übereinstimmung als andere aktulle Methoden. Des weiteren ist die Methode robust gegenüber verrauschten Signalen, nicht beobachteten Knoten in Netzwerken und variabler neurovaskulärer Kopplung bei BOLD Signalen.

Zusammenfassung

In dieser Arbeit wird eine neue Methode zur Schätzung effektiver Konnektivität aus den Kovarianzen neuronaler Aktivität präsentiert. Die Charakteristiken dieser Methode erlauben die Anwendung auf eine breite Klasse von Datentypen, einschließlich BOLD-Effekt basierender oder elektrophysiologischer Messungen. Die Anwendung auf fMRI Messungen mit hoher zeitlicher Auflösung zeigen plausible und kohärente Resultate.

Acknowledgements

During the time working on this thesis I benefited from help and advice of numerous colleagues and friends whom I would like to thank. First, Stefan, my supervisor, for his support, ideas and for the possibility to follow my own ideas which might not be directly related to my project. Also, I want to thank my co-worker Alexander and co-adviser Volker who provided a low thresholded start into my project and, of course, all other BCF PhD students for discussions and support, especially Benjamin and Robin for the bike trips and Michael for reading a lot of the things I wrote. For writing this thesis, I relied heavily on computing and organizational infrastructure at the BCF and want to thank all people involved in providing this. Finally, I want to thank my family and friends for help and support during these years, in particular my girlfriend Jana who helped my a lot during my PhD.

Declaration

- Chapter 2 is published as

  Schiefer J, Niederbühl A, Pernice V, Lennartz C, Hennig J, LeVan P and Rotter S. (2018) From correlation to causation: Estimating effective connectivity from zero-lag covariances of brain signals. PLoS Comput Biol 14(3): e1006056. https://doi.org/10.1371/journal.pcbi.1006056

  This study was conceived and designed by J. Schiefer, A. Niederbühl, V. Pernice, C. Lennartz, J. Hennig, P. LeVan and S. Rotter. The data was analyzed by J. Schiefer, A. Niederbühl, V. Pernice and C. Lennartz. The manuscript was written by J. Schiefer and reviewed and edited by A. Niederbühl, V. Pernice, C. Lennartz, J. Hennig, P. LeVan and S. Rotter.

- Chapter 3 is published as

  Lennartz C, Schiefer J, Rotter S, Hennig J and LeVan P (2018) Sparse Estimation of Resting-State Effective Connectivity from fMRI Cross-Spectra. *Front. Neurosci. 12:287.* doi: 10.3389/fnins.2018.00287

  The study was conceived and designed by C. Lennartz, J. Schiefer, S. Rotter, J. Hennig and P. LeVan. The data was analyzed by C. Lennartz and J. Schiefer. The manuscript was drafted and approved by C. Lennartz, J. Schiefer, S. Rotter, J. Hennig and P. LeVan.

- Chapter 4 will be submitted soon as

  > Schiefer J, Schäfer LM and Rotter S. Estimating connectivity from simulated ECoG signals.

  The study was conceived and designed by J. Schiefer, L. Schäfer and S. Rotter. The data was analyzed by J. Schiefer. The manuscript was written by J. Schiefer and reviewed and edited by L. M. Schäfer and S. Rotter.

- An implementation of the algorithm presented in this thesis, together with examples of how to use it, is available at `https://github.com/jonathanschiefer/connFinder`.

1 Introduction

In modern neuroscience, the human brain is studied as a network of networks formed by neurons and connected by synapses. Function and malfunction of the brain is tightly linked to the structure and activity of these networks. Especially neural diseases such as Parkinson's, Alzheimer's and epilepsy are caused by changes of the neuronal network of the brain (Wenk, 2003; Davie, 2008; Jacobs et al., 2012; Goldberg and Coulter, 2013; Franzmeier et al., 2018). This makes neural connectivity, independent of the scale on which it is studied, a prominent research subject.

Recently, Li et al. (2017) argued that knowledge of effective connectivity is crucial for the understanding of many neuropsychiatric disorders. Generally, monitoring of effective connectivity is a potential candidate for the detection of degenerative neural diseases. Most of these diseases start long before the first symptoms occur. Changes in connectivity, therefore, can be a hint on such diseases. As more and better data becomes available, it might be possible in the future to build classifiers for diseases based on connectivity. Additionally, for modeling purposes, knowledge of connectivity is essential. Regardless of whether one is aiming to model neural circuits (Potjans and Diesmann, 2014) or the whole human brain (Tiesinga et al., 2015), without knowledge of connectivity it is impossible for models to match experimental findings.

In the brain, connectivity is established by synapses which connect neurons. In recent years, it has been common to study three types of connectivity: *structural connectivity* which is based on anatomical connections between neural units, *functional connectivity* which is based on statistical dependencies of neural activity and *effective connectivity* which reflects causal interactions among neural units (Friston, 2011). Just like in every complex system, connectivity plays a key role in neural dynamics and their correlations (Kriener et al., 2008; Tetzlaff et al., 2008; Pernice et al., 2011; Trousdale et al., 2012). Based on the analysis of Hawkes processes (Hawkes, 1971a,b) the

inverse problem, estimating the (integrated) kernel matrix, resulting in effective connectivity, from correlations has been solved (Pernice and Rotter, 2013; Bacry and Muzy, 2014). It has been shown that for Hawkes processes the spike count correlation matrix is related to the Hawkes kernel matrix through a system of Wiener–Hopf integral equations (Bacry and Muzy, 2016). Another approach uses integrated cumulants for the estimation of integrated Hawkes kernels (Achab et al., 2017).

On a coarser scale, it is common to divide the neurons in the brain into populations or regions of interest (ROIs), either on an anatomical (Brodmann, 1909; Tzourio-Mazoyer et al., 2002), a connectional basis (Fan et al., 2016) or implicitly by the range of the electrodes recording the activity. These neural populations then form a network of networks, still being interconnected on neuronal level. Neural signals of these populations can then either be measured invasively, for example with electrocorticography (ECoG), or non-invasively with methods like functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). On population scale, the concept of Granger Causality (Granger, 1969), partial directed coherence (Baccalá and Sameshima, 2001), Dynamic Causal Modeling (DCM) (Friston et al., 2003) and a number of other methods (Smith et al., 2011) have been used to estimate connectivity from such signals (Kiebel et al., 2008; Friston, 2009; Friston et al., 2014; Korzeniewska et al., 2008, 2011). However, all of these methods have their shortcomings: for example DCM is limited to very small networks (< 20 nodes) and Granger Causality relies on high temporal resolution. As it turns out, the modeling of population dynamics is also important for connectivity inference. But this is itself challenging and has multiple degrees of freedom (Coombes, 2010; Einevoll et al., 2013; Mazzoni et al., 2015).

Now, the question arises whether it is possible to overcome the problems of the existing methods for the estimation of effective connectivity on population level. The challenge is, estimating connectivity from population activity for large networks (up to hundreds of nodes) independent of the temporal resolution of the data. Also, the application on different data types should be possible: ECoG data is usually recorded in a trial based format (short but many trials), whereas fMRI data is recorded in one trial for up to ten minutes. Additionally, the algorithm should be applicable to as many types of networks as possible: the network topology, density and the spectral radius of the adjacency matrix should not have an effect on the estimation. In the case of ECoG data one also faces the problem, that the electrode only covers a part of the brain. So

the estimation method should be feasible in a setting of latent populations which interact with the observed populations but are not recorded themselves.

1.1 The concept of causal inference in the brain

The idea of learning about causal relationships from any kind of data is rather old. The basis of what is today known as *causal models* is the link of probabilistic dependencies and graphs, which was introduced by Rebane and Pearl (1987). One of the essential assumptions in the concept of causality is temporal precedence. This means, that in a causal system the effect cannot occur before the cause. Unfortunately, this is not sufficient without the observation of all factors involved as illustrated by Pearl (2000):

> "the barometer falls before it rains yet does not cause the rain".

In Rebane and Pearl (1987), the key to overcoming this problem is presented by studying the three basic causal substructures:

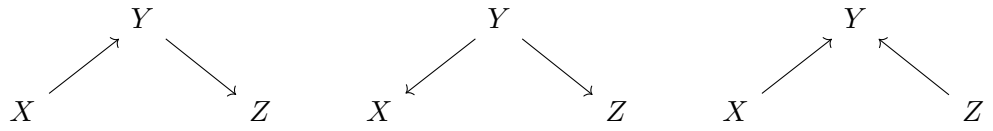

Figure 1.1: The three basic causal substructures: *chain*, *fork* and *collider*.

They ask the question whether causal influences in the underlying process can be infered by distinction among the dependencies of these structures. The determination of the direction of the causal relationship between X and Y is possible in the case of

> "the presence of a third variable Z that correlates with Y but not with X", Rebane and Pearl (1987).

This is the case in the collider structure $X \longrightarrow Y \longleftarrow Z$ which makes it a key structure in the concept of causal inference.

Nowadays, mathematical modeling of the brain uses the concept of networks of networks of neurons. The networks of neurons can be interpreted as finite directed graphs, where the neurons

correspond to the nodes and the synapses correspond to the edges. In the case of networks of neural populations, the populations correspond to the nodes and the connections between the populations to the edges. Causal interactions are represented by the transmission of currents through synapses evoked by action potentials generated by the neurons. The assumption of temporal precedence is justified in this setting: the depolarization or hyperpolarization of the post-synaptic neuron, depending on the type of the synapse, can only occur after the pre-synaptic has neuron fired an action potential.

On the scale of neural populations, this corresponds to the accumulated counterparts. For instance, a local field potential (LFP) signal is thought as a superposition of synaptic currents (Buzsáki et al., 2012; Einevoll et al., 2013). FMRI signals, on the other hand, are based on the coupling of cerebral blood flow, energy demand, and neural activity (Logothetis, 2003). In either case, the effect on the post-synaptic population cannot precede the cause in the pre-synaptic population. In case of LFPs for the same reason as for individual neurons, and for fMRI the cerebral blood flow cannot precede the energy demand, which can itself not precede the neural activity.

In the community based on the work of Judea Pearl, inference of causal relationships from weak structural assumptions, e.g. general directed acyclic graphs by studying statistical independence, is very popular. Unfortunately, even these assumptions are too strict for the graphs considered in neuroscience. As the brain is a highly recurrent network, the assumption of an acyclic graph is not feasible in our setting. Even in the visual system, which is a part of the brain with prominent feed-forward structure, feedback connections exist (Lamme and Roelfsema, 2000). This means, that even if one considers only sub-networks of the brain, the assumption of an acyclic graph is violated. Nevertheless, as I will show below, the investigation of the three basic causal structures, especially collider structures, can be beneficial for the concept of causality in the brain.

1.2 Structural, functional and effective connectivity

Structural connectivity describes the anatomical connections (more precisely existence and direction of connections) between neural units. Recently, lots of effort has been put into the

mapping of the human connectome, that is the structural connectivity of the complete human brain (Sporns et al., 2005). This is challenging, as invasive tracing methods can only be applied post mortem. Non-invasively, tractography algorithms on diffusion tensor imaging (DTI) data can reconstruct white matter tracts forming connections between regions in vivo (Wedeen et al., 2008; Reisert et al., 2013). However, this method has its limitations, especially for short-range connections. Also, as DTI data measures the diffusion of water molecules, it is not possible to reconstruct the direction of connections.

In Friston (2011), *functional connectivity* is defined as

> "... statistical dependencies among remote neurophysiological events".

Following this definition, the concept of functional connectivity is very general, applicable to any type of data. The accuracy of the estimate is, of course, highly dependent on the quality of the data, resulting from data length, sampling rate and artifacts contaminating the measurement such as breathing and heartbeat. It is independent of models underlying the dynamics. In contrast to structural and effective connectivity, functional connectivity is by definition, as covariance, or coherence in the frequency domain, a symmetric respectively undirected measure.

According to `https://pubmed.gov`[1] the term *effective connectivity* was first used by Melssen and Epping (1987) who used it to describe "combined influences of parallel pathways". In the following years, the usage of this term was not consistent across the field (Aertsen et al., 1989). Aertsen and Preissl (1991) came up with the following definition:

> "effective connectivity should be understood as the experiment and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons."

Today, this definition describes what usually is understood as effective connectivity (Friston, 2011). It has two elements which distinguish it from functional connectivity fundamentally: it is time dependent and relies on the model of interaction assumed. The approach of a "possible circuit diagram" matching the observation can be interpreted as a link to graphs as described in section 1.1.

[1] retrieved: February 21, 2018

A key difference between functional and effective connectivity can be illustrated as follows: if two nodes are not connected but have a common input, like in the fork $X \longleftarrow Y \longrightarrow Z$, the nodes X and Z will have statistical dependencies induced by the activity of Y, thus the functional connectivity between X and Z is non-zero. On the other hand, there is no causal interaction between X and Z, the effective connectivity is zero accordingly.

The differences in the concepts of structural, functional and effective connectivity show that the belief of getting a complete understanding of the brain by knowledge of the full human connectome has clear limitations. First, plasticity is present in the adult human brain. This means that the brain is constantly changing its structural connectivity, partially based on mechanisms as short-term depression and long-term potentiation, which are itself dependent of the neural activity. Also, the neural pathways alone do not allow much insight about function without taking activity on these neural pathways into consideration. The effect of a synapse, of course, depends on the activity of the pre-synaptic neuron. In the most extreme case, when the pre-synaptic neurons is completely silent, then the existence of a connection to another neuron has no effect on the post-synaptic neuron. On the other hand, knowledge about structural connectivity can be used as constraint for the estimation of effective connectivity (Gilson et al., 2016). However, the quality of the structural estimate is then fundamental for the estimation of the effective connectivity.

1.3 Inferring effective connectivity

As described in the section above, effective connectivity is model dependent. This induces constraints on possible validation procedures. In simulations, *anatomically* connected neural units are usually simulated with some kind of neural dynamics. The validation for such algorithms is then performed by comparing the estimated connections with structurally existing connections. This implicitly assumes that the underlying model of neural interactions fits the data generating process. For simulations, this problem is negligible as the interaction model of the estimation method can be used for simulating data. On the other hand, simulations also allow validating the robustness of the estimation method for different types of interaction models.

Biological neurons interact non-linearly. However, estimating connectivity of large networks

is not feasible without the assumption of linearity. Granger Causality, partial directed coherence as well as the method presented in this thesis assume linearity in the interaction of the neural units. This is a generic linear model of population activity

$$x(t) = (G * x)(t) + v(t) \tag{1.1}$$

where the vector $x(t)$ denotes the activity at time t, $G(t)$ the matrix of interaction kernels and $v(t)$ the fluctuating external inputs. More details about the estimation of the (integrated) linear interaction kernels G from the activity x can be found in Chapter 2.

For estimating effective connectivity from real data as fMRI, where no ground truth is available (Zaghlool and Wyatt, 2014; Ryali et al., 2016; Nie et al., 2017), the problem becomes more involved. Neither a ground truth for structural connectivity is available, nor is there a broadly accepted model for fMRI dynamics (Welvaert and Rosseel, 2014). It is only possible to validate the estimation methods with simulations based on various models, especially interaction models which are not the underlying model of the estimation method itself. Additionally, one can check the outcome of the algorithms on real data for plausibility and consistency with other estimation methods and measures, like structural connectivity reconstructed from DTI. Of course, this is only possible if the structural connectivity was not used as constraint for the effective connectivity. Clearly, the result of estimation methods can be coherent and plausible, but still totally wrong.

1.4 The scope of this thesis

In this thesis, I provide an algorithm to estimate effective connectivity for sparse networks. The algorithm is based on the same idea as Pernice and Rotter (2013), namely using a L^1-minimization to estimate connectivity which, based on a linear interaction model, fits the covariance structure of the data. This is done via a minimization on a matrix manifold with unitary constraint (Abrudan et al., 2009; Wen and Yin, 2013). It can be applied on networks with up to hundreds of nodes and is independent of the spectral radius of the networks adjacency matrix. Also, it can deal with the presence of latent nodes in networks. The applicability to zero-lag covariances as well as frequency-resolved cross-spectral densities allows applications on a broad range of data. By construction, this algorithm achieves reasonable accuracy only for

sparse networks, roughly up to 17% connection probability. This is due to the L^1-minimization which favors sparse solutions (Yuan and Lin, 2007). However, in presence of varying connection strengths, the method is able to estimate the strongest connection by estimating a sub-graph.

In chapter 2, the algorithm is introduced and intrinsic properties are studied based on noise-free covariance matrices. The performance of the estimation is also evaluated based on simulations of Ornstein-Uhlenbeck processes filtered with the canonical hemodynamic response function (Friston et al., 1998; Glover, 1999) as model for blood-oxygen-level dependent imaging (BOLD) data. The estimation is solely based on the zero-lag covariances of the data, making it independent of temporal precedence in the data. This makes it a good candidate for data where the sampling rate is much slower than the timescale of the neural interactions. Applications on MREG data, which is a fast fMRI sequence (Assländer et al., 2013), are shown. Also, the performance of the estimation method is compared with the Regularized Inverse Covariance method (Smith et al., 2011) which shows a superior performance of the new method.

In chapter 3 a variant of the estimation method is applied on MREG data and additional conditions are studied based on simulations. In this case, the estimation method is not applied on the zero-lag covariances but on cross-spectral densities. This leads to frequency resolved network, one for each frequency bin. The estimation of effective connectivity is validated with DTI data. Also, the effect of the sampling rate of the MRI sequence is studied, showing good performance of the estimation method for a MREG sampling rate (10 Hz) as well as standard fMRI sampling rates ($0.3 - 1$ Hz). The influence of variable hemodynamic response functions is studied showing better robustness of the proposed estimation method against it than Granger Causality. The influence of the strength of noise, type of noise and length of the measurement are studied based on cross-spectral densities conducted from simulated autoregressive processes.

Chapter 4 addresses the question of how crucial the assumption of linearity in the underlying neural interaction is. For this purpose, a simulation of an ECoG electrode is set up, based on non-linearly interacting leaky-integrate-and-fire neurons. Local field potentials are derived from the synaptic currents of the neurons and are used for the estimation of the connectivity of the neural populations underneath the electrodes. Analogue to chapter 2, the connectivity is estimated based on zero-lag covariances. As the structure of the data in this study is inspired by ECoG measurements, the covariance matrices are calculated for each trial and averaged before

the connectivity is estimated. The performance is compared to classical Granger Causality inference (Lütkepohl, 2005) and is shown to be superior for short trials. This study shows that even assuming a linear interaction model, the estimation method can reconstruct effective connectivity from data generated by non-linear processes.

In chapter 5 some aspects are explained in more detail and alternative approaches for parts of the algorithm are discussed. The method is compared to other estimation methods and the assumption of a linear interaction model is examined.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

From correlation to causation: Estimating effective connectivity from zero-lag covariances of brain signals

Jonathan Schiefer[1,2], Alexander Niederbühl[1,2], Volker Pernice[1,2], Carolin Lennartz[3,4], Jürgen Hennig[3,4], Pierre LeVan[3,4], Stefan Rotter[1,2]*

1 Bernstein Center Freiburg, Freiburg, Germany, **2** Faculty of Biology, University of Freiburg, Freiburg, Germany, **3** Department of Radiology, Medical Physics, Medical Center - University of Freiburg, Freiburg, Germany, **4** Faculty of Medicine, University of Freiburg, Freiburg, Germany

* stefan.rotter@biologie.uni-freiburg.de

Citation: Schiefer J, Niederbühl A, Pernice V, Lennartz C, Hennig J, LeVan P, et al. (2018) From correlation to causation: Estimating effective connectivity from zero-lag covariances of brain signals. PLoS Comput Biol 14(3): e1006056. https://doi.org/10.1371/journal.pcbi.1006056

Editor: Saad Jbabdi, Oxford University, UNITED KINGDOM

Received: September 12, 2017

Accepted: February 26, 2018

Published: March 26, 2018

Data Availability Statement: All data analysis and simulation software files are available at https://github.com/jonathanschiefer/connFinder. The fast fMRI data files are available from the Open Science Framework at https://osf.io/52mf4/.

Funding: Supported by the DFG (grant EXC 1086). The HPC facilities are funded by the state of Baden-Württemberg through bwHPC and DFG grant INST 39/963-1 FUGG. The article processing charge was covered by the open access publication fund of the University of Freiburg. The funders had no role in

Abstract

Knowing brain connectivity is of great importance both in basic research and for clinical applications. We are proposing a method to infer directed connectivity from zero-lag covariances of neuronal activity recorded at multiple sites. This allows us to identify causal relations that are reflected in neuronal population activity. To derive our strategy, we assume a generic linear model of interacting continuous variables, the components of which represent the activity of local neuronal populations. The suggested method for inferring connectivity from recorded signals exploits the fact that the covariance matrix derived from the observed activity contains information about the existence, the direction and the sign of connections. Assuming a sparsely coupled network, we disambiguate the underlying causal structure via L^1-minimization, which is known to prefer sparse solutions. In general, this method is suited to infer effective connectivity from resting state data of various types. We show that our method is applicable over a broad range of structural parameters regarding network size and connection probability of the network. We also explored parameters affecting its activity dynamics, like the eigenvalue spectrum. Also, based on the simulation of suitable Ornstein-Uhlenbeck processes to model BOLD dynamics, we show that with our method it is possible to estimate directed connectivity from zero-lag covariances derived from such signals. In this study, we consider measurement noise and unobserved nodes as additional confounding factors. Furthermore, we investigate the amount of data required for a reliable estimate. Additionally, we apply the proposed method on full-brain resting-state fast fMRI datasets. The resulting network exhibits a tendency for close-by areas being connected as well as inter-hemispheric connections between corresponding areas. In addition, we found that a surprisingly large fraction of more than one third of all identified connections were of inhibitory nature.

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Changes in brain connectivity are considered an important biomarker for certain brain diseases. This directly raises the question of accessibility of connectivity from measured brain signals. Here we show how directed effective connectivity can be inferred from continuous brain signals, like fMRI. The main idea is to extract the connectivity from the inverse zero-lag covariance matrix of the measured signals. This is done using L^1-minimization via gradient descent algorithm on the manifold of unitary matrices. This ensures that the resulting network always fits the same covariance structure as the measured data, assuming a canonical linear model. Applying the estimation method on noise-free covariance matrices shows that the method works nicely on sparsely coupled networks with more than 40 nodes, provided network interaction is strong enough. Applying the estimation on simulated Ornstein-Uhlenbeck processes supposed to model BOLD signals demonstrates robustness against observation noise and unobserved nodes. In general, the proposed method can be applied to time-resolved covariance matrices in the frequency domain (cross-spectral densities), leading to frequency-resolved networks. We are able to demonstrate that our method leads to reliable results, if the sampled signals are long enough.

Introduction

The networks of the brain are key to understanding its function and dysfunction [1]. Depending on the methods employed to assess structure and to record activity, networks may be defined at different levels of resolution. Their nodes may be individual neurons, linked by chemical or electrical synapses. Alternatively, nodes may also be conceived as populations of neurons, with links represented by the net effect of all synaptic connections that exist between two populations. In any case, this defines the structural substrate of brain connectivity, representing the physical (causal) basis of neuronal interactions. Nodes in a brain network influence each other by sending signals. For example, the activities of nodes in a network are generally not independent, and neuronal dynamics are characterized by correlations among the nodes involved in the network. This suggests an alternative perspective on active brain networks: Functional connectivity assigns a link to a pair of nodes to the degree to which their activities are correlated. It has been argued that this concept emphasizes connections that "matter", including the possibility that the same substrate may give rise to different networks, depending on how they are used. As a consequence, functional connectivity and structural connectivity are not equivalent. A well-known phenomenon is that two nodes may be correlated, even if there is no direct anatomical link between them. For example, a shared source of input to both nodes may generate such a correlation, which does not correspond to a direct interaction between the two nodes. Apart from that, correlation is a symmetric relation between two nodes, whereas a physical connection implies a cause-effect relation that is directed. There have, in fact, been multiple attempts to overcome the shortcomings of functional connectivity, especially the lack of directed interaction. The term *effective connectivity* has been suggested for this [2]. The idea is to bring the networks, inferred from activity measurements, closer to structural connectivity, which can only be inferred with anatomical methods. The dichotomy between structural and functional aspects of connectivity raises the general question whether it is possible to infer brain networks from recorded activity. We are only beginning to understand the forward link between structural connectivity and functional connectivity. As a consequence, it is possible to compute correlations from connectivity in certain simplified network

scenarios [3]. The correspondence between connectivity and correlation, however, is not one-to-one. Networks with different connectivity may lead to exactly the same correlations between nodes. As a consequence, the inverse problem of inferring connectivity from correlation is generally ill-defined. As we will demonstrate in this paper, additional assumptions about the connectivity can help to resolve the ambiguity. Specifically, we search for the network with the lowest number of nonzero edges (via L^1-minimization) to disambiguate the problem. Structural, functional and effective connectivity are not equally well accessible. Some aspects of the anatomical structure can be assessed *post mortem* by invasive tracing methods, or non-invasively by Diffusion Tensor Imaging, DTI. In contrast, functional connectivity is based on statistical relationships between the activity of neuronal populations and can be easily estimated from recorded signals. For estimating effective connectivity there are methods like Dynamic Causal Modelling, DCM [4, 5], Granger causality [6] and others [7–13]. Only few methods to infer effective connectivity, however, can deal with large numbers of nodes (40 or more) based on zero-lag correlation only. However, they are either limited to small networks [14], or to directed acyclic graphs [15]. Here, we are proposing a new method for the estimation of effective connectivity from population activity in the brain, especially BOLD-related signals. The new method is a variant of the procedure described in [16], based on a L^1-minimization. For the method proposed here it is sufficient to use zero-lag covariances to estimate directed effective connectivity.

Materials and methods

Estimation method

The main idea of our estimation method is inspired by the finding, "that the key to determining the direction of the causal relationship between X and Y lies in 'the presence of a third variable Z that correlates with Y but not with X,' as in the collider $X \rightarrow Y \leftarrow Z \ldots$" [17, 18]. Similarly, assuming a linear interaction model, the presence of a collider structure in a network (see Fig 1) produces specific entries in the corresponding inverse covariance (precision) matrix. Fig 1 shows a disconnected network in the left column, and a network which induces the same covariance matrix if all links have opposite direction in the middle column. In the latter case an estimation of the direction is impossible, because there is simply no information about it in the covariance matrix. Whenever a collider structure is present, however, the entry in the inverse covariance matrix for the two source nodes (here, 2 and 3) is non-zero. This is due to the fact that in a linear model the entry in the inverse covariance matrix depends not only on the connections of the nodes 2 and 3, but also whether these nodes have a common target. This means the presence of a collider structure allows us to disambiguate the direction of this particular connection.

We consider here a scenario, where the interaction between nodes is described by a generic linear model. Assuming stationarity, let the neural activity $x(t)$ be implicitly defined by the consistency equation

$$x(t) = (G * x)(t) + v(t) \tag{1}$$

where $G(t)$ is a matrix of causal interaction kernels and $v(t)$ denotes fluctuating external inputs ("driving noise"). All variables are also listed in Table 1. Fourier transformation of Eq (1) yields

$$\hat{x}(f) = \hat{G}(f)\hat{x}(f) + \hat{v}(f)$$

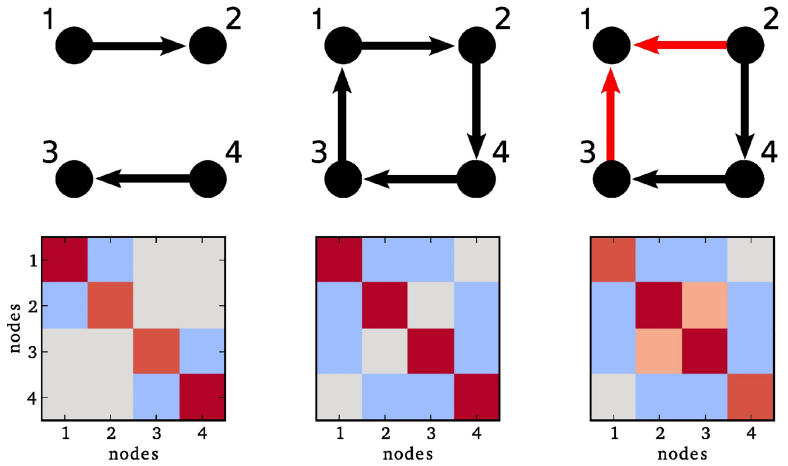

Fig 1. Collider structures are encoded in the inverse covariance matrix. Upper row: Three simple network architectures. Lower row: The corresponding inverse covariance matrices, red color represents positive entries, blue color stands for negative ones. In the left and middle column, the entries (2, 3) and (3, 2) are 0. The only difference between the right column and the middle column is that the connection between node 1 and 2 is flipped, such that nodes 1, 2 and 3 form a collider structure. Although there is still only an indirect connection between node 2 and 3, the entry in the corresponding inverse covariance matrix is now non-zero.

https://doi.org/10.1371/journal.pcbi.1006056.g001

and simple rearrangement leads to

$$\hat{x}(f) = \left(\mathbb{1} - \hat{G}(f)\right)^{-1} \hat{v}(f)$$

where $\hat{x}$ denotes the Fourier transform of x. The cross spectral density of the signals is then given by

$$\hat{C}(f) = \left(\mathbb{1} - \hat{G}(f)\right)^{-1} \hat{Z}(f) \left(\mathbb{1} - \hat{G}^*(f)\right)^{-1}$$

where $\hat{Z}(f)$ is the cross-spectral density of the external inputs. It follows

$$\begin{aligned} \hat{C}^{-1}(f) &= \left(\mathbb{1} - \hat{G}^*(f)\right)\hat{Z}^{-1}(f) \left(\mathbb{1} - \hat{G}(f)\right) \\ &= B^*(f)B(f) \end{aligned} \tag{2}$$

Table 1. Variables used for the estimation method and simulation.

Variable name	Symbol
node activity	$x(t)$
network connectivity	G
external inputs	$v(t)$
cross-spectral density	$\hat{C}(f)$
covariance of external input	$\hat{Z}(f)$
unitary matrix	U
L^1-norm cost function	Γ
gradient	d
initial matrix	B_0
Wiener process	$w(t)$
stationary covariance matrix	σ
simulation step	Δt
time constant of activity	τ
regularisation-controlling parameter for regularized ICOV	λ

https://doi.org/10.1371/journal.pcbi.1006056.t001

with $B(f) = \sqrt{\hat{Z}^{-1}(f)}\left(\mathbb{1} - \hat{G}(f)\right)$. In our model, we assume that the components of the external fluctuating input are pairwise stochastically independent for all nodes. Then, $\hat{Z}$ is a diagonal matrix, and we make the additional assumption that $\hat{Z} = \mathbb{1}$. For the linear model considered here, there is a relation between covariance and connectivity, which can be exploited for the estimation of connectivity from correlation. In the case $\hat{Z} = \mathbb{1}$ it is given by

$$\hat{C}^{-1}(f) = \left(\mathbb{1} - \hat{G}^*(f)\right)\left(\mathbb{1} - \hat{G}(f)\right) = \mathbb{1} - \hat{G}(f) - \hat{G}^*(f) + \hat{G}^*(f)\hat{G}(f)$$

where the last term contributes the information of the collider structures. If the matrix product $\hat{G}^*(f)\hat{G}(f)$ has a non-zero off-diagonal entry the corresponding nodes have outgoing connections terminating at the same node, which means these nodes form a collider. It is clear that for any unitary matrix $U \in \mathcal{U}(n)$ the product UB is still a solution of Eq (2), as $U^*U = \mathbb{1}$. We will resolve this ambiguity with an L^1 minimization which is known to prefer sparse solutions under certain conditions [19]. In order to find G from a given C we first fix an initial matrix B, and then search for a unitary matrix $U \in \mathcal{U}(n)$ such that $\|UB\|_1$ is minimal, so we are minimizing the function

$$\begin{aligned} \Gamma : \mathcal{U}(n) &\longrightarrow \mathbb{R} \\ U &\longmapsto \|UB\|_1. \end{aligned} \tag{3}$$

Gradient descent. To estimate the connectivity matrix from the covariance matrix we use a conjugate gradient descent algorithm similar to [20, 21] for minimizing the function $\Gamma(U)$ given in Eq (3), implemented in Python. For details please see S1 Text. For the gradient

$$d_{ij} = \frac{\partial \Gamma(U)}{\partial U_{i,j}}$$

of the cost function $\Gamma(U)$, $a = \frac{1}{2}(d - d^*)$ is skew-hermitian, and the matrix exponential of a skew-hermitian matrix is unitary. This means, starting in a point U_{act} and choosing an appropriate step size δ, we obtain a point $U_{\text{new}} = \exp(-\delta a)U_{\text{act}}$ with $\Gamma(U_{\text{new}}) < \Gamma(U_{\text{act}})$. In other words, the new point has a smaller L^1-norm than the old one and still satisfies the condition $\hat{C}^{-1} = B^* U_{\text{new}}^* U_{\text{new}} B$. Iterating this procedure until convergence leads to a point with locally minimal L^1-norm. The two conditions for convergence are inspired by [21]. The first one is a condition on the norm of the gradient. In each step, it is checked if

$$\|d - Ud^*U\|_F = \sqrt{\sum_{i,j} \left|d_{ij}(Ud^*U)_{ij}\right|^2} < \text{gtol}$$

is fulfilled, where $\|\ldots\|_F$ is the Frobenius norm and gtol > 0 is the convergence tolerance. As a second (alternative) condition, it is checked whether simultaneously

$$\frac{\|U - U_{\text{old}}\|_F}{\sqrt{N}} < \text{xtol} \qquad \text{and} \qquad \frac{|\Gamma(U_{\text{old}}) - \Gamma(U)|}{|\Gamma(U_{\text{old}} + 1)|} < \text{ftol}$$

are fulfilled. The values used are listed in Table 2. Before convergence the cost function typically oscillates around a certain value. To avoid stopping at a random phase of this oscillation, as a final step we apply a line-search, for details see S1 Text. The described gradient descent algorithm provides an efficient way for minimizing Eq (3). When calculating the gradient, we neglect the diagonal. Consequently, we also neglect the diagonal of the resulting estimated matrix, so we are not able to study self connections of the nodes.

