NeuroImage 64 (2013) 120–133

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A critical assessment of connectivity measures for EEG data: A simulation study

Stefan Haufe [a,b,*], Vadim V. Nikulin [c,d], Klaus-Robert Müller [a,b,e], Guido Nolte [f]

[a] Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, Franklinstr. 28/29, 10587 Berlin, Germany
[b] Bernstein Focus Neurotechnology, Berlin, Germany
[c] Neurophysics Group, Department of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, 12203 Berlin, Germany
[d] Bernstein Center for Computational Neuroscience, Berlin, Germany
[e] Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136–713, Republic of Korea
[f] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

ARTICLE INFO

Article history:
Accepted 14 September 2012
Available online 21 September 2012

Keywords:
EEG
Effective connectivity
Inverse source reconstruction
GC
PDC
PSI
WMN
S-FLEX
LCMV

ABSTRACT

Information flow between brain areas is difficult to estimate from EEG measurements due to the presence of noise as well as due to volume conduction. We here test the ability of popular measures of effective connectivity to detect an underlying neuronal interaction from simulated EEG data, as well as the ability of commonly used inverse source reconstruction techniques to improve the connectivity estimation. We find that volume conduction severely limits the neurophysiological interpretability of sensor-space connectivity analyses. Moreover, it may generally lead to conflicting results depending on the connectivity measure and statistical testing approach used. In particular, we note that the application of Granger-causal (GC) measures combined with standard significance testing leads to the detection of spurious connectivity regardless of whether the analysis is performed on sensor-space data or on sources estimated using three different established inverse methods. This empirical result follows from the definition of GC. The phase-slope index (PSI) does not suffer from this theoretical limitation and therefore performs well on our simulated data.

We develop a theoretical framework to characterize artifacts of volume conduction, which may still be present even in reconstructed source time series as zero-lag correlations, and to distinguish their time-delayed brain interaction. Based on this theory we derive a procedure which suppresses the influence of volume conduction, but preserves effects related to time-lagged brain interaction in connectivity estimates. This is achieved by using time-reversed data as surrogates for statistical testing. We demonstrate that this robustification makes Granger-causal connectivity measures applicable to EEG data, achieving similar results as PSI. Integrating the insights of our study, we provide a guidance for measuring brain interaction from EEG data. Software for generating benchmark data is made available.

Introduction

Due to its temporal resolution in the millisecond range as well as its noninvasiveness, portability and relatively low costs, electroencephalography (EEG) is a popular and widely used measurement technique for studying brain dynamics and interaction in humans. However, any neurophysiological interpretation of EEG data is hindered by the fact that the signals related to electrical activity in source brain regions are spread across the EEG sensors due to volume conduction in the head. The inversion of volume conduction is an ill-posed inverse problem. In the EEG-based analysis of information flow between brain regions (Friston, 1994; Horwitz, 2003; Jirsa and McIntosh, 2007), volume conduction poses a serious challenge, since multiple active sources are usually present, the contributions of which mix into all EEG sensors. Nevertheless, EEG recordings have been widely used in neuroscience to estimate brain connectivity (e.g., Astolfi et al., 2004; Babiloni et al., 2004; Babiloni et al., 2005; Blinowska et al., 2010; Kamiński et al., 1997; Silberstein, 2006; Srinivasan et al., 2007; Supp et al., 2007). Only recently, the fact that volume conduction has to be accounted for in EEG (as well as magnetoencephalography, MEG) based brain connectivity studies has been seriously acknowledged (Gómez-Herrero et al., 2008; Haufe, 2011; Haufe et al., 2010b; Nolte and Müller, 2010; Nolte et al., 2004, 2006, 2008; Schlögl and Supp, 2006; Schoffelen and Gross, 2009).

In this paper, we present results of a series of simulation experiments in which we systematically assessed common measures of effective (i.e. directed as opposed to undirected/functional) connectivity in terms of their ability to infer source interactions from pseudo-EEG recordings. Previous efforts in that direction have been carried out in Astolfi et al. (2007) and Silfverhuth et al. (2012). However, none of these studies considers biological noise (temporally and spatially correlated noise), which is present in any EEG recording as a result of the

* Corresponding author at: Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, Franklinstr. 28/29, 10587 Berlin, Germany. Fax: +49 3031478622.
E-mail address: stefan.haufe@tu-berlin.de (S. Haufe).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.09.036

brain's background activity, and has been reported to limit the detection accuracy of Granger-causal measures (Nolte et al., 2008, 2010) even if the signals-of-interest are directly observable. Even more severely, neither Astolfi et al. nor Silfverhuth et al. simulate spatial source mixing, which is another indisputable property of real EEG data caused by volume conduction in the head. Source mixing may affect both the correctness and the interpretability of the results, for which reason it is common practice to perform source demixing, e.g., using inverse source reconstruction approaches, prior to the connectivity estimation. Here, we also evaluate such various approaches.

We deliberately restrict ourselves to the analysis of simulated data, since we believe that any connectivity estimation should achieve reliable performance on appropriately designed artificial data before it can be applied to real data at all. Although our opinion does not seem to predominate in the field, there are authors who even emphasize "the importance of avoiding (that's right - avoiding) 'real' data". Those authors feel that "too many studies […] have been corrupted by the dogma that a methodology is not tested unless it is tested on real data", and they argue that a methodology "should not be permitted anywhere near real data, until it has been extensively tested in artificial data" (Theiler and Prichard, 1997). Many problems in neuroimaging (such as brain connectivity analysis) are inherently unsupervised, which means that the "ground truth" cannot be retrieved. In these cases, simulations are the only way to benchmark a method's ability to solve the task if theoretical results are not available, while a neuroscientific finding on real data that matches prior expectations should not be mistaken for a proof-of-concept of the method.

To our knowledge, our study is the first assessment of methods on standardized simulated EEG data that have been generated using a realistic model of volume conduction. The purpose of the paper, however, is not to provide an exhaustive numerical comparison of all these methods, but rather to highlight potential pitfalls of such analyses. In particular, we demonstrate that source mixing may easily mislead connectivity estimation depending on the type of measure used, while the success of inverse source reconstruction algorithms crucially depends on their ability to deal with the presence of multiple interdependent sources. Thus, even methodologies combining rather standard source reconstruction and connectivity estimation algorithms may not permit a correct neurophysiological interpretation. With this paper we want to bring such issues to the attention of the practitioners by visually demonstrating how certain properties of the data in combination with the characteristics of the various methods can potentially spoil connectivity analyses. We focus on a single minimalistic scenario, which however does ensure that the simulated data comprise some of the defining characteristics of real EEG data. Starting from this scenario, we subsequently apply certain modifications in order to demonstrate their effect on causal estimation.

As a methodological contribution we introduce the concepts of weak and strong asymmetries in multivariate time series in the context of causal modeling. Strong asymmetries arise only due to time-lagged interactions, while weak asymmetries may reflect general causal or non-causal interactions in the data. Weak asymmetries not related to time-lagged interaction frequently mislead certain types of connectivity analyses. As a remedy, we propose to use time-reversed data as surrogates in order to suppress the influence of weak data asymmetries when testing for causal influences.

We introduce relevant measures of effective connectivity as well as methods for inverse source reconstruction in Methods review section. Data asymmetries section introduces a novel methodology for improving connectivity estimates by distinguishing between weak and strong data asymmetries. Experiments section presents a series of four experiments, in which connectivity measures and pre-/postprocessing steps are evaluated within a common framework. The results are discussed in Discussion section, before we reach conclusions in Conclusions section.

Methods review

Measures of time-lagged effective connectivity

While there are multiple ways to define effective connectivity, the most widely accepted definition is based on a temporal argument: the cause must precede the effect. Algorithms implementing this definition are often subsumed under the term Granger-causal modeling (Valdes-Sosa et al., 2011), although the technique that is known as Granger causality is only one way to measure time-lagged influence.

Granger causality based on model errors

The linear multivariate autoregressive (MVAR) process (Brockwell and Davis, 1998) assumes that the present state of a time series can be approximated by a linear combination of its past P samples, i.e.,

$$\mathbf{x}(t) = \sum_{p=1}^{P} B(p)\mathbf{x}(t-p) + \varepsilon(t), \tag{1}$$

where the innovations $\varepsilon(t)$ are usually assumed to be independent and Gaussian distributed, and where $B(p)$ are matrices describing the time-delayed influences of $\mathbf{x}(t-\tau)$ on $\mathbf{x}(t)$. The off-diagonal parts $B_{i,j}(p)$, $i \neq j$ describe time-lagged influences between different time series x_i and x_j.

Granger causality (Granger, 1969) involves fitting an MVAR model for the full set $\mathbf{x}_{\{1,\ldots,M\}} = \mathbf{x}$, as well as for the reduced set $\mathbf{x}_{\{1,\ldots,M\}\setminus\{i\}}$ of available time series. Denoting the prediction errors of the full model by $\varepsilon^{\text{full}}$ and those of the reduced model by $\varepsilon^{\setminus i}$, the Granger score GC describing the influence of x_i on x_j is defined as the log-ratio of the mean-squared errors (MSE) of the two models with respect to x_j. That is,

$$\mathrm{GC}_{i,j} = \log\left(\frac{\sum_{t=P+1}^{T}\left[\varepsilon_j^{\text{full}}(t)\right]^2}{\sum_{t=P+1}^{T}\left[\varepsilon_j^{\setminus i}(t)\right]^2}\right). \tag{2}$$

By considering only two variables at a time (i.e. comparing a univariate and a bivariate AR model), a bivariate variant of Granger causality is obtained. There exists also nonlinear variants of Granger causality (Marinazzo et al., 2008; Vicente et al., 2011), which are not considered here.

