A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies

Stefan Haufe · Arne Ewald

Abstract Due to its high temporal resolution, electroencephalography (EEG) is widely used to study functional and effective brain connectivity. Yet, there is currently a mismatch between the vastness of studies conducted and the degree to which the employed analyses are theoretically understood and empirically validated. We here provide a simulation framework that enables researchers to test their analysis pipelines on realistic pseudo-EEG data. We construct a minimal example of brain interaction, which we propose as a benchmark for assessing a methodology's general eligibility for EEG-based connectivity estimation. We envision that this benchmark be extended in a collaborative effort to validate methods in more complex scenarios. Quantitative metrics are defined to assess a method's performance in terms of source localization, connectivity detection and directionality estimation. All data and code needed for generating pseudo-EEG data, conducting source reconstruction and connectivity estimation using baseline methods from the literature, evaluating performance metrics, as well as plotting results, are made publicly available. While this article covers only EEG modeling, we will also provide a magnetoencephalography (MEG) version of our framework online.

1 Introduction

Electromagnetic source imaging, that is the estimation of the activation time courses and locations of the neuronal populations (*sources*) contributing to an electro- or magnetoencephalographic (EEG/MEG) recording, is a challenging inverse problem. Lacking a unique solution, it can only

S. Haufe
Laboratory for Intelligent Imaging and Neural Computing, Columbia University, New York, NY, 10027, USA
Machine Learning Department, Technische Universität Berlin, 10587 Berlin, Germany
E-mail: stefan.haufe@tu-berlin.de

A. Ewald
Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany
E-mail: mail@aewald.net

be solved approximately using prior assumptions on the source activity. To date, a large variety of inverse solutions exist, accounting for the fact that different experimental settings may require different characterizations of the sources (Baillet et al., 2001a). Distributed solutions estimate the activity of the entire brain at once. Early approaches used spatial smoothness constraints to obtain a unique solution (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994). Other researchers introduced sparsity penalties to obtain spatially focal sources (Haufe et al., 2008, 2009; Ding and He, 2008; Owen et al., 2012). The advent of powerful computing solutions also enabled the simultaneous localization of entire time series, the spatio-temporal dynamics of which have been modeled using various combinations of penalty terms (e.g., Ou et al., 2009; Haufe et al., 2011; Gramfort et al., 2013b; Castaño Candamil et al., 2015). Another type of inverse procedures is based only on temporal assumptions, while the source activity is estimated separately for each brain location(Van Veen et al., 1997; Mosher and Leahy, 1999; Gross et al., 2001).

A related problem is the EEG/MEG-based analysis of brain connectivity, which is concerned with the estimation of interactions *between* sources based on their reconstructed time courses. Just as in EEG/MEG source imaging, a wealth of methods exist, which are based on different models of brain interaction. Among the most popular methods are approaches that define interaction either based on the cross-spectrum (Nunez et al., 1997; Nolte et al., 2004, 2008; Marzetti et al., 2008; Ewald et al., 2012; Chella et al., 2014; Shahbazi et al., 2015), Granger causality (Kamiński and Blinowska, 1991; Astolfi et al., 2006; Schelter et al., 2009; Wibral et al., 2011; Valdes-Sosa et al., 2011; Haufe et al., 2012a; Barnett and Seth, 2014), dynamic causal modeling (Kiebel et al., 2006, 2008; Stephan et al., 2007), and other non-linear relationships such as correlations of the phases or amplitudes of brain rhythms (Brookes et al., 2011, 2012; Siegel et al., 2012; Hipp et al., 2012). There is a host of literature employing such methods to research questions in basic and clinical neuroscience (e.g., Supp et al., 2007; Sasaki et al., 2013; van Mierlo et al., 2014; Barttfeld et al., 2014). Brain connectivity estimation however builds on the correct localization and time series reconstruction in the first stage. It is therefore more challenging than inverse source reconstruction alone and needs careful validation.

Evaluating the quality of inverse solutions and connectivity estimates is not straightforward, as an objective ground truth is typically not available. Inverse solutions have been compared to spatial maps obtained from functional magnetic resonance imaging (fMRI) (Lantz et al., 2001; Grova et al., 2008). Since the exact relationship between EEG/MEG activity and the BOLD signal measured by fMRI is, however, unclear, this approach provides only weak evidence for correct localization. Intracranial electrophysiological recordings as well as surgical outcome have been proposed as a possible means of validation in patients undergoing surgery (e.g. Benar et al., 2006; Vulliemoz et al., 2010; Zwoliński et al., 2010). Such invasive procedures are however limited by the fact that patterns of electrical current flow are strongly affected by tissue inhomogeneities such as lesions and skull holes, which are typically neglected when building the electrical volume conductor model in which source localization is carried out. Phantoms studies (Leahy et al., 1998; Baillet et al., 2001b) may offer a more controlled way of validating inverse methods and the underlying forward (volume conductor) model, but are technically challenging.

As the forward problem has a unique solution that can be well approximated using numerical methods (Sarvas, 1987; Vorwerk et al., 2014; Haufe et al., 2015; Huang et al., 2015), it is convenient to study the effect of inverse modeling in isolation for a fixed volume conductor model. This can be done without a phantom using numerical simulations. Here, the forward model is summarized as a 'lead field' matrix, which is applied to the simulated source time series to yield pseudo-EEG data. Simulations allow one to conveniently and objectively test methods under different signal-to-noise ratios (SNR), as well as for different spatial and dynamical characteristics of the sources, and are therefore widely used for benchmarking purposes in the source modeling literature (e.g., Darvas et al., 2004; Haufe, 2011; Gramfort et al., 2013b).

In the same way as EEG/MEG based brain connectivity analyses are more challenging than inverse source reconstruction alone, their validation is also more challenging. On one hand, validation in real data is less straightforward, as experimental settings, for which the location of the underlying sources as well as their connectivity is known, are more rare compared to the case when only the source locations need to be known. On the other hand, even when the 'ground truth' is available, it is difficult to define appropriate performance measures, as inevitable localization errors can render the subsequent assessment of source connectivity between the true source locations inappropriate. This problem also applies to simulations.

Presumably due to these difficulties, it has become common practice that authors conduct and report brain connectivity analyses of real data without or with only limited empirical validation of the employed methodology. In some studies, the applied connectivity measure is simply presented as the 'definition' of interaction. Of those simulations provided in the literature, many do not apply to the EEG/MEG case due to insufficient modeling or complete disregard of the linear source mixing caused by volume conduction in the head, as well as disregard of additive noise (e.g., Korzeniewska et al., 2003; Astolfi et al., 2006, 2007; Velez-Perez et al., 2008; Barrett et al., 2012; Silfverhuth et al., 2012; Wibral et al., 2013; Sameshima et al., 2015). Other simulation studies have started to explicitly include EEG forward models (e.g., Schoffelen and Gross, 2009; Haufe et al., 2010, 2012a; Ewald and Nolte, 2013; Rodrigues and Andrade, 2015; Cho et al., 2015).

Source crosstalk and correlated noise as caused by volume conduction can lead to large false detection rates for a number of established connectivity metrics, among them seemingly intuitive measures such as Granger Causality (see Nolte et al., 2004, 2008; Haufe et al., 2012b,a; Vinck et al., 2015; Winkler et al., 2015). These results underline the importance of testing methods under realistic simulated conditions in order to avoid unjustified conclusions on real data. Validating source connectivity methodologies is also a prerequisite for justifying subsequent analyses such as the study of network properties of the estimated connectivity graphs (De Vico Fallani et al., 2007; Rubinov and Sporns, 2010; Mullen et al., 2011; Barttfeld et al., 2014; Bola and Sabel, 2015), which crucially depend on the correctness of the estimated connections.

Considering the vast and steadily growing body of literature on EEG/MEG based brain connectivity analysis, we believe that the field would benefit from a standardized benchmark. We here present a first effort towards this, focusing on EEG. We present a simulation framework for generating realistic pseudo-EEG data from underlying interacting brain sources. Based on this framework, we develop a benchmark for validating the performance of inverse source reconstruction and brain

connectivity analyses. We define three simple criteria to quantify a methodology's success in terms of source localization, the detection of the presence of connectivity, and directionality estimation. Through being defined on the coarse level of eight regions of interest (ROI) identical to the brain octants, these measures are by construction fairly robust with respect to slight mislocalizations of the sources.

While the proposed benchmark is limited to what we consider a minimal realistic case of brain interaction, it can be extended to model more complex cases. Researchers employing EEG-based connectivity analyses may create versions of the benchmark that are specifically tailored to assessing the validity of their methodologies. To facilitate such efforts, all data and code needed for generating pseudo-EEG data, conducting source reconstruction and connectivity estimation using baseline methods from the literature, evaluating performance metrics, as well as plotting results, are made publicly available in Matlab format. Moreover, unlabeled data generated under the proposed minimal model of interaction are made available online. Researchers are invited to analyze these data as part of a data analysis challenge, which will be announced at http://bbci.de/supplementary/EEGconnectivity/.

2 Objectives

Our goal is to provide a simulation framework and benchmark with the following properties.

Realism

Simulated pseudo-EEG data should be sufficiently realistic to provide an undistorted view of the performance of EEG-based connectivity analyses. To this end, the following features are implemented.

- The use of a realistic finite element volume conductor model including six tissue types (Haufe et al., 2015; Huang et al., 2015), integrating recently formulated guidelines for volume conductor modeling of the human head (Vorwerk et al., 2014).
- The presence of interacting sources exerting time-delayed influence on another.
- Interactions being confined to a narrow frequency band.
- Realistic source locations being confined to the cortical manifold and emitting electrical currents perpendicular to the local surface.
- Variable locations, spatial extents and depths of the sources.
- The presence of independent background brain processes with 1/f (pink noise) spectra.
- The presence of measurement noise.
- Broad yet realistic SNR ranges.
- The availability of separate baseline measurements not containing interacting sources.

Standardization

The framework makes use of established standards. As many of the major software packages for EEG data analysis are based on Matlab (The Mathworks, Natick, MA) (Delorme and Makeig, 2004;

Acar and Makeig, 2010; Delorme et al., 2011; Tadel et al., 2011; Oostenveld et al., 2011), Matlab is adopted as the simulation platform. The generated data can be analyzed within existing packages, or using stand-alone code. Our standardization efforts moreover include

- The use of a highly detailed anatomical template of the average adult human head (Fonov et al., 2009, 2011; Haufe et al., 2015; Huang et al., 2015).
- The use of to the Montreal Neurological Institute (MNI) coordinate system (Evans et al., 1993; Mazziotta et al., 1995).
- The use of the 10/5 electrode placement system (Oostenveld and Praamstra, 2001).