Table 2. Parameter used for estimation.

Parameter	Value
xtol	$0.7 \cdot 10^{-2}$
ftol	$0.7 \cdot 10^{-4}$
gtol	$0.7 \cdot 10^{-2}$
κ	500
λ	5

https://doi.org/10.1371/journal.pcbi.1006056.t002

Initial condition. As starting condition for the gradient descent we use a matrix $B_0(f)$ such that

$$B_0^*(f)B_0(f) = \hat{C}^{-1}(f).$$

There are many ways to choose a B_0 with this property, we found the following choice efficient: As $\hat{C}$ is the cross-spectral density it is positive definite, and so is $\hat{C}^{-1}$. Thus, there is exactly one positive definite square root of $\hat{C}^{-1}$ [22] which can be calculated by

$$B_0 = \sqrt{\hat{C}^{-1}} = W\sqrt{E}W^* \tag{4}$$

where the columns of W are the eigenvectors of $\hat{C}^{-1}$, and E is the matrix with the corresponding eigenvalues of $\hat{C}^{-1}$ on the diagonal. Thus we initialize the gradient descent with $U_0 = \mathbb{1}$ and B_0 given by Eq (4).

Step size selection. A critical part of the optimization is the selection of an appropriate step size. If the step size is too large, one might miss the minimum. If the step size is too small, the optimization converges very slowly. For the gradient descent, we use an adaptive scheme inspired by [20], where the step-size depends on the largest eigenvalue of the actual gradient: Let $\lambda_{\max}$ be the largest eigenvalue of d_{act}, the step size is given by

$$\delta_{\text{act}} = \frac{2\pi}{|\lambda_{\max}| \cdot \kappa}$$

where κ is constant. The intuition behind that, is that smooth cost functions along a geodesic on the unitary manifold are almost periodic. So the step size should be a fraction of the period of this function. This is achieved by the scaling with the largest eigenvalue, which allows us to take a scale-invariant fraction of this period.

Validation methods

Noise-free covariance matrices. We assume that the interactions among neuronal populations can be described by a linear model, see Eq (2) with $Z = \mathbb{1}$. This model allows us to derive a relation between the connectivity matrix of the network G and the inverse cross-spectral density matrix $\hat{C}^{-1}$ of the measured activity

$$\hat{C}^{-1} = (\mathbb{1} - \hat{G}^*)(\mathbb{1} - \hat{G}) = \mathbb{1} - \hat{G}^* - \hat{G} + \hat{G}^*\hat{G}. \tag{5}$$

Given a sampled connectivity matrix G we can calculate the inverse covariance matrix directly using Eq (5). For all simulations, half of the connections were negative (inhibitory) connections, the absolute strength was the same for all connections and 20 repetitions were simulated. As connectivity profiles we used random Erdős-Rényi networks.

Ornstein-Uhlenbeck processes. To validate our inference procedure before applying it to the network inference from measurements of neuronal activity we simulated stationary signals. Since there is no gold-standard for simulations of fMRI data [23], we based our simulations on the Ornstein-Uhlenbeck process [24], which provides a simple linear model for neural activity.

$$dx(t) = Ax(t)dt + dW(t) \tag{6}$$

where A is a matrix and W a Wiener process. In our applications, we parametrize this matrix as $A = \frac{1}{\tau}(G - \mathbb{1})$ with real-valued connectivity matrix G and time constant τ. For this process, it is possible to calculate the stationary covariance matrix Σ from the continuous Lyapunov equation

$$\mathbb{1} = A\Sigma + \Sigma A^T.$$

In fact, we simulated the process in discrete time. In analogy with [25] we use a multivariate version of the exact update formula

$$x(t + \Delta t) = e^{A\Delta t}x(t) + n(t), \tag{7}$$

where $n(t) \sim N(0, \Sigma)$ is normally distributed, with Σ being the stationary covariance matrix described above. As a final step, we filter the time series $x(t)$ with the canonical hemodynamic response function (HRF) [26, 27]. To match the data obtained in brain scans sampled at a temporal resolution of $\Delta t = 0.1$ s, we used random connectivity profiles G with a connection probability $p = 0.1$ (Erdős-Rényi model), 50% negative entries, and a spectral radius of $\rho = 0.3$. All parameters used are listed, once more, in Table 3. Before calculating the covariance C, the data is standardized such that the mean is 0 and the variance is 1 for all components of the time series. We add normally distributed observation noise u_{obs} with a $\mathcal{N}(0, \sigma_{\text{obs}})$ distribution to the simulated signal. After the simulation we calculated the signal-to-noise ratio according to

$$\text{SNR} = \frac{\sigma_X^2}{\sigma_{\text{obs}}^2}$$

where σ_X^2 denotes the variance of the signal.

Performance measures

When estimating connectivity from simulations with known underlying network structure (ground truth), one can quantify the performance of the estimation. For measuring the accuracy of our estimation we employ three different methods. First, we use the area under the ROC-curve (AUC). The ROC (receiver operating characteristic) curve is obtained as following:

Table 3. Parameter used for simulations.

Parameter	Value
repetitions	20
network type	Erdős-Rényi
N	100
p	0.1
T	350000 s
dt	0.1 s
τ	0.1 s
ρ	0.3

https://doi.org/10.1371/journal.pcbi.1006056.t003

For each possible parameter value (in our case the threshold for the existence of a connection), the number of true-positives (TP) and false-positives (FP) is used to calculate the true-positive rate (or recall) TP/(TP+ FN) and the false-positive rate FP/(FP+ TN). The ROC curve is then obtained by plotting the true-positive-rate against the false-positive rate. Secondly, we use the average precision score (PRS) which is the area under the precision-recall curve. This also includes the false-negatives (FN) (precision: $\frac{TP}{TP+FP}$). If both AUC and PRS are equal to 1, the connections in the network are perfectly estimated. Sample curves are shown in Fig 2D. Thirdly, we calculate the Pearson Correlation Coefficient (PCC) which in contrast to the measures defined before also take the strength and the sign of the interactions into account. This also means that this measure is less suited to assess whether a connection exists or not. It rather measures whether the estimated connections have the same strength as the original ones. We consider all three performance measures simultaneously to establish the quality of our estimates.

Experimental fMRI data

Seven healthy subjects underwent a 20-minute resting-state fMRI experiment on a 3 T Siemens Prisma scanner. The data was acquired using the MREG sequence [28], yielding a high temporal resolution (TR = 0.1 s, 12000 time points) that facilitates functional connectivity analyses [29]. The other sequence parameters were TE = 36 ms, FA = 25˚, 64 × 64 × 50 matrix and 3mm isotropic voxel size. Additionally, cardiac and respiratory signals were recorded with the ECG and abdominal breathing band from the scanner's physiological monitoring unit. Motion correction was done with FSL and physiological noise correction was performed with RETROICOR [30]. Average CSF and white matter signals were regressed out, but no global signal regression was performed. Following image normalization to MNI space, voxels were parcellated according to the AAL atlas (excluding the cerebellum), and the mean activity within each atlas region was calculated. The connectivity was then estimated using zero-lag covariances of the standardized signals.

Ethics statement. The experiments have been approved by the Ethics Committee of the University Medical Center Freiburg.

Results

Noise-free covariance matrices

Intrinsic properties of our new estimation procedure can be identified by studying the performance of the method for perfectly estimated (noise-free) covariance matrices. This way we address properties that do not depend on any particular feature of the underlying data, and that are not due to the success of the measurement process. In particular, we show for which types of networks our estimation procedure gives good results on technical grounds, with a wide range of networks hopefully including those arising in applications. We used random Erdős-Rényi connectivity profiles for all simulations. The macro-connectivity between neuronal populations has to satisfy certain conditions in order to be tractable by our methods. Two of these conditions concern the dynamic stability of the network and the strength of the interactions. There is a trade-off between the number of physical links and the resulting strength of macro-connections, and the dynamic stability of the network. To study the performance of our method in these various regimes, we separately varied the network size N, the connection probability p, and the absolute strength of connections $|J|$ in the connectivity matrix G, while the fraction of inhibitory couplings was kept at 50%. The spectral radius ρ of

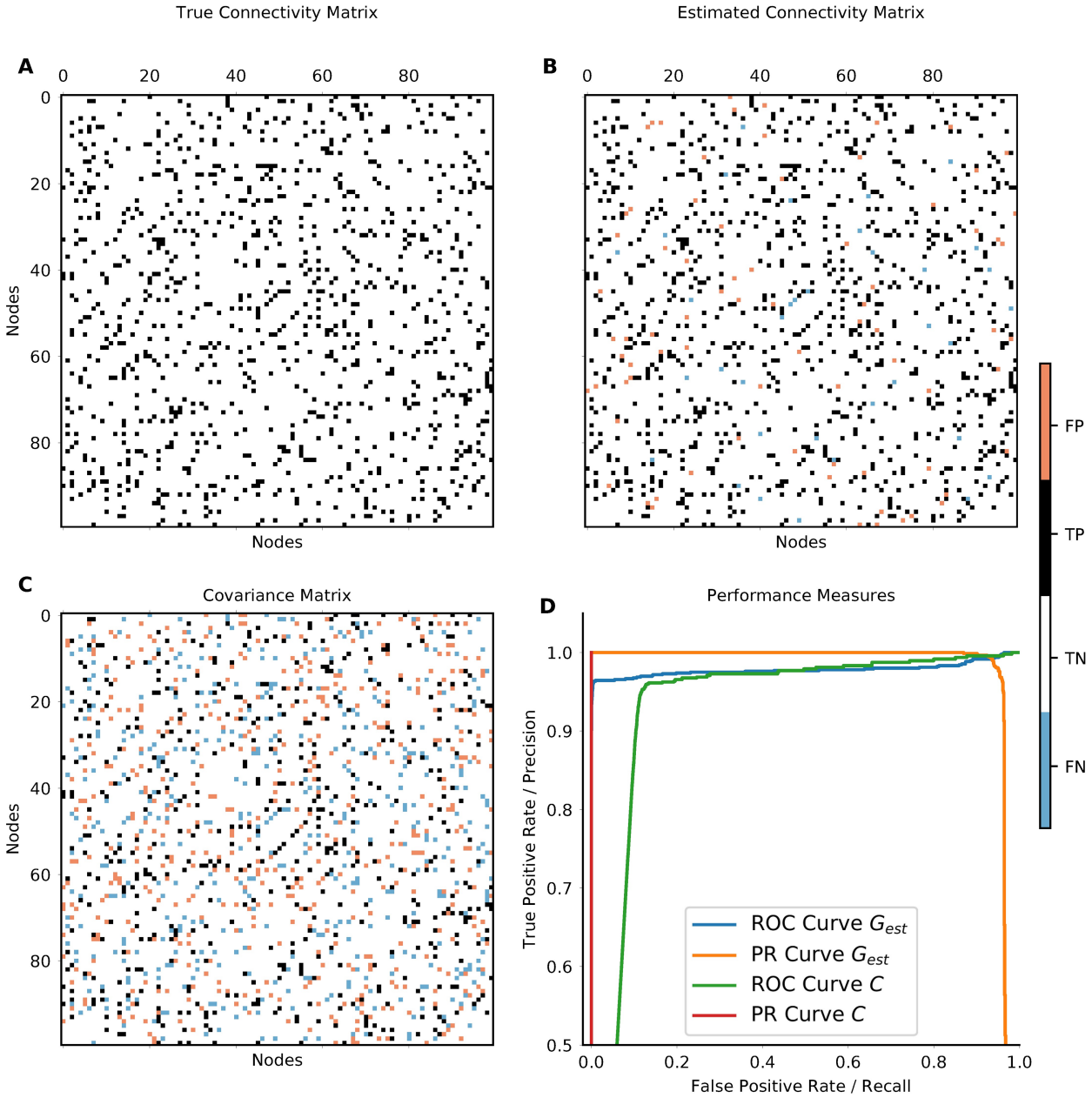

Fig 2. Networks inferred from a simulated Ornstein-Uhlenbeck process. A shows the original network. **B** shows the network inferred with our new method from the zero-lag covariances. White and black entries indicate true negative (TN) and true positive (TP) connections, blue and red entries indicate false negative (FN) and false positive (FP) connections, respectively. In this example, the performance measures are AUC = 0.98, PRS = 0.97 and PCC = 0.95. **C** depicts the sample covariance (functional connectivity) matrix directly estimated from the data. In **C**, as a consequence of symmetry, the number of wrongly estimated connections is quite high, the performance measures are AUC = 0.93, PRS = 0.54, and PCC = 0.29. **D** shows the Receiver Operating Characteristic Curve and the Precision Recall Curve for the networks estimated from zero-lag covariance G_{est} in blue/orange and of the functional connectivity C in red/cyan. The areas under these curves are the AUC and PRS, respectively.

https://doi.org/10.1371/journal.pcbi.1006056.g002

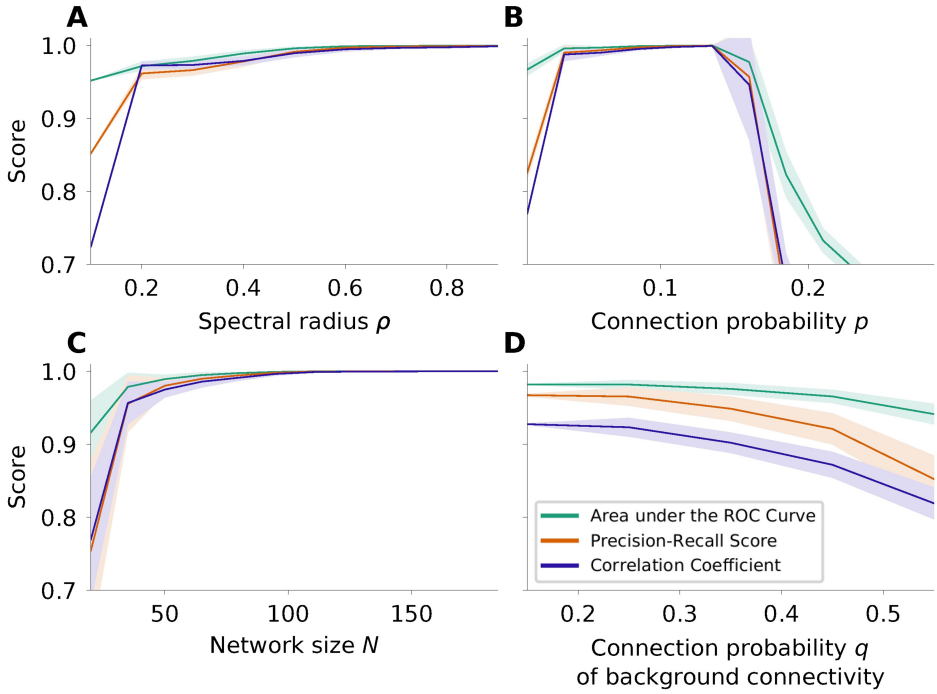

Fig 3. Effects of spectral radius ρ, connection probability p and network size N. Here we consider the case of noise-free covariance matrices, which were created based on the theory of the underlying model. The quantities considered are the area under the ROC curve (AUC; green), the precision recall score (PRS; orange) and the Pearson correlation coefficient (PCC; purple). The shaded areas indicate the mean ± standard deviation computed over 20 realizations. **A** If the network interaction is larger than $\rho_{\min}$, it has relatively little effect on the performance of the estimation. Even in the extreme case, where ρ is close to 1, the estimation works well. **B** Performance of the estimation for different sparsity levels, encoded by the respective connection probabilities p. As expected, for non-sparse networks the performance of the algorithm degrades dramatically. **C** Performance of the estimation for increasing network size. Our results indicate clearly that bigger networks can be better reconstructed. Applicability may be limited by the numerical effort associated with the optimization. **D** Performance of the estitmation in presence of weak background connections. It is nevertheless possible to infer the skeleton of strong connections with high fidelity.

https://doi.org/10.1371/journal.pcbi.1006056.g003

the bulk eigenvalue spectrum is approximately given by

$$\rho^2 = J^2 p(1-p)N. \tag{8}$$

The default values of the parameters used in our study were $N = 100$, $p = 0.1$ and $\rho = 0.7$, where only one of them at a time was systematically varied. Low values of the spectral radius ρ correspond to networks with weak recurrent interaction and high values to networks with strong interaction, respectively. According to the model of network interaction assumed here, the networks need to have a spectral radius $\rho > 0$ for network interaction to be present and $\rho < 1$ for the dynamics to be stable. First, our results in Fig 3A indicate that a certain minimal level of interaction is necessary to be able to estimate the connections reliably. Above a value of $\rho_{\min} = 0.2$, the influence of the spectral radius on the performance of the estimation is weak, but the larger the spectral radius is the better the estimation gets.

Secondly, the connection probability of the network influences the quality of the estimation. For all connection probabilities tested here the network size was kept constant at $N = 100$ nodes. The networks were constructed such that the strength $|J|$ of all connections was the same and such that the spectral radius ρ was constant according to Eq (8). Fig 3B shows that the estimation works very well for sparse matrices with a connection probability in the range

between 5% and 15%. For networks with higher connection probability and equally strong connections, the performance decreases as expected, due to the bias associated with L^1-minimization. But even for a connection probability of $p = 0.21$, a fraction of 14.2% of the estimated connections are false negative, and 3.3% are false positive. More than 90% of the correctly estimated connections have the correct sign. In applications, the focus of the estimation often lies on the strongest connections in the network. In networks with a background of weak connections and a sparse skeleton of stronger connections, it is possible to selectively estimate these strong links although, strictly, the assumption of a sparse network is violated. Fig 3D shows the performance of our method for such networks: the networks consist of a skeleton of strong connections with connection probability $p = 10\%$ and a connection strength derived from Eq (8) for $\rho = 0.7$. Additionally, we created a second network with weaker connections for various connection probabilities q. The two networks were combined by adding the connectivity matrices. The connection strength of this weaker connections is also derived from Eq (8), with a spectral radius of the background network being 20% of the spectral radius of the skeleton network. Then the performance of the estimation is calculated with respect to the skeleton of strong connections.

Thirdly, to be applicable to a broad range of data types, a method of connectivity estimation should perform stable for different network sizes N. For most common types of non-invasive recordings of population activity the number of nodes considered is in the range between 30 and 150. It is, of course, possible to consider larger networks, although the estimation becomes computationally more expensive. The runtime of the algorithm for networks with 200 nodes still in the range of seconds on a state-of-the-art desktop computer, but even networks with 1000 nodes or more are tractable. The strength of the connections $|J|$ are set such that the spectral radius ρ of G is constant; the connection probability is constant at $p = 0.1$. Fig 3C shows that our method performs better for bigger networks. We have observed that the L^1 cost landscape becomes smoother for larger networks.

Ornstein-Uhlenbeck processes as model for BOLD signals

In order to create surrogate data which fit fast fMRI data [28], we simulated interacting stochastic processes known as Ornstein-Uhlenbeck processes. In this case, the performance of the network inference depends on how well the inverse covariance matrix, which is the basis of the estimation, can be derived from the data. In addition to finite size effects, we studied the impact of observation noise on the performance, see Fig 4. We used $N = 100$, $p = 0.1$, dt = 0.1 s, $\rho = 0.74$ and $\tau = 0.1$ s as default values of the parameters. Generally, it seems natural to use Welch's method to calculate cross-spectral densities directly, and then to estimate the connectivity for each frequency band separately. For the data described here, however, we can estimate the connectivity from zero-lag sample covariances in the time domain. This is possible when the mass of the covariance function is concentrated very close around lag 0. Then lag 0 is the only one contributing to the integral of the covariance function, which corresponds to the cross-spectral density $\hat{C}(0)$. As shown in Fig 4A, with noisy data the AUC is still good, but the PRS is lower than in the case, where the covariance is known without error. However, for a signal-to-noise ratio above 1 the performance improves very quickly.

In the case of fMRI usually the whole brain is scanned, and there are no unobserved nodes in the network. However, for other data types (e.g. fNIRS) only parts of the brain can be observed. The question then is, whether this sub-network can nevertheless be reconstructed from the recorded signals. To model this scenario, we took simulated data and removed randomly a certain subset of components from the dataset. The interaction of the removed nodes is then not part of the covariance matrix of the reduced dataset, although the unobserved

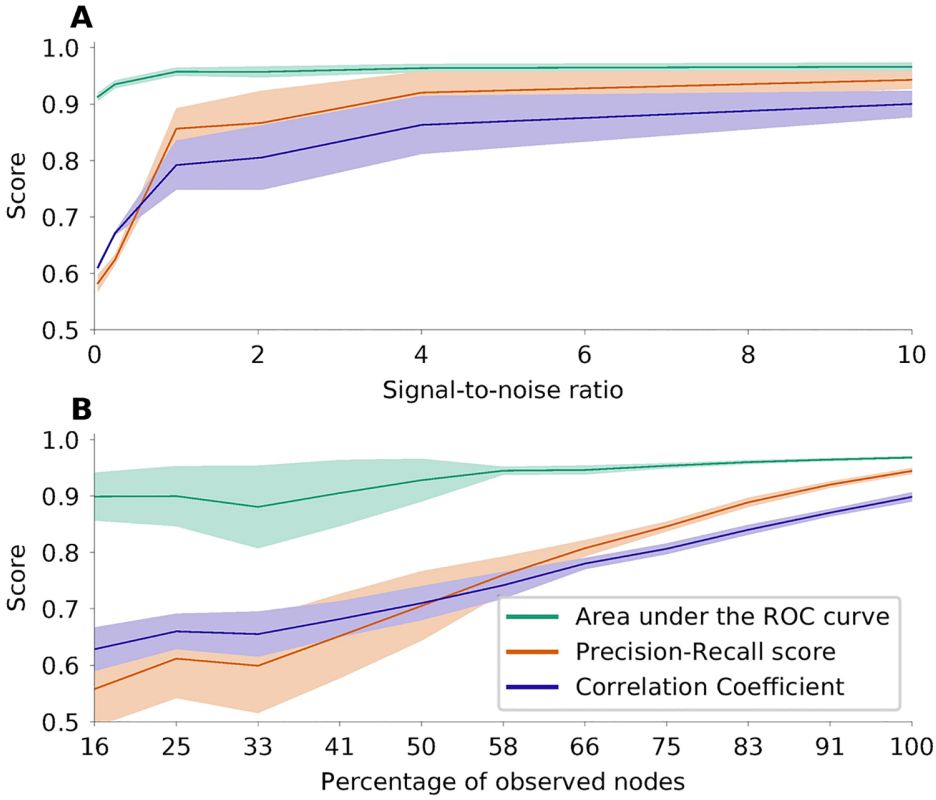

Fig 4. Performance of network inference based on simulated Ornstein-Uhlenbeck processes. Same colors as in Fig 3. **A** Performance of the estimation when measurement noise is added. **B** Performance of the estimation if only parts of the network are observable. The fraction of observed nodes in a network are indicated on the *x*-axis. The total number of nodes in the network was $N = 180$.

https://doi.org/10.1371/journal.pcbi.1006056.g004

nodes of course still exert their influence on the observed ones. The performance of the estimation of the sub-network based on the reduced dataset is shown in Fig 4B. Our analysis shows very clearly that the estimation still leads to reasonable results under these conditions. In fact, we can demonstrate that we are inferring causal connections only: For unconnected observed nodes X, Y and a latent node L connected to both X and Y, our method does not erroneously indicate a link between X and Y. One key factor for a reliable estimation of the covariance matrix is the amount of data available. This depends on the length of the measurement or simulation, and on the sampling rate. Since fast fMRI time series are obtained by measuring the BOLD response as a proxy of neuronal activity, the time scale of the measured data is relatively slow compared to the time scale of the underlying neuronal activity. Fig 5 shows the performance of network inference depending on the amount of data available, and on the time-scale of the neuronal activity. Not surprisingly, the more voluminous the dataset is, the better the estimation gets. On the other hand, it shows that the estimation generally leads to better results for slower temporal dynamics. Also, for data of sufficient length with a fairly good signal-to-noise ratio, the estimation of the connectivity is possible even when only a part of the network is observed. To allow comparison of our new method with other known methods for network inference [31–33], we applied it to the *NetSim* dataset provided by [34]. For details on the result of this, please see Fig A in S1 Text.

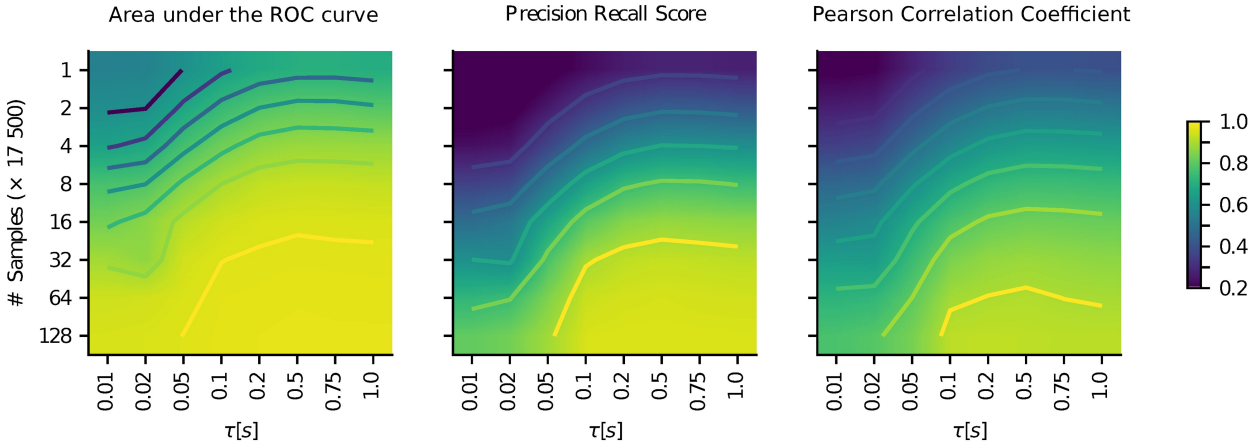

Fig 5. Performance (color coded) of the estimation depending on data length (*y*-axis) and time scale of the activity (*x*-axis). Both scales are logarithmic. For interpolation a bilinear method is used.

https://doi.org/10.1371/journal.pcbi.1006056.g005

fMRI data

We estimated connectivity from seven fast fMRI datasets, for details see the methods section. The resulting networks, after a threshold of 10% was applied, consist of 810 connections for each dataset. The threshold of 10% was chosen arbitrarily. In Fig B in S1 Text we show the histogram of estimated connection strengths for all seven reconstructed networks before thresholding. The threshold is derived from the 10% strongest connections, disregarding their signs. As there is generally no full ground truth for the connectivity inferred from human fMRI recordings available [31, 32, 35], we cannot definitely assess the degree to which the result of our inference are correct. We can, however, establish whether they are plausible. One representative connectivity matrix is shown in Fig 6. On average, 34% of the connections were inhibitory, with negligible variability across subjects. Of all connections found, 301(37%) were found in four subjects or more, and 4872 out of 8100 possible connections were absent in all subjects. On average, 245 of the connections were bi-directional and 565 connections were identified only for one direction. In general, close-by areas are more likely to be connected than more distant ones. This fact is (approximately) represented by a concentration of connections along secondary diagonals in the within-hemisphere blocks. Also, there are frequent inter-hemispheric connections between corresponding areas. This fact is represented by the diagonal entries in the across-hemisphere blocks.

Comparison with the Regularized Inverse Covariance (RIC) method

As mentioned above, different heuristics have been suggested to reconstruct networks from neuronal signals. In Fig 7 we compare the performance of the new method we propose here and the established method of Regularized Inverse Covariance [34], based on the implementation provided at https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets. Our comparison clearly shows that our new method performs significantly better than the Regularized Inverse Covariance method, mainly, because the latter cannot establish the direction of connections. The superior performance of the new method is reflected in higher values for all three performance measures, in particular PRS and PCC. As regularization parameter required by the software toolbox, we used $\lambda = 5$.