Granger causality based on AR coefficients

A causal dependence of time series x_j on time series x_i in Granger's sense is sufficiently evidenced if any of the P coefficients $B_{i,j}(p)$ of the MVAR model fitted on the full set of available time series is significantly different from zero. This is assessed per frequency by partial directed coherence (PDC, Baccalá and Sameshima, 2001). Let $\tilde{B}(f)$ be the Fourier transform of the coefficients of a multivariate AR model fitted on all available time series, and $\tilde{\bar{B}}(f) = I_M - \tilde{B}(f)$, PDC is defined as

$$\mathrm{PDC}_{i,j}(f) = \left|\frac{\tilde{\bar{B}}_{i,j}(f)}{\sqrt{\tilde{\bar{\mathbf{B}}}_j^H(f)\tilde{\bar{\mathbf{B}}}_j(f)}}\right|^2, \tag{3}$$

which is the squared absolute value of the complex-valued quantity introduced in Baccalá and Sameshima (2001) measuring the strength of the information flow from x_i to x_j at frequency f. Partial directed coherence has similarities to so-called directed transfer function (DTF, Kamiński and Blinowska, 1991), with which it coincides in the bivariate case. Unlike GC, PDC and DTF are not independent of the scale of the data, and are usually applied on data that has been transformed to have zero mean and unit variance.

The phase-slope index

Another measure of interaction at a specific frequency is coherency (denoted by *CHY*) (Nunez et al., 1997, 1999), defined as

$$\mathrm{CHY}_{i,j}(f) = \frac{S_{i,j}(f)}{\left(S_{i,i}(f)S_{j,j}(f)\right)^{\frac{1}{2}}}, \tag{4}$$

where

$$S_{i,j}(f) = \frac{1}{K}\sum_{k=1}^{K} \tilde{x}_{i,k}(f)^{*}\tilde{x}_{j,k}(f) \tag{5}$$

denotes the empirical cross-spectrum. The absolute value of coherency is often used to quantify the strength of functional connections. However, this measure makes no distinction between instantaneous and truly time-delayed correlation and is hence dominated by effects of volume conduction when applied to EEG data. As a remedy, one can look at the imaginary part of coherency only (Nolte et al., 2004), which is systematically different from zero only for nonzero phase lags.

In general, a positive imaginary part of $CHY_{i,j}(f)$ suggests that z_i is earlier than x_j. However, for oscillations with period lengths of the order of the delay, "earlier" and "later" are ambiguous due to the periodicity of the processes. In order to resolve this ambiguity, the information contained in nonzero phase lags can be aggregated within a frequency band-of-interest, which is the idea of the phase-slope index (PSI, Nolte et al., 2008). Denoting by $\mathcal{F}$ a contiguous set of frequencies and by δf the frequency resolution, PSI is defined (disregarding the standardization proposed in Nolte et al. (2008) as

$$\Psi_{i,j} = \Im\left(\sum_{f\in\mathcal{F}} \mathrm{CHY}^{*}_{i,j}\mathrm{CHY}_{i,j}(f+\delta f)\right). \tag{6}$$

From the Hermitianity of the cross-spectral matrices it follows that PSI is antisymmetric, i.e., $\Psi_{i,j} = -\Psi_{j,i}$. Moreover, due to the normalization of the cross-spectrum, PSI is invariant with respect to rescaling of the data.

Statistical testing

In order to infer the presence of causal interactions according to GC, PDC or PSI, the connectivity scores obtained from these measures have to be tested against a suitable null hypothesis related to the absence of any interaction. Since PSI is a signed quantity, the null hypothesis is simply that the scores are drawn from a zero mean distribution. Both GC and PDC are strictly positive quantities, for which the formulation of a null hypothesis is more involved. A simple alternative is to consider tnet interactions, which amounts to testing the predominant direction of information flow for each pair of time series. For a connectivity measure M the net scores $M^{\mathrm{net}}_{ij} = M_{i,j} - M_{j,i}$ are obtained by antisymmetrizing the connectivity matrix M. As for PSI scores, the null hypothesis for net connectivity scores is simply the presence of a zero mean distribution. Antisymmetrization of Granger-causal measures is occasionally used in the literature (e.g. Roebroeck et al., 2005).

If antisymmetrization is to be avoided, an analytic procedure using *F*-tests can be used for Granger causality to assess whether the full MVAR model adds significant explanatory power compared to the reduced model (e.g. Seth, 2010). However, there is no similar procedure for PDC. Therefore, one has to resort to using so-called surrogate data in order to establish a null distribution. A common choice for surrogates is the original data, in which the temporal order has been randomly permuted separately for each time series (e.g., Babiloni et al., 2005; Kamiński et al., 2001). This procedure is often considered an application of the "method of surrogate data" introduced by Theiler et al. (1992), although the original method proposed by Theiler et al. differs from that approach in certain aspects.

Methods for inverse source reconstruction

The most general generative model of EEG data is given by

$$\mathbf{x}(t) = A\mathbf{s}(t) + \eta(t), \tag{7}$$

where $\mathbf{x}(t)\in\mathbb{R}^{M}$ is the signal measured at M EEG electrodes at time t, $\mathbf{s}(t)\in\mathbb{R}^{K}$ is the activity of K brain sources at time t and $A\in\mathbb{R}^{M\times K}$ is a matrix representing instantaneous source mixing due to volume conduction. The noise term $\eta(t)$ comprises uncorrelated measurement (sensor) noise as well as correlated noise, which could be due to non-task-related background activity or artifacts. Notably, EEG activity of cerebral origin is always instantaneously correlated due to volume conduction, which is modeled here explicitly using the matrix A. In reality, xtitA describes a physical process, namely the propagation of the brain electric currents from the source regions to the EEG electrodes. Given the geometry and electrical conductivities of the various tissues in the head, the columns of A (corresponding to the mixing patterns of idealized brain sources) can be computed. This step is called forward modeling. Inverse source reconstruction is concerned with the estimation of $\mathbf{s}$ given $\mathbf{x}$ and A, which amounts to solving the so-called electromagnetic inverse problem (Baillet et al., 2001; Nunez and Srinivasan, 2006). In distributed inverse imaging, dipolar sources are modeled at many locations within the brain, and the activity at all those locations is estimated jointly. To overcome the ambiguity of the solution, it is crucial to constrain the solution to be consistent with prior domain knowledge. Depending on the type of constraint, the solution can be linear or nonlinear in the observations. A second class of inverse methods are beamformers. Here, the activity at each voxel is estimated using a linear spatial filter that is optimized for that voxel.

The weighted minimum-norm estimate

The weighted minimum-norm (WMN) source estimate (Hämäläinen and Ilmoniemi, 1994; Ioannides et al., 1990; Jeffs et al., 1987) is the source distribution with minimal power that explains the EEG measurement. As it is typical for linear methods (Grave de Peralta-Menendez and Gonzalez-Andino, 1998; Haufe et al., 2008; Matsuura and Okabe, 1995), the WMN solution tends to be very blurred, and may not resolve multiple sources. To counteract a location bias in the estimation, we consider a depth-compensation as proposed in Haufe et al. (2008).

Sparse basis field expansions (S-FLEX)

Inverse source reconstruction via sparse basis field expansions (S-FLEX, Haufe et al., 2009, 2011) achieves a compromise between smoothness and focality of the source current distribution and is thereby able to model the simultaneous occurrence of multiple extended sources of different sizes and shapes. The idea of S-FLEX is to expand the current density at time t as a linear combination of (potentially many) spatial basis fields, which are defined as the outer products of scalar Gaussian basis functions and 3-dimensional coefficient vectors. The assumption made by S-FLEX is that the current density can be well approximated by a small number of basis fields, which is encoded by means of an $\ell_{1,2}$-norm penalty on the coefficients. The $3T$ coefficients related to a single basis function are tied under a common ℓ_2-norm and can only be pruned to zero at the same time. Thus, the selection of basis functions which contribute coherently to the entire EEG time series is facilitated.

Linearly constrained minimum-variance (LCMV) beamforming

The idea of beamforming is to find a spatial projection of the observed signal, such that signals from a specific location in the brain are preserved, while contributions from all other locations as well as noise contributions are maximally suppressed. The linearly constrained minimum-variance (LCMV) spatial filter (Van Veen et al., 1997) does that by minimizing the variance of the filtered signal subject

to a unit-gain constraint (that is, the product of filter and forward matrix at the desired location is enforced to be unity).

Data asymmetries

Many measures of effective connectivity are based on the principle that the cause precedes the effect. However, it would be misleading to assume that temporal ordering is necessarily the dominant factor which affects the estimation of causal relationships. Rather, most methods are based on general asymmetries between two (or more) signals out of which the temporal order is just one specific feature. Other asymmetries, like different signal-to-noise ratios, different overall power or spectral details may in general also affect causal estimates depending on which method is used.