Availability

To facilitate easy adoption, all involved data and codes are made publicly available. This includes lead fields, surface meshes and electrode coordinates, as well as codes for generating simulated data, evaluating performance metrics, and conducting source reconstructions and connectivity estimations using baseline methods from the literature. We also provide routines for plotting EEG scalp potentials as well as source distributions.

Minimality

To provide an initial assessment of any method, we propose a benchmark that focuses on a minimal realistic case of brain interaction. For this benchmark we restrict ourselves to

- The presence of only two interacting sources.
- Linear interaction.
- Uni-directional information flow.
- Spatially non-overlapping sources being confined to different brain octants.

Unability of a method to deal with this case will raise doubt on its eligibility for EEG-based connectivity estimation in general.

Quantitative evaluation

All evaluations are carried out on the coarse level of eight regions of interest (ROIs) that are identical to the octants of the brain. Quantitative measures for assessing source localization error, the detection of brain connectivity, and the estimation of interaction directionality are provided.

Extendability

The framework can serve as a basis to model and benchmark more complex types of brain interaction in order to enable researchers to test their preferred methods on appropriately designed data. A number of potential extensions are outlined in Section 6.1.

3 A realistic head model for simulating EEG data

The basis of the proposed work is are several Matlab script for simulating the activity of neuronal populations in specific brain locations, and mapping that activity to EEG sensors using a realistic model of electrical current propagation in the head.

3.1 Forward model of EEG data

The so-called forward model of EEG data describes how neural activity in the brain maps to the EEG sensors. Specifically, it describes the flow of the extra-cellular ionic return currents emerging in response to the intra-cellular neuronal activity. The forward model comprises information about the geometries of the various tissue compartments (gray matter, white matter, cerebrospinal fluid, skull, skin) in the studied head, as well as the conductive properties of these tissues. Under the assumption that the quasi-static approximation of Maxwell's equations holds for the frequencies typically studied in EEG, it is linear (Sarvas, 1987; Baillet et al., 2001a). In its discretized form it reads

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\varepsilon}(t) \ , \tag{1}$$

where the time-dependent $3R$-dimensional vector $\mathbf{j}(t)$ represents the directed primary currents at R distinct locations on the cortical surface, the $M \times 3R$ *lead field* matrix $\mathbf{L}$ describes the relationship between primary currents and the observable scalp potentials at M sensors, and $\boldsymbol{\varepsilon}(t)$ is an M-dimensional noise vector.

Inverse source reconstruction is concerned with the estimation of the source primary currents $\mathbf{j}(t)$ given the measurements $\mathbf{x}(t)$ in a given head model $\mathbf{L}$, while functional or effective brain connectivity estimation is concerned with the estimation of the information flow between brain sites from either $\mathbf{j}(t)$ or (less commonly) $\mathbf{x}(t)$. In order to simulate physically realistic EEG data for the purpose of benchmarking, source time series $\mathbf{j}(t)$ mimicking the behavior of macroscopic neuronal populations are generated based on prior assumptions on the dynamic properties (e. g. coupling structure) of the populations. These source time series, which are localized on the cortical surface, are then mapped to the EEG sensors by means of the lead field $\mathbf{L}$. Finally, a correlated noise vector $\boldsymbol{\varepsilon}(t)$ comprising measurement noise as well as artifactual (such as line noise or muscular) and background brain activity is added.

3.2 Head model

We here use lead fields that were precomputed in the so-called New York Head (Haufe et al., 2015; Huang et al., 2015). The New York Head model combines a highly detailed magnetic resonance (MR) image of the average adult human head with state-of-the-art finite element electrical modeling. It is based on the ICBM152 anatomy provided by the International Consortium for Brain Mapping (ICBM), which is a nonlinear average of the T1-weighted structural MRI of 152 adults. A detailed

segmentation of the New York Head into six tissue types (scalp, skull, CSF, gray matter, white matter, air cavities) was performed at the native MRI resolution of $0.5\,\mathrm{mm}^3$. Based on this segmentation, a finite element model (FEM) was solved to generate the lead field. The model was evaluated for 231 electrode positions including 165 positions defined in the 10-5 electrode placement system (Oostenveld and Praamstra, 2001), and for 75 000 nodes of a mesh of the cortical surface. The cortical surface was extracted using the BrainVISA Morphologist toolbox (Rivière et al., 2003; Geffroy et al., 2011). For conducting the simulations and evaluations proposed in this work, subsets of 108 electrodes and 2 000 cortical locations were selected for practical reasons. The lead field was referenced to the common average of the 108 selected electrodes. For plotting purposes in this paper, results are mapped onto a smoothed version (Tadel et al., 2011) of the full cortical mesh. Figure 1 depicts the New York Head surface with 108 electrodes mounted, as well as the original and smoothed cortical surface.
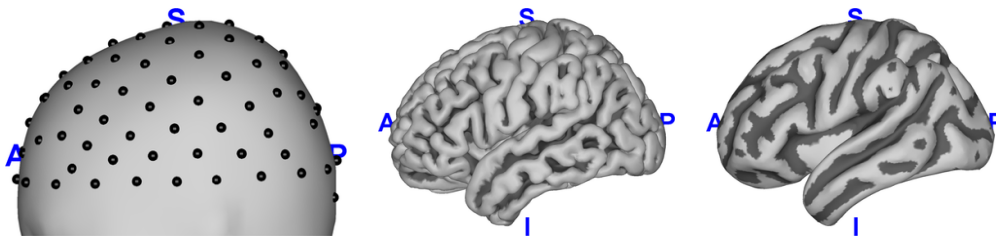

Fig. 1: The New York Head anatomy Left: head surface with 108 electrodes placed. Center: high-resolution cortical surface. Right: smoothed cortical surface used for plotting. Cortical sulci are marked in dark color.

3.3 Regions of interest

Eight regions of interest (ROIs) identical to the octants of the brain are defined. To ensure that octants cover brain areas of roughly similar size, the origin of the coordinate system was shifted based on cutting the cortical mesh into two halves each containing an equal number of nodes. The cutting planes obtained this way are defined by the equations $x = 0\,\mathrm{mm}$ (separating left and right hemispheres), $y = $ -18.7 mm (separating anterior and posterior hemispheres) and $z = 12.8\,\mathrm{mm}$ (separating superior and inferior hemispheres), where all coordinates given in MNI space. The combination of the three hyperplanes defines eight octants as shown in Figure 2.

4 A minimal connectivity benchmark

In the following, we provide the specifics of a benchmark intended to provide a first assessment of EEG-based brain connectivity estimation pipelines in a minimal setting involving brain interaction.

|	Octant	MNI coordinates						Color
RAI	Right Anterior Inferior	$x \geq 0$ mm	&	$y \geq$ -18.7 mm	&	$z < 12.8$ mm		Gray
RAS	Right Anterior Superior	$x \geq 0$ mm	&	$y \geq$ -18.7 mm	&	$z \geq 12.8$ mm		Pink
RPI	Right Posterior Inferior	$x \geq 0$ mm	&	$y <$ -18.7 mm	&	$z < 12.8$ mm		Brown
RPS	Right Posterior Superior	$x \geq 0$ mm	&	$y <$ -18.7 mm	&	$z \geq 12.8$ mm		Yellow
LAI	Left Anterior Inferior	$x < 0$ mm	&	$y \geq$ -18.7 mm	&	$z < 12.8$ mm		Orange
LAS	Left Anterior Superior	$x < 0$ mm	&	$y \geq$ -18.7 mm	&	$z \geq 12.8$ mm		Green
LPI	Left Posterior Inferior	$x < 0$ mm	&	$y <$ -18.7 mm	&	$z < 12.8$ mm		Blue
LPS	Left Posterior Superior	$x < 0$ mm	&	$y <$ -18.7 mm	&	$z \geq 12.8$ mm		Red

Table 1: Division of the brain into eight octants.
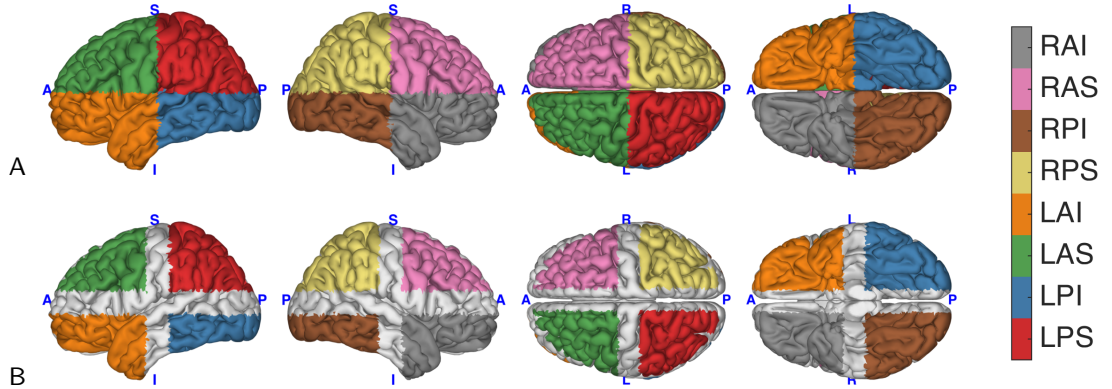
Fig. 2: Division of the cortex into eight octants. A: complete tesselation. B: seed regions used to sample source centers.

4.1 Spatial structure of the sources

In each instance of the experiment, two distinguished brain sources, the driver-receiver relationship of which is to be analyzed, are modeled. Each of these sources is confined to a different randomly assigned brain octant. Within each octant, a random node of the cortical mesh is picked as the center of the source activity. The locations $\mathbf{d}_1$ and $\mathbf{d}_2$ of these center nodes are required to be at least 10 mm away from the octant boundaries. Note that the randomized sampling of source locations leads to a considerable variation of source depth with sources in inferior regions being deeper than sources in corresponding superior regions, and sources in posterior regions being deeper than corresponding sources in anterior regions (see Figure 3). Here, depth is defined as the mean Euclidean distance of the center node from all scalp electrodes.