Furthermore, we applied the RIC method on all seven MREG datasets described before. A threshold was applied, such that only the 10% strongest connections are retained. To compare

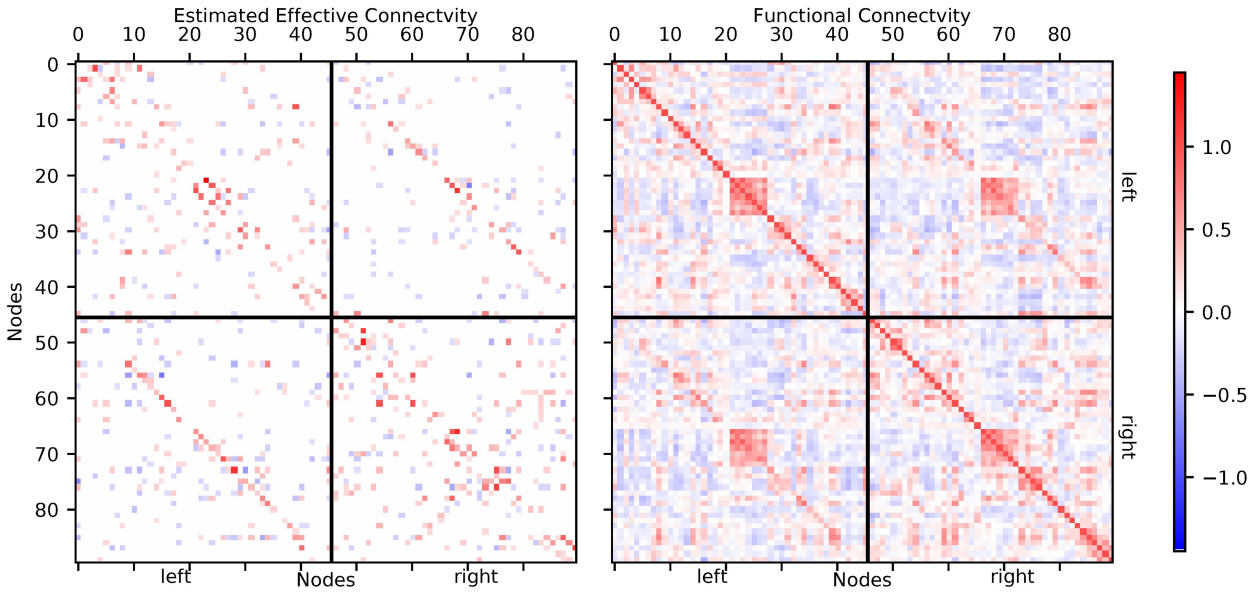

Fig 6. Left panel: Directed connectivity estimated with our new method from one sample MREG data set. Voxels were parceled using the AAL90-atlas. In the top-left block of the connectivity matrix connections within the left hemisphere are shown, in the lower-right block connections within the right hemisphere. The off-diagonal blocks represent the inter-hemispheric connections from the left to the right hemisphere (lower left) and from the right to the left hemisphere (top right). The strength of all connections is color coded, with red representing positive (excitatory) connections and blue representing negative (inhibitory) connections. Only the strongest 10% of connections are shown. Right panel: Functional connectivity matrix derived from the same data.

https://doi.org/10.1371/journal.pcbi.1006056.g006

the outcome of both methods, we only condidered the existence of connections (binary and symmetric connectivity) and disregarded weights and directions (weighted nonsymmetric connectivity). One representative example of the comparison of both methods is shown in Fig 8. For RIC, 376 out of 810 possible connections where identified in four subjects or more out

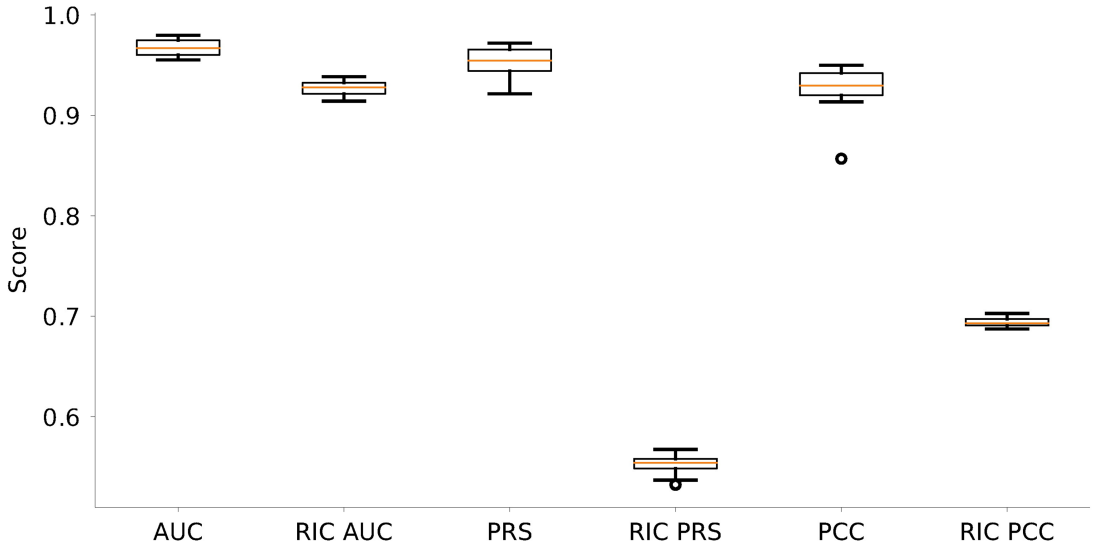

Fig 7. Comparison of performance with the Regularized Inverse Covariance (RIC) method based on numerical simulations of Ornstein-Uhlenbeck processes. Shown are the results from the reconstruction of 20 different networks with Erdős-Rényi connectivity profiles as described before (cf. Fig 2). AUC, PRS and PCC of our new method and of the RIC method, respectively, are shown side-by-side.

https://doi.org/10.1371/journal.pcbi.1006056.g007

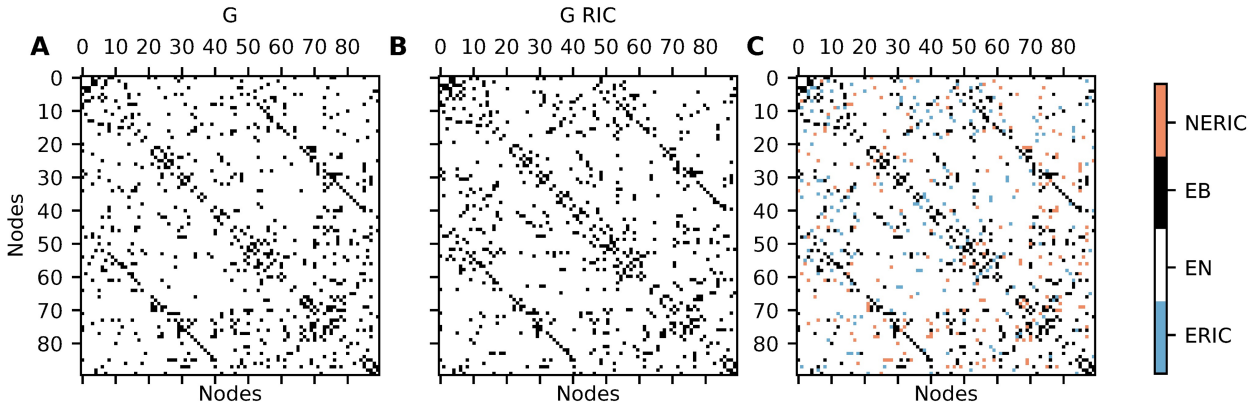

Fig 8. Estimated networks for one representative MREG dataset. The left panel shows the symmetrized network reconstructed with our estimation method, the middle panel shows the network found with the RIC method. The right panel shows the connections which are identified by both methods (EB, black), by none of the methods (EN, white), the connections found only by the RIC (ERIC, blue) and the connections found only by our method, but not by the RIC method (NERIC, red).

https://doi.org/10.1371/journal.pcbi.1006056.g008

of seven, the corresponding number for our method is 392 out of 810 possible connections. If any method produced directed networks with 10% connection probability at random, this would yield an average count of less than 25 connections (3% of 810 connections) that agree for least four out of seven independently generated networks. On average, 290.5 out of 4050 possible connections (undirected) are identified by both methods, 3530.5 connections were found by neither of the methods. This means that both methods agree on 3821 out of 4050 connections on average. The two methods disagreed on the remaining 229 connections.

Discussion

With the described method we can estimate directed and signed effective connectivity between neural populations from measured brain signals, based on zero-lag covariances only. To investigate the reliability of our estimated connections we used simulations of Ornstein-Uhlenbeck processes mimicking BOLD-related signals generated by interacting neuronal populations. Our method shows very good performance, if enough data is available and the observation noise is not too strong. Also, even in cases with relatively poor performance (e.g. if the network is too dense) more than 90% of the estimated connections have the correct sign. Applying the method on measured fast fMRI data, we found that about 34% of all identified connections have an inhibitory effect on their respective target population. In general, inhibitory synapses are mainly formed within local populations, and typically do not project to distant targets. An inhibitory connection between populations, however, can also be achieved by excitatory neurons preferentially terminating on the inhibitory neurons of the target region. The comparison with the Regularized Inverse Covariance (RIC) method shows good agreement with regard to the existence of connections. Directions cannot be disambiguated with the RIC method. Our results based on simulated surrogate data reflect what one would expect from the design of an estimation procedure. For large, sparse networks with sufficiently strong interaction, our network estimation procedure works reliably. However, as expected if the network is not sparse, or the time series is too short, the quality of the estimate drops. Nevertheless, in most cases the main interest lies on the strongest connections, which can be reliably estimated with our method even when the network is not sparse. For the experimental data shown, individual connections may be unreliable because of the limited size of the dataset. Also, it is unclear whether the biological network to be analyzed is really sparse, and if the assumption of

pairwise independent external input is really justified. On the other hand, due to the higher likelihood of a coupling between close-by areas and between inter-hemispheric counterparts, the resulting network looks plausible. For interpreting individual connections longer recordings would certainly be beneficial. Also, one could then use temporal information from additional frequency bands. Of high interest is also the comparison with structural measures as the ones obtained by diffusion tensor imaging. To the best knowledge of the authors, this is the first time that effective whole-brain connectivity has been estimated from zero-lag covariances. Other methods [8] rely on lagged covariances, where the correct lag parameter is critical, and needs to be inferred from the exponential decay of the observed auto-covariances. Also, our proposed method is the only one that can detect directed inhibitory connections on the whole-brain scale. The estimation procedure is fast and easy to apply. As it uses no temporal information, our method can also be applied on other data types that rely on the BOLD effect, e.g. fNIRS, but also data types measuring electrical population activity directly. This makes it a good candidate for, among other things, studying changing connectivity in neurodegenerative diseases, like Parkinson's or Alzheimer's.

Conclusion

With the presented method we can estimate directed effective connectivity on a whole-brain scale. Also we are able to detect whether connections are excitatory or inhibitory. The estimation is possible based on zero-lag covariances, but can also be applied to frequency-resolved cross spectral densities.

Supporting information

S1 Text. Supporting information.
(PDF)

Acknowledgments

We thank Uwe Grauer from the Bernstein Center Freiburg as well as Bernd Wiebelt and Michael Janczyk from the Freiburg University Computing Center for their assistance with HPC issues.

Author Contributions

Conceptualization: Jürgen Hennig, Pierre LeVan, Stefan Rotter.

Data curation: Carolin Lennartz, Pierre LeVan.

Formal analysis: Jonathan Schiefer, Alexander Niederbühl, Volker Pernice, Carolin Lennartz.

Funding acquisition: Jürgen Hennig, Stefan Rotter.

Investigation: Jonathan Schiefer, Alexander Niederbühl, Volker Pernice, Carolin Lennartz, Pierre LeVan.

Methodology: Volker Pernice.

Project administration: Stefan Rotter.

Resources: Jürgen Hennig, Stefan Rotter.

Software: Jonathan Schiefer, Alexander Niederbühl, Volker Pernice, Carolin Lennartz.

Supervision: Volker Pernice, Jürgen Hennig, Pierre LeVan, Stefan Rotter.

Validation: Jonathan Schiefer, Alexander Niederbühl, Volker Pernice.

Visualization: Jonathan Schiefer, Alexander Niederbühl.

Writing – original draft: Jonathan Schiefer.

Writing – review & editing: Jonathan Schiefer, Alexander Niederbühl, Volker Pernice, Carolin Lennartz, Pierre LeVan, Stefan Rotter.

References

1. Li B, Razi A, Friston KJ. Editorial: Mapping Psychopathology with fMRI and Effective Connectivity Analysis. Front. Hum. Neurosci. 2017; 11:151. https://doi.org/10.3389/fnhum.2017.00151 PMID: 28408874
2. Friston KJ. Functional and effective connectivity: a review. Brain connectivity. 2011; 1(1):13–36. https://doi.org/10.1089/brain.2011.0008 PMID: 22432952
3. Pernice V, Staude B, Cardanobile S, Rotter S. How structure determines correlations in neuronal networks. PLoS Comput. Biol. 2011; 7,. https://doi.org/10.1371/journal.pcbi.1002059 PMID: 21625580
4. Havlicek M, Roebroeck A, Friston KJ, Gardumi A, Ivanov D, Uludag K. On the importance of modeling fMRI transients when estimating e ff ective connectivity: A dynamic causal modeling study using ASL data. NeuroImage. 2017; 155(July 2016):217–233. https://doi.org/10.1016/j.neuroimage.2017.03.017 PMID: 28323165
5. Stephan KE, Friston KJ. Analyzing effective connectivity with fMRI. Wiley interdisciplinary reviews Cognitive science. 2010; 1(3):446–459. https://doi.org/10.1002/wcs.58 PMID: 21209846
6. Smith JF, Pillai A, Chen K, Horwitz B. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. NeuroImage. 2010; 52 (3):1027–1040. https://doi.org/10.1016/j.neuroimage.2009.11.081 PMID: 19969092
7. Freestone DR, Karoly PJ, Nešić D, Aram P, Cook MJ, Grayden DB. Estimation of effective connectivity via data-driven neural modeling. Front. Neurosci. 2014; 8:383. https://doi.org/10.3389/fnins.2014.00383 PMID: 25506315
8. Gilson M, Moreno-Bote R, Ponce-Alvarez A, Ritter P, Deco G. Estimation of Directed Effective Connectivity from fMRI Functional Connectivity Hints at Asymmetries of Cortical Connectome. PLoS computational biology. 2016. https://doi.org/10.1371/journal.pcbi.1004762
9. Ting CM, Seghouane AK, Member S, Salleh SH, Noor AM. Estimating Effective Connectivity from fMRI Data Using Factor-based Subspace Autoregressive Models. IEEE Signal Processing Letters. 2015; 22 (6):757–761. https://doi.org/10.1109/LSP.2014.2365634
10. Roebroeck A, Formisano E, Goebel R. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. NeuroImage. 2011; 58(2):296–302. https://doi.org/10.1016/j.neuroimage.2009.09.036 PMID: 19786106
11. Mehta-Pandejee G, Robinson PA, Henderson JA, Aquino KM, Sarkar S. Inference of direct and multistep effective connectivities from functional connectivity of the brain and of relationships to cortical geometry. Journal of Neuroscience Methods. 2017; 283:42–54. https://doi.org/10.1016/j.jneumeth.2017.03.014 PMID: 28342831
12. Marrelec G, Krainik A, Duffau H, Doyon J, Benali H. Partial correlation for functional brain interactivity investigation in functional MRI. NeuroImage. 2006; 32:228–237. https://doi.org/10.1016/j.neuroimage.2005.12.057 PMID: 16777436
13. Timme M, Casadiego J. Revealing networks from dynamics: an introduction. Journal of Physics A: Mathematical and Theoretical. 2014; 47(34):343001. https://doi.org/10.1088/1751-8113/47/34/343001
14. Gates KM, Molenaar PCM. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. NeuroImage. 2012; 63(1):310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026 PMID: 22732562
15. Ramsey J, Glymour M, Sanchez-Romero, Glymour C. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. International Journal of Data Science and Analytics. 2016. https://doi.org/10.1007/s41060-016-0032-z PMID: 28393106
16. Pernice V, Rotter S. Reconstruction of sparse connectivity in neural networks from spike train covariances. Journal of Statistical Mechanics: Theory and Experiment. 2013; 2013(03):P03008. https://doi.org/10.1088/1742-5468/2013/03/P03008
17. Pearl J. Causality: Models, Reasoning, and Inference. New York, NY, USA: Cambridge University Press; 2000.

18. Rebane G, Pearl J. The recovery of causal poly-trees from statistical data. Proceedings of the Third Workshop on Uncertainty in AI; 1987; 222–228
19. Candes EJ, Tao T. Decoding by Linear Programming. IEEE Trans Inf Theor. 2005; 51(12):4203–4215. https://doi.org/10.1109/TIT.2005.858979
20. Abrudan T, Eriksson J, Koivunen V. Conjugate Gradient Algorithm for Optimization Under Unitary Matrix Constraint. Signal Processing. 2009; 89:1704–1714. https://doi.org/10.1016/j.sigpro.2009.03.015
21. Wen Z, Yin W. A feasible method for optimization with orthogonality constraints. Mathematical Programming. 2013; 142(1):397–434. https://doi.org/10.1007/s10107-012-0584-1
22. Horn RA, Johnson CR. Matrix Analysis. Cambridge University Press, Cambridge, MA, 1985
23. Welvaert M, Rosseel Y. A review of fMRI simulation studies. PLoS ONE. 2014; 9(7):e101953. https://doi.org/10.1371/journal.pone.0101953 PMID: 25048024
24. Gardiner CW. Handbook of stochastic methods for physics, chemistry and the natural sciences. vol. 13 of Springer Series in Synergetics. 3rd ed. Berlin: Springer-Verlag; 2004.
25. Gillespie DT. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. Phys Rev E. 1996; 54:2084–2091. https://doi.org/10.1103/PhysRevE.54.2084
26. Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. Event-Related fMRI: Characterizing Differential Responses. NeuroImage. 1998; 7(1):30–40. https://doi.org/10.1006/nimg.1997.0306 PMID: 9500830
27. Glover GH. Deconvolution of Impulse Response in Event-Related BOLD fMRI1. NeuroImage. 1999; 9(4):416–429. https://doi.org/10.1006/nimg.1998.0419 PMID: 10191170
28. Assländer J, Zahneisen B, Hugger T, Reisert M, Lee HL, LeVan P, Hennig J. Single shot whole brain imaging using spherical stack of spirals trajectories. NeuroImage. 2013; 73:59–70. https://doi.org/10.1016/j.neuroimage.2013.01.065
29. LeVan P, Akin B, Hennig J. Fast imaging for mapping dynamic networks. NeuroImage. 2017 https://doi.org/10.1016/j.neuroimage.2017.08.029
30. Glover GH, Li TQ, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magnetic Resonance in Medicine. 2000; 44(1):162–167. https://doi.org/10.1002/1522-2594(200007)44:1%3C162::AID-MRM23%3E3.0.CO;2-E PMID: 10893535
31. Nie L, Yang X, Matthews PM, Xu ZW, and Guo YK. Inferring functional connectivity in fMRI using minimum partial correlation. International Journal of Automation and Computing, 14:371–385, 2017. https://doi.org/10.1007/s11633-017-1084-9
32. Ryali S, Chen T, Supekar K, Tu T, Kochalka J, Cai W and Menon V. Multivariate dynamical systems-based estimation of causal brain interactions in fMRI: Group-level validation using benchmark data, neurophysiological models and human connectome project data. Journal of Neuroscience Methods, 268:142–153, 2016. https://doi.org/10.1016/j.jneumeth.2016.03.010 PMID: 27015792
33. Hyvärinen A and Smith SM. Pairwise likelihood ratios for estimation of non-gaussian structural equation models. Journal of Machine Learning Research, 14(1):111–152, 2013.
34. Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. Network modelling methods for fMRI. NeuroImage 2011; 54;2:875–891. https://doi.org/10.1016/j.neuroimage.2010.08.063 PMID: 20817103
35. Zaghlool SB and Wyatt CL. Missing data estimation in fMRI dynamic causal modeling. Frontiers in Neuroscience, 8:191, 2014. https://doi.org/10.3389/fnins.2014.00191 PMID: 25071435

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 08 May 2018
doi: 10.3389/fnins.2018.00287

Sparse Estimation of Resting-State Effective Connectivity From fMRI Cross-Spectra

Carolin Lennartz[1,2]*, *Jonathan Schiefer*[2,3], *Stefan Rotter*[2,3], *Jürgen Hennig*[1,2] and *Pierre LeVan*[1,2]

[1] Department of Radiology, Medical Physics, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [2] BrainLinks-BrainTools Cluster of Excellence, University of Freiburg, Freiburg, Germany, [3] Bernstein Center Freiburg & Faculty of Biology, University of Freiburg, Freiburg, Germany

In functional magnetic resonance imaging (fMRI), functional connectivity is conventionally characterized by correlations between fMRI time series, which are intrinsically undirected measures of connectivity. Yet, some information about the directionality of network connections can nevertheless be extracted from the matrix of pairwise temporal correlations between all considered time series, when expressed in the frequency-domain as a cross-spectral density matrix. Using a sparsity prior, it then becomes possible to determine a unique directed network topology that best explains the observed undirected correlations, without having to rely on temporal precedence relationships that may not be valid in fMRI. Applying this method on simulated data with 100 nodes yielded excellent retrieval of the underlying directed networks under a wide variety of conditions. Importantly, the method did not depend on temporal precedence to establish directionality, thus reducing susceptibility to hemodynamic variability. The computational efficiency of the algorithm was sufficient to enable whole-brain estimations, thus circumventing the problem of missing nodes that otherwise occurs in partial-brain analyses. Applying the method to real resting-state fMRI data acquired with a high temporal resolution, the inferred networks showed good consistency with structural connectivity obtained from diffusion tractography in the same subjects. Interestingly, this agreement could also be seen when considering high-frequency rather than low-frequency connectivity (average correlation: $r = 0.26$ for $f < 0.3$ Hz, $r = 0.43$ for $0.3 < f < 5$ Hz). Moreover, this concordance was significantly better ($p < 0.05$) than for networks obtained with conventional functional connectivity based on correlations (average correlation $r = 0.18$). The presented methodology thus appears to be well-suited for fMRI, particularly given its lack of explicit dependence on temporal lag structure, and is readily applicable to whole-brain effective connectivity estimation.

Keywords: effective connectivity, functional connectivity, structural connectivity, fMRI, resting state, correlation

OPEN ACCESS

Edited by:
Jorge Bosch-Bayard, Universidad Nacional Autonoma de Mexico, Mexico

Reviewed by:
Jan Casper De Munck, VU University Amsterdam, Netherlands
Felix Carbonell, Biospective Inc., Canada

***Correspondence:**
Carolin Lennartz
carolin.lennartz@uniklinik-freiburg.de

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *31 August 2017*
Accepted: *11 April 2018*
Published: *08 May 2018*

Citation:
Lennartz C, Schiefer J, Rotter S, Hennig J and LeVan P (2018) Sparse Estimation of Resting-State Effective Connectivity From fMRI Cross-Spectra. Front. Neurosci. 12:287. doi: 10.3389/fnins.2018.00287

INTRODUCTION

In recent years, brain connectivity analysis of functional magnetic resonance imaging (fMRI) data has become of high interest, particularly as many diseases such as Alzheimer's and epilepsy are now understood as cerebral network malfunctions (Fisher et al., 2017; Ofer et al., 2018). Functional MRI is a non-invasive method that can monitor whole-brain functional activity. In

resting state fMRI (rs-fMRI), relationships between intrinsic fluctuations across multiple brain areas are analyzed, giving rise to the concept of the brain as a network (Biswal et al., 1995). For connectivity analyses, it is common to consider *functional connectivity* (FC), which is retrieved by calculating the correlation between the time series from different brain areas. However, this approach exhibits some limitations (Stephan, 2004; Petersen and Sporns, 2015) as it yields only symmetric connections, with no information on their direction. Moreover, conventional approaches using raw correlations may reflect indirect connections between brain areas that are not actually directly linked.

Of great interest are methods that retrieve information about the existence and direction of connections, and which can rule out indirect connections. The *effective connectivity* (EC) describes "the influence one neural system exerts over another" (Friston, 1994), or as Aertsen and Preißl (1991) put it, EC is "the simplest possible circuit diagram that would replicate the observed timing relations" between observed responses and therefore describes directed connectivity. Although several different approaches have been suggested to estimate EC, the most widely used methods for fMRI data are Granger Causality (Bressler and Seth, 2011) and Dynamic Causal Modeling (Friston et al., 2003).

Granger causality (GC) exploits temporal precedence between two time series to estimate the direction of the connections. It exists both for time domain (Geweke, 1982) and frequency domain (Geweke, 1984; Baccalá and Sameshima, 2001) data. In order to estimate GC, usually a vector autoregressive process is fit to the data, which can be problematic as fMRI signals typically have a temporal resolution of 1–3 s (Lin et al., 2014), whereas characteristic time scales of neuronal processes are in the order of tens to hundreds of milliseconds. Moreover, temporal relationships between cerebral areas are confounded by the spatial variability of the hemodynamic response function (Handwerker et al., 2004). Although MR acquisition sequences with faster temporal resolutions are becoming increasingly common (Feinberg et al., 2010; Posse et al., 2012; Akin et al., 2017; LeVan et al., 2017), neuronal processes still undergo considerable downsampling in fMRI time series, affecting the reliability of GC estimates (Seth et al., 2013; Friston et al., 2014a).

Dynamic causal modeling (DCM) is a framework fitting differential equations to the fMRI data to yield parameters for the strength of connections, as well as the strength of the influence of external stimuli on connectivity. In the classical deterministic DCM, but also stochastic DCM (Li et al., 2011), the neuronal activity underlying the BOLD response is determined by a bilinear model, whereas the hemodynamic response is estimated using the Balloon model (Buxton et al., 1998; Friston et al., 2000). DCM requires to define a model a priori to test different specific hypotheses, which can then be compared via Bayesian model comparison (Penny et al., 2004, 2010; Penny, 2012). While the classical or stochastic DCM is only suited for task data with known input functions, a DCM for resting state data was developed recently (Friston et al., 2014b), which fits a model to the cross-spectrum of the data. However, due to the computational complexity of the differential equations, DCM is not suited for whole-brain connectivity analysis. Furthermore, with growing size of the models, non-identifiability becomes an issue of increasing severity (Arand et al., 2015; Frässle et al., 2015).

Aiming to overcome some of the issues outlined above, we present a methodology to estimate the EC from the frequency-domain cross-spectral density (CSD). Similar to the GC approach, the fMRI data are expressed as a multivariate autoregressive process, which is computationally suitable to model a large number of nodes in whole-brain datasets. However, unlike GC or other similar lag-based methods (see Smith et al., 2011 for a review of several such methods), we do not make use of temporal precedence to define the directionality of the estimated connections, thus partially circumventing hemodynamic confounds on the lag structure of fMRI time series. Rather, a directed and potentially asymmetric network is estimated in such a way as to explain the observed cross-spectral density matrix. As temporal precedence is not enforced, this is an underdetermined problem with a potentially infinite number of solutions, so we additionally constrain the network to have the smallest number of non-zero connections using an L1 minimization on the entries of the connectivity matrix.

One issue when validating EC estimation in real fMRI data is the lack of an ideal ground truth. One popular approach is to use information from structural connectivity (SC), which can be estimated using diffusion-tensor imaging (DTI). Using tractography algorithms (Wedeen et al., 2008; Reisert et al., 2013) on the DTI data, the white matter tracts forming connections between different regions can be reconstructed and the number of "fibers" (streamlines) can be used as a proxy for the strength of these connections. SC is commonly used to constrain the estimation (Gilson et al., 2016; Crimi et al., 2017; Dang et al., 2018), or may be used independently to validate the estimated EC (Uddin et al., 2011; Bringmann et al., 2013). However, there are also clear limitations to such approaches, as EC is dynamic and potentially brain-state-dependent as opposed to static SC. As such, while SC is often used as a proxy for connectivity and reasonable agreement is found between FC and SC (Li et al., 2012; Finger et al., 2016), we should not expect complete concordance between SC and EC, although the two measures should still be consistent with each other.

In the remaining sections we will briefly explain the mathematical background and implementation of the method. In a simulation study, the influence of several parameters on the estimation will be analyzed. Finally, we will apply the methodology to real resting-state fMRI data. In the absence of ideal validation measures in real data, the consistency of the estimated effective connectivity with the structural connectivity from white matter tracts will then be assessed.

MATERIALS AND METHODS

Methodology

Mathematical and Algorithmic Background

We consider networks of n interconnected neuronal populations. Each population is characterized by neuronal activity $y_i(t)$ with

$i \in [1, n]$. Similar to the GC framework, we assume that the neuronal activity follows a generic multivariate autoregressive process

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{x}(t) + \int_{-\infty}^{t} \boldsymbol{G}(t-u)\,\boldsymbol{y}(u)\,du \\ &= \boldsymbol{x}(t) + \boldsymbol{G}*\boldsymbol{y}(t) \end{aligned} \quad (1)$$

which describes how the neuronal activity $\boldsymbol{y}(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T$ at time point t in each population depends on the driving "noise" (or external stimuli) $\boldsymbol{x}(t)$ and the activity in other populations with time lag u via the linear coupling kernel $\boldsymbol{G}(t)$, where $\boldsymbol{G}_{ij}(t)$ describes the influence of node j on node i. The coupling can be described by a convolution ("*") of $\boldsymbol{G}(t)$ with the neuronal activity $\boldsymbol{y}(t)$.

Now, in the GC framework, a causal system would then be assumed by additionally setting $\boldsymbol{G}(t) = 0$ for negative time lags $t < 0$, and the remaining coefficients of $\boldsymbol{G}$ could then be fitted by linear regression, with non-zero coefficients indicative of a directed influence of one node on another inferred from their temporal precedence relationship (Goebel et al., 2003; Duggento et al., 2016). This approach can also been extended to support non-linear interactions (Harrison et al., 2003) and couplings that are dynamically fluctuating over time (Smith et al., 2013; Park et al., 2017; Samdin et al., 2017). However, as outlined above, fMRI only indirectly measures neuronal activity in the form of the BOLD signal, yielding low temporal resolutions and spatially variable lag structure that confound GC estimates (Deshpande et al., 2009; Rogers et al., 2010).

Circumventing these issues, we deviate from the GC framework and do not enforce the causality of $\boldsymbol{G}$ and thus do not rely on temporal precedence relationships to identify directed connections. Rather, we rely on the observation that cross-correlations, which are symmetric and thus undirected, nevertheless contain information about the underlying directed (and thus potentially asymmetric) network, notably the presence of so-called "collider" structures (Ramsey et al., 2010; Pernice and Rotter, 2013). Based on frequency-domain cross-spectra, we thus estimate a directed network independently of temporal precedence relationships.

Applying the Fourier transform to Equation 1, we get $\widehat{\boldsymbol{y}}(f) = \widehat{\boldsymbol{x}}(f) + \widehat{\boldsymbol{G}}(f)\widehat{\boldsymbol{y}}(f)$, where $\widehat{\cdot}$ depicts the Fourier transform of the respective variable. Assuming that both the intrinsic noise $\boldsymbol{x}(t)$ and the neuronal activity $\boldsymbol{y}(t)$ are stationary stochastic processes, the cross-spectral density can be derived (Hawkes, 1971; Pernice and Rotter, 2013):

$$\langle \widehat{\boldsymbol{y}}(f)\widehat{\boldsymbol{y}}^*(f)\rangle = \widehat{\boldsymbol{C}}(f) = \left[\mathbf{1} - \widehat{\boldsymbol{G}}(f)\right]^{-1} \widehat{\boldsymbol{X}}(f) \left[\mathbf{1} - \widehat{\boldsymbol{G}}^*(f)\right]^{-1} \quad (2)$$

$\widehat{\boldsymbol{G}}(f)$ is the frequency-dependent coupling matrix, 1 the identity matrix, $\widehat{\boldsymbol{X}}(f) = < \widehat{\boldsymbol{x}}(f)\widehat{\boldsymbol{x}}^*(f) >$ depends on the driving noise, and $<,>$ is the time expectation operator. Noise is assumed to be independent and Gaussian, so that $\widehat{\boldsymbol{X}}(f)$ is a diagonal matrix of (unknown) noise variances.

We are ultimately interested in recovering the effective connectivity $\widehat{\boldsymbol{G}}(f)$ of the network of neuronal populations, given only $\widehat{\boldsymbol{C}}(f)$, the cross-spectral density matrix of the measured activity $\boldsymbol{y}(t)$. Taking the inverse of the CSD [2] we get

$$\widehat{\boldsymbol{C}}^{-1}(f) = \left[\mathbf{1} - \widehat{\boldsymbol{G}}^*(f)\right] \widehat{\boldsymbol{X}}^{-1}(f) \left[\mathbf{1} - \widehat{\boldsymbol{G}}(f)\right] = \boldsymbol{B}^*(f)\boldsymbol{B}(f) \quad (3)$$

with $\boldsymbol{B}(f) = \sqrt{\widehat{\boldsymbol{X}}(f)}^{-1} \left[\mathbf{1} - \widehat{\boldsymbol{G}}(f)\right]$.
Given an estimate of $\boldsymbol{B}(f)$, the coupling matrix

$$\widehat{\boldsymbol{G}}(f) = \mathbf{1} - \sqrt{\widehat{\boldsymbol{X}}(f)}\,\boldsymbol{B}(f) \quad (4)$$

can be estimated only up to a positive factor $\sqrt{\widehat{\boldsymbol{X}}(f)}$ as the covariance of the intrinsic noise is not known. The matrix $\boldsymbol{B}(f)$, nonetheless, gives information about strength, sign and direction of connections since $\widehat{\boldsymbol{X}}(f)$ is diagonal, although it may affect the scaling of the estimated weights (For better readability the dependency of the variables on the frequency is dropped from here on).

The computation of $\boldsymbol{B}$ from the CSD is, however, not straight forward, because it is not uniquely defined: Many different network topologies can give rise to the same CSD. More precisely, the decomposition of the CSD is only defined up to an arbitrary unitary transformation $\boldsymbol{U}$ since

$$\widehat{\boldsymbol{C}}^{-1} = \boldsymbol{B}^*\boldsymbol{B} = \boldsymbol{B}^*\boldsymbol{U}^*\boldsymbol{U}\boldsymbol{B}. \quad (5)$$

To resolve this ambiguity, we assume that the network formed by the neuronal populations is sparse, which entails minimizing the L1-norm of the entries of the matrix $\boldsymbol{UB}$. The corresponding cost function is

$$\Gamma(\boldsymbol{U}\boldsymbol{B}_0) = \|\boldsymbol{U}\boldsymbol{B}_0\|_1 = \sum\nolimits_{i\neq j} |(\boldsymbol{U}\boldsymbol{B}_0)_{ij}| = \sum\nolimits_{i\neq j} \left|\sum\nolimits_k U_{ik}B_{0,kj}\right| \quad (6)$$

where $\boldsymbol{B}_0$ is the initial guess of the decomposition. So the problem is to find the unitary transform $\boldsymbol{U}$ minimizing the cost function $\Gamma(\boldsymbol{U}\boldsymbol{B}_0)$ (Pernice and Rotter, 2013; Schiefer and Rotter, 2016)

$$\begin{aligned} &\operatorname{argmin}_U \Gamma(\boldsymbol{U}\boldsymbol{B}_0) \\ &s.t. \boldsymbol{U}\boldsymbol{U}^* = \mathbf{1} \end{aligned}$$

Geometrically this optimization can be viewed as a complex rotation of the cross-spectral density matrix, which can be implemented using a conjugate gradient descent algorithm (Abrudan et al., 2008, 2009).

For the estimation of the effective connectivity, each frequency bin of the CSD is treated separately, leading to a frequency-dependent connectivity. As starting point $\boldsymbol{B}_0$ for the estimation, the positive definite matrix square root of $\widehat{\boldsymbol{C}}^{-1}(f)$ is chosen.