We here propose to distinguish two kinds of asymmetries. We call the first type "strong asymmetries" defined as asymmetries in the relation between two (or more) signals like the temporal ordering. The second type is called "weak asymmetry" and denotes different univariate properties as given, e.g., by the spectral densities. Weak asymmetries can hence be detected from two signals without estimating any functional relationship between them whereas a strong asymmetry is a property of that functional relationship.

We restrict ourselves in the following to the discussion of stationary and Gaussian distributed data. Let $x_j(t)$ be the signal in channel j at time t. Then the statistical properties are completely defined by the cross-covariance matrices

$$C(p) = \left\langle (\mathbf{x}(t) - \hat{\mu}_{\mathbf{x}})(\mathbf{x}(t-p) - \hat{\mu}_{\mathbf{x}})^{\top} \right\rangle, \tag{8}$$

where $\langle \cdot \rangle$ denotes expectation. The process is now said to contain a strong asymmetry if for some i, j and some p it is found that $C_{i,j}(p) \neq C_{j,i}(p)$, i.e. $C(p)$ is asymmetric for at least one p. The process is said to contain a weak asymmetry if for some i, j and some p it is found that $C_{i,i}(p) \neq C_{j,j}(p)$, i.e. the diagonals are not all equal.

Weak asymmetries can be detected more robustly than strong asymmetries, but can also be considered weaker evidence for causal relations. In particular, they arise inevitably in real EEG data due to volume conduction, even if the underlying sources are statistically independent. In this case all cross-covariances are weighted sums of the auto-covariances of the sources. Since auto-covariances are always symmetric functions of the delay p and since generally $C(-p) = C^{\top}(p)$ it follows that $C(p) = C^{\top}(p)$ for mixtures of independent sources (Nolte et al., 2006). Hence, such mixtures can only contain weak asymmetries but not strong ones. On the other hand, two sources having a time-delayed influence on another, e.g., through a bivariate AR model, always do exhibit a strong asymmetry, which can still be observed after linearly mapping the source activity to EEG sensor space. However, this mixing additionally introduces weak asymmetries, which may mask the detection of the strong asymmetry depending on the method used.

For methods which are sensitive to both weak and strong asymmetries it is in general difficult to tell on what property of the data an estimated causal relation is based. The ambiguity can be resolved by testing against appropriately designed surrogate data. Recall that in permutation testing, the aim of using surrogate data is to obtain a null distribution related to the lack of any structure in the data. The underlying assumption here is that any significant deviation of a connectivity score from the null distribution is indeed related to causal interaction between the corresponding time series. However, this assumption does not hold in the presence of weak data asymmetries.

In a broader sense the idea of surrogate data is to create data which differ from the data under study in exactly the property that is being analyzed, but agree with them in as many as possible other aspects. In the context of the discovery of time-delayed interactions it is therefore desirable to design surrogate data such that they share weak asymmetries with the original data but lack the strong asymmetries. To this end, we here suggest to use time-reversed signals. This corresponds to the general intuitive idea that, if temporal order is crucial to tell a driver from recipient, the result can be expected to be reversed if the temporal order is reversed. The mathematical basis for this is the simple observation that the cross-covariance for the time inverted signals, say $\tilde{C}(p)$, is given as

$$\tilde{C}(p) = C(-p) = C^{\top}(p), \tag{9}$$

implying that our approach improves upon permutation testing in two aspects. First, all weak asymmetries are exactly preserved. Moreover, strong asymmetries are not only eliminated, but exactly inverted. This leads to greater statistical power of time inversion testing compared to permutation testing in the presence of weak asymmetries, which has been evidenced in a simulation setting comprising mixed noise (Haufe et al., 2012). Haufe et al. (2012) further outline a strategy for rejecting false positives based on the evaluation of a measure that quantifies the ratio of weak and strong asymmetries.

Note that as a result of the Hermitianity of the cross-spectrum it not only follows that PSI is antisymmetric, but also that this measure exactly flips its sign when being applied to time-reversed data. Hence, time inversion testing is formally equivalent to the standard approach of testing PSI for nonzero mean without using surrogate data.

Experiments

In the following we present a series of four experiments designed to assess state-of-the-art approaches to EEG-based brain effective connectivity analysis. We focus on the most-simple model that includes source interaction, namely a two-dipole model with linear dynamics and a time-delayed linear influence of one source on the other. In the first experiment, measures of effective connectivity are applied to simulated unmixed source time series. The second experiment deals with simulated pseudo-EEG comprising realistic effects of volume conduction and noise. The third experiment demonstrates the influence of the signal-to-noise ratio and the choice of the reference electrode. The last experiment assesses inverse source reconstruction techniques regarding their aptitude as a preprocessing for source connectivity analysis.

Experiment 1: two interacting sources

We simulate a system of two simulated brain sources $s_{1/2}(t)$ following a stable bivariate AR process of order $P = 5$, from which we generate $T = 10000$ samples. The AR coefficients are drawn independently from $\mathcal{N} \sim (0, 0.01)$. By setting the off-diagonal coefficients $B_{1,2}(p), 1 \leq p \leq P$ to zero while all other coefficients remain nonzero, unidirectional flow from s_1 to s_2 is modeled. The innovations $\varepsilon(t)$ of the source AR process are drawn from the univariate standard Normal distribution. We perform 100 repetitions of the experiment. For each repetition, a dataset comprising distinct innovation terms and source AR coefficients is generated. Note that we consider the noiseless case here, while the influence of noise has been studied in detail in the literature (Haufe et al., 2012; Nolte et al., 2008, 2010). Moreover, there is no source or noise mixing due to volume conduction in this first simulation. Thus, it mainly serves as a proof-of-concept for effective connectivity measures under ideal (but unrealistic) conditions, and a baseline for source demixing algorithms.

We apply Granger causality, partial directed coherence and the phase-slope index to normalized source time series, in which the activity in each channel is transformed to have zero mean and unit variance. The phase-slope index is computed using an implementation provided by Nolte et al. (2008),[1] while the "Granger Causal

[1] http://ml.cs.tu-berlin.de/causality.

Connectivity Analysis" toolbox (Seth, 2010)[2] is used to compute GC, and the MVARICA toolbox (Gómez-Herrero et al., 2008) is used to compute PDC. The AR model underlying the computation of PDC is estimated using the ARFIT package (Neumaier and Schneider, 2001).[3] For PDC, we average the connectivity scores across all frequency bins to obtain a global measure of interaction.

We consider three statistical testing strategies, which are applicable to all connectivity measures. In antisymmetrization testing, two-sided *t*-tests are used to assess whether net connectivity scores are significantly different from zero across repetitions. In permutation testing, the same *t*-test is performed on the differences of scores obtained on original and permuted data. Finally, time-reversed data are used as surrogates in time inversion testing instead of temporally permuted data. Note that for PSI time inversion and antisymmetrization testing are formally equivalent to testing the original connectivity scores for nonzero mean.

As a postprocessing to each statistical test, the resulting *t*-scores are converted into *z*-scores. This makes it easier to compare the results of different experiments, in which the degrees of freedom of the t-distribution may be different. The transformation is performed per $z = \mathrm{cdf}_z^{-1}(\mathrm{cdf}_t(t,\nu))$, where cdf_t is the cumulative distribution function (cdf) of Students *t*-distribution with ν degrees of freedom Gosset (1908) and cdf_z^{-1} is the inverse of the cdf of the univariate standard normal distribution. Throughout the paper, connections are reported as significant if the *p*-value associated with a *z*-score (corrected for the testing of multiple entries of a connectivity matrix using the Bonferroni method) falls below 0.05. In this experiment, this is the case for *z*-scores with absolute values greater than 1.96.

Experiment 2: simulated EEG

We next consider simulated EEG data comprising effects of volume conduction and noise. The artificial EEG signal is generated according to

$$\mathbf{x}(t) = \frac{(1-\gamma)\left(\frac{1}{2}\sum_{i=1}^{2}\frac{\mathbf{a}_i s_i(t)}{\|(S^\top)_i\|_2}\right)}{\|\mathbf{vec}(S)\|_2} + \gamma\frac{\eta(t)}{\|\mathbf{vec}(E)\|_2}, \tag{10}$$

where $\mathbf{x}$ is the EEG signal, $s_{1/2}$ are the source time series, $\mathbf{a}_{1/2}$ are the spread patterns of the dipolar sources evaluated at 59 EEG electrode at standard positions as defined in the extended international 10–20 system (Chatrian et al., 1988), η is noise and $\gamma, 0 \le \gamma \le 1$ is a parameter that adjusts the signal-to-noise ratio (SNR). Moreover, $E = (\eta(1),\ldots,\eta(T))$ and $S = (\mathbf{s}(1),\ldots,\mathbf{s}(T))$. The normalizing terms $\|(S^\top)_i\|_2$ are used to equalize the power of driver and receiver time series, while the normalization by $\|\mathbf{vec}(S)\|_2$ and $\|\mathbf{vec}(E)\|_2$, respectively, allows precise adjustment of the SNR by means of γ. Here, we set $\gamma = 0.5$, corresponding to a balanced SNR.