The spatial distribution of the source current amplitudes is modeled by a Gaussian function, where the geodesic distance between nodes of the cortical mesh is used as the distance metric. The spatial standard deviation of the amplitude distributions is sampled uniformly between 10 mm and 40 mm. The amplitude at nodes located outside the seed octant is set to zero, such that no 'leakage' of activity across octant borders occurs, and the true connectivity between octants can always be defined unambigously. The amplitude distributions are divided by their ℓ_2-norm for each

source separately. The orientation of the neuronal current at each node is defined as the normal vector w. r. t. the mesh surface at that node. Scalp topographies for each source are computed by multiplying the 3D current distribution $\mathbf{j}(t)$ (the product of amplitude and orientation) with the lead field $\mathbf{L}$, that is, by summing up the contributions from all nodes of the source octant using Eq. 1. Figure 3A depicts the source amplitude distributions, as well as the resulting scalp potentials, for two representative sources.
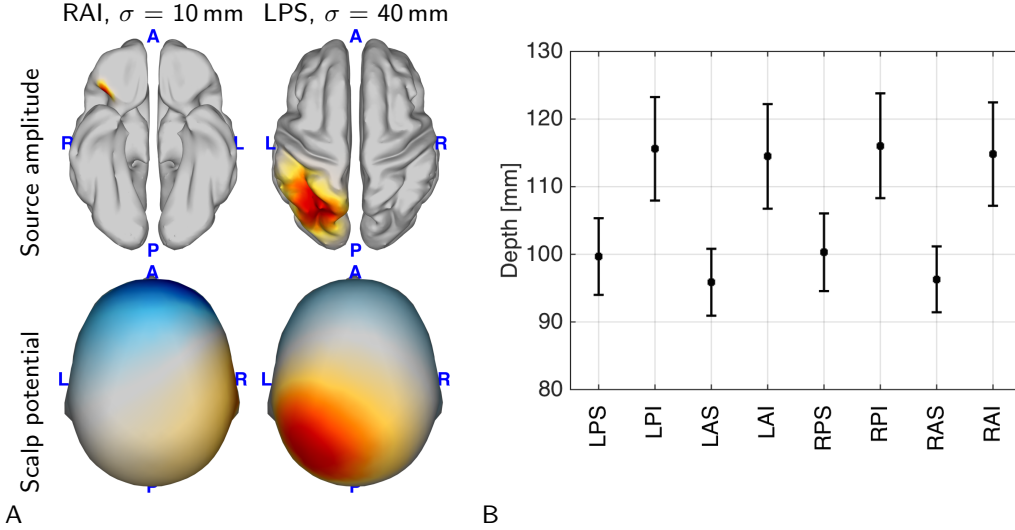

Fig. 3: A: Examples of the spatial structure of the simulated brain sources. Left: source with small spatial extent (spatial standard deviation along cortical manifold $\sigma = 10\,\mathrm{mm}$) in the right anterior inferior octant of the brain. Right: source with large ($\sigma = 40\,\mathrm{mm}$) spatial extent in the left posterior superior octant of the brain. Upper panel: source amplitude distribution. Note that sources do not extend into neighboring octants. Lower panel: resulting EEG field potentials calculated in the New York Head head model assuming currents oriented perpendicular to the cortical surface. B: Depth distribution (mean and standard deviation) of the cortical surface points belonging to each of the 8 ROIs. The depth of a point is defined as the mean Euclidean distance from all EEG sensors. Inferior regions are deeper than corresponding superior regions, and posterior regions are deeper than corresponding anterior regions.

4.2 Source dynamics

The time courses of the two distinguished sources are modeled using bivariate linear autoregressive (AR) models of the form

$$\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} a_{11}(p) & a_{12}(p) \\ a_{21}(p) & a_{22}(p) \end{bmatrix} \begin{bmatrix} z_1(t-p) \\ z_2(t-p) \end{bmatrix} + \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{bmatrix}, \tag{2}$$

where $a_{ij}(p)$, $i,j \in \{1,2\}, p \in \{1,\ldots,P\}$ are linear AR coefficients, and $\epsilon_i(t)$, $i \in \{1,2\}$ are uncorrelated standard normal distributed noise variables (innovations). Importantly, the off-diagonal entries $a_{12}(p)$ and $a_{21}(p)$ describe time-delayed linear influences of one source on another. A sampling rate of 100 Hz, and an AR model order of $P = 5$ is used.

In each instance of the simulation, either one out of two possible variants of the linear dynamical system $\mathbf{z}(t) = [z_1(t), z_2(t)]^\top$ is constructed. The probability to determine which variant is generated is set to 50 %. For the first variant, $\mathbf{z}^{\mathrm{int}}(t)$, $a_{12}(p)$, $p \in \{1,\ldots P\}$ is set to zero for all lags p, while $a_{21}(p)$, $p \in \{1,\ldots P\}$ is generally nonzero and randomly drawn. Thus, there is a unidirectional time-delayed influence of $z_1^{\mathrm{int}}(t)$ on $z_2^{\mathrm{int}}(t)$. For the second variant, $\mathbf{z}^{\mathrm{nonint}}(t)$, all offdiagonal coefficients $a_{12}(p)$ and $a_{21}(p)$, $p \in \{1,\ldots P\}$ are set to zero, leaving the two time series $z_1^{\mathrm{nonint}}(t)$ and $z_2^{\mathrm{nonint}}(t)$ completely mutually independent. In all cases, the AR coefficients are sampled from the univariate standard normal distribution. Only stable AR systems are selected. Moreover, we require that the combined spectral power of the two sources in the alpha band (8–13 Hz), normalized by the width of the alpha band exceeds the overall normalized power by a factor of 1.2. Sources are bandpass-filtered in the alpha band using an acausal third-order Butterworth filter with zero phase delay. The generated time series therefore represent alpha oscillations that are either mutually statistically independent or characterized by a clearly defined sender-receiver relationship.

4.3 Generation of pseudo-EEG data

Pseudo-EEG data are created from simulated underlying brain sources. We describe the procedure for the interacting sources $\mathbf{z}^{\mathrm{int}}(t)$, while analogous steps are taken for the independent sources $\mathbf{z}^{\mathrm{nonint}}(t)$. A total of $T = 18\,000$ data points of the source time series $\mathbf{z}^{\mathrm{int}}(t)$ are sampled from the bivariate AR process described above, corresponding to a 3 minute recording. These source time courses are then placed onto two patches of the cortical surface specified by the locations, orientations and Gaussian amplitude distributions described in Section 4.1, resulting in the source distribution $\mathbf{j}^{\mathrm{int}}(t)$. Additionally, 500 mutually statistically independent brain noise time series characterized by $1/f$-shaped (pink noise) power and random phase spectra are generated, and placed randomly at 500 locations sampled from the entire cortical surface. The resulting noise current density is denoted by $\mathbf{j}^{\mathrm{noise}}(t)$. The signal and noise sources from the brain are then normalized by their Frobenius norm in the alpha band, and summed up

$$\mathbf{j}(t) = \alpha \frac{\mathbf{j}^{\mathrm{int}}(t)}{\|\mathbf{j}^{\mathrm{int}}(t)\|_F} + (1-\alpha) \frac{\mathbf{j}^{\mathrm{noise}}(t)}{\|\tilde{\mathbf{j}}^{\mathrm{noise}}(t)\|_F} \; , \tag{3}$$

where $\tilde{\mathbf{j}}^{\mathrm{noise}}(t)$ is the result of filtering $\mathbf{j}^{\mathrm{noise}}(t)$ in the alpha band. The signal-to-noise (SNR) parameter α is drawn uniformly from the interval $[0.1, 0.9]$. The resulting source distribution $\mathbf{j}(t)$ is projected to the EEG sensors through multiplication with the lead field $\mathbf{L}$, giving rise to the brain contribution $\mathbf{x}^{\mathrm{brain}}(t)$ of the EEG. Spatially and temporally uncorrelated vectors $\mathbf{x}^{\mathrm{noise}}(t)$ mimicking measurement noise are sampled from a univariate standard normal distribution. The overall pseudo-EEG data are generated according to

$$\mathbf{x}^{\mathrm{int}}(t) = 0.9 \frac{\mathbf{x}^{\mathrm{brain}}(t)}{\|\mathbf{x}^{\mathrm{brain}}(t)\|_F} + 0.1 \frac{\mathbf{x}^{\mathrm{noise}}(t)}{\|\mathbf{x}^{\mathrm{noise}}(t)\|_F} \; . \tag{4}$$

Lastly, a highpass filter (third order acausal Butterworth) with cutoff frequency at 0.1 Hz is applied to $\mathbf{x}^{\mathrm{int}}(t)$.

In the same way as $\mathbf{x}^{\mathrm{int}}(t)$, $\mathbf{x}^{\mathrm{nonint}}(t)$ is generated by replacing the interacting sources $\mathbf{z}^{\mathrm{int}}(t)$ with their non-interacting counterparts $\mathbf{z}^{\mathrm{nonint}}(t)$. Thereby, it is ensured that $\mathbf{x}^{\mathrm{nonint}}(t)$ does not represent any form of brain interaction at all. In addition to the data set either containing interacting or non-interacting sources, an analogous baseline dataset $\mathbf{x}^{\mathrm{bl}}(t)$ lacking any contribution of the two alpha sources is constructed. This measurement mimicks (the absence of) a task involving the sources of interest, and can be used to aid source localization.

A complete instance of an experiment instance consists of two measurements: Either $\mathbf{x}^{\mathrm{int}}(t)$ or $\mathbf{x}^{\mathrm{nonint}}(t)$, and $\mathbf{x}^{\mathrm{bl}}(t)$. Examples of the spectral properties as well as the eigenstructure of the three measurements are depicted in Figure 4. The spectra of the measurements $\mathbf{x}^{\mathrm{int}}(t)$ and $\mathbf{x}^{\mathrm{nonint}}(t)$ resemble real EEG measurements in that the overall $1/f$-shaped spectrum is superimposed by a peak in the alpha band (Figure 4A and B). Since the alpha sources are not contained in the baseline measurement, this peak is lacking in $\mathbf{x}^{\mathrm{bl}}(t)$. The singular values of all three pseudo-measurement are almost indistinguishable (Figure 4C).
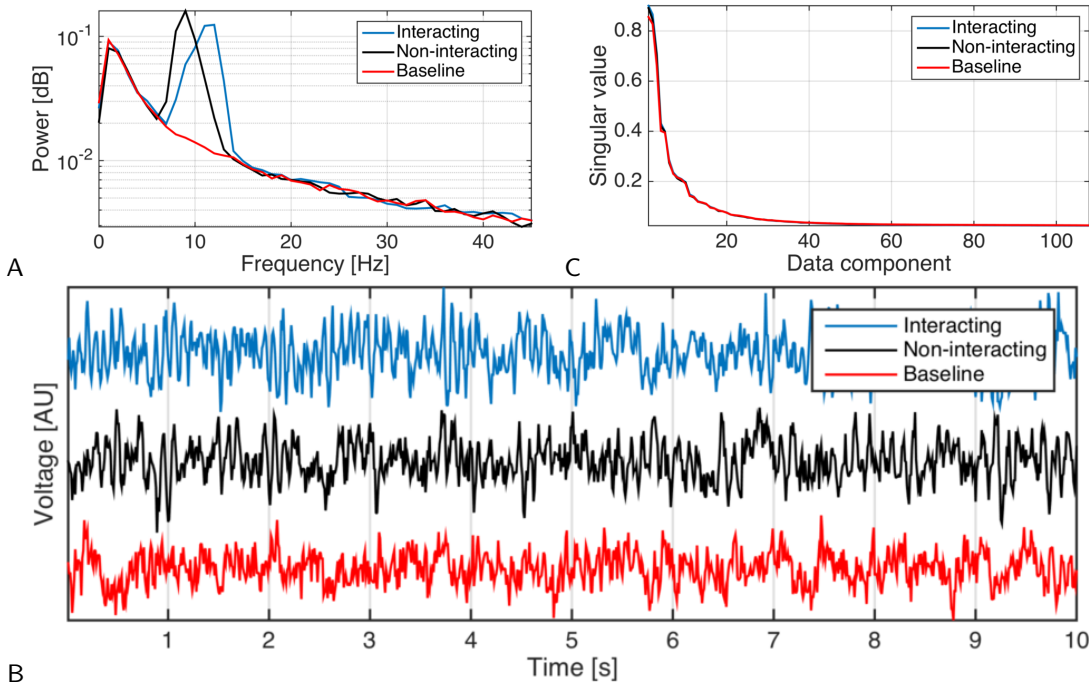