Threshold From Null Distribution

To exclude statistically non-significant connections in the estimated connectivity matrix, a threshold for each frequency is derived from a null distribution. The null distribution is computed by first splitting the time series into equal segments, shuffling the segments randomly and differently for each time

series, and finally computing the CSD on the shuffled segments using Welch's method

$$CSD_{ij,\ null}(f) = \frac{1}{n}\sum_{k=1}^{n} \widehat{y}_{ik}\left(f\right) \cdot \widehat{y}_{j\sigma(k)}(f) \tag{7}$$

where $\sigma\left(k\right)$ is a permutation mapping. The shuffling will only affect the cross-spectra, while the power spectra (diagonal of the CSD matrix) will be preserved. Calculating connectivities from this null CSD and assuming that these values are to a great extent independent, a distribution of effective connectivity values is then derived. The 2.5 and 97.5% quantiles then yield $p < 0.05$ lower and upper thresholds for the connectivity matrices.

Confidence Intervals From Bootstrapping

We also derive confidence intervals for the connection strengths using bootstrapping. This could also be used to exclude connections that include zero in their confidence interval.

To derive the confidence intervals, the time series are again split into segments which are Fourier transformed. These segments are then drawn randomly with replacement and the CSD is calculated with the order of the segments kept identical for each time series

$$\mathrm{CSD_{ij,\ bootstrap}(f)} = \frac{1}{\mathrm{n}}\sum_{\mathrm{k=1}}^{\mathrm{n}} \widehat{\mathrm{y}}_{\mathrm{i}\gamma(\mathrm{k})}\left(\mathrm{f}\right) \cdot \widehat{\mathrm{y}}_{\mathrm{j}\gamma}(\mathrm{k})(\mathrm{f}) \tag{8}$$

where $\Gamma(k)$ is a permutation with replacement, which is the same for both time series $\widehat{y}_i$ and $\widehat{y}_j$. Calculating several bootstrap CSDs and estimating the connectivity thereof, a distribution of connection strength can be derived for each connection. Assuming an asymptotic Gaussian distribution of the parameter values, confidence intervals can then be determined.

Data Acquisition

fMRI Acquisition and Pre-processing

For experiments with real data, all measurements were performed on a 3 T Prisma scanner (Siemens Healthineers, Erlangen, Germany). Seven healthy volunteers, five male and two female in the age between 18 and 49, underwent a 20 min resting-state fMRI scan using the MREG sequence (Hugger et al., 2011; Assländer et al., 2013) with $TR = 0.1$ s, $TE = 36$ ms, $FA = 25^{\circ}$, $64 \times 64 \times 50$ matrix and 3 mm isotropic voxel size. T1-weighted MPRAGE images ($TR = 2{,}000$ ms, $TE = 4.11$ ms, FOV = 256 mm, 256×256 matrix, 160 sagittal slices, 1 mm slice thickness) were acquired for anatomic reference. Cardiac and respiratory fluctuations were additionally recorded with ECG and abdominal breathing band from the scanner's physiological monitoring unit. This study was approved by the Ethics Committee of the University Medical Center Freiburg. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The data is available via the Open Science Framework repository (https://osf.io/52mf4/).

The fMRI data was motion corrected using FSL. Physiological noise correction was conducted with RETROICOR (Glover et al., 2000). The fMRI data sets were registered to their corresponding T1-images, which were in turn registered to MNI space. The registered fMRI data sets were parcellated according to the AAL-atlas and mean activity was calculated within each atlas region, excluding the cerebellum. The CSD was calculated for each dataset using Welch's method with a Hanning window with 50% overlap between windows.

To ensure that the CSD has full rank to be invertible, the number of frequency bins needs to be smaller than the degrees of freedom, i.e., the number of Fast Fourier Transform bins $NFFT < \frac{\#\ time\ points}{\#\ nodes}$. As the convolution with the HRF further reduces the degrees of freedom, the number of frequency bins was further decreased to the next lower power of two. Finally, the effective connectivity was extracted from the CSD for each frequency by sparse optimization as described in section Mathematical and Algorithmic Background.

DTI Acquisition and Pre-processing

In the absence of a gold standard for validation, a comparison with structural connectivity was performed. Thus, diffusion-weighted data was also acquired during the MRI sessions (61 diffusion directions, $TR = 6.6$ s, $TE = 80$ ms, $b = 1{,}000$ s/mm^2, 60 slices, 2 mm isotropic voxel size). Using a global fiber tractography algorithm (Reisert et al., 2013) the structural connectivity could be extracted by counting streamlines connecting each pair of brain regions. Fiber endpoints lying in brain areas not covered by a region in the AAL atlas were reassigned to the nearest AAL area.

A summary SC matrix across all subjects was also generated from the individual SC matrices by considering connections existing in at least two thirds of the subjects.

Simulation Study

As a proof of principle we first applied the method to simulated fMRI data. Moreover, we investigated the influence of several parameters on the estimation of the effective connectivity.

For this purpose, a vector autoregressive process of order 50 (VAR[50]) corresponding to a maximum conduction delay of 5 s was used with an additional contemporaneous term to model instantaneous self-excitation effects in each node and driving noise $\boldsymbol{e}(t)$:

$$\boldsymbol{y}(t) = \sum_{p=0}^{50} \boldsymbol{G}(p)\boldsymbol{y}(t-p) + \boldsymbol{e}(t)$$

where $\mathbf{G}(p = 0)$ is the identity matrix. To simulate the oscillatory nature of resting-state fMRI data, the intrinsic activity in each node of the network was modeled as a noisy superposition of harmonic oscillations with different phases and frequencies. We chose a connection probability of each pair of nodes of 15% to model a sparse network. The coupling matrices $\mathbf{G}(p)$, which were modeled using random Erdős-Rényi networks, were the same for every lag p, however, they decreased in strength following a logistic decay. As a last step, the time series were convolved with the canonical hemodynamic response function (HRF) to simulate BOLD responses.

Functional magnetic resonance imaging (fMRI) observational noise is made up of several noise sources like scanner, physiological and temporal noise. The scanner noise is inherent in all fMRI data and can be modeled by Gaussian white noise (Gudbjartsson and Patz, 1995; Welvaert et al., 2011) given

sufficient signal to noise ratio (SNR). Structured physiological noise corresponds to respiratory and cardiac oscillations. As we have a relatively high temporal resolution in our data, we assumed that most of the physiological noise could be filtered out. Therefore, it was not modeled in the simulated data. Temporal noise accounts for various sources of fluctuations with temporal autocorrelation (Purdon and Weisskoff, 1998). This was modeled using an AR[1] with coupling strength 0.5. Thus, observational noise with both white "scanner" noise and "temporal" noise in the form of an AR[1] were added.

We performed 20 simulations with 100 nodes. For each simulation the same network connectivity was used, but with different noise realizations. Furthermore, because a fast fMRI sequence with a $TR = 0.1$ s was used for the real fMRI measurements, our simulations of the VAR-process and the convolution with the HRF were performed with this temporal resolution with 51,000 data points corresponding to a 85 min measurement.

Connectivity Analysis

Following the procedure outlined in section Methodology, the connectivity matrices were estimated for each frequency f from the CSD. The raw estimated connectivities were used directly for the analysis of the influence of parameters such as length of the time series. For the final performance analysis, however, the connectivities were additionally thresholded using the previously described null distribution and confidence intervals for each frequency bin. For the null distribution (Equation 7), as the "connections" in the null connectivity matrices are independent, only 10 cross-spectral density matrices were calculated per frequency and all derived connections were pooled to build a null distribution with 100,000 entries in the histogram. For the confidence intervals (Equation 8), 1,000 bootstrapped CSDs were calculated for each frequency.

Simulation study

Various simulations were performed to investigate the influence of the following parameters:

Length of the time series

Lengths were varied between 3,000 and 51,000 data points in steps of 6,000 (= 10 min). Furthermore, the SNR (ratio of signal variance to noise variance) was varied between 1 and 5.

Type of observational noise

Data were simulated with either pure white noise, a more realistic temporally correlated noise, or a mixture of both

$$e = a(1) \cdot e_{white} + a(2) \cdot e_{temp}$$

where a (1) is 0.3 and a (2) is 0.7. We again varied the SNR between 1 and 5 and used time series lengths of 20,000 (∼35 min) or 40,000 (∼70 min) data points.

Number of observed nodes

If connectivity analyses are performed in a given subnetwork of interest rather than the whole brain, hidden nodes exerting an influence on the observed nodes might yield erroneous results. To investigate this, connectivity estimations were performed within various fractions of the whole network in steps of 10 nodes, using either 20,000 or 40,000 data points and an SNR of 5.

Hemodynamic variability

In order to investigate the sensitivity of the estimation to hemodynamic variability, the activity at each node was subjected to a different random HRF instead of the canonical HRF. The random HRFs were generated using a double-gamma model, where the onset times of each gamma function was varied by up to 5 s, while the dispersion and amplitude parameters were varied by a factor of up to 5.

Finally, the recovery of an "average" group-level network was investigated using the 20 simulated realizations of the same network with an SNR of 5 and a length of 40,000 data points. After using the null distribution and confidence intervals to remove non-significant connections, averaging the resulting networks was not possible since connections did not necessarily exist in all datasets, so the mean network was defined as connections existing in at least half of the 20 data sets.

The comparison of the estimated EC with the true connectivity was done using correlation between the connectivity matrices. However, the correlation could be high even if many erroneous connections were detected, as long as all true connections are also found. Therefore, the area under the receiver-operator characteristic (ROC) curve (AUC) was also calculated to gain information about sensitivity of the estimation, where the ROC curve was obtained by varying the threshold on the estimated EC matrices.

fMRI data

For real fMRI data, the effective connectivity matrices were also derived for each frequency and non-significant connections were removed by calculating the threshold from the null distribution and deriving the confidence intervals.

To analyse the variability of the derived networks over subjects, the correlation and area under the ROC curve were calculated for the EC and SC networks between all subjects for each frequency. To compute an average connectivity over all subjects, only connections which existed in at least half of the subjects (4 in this case) were kept in the connectivity matrix. While the true underlying connectivity is not known, consistency was nevertheless assessed between the estimated EC and SC from DTI. However, because SC is symmetric, the estimated EC networks were first "mirrored" by adding the transposed connectivity matrix to the normal connectivity matrix. It was thus not possible to strictly validate the directionality of the estimated connections; in the absence of suitable gold standard, this approach is nevertheless expected to provide a limited degree of validation in real data.

Furthermore, EC was also compared to standard functional connectivity, represented by the raw cross-spectral density between time series from the various regions of interest. The correlation and AUC values for the comparison between EC and SC and FC and SC were calculated for each frequency and each subject. Furthermore, the agreement of SC and EC/FC was compared by determining the percentage of connections

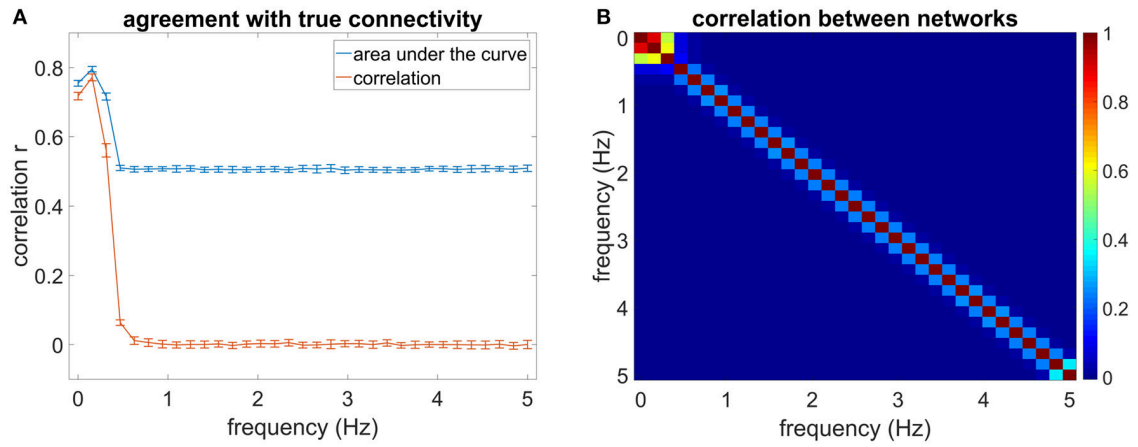

FIGURE 1 | (A) Shows the mean correlation of the estimated connectivity matrices and the true connectivity and the area under the ROC-curve for 20 data sets at different frequencies. The error bars show the standard deviation of the estimated correlation or AUC values over data sets. **(B)** Shows the correlation between estimated networks at different frequencies.

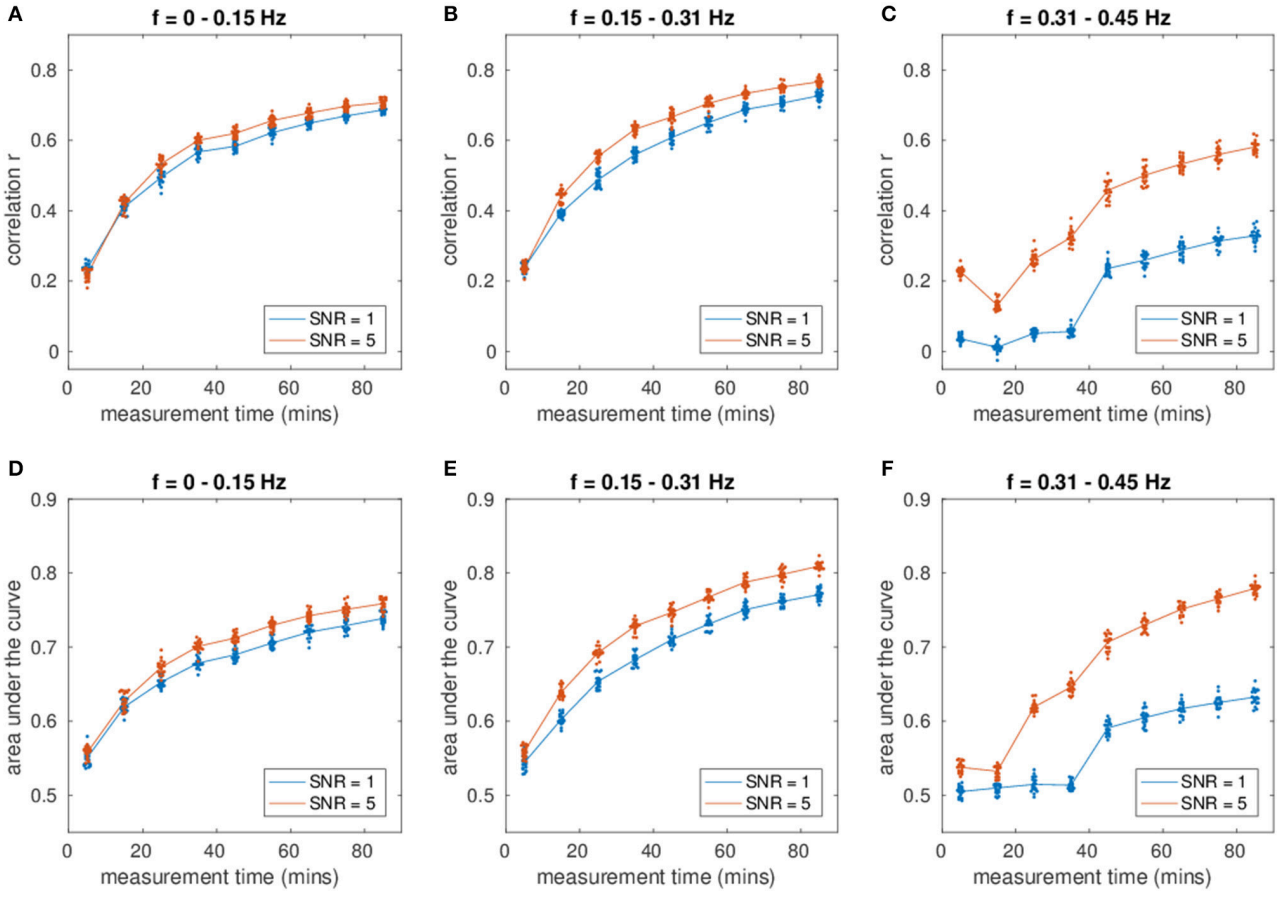

FIGURE 2 | (A–C) Shows the correlation of the estimated connectivity matrices and the true connectivity with varying dataset length. **(D–F)** Shows the area under the ROC-curve (AUC) for the different lengths. Different colors correspond to different SNR values.

fulfilling each of the three following cases: (1) The connection is present in both SC and EC/FC, (2) the connection is present in EC/FC, but not in SC, and (3) the connection is present in SC, but not EC/FC. All calculated percentages were relative to the number of connections present either in SC and/or EC/FC. For FC, connections were thresholded using the null distribution

in section Threshold From Null Distribution and the mean FC network was derived by taking only connections existing in at least half of the subjects. Since the network densities of EC and FC do not agree, an additional analysis was performed using custom thresholds on the mirrored EC and FC networks, set to yield a 10% false positive rate of connections present in EC/FC but not SC.

As the temporal resolution given by the repetition time (TR) was much higher than in conventional fMRI (1–3 s), the influence of the TR was also analyzed by downsampling the time series to TRs between 0.1 and 3 s.

Finally, the default-mode network (DMN) was analyzed as an example of a well-studied network in literature. The DMN consists of three main brain areas in each hemisphere: the medial prefrontal cortex (mPFC), the inferior parietal cortex (IPC), and the posterior cingulate cortex (PCC). Furthermore, the hippocampus (HIP) and the temporal cortex (TC) are also sometimes included in the network. Note that the network estimation was still performed at the whole-brain level, after which only the connections within the DMN were examined in more detail.

RESULTS

Simulation Study

For the simulated data, only the low-frequency bins were analyzed as the higher frequency bins contained mainly noise due to the convolution with the canonical HRF. This left three frequency bins with $0 < f < 0.45\ Hz$ (cf. **Figure 1**).

Influence of Length of Time Series and SNR

Figures 2A–C show the mean correlation of the estimated network with the true network for the first three frequency bins according to measurement time and SNR. **Figures 2D–F** show the mean AUC values for the first three frequency bins. The correlation and AUC increases monotonically with increasing length of the time series. Moreover, an increase in SNR improves the estimation.

With increasing measurement time, the correlation of the networks and the AUC increases strongly. After a measurement time of ∼35 min the slope of the correlation and AUC plot is shallower. Most of the estimation power is concentrated in the frequency range between 0 and 0.31 Hz, yielding higher correlation of the estimated networks with the true networks than the higher frequency bins (see **Figure 1**). For the first frequency bin, the correlation reaches a level of close to 0.8 for an SNR of 5. For the second frequency bin (0.15–0.31 Hz), the correlation and AUC even go beyond 0.8 for high SNR. For the third frequency bin, the correlation ranges for a measurement time of 85 min between 0.4 for the lower SNR and 0.75 for the higher SNR. The AUC varies between 0.6 and 0.75, where a value of 0.5 equals pure chance.

The SNR has a strong influence on the estimation: An increase of SNR improves the estimation. Higher frequencies are especially sensitive to measurement time and SNR (cf. 2 C/F). While for a SNR of 5 the estimation is still quite good, especially for long measurements of 40 min and more, the estimation for the lower SNR declines for measurements shorter than approximately 45 min. At shorter measurements the noise predominates in the CSD. However, at approximately 45 min there is a prominent jump in the correlation and AUC values. At such long measurement times, sufficient noise averaging occurs and the true covariance structure can be retrieved fairly well.

TABLE 1 | Influence of type of noise.

Noise type\SNR	1	5
LENGTH OF TS: 20,000 DATA POINTS		
White	0.462 ± 0.003	0.588 ± 0.003
Mixture	0.414 ± 0.006	0.565 ± 0.004
Temporally correlated	0.410 ± 0.006	0.562 ± 0.005
LENGTH OF TS: 40,000 DATA POINTS		
White	0.558 ± 0.003	0.710 ± 0.003
Mixture	0.505 ± 0.002	0.686 ± 0.003
Temporally correlated	0.499 ± 0.001	0.682 ± 0.004

Mean correlation over 20 datasets and the first three frequency bins for different noise compositions, SNRs and different length of measurement time.

Influence of Type of Noise

Table 1 shows results from the networks simulated using different noise types, averaged over the three low-frequency bins.

At low SNR, the differences for the various types of noise are more prominent than at high SNR, where the differences start to vanish. The difference between the mixture of white and temporal noise and pure temporal noise, however, is not so prominent. An increase of measurement time improves the estimation itself, but does not have an influence on the observed differences between noise types.

Influence of Missing Nodes on the Estimation

Figure 3 shows an increase of estimation power with increasing fraction of observed nodes (**Figure 3A**, correlation and **Figure 3B**, AUC), indicating the importance of the missing nodes on the network.

Regarding the variance of the estimation for different network sizes, a strong decrease in variance with decreasing fraction of missing nodes can be observed, demonstrating the beneficial influence of more nodes and therefore more information to recover the network.

Influence of Hemodynamic Variability

Figure 4 shows the estimation performance under various degrees of HRF variability as well as a comparison with multivariate GC (Barnett and Seth, 2014). While both methods perform well with fixed HRFs, they also show a clear degradation under variable HRF conditions, although GC is more susceptible to the confounding influence of the HRF on temporal precedence information.

Mean Network From All Data Sets

Figure 5 shows the comparison between the true network (**Figure 5A**) and the mean estimated network (**Figures 5B–D**)

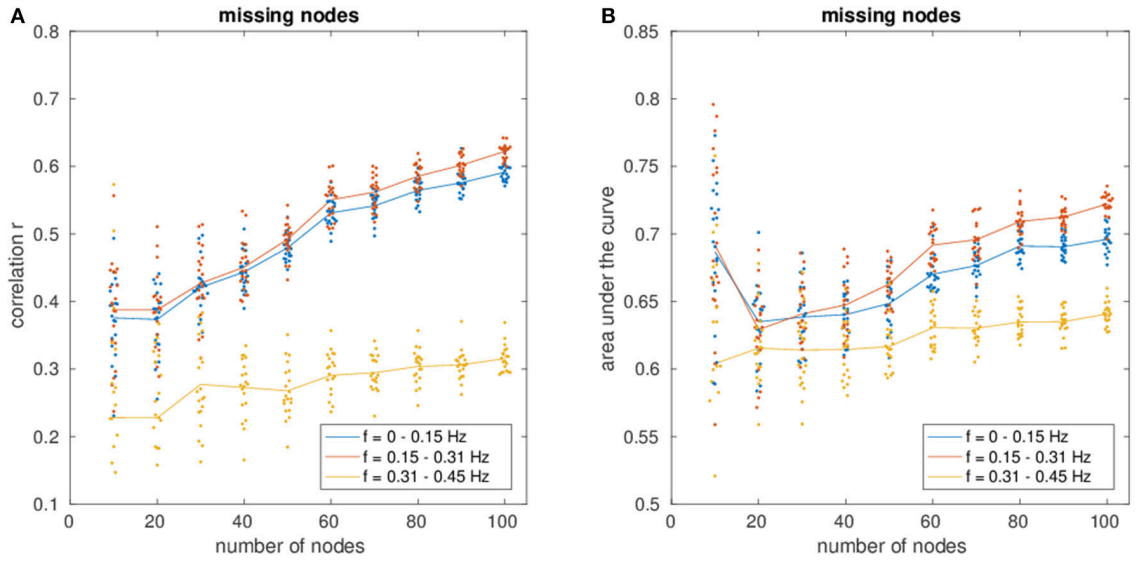

FIGURE 3 | (A) Shows the correlation of the estimated connectivity matrices and the true connectivity as a function of varying network size with missing nodes present. **(B)** Shows the corresponding area under the ROC-curve (AUC). Different colors correspond to different frequency bins.

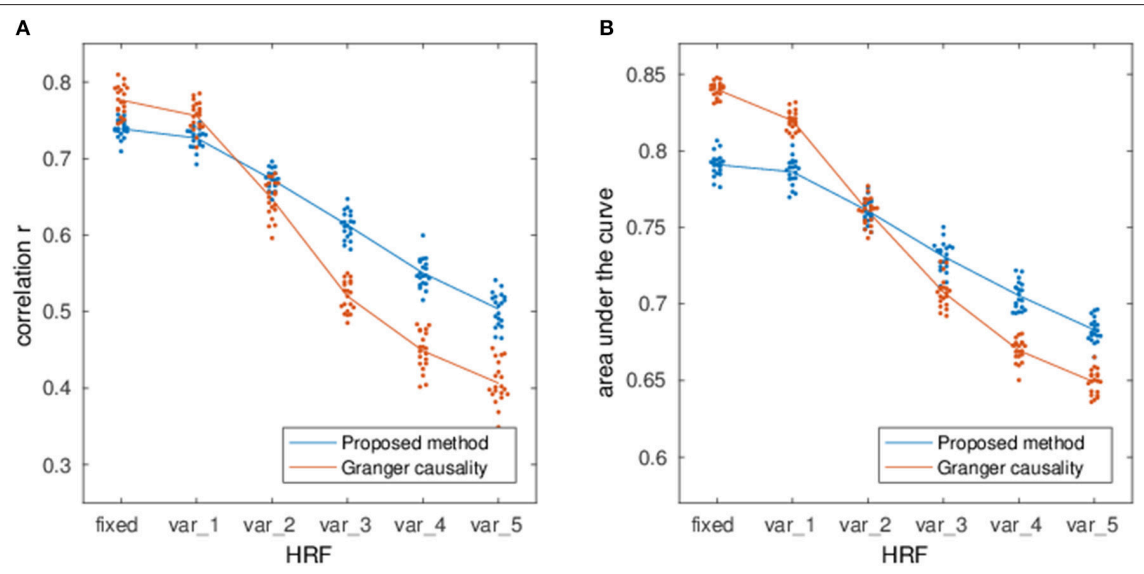

FIGURE 4 | Correlation **(A)** and AUC **(B)** of the EC estimation method as a function of HRF variability (blue line). The labels var_*k* denote simulations performed with double-gamma HRFs where the onset of the gamma functions was randomly varied between ±*k* seconds and the dispersion and amplitude parameters were randomly varied by a factor of up to *k* with respect to the canonical HRF. For comparison, the red line shows the performance of multivariate Granger Causality estimation.

for the lowest three frequency bins. Each entry in the matrix plot corresponds to a directed connection, where the connection goes from column to line. Red entries in the matrix plot correspond to positive (excitatory) connections and blue ones to negative (inhibitory) connections. Moreover, the hue of the color depicts the strength of the connections. Due to the random nature of the simulations, there is no structure in the network that may facilitate visual inspection, but it can still be observed that strong connections are especially well estimated.

The titles from **Figures 5B–D** show the correlation and AUC values for the mean networks without thresholding by the null distribution (AUC and r_m) and correlation with applied thresholds to exclude non-significant connections (r_{tm}). Taking solely the mean over all networks yields high correlations between 0.8 and 0.9 for the first three frequency bins. By removing the non-significant connections from the network using the threshold from the null distribution and taking only connections which exist in at least 50% of the estimated networks increases the correlation even

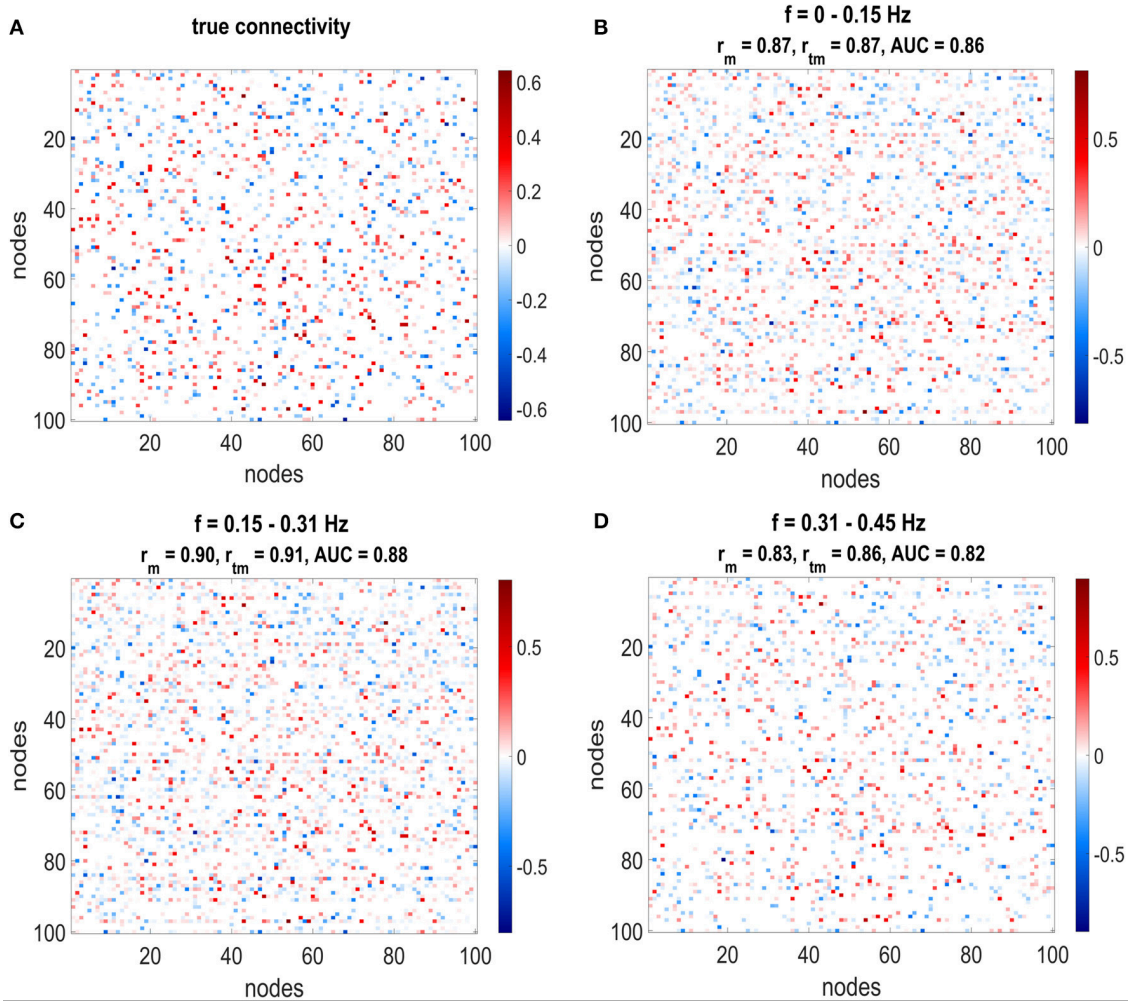

FIGURE 5 | The estimated connectivity matrices and the true connectivity for three different frequency ranges are depicted. **(A)** Shows the true connectivity used to simulate the data, **(B–D)** Shows the estimated connectivity matrices for the different frequency ranges. The plots show the weighted adjacency matrices of the network, where red entries stand for connections with positive weights, blue entries for connections with negative weights and white depicts no connection. Furthermore, the hue of the color depicts the strength of the connection. Correlation and AUC values are given in the title of each plot: r_m stands for correlation with mean network without threshold from null distribution, r_{tm} corresponds to correlation with mean network with threshold from null distribution. All correlations and AUC values are above 0.8.

further. The area under the ROC curve lies between 0.8 and 0.9.

Real fMRI Data

Applying the threshold from the null distribution and deriving the confidence intervals from the bootstrapped networks, sparse networks were achieved in the real fMRI data (see **Figure 6**). We separately consider a low frequency band (0–0.31 Hz) and a high frequency band (0.31–5 Hz). The estimated average low and high frequency networks are quite similar ($r = 0.6$) and, moreover show a strong similarity between hemispheres ($r = 0.8$ for the low frequency network, $r = 0.9$ for high frequencies).

The results from the analysis of the variability of the SC and EC networks over subjects can be seen in **Table 2**. EC shows strong variability both for low and high frequencies with a correlation around 0.2–0.3 and AUC of around 0.6. SC, however, shows a strong agreement between subjects ($r = 0.85$ and AUC = 0.80), showing the high stability of the SC across subjects.

Comparison With DTI Tractography

Due to the lack of a gold standard the estimated EC networks were compared to SC from DTI. The resulting correlation and AUC values are displayed in **Table 3**.