We use a head model with realistically shaped brain, skull and skin shells (Holmes et al., 1998), and assume a nose reference. The source dipoles are placed in the left and right hemispheres of the brain, 3 cm below C3 (s_1) and C4 (s_2), respectively. The positions of these dipoles are marked by circles in Fig. 1. The current moment vectors of both dipoles are tangentially oriented, leading to bipolar field patterns. The sources are designed to reproduce field patterns of N20 event-related potentials observed after median nerve stimulation at the hands in real EEG (Haufe et al., 2008). Similar patterns are also frequently extracted by common spatial patterns (CSP) analysis of mu-rhythm oscillations related to idling of the hand motor system Blankertz et al. (2008). Thus, our scenario resembles sources in the left and right sensorimotor cortices, where information flows from the left to the right sensorimotor cortex. The field patterns $\mathbf{a}_{1/2}$ describing the spread of the source dipoles to the EEG sensors are computed according to Nolte and Dassios (2005). Both dipolar sources and the corresponding field patterns are depicted in Fig. 1.

The noise terms $\eta(t) = \eta^{\mathrm{sensor}} + A^{\mathrm{biol.}}\eta^{\mathrm{biol.}}$ are composed *of sensor noise* η^{sensor}, which is drawn independently for each sensor and time point from a Gaussian distribution. Additionally, we include ten sources of biological noise, the time courses $\eta_1^{\mathrm{biol.}}(t),\ldots,\eta_{10}^{\mathrm{biol.}}(t)$ of which are generated using random stable univariate AR models of order $P^{\mathrm{biol.}} = 10$, and are mixed by a spread matrix $A^{\mathrm{biol.}}$ representing the spread of ten randomly placed dipoles with random current moment vectors. Sensor and biological noises are scaled to contribute equally to the overall noise η.

We generate artificial EEG datasets using the source time series introduced in Experiment 1. The underlying source dipole locations and orientations are left constant across experiments, while distinct innovations of the noise AR processes, as well as distinct noise AR coefficients and noise dipole locations are drawn. Connectivity is assessed between normalized sensor-space time series, which is common practice in parts of the literature. While PSI is by definition a bivariate measure, GC and PDC are usually estimated using all available time series in order to reduce the deceptive influence of causal confounders common to more than one time series. Here we test the effect of the number of variables included in the MVAR estimations underlying the calculation of GC and PDC. That is, apart from the bivariate measures, we also assess their multivariate counterparts which use either all 59 EEG channels or just the subset of 19 channels defined in the international 10–20 system within the MVAR estimation. The result of each connectivity analysis is either a 19×19 or a 59×59 matrix of connectivity scores related to pairs of sensors. Statistical significance of each entry of these matrices is assessed using antisymmetrization, permutation, and time inversion testing.

Experiment 3: influence of reference electrode and SNR

Sensor-space connectivity patterns necessarily depend on the spatial spread patterns of the underlying interacting electrical brain sources, which are functions of their location and spatial orientation as described by the forward model. Since these parameters are unknown in practice, the interpretability of sensor-space connectivity maps is naturally limited. An even more indirect factor influencing sensor-space results is the dependence of field spread patterns on the choice of the reference electrode. The purpose of this experiment is to demonstrate that a change of reference can have an impact on sensor-space connectivity estimation. To this end, we re-reference the data used in Experiment 2 by subtracting the activity of either the TP9 or the TP10 electrode. Since these are the electrodes closest to the ears, doing so amounts to simulating the reference electrode to be located either at the left or the right mastoid. Both are standard choices in practice. To restore the full rank of the resulting data matrices (the post-hoc re-referenciation decreases the rank by one), a small amount of spatially and temporally independent Gaussian distributed noise is added. In addition to re-referenciation, we here also study the influence of the signal-to-noise ratio adjusted by γ. We present results for two additional SNRs of $\gamma = 0.25$ and $\gamma = 0.75$ using the original nose-referenced dataset. We restrict ourselves here to the application of PSI.

Experiment 4: inverse source reconstruction

A number of studies have investigated the (effective or functional) connectivity of source estimates obtained from linear inverse imaging (Astolfi et al., 2006; Babiloni et al., 2005; Gow et al., 2008) or beamforming (Brookes et al., 2011; Martino et al., 2011; Wibral et al., 2011). Here, we compare distributed inverse imaging according to WMN and S-FLEX as preprocessings for EEG-based source connectivity

[2] http://www.informatics.sussex.ac.uk/users/anils/aks_code.htm.
[3] http://www.gps.caltech.edu/~tapio/arfit.

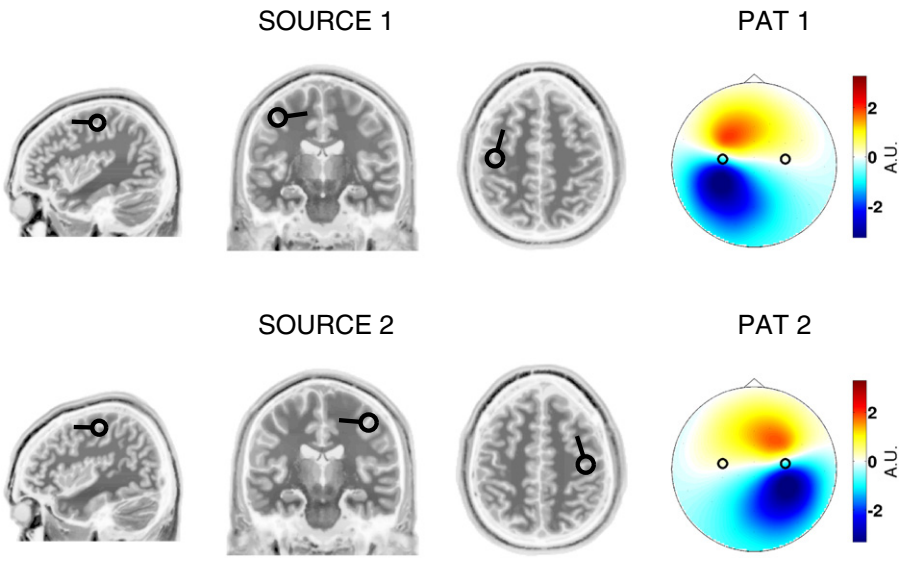

Fig. 1. Two simulated dipolar sources (SOURCE 1/2) and their corresponding EEG field patterns (PAT 1/2). Sources are placed 3 cm below the C3 (left) and C4 (right) electrodes and are oriented tangentially to the scalp.

analysis with beamforming via LCMV. Sources are reconstructed in the same head model in which the data are simulated. To this end the interior of the whole brain shell is partitioned into $N = 2142$ voxels of 10 mm side length. In the center of each voxel a dipolar source is modeled, the current moment vector of which is estimated for each time point. Note that this source space might be considered too rich, since the EEG signal is believed to mainly originate from cortical gray matter. However, in our simulated scenario, this detail plays a minor role regarding the evaluation of inverse methods and source connectivity estimation in general.

We use our own implementations of WMN, S-FLEX and LCMV for transforming the pseudo-EEG measurements into source time series. The regularization parameter of WMN is selected using 5-fold cross-validation, which is implemented by splitting the set of electrodes randomly into five parts. The same regularization parameter is selected for all time indices t. Since the WMN solution is linear in the data, the source distribution is easily acquired even for long time series data by means of a matrix multiplication. This is different for S-FLEX, which requires estimating all source variables (the coefficients of the basis field expansion related to all measurements) jointly using nonlinear optimization. Doing so for 10 000 samples is prohibitive due to excessive memory requirements, for which reason a two-step procedure is adopted, which restricts the number of variables involved in each step. In the first step, S-FLEX is applied to 100 randomly selected samples. Using only the basis functions with nonzero corresponding coefficients, the second estimation is performed for all time samples. We apply S-FLEX using Gaussian basis functions with spatial standard deviations $\varsigma_1 = 0.75$, $\varsigma_2 = 1$ and $\varsigma_3 = 1.25$. The regularization parameter in both steps is adjusted such that the S-FLEX solution achieves the same goodness-of-fit as the corresponding cross-validated WMN estimate. In contrast to both distributed inverses, the LCMV beamformer estimates much less parameters than samples, and regularization is less of an issue. To ensure numerical stability of the inversion of the data covariance matrix C, we here use a slightly regularized estimate $\tilde{C} = C/||C|| + 0.01I/||I||$, where I is the identity matrix. Moreover, to counteract potential locations biases, we normalize the source power map with an estimate of the noise source power using the identity matrix as an approximation for the sensor-space noise covariance matrix (Van Veen et al., 1997). The quality of the source reconstructions is measured using the earth mover's distance (EMD) metric (Rubner et al., 2000), by which it is possible to objectively compare the simulated two-dipole source configuration with the estimated current densities (Haufe et al., 2008). Additionally, the largest two local maxima of the current amplitude distribution are sought for each method, and the average distance to the corresponding source dipoles is calculated.

In order to reduce the complexity of the source connectivity analysis, we define regions-of-interest (ROIs) similar to Babiloni et al. (2005) and Astolfi et al. (2006). Since we are only concerned with simulated data here, these regions are not defined anatomically but by partitioning the source space according to the closest (in the Euclidean sense) of the 19 EEG electrodes defined in the 10–20 system. This enables us to present the results in the same way as sensor-space results. Dipoles that are further than 5 cm away from any electrode are not assigned to any region. The source activity within each ROI is summed separately for each spatial dimension to yield a $3 \cdot 19 = 57$-dimensional time series. Since this time series might be singular a small amount of spatially and temporally independent Gaussian distributed noise is added in order to establish full rank. The resulting time series are normalized and subjected to effective connectivity analysis using PSI as well as four variants of GC employing either bivariate or 57-dimensional MVAR modeling in combination with either permutation or time inversion statistical testing.