Fig. 4: Spectral properties of instances of the three pseudo EEG measurements $\mathbf{x}^{\mathrm{int}}(t)$ (interacting), $\mathbf{x}^{\mathrm{nonint}}(t)$ (non-interacting), and $\mathbf{x}^{\mathrm{bl}}(t)$ (baseline) at an SNR of 0 dB ($\alpha = 0.5$). A: Power spectra at an electrode with high alpha power visible as a distinct peak in $\mathbf{x}^{\mathrm{int}}(t)$ and $\mathbf{x}^{\mathrm{nonint}}(t)$. B: Time-series trace at that electrode. Note the absence of an alpha oscillation in $\mathbf{x}^{\mathrm{bl}}(t)$. C: Singular value spectra of all three measurements are almost indistinguishable.

4.4 Task

For each given instance of the experiment, the following three questions are asked:

1. **Localization:** In which two brain octants are the alpha sources of interest located?

2. **Connectivity:** Does the data set contain interaction between the two simulated brain sources $z_1(t)$ and $z_2(t)$?

3. **Direction:** If so, which of the two octants contains the sending source $z_1^{\mathrm{int}}(t)$, and which one contains the receiving source $z_2^{\mathrm{int}}(t)$?

4.5 Performance measures

The correctness of the answers to the above-mentioned questions is assessed quantitatively using the following three performance measures.

LOC

This measure compares the true octants containing the alpha sources with the estimated ones. Each source octant estimated correctly yields the score of 1/2. Each octant estimated wrongly yields the score of -1/2. A researcher can also choose to not estimate either one or both locations. For each case for which the estimation of a source octant is refused, 0 points are scored. According to these rules, the expected value under random guessing is -1/2.

CONN

This measure evaluates the correctness of the estimation of the presence of interaction in a particular data set. Correct estimates lead to a score of +1, whereas incorrect estimates are penalized with a score of -2. Again, researchers can refuse to make a decision, which leads to a score of 0. Thus, CONN may take one of the values -2, 0 and +1. The expected value under random guessing is -1/2. Please note that this measure is independent of the estimation of source locations.

DIR

This measure evaluates the correct assessment of interaction direction by comparing the estimated connectivity between estimated sources with the true connectivity between simulated sources. A correct estimation of directionality yields a score of 1, while a wrong estimation yields a score of -2. Refusal to estimate directionality yields 0 points. Note that the performance measure DIR depends on LOC and CONN estimates. A point can only be obtained if a) the source locations are estimated correctly and b) if interaction is present and correctly detected.

Note that for all three measures, the expected score under random guessing is smaller than the score obtained by not providing an answer. Refusing a decision in case of weak evidence is therefore a promising strategy to improve each score.

5 Example

5.1 A source connectivity estimation pipeline using LCMV, ImCoh and PSI

We implemented an example processing pipeline for answering the three questions outlined above using established methods from the literature. In this pipeline, we first calculate signal power on sensor level in order to estimate the signal-to-noise-ratio (SNR) within the alpha band. To compute the cross-spectrum, from which power and coherency can be calculated, we split the data in segments of one second (100 samples) length and use a segment overlap of 50 %. This results in a frequency resolution of 1 Hz. To obtain a scalar value as an estimate for signal power, we first take the mean power over all channels of the data (denoted as $p_d(f)$) at each frequency $8\,\text{Hz} \leq f \leq 13\,\text{Hz}$. Second, we take the maximum of $p_d(f)$ over all frequencies in the alpha band, i.e. $\max\left[p_d\left(8\,\text{Hz}\right)\ldots p_d(13\,\text{Hz})\right]$ which corresponds to the value of alpha peak in the power spectrum. As a noise power estimate we consider the mean over frequencies of the mean over channels of the provided baseline data, i.e., $\text{mean}\left[p_b\left(8\,\text{Hz}\right),\ldots,p_b(13\,\text{Hz})\right]$. This finally leads to the estimate of the SNR given by

$$\overline{\text{SNR}} = \frac{\max\left[p_d\left(8\text{Hz}\right)\ldots p_d(13\text{Hz})\right]}{\text{mean}\left[p_b\left(8\text{Hz}\right)\ldots p_b(13\text{Hz})\right]} \ . \tag{5}$$

The value $\overline{\text{SNR}}$ is used to decide if an estimate of the source locations is provided. Only for $\overline{\text{SNR}} > 1.5$, we consider the signal power as sufficiently strong to provide a reliable source estimate. Please note that the value of 1.5 is set rather arbitrarily and could be further optimized. However, our aim here is not to establish an optimal solution but to give an example of a possible processing pipeline with practical relevance. The imaginary part of coherency (ImCoh, Nolte et al., 2004) on sensor level is used to estimate if interaction is present in a data set. As a signal estimate of the ImCoh (ImCoh_d), we consider the absolute value of the maximum ImCoh over all channel pairs and frequency bins in the alpha band computed from the data set. A noise level is estimated as the same value computed on the baseline data set (ImCoh_b). If the difference $\text{ImCoh}_d - \text{ImCoh}_b$ exceeds the value 0.1, we here assume that there is interaction present in the data set. Similar to the threshold used before, proper statistics, e.g. by resampling procedures, might enhance the performance both in terms of sensitivity and by avoiding false positives. As an inverse filter we use a Linear-constrained Minimum Variance beamformer (LCMV, Van Veen et al., 1997) applied on the real part of the cross-spectrum averaged over all frequency bins in the alpha band. Given the inverse filter, the voxel-wise power is computed for the data and the baseline data. Denoting those powers by $\mathbf{p}_d$ and $\mathbf{p}_b$, a voxel-wise index of baseline-normalized alpha-band source power is defined as

$$\mathbf{p} = \frac{\mathbf{p}_d}{\mathbf{p}_b} \ . \tag{6}$$

The overall alpha power in each of the eight brain octants is computed by summing over the appropriate parts of $\mathbf{p}$. The two octants displaying highest overall alpha power are selected as the assumed source regions of the alpha oscillations. From each of these two regions, the voxel the highest power is selected. Next, for these two voxels, the source time series are are calculated by

beamforming the sensor level data to source space. Using the phase-slope index (PSI, Nolte et al., 2008), time-delayed interaction is assessed between this particular pair of selected voxels of the two octants. Here, we again use a 1 s window, amounting to a 1 Hz frequency resolution, and evaluate phase slopes only in the alpha frequency range between 8 and 13 Hz. Through a jackknife procedure, PSI provides a standardized signed measure of interaction direction, whose significance can be tested using a one-sample z-test. Only if the p-value is smaller than an alpha level of 0.01, we consider the interaction direction as reliable and provide an estimate.

To summarize the applied heuristic: a) only if the power SNR exceeds the threshold of 1.5, source location estimates are provided; b) only if the difference in ImCoh between data and baseline exceeds the threshold of 0.1, we consider an interaction being present in the data set; c) only if both prior conditions meet and PSI yields a significant result, we provide an estimate of interaction direction.

5.2 Results

We applied the analysis pipeline outlined above to 100 data sets generated using the procedures outlined in Sections 4.1-4.3, and obtained the following performance scores: LOC = 0.54 ± 0.05, CONN = 0.52 ± 0.11, DIR = 0.00 ± 0.05 (mean $\pm$ standard error).

Figure 5 shows the results obtained for one particular instance of the experiment generated with an SNR of $\alpha = 0.5$. The summed amplitude distribution of the two simulated sources is depicted in the first row on the left as a color-coded heat map. The source centers $\mathbf{d}_1$ and $\mathbf{d}_2$ are marked as black spheres. In this case, the simulated sources are located in the left parietal superior (LPS) and right anterior superior (RAS) octants of the brain, with the source in RAS exerting a unidirectional time-delayed influence on the source in LPS. The middle left part of the figure shows the source amplitude distribution $\mathbf{p}$ as estimated using LCMV beamforming. Both source locations are recovered, although the source in LPS is barely observable. The lower part of Figure 5 shows the power spectrum where a peak in the alpha range is observable compared to the baseline measurement. Here, the power offset leads to an estimated SNR greater than 1.5 such that a source estimated is provided. Therefore, the attained localization score is LOC = 1.

In the lower right, the ImCoh for all channel pairs is shown as a butterfly plot. Furthermore, the maximum over channel pairs of the absolute value of the ImCoh is displayed as a black dashed line. Within the frequency band of interest (shown as vertical green lines), the maximum ImCoh strongly exceeds the noise level. Hence, this data set is correctly classified to contain interaction yielding a score of CONN = 1.

The upper right part shows the phase-slope index for all voxels calculated with respect to $\mathbf{d}_1$. The positive values, indicated in red, show the sender of information. Hence, the analysis yields a significant information flow from RAS to LPS. As also all other prerequisite are fulfilled (sufficiently large SNR and interaction present), the attained direction score for this example is DIR = 1.