The low frequency network shows some correlation with SC ($r = 0.24$ and AUC = 0.56), which is only slightly increased for the mirrored network ($r = 0.26$ and AUC = 0.55). For the high frequency band the agreement between SC and EC is much more pronounced. For the mirrored network we have a correlation of $r = 0.43$ and even for the normal EC the correlation is quite high

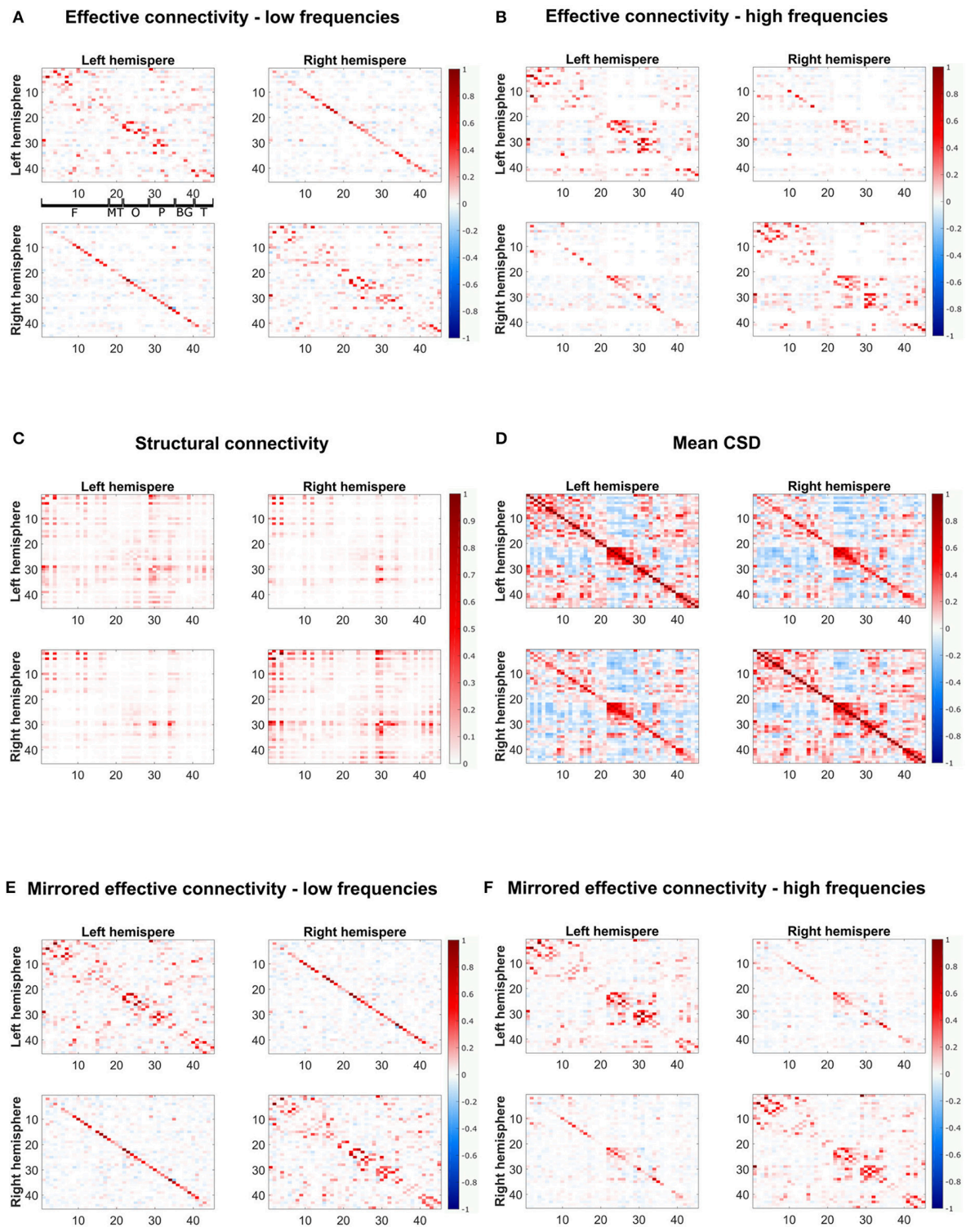

FIGURE 6 | The estimated effective connectivity matrices from fMRI data for low- and high-frequency bands and their mirrored networks and the structural connectivity from DTI are depicted. **(A,B)** Show the effective connectivity for low- and high-frequency bands, respectively, **(E,F)** Show the mirrored estimated connectivity matrices for the different frequency ranges. For visualization purposes, only connections present in at least two thirds of frequencies in the respective frequency bands are shown. **(C)** Shows the structural connectivity and **(D)** the mean CSD over all subjects and frequencies. The plots show the weighted adjacency matrices of the network, where red entries stand for positive connections, blue entries for negative connections and white depicts no connection. Furthermore, the color hue depicts the strength of the connections. The axes refer to indices from the AAL atlas, separated between left and right hemispheres, corresponding to the regions indicated in the topmost left plot: F, frontal; MT, mesial temporal; O, occipital; P, parietal; BG, basal ganglia; T, temporal.

TABLE 2 | Variability of networks between subjects.

	Correlation r	AUC of ROC
Low-frequency band	0.228 ± 0.054	0.552 ± 0.022
High-frequency band	0.256 ± 0.060	0.637 ± 0.035
DTI	0.845 ± 0.049	0.804 ± 0.042

Correlation and AUC values (mean and standard deviations) calculated between estimated networks from fMRI and structural connectivity from DTI between subjects.

TABLE 3 | Agreement between EC/FC and SC.

	Correlation r	AUC of ROC
LOW FREQUENCY		
$G+G^T$	0.26	0.55
G	0.24	0.56
HIGH FREQUENCY		
$G+G^T$	0.43	0.62
G	0.41	0.60
MEAN CSD		
CSD	0.18	0.56

Correlation and AUC values between mean estimated effective connectivity networks and functional connectivity networks from fMRI and structural connectivity from DTI.

with 0.41 (cf. **Table 3**). However, there are still clear discordances, as seen in the AUC values of 0.62 for the mirrored and 0.60 for the normal EC.

Figure 6 shows both normal and mirrored estimated EC networks for the low- and the high-frequency band (**Figures 6A,B,E,F**), the SC network (**Figure 6C**) and the mean CSD over all subjects and frequencies (**Figure 6D**). In agreement with SC, the EC shows strong connectivity in the frontal and and parietal regions. Moreover, the lack of connections between frontal and mesial temporal and occipital regions is correctly identified. Discrepancies are mainly visible along the diagonals of the top-right and bottom-left quadrants corresponding to interhemispheric connections between homologous regions. Differences are also seen close the diagonal of the top-left and bottom-right quadrants because local short-range connections are not easily recovered by diffusion tractography. The high frequency EC network, however, has less pronounced interhemispheric connections, therefore also yielding higher correlation with SC.

Comparison With Raw Cross-Spectral Density

The results of EC were also compared to standard functional connectivity, represented by the cross spectral density (**Figure 6D**). Mean CSD shows lower agreement with SC ($r = 0.18$ and AUC $= 0.56$). It can be seen that the EC networks are much sparser than the CSD functional network. Furthermore, the block of massive inter-hemispheric connections in the occipital lobe present in the CSD (between regions 20–30 on the top-right and bottom-left quadrants) vanished in the low frequency EC network and is very much reduced in the high frequency EC network.

Figures 7A–C shows coincidence maps, where green entries depict agreement between SC and EC/FC, bright red entries connections that are only present in EC/FC but not SC, and pale red entries connections that are only present in SC but not in EC/FC. For comparison purposes, the EC/FC networks were thresholded to have a false positive rate (FPR) of 10% to gain an insight into the agreement of SC with EC/FC given a fixed FPR. At this FPR, many SC connections are not reflected in EC/FC. Nevertheless, the high-frequency EC shows the highest agreement with SC (**Figure 7B**). Low-frequency EC and FC show similar agreement (cf. **Table 4**), although for low-frequency EC the agreeing connections are very scattered while for FC the agreeing connections tend to form clusters.

Figures 7D–F shows histograms of the connection strengths for connections not present in SC but in EC/FC. Such false connections (without an underlying structural basis) for low- and high-frequency EC networks are very weak even though they were statistically significant. In contrast, the connection strengths of the false connections for CSD range from −0.3 up to 0.6.

To quantify the agreement between SC and EC/FC, the percentage of connections agreeing between SC and EC/FC, connections only present in EC/FC, and connections only present in SC were calculated (see **Table 4**). For non-adapted network densities, SC and EC correspond better than CSD, which is mainly due to the higher number of connections in EC than FC. For connections present in EC/FC, but not in SC, the percentage is similar for normal EC and SC. Looking at the networks with adapted densities, mirrored EC for high frequencies and SC have the highest agreement; FC and EC for low frequencies have similar but lower agreement with SC. However, as seen in **Figures 7D–F**, "wrong" connections, which are present in EC/FC but not SC, cover a much broader range for FC than for EC.

Figure 8 depicts the mean correlation and AUC values of all subjects over all frequencies of EC/FC with SC from DTI. The correlation and AUC is always higher for mirrored EC networks than for normal EC networks. The correlation is also higher when compared to the raw CSD. However, for the AUC there is high variability and strong overlap between mirrored EC and CSD, although mirrored EC is still mostly above CSD.

This can also be seen in **Table 5**, where the statistical significance of the *t*-test of EC correlation and AUC values vs. those from CSD are tested, where correlation values were z-transformed prior to the *t*-test. For the correlation at low frequencies, mirrored EC is better than CSD, but other measures are not significantly different. However, at high frequencies, both normal and mirrored EC correlate significantly better than CSD, but only the mirrored EC shows significantly higher AUC than the CSD.

Influence of TR/Sampling Rate

The results of the analysis with different TRs are depicted in **Figure 9** for low frequencies (A) and high frequencies (B). While low frequencies are available at all examined TRs, high frequencies could only be analyzed at shorter TRs. Each plot shows the mean correlation of the estimated mirrored EC with SC for different TRs (blue solid line) and correlation of the mean network over all subjects with SC (red dashed line). While the variability between subjects is quite high, the estimated

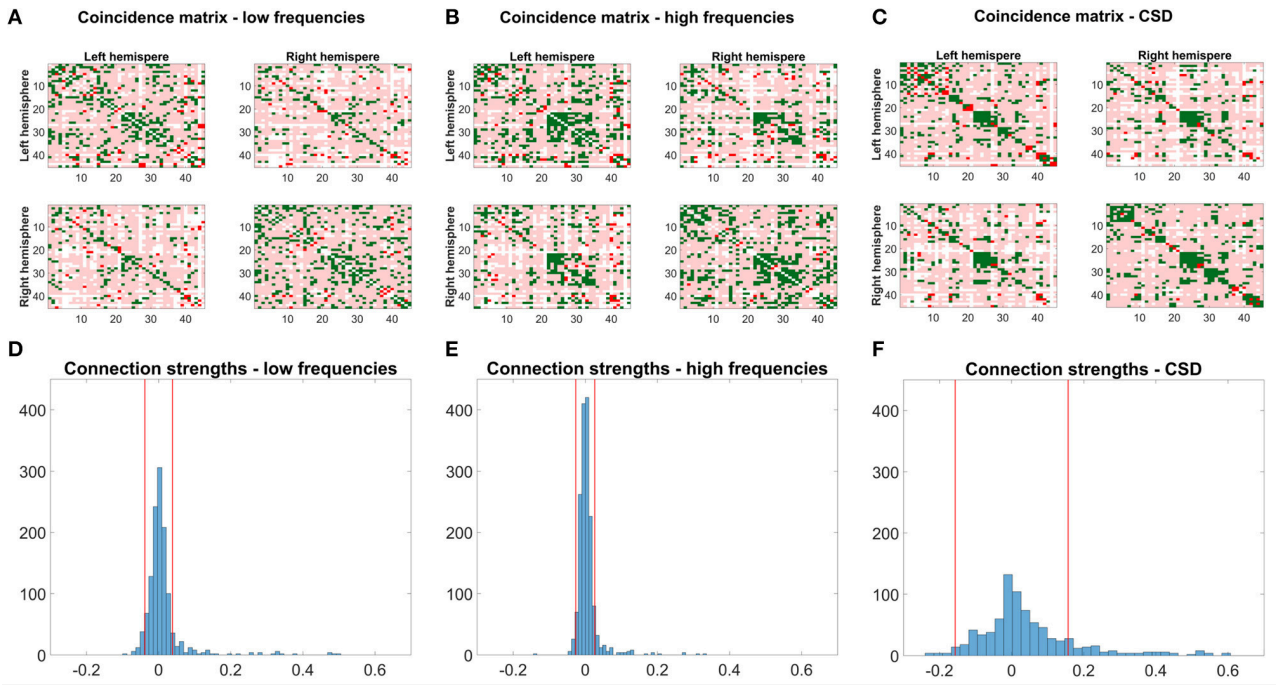

FIGURE 7 | Coincidence maps between structural connectivity from DTI and the mirrored low- **(A)** and high- **(B)** frequency networks and the cross-spectral density **(C)**. EC and FC networks were thresholded to yield a false positive rate (FPR) of 10%. Green entries depict connections existing in SC and EC/FC, bright red entries depict connections only existing in EC/FC. Pale red entries denote connections present in SC but not in EC/FC. **(D–F)** Shows a histogram of the connection strengths for connections not present in SC. The red lines depict the threshold chosen to obtain a FPR of 10%.

TABLE 4 | Coincidence of connections in SC and EC/FC.

	Connections in SC and EC/FC(%)	Connections in EC/FC, not SC(%)	Connections in SC, not EC/FC(%)	Network density EC/FC(%)	Network density SC(%)
Low frequencies	60.6	15.5 (9.9)	22.7	75.7 (43.6)	78
High frequencies	77.1	20.6 (11.6)	2.28	95.1 (49.9)	78
CSD	55.8	11.9	32.3	59.9	78
Low frequencies–adapted FPR	18.5	2.8	77.8	17.1	78
High frequencies–adapted FPR	25.0	2.8	71.3	22.3	78
CSD-adapted FPR	18.6	2.7	78.7	17.1	78

Percentage of connections agreeing between SC and EC/FC. EC networks are mirrored; values for non-mirrored EC are given in brackets. In the last three rows, the percentages of agreeing connections are shown for mirrored EC and FC whose false positive rate (FPR) was adapted to 10% for a better comparison.

network barely changes with decreasing TR, which is portrayed by a basically horizontal line for the correlation across different repetition times, indicating that the method is also appropriate at slower TRs. Nevertheless, the higher agreement between EC and SC found at higher frequencies can only be observed at TRs sufficiently short to observe such frequencies.

Default-Mode Network

From the estimated networks, the default-mode network was examined more closely. In **Figure 10**, the DMN is presented for both the low- and high-frequency bands. In **Figures 10A,B**, the low- and high-frequency band networks are shown with dots for the brain regions connected by red (positive weights) and blue (negative weights) lines. The width of the lines is proportional to the connection strength. As both networks are normalized to their respective strongest connection, the line thickness give only relative information about connection strength and cannot be compared directly across frequency bands.

For the low-frequency band the homologous brain areas were much more connected than for the high-frequency band. In the low-frequency network, all regions except TC were connected to their homologous regions. At high frequencies, only TC and mPFC had connections between homologous brain regions. Furthermore, the low-frequency network was much sparser, while the high-frequency network showed many quite strong connections, which were ordered in a symmetric fashion. For both networks, mPFC projects to other regions but does not receive input except from its homologous region in the other

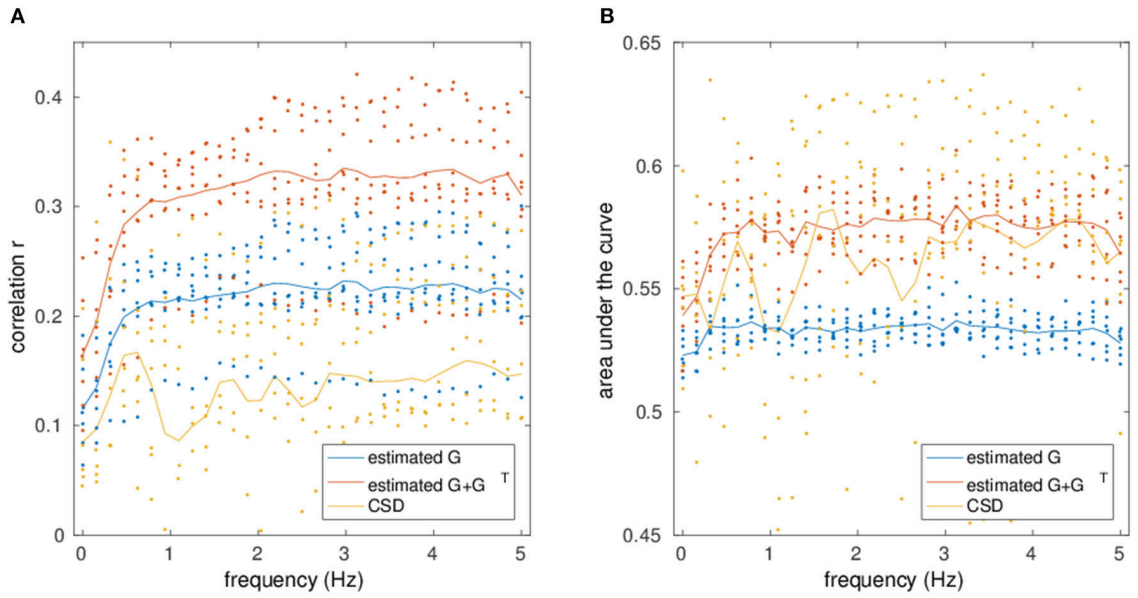

FIGURE 8 | Correlation **(A)** and AUC **(B)** values for the comparison between the estimated networks from fMRI/ CSD matrices and structural connectivity from DTI is shown at different frequencies.

TABLE 5 | Significance of agreement between EC/FC and SC.

	Correlation r		AUC of ROC	
Test	**Significant**	***p*-value**	**Significant**	***p*-value**
LOW FREQUENCY				
$G+G^T$ vs. CSD	1	0.01	0	0.79
G vs. *CSD*	0	0.84	0	0.24
HIGH FREQUENCY				
$G+G^T$ vs. CSD	1	<0.001	1	0.02
G vs. *CSD*	1	<0.001	0	0.15

Test for significance between estimated EC networks from fMRI and pure cross-spectral density matrices. The correlation and AUC values between estimated networks/CSD and SC from DTI were used and the mean value calculated for the high and the low frequency band for each subject.

hemisphere. Also IPC only projects to other regions; however, it receives input from its homologous region only at low frequencies. PCC both receives input from mPFC and IPC and projects further to HIP. The TC mainly receives input but for high frequencies it also projects to PCC.

DISCUSSION

We presented a method to estimate the effective connectivity from fMRI data, based on the symmetric cross-spectral density matrix of the acquired time series. As many different topologies can give rise to the same cross-spectral structure, the ambiguity in the estimation is resolved by using a L1-regularization preferring sparse networks. This is also a popular assumption in the GC framework, particularly in the case of voxel-wise connectivity estimation, due to the resulting large number of network nodes (Valdés-Sosa et al., 2005; Haufe et al., 2010; Garg et al., 2011; Tang et al., 2012), and is supported by the observation that cerebral connections tend to be highly selective (Valdés-Sosa et al., 2005; Sanchez-Bornot et al., 2008).

Simulated Data

As a proof of principle, the method was first applied to simulated data, where the influence of several parameters was analyzed.

Considering the length of the dataset and the SNR, increasing the SNR improved the estimation. An increase in the length of the time series also improved the estimation considerably up to a measurement time of 40 min, after which the increase became slower (cf. **Figure 2**). Although the methodology only depends on the cross-spectral density and not directly on the time course of the neuronal activity, increasing the length of the time series yields better estimates of the sample CSD. Furthermore, increasing the length of the time series allows for a finer frequency resolution of the CSD. However, long measurements might be problematic from the point of view of subject comfort and motion artifacts. Moreover, this assumes data stationarity, which is questionable for long measurements, particularly given the prevalence of dynamic connectivity states (Calhoun et al., 2014; Preti et al., 2016). Hence, a possible solution would be the acquisition of several shorter measurements and taking the mean cross-spectrum over the measurements and over connectivity states.

To further analyse the influence of the type of noise on the estimation, EC was estimated for pure white noise, pure pink noise and a mixture of both for different length of time series and SNR. Pure pink noise and the mixture gave similar results, mainly due to the high degree of pink noise in the mixed noise.

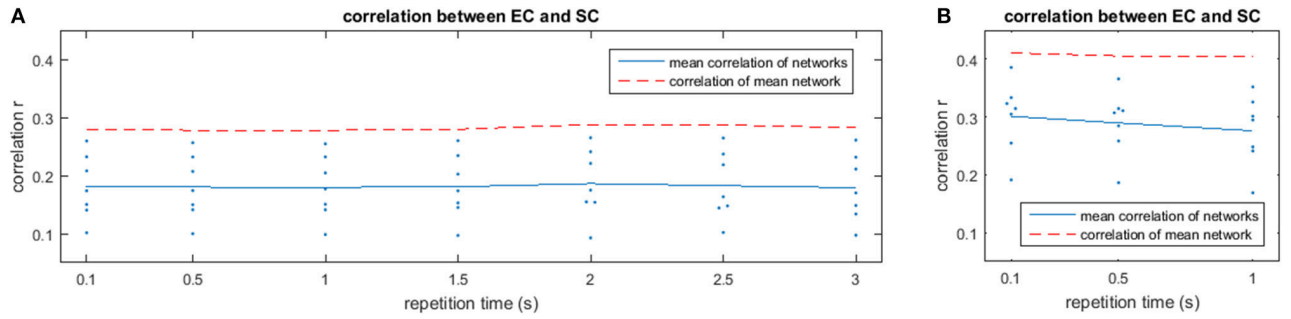

FIGURE 9 | Correlation values for the comparison between the estimated networks from fMRI matrices and structural connectivity from DTI is shown for different repetition times (TR) in the low-frequency **(A)** and high-frequency **(B)** bands. The red dashed line shows the correlation of the mean network over subjects with the structural connectivity from DTI.

For increasing SNR, the difference in estimation power for white, pink or mixed noise decreases, giving good results also for more realistic noise. But even for a lower SNR of 1, the difference is not very striking (cf. **Table 1**). The loss of estimation power for temporal noise is due to the correlation between adjacent data points, which leads to a loss of degrees of freedom and therefore a loss of information in the CSD. However, the connectivity information can nonetheless be retrieved from the data unless data sets are very short and with high temporal autocorrelation. For higher SNR the information from the true signal dominates the data, leading to better estimation results.

Often, effective connectivity analyses will only be performed on a small number of brain nodes of interest. This, however, poses the problem of missing nodes in the estimation of networks, which might lead to an erroneous estimation of the connectivity (Eichler, 2005; Daunizeau et al., 2011; Waldorp et al., 2011). For example, an indirect influence from a missing node on two nodes of interest might be interpreted as a spurious link between those two nodes. Therefore, the proposed method was applied to various network sizes that were part of larger networks with unobserved nodes. Not surprisingly, the estimation power increases monotonically with increasing fraction of observed nodes of the network (cf. **Figure 3**). This demonstrates the importance of minimizing the number of missing nodes. As fMRI data sets are usually whole brain scans and the computational efficiency of the proposed method allows for a high number of nodes (estimation time of a few minutes for a network with a hundred nodes on a standard computer), this problem can be overcome by estimating the full network and retrieving the partial network of interest afterwards.

The reduced dependence on HRF variability in comparison to lag-based Granger causality (cf. **Figure 4**) is especially relevant for fMRI given the indirect nature of the measured hemodynamic signals. This further confirms previous results on the possibility to estimate sparse networks from lag-free covariances (Pernice and Rotter, 2013; Schiefer et al., 2018). Nevertheless, it is clear that hemodynamic variability still acts as an important confounder on observed time-series correlations, so that integrating hemodynamic information, either via separate HRF estimates (Wu et al., 2013; Proulx et al., 2014) or by specifically including hemodynamics in the generative model (Ryali et al., 2011; Friston et al., 2014b) would be beneficial to EC estimation.

All in all, the analysis of the proposed method on simulated data proved quite successful, showing high agreement between the estimated and the true networks (cf. **Figure 5**). This might in part be due to the method used to generate the simulated data, which was closely matched to the estimation model. However, Pernice and Rotter (2013) also demonstrated good results in the estimation of networks of leaky integrate-and-fire neurons, suggesting that the analysis is applicable to a wide variety of different data types.

fMRI Data

In a second step, the proposed method was applied to fast fMRI data. Correlating the estimated networks over frequencies suggested a clustering in two frequency bands: A low-frequency band from $f = 0$–0.31 Hz and a high-frequency band $f = 0.31$–5 Hz. Interestingly, high-frequency BOLD signal fluctuations above $0.1-0.2$ Hz have rarely been considered in conventional functional connectivity analyses. However, recently emerged fast fMRI sequences allow to analyse such higher frequencies, with multiple studies suggesting that they contain relevant information (Lee et al., 2013; Yuan et al., 2014; Trapp et al., 2017). High-frequency connectivity could not be examined in our simulations as it was completely attenuated by the convolution with the canonical HRF, but it has been recently reported that resting-state fMRI may be driven by narrower HRFs with non-negligible contributions at high frequencies (Chen and Glover, 2015). This is also in line with another recent study that found that information in Granger causality estimates is carried at frequencies up to 3 Hz in fMRI data (Lin et al., 2015).

The estimated networks showed a strong similarity between hemispheres ($r = 0.8$ for the low frequency network, $r = 0.9$ for high frequencies), which would be expected. Both networks showed strong intra-hemispheric connections in the frontal, occipital and parietal lobes (see **Figure 6**). Although the networks for low- and high-frequencies were similar, the low-frequency network was much sparser and less symmetric than the high-frequency network. Moreover,

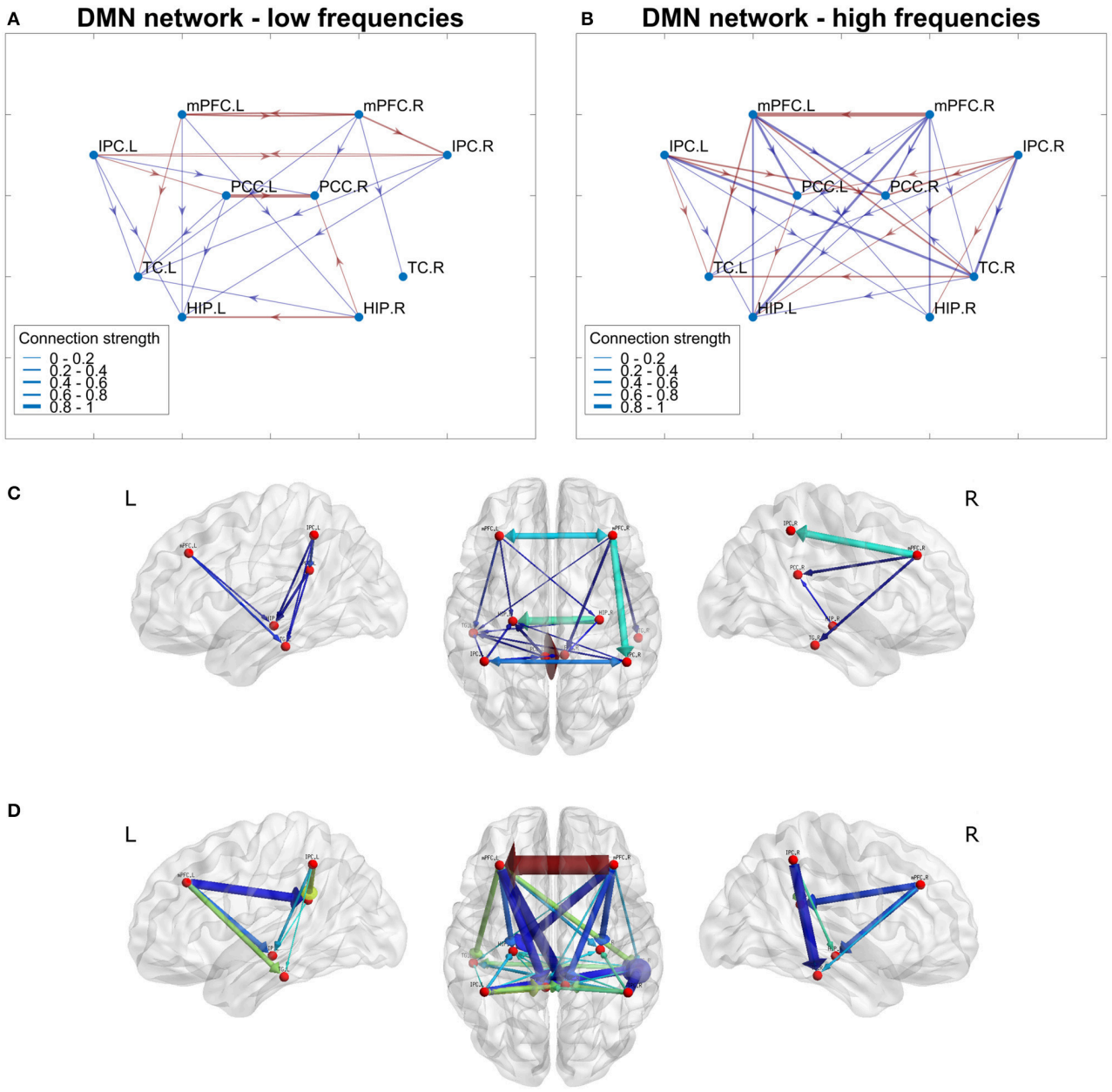

FIGURE 10 | Networks for low- and high-frequency bands of the default-mode network (DMN). In **(A,B)**, network connections are depicted by lines between regions. Red lines correspond to positive connections and blue lines to negative connections. Connection strength ranges are illustrated by the line thickness. The connection strengths for both networks are normalized to the strongest connection in the network. **(C,D)** Show the same networks (low frequencies **C**, high frequencies **D**) where the network is projected onto a generic brain. Both color and arrow thickness represents absolute connection strengths (The brain networks in **(C,D)** were visualized with BrainNet Viewer (http://www.nitrc.org/projects/bnv/) (Xia et al., 2013).

the low-frequency network showed strong inter-hemispheric connections between homologous brain regions in both hemispheres, which is less pronounced in the high-frequency network.

Due to the lack of gold standard, the estimated effective connectivity networks were compared to structural connectivity from DTI. While the SC networks were very stable over subjects, the EC networks showed quite high variability. This is not very surprising, since the SC network corresponds to the "hard wiring" of the brain, which is expected to be similar for different individuals. The EC, however, is estimated from resting-state data. Although some general patterns evolve in resting-state data, the processing network might vary strongly between subjects (Mueller et al., 2013). Furthermore, during the measurement,

subjects might not always be in perfect resting state, but might let their mind wander leading to an altered network (Kucyi, 2017).

The comparison between EC and SC yielded relatively low correlation values, which could be partly attributed to the symmetry of undirected SC measures. The correlation increased significantly after mirroring the estimated EC. The agreement between the high-frequency network and SC was significantly higher than for the low-frequency network. In part this is due to the presence of strong inter-hemispheric connections between almost all homologous brain regions in the low-frequency band, which is a typical point of discordance with SC (Messé et al., 2014). For the high-frequency band, however, these inter-hemispheric connections were less pronounced yielding much higher agreement between EC and SC. Furthermore, the low-frequency network is much sparser than both the high-frequency network and SC, hinting to the idea that for low frequencies much fewer structural connections are active than for high frequencies due to relatively short conduction delays between brain areas leading to activity at higher frequencies. We emphasize again, however, that comparison with SC cannot be considered a strict validation, since the static physical connections in SC cannot represent the dynamic connections active in any given time or brain state. Moreover, diffusion tractography itself only provides an imperfect estimate of SC and may miss major structural links, notably interhemispheric connections (Robinson et al., 2014). Nevertheless, consistency with SC can still provide some evidence of a successful EC estimation, especially when considering functional connections not supported by an underlying structural connection. However, we cannot draw firm conclusions on the performance of the method regarding the converse situation, that is, structural connections that may or may not result in an identified functional connection. A true validation of EC in humans would require invasive approaches such as intracranial EEG, which is employed for clinical purposes in some epilepsy patients. Unlike the non-invasive DTI approach used in the current study, intracranial EEG can provide directed measures of effective connectivity (Wendling et al., 2010; Entz et al., 2014) and would be well worth investigating in the future.

As the most widespread technique for functional connectivity analysis is the computation of undirected, potentially band-limited correlations, the proposed method was also compared to the pure cross-spectral density and its agreement with SC. FC was actually better than the normal, unmirrored EC for AUC ($p_{AUC} = 0.02$). However, when considering the mirrored EC network, there was significantly better agreement than FC. Moreover, CSD exhibited very high variance both for correlation and AUC compared to normal and mirrored EC. Falsely identified connections (without an underlying structural connection) were very weak for EC but covered a broad range of connectivity strengths for FC (cf. **Figure 7**). Thus, the estimated EC networks showed more consistency with DTI than the functional networks from CSD. The significance values were however relatively low, which is due to the low number of subjects used to estimate the networks.