The application of connectivity measures to source time series in each ROI yields $3 \cdot 19 \times 3 \cdot 19$ connectivity matrices consisting of 3×3 blocks $R(i,j)$, which describe the interactions between the i-th and the j-th ROI in all three spatial dimensions. We calculate the total flow from the i-th to the j-th voxel by summing over all entries of $R(i,j)$. This yields a 19×19 matrix, the entries of which are subjected to statistical significance testing.

Results

Two interacting sources

In the noiseless and unmixed scenario, all three connectivity measures correctly indicate highly significant information flow from s_1 to s_2 with $z_{1,2} > 10$ regardless of the testing procedure used. If permutation testing is used, there is no significant flow from s_2 to s_1 ($|z_{2,1}| < 2$) for GC and PDC, while there is highly significant ($z_{2,1} < 10$) negative flow from s_2 to s_1 for PSI owing to the intrinsic antisymmetry of this measure. Both results are correct. A significant negative flow from s_2

to s_1 is also obtained if antisymmetrization testing is used for GC and PDC. Moreover, we observe significant ($z_{2,1}<10$) negative flow from s_2 to s_1 for all three measures in combination with time inversion testing.

Simulated EEG

The sensor-space connectivity matrices estimated from noisy pseudo-EEG data by GC, PDC and PSI in combination with three different variants of statistical testing are visualized in Fig. 2 as head-in-head plots (Nolte et al., 2004, 2008). Each head-in-head plot consists of 19 small circles representing electrodes on the human scalp. These are arranged within one large scalp plot according to the positions of the 19 electrodes of the 10–20 electrode placement system. Each of the small scalp plots shows the estimated interaction of the corresponding electrode to the other 18 electrodes defined in the 10–20 system (an exception is made for GC and PDC estimates calculated using 59-dimensional MVAR models, for which the connectivity to all 58 remaining electrodes is shown). Red and yellow colors ($z>0$) stand for information outflow and blue and cyan colors ($z<0$) stand for information inflow. The Bonferroni factor is the number of electrode pairs visualized, which is $19\cdot(18)/2=171$ ($19\cdot(58)/2=551$ for the 59-dimensional case). The corresponding Bonferroni-corrected significance threshold of $z=3.62$ (3.91 for the 59-dimensional case) is indicated by a thin black line in the colorbar, while the *z*-score corresponding to an uncorrected *p*-value ($z=1.96$) is indicated by a thick black line.

The phase-slope index correctly reveals general significant information flow from the left to the right hemisphere regardless of the statistical testing procedure used. The observed connectivity matrices resemble the true field patterns of the underlying sources in that information flow is estimated to pass predominantly from electrodes in the right hemisphere to electrodes in the left hemisphere. The most significant flow passes from those regions in which the driving source is most strongly expressed to those in which the receiving source is most strongly expressed.

The results according to GC and PDC greatly vary depending on the statistical testing strategy used and the number of variables included in the estimation of the underlying MVAR models. Using the bivariate variants of both methods in combination with permutation testing, almost all causal connections (in both directions) exceed the Bonferroni-corrected significance threshold. For both measures the number of significant connections decreases gradually the more variables are included in the MVAR estimation. The remaining few connections in the 59-dimensional multivariate case, however, do not reflect the simulated interhemispheric information flow. Rather, interactions are estimated predominantly between neighboring channels in both hemispheres. Note that the same situation occurs, if analytic testing by means of *F*-tests is performed for GC.

Using time inversion testing, the bivariate and 19-dimensional multivariate variants of GC and PDC correctly estimate general flow from the left to the right hemisphere similarly to PSI. The number of significant connections is larger in the bivariate case. In the 59-dimensional multivariate case, only very few significant connections are found. These do not give rise to a neurophysiological interpretation. Antisymmetrization testing for the bivariate variants of GC and PDC yields a similar result as time inversion testing in that general significant net flow from the left to the right hemisphere is observed. However, in combination with the multivariate variants of GC and PDC, antisymmetrization testing yields a qualitatively different result. Here, significant flow is observed to pass predominantly from higher-SNR channels to lower-SNR channels (where the SNR of a single channel is defined as the combined strength of contributions of the two sources in that channel) regardless of the laterality of the involved electrodes. These spurious connections are slightly weaker (less significant) for GC than for PDC, and remain significant only for the latter method in the 59-dimensional multivariate case.

Influence of reference electrode and SNR

Fig. 3 depicts sensor-space connectivity maps resulting from the analysis of data generated using different simulated electrical references as well as different signal-to-noise ratios. Depending on the reference, the estimated connectivity varies substantially. Most notably, the left-to-right flow observed on nose-referenced data is not predominant anymore when the data is in TP9 or TP10 reference. Rather, we also observe information flow from right frontal to right parietal areas in TP9-referenced data, and flow from left parietal to left frontal areas in TP10-referenced data, which is in conflict with the simulated information flow passing exclusively from the left to the right hemisphere.

In the low SNR regime ($\gamma=0.25$), we observe much less significant connections than previously. Importantly, the only significant connections are estimated between posterior channels. Analogously, for $\gamma=0.75$, larger areas than previously show significant interaction.

Inverse source reconstruction

Results of inverse source reconstructions are depicted in the right part of Fig. 4 as heat maps showing estimated source current amplitudes averaged over time instants and repetitions (and divided by the estimated source current amplitude of white sensor noise in case of LCMV). The plot is overlaid with arrows representing the simulated interacting dipolar sources. The source activity estimated by WMN is spread over the entire brain. The true sources are not well separated, as the source amplitude appears to have only one center of activity located in between the two simulated interacting dipoles. The amplitude distribution according to LCMV is similarly spread-out as the WMN estimate. However, there are two distinguished local maxima close to the simulated interacting sources. The S-FLEX solution is more focal than the other two, exhibiting two dominant local patches of activity located near (but slightly deeper than) the two simulated interacting dipoles. In the right part of Fig. 4, the amplitude of the summed estimated source signals in each ROI is depicted as a heat scalp map. All three methods correctly exhibit two local maxima at the ROIs below the C3 and C4 electrodes, while there is more energy in the remaining ROIs for WMN and LCMV compared to S-FLEX.

The qualitative assessment of the reconstruction performance of the three methods is supported by the numerical evaluation of EMD scores and distances between the locations of the true dipoles and the local maxima of the estimated source current amplitude distributions, which are presented in Tables 1 and 2. Source imaging according to S-FLEX yields better reconstruction of the two simulated dipoles than WMN and LCMV in all three SNR regimes, as evidenced by significantly lower EMD scores. The distance between the peaks of the source amplitude distribution and the true dipole location is smaller for LCMV compared to S-FLEX in low and medium SNR ranges, and is as small as possible (given the discreteness of the source grid) for high SNR. In the high-SNR regime, S-FLEX also achieves minimum peak-to-dipole distances.

The estimated effective connectivity between ROIs is depicted in Fig. 5 using head-in-head plots. Note that the interpretation of these plots here is much easier as in previous experiments, since the depictions do not represent interactions between electrodes but interactions between source space regions-of-interest below these electrodes. Since we simulated the true interacting dipoles to lie exactly below C3 and C4, it is sufficient to assess whether (only) the flow from (the region below) C3 to (the region below) C4 is present when evaluating source connectivity head-in-head plots. Since information flow is only meaningful if it is estimated between active regions, the information about the strength of the source activity (as visualized in the right part of Fig. 4) is encoded in the visualization by means of the alpha (transparency) value. Here, the ROI with maximal strength is drawn with full opacity, while the