Note that this particular example was chosen mainly to demonstrate how directed connectivity can be assessed and benchmarked. The reported results should not be regarded as a definite statement about the performance of LCMV beamforming, ImCoh and PSI in general, as they

were obtained using a multitude of ad-hoc heuristics, e.g. related to the usage of the separate baseline measurement, arbitrarily chosen thresholds, and the usage of single voxel estimates in case of localization and directionality estimation. In contrast, these techniques could also be used in a multivariate fashion. In fact, rather than trying to optimize performance, our goal here was to provide a comprehensible analysis pipeline using conventional tools in order to provide a baseline.
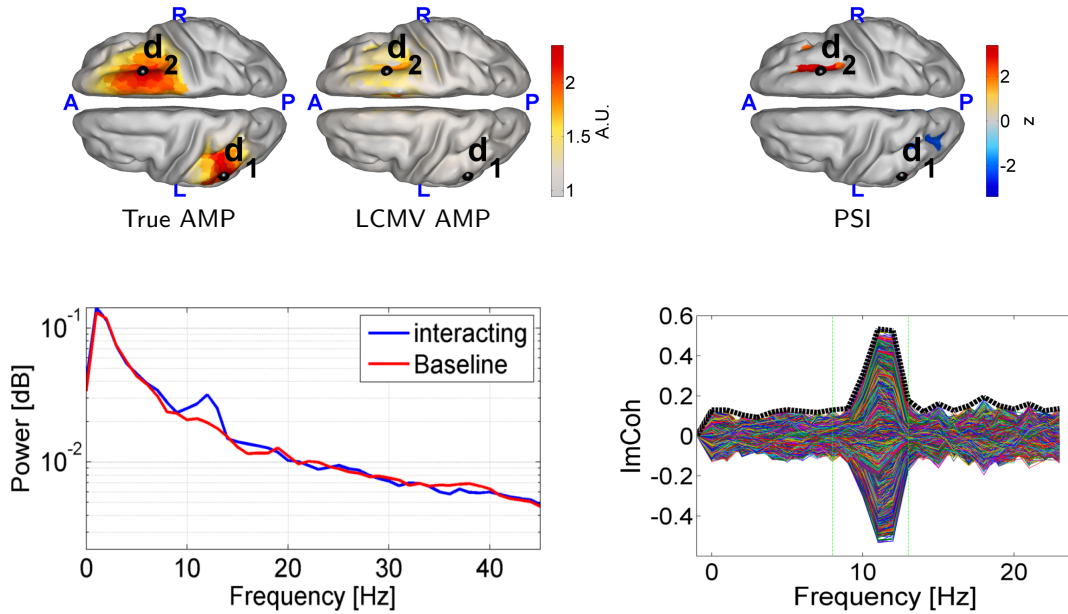

Fig. 5: Results obtained for a particular instance of the experiment with $\alpha = 0.5$ using LCMV beamforming, ImCoh and the phase-slope index (PSI). Left upper part: amplitude of the simulated sources, and of the sources estimated using LCMV beamforming. Centers $\mathbf{d}_1$ and $\mathbf{d}_2$ of the simulated sources are marked as black spheres. Upper right part: The phase slope index (PSI) with $\mathbf{d}_1$ as a reference. Lower left part: The power spectrum for the baseline condition and the data containing oscillatory interaction. Lower right part: The imaginary part of coherency for all channel pairs and the maximum of the ImCoh (black dashed line).

6 Discussion

Brain connectivity analysis using EEG/MEG is an emerging field of obvious relevance for basic neuroscience and clinical research. Nonetheless, there is currently a lack of validation, and a resulting confusion about the reliability of the available methods. It is in the interest of researchers using connectivity analyses to address research questions that all parts of an analysis pipeline including head modeling, source reconstruction, connectivity estimation, graph-theoretical analysis, and statistical testing are rigorously validated under circumstances as realistic as possible. In the absence of a clear external ground truth for real data, simulations using realistically modeled artificial data

are a suitable way to implement such a testbed. We here provide a simulation framework that can be used as a starting point for a systematic (re)evaluation of popular estimation pipelines. Here we first provide a discussion of our work including the limitations and possible extensions of the proposed framework. We then go on to a general discussion of the major problems in EEG/MEG-based brain connectivity estimation, and their potentials solutions.

6.1 Simulation framework and benchmark

Realism

Our aim was to create pseudo-EEG data as realistic as necessary to objectively benchmark connectivity analyses, but at the same time as simple and controlled as possible. To this end, we use a highly-detailed FEM volume conductor model of an average human head to map the activity of all simulated brain sources into EEG sensor space. The dynamics of all background brain sources are modeled by pink noise processes, whereas the interacting sources are generated as band-limited linear AR processes. The pseudo-EEG data obtained that way are very similar to real EEG data in terms of power spectra and spatial correlation structure.

As our model is rather simplistic compared to the complexity of the head and brain, any of our modeling choices is of course arguable. For example, the quasi-static approximation of Maxwell's equations underlying our forward modeling is considered accurate by the majority of researchers in the field, but has recently been questioned (Gomes et al., 2016). Other parameters of our simulation, such as those concerning the distributions and dynamics of the modeled brain and noise sources are also debatable. For example, one might question the linearity of the source interaction and the absence of any non-Gaussian signal or noise component, as well as the fact that our signal sources are restricted to the alpha band rather than obeying general 1/f spectra. To this end, we would like to note that none of these choices is meant to be definite. Rather, the proposed benchmark is is supposed to be just an example, and should be extended in the future to include more realistic pseudo-EEG data as well as to deal with more complex estimation tasks.

Regarding the present task, one might argue that the division of the brain into octants is artificial and does not represent the structural or functional organization of the brain. The actual anatomy of the brain is however irrelevant when dealing with simulated data as long as a realistic physical model of volume conduction is employed. The division into octants was preferred here over functionally-informed brain parcellations into units such as lobes or Brodman areas, as it facilitates a simplified and more flexible performance evaluation.

Further limitations and possible extensions

Here we discuss potential extension to our benchmark that should be implemented in the future to deal with more complex cases of brain interaction.

Non-linearity and non-Gaussianity We here only consider linear interaction using autoregressive models. To demonstrate advantages of non-linear or general non-parametric approaches to connec-

tivity estimation (Wibral et al., 2011; Marinazzo et al., 2011), it may be useful to create a separate branch of the benchmark including particular non-linearities such as neural mass models (Spiegler et al., 2010) or Kuramoto dynamics (Blythe et al., 2014; Rodrigues and Andrade, 2015). Nonlinear source dynamics can be easily included in the current framework by replacing the function that generates the bivariate source signals (see Appendix B in the Supplementary Material). As for the linear case, both an interacting and a non-interacting version of the source time series should be provided, and both should coincide in all essential data properties except the source coupling. Another limitation of the current framework is that all noise sources obey Gaussian distributions, while introducing a certain amount of non-Gaussian noise mimicking realistic artifacts would be more realistic.

Bi-directional interaction While we here only consider unidirectional information flow, as modeled, e.g., through bivariate AR models with all nonzero off-diagonal coefficients $a_{12}(p)$. Studying the more realistic bidirectional case is also worthwhile (Vinck et al., 2015). One way to do this would be to draw uni- and bidirectional source AR models with 50% probability, so that one of the following three cases can occur: only source 1 is sending, only source 2 is sending, both sources are sending information to each other. There are two possible ways in which the directionality measure DIR could be adapted to deal with the bidirectional case. First, the full score of 1 per spatial direction is only given if the correct one of the three possible cases is determined; otherwise the score is -2. Second, for each correctly estimated interaction direction (flow from octant 1 to octant 2 and flow from octant 2 to octant 1), a sub-score of $+1/2$ is given, while -1 is given for incorrect estimations.

More than two interacting sources and network analyses Graph-theoretical analyses of estimated source connectivity networks are becoming increasingly popular. Such approaches rely on the presence of dense network structures involving more than two interacting sources. To benchmark them, it will be necessary to extend the current simulation framework, as well as the performance measures LOC and DIR, accordingly.

Magnetoencephalography While this paper focuses on EEG, we are currently conducting an equivalent effort for MEG. To this end, we will provide lead fields for a common MEG system in the same anatomy considered here for EEG. This effort will not only make our framework useful for a larger community. It will also allow for direct comparisons of the two modalities in terms of answering the localization, connectivity presence and directionality questions on the same source configurations.

Alternative head models Finally, while we here provide a head model that can be used both to simulate and reconstruct sources, it is theoretically more appropriate to conduct the reconstruction part using a different model to avoid committing the so-called 'inverse crime' (Colton and Kress, 1997). As we here use a standardized geometry based on the MNI coordinate space and the 10-5 EEG electrode positioning system, this can be easily achieved using the templates readily available in software packages such as Brainstorm or Fieldtrip without the need to perform certain preprocessing steps such as MR segmentation and surface extraction.

Strategies

There are multiple ways in which the location and connectivity of the two simulated alpha sources could be estimated. The strategy pursued in the provided example (Section 5) is to apply inverse source reconstruction first, define the two brain octants containing alpha sources based on alpha power, and then analyze the directionality between those octants in a second step. Another valid way would be to first analyze the full connectivity graph after source localization, and then determine source octants based on maximal connectivity. Approaches using blind source separation techniques (Gómez-Herrero et al., 2008; Marzetti et al., 2008; Haufe et al., 2010; Ewald et al., 2012) or avoiding source representations entirely are also in principle valid as long as they lead to source octant and connectivity estimates. Another interesting strategy could be to use large-scale machine learning algorithms to learn localizations and connectivity structures from massive amounts of labeled training data that could be generated within our framework. The downsides of this approach may however be a general lack of insight into the estimation process as well as a lack of scalability to more complex estimation tasks possibly involving more than two sources and many more than the currently considered eight regions of interest.

Dissemination

All data as well as Matlab scripts required to generate new datasets, to run analyses using standard methods, and to evaluate performance measures are made publicly available at `http://bbci.de/supplementary/EEGconnectivity/`. We here chose Matlab as the development platform, as it is used in many of the most common software packages for EEG/MEG source analysis such as EEGlab, Fieldtrip and Brainstorm (Delorme et al., 2011; Tadel et al., 2011; Oostenveld et al., 2011). However, packages based on Python, such as MNE (Gramfort et al., 2013a), are gaining more and more popularity, as Python is a powerful and free alternative to Matlab. In this regard it is worth noting that the Matlab data structures provided here can also be read and processed using Python software.

In addition, we provide data generated in 100 randomized instances of the proposed experiment involving minimal examples of brain interaction (4.1–4.3), where the ground truth about the locations and connectivity structure of the sources of interest is withheld. As a *data analysis challenge*, researchers are invited to analyze these data with respect to source locations, the presence of connectivity, and the direction of information flow. Results that are sent to us will be evaluated in terms of the performance measures LOC, CONN and DIR, and the achieved scores will be announced online as well as at an upcoming conference. To finetune their methods, researchers may use the provided code to generate as many labeled training instances generated under the same distribution as they wish. An overview of the provided data and code is given in Appendices A and B in the Supplementary Material.

6.2 Problems caused by volume conduction

Various commonly used connectivity measures suffer from large numbers of false positive detections when applied to linear mixtures of source signals. That is, they may infer interactions with high confidence even if the underlying sources are entirely statistically independent. We call such methods 'non-robust' (w.r.t. volume conduction). Non-robustness is a serious problem, as source mixing due to volume conduction not only occurs on the EEG/MEG sensor level but is also present in any source estimate due to the inevitable spatial leakage introduced by the inverse operator. Benchmarking connectivity measures as part of an entire analysis chain starting from realistically generated pseudo-EEG data as opposed to, e.g., merely testing their ability to reconstruct the known connectivity structure of simulated sources is therefore of particular importance.