As typical fMRI sequences have much lower temporal resolution than the sequence used in this study, the network estimation was performed on datasets that were retrospectively downsampled to longer TRs to analyse the influence of the temporal resolution on the estimation. Correlating the estimated networks with SC showed a relatively stable estimation of the networks even at lower temporal resolutions. The overall results suggest that the measurement length is more important than the number of data points for a given scan time. Note however, that only low-frequency networks can be recovered at long TRs. Higher temporal resolutions was still beneficial for the estimation of high-frequency connectivity (see **Figure 9B**), which showed better estimation performance than low-frequency networks, as well as preprocessing advantages such as better physiological noise removal (Lin et al., 2012; Jacobs et al., 2014; Korhonen et al., 2014).

Finally, the directed connectivity for the default mode network was retrieved from the estimated network (**Figure 10**). The low frequency network is much sparser and less symmetrical than the high frequency network. Compared to the results of Miao et al. (2010), who did a Granger causality analysis on the DMN, similar results are obtained, notably the strong connections from all other regions to the PCC. However, differences are also observed in the mPFC where we found mostly outgoing rather than ingoing connections. One potential cause of this discrepancy may be the particular sensitivity of the employed MREG sequence to off-resonances in the mPFC, leading to potential artifacts (Zahneisen et al., 2012; Assländer et al., 2013). Future work will focus on further validation of the inferred directed networks.

CONCLUSION

In this paper we presented a method to estimate the effective connectivity from whole brain resting-state fMRI scans from the cross-spectral density in the frequency domain. The influence of different measurement parameters was analyzed in simulated fMRI data, notably showing a reduced dependency on hemodynamic variability compared to lag-based methods such as Granger Causality. The proposed method was further applied to resting-state fMRI data, showing improved consistency with the underlying structural connectivity networks obtained from DTI tractography in comparison to conventional functional connectivity.

AUTHOR CONTRIBUTIONS

CL, JS, SR, JH, and PL: Conceived and designed the study; CL and JS: Analyzed the data; CL, JS, SR, JH, and PL: Drafted and approved the manuscript.

FUNDING

This work was funded by the DFG Excellence Cluster BrainLinks-BrainTools (grant EXC 1086) and the DFG Grant HE 1875/28-1. The HPC facilities are funded by the state of Baden-Württemberg through bwHPC and DFG grant INST 39/963-1 FUGG. The article processing charge was covered by the open access publication fund of the University of Freiburg.

REFERENCES

Abrudan, T. E., Eriksson, J., and Koivunen, V. (2008). Steepest descent algorithms for optimization under unitary matrix constraint. *IEEE Trans. Signal Process.* 56, 1134–1147. doi: 10.1109/TSP.2007.908999

Abrudan, T., Eriksson, J., and Koivunen, V. (2009). Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Process.* 89, 1704–1714. doi: 10.1016/j.sigpro.2009.03.015

Aertsen, A., and Preißl, H. (1991). *Dynamics of Activity and Connectivity in Physiological Neuronal Networks*. Weinheim: VCH Verlag.

Akin, B., Lee, H.-L., Hennig, J., and LeVan, P. (2017). Enhanced subject-specific resting-state network detection and extraction with fast fMRI: enhanced subject-specific RSN detection. *Hum. Brain Mapp.* 38, 817–830. doi: 10.1002/hbm.23420

Arand, C., Scheller, E., Seeber, B., Timmer, J., Klöppel, S., and Schelter, B. (2015). Assessing parameter identifiability for dynamic causal modeling of fMRI data. *Front. Neurosci.* 9:43. doi: 10.3389/fnins.2015.00043

Assländer, J., Zahneisen, B., Hugger, T., Reisert, M., Lee, H.-L., LeVan, P., et al. (2013). Single shot whole brain imaging using spherical stack of spirals trajectories. *Neuroimage* 73, 59–70. doi: 10.1016/j.neuroimage.2013.01.065

Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990

Barnett, L., and Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* 223, 50–68. doi: 10.1016/j.jneumeth.2013.10.018

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541. doi: 10.1002/mrm.1910340409

Bressler, S. L., and Seth, A. K. (2011). Wiener-Granger causality: a well established methodology. *Neuroimage* 58, 323–329. doi: 10.1016/j.neuroimage.2010.02.059

Bringmann, L. F., Scholte, H. S., and Waldorp, L. J. (2013). Matching structural, effective, and functional connectivity: a comparison between structural equation modeling and ancestral graphs. *Brain Connect.* 3, 375–385. doi: 10.1089/brain.2012.0130

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39, 855–864. doi: 10.1002/mrm.1910390602

Calhoun, V. D., Miller, R., Pearlson, G., and Adali, T. (2014). The Chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015

Chen, J. E., and Glover, G. H. (2015). BOLD fractional contribution to resting-state functional connectivity above 0.1Hz. *Neuroimage* 107, 207–218. doi: 10.1016/j.neuroimage.2014.12.012

Crimi, A., Dodero, L., Murino, V., and Sona, D. (2017). "Case-control discrimination through effective brain connectivity," in *Proceedings of IEEE 14th International Symposium on Biomedical Imaging* (Melbourne), 970–973.

Dang, S., Chaudhury, S., Lall, B., and Roy, P. K. (2018). Tractography-based score for learning effective connectivity from multimodal imaging data using dynamic bayesian networks. *IEEE Trans. Biomed. Eng.* 65, 1057–1068. doi: 10.1109/TBME.2017.2738035

Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage* 58, 312–322. doi: 10.1016/j.neuroimage.2009.11.062

Deshpande, G., LaConte, S., James, G. A., Peltier, S., and Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. *Hum. Brain Mapp.* 30, 1361–1373. doi: 10.1002/hbm.20606

Duggento, A., Bianciardi, M., Passamonti, L., Wald, L. L., Guerrisi, M., Barbieri, R., et al. (2016). Globally conditioned Granger causality in brain-brain and brain-heart interactions: a combined heart rate variability/ultra-high-field (7 T) functional magnetic resonance imaging study. *Philos. Trans. A Math. Phys. Eng. Sci.* 374:20150185. doi: 10.1098/rsta.2015.0185

Eichler, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 953–967. doi: 10.1098/rstb.2005.1641

Entz, L., Tóth, E., Keller, C. J., Bickel, S., Groppe, D. M., Fabó, D., et al. (2014). Evoked effective connectivity of the human neocortex. *Hum. Brain Mapp.* 35, 5736–5753. doi: 10.1002/hbm.22581

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., et al. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS ONE* 5:e15710. doi: 10.1371/journal.pone.0015710

Finger, H., Bönstrup, M., Cheng, B., Messé, A., Hilgetag, C., Thomalla, G., et al. (2016). Modeling of large-scale functional brain networks based on structural connectivity from DTI: comparison with EEG derived phase coupling networks and evaluation of alternative methods along the modeling path. *PLoS Comput. Biol.* 12:e1005025. doi: 10.1371/journal.pcbi.1005025

Fisher, R. S., Cross, J. H., French, J. A., Higurashi, N., Hirsch, E., Jansen, F. E., et al. (2017). Operational classification of seizure types by the international league against epilepsy: position paper of the ILAE commission for classification and terminology. *Epilepsia* 58, 522–530. doi: 10.1111/epi.13670

Frässle, S., Stephan, K. E., Friston, K. J., Steup, M., Krach, S., Paulus, F. M., et al. (2015). Test-retest reliability of dynamic causal modeling for fMRI. *Neuroimage* 117, 56–66. doi: 10.1016/j.neuroimage.2015.05.040

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78. doi: 10.1002/hbm.460020107

Friston, K. J., Bastos, A. M., Oswal, A., van Wijk, B., Richter, C., and Litvak, V. (2014a). Granger causality revisited. *Neuroimage* 101, 796–808. doi: 10.1016/j.neuroimage.2014.06.062

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K. J., Kahan, J., Biswal, B., and Razi, A. (2014b). A DCM for resting state fMRI. *Neuroimage* 94, 396–407. doi: 10.1016/j.neuroimage.2013.12.009

Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. (2000). Nonlinear responses in fMRI: the balloon model, Volterra Kernels, and other hemodynamics. *Neuroimage* 12, 466–477. doi: 10.1006/nimg.2000.0630

Garg, R., Cecchi, G. A., and Rao, A. R. (2011). Full-brain auto-regressive modeling (FARM) using fMRI. *Neuroimage* 58, 416–441. doi: 10.1016/j.neuroimage.2011.02.074

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* 77, 304–313. doi: 10.1080/01621459.1982.10477803

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* 79, 907–915. doi: 10.1080/01621459.1984.10477110

Gilson, M., Moreno-Bote, R., Ponce-Alvarez, A., Ritter, P., and Deco, G. (2016). Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput. Biol.* 12:e1004762. doi: 10.1371/journal.pcbi.1004762

Glover, G. H., Li, T. Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* 44, 162–167. doi: 10.1002/1522-2594(200007)44:1<162::AID-MRM23>3.0.CO;2-E

Goebel, R., Roebroeck, A., Kim, D. S., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging.* 21, 1251–1261. doi: 10.1016/j.mri.2003.08.026

Gudbjartsson, H., and Patz, S. (1995). The rician distribution of noisy MRI data. *Magn. Reson. Med.* 34, 910–914. doi: 10.1002/mrm.1910340618

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651. doi: 10.1016/j.neuroimage.2003.11.029

Harrison, L., Penny, W. D., and Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage* 19, 1477–1491. doi: 10.1016/S1053-8119(03)00160-5

Haufe, S., Nolte, G., Müller, K. R., and Krämer, N. (2010). Sparse causal discovery in multivariate time series. *J. Mach. Learn. Res.* 6, 97–106.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 83–90. doi: 10.1093/biomet/58.1.83

Hugger, T., Zahneisen, B., LeVan, P., Lee, K. J., Lee, H.-L., Zaitsev, M., et al. (2011). Fast undersampled functional magnetic resonance imaging using nonlinear regularized parallel image reconstruction. *PLoS ONE* 6:e28822. doi: 10.1371/journal.pone.0028822

Jacobs, J., Stich, J., Zahneisen, B., Assländer, J., Ramantani, G., Schulze-Bonhage, A., et al. (2014). Fast fMRI provides high statistical power

in the analysis of epileptic networks. *Neuroimage* 88, 282–294. doi: 10.1016/j.neuroimage.2013.10.018

Korhonen, V., Hiltunen, T., Myllylä, T., Wang, X., Kantola, J., Nikkinen, J., et al. (2014). Synchronous multiscale neuroimaging environment for critically sampled physiological analysis of brain function: hepta-scan concept. *Brain Connect*. 4, 677–689. doi: 10.1089/brain.2014.0258

Kucyi, A. (2017). Just a thought: how mind-wandering is represented in dynamic brain connectivity. *Neuroimage*. doi: 10.1016/j.neuroimage.2017.07.001. [Epub ahead of print].

Lee, H.-L., Zahneisen, B., Hugger, T., LeVan, P., and Hennig, J. (2013). Tracking dynamic resting-state networks at higher frequencies using MR-encephalography. *Neuroimage* 65, 216–222. doi: 10.1016/j.neuroimage.2012.10.015

LeVan, P., Akin, B., and Hennig, J. (2017). Fast imaging for mapping dynamic networks. *Neuroimage*. doi: 10.1016/j.neuroimage.2017.08.029. [Epub ahead of print].

Li, B., Daunizeau, J., Stephan, K. E., Penny, W., Hu, D., and Friston, K. (2011). Generalised filtering and stochastic DCM for fMRI. *Neuroimage* 58, 442–457. doi: 10.1016/j.neuroimage.2011.01.085

Li, L., Rilling, J. K., Preuss, T. M., Glasser, M. F., Damen, F. W., and Hu, X. (2012). Quantitative assessment of a framework for creating anatomical brain networks via global tractography. *Neuroimage* 61, 1017–1030. doi: 10.1016/j.neuroimage.2012.03.071

Lin, F.-H., Ahveninen, J., Raij, T., Witzel, T., Chu, Y.-H., Jääskeläinen, I. P., et al. (2014). Increasing fMRI sampling rate improves granger causality estimates. *PLoS ONE* 9:e100319. doi: 10.1371/journal.pone.0100319

Lin, F. H., Chu, Y. H., Hsu, Y. C., Lin, J. F., Tsai, K. W., Tsai, S. Y., et al. (2015). Significant feed-forward connectivity revealed by high frequency components of BOLD fMRI signals. *Neuroimage* 121, 69–77. doi: 10.1016/j.neuroimage.2015.07.036

Lin, F.-H., Nummenmaa, A., Witzel, T., Polimeni, J. R., Zeffiro, T. A., Wang, F.-N., et al. (2012). Physiological noise reduction using volumetric functional magnetic resonance inverse imaging. *Hum. Brain Mapp*. 33, 2815–2830. doi: 10.1002/hbm.21403

Messé, A., Rudrauf, D., Benali, H., and Marrelec, G. (2014). Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. *PLoS Comput. Biol*. 10:e1003530. doi: 10.1371/journal.pcbi.1003530

Miao, X., Chen, K., Li, R., Wen, X., Yao, L., and Wu, X. (2010). "Application of Granger causality analysis to effective connectivity of the default-mode network," in *IEEE/ICME International Conference on Complex Medical Engineering* (Gold Coast), 156–160.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., et al. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron* 77, 586–595. doi: 10.1016/j.neuron.2012.12.028

Ofer, I., Jacobs, J., Jaiser, N., Akin, B., Hennig, J., Schulze-Bonhage, A., et al. (2018). Cognitive and behavioral comorbidities in rolandic epilepsy and their relation with default mode network's functional connectivity and organization. *Epilepsy Behav.* 78, 179–186. doi: 10.1016/j.yebeh.2017.10.013

Park, H. J., Friston, K. J., Pae, C., Park, B., and Razi, A. (2017). Dynamic effective connectivity in resting state fMRI. *Neuroimage*. doi: 10.1016/j.neuroimage.2017.11.033. [Epub ahead of print].

Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage* 59, 319–330. doi: 10.1016/j.neuroimage.2011.07.039

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of dynamic causal models. *PLoS Comput. Biol*. 6:e1000709. doi: 10.1371/journal.pcbi.1000709

Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Comparing dynamic causal models. *Neuroimage* 22, 1157–1172. doi: 10.1016/j.neuroimage.2004.03.026

Pernice, V., and Rotter, S. (2013). Reconstruction of sparse connectivity in neural networks from spike train covariances. *J. Stat. Mech. Theory Exp*. 2013:P03008. doi: 10.1088/1742-5468/2013/03/P03008

Petersen, S. E., and Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron* 88, 207–219. doi: 10.1016/j.neuron.2015.09.027

Posse, S., Ackley, E., Mutihac, R., Rick, J., Shane, M., Murray-Krezan, C., et al. (2012). Enhancement of temporal resolution and BOLD sensitivity in real-time fMRI using multi-slab echo-volumar imaging. *Neuroimage* 61, 115–130. doi: 10.1016/j.neuroimage.2012.02.059

Preti, M. G., Bolton, T. A., and Van de Ville, D. (2016). The dynamic functional connectome: state-of-the-art and perspectives. *NeuroImage* 160, 41–54. doi: 10.1016/j.neuroimage.2016.12.061

Proulx, S., Safi-Harb, M., Levan, P., An, D., Watanabe, S., and Gotman, J. (2014). Increased sensitivity of fast BOLD fMRI with a subject-specific hemodynamic response function and application to epilepsy. *Neuroimage* 93(Pt 1), 59–73. doi: 10.1016/j.neuroimage.2014.02.018

Purdon, P. L., and Weisskoff, R. M. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum. Brain Mapp*. 6, 239–249. doi: 10.1002/(SICI)1097-0193(1998)6:4<239::AID-HBM4>3.0.CO;2-4

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. A., and Glymour, C. (2010). Six problems for causal inference from fMRI. *Neuroimage* 49, 1545–1558. doi: 10.1016/j.neuroimage.2009.08.065

Reisert, M., Mader, I., Umarova, R., Maier, S., Tebartz van Elst, L., and Kiselev, V. G. (2013). Fiber density estimation from single q-shell diffusion imaging by tensor divergence. *Neuroimage* 77, 166–176. doi: 10.1016/j.neuroimage.2013.03.032

Robinson, P. A., Sarkar, S., Pandejee, G. M., and Henderson, J. A. (2014). Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 90:012707. doi: 10.1103/PhysRevE.90.012707

Rogers, B. P., Katwal, S. B., Morgan, V. L., Asplund, C. L., and Gore, J. C. (2010). Functional MRI and multivariate autoregressive models. *Magn. Reson. Imaging*. 28, 1058–1065. doi: 10.1016/j.mri.2010.03.002

Ryali, S., Supekar, K., Chen, T., and Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fMRI. *Neuroimage* 54, 807–823. doi: 10.1016/j.neuroimage.2010.09.052

Samdin, S. B., Ting, C. M., Ombao, H., and Salleh, S. H. (2017). A unified estimation framework for state-related changes in effective brain connectivity. *IEEE Trans. Biomed. Eng.* 64, 844–858. doi: 10.1109/TBME.2016.2580738

Sanchez-Bornot, J. M., Martinez-Montes, E., Lage-Castellanos, A., Vega-Hernandez, M., and Valdes-Sosa, P. A. (2008). Uncovering sparse brain effective connectivity: a voxel-based approach using penalized regression. *Stat. Sin*. 18, 1501–1518.

Schiefer, J., Niederbühl, A., Pernice, V., Lennartz, C., Hennig, J., LeVan, P., et al. (2018). From correlation to causation: estimating effective connectivity from zero-lag covariances of brain signals. *PLoS Comput. Biol.* 14:e1006056. doi: 10.1371/journal.pcbi.1006056

Schiefer, J., and Rotter, S. (2016). "Inference of cerebral network structure," in *Presented at the 2nd International Conference on Mathematical NeuroScience* (Juan-les-Pins).

Seth, A. K., Chorley, P., and Barnett, L. C. (2013). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *Neuroimage* 65, 540–555. doi: 10.1016/j.neuroimage.2012.09.049

Smith, J. F., Chen, K., Pillai, A. S., and Horwitz, B. (2013). Identifying effective connectivity parameters in simulated fMRI: a direct comparison of switching linear dynamic system, stochastic dynamic causal, and multivariate autoregressive models. *Front Neurosci.* 7:70. doi: 10.3389/fnins.2013.00070

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Stephan, K. E. (2004). On the role of general system theory for functional neuroimaging. *J. Anat.* 205, 443–470. doi: 10.1111/j.0021-8782.2004.00359.x

Tang, W., Bressler, S. L., Sylvester, C. M., Shulman, G. L., and Corbetta, M. (2012). Measuring granger causality between cortical regions from voxelwise fMRI BOLD signals with LASSO. *PLoS Comput Biol.* 8:e1002513. doi: 10.1371/journal.pcbi.1002513

Trapp, C., Vakamudi, K., and Posse, S. (2017). On the detection of high frequency correlations in resting state fMRI. *NeuroImage* 164, 202–213. doi: 10.1016/j.neuroimage.2017.01.059

Uddin, L. Q., Supekar, K. S., Ryali, S., and Menon, V. (2011). Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development. *J. Neurosci.* 31, 18578–18589. doi: 10.1523/JNEUROSCI.4465-11.2011

Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., et al. (2005). Estimating brain functional

connectivity with sparse multivariate autoregression. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 360, 969–981. doi: 10.1098/rstb.2005.1654

Waldorp, L., Christoffels, I., and van de Ven, V. (2011). Effective connectivity of fMRI data using ancestral graph theory: dealing with missing regions. *Neuroimage* 54, 2695–2705. doi: 10.1016/j.neuroimage.2010.10.054

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., et al. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277. doi: 10.1016/j.neuroimage.2008.03.036

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011). neuRosim: an R package for generating fMRI data. *J. Stat. Softw.* 44, 1–18. doi: 10.18637/jss.v044.i10

Wendling, F., Chauvel, P., Biraben, A., and Bartolomei, F. (2010). From intracerebral EEG signals to brain connectivity: identification of epileptogenic networks in partial epilepsy. *Front. Syst. Neurosci.* 4:154. doi: 10.3389/fnsys.2010.00154

Wu, G. R., Liao, W., Stramaglia, S., Ding, J. R., Chen, H., and Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* 17, 365–374. doi: 10.1016/j.media.2013.01.003

Xia, M., Wang, J., and He, Y. (2013). BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS ONE* 8:e68910. doi: 10.1371/journal.pone.0068910

Yuan, B.-K., Wang, J., Zang, Y.-F., and Liu, D.-Q. (2014). Amplitude differences in high-frequency fMRI signals between eyes open and eyes closed resting states. *Front. Hum. Neurosci.* 8:503. doi: 10.3389/fnhum.2014.00503

Zahneisen, B., Hugger, T., Lee, K. J., LeVan, P., Reisert, M., Lee, H.-L., et al. (2012). Single shot concentric shells trajectories for ultra fast fMRI. *Magn. Reson. Med.* 68, 484–494. doi: 10.1002/mrm.23256

Conflict of Interest Statement: 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.

Estimating connectivity from simulated ECoG signals

Jonathan Schiefer, Lisa M. Schäfer and Stefan Rotter

April 18, 2018

Abstract

Estimating brain connectivity has been a widely studied field in recent neuroscience. Lacking access to the ground truth is a major challenge for the validation of the resulting estimates. In this study we introduce a detailed simulation of electrocorticographic (ECoG) signals and use them to validate the performance of estimation methods on these signals. We simulate a network of networks consisting of leaky-integrate-and-fire (LIF) neurons from which we derive local field potentials. Then, we estimate the connectivity by fitting autoregressive processes (AR) to the simulated signals of the neural populations. Also, we introduce a eigenvalue-spectrum-based distance of networks to quantify the difference of arbitrary networks. We show that methods based on temporal information as well as covariance based estimation methods can achieve good estimations of neural population connectivity. We also show that the spectral distance of networks reflects the same characteristics of differences of networks like the area under the ROC curve and precision recall score. Our results indicate that autoregressive processes reflect the key features of the dynamics of neural population and the estimation of connectivity based on such processes lead to valid estimates.

Introduction

Estimating connectivity reflects multiple challenges of modern neuroscience: it is multi-scale and can be defined and described in multiple ways: as structural, functional and effective connectivity [1]. Access to the connectivity of distributed networks of interacting brain systems is thought of being a key to understand function and dysfunction of the brain [2]. Originating from the concept of Granger causality [3], there are multiple methods to estimate effective connectivity based on fitting an autoregressive process to the data. The original method is based on the coherence derived from the partial cross-spectrum, variants of it use the temporal information directly [4]. Others are based on the covariance of the measured data [5, 6, 7, 8]. All of them work fairly well on linear processes.

The estimation method used should always fit the type of data available, and match the temporal and spatial resolution as well as the amount of data which was recorded. Some of them have been applied on measured electrocorticography data [7, 9]. However, it remains unclear, whether fitting an autoregressive process is a suitable method for ECoG data which consists of accumulated post-synaptic currents [10, 11, 12]. This implies that the dynamics of neural populations, which consist of non-linearly interacting neurons, can be modeled with linear processes. Also, the lack of access to the ground truth induces a challenge of validating resulting networks. Often, changes in connectivity over time are studied, but an objective quantification of the difference of networks is still lacking. For the comparison of resulting networks, measures like area under the ROC curve and precision recall score are used. However, these measures assume access to the ground truth. Other comparisons of networks like [13] are only applicable for undirected networks of the same size. In order to achieve insights about function and dysfunction of the brain based on changes in connectivity, there is the

need for validation of the estimation methods as well as a comparable quantification of the changes in the connectivity.

In this study, we introduce a simulation of ECoG signals based on leaky-integrate-and-fire neurons. We focus on the estimation of effective connectivity based on data with high temporal resolution. Our simulation consists of 64 interconnected neural networks representing the neural populations underneath the ECoG electrode. Based on this simulation, we study the performance of a temporal based least-square estimation of AR coefficients and compare the results to a covariance based method. Also, based on the eigenvalue spectra of networks, we introduce an universal procedure to compare directed networks of arbitrary sizes (including networks of different sizes).

Materials and methods

Simulating ECoG data

Our goal is to establish a computational model for signals recorded by electrocorticography (ECoG) electrodes. ECoG electrodes are usually placed on the exposed surface of the cortex and record a composition of post-synaptic potentials [10, 11, 12]. ECoG electrodes differ in size and the amount of electrodes, for our model we use a 64 electrode (8×8) grid as template. Existing models for ECoG signals [14] are not based on the neuronal activity but on a multivariate tuning model. In contrast to this, our modeling approach consists of a network of networks of neurons, each sub-network representing the neural population recorded by one ECoG electrode. In our case, this means we model 64 neural populations, each consisting of 1 000 neurons. Each neural population is a scaled version of a Brunel network [15] consisting of 1 000 LIF neurons, where 800 neurons are excitatory and 200 neurons are inhibitory. For all neurons we use a membrane time constant of $\tau_{\mathrm{m}} = 20\,\mathrm{ms}$ and a reset potential of $V_{\mathrm{reset}} = -70\,\mathrm{mV}$. Each neuron receives individual Poisson input with a constant rate. All simulations are performed using NEST [16] and Python [17]. All parameters used in the simulation can be found in Table 1.

The neurons within each population are randomly connected with a connection probability of 10%, purely by chemical synapses with α-shaped post-synaptic potentials. For all synapses we use time constants $\tau_{\mathrm{syn}} = 2\,\mathrm{ms}$, a synaptic delay of 1.5 ms and connection weight of $w = 10\,\mathrm{mV}$. The ratio of inhibitory connection strength to excitatory connection strength is $g = 5.0$. Additionally, the excitatory neurons also form long-range connections to neurons in other populations according to a pre-defined connectivity scheme, cf. Fig 1. While having the same weights as the synapses within a population, the synaptic delays for these connections are 3 ms. Inhibitory neurons have been reported to have local connectivity [18], so in our model they form connection within populations but not to other populations.

For the meta connectivity of the populations we used a network with both, a distance dependent component and a random Erdős-Rényi component. The connectivity matrix is shown in the left panel of Fig 2. Each population is connected with its direct neighbors and also with neighboring populations up to two steps away. These connections are represented by the off-diagonal entries in the connectivity matrix. We additionally created a random Erdős-Rényi network without any distance dependency and a connection probability of 10%. The connections of both meta connectivity matrices, the distance dependent one and the Erdős-Rényi one, are realized by connecting random neurons of the corresponding populations.

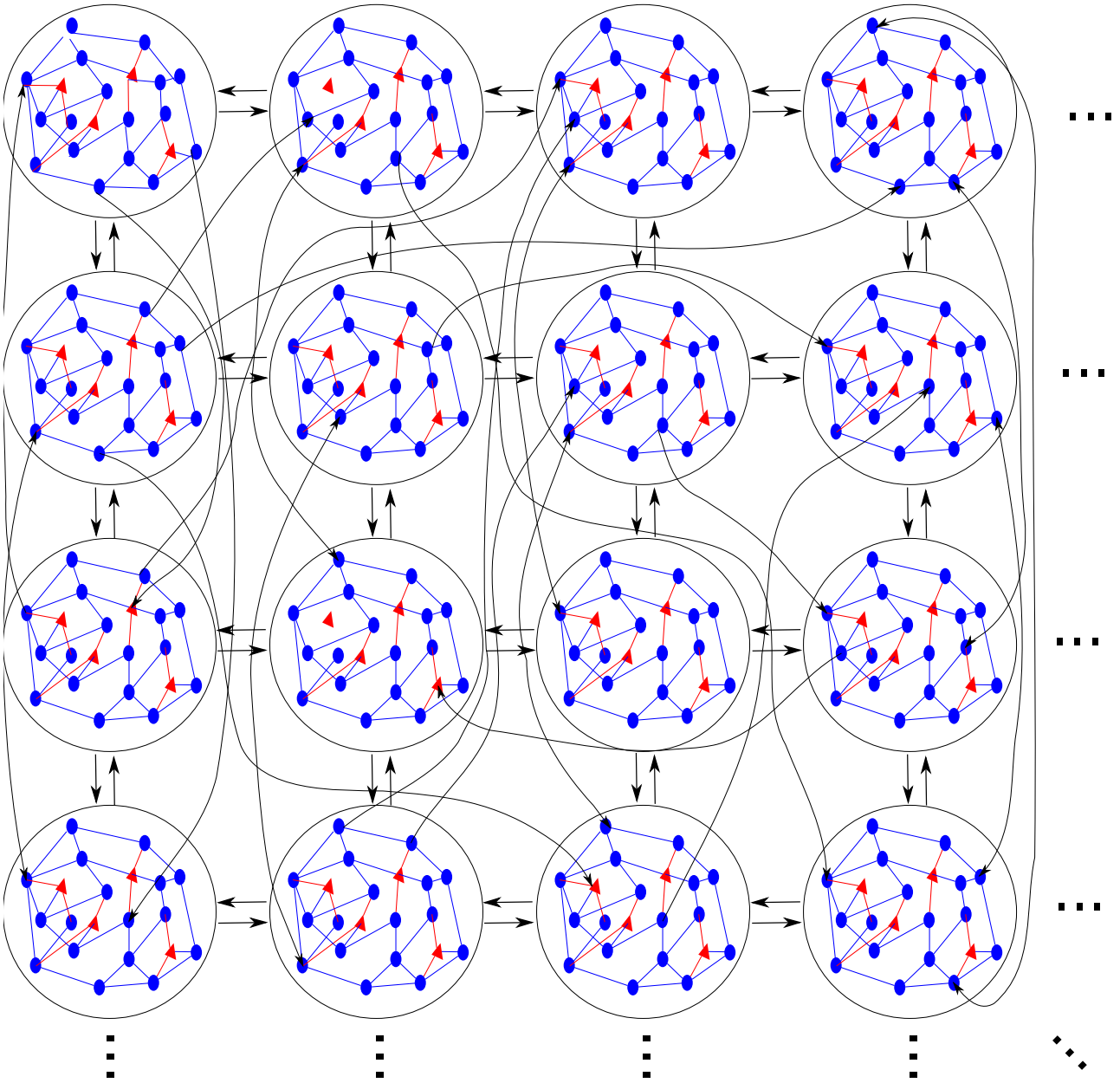

Figure 1: Scheme of the simulated network. Each circle represents the neural population underneath one ECoG electrode. Within the populations there is a random connectivity of 10%, excitatory neurons additionally form long-range connections to other populations following a pre-defined meta connectivity. Here, the inter-population connectivity of upper-left 4×4 electrodes are shown, in total we simulated 8×8 electrodes.

As local field potentials (LFPs) are thought of being generated by postsynaptic currents [10, 11, 12], we calculate the signal of each electrode similar to [19] by recording the absolute value of the summed synaptic currents of synapses terminating at the excitatory populations.

Inhomogeneous input

In addition to the model described above, we simulated an extended version of the model including inhomogeneous inputs to the neurons. Inspired by [20], we choose to use inhomogeneous input to the neurons such that the power-spectra of the simulated signals are self-consistent. The idea is to iteratively use the power-spectrum of the simulated signals to create a random signals with the

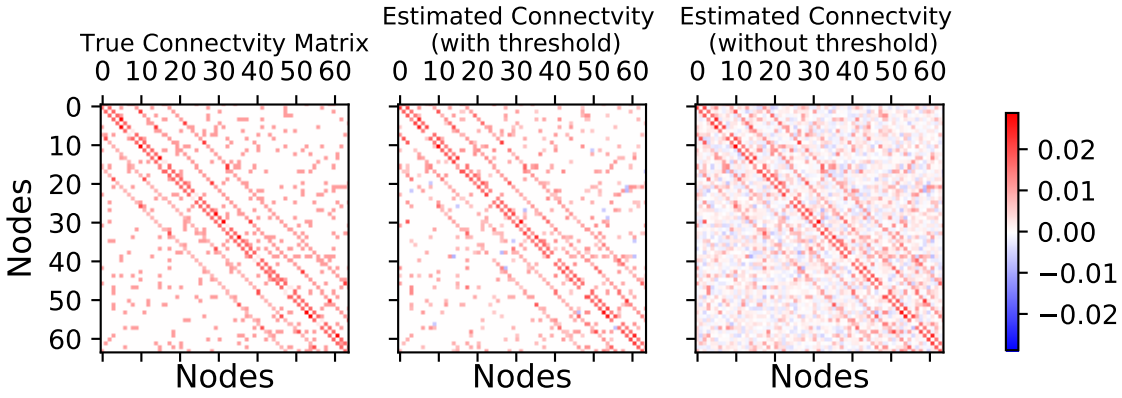

Figure 2: Example for the estimation of the connectivity from the data. The left panel shows the true connectivity matrix of the 64 populations (the connection strengths are color coded and have been scaled for better visibility). The middle and the right panel show the estimated connectivity of the populations for AR process of order 2 fitted to the data. In the middle panel only values bigger than 0.0054 are shown, in the right panel there is no threshold applied. The area under the ROC curve (AUC) is 0.99 and the precision recall score (PRS) is 0.97.