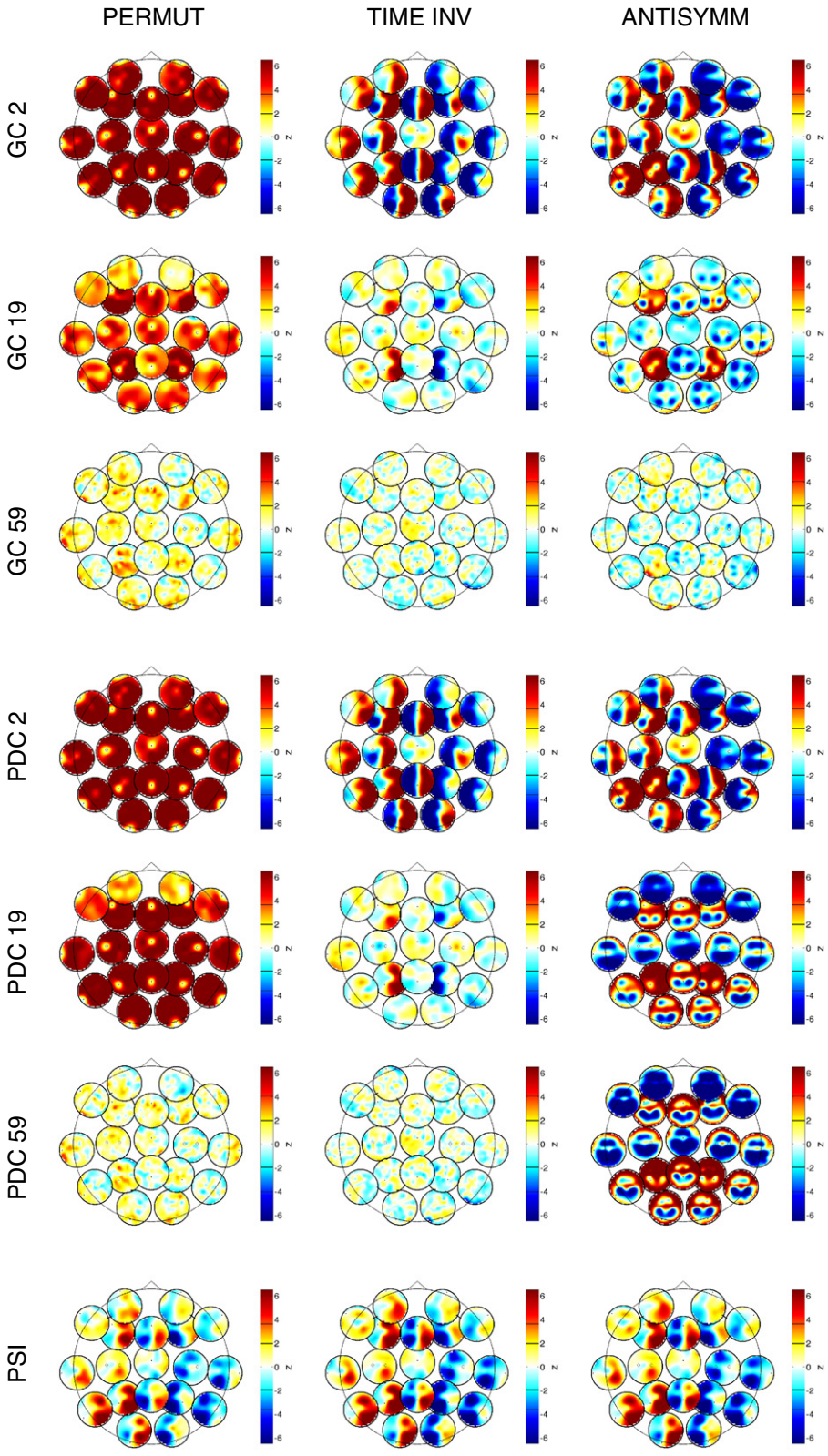

Fig. 2. Comparison of effective connectivity between simulated EEG sensor measurements as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) using permutation testing (PERMUT), time inversion testing (TIME INV) and antisymmetrization testing (ANTISYMM). Granger-causal measures are calculated using bivariate MVAR modeling (GC/PDC 2), as well as multivariate modeling of either 19 or 59 channels (GC/PDC 19/59), respectively. Information flow from the left (below C3) to the right (below C4) source is modeled by means of a bivariate AR model. The simulated EEG is superimposed by non-interacting biological and sensor noise (SNR = 1). The significance of estimated interactions is measured in terms of z-scores and visualized as head-in-head plots, where red and yellow colors ($z > 0$) stand for information outflow and blue and cyan colors ($z < 0$) stand for information inflow. The Bonferroni-corrected significance level is indicated by a thin black line in the colorbar, while the uncorrected significance level is indicated by a thick black line.

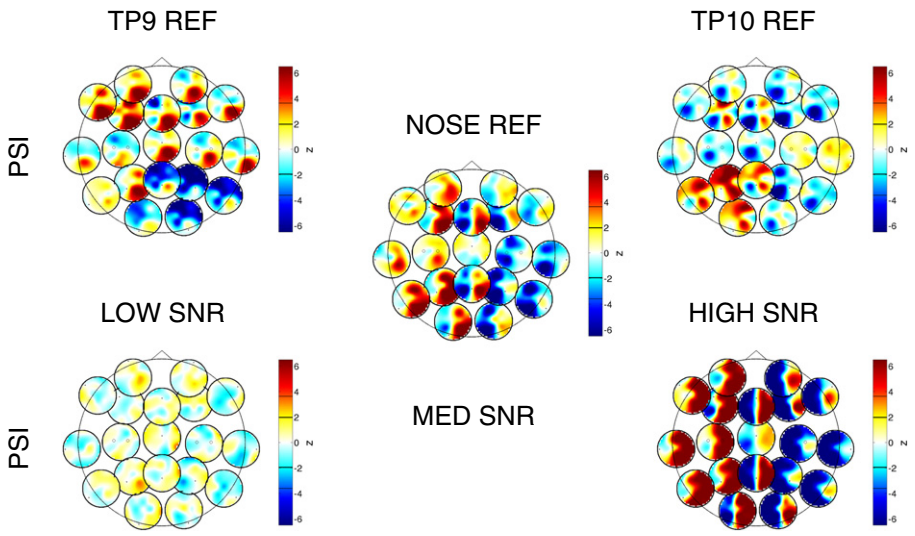

Fig. 3. Comparison of effective connectivity between EEG sensor according to the phase-slope index (PSI) for three different choices of the reference electrode, as well as for three different signal-to-noise ratios.

ROI with minimal strength is drawn with full transparency. Alpha values for ROIs in between are assigned using a monotonous sigmoidal nonlinearity.

The application of PSI as well as bivariate GC in combination with time inversion testing on estimated source time courses reveals the underlying interhemispheric left-to-right flow for all three inverse source reconstruction methods. In combination with S-FLEX, the interaction is predominantly observed between the true generating brain areas, namely the regions below C3 and C4. Beamforming according to LCMV estimates broader active areas, in which some of the activity of the simulated interacting sources seems to leak. Consequently, both connectivity measures do not only detect flow from

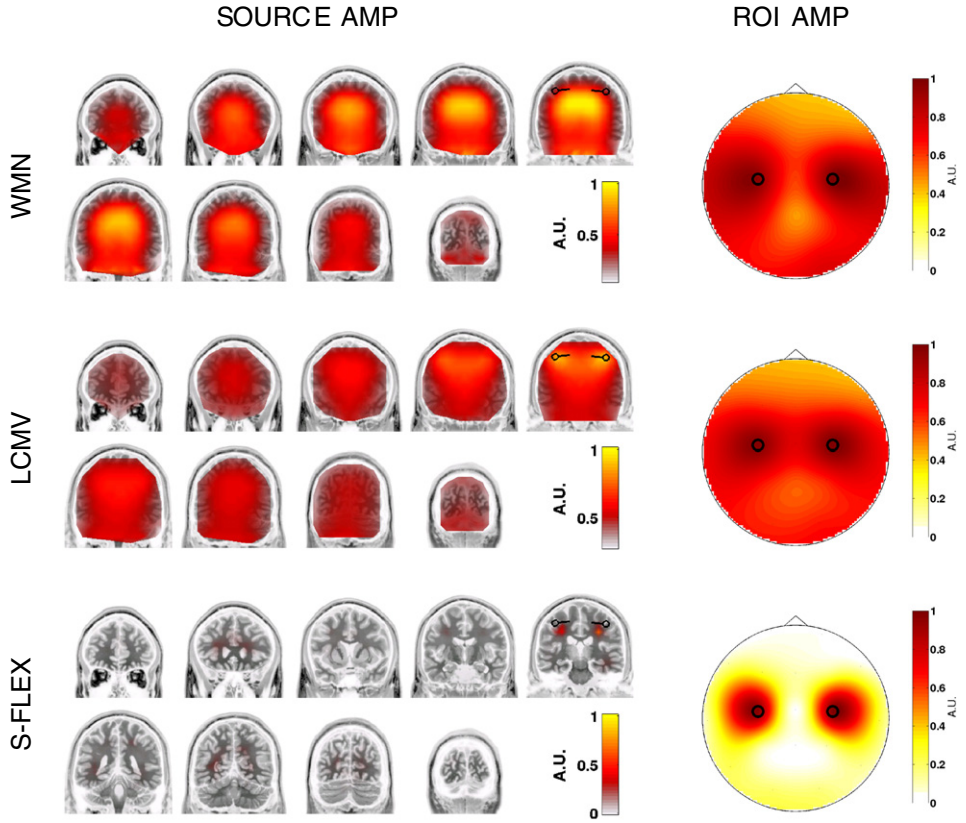

Fig. 4. Source amplitude distributions obtained by weighted minimum-norm (WMN) and sparse basis field expansions (S-FLEX) distributed inverse imaging, as well as linearly constrained minimum-variance (LCMV) beamforming. SOURCE AMP: voxelwise amplitudes. ROI AMP: amplitudes of summed activity in regions-of-interest defined based on the closest EEG electrode.

Table 1
Earth mover's distance (mean over 100 repetitions ± standard error of the mean) between the two simulated dipoles and the source current amplitude distributions as estimated in single repetitions by means of the weighted minimum-norm (WMN) and sparse basis field expansions (S-FLEX) inverse imaging estimates, as well as the linearly constrained minimum-variance (LCMV) beamforming estimate for three different signal-to-noise ratios. Lower EMD scores indicate better source reconstruction. Entries marked in bold indicate superior performance.