A counterexample for PLV, Coherence, Granger Causality, Transfer Entropy, PDC and DTF

Consider an example in which a single brain source $s(t)$ spreads to two EEG electrodes $x_1(t)$ and $x_2(t)$ as a result of volume conduction (alternatively, to two brain voxels $\hat{s}_1(t)$ and $\hat{s}_2(t)$ as a result of a source reconstruction procedure). Assume moreover that these two channels (EEG electrodes or brain voxels) pick up noises that are not completely correlated as a result of the measurement process or by capturing (the reconstructions of) additional independent brain processes. No interaction takes place. Yet, both channels are highly correlated due to being dominated by the same source's activity $s(t)$. This will be reflected in a stable phase-delay as measured by the phase-locking value (PLV, Lachaux et al., 1999) or coherence (the absolute value of complex coherency, see Nunez et al., 1997), and might be misinterpreted as brain interaction. The presence of linearly independent noise contributions in the two channels moreover allows both to mutually improve predictions of the other channel's future from past values. Thus, the two channels 'causally interact' in Granger's sense (Granger, 1969), as measured by varieties of Granger Causality (Granger, 1969; Geweke, 1982; Marinazzo et al., 2011), Transfer Entropy (Vicente et al., 2011; Wibral et al., 2011), Directed Transfer Function (DTF, Kamiński and Blinowska, 1991), and Partial Directed Coherence (PDC, Baccalá and Sameshima, 2001). Figure 6 depicts this situation. These measures are therefore non-robust.

Robust connectivity measures

Imaginary part of coherency Without further knowledge about the (e.g. mixing properties of the) data, is impossible to distinguish interactions happening with zero phase delay from trivial correlations that are artifacts of volume conduction. Non-robust measures make no distinction between zero and nonzero phase delays and are therefore prone to misinterpreting the often dominant instantaneous correlations in EEG/MEG data as genuine interaction. For methods based on quantifying phase-relations, a simple remedy is to analyze only the imaginary part of the cross-spectrum (Nolte et al., 2004, 2008), which is equivalent to requiring the phase delay between two interacting signals to be non-zero. Using this modification, measures such as coherence and the phase-locking value can be 'robustified'.
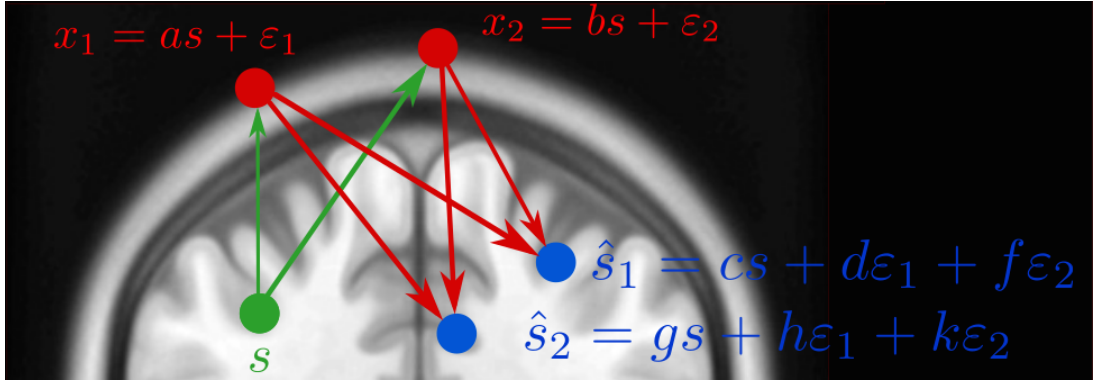
Fig. 6: Even in the case of only one brain source $s(t)$, sensors $x_1(t)$ and $x_2(t)$ as well as estimated sources $\hat{s}_1(t)$ and $\hat{s}_2(t)$ are highly correlated due to the mixing introduced by volume conduction (green arrows), as well as inverse source mapping (red arrows). This lead to high values of coherence, as well has high phase-locking values. As sensors and source estimates moreover pick up different noise realizations $\varepsilon_1(t)$ and $\varepsilon_2(t)$ (or linear combinations therefor), $x_1(t)$ and $x_2(t)$ as well as estimated sources $\hat{s}_1(t)$ and $\hat{s}_2(t)$ moreover 'Granger-cause' each other.

Time reversal An approach suitable for a larger class of interaction measures is time reversal. Its idea is to reverse the temporal order of the data, which can be done either in sensor- or source space without the need for decomposing the data using prior knowledge. Time-reversed data possess the same instantaneous correlations as the original data, whereas their connectivity structure is disrupted or even reversed. Statistical tests contrasting connectivity scores obtained on original and time-reversed data can effectively cancel out those parts of the effect that are due to linear mixing and thereby common to original and time-reversed data. It can be proven that time-reversal robustifies every measure that is based only on second-order statistics of the data. Notable, this includes many variants of Granger causality such as DTF and PDC (Haufe et al., 2012b,a). Moreover, theoretical results for the validity of time reversed Granger causality *in the presence of interaction* exist (Winkler et al., 2015). Applied to complex coherency, it can further be shown that the imaginary part is preserved under time reversal, while the real part cancels out.

Surrogate data approach Non-robustness of any connectivity measures can also be regarded as a result of improper statistical testing. Statistical significance of a measure is typically assessed by estimating its distribution and comparing it to a baseline. For small sample sizes, this approach may however result in low statistical power. An alternative is to compare a single measurement to a distribution obtained under a *null hypothesis.* While samples of the null distribution are often not available, it is possible to create arbitrary amounts of artificial null data by randomly manipulating the original data. This concept is known as the 'method of surrogate data' (Theiler et al., 1992). In brain connectivity studies, surrogate data must be consistent with the null hypothesis of no brain interaction while sharing all other properties of the original data. Common approaches include the use of randomly permuted samples (e.g., Muthuraman et al., 2014) or phase randomized signals (e.g., Astolfi et al., 2005). These approaches however destroy all dependencies between time series,

while it is easy to see (e.g., through the counterexample Figure 6) that EEG/MEG data as well as their source estimates *do* exhibit significant correlations under the null hypothesis of independent sources. Not accounting for these correlations results in too liberal tests for interaction.

The generation of surrogate data with realistic correlation structure is subject of ongoing research. By using the same random phase offsets for all variables, Prichard and Theiler (1994) extend the original phase-randomization approach to multivariate data in a way such that correlations between variables are preserved. Dolan and Neiman (2002) describe an improved procedure that preserves power spectra as well as coherence functions. Breakspear et al. (2003, 2004) use wavelets to approximate the spatio-temporal correlation structure of the original data, while Palus (2008) use a similar multiscale decomposition to even preserve certain non-linearities, which is of interest in resampling approaches. The problem of potential non-Gaussianity in phase-randomized data is discussed in Rath et al. (2012), while an overview over a number of surrogate approaches is provided in Marin Garcia et al. (2013).

What is common to all these approaches is that the correlation structure is estimated from the sensor data, and reproduced to the highest possible degree in the surrogates. Thereby, it might be difficult for these algorithms to distinguish between instantaneous correlation caused solely by volume conduction, and correlations that are side-effects of actual time-delayed or non-linear interaction. This potential problem can however be circumvented by considering that the spatial correlation of EEG/MEG data under the Null hypothesis is solely determined by the lead field.

We here propose to generate *physically realistic* surrogate EEG/MEG data as follows. First, the sensor-space EEG/MEG data are mapped into source space using a 'genuine' linear inverse solution $\mathbf{P}$ with $\mathbf{LP} = \mathbf{I}$. 'Genuine' here refers to methods providing an actual estimate of the distributed current density $\mathbf{j}(t)$ as opposed to a voxel-wise activity index, and includes popular approaches such as the weighted minimum-norm estimate (Hämäläinen and Ilmoniemi, 1994) and eLORETA (Pascual-Marqui, 2007), but not beamformers. Next, potential interactions between sources are destroyed while maintaining other statistical properties of the data such as the power spectrum by using e. g., the method of Theiler et al. (1992). This gives rise to quasi-independent sources $\breve{\mathbf{j}}(t)$. Finally, using the the lead field $\mathbf{L}$, the quasi-independent time series are mapped back to sensor space to obtain *surrogate sensor-space data* $\tilde{\mathbf{x}}(t) = \mathbf{L}\breve{\mathbf{j}}(t)$. These surrogates possess the spatio-spectral correlation structure of real EEG/MEG data, but by construction contain no interaction. They are then analyzed using the same pipeline that is applied to the original data. A Null distribution of the statistics of interest (such as Granger scores) is established by repeating the analysis for sufficiently many surrogate EEG/MEG datasets derived from different phase-randomizations of the sources.

The procedure depends on the choice of a volume conductor model $\mathbf{L}$. As correct anatomical information is not crucial for generating the surrogates, generic head models (e.g., Huang et al., 2015) can be used. The criterion $\mathbf{LP} = \mathbf{I}$ ensures that the procedure will recover the original data when leaving out the phase randomization step, and therefore does not introduce any artificial bias. It can always be met by setting a method's regularization to zero. The application of the regular estimation pipeline to the sensor-space surrogates may in general involve another inverse source reconstruction step $\tilde{\mathbf{j}}(t) = \mathbf{P}^*\tilde{\mathbf{x}}(t)$. Here, $\mathbf{P}^*$ is not necessarily identical to $\mathbf{P}$ as it may involve regularization or amount to a beamformer. What is important to note is that the *source-space*

surrogates obtained that way fundamentally differ from the quasi-independent sources $\check{\mathbf{j}}(t)$ obtained after phase-randomization in that they possess a realistic spatial correlation structure while still perfectly complying with the null hypothesis.

The proposed method makes use of a volume conductor model to obtain surrogate data with realistic spatio-temporal correlation structure. By generating these surrogates as realistic mixtures of quasi-independent sources, it is per construction *robust*. Its statistical power in revealing non-trivial interactions however needs to be rigorously quantitatively assessed and be compared to related methods (see above), which is subject of our future work. One particular method that will be included in the comparison is the method of Shahbazi et al. (2010). By employing blind source separation (BSS) instead of a source imaging procedure, their method is capable of generating surrogate data with similar theoretical properties, while bearing the further advantage that no physical volume conductor model is required.

Potential remaining confounds

Significant findings by robust connectivity measures must relate to genuine time-delayed interaction that cannot be explained by linear mixing of independent sources. Whether such findings can be safely interpreted as reflecting neural as opposed to non-neural interaction however requires further validation efforts. One problem may be the occurrence of hidden causes, which cannot be excluded in practice due to the fact that EEG/MEG do not measure the entirety of human brain activity but only a tiny fraction of it. It is moreover conceivable that detected time-lagged interactions are caused by external factors like hemodynamics affecting different parts of the brain at different time lags. The influence of such potential confounds should be investigated in further simulation studies. To prevent inexact claims, it is further advisable to report results in terms of the measured quantities (such as the phase slope index) only, and to be very careful when interpreting these results in terms of their hypothesized underlying physiological phenomena such as neural interaction.