Table 1: **Parameters used for the simulation of ECoG data**

Variable name , Symbol	**Value**
Number of Electrodes, N	64
Simulation length, T	39 997ms
Bin size inhomogeneous input	5.0 ms
Number of inter Electrode connections	3835320
Number of neurons per population	1000
Percentage of excitatory connections	80%
Rate of poissonian input, pois rate	90.75 Hz
Neuron model	iaf_psc_alpha
Synaptic time constant excitatory, $\tau_{\text{syn}}^{\text{ex}}$	2.0 ms
Synaptic time constant inhibitory, $\tau_{\text{syn}}^{\text{in}}$	2.0 ms
Membrane time constant τ_{m}	20.0 ms
Spike threshold, V_{th}	−55.0 mV
Connection propability within populations p	10%
Inter-electrode connection weight	10 mV
Connection weight excitatory, w_{ex}	10 mV
Reset potential, V_{reset}	−70 mV
Synaptic delay	1.5 ms
Inter-electrode synaptic delay	3 ms
Ratio inhibition to excitation, g	5.0

same power-spectrum. These signals are then used as rate profiles for the time-varying inputs to the neurons in the next iteration. Each neuron gets a private realization of a Poisson process with varying intensity as input. The intensity is piecewise constant, changing every 5 ms. This is repeated until the power-spectrum of the signals converge towards the power-spectrum of the iteration before, which means self-consistency for the power-spectra.

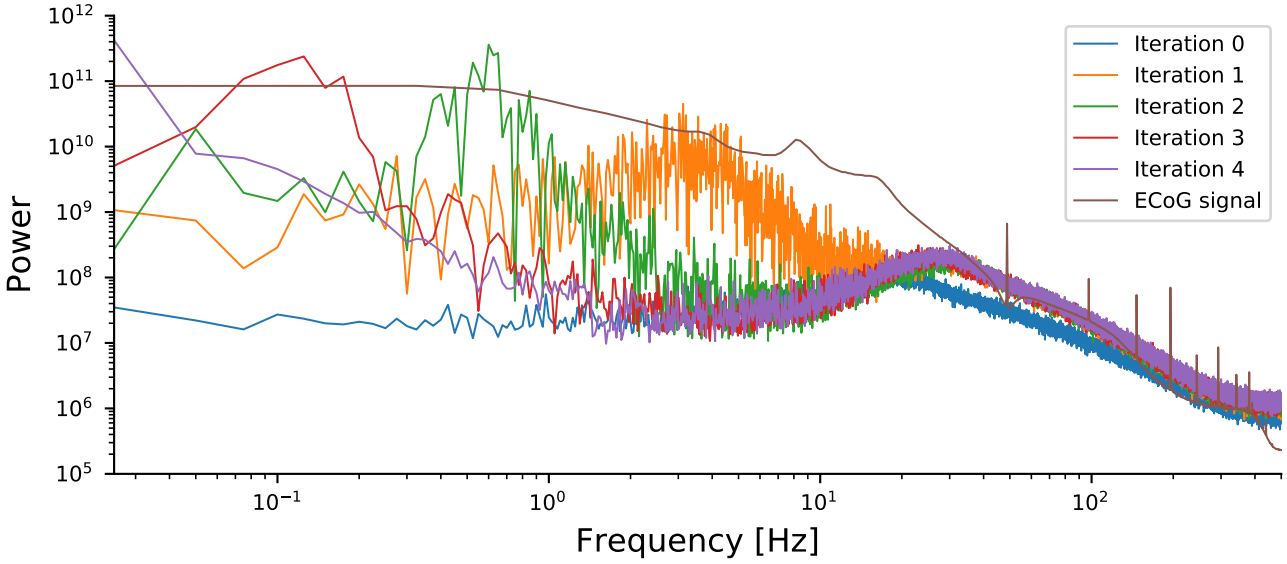

Figure 3: Powerspectra of measured (brown) and simulated ECoG signals for 5 iterations. The powerspectra have been calculated based on the simulated signals divided by a factor of 1000.

The presence of inhomogeneous inputs leads to higher firing rates of the neurons. To compensate for this effect, we reduce the homogeneous Poissonian input each neuron receives. This approach can be interpreted as a model for input of ongoing activity from brain regions which are not recorded to the neural populations.

Estimating connectivity

Usually, the temporal resolution of ECoG is in the same order of magnitude as the timescales of neuronal population activity. Other then the methods in [5] and [6], this allows us to use methods for the estimation of connection which rely on the temporal information in the data. We use a simple linear regression method described in detail in [4]. The main idea is, as is in Granger causality, to fit an autoregessive process to the data. Here, an L^2 approach is used to fit the AR process to the data, Assuming the population signals $x(t)$ can be described linearly by an AR process, we have

$$x(t) = v(t) + \sum_{i=0}^{p} G(i)x(t-i), \tag{1}$$

where $v(t) \sim N(0, \sigma)$ is the driving noise of the process, $G(i)$ are the matrices of interaction kernels and p is the order of the process. With an easy calculation which can be found in [4], one can show that

$$\hat{\boldsymbol{G}} = \boldsymbol{x}\boldsymbol{Z}^*(\boldsymbol{Z}^*\boldsymbol{Z})^{-1}$$

is the least square estimator for $\boldsymbol{G} = (G(1), \ldots, G(p))$ with $\boldsymbol{Z} = (Z_0, \ldots, Z_{t-1})$ and $Z_t = \begin{pmatrix} x(t) \\ x(t-1) \\ \vdots \\ x(t-p+1) \end{pmatrix}$.

This product with the pseudo inverse of $\boldsymbol{Z}$ is easy to compute, as long as the networks are not too big (up to 100 nodes is easily feasible on a normal desktop machine).

Pseudo-metric on graphs

For analyzing the resulting networks of the method described above, we now introduce a pseudo-metric on graphs. The idea is similar to the method presented in [13], but instead of using the Laplacian spectrum of the networks, we use the eigenvalue spectrum of the adjacency matrix of the networks. With this it is possible to study directed graphs instead of only undirected graphs as with the Laplacian.

For a given network with n nodes, we can calculate the eigenvalues of the corresponding adjacency matrix. These form a discrete set of points, denoted by $\mathcal{S}$. We can write these sets as Dirac-delta functions

$$E(x) = \sum_{k=1}^{n} \delta(x - x_k)$$

with $\mathcal{S} = \{x_1, \ldots, x_n\}$. The convolution with a Gaussian kernel g,

$$(E * g)(x) = \int E(x - y)g(y)dy = \sum_k g(x - x_k),$$

is an element in L^2, so we can define a composition

$$\begin{aligned} \mathbb{C} &\xrightarrow{\Gamma_g} \Delta \xrightarrow{*} L^2 \\ \mathcal{S} &\longmapsto E \longmapsto E * g \end{aligned}$$

where Δ denotes the set of delta-functions.

For two given sets of eigenvalues $\mathcal{S}$ and $\mathcal{S}'$ of graphs with n and m nodes, respectively, possibly $n \neq m$, with the corresponding sums of delta-functions E and E' and a fixed kernel $g \in L^2$ with $g > 0$ we can now define an inner product

$$\langle E, E' \rangle_g = \langle E * g, E' * g \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of L^2,

$$\langle f, h \rangle = \int f(t)h(t)\mathrm{d}t,$$

with $f,\ h \in L^2$. Thus, we can calculate the inner product of two sets of eigenvalues, $\mathcal{S} = \{ x_1, \ldots, x_n\}$, $\mathcal{S}' = \{ x'_1, \ldots, x'_m\}$ with corresponding sums of delta-functions

$$E(x) = \frac{1}{n} \sum_{k=1}^{n} \delta(x - x_k) \text{ and } E'(x) = \frac{1}{m} \sum_{j=1}^{m} \delta(x - x_j)$$

via

$$\begin{aligned} \langle E, E' \rangle_g &= \langle E * g, E' * g \rangle \\ &= \langle \sum_{k=1}^{n} g(x - x_k), \sum_{j=1}^{m} g(x - x'_j) \rangle \\ &= \int \sum_{k=1}^{n} g(x - x_k) \sum_{j=1}^{m} g(x - x'_j)\mathrm{d}x \\ &= \int \sum_{j,k} g(x - x_k)g(x - x'_j)\mathrm{d}x \\ &= \sum_{j,k} \int g(x - x_k)g(x - x'_j)\mathrm{d}x \end{aligned}$$

where $\int g(x-x_k)g(x-x'_j)\mathrm{d}x$ is the autocorrelation of the kernel g evaluated at $x_k - x'_j$. The inner product defined above can be used to define a distance between two spectra of graphs:

$$d_g(E, E') = \sqrt{\langle E - E', E - E' \rangle_g}.$$

Consider $\widehat{\mathcal{S}} = span\ \Gamma_g(\mathcal{S}) \subset \Delta$. It is clear that the delta-functions

$$f(x') \ = \ \delta(x - x')$$

for all $x' \in \mathbb{C}$ form a basis of $\widehat{\mathcal{S}}$. So, for $x', x'' \in \mathbb{C}$ the inner product $\langle \cdot, \cdot \rangle_g$ is uniquely given by

$$\langle f(x'), f(x'') \rangle_g = \langle \delta(x - x'), \delta(x - x'') \rangle_g. \tag{2}$$

The reason why this methods leads to a pseudo-metric and not a metric is, that there are isospectral graphs, which are non-isomorphic graphs with the same spectrum. Of course, isospectral graphs with corresponding delta-functions E and E' have distance zero: $d_g(E, E') = 0$ for all $g \in L^2$. This means that d_g is not a metric on the space of graphs. However, the other direction holds: if $d_g(E, E') = 0$ for all $g > 0$, then the corresponding graphs to E and E' are isospectral. This, and the fact that for a growing number of vertices the number of isospectral graphs is decreasing [21], makes this method a promising candidate as an approximation of a metric on large graphs.

Results

Simulated ECoG data without inhomogeneous input

The raster plot of the spiketrains shown in Fig 4 A indicates that the neurons fire asynchronous irregularly, the average firing-rate is 11.25 Hz. Fig 4 C shows the distributions of the firing rates, coefficient of variation (CV) of individual spike trains (the mean is 0.59) and correlation coefficients (CC) for firing of pairs of neurons. The power-spectrum of the simulated signals, averaged over all channels, is shown as iteration 0 in Fig 3.

Simulated ECoG data with inhomogeneous input

The presence of inhomogeneous input to the neurons changes the characteristics of the firing of the neurons. The firing rate with inhomogeneous input, being 18.07 Hz, is higher than in the case without inhomogeneous input. Fig 4 B shows a raster plot for the same neurons as in Fig 4 A, it still indicates asynchronous irregular firing of the neurons. As shown in Fig 4 D, the firing of neurons is more regular than without inhomogeneous input, the mean CV decreased to 0.51. Also, correlation coefficients for the firing of pairs of neurons increased (to a mean value of 0.03 calculated for bins of 20 ms).

As repeatedly reported in the literature [11, 12, 22], the powerspectra of measured LFP data are decreasing with $1/f$ between 10 Hz and 100 Hz as shown in Fig 3. The simulated data also shows a decrease, independent of the inhomogeneous input. However, in the presence of inhomogeneous input, the decrease is amplified. Additionally, when inhomogeneous input is present the power in low frequencies is increased and closer to measured ECoG data. As shown in Fig 3 the difference in the inhomogeneous inputs lead to different power for frequencies under 10 Hz whereas for frequencies higher than 10 Hz the power-spectra are very similar.

Estimating connectivity

We used the estimation method described in the materials section to estimate the connectivity of the neural populations for both data types, without and with inhomogeneous inputs. The simulation length was in total $T = 39.997$ s. Often, ECoG recordings are recorded during trial based activities. The task the subjects are asked to perform is repeated several times leading to data sets with multiple

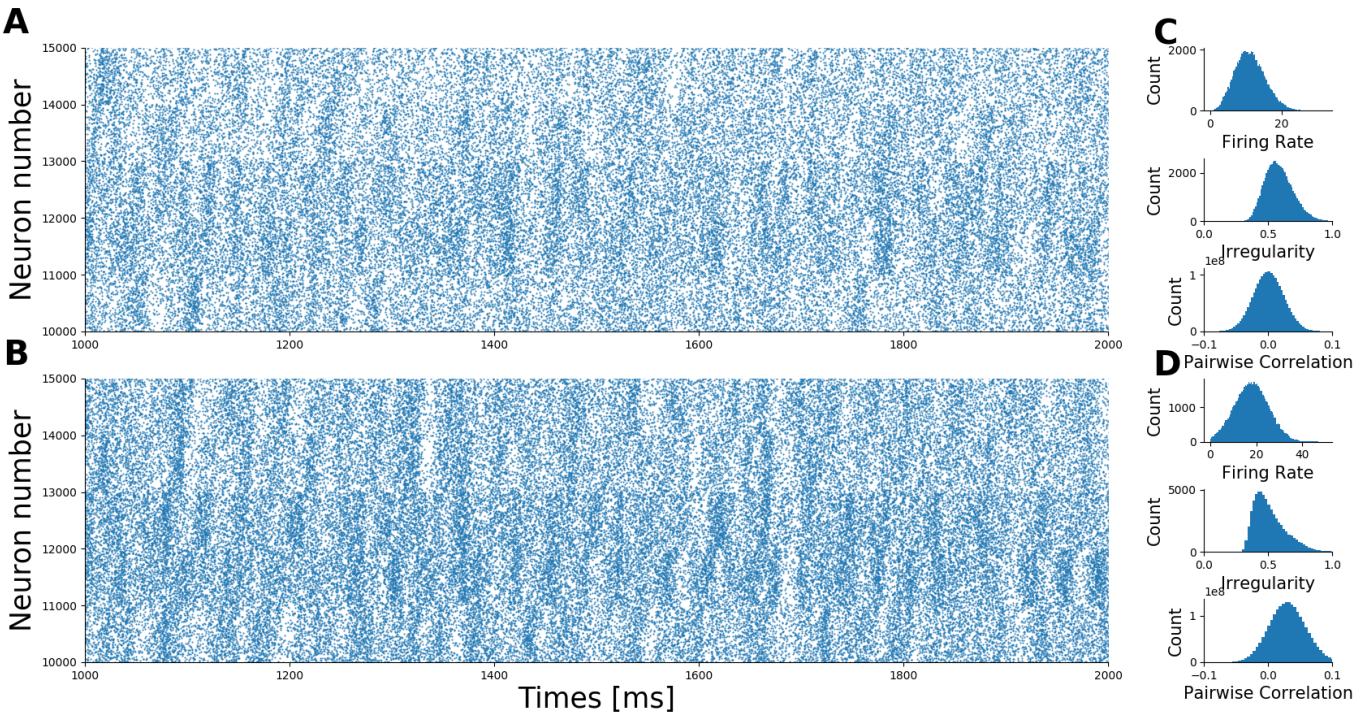

Figure 4: **A** Rasterplot of five populations (5 000 neurons) without inhomogeneous input. On average, the neurons fire at 11.25 Hz and have a mean CV of 0.59. **B** Rasterplot for the same neurons with inhomogeneous input. The mean firing rate is 18.07 Hz and the mean CV 0.51. **C** and **D** show the distribution of firing rates, coefficients of variation and pairwise correlation coeffiecients for the spiking data with and without inhomogeneous input, respectively.

trials each around 3 or 4 s. To ensure the applicability of the estimation to such data sets, we divided the data into 39 chunks with a length of 1 000 ms each. The estimation was then performed on each of the chunks separately. The resulting AR coefficients were averaged over all trials. An example of the estimation is shown in Fig 2, where the AR coefficients were averaged over all orders to end up with one resulting estimated connectivity matrix.

In general, the estimation works close to perfect, the very high scores for Area Under the ROC Curve (AUC) and Precision Recall Score (PRS) of 0.99 and 0.97, respectively, where 1.0 is the maximum achieved by a perfect estimation. The values of our estimation indicate that there is a threshold for which the estimated connectivity and the true connectivity match almost perfectly. A key factor for the estimation of the networks is which order is used, cf. variable p in Eq (1). In both cases, the order were the estimation works best is 2. The results for the estimation studying the effect of the order of the AR process fitted to the data are shown in Fig 5 A and B, respectively.

Additional to AUC and PRS we also calculated the distance of the estimated connectivity matrices to the true connectivity matrices using the method introduced in the methods section. The orange curves (which corresponds to the right y-axis) in Fig 5 show the spectral distances of the networks, depending on the order used for the estimation, the red curves show the Euclidean distances, respectively. For calculating the distances, the true network and the estimated networks were normalized by their maximal value before calculating the distance to achieve comparable values for the distances. As expected, both sorts of distances show lower values for networks with higher AUCs and PRS and higher values for estimated networks with lower values for AUC and PRS. Also, their slope is qualitatively the same. However, the values for the distance have an offset of about 6.

As we simulate non-linear neurons and derive a population activity by summing up post-synaptic

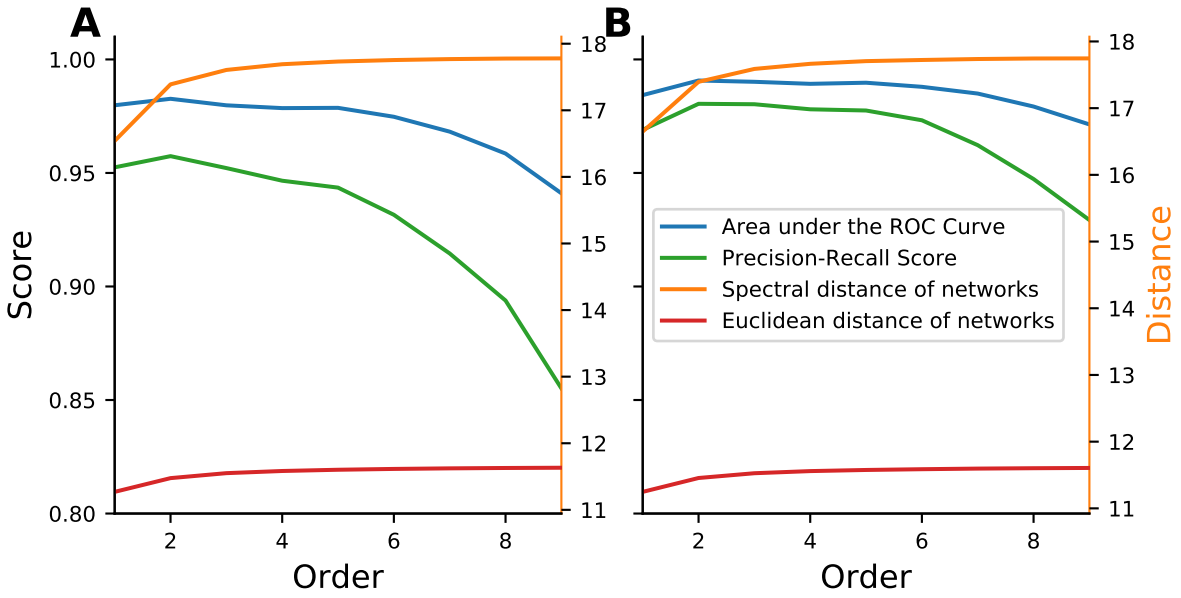

Figure 5: **A** Performance of network estimation on simulated data without inhomogeneous input. The blue curve shows the Area Under the ROC curve, the green curve shows the Precision-Recall score. The orange curve shows the spectral distance ($\sigma = 0.1$) between the estimated network and the true network and the red curve shows the Euclidean distance between the estimated network and the true network. For the distance the right y-axis applies. **B** Same as A but for data with inhomogeneous input.

currents, the question arises what kind of linear activity is fitted to the simulated data by the least-square estimator. To study the kernels, we used the same estimations as above, with the only difference that we don't average over the order. So, as above, the kernels were estimated for 1 000 ms data chunks and then averaged over all 39 chunks. The estimated kernels are shown in Fig 6. Independent of the order used for the estimation, all of them show high values for order $p = 1$ and a rapid decrease for the following orders. Even though the kernels estimated for higher orders increase again to a local peak for order $p = 5$, this indicates that the dynamics of neural populations act on very fast time scales. In fact, the estimated kernels indicate that the interactions of neuronal populations occur almost instantaneous. Qualitatively, the kernels of the data with inhomogeneous input does not differ from the data without inhomogeneous input.

For calculating the spectral distance, we use a Gaussian kernel with width $\sigma = 0.1$ as kernel g in Eq (2). Of course, this value influences the resulting distance and needs to be chosen adequately for the network size.

Minimal data length necessary for connectivity estimation

A key factor for the accuracy of the estimation is the amount of data available. In the setting described above, this has two components: the amount of trials recorded and the length of each trial. Also, if one wants to study the changes of connectivity over time, one might use sliding windows with each windows being only a fraction of the trial. Each estimation method has its own characteristics regarding what kind of data it is usable for. Methods which rely on the temporal information (like the one described above) need a sufficient length for each data chunk to be able to achieve a reasonable estimation for each chunk before estimated over all chunks. Other methods which are based on the covariance of the dynamics and do not rely on the temporal information itself, like [7, 8] or [5] , can perform better on very short data chunks but might not achieve the same performance as temporal based methods for longer pieces of data.

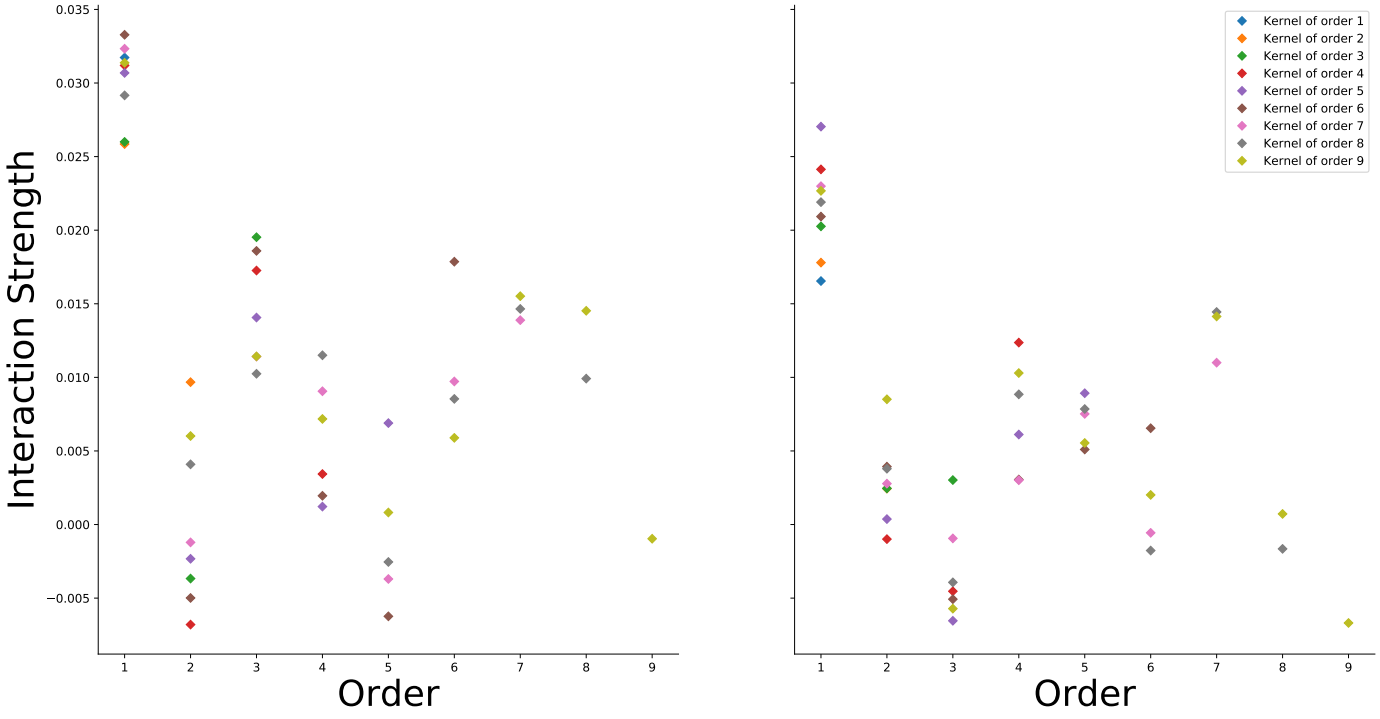

Figure 6: Estimated linear interaction kernels. The left panel shows the kernels for the simulated data without inhomogeneous input, the right panel the kernels for the data with inhomogeneous input. The same estimations as for Fig 5 were used.

We studied the estimation for different lengths of data chunks available, the results are shown in Fig 7. The amount of data available in total was kept constant (in total 39 997 ms), but we varied the length of the data chunks. Depending on the length of the chunks, the amount of trials available varied ($^{39\,997}/_{\text{length of chunks}} = \#\text{trials}$). For the least-square estimator of the AR process, we estimated the AR kernels for each data chunk separately and then averaged over all trials. For calculating the performance of the estimation we used a fit of order $p = 2$ and averaged then the two kernel estimates. We compared this estimation to the estimation method introduced in [5]. This method is a covariance-based approach using a L^1-minimization. We calculated the instant covariance matrix for each trial and then averaged them. The resulting averaged covariance matrix was then used for the connectivity estimation.

The results shown in Fig 7 indicate that the covariance approach is superior for very short data chunks whereas for longer data chunks the least-square estimator achieves better results. In fact, the AUC and PRS of the L^1-based method are 0.95 and 0.86, respectively, for 333 data chunks of 75 ms each. The least-square estimation method needs data chunks of 300 ms (resulting in 133 trials) to achieve the same performance. On the other hand, the least-square method achieves AUC and PRS of 0.99 and 0.97 for data chunks of 1 000 ms, whereas the maximum performance of the L^1-based method are 0.97 and 0.91 for AUC and PRS, respectively. Overall, the presence of inhomogeneous input does not affect either of the methods.

Discussion

In this study, we introduced a detailed simulation for ECoG signals. We used this simulation for validating the results of different connectivity estimation procedures. Also, we extended an existing

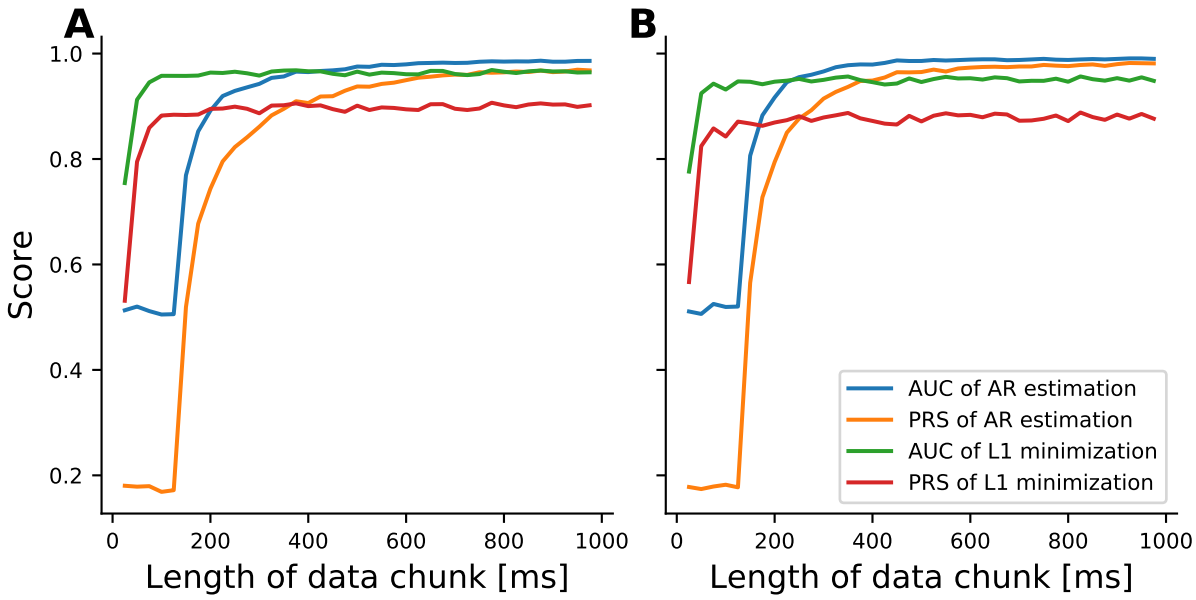

Figure 7: Performance of the estimation methods depending on the length of data chunks used (total amount of data was kept constant). **A** shows the performances for the simulated ECoG data without inhomogeneous inputs and **B** the performances for data with inhomogeneous inputs. The blue and orange curves represent the AUC and PRS for the least-square estimation of AR coefficients, the green and the red curves represent the AUC and PRS for the L^1 minimization based on the covariance matrix described in [5].

tool for the comparison of networks to an universal tool for the quantification of differences in estimated networks.

The simulation of ECoG signals are based on leaky-integrate-and-fire neurons, which is a realistic neuron model reflecting fundamental non-linear features of biological neurons and still is computationally feasible on this scale. Both, the L^1 and the L^2 based estimation methods, show reasonable results on the simulated ECoG signals, cf. Fig 7. Also, the spectral distance of networks reflects the same characteristics (Fig 5) of differences of networks as the area under the ROC curve, precision recall score and Euclidean distance which are only applicable for networks of the same size.

The estimation methods assume linear interactions among the neural populations. Although they were applied on signals which are based on non-linear interactions of neurons, the resulting networks almost don't differ from the true networks. The homogeneity of external inputs has no effect on the estimation procedures. All this indicates that it is reasonable to use estimation methods which are based on linear interactions for estimating connectivity on neural population scale. Of course, the estimation method chosen should be adapted on the data available (total amount of data, amount of trials). A limitation of our model is, that it does not include volume conduction, which can play a role in the generation of ECoG signals and local field potentials in general [11]. This can possibly affect the estimation of connectivity through superposition of signals. However, the order of magnitude of volume conduction depends fundamentally on the design of the electrode. Also, the leaky-integrate-and-fire model does not account for the spatial extension of biological neurons. As the bi-polar nature of neurons and the spatial alignment of the neurons are key factors of local field potentials, this is another limitation of our model. A third limitation of our simulation is, that for real ECoG measurements the electrodes only conducts signals from parts of the brain which are embedded in a much bigger network whereas our simulation observes all simulated neurons. However, we compensated for that by using inhomogeneous inputs to the neurons. Also, as shown in [5] the covariance based estimation method is robust against unobserved nodes .

Although extensively used in previous studies, this is, to the authors knowledge, the first attempt

to validate estimation procedures for ECoG data or local field potentials in general based on the simulation of single neurons. Also, we adapted an existing approach to compare estimated networks [13] to the more general case where directed networks and even networks of different sizes are comparable.

Conclusion

In this study we showed, that estimating effective connectivity by fitting an autoregressive process to the data is a suitable approach for neural populations. We showed that depending on the data available and the research question asked, different methods for the estimation of connectivity should be considered.

Acknowledgments

Supported by the DFG (grant EXC 1086). The HPC facilities are funded by the state of Baden-Württemberg through bwHPC and DFG grant INST 39/963-1 FUGG. We thank Tonio Ball and Markus Kern for sharing the measured ECoG data with us. Also, we thank Uwe Grauer from the Bernstein Center Freiburg as well as Bernd Wiebelt and Michael Janczyk from the Freiburg University Computing Center for their assistance with HPC issues.

References

[1] Karl J. Friston. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36, 2011.

[2] Baojuan Li, Adeel Razi, and Karl J. Friston. Editorial: Mapping Psychopathology with fMRI and Effective Connectivity Analysis. *Frontiers in Human Neuroscience*, 11, 2017.

[3] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[4] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg, 2005.

[5] Jonathan Schiefer, Alexander Niederbühl, Volker Pernice, Carolin Lennartz, Jürgen Hennig, Pierre LeVan, and Stefan Rotter. From correlation to causation: Estimating effective connectivity from zero-lag covariances of brain signals. *PLoS Computational Biology*, 14(3):e1006056, 2018.

[6] Carolin Lennartz, Jonathan Schiefer, Stefan Rotter, Jürgen Hennig, and Pierre LeVan. Sparse Estimation of Resting-State Effective Connectivity from fMRI Cross-Spectra. *Frontiers in Neuroscience*, 2018.

[7] Anna Korzeniewska, Ciprian M. Crainiceanu, Rafał Kuś, Piotr J. Franaszczuk, and Nathan E. Crone. Dynamics of event-related causality in brain electrical activity. *Human Brain Mapping*, 29(10):1170–1192, 2008.

[8] Mingzhou Ding, Steven L. Bressler, Weiming Yang, and Hualou Liang. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological Cybernetics*, 83(1):35–45, 2000.

[9] Anna Korzeniewska, Piotr J. Franaszczuk, Ciprian M. Crainiceanu, Rafał Kuś, and Nathan E. Crone. Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG). *NeuroImage*, 56(4):2218–2237, 2011.

[10] Nikos K Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of neuroscience*, 23(10):3963–3971, 2003.

[11] György Buzsáki, Costas A Anastassiou, and Christof Koch. The origin of extracellular fields and currents - EEG , ECoG , LFP and spikes. *Nature Reviews Neuroscience*, 13(June):407–420, 2012.