	WMN	LCMV	S-FLEX
$\gamma = 0.25$	56.0 ± 0.4	51.2 ± 0.3	32.4 ± 0.8
$\gamma = 0.50$	47.8 ± 0.2	46.5 ± 0.3	15.2 ± 0.5
$\gamma = 0.75$	36.4 ± 0.2	42.7 ± 0.2	4.5 ± 0.2

below C3 to below C4, but also between the areas below electrodes neighboring C3 and C4. For WMN inverse source reconstruction, almost all ROIs are estimated to significantly participate in the general left-to-right interaction.

Granger causality using 57-dimensional MVAR modeling and time inversion testing does not yield any significant information flow between ROIs for WMN and LCMV inverse source reconstruction preprocessing. However, it does correctly reveal significant information flow from the area below C3 to the area below C4 for S-FLEX. However, these results are not as significant as for bivariate GC using the same testing procedure. The application of GC using bivariate MVAR modeling and permutation testing does not yield correct results regardless of the type inverse source reconstruction preprocessing used, and thus does not provide an improvement over sensor-space results. As in sensor-space, significant bidirectional information flow is estimated between most of the active brain areas. In particular, bidirectional positive flow between the areas below C3 and C4 is incorrectly inferred for all three inverse methods. If GC with 57-dimensional MVAR modeling is used instead of the bivariate variant, no significant interactions are found for WMN and LCMV. Interestingly, a correct result is obtained for S-FLEX, namely a truly unidirectional flow from the area under C3 to the area under C4. However, the spurious flow in the opposite direction becomes significant at higher SNR also for this combination of methods.

Discussion

Avoiding spurious connectivity caused by volume conduction in Granger-causal analyses

Our study illustrates challenges in the estimation of brain interaction from EEG measurements. We tested three measures of effective connectivity in combination with three different approaches to assessing the statistical relevance of the results. All combinations correctly reveal the direction of information flow when being applied directly to (noiseless) source time series. However, real EEG data is always characterized by source mixing and the presence of brain and sensor noise. In this case it is easy to show that all pairs of electrodes are causally related to each other according to Granger's definition, even if only a single brain source is active. The reason for this is that the autoprediction of any noisy channel is improved if the prediction model is augmented by another channel measuring the same signal but different noise realizations. Our analyses confirm this theoretical consideration for both connectivity measures based on the idea of Granger causality (GC and PDC), demonstrating that these measures are not applicable to EEG data the way they are commonly applied using either permutation or analytic statistical testing.

Table 2
Average Euclidean distance between the two simulated dipoles and the corresponding two strongest peaks of the mean source current amplitude distributions (taken over 100 repetitions) as estimated by WMN, S-FLEX and LCMV. Smaller distances indicate better source reconstruction.

	WMN	LCMV	S-FLEX
$\gamma = 0.25$	11.1 cm	4.9 cm	6.0 cm
$\gamma = 0.5$	3.7 cm	0.5 cm	1.9 cm
$\gamma = 0.75$	2.6 cm	0.5 cm	0.5 cm

In this work we have pointed out that spurious results of Granger-causal analyses can be attributed to weak data asymmetries caused, e.g., by linear mixing of the interacting sources, as opposed to strong asymmetries related to genuine time-lagged information flow. Our novel approach of time inversion testing is designed to alleviate the influence of weak asymmetries on the result of any causal measure, while maintaining (or even amplifying) the contribution of strong asymmetries. Our results demonstrate that this approach effectively makes GC and PDC robust to artifacts of volume conduction. As a result, the simulated information flow from the left to right hemisphere is correctly recovered. Moreover, the fact that the corresponding negative flow in the opposite direction is as well recovered indicates that GC and PDC are capable of detecting reverse flow from reversed time series, further increasing the statistical power of the time inversion testing procedure. Note that this is not a trivial finding, since the dynamics of the time-reversed simulated data cannot necessarily be described by the finite-order linear MVAR process assumed by GC and PDC.

As a general remark, note that in a multivariate setting it would in principle be desirable to be able to destroy (or rather invert) strong asymmetries while preserving weak asymmetries only for a specific connection. This is not possible using time reversal since, e.g., inversion of only some of the time series would amount to cutting some connections completely. In practice, (complete) time inversion often leads to almost identical results for Granger-causal methods on real EEG data, indicating that partial manipulation would not have any effect either.

A third testing approach based on assessing net connectivity scores achieves good results for bivariate variants of GC and PDC, but breaks down for multivariate variants in the sense that predominantly spurious interactions are estimated. These spurious connections indicate that there are systematic differences in the GC/PDC scores of most of the channel pairs. Those are not related to the actual interaction but rather to asymmetries in the channelwise SNR (and even to the scale of the data in the case of PDC). As a general consequence of our analyses, we suggest to use Granger-causal measures only in combination with time inversion testing in the context of EEG connectivity studies. Importantly, it is possible that this combination even yields advantages over in settings comprising more than two interacting sources. Our experience shows that time inversion can be safely combined with antisymmetrization, if an analysis of net connectivity scores is desired. In order to better understand the behaviour of pure antisymmetrization testing, a thorough theoretical investigation of the properties of multivariate GC and PDC scores under spatial source mixing has to be conducted.

We observe that in most of our analyses the bivariate variants of GC and PDC achieve more significant results than their multivariate counterparts. Moreover, the result obtained using antisymmetrization testing is qualitatively better in the bivariate case than in the multivariate case. One reason for the inferior performance of multivariate MVAR modeling might be overfitting caused by the fact that the number of parameters of these models grows quadratically with the number of variables included. Suitable regularization of the MVAR model fitting (e.g., Haufe et al., 2010a) might therefore improve the results. In order to avoid biased results due to overfitting, Granger scores should moreover generally be evaluated on data that have not been used for fitting the underlying autoregressive models. Such potential biases are however not the predominant cause for the detection of spurious connectivity, as demonstrated here using bivariate Granger-causal models. Note that multivariate methods are commonly employed based on the consideration that the inclusion of more time series helps to rule out indirect connectivity between channels that are caused by a common confounder. However, that argument does not apply to EEG data, where all causal confounders contribute to all channels due to source mixing.

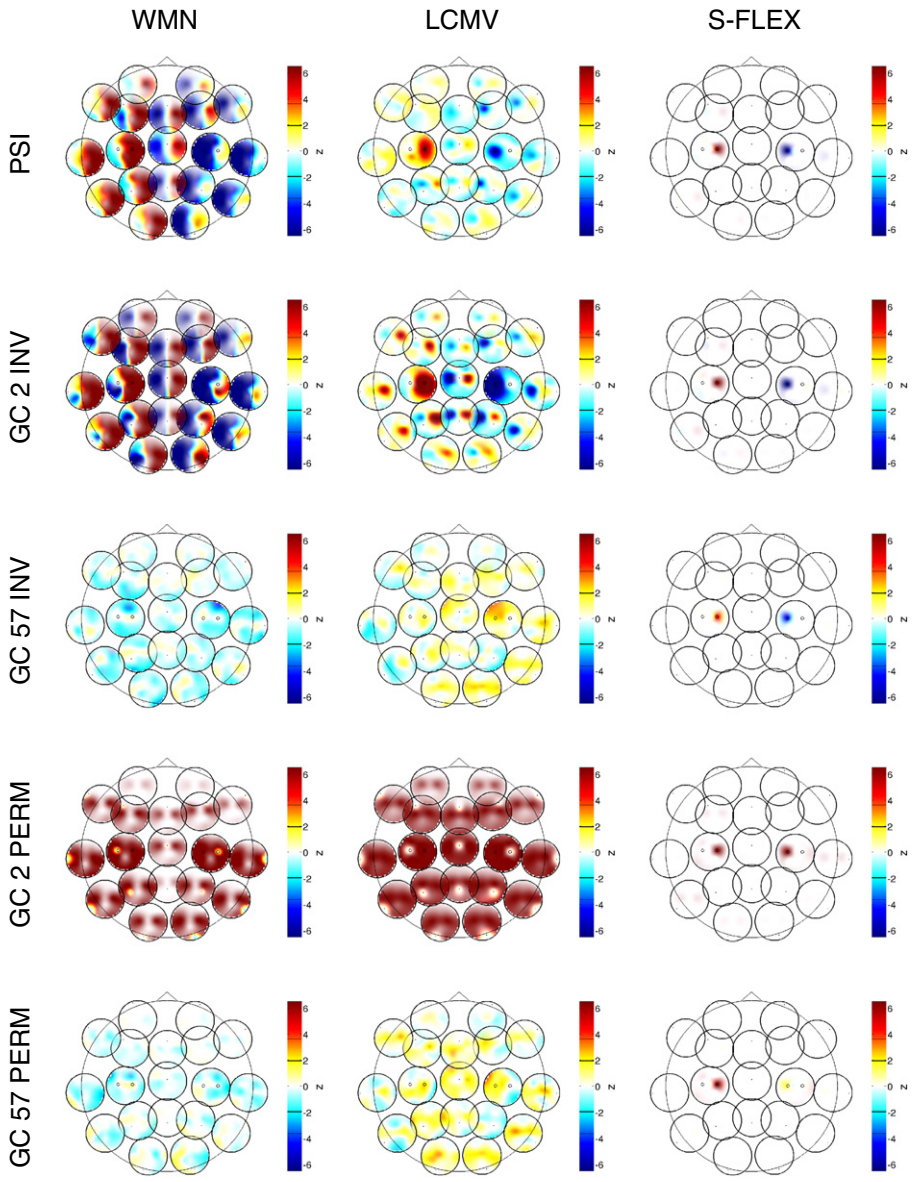

Fig. 5. Comparison of estimated effective connectivity between regions-of-interest defined based on the closest EEG electrode. Sources are estimated according to weighted minimum-norm (WMN) and sparse basis field expansions (S-FLEX) distributed inverse imaging, as well as linearly constrained minimum-variance (LCMV) beamforming. Connectivity is estimated using the phase-slope index (PSI), as well as variants of Granger causality using either bivariate (GC 2) or multivariate (GC 57) MVAR modeling in combination with either time inversion testing (INV) or permutation testing (PERM).