Contrasting experimental conditions

It is often assumed that by comparing connectivity scores between conditions the problem of spurious connectivity can be overcome, as the 'effect of volume conduction cancels out'. However, this is not the case (see also Schoffelen and Gross, 2009). Volume conduction causes a linear superposition of the *active* brain sources. It is therefore not a static factor but highly dependent on the actual source activity. The value of any connectivity measure – even robust ones – depends not only on the strength of the coupling of the interacting sources, but to a large degree also on the SNR, which is a function of the strength of all sources including those not participating in the interaction. To effectively rule out the effect of volume conduction and SNR, one would require that the all brain and noise sources are identically active across conditions. This is however an unrealistic assumption, as lower-level EEG/MEG features such as the power of certain brain rhythms are much more likely to be measurably modulated by the experimental condition than the strength of the interaction between such rhythms. As a result, even robust measures are unable to objectively estimate the strength of an interaction unless they can be perfectly isolated through

signal preprocessing. As a consequence, statements about the relative strength of an interaction between experimental conditions have to be made very carefully. In the present benchmark, each generated dataset provides a baseline measurement and a measurement that may or may not contain brain interaction. Using these data it is possible to demonstrate that contrasting measurements is not sufficient to rule out spurious connectivity for non-robust connectivity measures.

Analysis of network properties

Results of brain connectivity analyses are often further analyzed to infer structural properties of the connectivity graphs such as 'small-worldness'. The validity of such analyses obviously depends on the correctness of the identified connections. For non-robust measures, the estimated network may to a large degree reflect factors unrelated to connectivity, which can of course render graph measures meaningless. Due to the non-random structure of the connectivity graph induced by these factors, the popular way of obtaining Null distributions by randomly permuting the edges (e.g., Bola and Sabel, 2015) is unlikely to lead to sufficiently robust statistics revealing genuine interaction. But even for robust measures it is conceivable that, while all detected interactions are genuine, the structure of the connectivity graph reflects rather trivial properties related to volume conduction such as the SNR distribution.

7 Conclusion

We present an open and extendable software framework for benchmarking EEG-based brain connectivity estimators using simulated data. Code and benchmark data are made available. We hope that this work will help to resolve the present confusion about the reliability of commonly used data analysis pipelines, and to establish realistic simulations as a standard way of validating methodologies before applying them to neuroscience questions.

References

Acar, Z. A., Makeig, S., Jul. 2010. Neuroelectromagnetic forward head modeling toolbox. Journal of Neuroscience Methods 190 (2), 258–270.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccalà, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Babiloni, F., 2006. Assessing cortical functional connectivity by partial directed coherence: simulations and application to real data. IEEE Trans Biomed Eng 53, 1802–1812.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Edgar, J. C., Miller, G. A., He, B., Babiloni, F., Feb 2007. Comparison of different cortical connectivity estimators for high-resolution EEG recordings. Hum Brain Mapp 28 (2), 143–157.

Astolfi, L., et al., 2005. Assessing cortical functional connectivity by linear inverse estimation and directed transfer function: simulations and application to real data. Clin Neurophysiol 116 (4), 920–932.

Baccalá, L. A., Sameshima, K., 2001. Partial directed coherence: a new concept in neural structure determination. Biol Cybern 84, 463–474.

Baillet, S., Mosher, J. C., Leahy, R. M., 2001a. Electromagnetic brain mapping. IEEE Signal Proc Mag 18, 14–30.

Baillet, S., Riera, J. J., Marin, G., Mangin, J. F., Aubert, J., Garnero, L., Jan 2001b. Evaluation of inverse methods and head models for EEG source localization using a human skull phantom. Phys Med Biol 46 (1), 77–96.

Barnett, L., Seth, A. K., Feb 2014. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. J. Neurosci. Methods 223, 50–68.

Barrett, A. B., Murphy, M., Bruno, M. A., Noirhomme, Q., Boly, M., Laureys, S., Seth, A. K., 2012. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. PLoS ONE 7 (1), e29072.

Barttfeld, P., Petroni, A., Baez, S., Urquina, H., Sigman, M., Cetkovich, M., Torralva, T., Torrente, F., Lischinsky, A., Castellanos, X., Manes, F., Ibanez, A., 2014. Functional connectivity and temporal variability of brain connections in adults with attention deficit/hyperactivity disorder and bipolar disorder. Neuropsychobiology 69 (2), 65–75.

Benar, C. G., Grova, C., Kobayashi, E., Bagshaw, A. P., Aghakhani, Y., Dubeau, F., Gotman, J., May 2006. EEG-fMRI of epileptic spikes: concordance with EEG source localization and intracranial EEG. Neuroimage 30 (4), 1161–1170.

Blythe, D. A. J., Haufe, S., Müller, K.-R., Nikulin, V. V., 2014. The effect of linear mixing in the EEG on hurst exponent estimation. NeuroImageIn press.

Bola, M., Sabel, B. A., Mar 2015. Dynamic reorganization of brain functional networks during cognition. Neuroimage.

Breakspear, M., Brammer, M., Robinson, P. A., Aug. 2003. Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. Physica D Nonlinear Phenomena 182, 1–22.

Breakspear, M., Brammer, M. J., Bullmore, E. T., Das, P., Williams, L. M., Sep 2004. Spatiotemporal wavelet resampling for functional neuroimaging data. Hum Brain Mapp 23 (1), 1–25.

Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., Owen, J. P., Morris, P. G., Nagarajan, S. S., Jun 2011. Measuring functional connectivity using MEG: methodology and comparison with fcMRI. Neuroimage 56 (3), 1082–1104.

Brookes, M. J., Woolrich, M. W., Barnes, G. R., 2012. Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage. Neuroimage 63 (2), 910–920.

Castaño Candamil, S., Höhne, J., Martinez-Vargas, J.-D., An, X.-W., Castellanos-Dominguez, G., Haufe, S., 2015. Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints. Neuroimage (in press).

Chella, F., Marzetti, L., Pizzella, V., Zappasodi, F., Nolte, G., 2014. Third order spectral analysis robust to mixing artifacts for mapping cross-frequency interactions in EEG/MEG. Neuroimage 91, 146–161.

Cho, J. H., Vorwerk, J., Wolters, C. H., Knosche, T. R., 2015. Influence of the head model on EEG and MEG source connectivity analyses. Neuroimage 110, 60–77.

Colton, D., Kress, R., 1997. Inverse Acoustic and Electromagnetic Scattering Theory. Applied Mathematical Sciences. Springer Berlin Heidelberg.

Darvas, F., Pantazis, D., Kucukaltun-yildirim, E., Leahy, R. M., 2004. Mapping human brain function with meg and eeg: methods and validation. NeuroImage 23, 289–299.

De Vico Fallani, F., Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Salinari, S., Kurths, J., Gao, S., Cichocki, A., Colosimo, A., Babiloni, F., 2007. Cortical functional connectivity networks in normal and spinal cord injured patients: Evaluation by graph analysis. Hum Brain Mapp 28 (12), 1334–1346.

Delorme, A., Makeig, S., Mar 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., Makeig, S., 2011. EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. Comput Intell Neurosci 2011, 130714.

Ding, L., He, B., 2008. Sparse source imaging in EEG with accurate field modeling. Hum Brain Mapp 29, 1053–1067.

Dolan, K. T., Neiman, A., Feb 2002. Surrogate analysis of coherent multichannel data. Phys Rev E Stat Nonlin Soft Matter Phys 65 (2 Pt 2), 026108.

Evans, A., Collins, D., Mills, S. R., Brown, E. D., Kelly, R. L., Peters, T., Oct. 1993. 3D statistical neuroanatomical models from 305 MRI volumes. In: Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record. pp. 1813–1817 vol.3.

Ewald, A., Marzetti, L., Zappasodi, F., Meinecke, F. C., Nolte, G., 2012. Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space. NeuroImage 60 (1), 476 – 488.

Ewald, A., Nolte, F. S. G., 2013. Identifying causal networks of neuronal sources from eeg/meg data with the phase slope index: a simulation study. Biomed Tech 22, 1 – 14.

Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D., Jul. 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage 47, S102.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., Brain Development Cooperative Group, Jan. 2011. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54 (1), 313–327.

Geffroy, D., Rivière, D., Denghien, I., Souedet, N., Laguitton, S., Cointepas, Y., Aug. 2011. Brainvisa: a complete software platform for neuroimaging. In: Python in Neuroscience workshop. Paris.

Geweke, J., 1982. Measurement of Linear Dependence and Feedback Between Multiple Time Series. Journal of the American Statistical Association 77 (378), 304–313.

Gomes, J. M., Bedard, C., Valtcheva, S., Nelson, M., Khokhlova, V., Pouget, P., Venance, L., Bal, T., Destexhe, A., Jan 2016. Intracellular Impedance Measurements Reveal Non-ohmic Properties of the Extracellular Medium around Neurons. Biophys. J. 110 (1), 234–246.

Gómez-Herrero, G., Atienza, M., Egiazarian, K., Cantero, J. L., 2008. Measuring directional coupling between EEG sources. NeuroImage 43, 497–508.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hamalainen, M., Dec 2013a. MEG and EEG data analysis with MNE-Python. Front Neurosci 7, 267.

Gramfort, A., Strohmeier, D., Haueisen, J., Hamalainen, M. S., Kowalski, M., Apr 2013b. Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations. Neuroimage 70, 410–422.

Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., Jan 2001. Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U.S.A. 98 (2), 694–699.

Grova, C., Daunizeau, J., Kobayashi, E., Bagshaw, A. P., Lina, J. M., Dubeau, F., Gotman, J., Jan 2008. Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes. Neuroimage 39 (2), 755–774.

Hämäläinen, M., Ilmoniemi, R., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32, 35–42.

Haufe, S., 2011. Towards EEG *source* connectivity analysis. Ph.D. thesis, Berlin Institute of Technology.

Haufe, S., Huang, Y., Parra, L. C., 2015. A highly detailed FEM volume conductor model of the ICBM152 average head template for EEG source imaging and tCS targeting. In: Conf Proc IEEE Eng Med Biol Soc. In Press.

Haufe, S., Nikulin, V., Ziehe, A., Müller, K.-R., Nolte, G., 2008. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage 42, 726–738.

Haufe, S., Nikulin, V. V., Müller, K.-R., Nolte, G., 2012a. A critical assessment of connectivity measures for EEG data: a simulation study. NeuroImage 64, 120–133.