[12] Gaute T Einevoll, Christoph Kayser, Nikos K Logothetis, and Stefano Panzeri. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature reviews. Neuroscience*, 14(11):770–85, 2013.

[13] Siemon C de Lange, Marcel a de Reus, and Martijn P van den Heuvel. The Laplacian spectrum of neural networks. *Frontiers in computational neuroscience*, 7(189), 2014.

[14] M. S. Fifer, G. W. Milsap, E. Greenwald, D. P. McMullen, W. S. Anderson, N. V. Thakor, N. E. Crone, and R. Vinjamuri. Design and implementation of a human ECoG simulator for testing brain-machine interfaces. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1311–1314, 2014.

[15] Nicolas Brunel. Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. *Journal of Computational Neuroscience*, 208:183–208, 2000.

[16] Hannah Bos, Abigail Morrison, Alexander Peyser, Jan Hahne, Moritz Helias, Susanne Kunkel, Tammo Ippen, Jochen Martin Eppler, Maximilian Schmidt, Alex Seeholzer, Mikael Djurfeldt, Sandra Diaz, Janne Morén, Rajalekshmi Deepu, Teo Stocco, Moritz Deger, Frank Michler, and Hans Ekkehard Plesser. Nest 2.10.0, December 2015.

[17] Guido Rossum. Python reference manual. Technical report, Amsterdam, The Netherlands, The Netherlands, 1995.

[18] Elodie Fino and Rafael Yuste. Dense Inhibitory Connectivity in Neocortex. *Neuron*, 69(6):1188–1203, March 2011.

[19] Alberto Mazzoni, Stefano Panzeri, Nikos K. Logothetis, and Nicolas Brunel. Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS Computational Biology*, 4(12), 2008.

[20] Benjamin Dummer, Stefan Wieland, and Benjamin Lindner. Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity. *Frontiers in Computational Neuroscience*, 8(September):1–12, 2014.

[21] Edwin R. van Dam and Willem H. Haemers. Which graphs are determined by their spectrum? *Linear Algebra and its Applications*, 373(0):241 – 272, 2003.

[22] Claude Bédard and Alain Destexhe. Macroscopic Models of Local Field Potentials and the Apparent 1/f Noise in Brain Activity. *Biophysical Journal*, 96(April):2589–2603, 2009.

5 Discussion

5.1 Optimization on unitary manifolds

The method proposed in this thesis uses the exponential map for matrices to project the gradient back on the unitary manifold. To recap: the gradient of the map

$$\begin{aligned} \Gamma \colon \mathcal{U}(n) &\longrightarrow \mathbb{R} \\ U &\longmapsto \|UB\|_1 \end{aligned} \tag{5.1}$$

with respect to U is given by

$$d_{ij} = \frac{\partial \Gamma(U)}{\partial U_{ij}}.$$

Then, for the matrix

$$a = \frac{1}{2}\left(d - d^*\right)$$

it follows

$$a^* = \frac{1}{2}\left(d - d^*\right)^* = \frac{1}{2}\left(d^* - d\right) = -a \tag{5.2}$$

which means that a is skew-hermitian. With this it follows that

$$\exp(a)\exp(a^*) = \exp(a)\exp(-a) = \exp(a-a) = \exp(0) = \mathbb{1}$$

which means that $\exp(a)$ is unitary. In other words, the space of skew-hermitian matrices forms the Lie-Algebra $\mathfrak{u}(n)$ corresponding to the Lie-group $\mathcal{U}(n)$. This is the fundamental principle of the update scheme used in the presented algorithm.

Now the question arises, why with the procedure described above, the update step leads to a matrix with smaller L^1-norm $\Gamma\left(U_s\right) \geqslant \Gamma\left(U_{s+1}\right)$ for any step s and why it makes sense to take the skew-hermitian part of the derivative d. I briefly recap the underlying principles, for details

please see Absil et al. (2008). In the following, let $\mathcal{M}$ be a manifold and $f\colon \mathcal{M} \longrightarrow \mathbb{R}$ a function defined on $\mathcal{M}$. The set of smooth real-valued functions defined on a neighborhood of $x \in \mathcal{M}$ is denoted by $\mathfrak{F}_x(\mathcal{M})$. A smooth map $\gamma\colon \mathbb{R} \longrightarrow \mathcal{M}$ is called a *curve in* $\mathcal{M}$.

The *directional derivative of* f *at* x is defined as

$$\mathrm{D}f(x)[\eta] = \lim_{t\to 0} \frac{f(x+t\eta) - f(x)}{t}. \tag{5.3}$$

With Γ defined in (5.1) and $f = \Gamma$, this corresponds to

$$\mathrm{D}\Gamma(x)[\eta] = \lim_{t\to 0} \frac{\Gamma(x+t\eta) - \Gamma(x)}{t}.$$

Alternatively, for a real-valued function defined on a neighborhood of x, i.e. $f \in \mathfrak{F}_x(\mathcal{M})$, the *tangent vector of the curve* γ *at* $t = 0$, is defined as a mapping

$$\begin{aligned} \dot{\gamma}(0) : \mathfrak{F}_x(\mathcal{M}) &\longrightarrow \mathbb{R} \\ f &\longmapsto \frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\Big|_{t=0} \end{aligned}$$

for γ a curve through x at $t = 0$. Here it is exploited that $f \circ \gamma\colon \mathbb{R} \longrightarrow \mathbb{R}$ is a smooth function with a well-defined derivative. In the setting of the unitary manifold

$$\mathcal{U}(n) = \{U \in \mathbb{C}^{n\times n}\colon U^*U = \mathbb{1}\},$$

let $U(t) = x + t\eta$ be a curve in $\mathcal{U}(n)$ through $x = \mathbb{1}$. From

$$U^*(t)\,U(t) = \mathbb{1}_n \tag{5.4}$$

for all t, by differentiating Eq (5.4) we get

$$\dot{U}^*(t)\,U(t) + U^*(t)\,\dot{U}(t) = 0 \tag{5.5}$$

where

$$\begin{aligned} \dot{U}(0)\colon \mathfrak{F}_{\mathbb{1}}(\mathcal{U}(n)) &\longrightarrow \mathbb{R} \\ \Gamma &\longmapsto \frac{\mathrm{d}(\Gamma(\mathrm{U}(\mathrm{t})))}{\mathrm{d}t}\Big|_{t=0} \end{aligned}$$

Next, we define a *tangent vector* ξ_x to a manifold $\mathcal{M}$ at point x as a map

$$\xi_x\colon \mathfrak{F}_x(\mathcal{M}) \longrightarrow \mathbb{R}$$

such that there exists a curve γ on $\mathcal{M}$ with $\gamma(0) = 0$ satisfying

$$\xi_x f = \dot{\gamma}(0) f \coloneqq \left.\frac{\mathrm{d}(f(\gamma(t)))}{\mathrm{d}t}\right|_{t=0}$$

for all $f \in \mathfrak{F}_x(\mathcal{M})$. Using this definition, it can be shown that the *tangent space* $T_x\mathcal{M}$ which is the set of all tangent vectors, is a vector space. In the unitary case, this is clear since we have

$$\dot{U}(0)\Gamma = \left.\frac{\mathrm{d}(\Gamma(\mathrm{U}(\mathrm{t})))}{\mathrm{d}t}\right|_{t=0}.$$

Defining a derivative of a curve $\gamma(t)$ as

$$\gamma'(t) \coloneqq \lim_{\tau \to 0} \frac{\gamma(t+\tau) - \gamma(t)}{\tau}$$

requires a vector space structure to compute $\gamma(t+\tau) - \gamma(t)$. Thus, this is not possible on abstract manifolds. However, given a vector space $\mathcal{E}$, it has the structure of a linear manifold in a natural way. For a tangent vector ξ_x to $\mathcal{E}$ at x we get

$$\xi_x f = \mathrm{D}f(x)[\gamma'(0)]$$

for $f \in \mathfrak{F}_x(\mathcal{M})$. The derivative γ' does not depend on the curve that realizes ξ_x. Thus $\xi_x \mapsto \gamma'(0)$ defines a linear map which identifies

$$T_x\mathcal{E} \simeq \mathcal{E}. \tag{5.6}$$

Due to the group structure of $\mathcal{U}(n)$ we can exploit Eq (5.6) which gives us $T_x\mathcal{U}(n) \simeq \mathcal{U}(n)$. From Eq (5.5) it follows that $\dot{U}(0) \in \{Z \in \mathbb{C}^{n\times n} \colon U_0^* Z + Z^* U_0 = 0\}$. Thus, for the tangent space it holds

$$T_{U_0}\mathcal{U}(n) \subset \{Z \in \mathbb{C}^{n\times n} \colon U_0^* Z + Z^* U_0 = 0\}.$$

Also, it holds

$$0 = \dot{U}^*(0) U(0) + U^*(0) \dot{U}(0) = \dot{U}^*(0) \mathbb{1} + \mathbb{1}\dot{U}(0)$$

and thus it follows

$$\dot{U}^*(0) = -\dot{U}(0)$$

which means that the matrix $\dot{U}(0)$ is skew-hermitian.

Now, with tangent vectors as defined above, we have the notion of directional derivatives for manifolds. The goal is to find the motion from a point U on the unitary manifold for minimizing

Γ. This means, we want to find the motion where Γ has the steepest descent. For doing so, we need to add a notion of length for tangent vectors. This is the case for *Riemannian manifolds*, which are real, smooth manifolds for which the tangent space at each point is equipped with an inner product and, with this, an induced metric. Since we are focusing on the unitary manifold, we can use the inner product of $\mathbb{C}$ together with the identification $\mathbb{C} \simeq \mathbb{R}^2$. With this, we can define the *gradient* of f at x, denoted by grad $f(x)$, as an element of $T_x\mathbb{C}^{n\times n}$ that satisfies

$$\langle \mathrm{grad} f(x), \xi \rangle_x = \mathrm{D}f(x)[\xi] \qquad \forall \xi$$

which is unique. In the unitary case, this gives us

$$\langle \mathrm{grad}\Gamma(U), \xi \rangle_U = \mathrm{D}\Gamma(U)[\xi] \qquad \forall \xi$$

We then get the result that the direction of the gradient grad $\Gamma(U)$ is the steepest-ascent direction of Γ at U:

$$\frac{\mathrm{grad}\ \Gamma(U)}{\| \mathrm{grad}\ \Gamma(U) \|} = \underset{\xi \in T_U\mathcal{U}(n):\ \|\xi\|=1}{\arg\max} \mathrm{D}\Gamma(U)[\xi].$$

The steepest-ascent property follows from the steepest-ascent properties of the derivative in $\mathbb{C}^{n\times n}$.

In practice, this means we can calculate the gradient of $\Gamma(U) = \| UB \|_1$ by calculating

$$d = \frac{\partial \Gamma(U)}{\partial U} = \begin{pmatrix} \frac{\partial \Gamma(U)}{\partial U_{11}} & \cdots & \frac{\partial \Gamma(U)}{\partial U_{1n}} \\ \vdots & \vdots & \vdots \\ \frac{\partial \Gamma(U)}{\partial U_{n1}} & \cdots & \frac{\partial \Gamma(U)}{\partial U_{nn}} \end{pmatrix}$$

and take the skew-hermitian part $a = \frac{1}{2}(d - d^*)$. This then gives us then the element in the tangent space $T_U\mathcal{U}(n)$ which can be projected back on the manifold by taking the matrix exponential $U_{s+1} = \exp(-\lambda a_s) U_s$ for a step size $\lambda \in \mathbb{R}$. With this, we get $\Gamma(U_{s+1}) \leqslant \Gamma(U_s)$ for all s.

Alternative update scheme

As shown in Wen and Yin (2013), there are other possibilities to project the gradient matrix a back on the orthogonal manifold. All of the results shown in this paper hold for the unitary case, which will be considered here. They show that for any skew-hermitian matrix a, the matrix

$$q = (\mathbb{1} + a)^{-1} (\mathbb{1} - a)$$

is well-defined and unitary. This is easy to see, using

$$(\mathbb{1} + M)(\mathbb{1} - M) = \mathbb{1} - M + M - M^2 = (\mathbb{1} - M)(\mathbb{1} + M)$$

for any matrix M. It follows

$$q^* q = \left((\mathbb{1} + a)^{-1}(\mathbb{1} - a)\right)^* \left((\mathbb{1} + a)^{-1}(\mathbb{1} - a)\right) = \mathbb{1}$$

which means q is unitary. For a given point X on the unitary manifold and a skew-hermitian matrix a the so-called Cayley transform given by

$$Y(\tau) = \left(\mathbb{1} + \frac{\tau}{2}a\right)^{-1} \left(\mathbb{1} - \frac{\tau}{2}a\right) X \tag{5.7}$$

gives a feasible update formula. Due to the computationally easier calculation, an algorithm based on the Cayley transform update scheme reaches a minimum faster than with an exponential update scheme. However, even after pronounced investigation of the update scheme and an appropriate step-size selection, it turns out to not be as accurate as the exponential case. For reasons I have not been able to work out, the resulting estimated connectivity matrices have a strong tendency to be symmetric. This corresponds to not being able to estimate the direction of the connections properly. On the other hand, the fast computability of the update step for different step sizes, this update formula offers the possibility to study the stucture of the underlying manifold in the direction of the gradient. An example of this is shown in Fig 5.1 for a collider structure with different connection strengths. The gradients are calculated at the initial point of the optimization and then the cost is evaluated based on Eq (5.7) for different step sizes. Notably, for connections that are too strong, the global minimum is in the direction of negative step sizes from the starting point. In these cases, the optimization does not estimate the correct network.

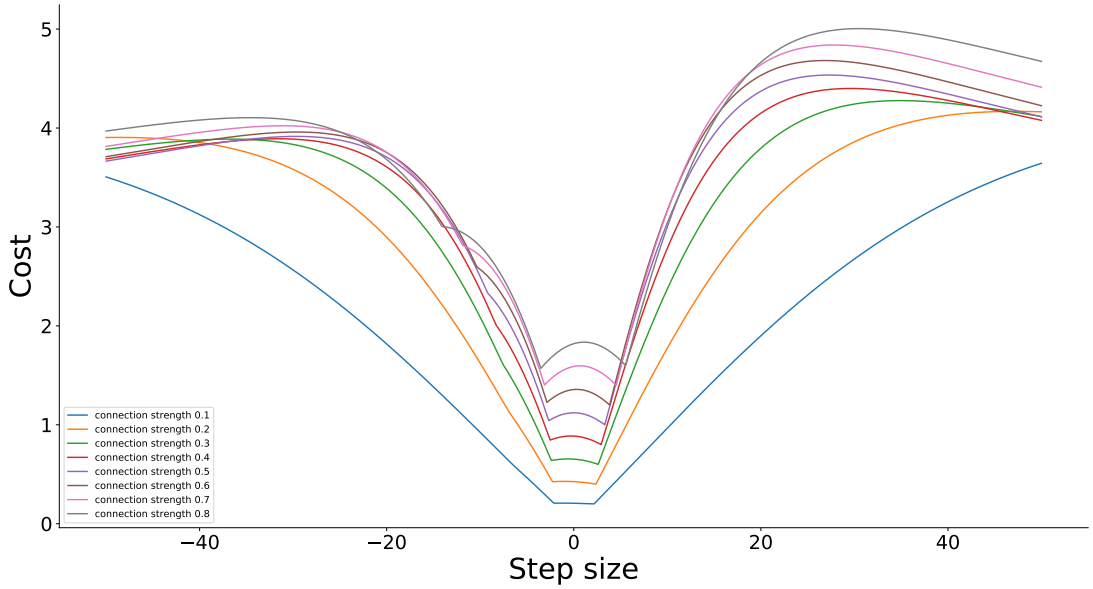

Figure 5.1: Cost landscapes from starting point of the optimization for varying step sizes. Each curve corresponds with a collider structure, where the connection strengths of the two connection vary.

5.2 L^1-minimization as basis for estimating connectivity

The L^1-minimization used for the connectivity estimation is a key factor for the intrinsic properties of the methods described in chapter 2. A major constraint induced by using L^1 as norm is the limitation of the method to sparse networks. One could ask, why not use the L^2 norm instead, which corresponds with minimizing the function

$$\begin{aligned} \Gamma_2 \colon \mathcal{U}(n) &\longrightarrow \mathbb{R} \\ U &\longmapsto \|UB\|_2 \end{aligned}$$

to overcome the sparsity constraint. For any unitary matrix $U \in \mathcal{U}(n)$ and $B \in \mathbb{C}^{n\times n}$ we get

$$\|UB\|_2^2 = \langle UB, UB\rangle = \langle B, U^*UB\rangle = \langle B, B\rangle = \|B\|_2^2$$

for $\langle \cdot, \cdot \rangle$ denoting the inner product. This means that multiplication with a unitary matrix preserves the L^2 of any matrix and is therefore not feasible for our purpose.

Alternatively, one can use a variant of the Lasso (Tibshirani, 1996) for solving the regularization problem

$$||\hat{C}^{-1} - B^*B||_F + \lambda \sum_{i,j} |B_{ij}|$$

with respect to B for a regularization parameter λ. It produces some coefficients that are exactly 0 due to the nature of this constraint, which means it also does not overcome the sparsity constraint. Nonetheless, it has the advantage of being computationally easier to solve than the optimization on the unitary manifold used in the presented estimation method. The minimization on the matrix manifold makes a calculation of the matrix exponential necessary in each step, which is computationally expensive. For the regularization problem, this is not necessary, which makes it feasible to even larger networks and faster convergence. However, the accuracy does not reach the level of the method proposed in this thesis. This, together with the fact, that most data acquisition methods produce networks in a feasible range of the proposed method, and the computation time being in range of seconds, speaks in favor of the proposed method.

5.3 Linear interactions as model for population activity

As explained in chapter 1, the underlying interaction model is a key ingredient of the concept of effective connectivity. The estimation method presented in this thesis relies on a linear interaction model, given by Eq (1.1). The validation and evaluation of intrinsic properties of this estimation method presented in chapters 2 and 3 are based on noise-free covariance matrices and dynamical processes derived from this linear interaction model. They show that for nodes following this linear interaction model, the estimation of connectivity is possible very accurately for large, sparsely coupled networks with reasonable network interaction, also in the presence of noise and latent nodes.

In contrast to the linear processes used in chapters 2 and 3, the processes used for validating the connectivity estimation in chapter 4 are based on leaky-integrate-and-fire neurons. The signals are then derived from the currents flowing through synapses terminating at excitatory post-synaptic targets. Due to the non-linear generation of action potentials in the leaky-integrate-and-fire neurons, these signals are also non-linear. Although both methods studied in chapter 4

are based on a linear interaction model, the estimation of effective connectivity from these signals shows a very good performance. This indicates that the assumption of linearity for modeling population activity, at least for the purpose of estimating connectivity, seems to be reasonable.

5.4 Comparability to other methods

For decades, methods for the estimation of connectivity have been a popular research subject (Granger, 1969; Melssen and Epping, 1987; Baccalá and Sameshima, 2001; Friston et al., 2003; Kiebel et al., 2008; Korzeniewska et al., 2008; Wedeen et al., 2008; Friston, 2009, 2011; Korzeniewska et al., 2011; Pernice and Rotter, 2013; Reisert et al., 2013; Friston et al., 2014; Zaghlool and Wyatt, 2014; Gilson et al., 2016; Ryali et al., 2016; Nie et al., 2017). However, the method presented in this thesis has several advantages over previously existing methods. First, the method is feasible if solely the zero-lag covariances of the data are available, which makes it independent of the sampling rate of the recording. This is a big advantage in case of BOLD related signals. The estimation method presented in Gilson et al. (2016) uses the same approach of estimating effective connectivity from measured covariances. This method does not use any sparsity constraint, but solves an optimization problem based on the Lyapunov equation. However, this method relies on time-shifted covariances matrices to estimate the direction of connections, where access to the correct shift parameter is crucial for the estimation. For fMRI data, this parameter needs to be accessed experimentally. Due to limited models for fMRI data, this also complicates the validation of the method. Thus, the independence of time-shifted covariances a is feature which should not be underestimated.

Secondly, the method is applicable on various types of data. In case of fMRI data, one usually has a single recording of a measurement lasting up to 10 minutes. The repetition time (TR) determines the temporal resolution of the measurement and through this, the total amount of samples. In the case of the MREG sequence, a dataset can be up to 12 000 samples per node. Also, for fMRI signals one needs to take the hemodynamic response function into consideration. The method presented here is feasible for such kind of data, as shown in chapter 2 and chapter 3. In case of ECoG signals, the datasets usually consist of short (up to 4 000ms) but many (possibly thousands) trials with a sampling rate around 1 kHz. Often, such data is recorded with a movement task for the subject. For such kind of data, the changes in connectivity

within the trials are of special interest. Consequently, one needs to estimate connectivity based on short data chunks. Assuming stationarity across trials, one can use every trial for estimating the covariance or connectivity matrix and average over all of these estimates. As shown in chapter 4 the method presented in this thesis is superior to Granger Causality (Granger, 1969), as presented in Lütkepohl (2005), in such a setting.

Thirdly, quite a lot of the existing methods, like the Regularized Inverse Covariance method studied in Smith et al. (2011), are not able to estimate the direction of connections but are limited to undirected networks. This is, due to the nature of the connections in the brain being formed by directed synapses, a clear limitation.

Finally, a major limitation of DCM (Friston et al., 2003) is the limitation on small networks (≤ 20 nodes). This limits applications of DCM to subnetworks of the human brain. Additionally, as DCM is based on Bayesian estimator, one needs to have access to a prior network and then estimate the probability distribution of the connection parameters, which limits the applications of the method to even more specific cases.

5.5 Conclusion

In this thesis a method for estimating effective connectivity from signals of neural populations is presented. The validations on simulations in chapters 2, 3 and 4 show that the estimation method works reliably for a broad range of networks. These simulations also show, that due to the nature of the estimation method it can be applied on a broad range of data: using zero-lag covariances, cross-spectral densities or averaged covariances from multiple trials allow a very high flexibility. Additionally, the simulations in chapter 4 show that the estimation method can be applied on data which does not follow linear interaction models.

The application of the estimation method on fMRI data, which is a prominent example of data not following linear interaction models, leads to reasonable results. This is supported by comparing the results to established estimation methods as the Regularized Inverse Covariance or the study of the Default Mode Network. Also, the analysis of DTI data in chapter 3 shows better agreement than other methods do for functional connectivity.

Of course, the lack of a ground truth of connectivity on neural population levels limits the possibility of validating these results. Even matching results of different methods only show coherence and are no proof that the estimated networks are the true networks in the brain. DTI and other structural methods themselves have problems of accuracy or applicability. This is also expressed in the lack of knowledge of connection probabilities on population scale. Connection probability, or more precisely, sparsity of the network, is the only fundamental constraint of the presented estimation method. Due to this, the resulting networks should only be interpreted as the strongest connections in the network, and not as the complete network. This reduces the possibility of declaring a connection as not existing to "no strong connection existing".

The link of changes in connectivity and development of function and dysfunction of the brain together with the possibility of accessing the connectivity indicates a enormous potential of such methods. Of course, further investigation has to be carried out before applying it in a clinical context. This could be done by investigating data which is already available, such as fMRI data for different stages of Alzheimer's disease, or on, for example, data of stroke patients where usually more than one fMRI measurement per subject is available. Also, other non-invasive data sources, such as fNRIS, could be studied in detail. The potential of classification of neural diseases based on connectivity estimated from non-invasive measurements is very promising and should undergo further scientific investigation.

The method for estimating connectivity of neural populations based on data of their activity presented in this thesis, overcomes several shortcomings of previously existing methods. The method can be applied on a broad range of data types and is feasible for almost all sparse networks which are biologically realistic. A corresponding implementation of the algorithm is easy to apply and available as free software.

Bibliography

Abrudan, T., Eriksson, J., and Koivunen, V. (2009). Conjugate Gradient Algorithm for Optimization Under Unitary Matrix Constraint. *Signal Processing*, 89:1704–1714.

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press.

Achab, M., Bacry, E., Gaïffas, S., Mastromatteo, I., and Muzy, J.-F. (2017). Uncovering causality from multivariate Hawkes integrated cumulants. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1–10. PMLR.

Aertsen, A. and Preissl, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. In *Nonlinear Dynamics and Neuronal Networks*, pages 281–301.

Aertsen, A. M., Gerstein, G. L., Habib, M. K., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *Journal of Neurophysiology*, 61(5):900–917.

Assländer, J., Zahneisen, B., Hugger, T., Reisert, M., Lee, H.-L., LeVan, P., and Hennig, J. (2013). Single shot whole brain imaging using spherical stack of spirals trajectories. *NeuroImage*, 73:59 – 70.

Baccalá, L. A. and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474.

Bacry, E. and Muzy, J.-F. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166.

Bacry, E. and Muzy, J. F. (2016). First- and Second-Order Statistics Characterization of Hawkes Processes and Non-Parametric Estimation. *IEEE Transactions on Information Theory*, 62(4):2184–2202.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG , ECoG , LFP and spikes. *Nature Reviews. Neuroscience*, 13(June):407–420.

Coombes, S. (2010). Large-scale neural dynamics: Simple and complex. *NeuroImage*, 52(3):731–739.

Davie, C. A. (2008). A review of Parkinson's disease. *British Medical Bulletin*, 86(1):109–127.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature reviews. Neuroscience*, 14(11):770–85.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., and Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex (New York, NY)*, 26(8):3508–3526.

Franzmeier, N., Düzel, E., Jessen, F., Buerger, K., Levin, J., Duering, M., Dichgans, M., Haass, C., Suárez-Calvet, M., Fagan, A. M., Paumier, K., Benzinger, T., Masters, C. L., Morris, J. C., Perneczky, R., Janowitz, D., Catak, C., Wolfsgruber, S., Wagner, M., Teipel, S., Kilimann, I., Ramirez, A., Rossor, M., Jucker, M., Chhatwal, J., Spottke, A., Boecker, H., Brosseron, F., Falkai, P., Fliessbach, K., Heneka, M. T., Laske, C., Nestor, P., Peters, O., Fuentes, M., Menne, F., Priller, J., Spruth, E. J., Franke, C., Schneider, A., Kofler, B., Westerteicher, C., Speck, O., Wiltfang, J., Bartels, C., Caballero, A., Ángel, M., Metzger, C., Bittner, D., Weiner, M., Lee, J.-H., Salloway, S., Danek, A., Goate, A., Schofield, P. R., Bateman, R. J., and Ewers, M. (2018). Left frontal hub connectivity delays cognitive impairment in autosomal-dominant and sporadic Alzheimer's disease. *Brain*.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-Related fMRI: Characterizing Differential Responses. *NeuroImage*, 7(1):30 – 40.

Friston, K. J. (2009). Causal Modelling and Brain Connectivity in Functional Magnetic Resonance Imaging. *PLoS Biology*, 7(2):e1000033.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.

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

Friston, K. J., Kahan, J., Biswal, B., and Razi, A. (2014). A DCM for resting state fMRI. *NeuroImage*, 94:396–407.

Gilson, M., Moreno-bote, R., Ponce-alvarez, A., and Ritter, P. (2016). Estimation of Directed Effective Connectivity from fMRI Functional Connectivity Hints at Asymmetries of Cortical Connectome. *PLoS Computational Biology*, pages 1–30.

Glover, G. H. (1999). Deconvolution of Impulse Response in Event-Related BOLD fMRI. *NeuroImage*, 9(4):416 – 429.

Goldberg, E. M. and Coulter, D. A. (2013). Mechanisms of epileptogenesis: a convergence on neural circuit dysfunction. *Nature reviews. Neuroscience*, 14(5):337–349.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438.

Hawkes, A. G. (1971a). Point Spectra of Some Mutually Exciting Point Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(3):438–443.

Hawkes, A. G. (1971b). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58(1):83–90.

Jacobs, H. I. L., Van Boxtel, M. P. J., Heinecke, A., Gronenschild, E. H. B. M., Backes, W. H., Ramakers, I. H. G. B., Jolles, J., and Verhey, F. R. J. (2012). Functional integration of parietal lobe activity in early Alzheimer disease. *Neurology*, 78(5):352–360.

Kiebel, S. J., Garrido, M. I., Moran, R. J., and Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics*, 2(2):121–136.

Korzeniewska, A., Crainiceanu, C. M., Kuś, R., Franaszczuk, P. J., and Crone, N. E. (2008). Dynamics of event-related causality in brain electrical activity. *Human Brain Mapping*, 29(10):1170–1192.

Korzeniewska, A., Franaszczuk, P. J., Crainiceanu, C. M., Kuś, R., and Crone, N. E. (2011). Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG). *NeuroImage*, 56(4):2218–2237.

Kriener, B., Tetzlaff, T., Aertsen, A., Diesmann, M., and Rotter, S. (2008). Correlations and Population Dynamics in Cortical Networks. *Neural Computation*, 20(9):2185–2226.

Lamme, V. A. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11):571–579.

Lennartz, C., Schiefer, J., Rotter, S., Hennig, J., and LeVan, P. (2018). Sparse estimation of resting-state effective connectivity from fmri cross-spectra. *Frontiers in Neuroscience*, 12:287.

Li, B., Razi, A., and Friston, K. J. (2017). Editorial: Mapping psychopathology with fmri and effective connectivity analysis. *Frontiers in Human Neuroscience*, 11:151.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971.

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

Mazzoni, A., Lindén, H., Cuntz, H., Lansner, A., Panzeri, S., and Einevoll, G. T. (2015). Computing the local field potential (lfp) from integrate-and-fire network models. *PLoS Computational Biology*, 11(12):1–38.

Melssen, W. J. and Epping, W. J. M. (1987). Detection and estimation of neural connectivity based on crosscorrelation analysis. *Biological Cybernetics*, 57(6):403–414.

Nie, L., Yang, X., Matthews, P. M., Xu, Z. W., and Guo, Y. K. (2017). Inferring functional connectivity in fMRI using minimum partial correlation. *International Journal of Automation and Computing*, 14:371–285.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA.

Pernice, V. and Rotter, S. (2013). Reconstruction of sparse connectivity in neural networks from spike train covariances. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03008.

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):e1002059.

Potjans, T. C. and Diesmann, M. (2014). The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. *Cerebral Cortex (New York, NY)*, 24(3):785–806.

Rebane, G. and Pearl, J. (1987). The recovery of causal poly-trees from statistical data. *Proceedings of the Third Workshop on Uncertainty in AI*, pages 222–228.

Reisert, M., Mader, I., Umarova, R., Maier, S., Tebartz van Elst, L., and Kiselev, V. G. (2013). Fiber density estimation from single q-shell diffusion imaging by tensor divergence. *NeuroImage*, 77:166–176.

Ryali, S., Chen, T., Supekar, K., Tu, T., Kochalka, J., Cai, W., and Menon, V. (2016). Multivariate dynamical systems-based estimation of causal brain interactions in fMRI: Group-level validation using benchmark data, neurophysiological models and human connectome project data. *Journal of Neuroscience Methods*, 268:142–153.

Schiefer, J., Niederbühl, A., Pernice, V., Lennartz, C., Hennig, J., LeVan, P., and Rotter, S. (2018a). From correlation to causation: Estimating effective connectivity from zero-lag covariances of brain signals. *PLoS Computational Biology*, 14(3):e1006056.

Schiefer, J., Schäfer, L. M., and Rotter, S. (2018b). Estimating connectivity from simulated ECoG signals. *unpublished*.

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

Sporns, O., Tononi, G., and Kötter, R. (2005). The Human Connectome: A Structural Description of the Human Brain. *PLoS Computational Biology*, 1(4):e42.

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., and Diesmann, M. (2008). Dependence of Neuronal Correlations on Filter Characteristics and Marginal Spike Train Statistics. *Neural Computation*, 20(9):2133–2184.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tiesinga, P., Bakker, R., Hill, S., and Bjaalie, J. G. (2015). Feeding the human brain model. *Current Opinion in Neurobiology*, 32:107–114.

Trousdale, J., Hu, Y., Shea-Brown, E., and Josić, K. (2012). Impact of Network Structure and Cellular Response on Spike Time Correlations. *PLoS Computational Biology*, 8(3):e1002408.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1):273–289.

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., and de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *NeuroImage*, 41(4):1267–1277.

Welvaert, M. and Rosseel, Y. (2014). A review of fMRI simulation studies. *PLoS ONE*, 9(7).

Wen, Z. and Yin, W. (2013). A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1-2):397–434.

Wenk, G. L. (2003). Neuropathologic Changes in Alzheimer's Disease. *The Journal of Clinical Psychiatry*, 64(suppl 9):7–10.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Zaghlool, S. B. and Wyatt, C. L. (2014). Missing data estimation in fMRI dynamic causal modeling. *Frontiers in Neuroscience*, 8:191.