Haufe, S., Nikulin, V. V., Nolte, G., 2012b. Alleviating the influence of weak data asymmetries on Granger-causal analyses. In: Theis, F., Cichocki, A., Yeredor, A., Zibulevsky, M. (Eds.), Latent Variable Analysis and Signal

Separation. Vol. 7191 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 25–33.

Haufe, S., Nikulin, V. V., Ziehe, A., Müller, K.-R., Nolte, G., 2009. Estimating vector fields using sparse basis field expansions. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems 21. MIT Press, pp. 617–624.

Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., Müller, K.-R., 2011. Large-scale EEG/MEG source localization with spatial flexibility. NeuroImage 54, 851–859.

Haufe, S., Tomioka, R., Nolte, G., Müller, K.-R., Kawanabe, M., 2010. Modeling sparse connectivity between underlying brain sources for EEG/MEG. IEEE Trans Biomed Eng 57, 1954–1963.

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., Engel, A. K., Jun 2012. Large-scale cortical correlation structure of spontaneous oscillatory activity. Nat. Neurosci. 15 (6), 884–890.

Huang, Y., Parra, L. C., Haufe, S., 2015. The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting. NeuroImage. In Press.

Kamiński, M. J., Blinowska, K. J., 1991. A new method of the description of the information flow in the brain structures. Biol Cybern 65, 203–210.

Kiebel, S. J., David, O., Friston, K. J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. Neuroimage 30 (4), 1273–1284.

Kiebel, S. J., Garrido, M. I., Moran, R. J., Friston, K. J., 2008. Dynamic causal modelling for EEG and MEG. Cogn Neurodyn 2, 121–136.

Korzeniewska, A., Mańczak, M., Kamiński, M., Blinowska, K. J., Kasicki, S., May 2003. Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. J. Neurosci. Methods 125 (1-2), 195–207.

Lachaux, J. P., Rodriguez, E., Martinerie, J., Varela, F. J., 1999. Measuring phase synchrony in brain signals. Hum Brain Mapp 8 (4), 194–208.

Lantz, G., Spinelli, L., Menendez, R. G., Seeck, M., Michel, C. M., Jul 2001. Localization of distributed sources and comparison with functional MRI. Epileptic Disord Special Issue, 45–58.

Leahy, R. M., Mosher, J. C., Spencer, M. E., Huang, M. X., Lewine, J. D., Aug 1998. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. Electroencephalogr Clin Neurophysiol 107 (2), 159–173.

Marin Garcia, A. O., Muller, M. F., Schindler, K., Rummel, C., Oct 2013. Genuine cross-correlations: which surrogate based measure reproduces analytical results best? Neural Netw 46, 154–164.

Marinazzo, D., Liao, W., Chen, H., Stramaglia, S., 2011. Nonlinear connectivity by Granger causality. Neuroimage 58 (2), 330–338.

Marzetti, L., Del Gratta, C., Nolte, G., 2008. Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. NeuroImage 42, 87–98.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., Lancaster, J., Jun. 1995. A probabilistic atlas of the human brain: theory and rationale for its development. the international consortium for brain mapping (ICBM). NeuroImage 2 (2), 89–101.

Mosher, J. C., Leahy, R. M., 1999. Source localization using recursively applied and projected (RAP) MUSIC. IEEE Trans Signal Proces 47, 332–340.

Mullen, T., Acar, Z. A., Worrell, G., Makeig, S., 2011. Modeling cortical source dynamics and interactions during seizure. Conf Proc IEEE Eng Med Biol Soc 2011, 1411–1414.

Muthuraman, M., et al., 2014. Beamformer source analysis and connectivity on concurrent EEG and MEG data during voluntary movements. PLoS ONE 9 (3), e91441.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., Hallett, M., 2004. Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin Neurophysiol 115, 2292–2307.

Nolte, G., Ziehe, A., Nikulin, V. V., Schlögl, A., Krämer, N., Brismar, T., Müller, K. R., 2008. Robustly estimating the flow direction of information in complex physical systems. Phys Rev Lett 100, 234101.

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., Cadusch, P. J., 1997. EEG coherency. I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr Clin Neurophysiol 103, 499–515.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience 2011, 1–9.

Oostenveld, R., Praamstra, P., Apr. 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 112 (4), 713–719.

Ou, W., Hämäläinen, M. S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG inverse solver. NeuroImage 44, 932–946.

Owen, J. P., Wipf, D. P., Attias, H. T., Sekihara, K., Nagarajan, S. S., Mar 2012. Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data. Neuroimage 60 (1), 305–323.

Palus, M., Sep 2008. Bootstrapping multifractals: surrogate data from random cascades on wavelet dyadic trees. Phys. Rev. Lett. 101 (13), 134101.

Pascual-Marqui, R., Michel, C., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18, 49–65.

Pascual-Marqui, R. D., Oct. 2007. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization.

Prichard, D., Theiler, J., 1994. Generating surrogate data for time series with several simultaneously measured variables. Phys. Rev. Lett. 73 (7), 951–954.

Rath, C., Gliozzi, M., Papadakis, I. E., Brinkmann, W., Oct 2012. Revisiting algorithms for generating surrogate time series. Phys. Rev. Lett. 109 (14), 144101.

Rivière, D., Régis, J., Cointepas, Y., Papadopoulos-Orfanos, D., Cachia, A., Mangin, J.-F., 2003. A freely available Anatomist/BrainVISA package for structural morphometry of the cortical sulci. In: Proc. 9th HBM. Neuroimage 19(2). New York, p. 934.

Rodrigues, J., Andrade, A., 5 2015. Synthetic neuronal datasets for benchmarking directed functional connectivity metrics. PeerJ 3, e923.

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

Sameshima, K., Takahashi, D. Y., Baccalá, L. A., 2015. On the statistical performance of granger-causal connectivity estimators. Brain Informatics, 1–15 (in press).

Sarvas, J., Jan. 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Physics in Medicine and Biology 32 (1), 11.

Sasaki, T., Abe, M., Okumura, E., Okada, T., Kondo, K., Sekihara, K., Ide, W., Kamada, H., 2013. Disturbed resting functional inter-hemispherical connectivity of the ventral attentional network in alpha band is associated with unilateral spatial neglect. PLoS ONE 8 (9), e73416.

Schelter, B., Timmer, J., Eichler, M., Apr 2009. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. J. Neurosci. Methods 179 (1), 121–130.

Schoffelen, J. M., Gross, J., 2009. Source connectivity analysis with MEG and EEG. Hum Brain Mapp 30, 1857–1865.

Shahbazi, F., Ewald, A., Nolte, G., May 2015. Self-Consistent MUSIC: An approach to the localization of true brain interactions from EEG/MEG data. Neuroimage 112, 299–309.

Shahbazi, F., Ewald, A., Ziehe, A., Nolte, G., 2010. Constructing surrogate data to control for artifacts of volume conduction for functional connectivity measures. In: Magjarevic, R., Nagel, J. H., Supek, S., Susac, A. (Eds.), 17th International Conference on Biomagnetism Advances in Biomagnetism - Biomag2010. Vol. 28 of IFMBE Proceedings. Springer Berlin Heidelberg, pp. 207–210.

Siegel, M., Donner, T. H., Engel, A. K., 2012. Spectral fingerprints of large-scale neuronal interactions. Nat. Rev. Neurosci. 13 (2), 121–134.

Silfverhuth, M. J., Hintsala, H., Kortelainen, J., Seppanen, T., Jul 2012. Experimental comparison of connectivity measures with simulated EEG signals. Med Biol Eng Comput 50 (7), 683–688.

Spiegler, A., Kiebel, S. J., Atay, F. M., Knosche, T. R., 2010. Bifurcation analysis of neural mass models: Impact of extrinsic inputs and dendritic time constants. Neuroimage 52 (3), 1041–1058.

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., Friston, K. J., 2007. Dynamic causal models of neural system dynamics:current state and future extensions. J. Biosci. 32 (1), 129–144.

Supp, G. G., Schlögl, A., Trujillo-Barreto, N., Müller, M. M., Gruber, T., 2007. Directed cortical information flow during human object recognition: analyzing induced EEG gamma-band responses in brain's source space. PLoS ONE 2, e684.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., Leahy, R. M., Apr. 2011. Brainstorm: A user-friendly application for MEG/EEG analysis. Computational Intelligence and Neuroscience 2011.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J. D., 1992. Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena 58 (14), 77 – 94.

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., Friston, K., Sep 2011. Effective connectivity: influence, causality and biophysical modeling. Neuroimage 58 (2), 339–361.

van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., Marinazzo, D., Oct 2014. Functional brain connectivity from EEG in epilepsy: seizure prediction and epileptogenic focus localization. Prog. Neurobiol. 121, 19–35.

Van Veen, B. D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44, 867–880.

Velez-Perez, H., Louis-Dorr, V., Ranta, R., Dufaut, M., 2008. Connectivity estimation of three parametric methods on simulated electroencephalogram signals. Conf Proc IEEE Eng Med Biol Soc 2008, 2606–2609.

Vicente, R., Wibral, M., Lindner, M., Pipa, G., Feb 2011. Transfer entropy–a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci 30 (1), 45–67.

Vinck, M., Huurdeman, L., Bosman, C. A., Fries, P., Battaglia, F. P., Pennartz, C. M., Tiesinga, P. H., 2015. How to detect the Granger-causal flow direction in the presence of additive noise? Neuroimage 108, 301–318.

Vorwerk, J., Cho, J.-H., Rampp, S., Hamer, H., Knösche, T. R., Wolters, C. H., Oct. 2014. A guideline for head volume conductor modeling in EEG and MEG. NeuroImage 100, 590–607.

Vulliemoz, S., Lemieux, L., Daunizeau, J., Michel, C. M., Duncan, J. S., Apr 2010. The combination of EEG source imaging and EEG-correlated functional MRI to map epileptic networks. Epilepsia 51 (4), 491–505.

Wibral, M., Pampu, N., Priesemann, V., Siebenhuhner, F., Seiwert, H., Lindner, M., Lizier, J. T., Vicente, R., 2013. Measuring information-transfer delays. PLoS ONE 8 (2), e55809.

Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., Kaiser, J., Mar 2011. Transfer entropy in magnetoencephalographic data: quantifying information flow in cortical and cerebellar networks. Prog. Biophys. Mol. Biol. 105 (1-2), 80–97.

Winkler, I., Panknin, D., Bartz, D., Müller, K.-R., Haufe, S., 2015. Validity of time reversal for testing Granger causality. IEEE Trans Sig Process. In Press.

Zwoliński, P., Roszkowski, M., Żygierewicz, J., Haufe, S., Nolte, G., Durka, P. J., 2010. Open database of epileptic EEG with MRI and postoperational assessment of foci—a real world verification for the EEG inverse solutions. Neuroinformatics 8, 285–299.