Interpretability issues in EEG sensor-space connectivity analysis

Since PSI inherently implements the ideas of antisymmetrization and time inversion testing, it yields results qualitatively similar to those of GC/PDC with time inversion testing in all our experiments. In a quantitative evaluation conducted on data provided by the "Signal Separation Evaluation Campaign" (SiSec) challenge 2011,[4] PSI however outperforms all competing methods including a variant of GC involving time inversion testing (Haufe et al., 2012).

Our simulated example has been designed such that the correctness of sensor-space connectivity estimates could be assessed in terms of the simulated left-to-right flow also on sensor level. However, in general the interpretation of sensor-space connectivity is much more difficult even when using robust measures, and generally requires knowledge of the field patterns of the underlying interacting sources. As one example, if it was unknown that our simulated sources give rise to bipolar field patterns, one would possibly incorrectly infer the presence of four sources (two sending and two receiving ones) located in frontal and parietal areas, while in fact there are only two simulated sources located centrally. One can easily imagine that interpretation becomes even harder if the source have more complex (and potentially more strongly overlapping) field spread patterns, and if there are more than two interacting sources. In addition, our analysis shows that even comparably small changes in the measurement setup such as the change of the reference electrode produce considerable variations in the resulting sensor-space connectivity maps, since such modifications affect the relative strengths with which driving and receiving sources are expressed in the sensors. A problematic influence of the choice of the reference has previously been mentioned by Nunez et al. (1997) and Marzetti et al. (2007).

The signal-to-noise ratio affects the results of all methods only quantitatively. As a result, however, the extent of the scalp regions between which significant interactions are detected may vary, which

[4] http://sisec.wiki.irisa.fr.

in turn may alter the qualitative assessment made by the data analyst. For example, analysis of the high-SNR scenario in Experiment 2 might lead to the conclusion that almost all brain sites are interacting, while the same analysis for lower SNR might infer interactions only between a number of parietal channels.

Benefit of source connectivity analysis

It is possible to transform EEG data into a reference-free representation by computing the scalp Laplacian (Kayser and Tenke, 2006). However, doing so does not ensure interpretability of sensor-space results, nor does it solve more general problems related to volume conduction, since the scalp Laplacian is incapable of extracting other than superficial radially oriented current sources. General source reconstructions via inverse methods or blind source separation (BSS) techniques are as well independent of the choice of the reference, while not exhibiting this limitation. They are therefore widely considered to be appropriate preprocessings for EEG-based connectivity analysis. While an evaluation of BSS methods is provided in Haufe (2011), we here assessed three inverse source reconstruction algorithms, two of which have been used in EEG source connectivity studies before. The advantage of working on inverse solutions compared to BSS methods is that the former allow one to study brain interaction directly in terms of the estimated signal-generating brain structures.

The quality of the overall source connectivity estimation relies on the ability of the employed inverse method to recover the spatial distribution of the sources. Importantly, we observe that large estimated active brain areas also translate into large areas between which significant connectivity (according to PSI or GC with time inversion) is estimated. The sources estimated by WMN and LCMV are more distributed around the simulated dipoles than it is the case for the S-FLEX estimate. Consequently, S-FLEX is more successful in narrowing down the interacting brain regions to the areas below C3 and C4, followed by LCMV. Note that the accuracy of the source reconstruction and thereby the quality of the source connectivity estimation increases with higher SNR for all source reconstruction methods.

It is often assumed that the application of (any) inverse source reconstruction eliminates the effects of volume conduction in the data such that connectivity measures designed to work well on unmixed data can be applied on reconstructed sources without hesitation. Our results, however, demonstrate that this is not the case due to artifacts of volume conduction persisting in all source estimates. Precisely, the application of GC with permutation testing on reconstructed sources leads to the detection of spurious connectivity regardless of the inverse method used.

Degree of realism of the simulations

In this study we modeled the simplest case that we consider realistic in the sense that the simulated EEG data comprises noise with temporal and spatial structure, as well as source mixing due to volume conduction. All remaining parts of our simulations were maximally simplified to the extent that meaningful connectivity analysis is still possible. Undoubtedly, more complex simulations involving more interacting sources with more strongly overlapping field patterns etc. would further aggravate connectivity estimation which, as we show, is already a non-trivial task even in the case of one pair of interacting sources. Similarly, linear dynamics as modeled here is an oversimplification of what can be expected to take place within the brain. Nevertheless, a linear source MVAR model does represent a possible mechanism of information transfer in the brain. Here, the source innovation sequences correspond to local brain activity, which is generated independently at various distant locations in the brain. The MVAR matrices relate to a stationary, but task-dependent brain network, which distributes the local information to all brain sites involved in the mental task with certain delays. The incoming information is fused at the various brain sites with the current local information to produce the source electrical activity that is indirectly observed in the EEG through the forward mapping.

Benchmarking

We would like to propose our data as a benchmark dataset for the evaluation of EEG-based connectivity estimation algorithms. Data-generating code can be downloaded.[5]

Establishing a standardized benchmarking protocol is more involved than proposing a dataset, since the performance measures have to be chosen depending on the general type of algorithm used for source and connectivity estimation. The statistics reported in this paper merely relate to the significance of the estimated connections rather than their correctness. Consequently, connectivity results have been presented rather qualitatively using the visualizations that would in reality be interpreted by neurophysiologists. We hereby encourage attempts on developing a generalized quantitative evaluation scheme for EEG-based connectivity analysis, possibly making use of the dataset proposed here. However, it is rather challenging to accommodate the various approaches including sensor-space, inverse source reconstruction and potentially blind source separation approaches into a common evaluation framework. Since sensor-space approaches lack the notion of sources, it is hard to assess their performance in terms of source connectivity estimation in general. In genuine source connectivity analysis, the overall estimation is split into source and connectivity estimation steps, and so must be the evaluation. We here quantified source estimation accuracy using the earth mover's distance in the inverse source reconstruction setting. A meaningful quantitative evaluation of the subsequent connectivity estimation is only possible if the estimated sources can be matched to the true sources one-to-one, which is not always easily possible.

Given the simplicity of the current simulation scenario, the fact that a method performs well here does of course not guarantee that it will do so on any other (in particular real) dataset. However, results obtained for a number of different methods do provide an indication of which method should be preferred. To increase the confidence in a methodology (or to possibly invalidate it), it is necessary to perform additional experiments reflecting prior assumptions as closely as possible. In particular, the influence of inaccuracies in the forward model should be investigated, where numerical simulations may also be complemented with phantom studies. Finally, it would be desirable to obtain a real-world benchmark dataset for which the active brain regions as well as the predominant information flow between them is known with high confidence.

Conclusions

In this paper we have studied simulated EEG data generated from a simple computational model of brain interaction. We have demonstrated that standard Granger-causal measures are not able to deal with such data due to the inherent source mixing caused by volume conduction, and have proposed a practical remedy based on time inversion statistical testing. However, even for measures robust to volume conduction the interpretation of EEG sensor-space connectivity is difficult. In order to obtain better interpretable results, it is helpful to conduct connectivity analysis on source estimates. However, the EEG inverse problem is ill-posed; therefore, the assumptions made by a source estimation algorithm must match the properties of the sources to be reconstructed. Inverse source reconstructions algorithms, for example, should be able to spatially separate multiple

[5] http://bbci.de/supplementary/EEGconnectivity/.

distinct sources, while being applicable to entire EEG time series. A parallel field of research not covered here is the development of BSS methods suitable for source connectivity analysis (Gómez-Herrero et al., 2008; Haufe et al., 2010b).

Acknowledgments

This work was financially supported by the Bundesministerium für Bildung und Forschung (BMBF grant nos. 16SV2234 and 01GQ0850), the Deutsche Forschungsgemeinschaft (DFG: SFB 936), the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT-216886, and by the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology, under Grant R31-10008.

References

Astolfi, L., Cincotti, F., Mattia, D., Salinari, S., Babiloni, C., Basilisco, A., Rossini, P.M., Ding, L., Ni, Y., He, B., Marciani, M.G., Babiloni, F., 2004. Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution EEG. Magn. Reson. Imaging 22, 1457–1470.

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Marciani, M.G., Babiloni, F., 2006. Estimation of the cortical connectivity patterns during the intention of limb movements. IEEE Eng. Med. Biol. 25, 32–38.

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.

Babiloni, C., Ferri, R., Moretti, D.V., Strambi, A., Binetti, G., Dal Forno, G., Ferreri, F., Lanuzza, B., Bonato, C., Nobili, F., Rodriguez, G., Salinari, S., Passero, S., Rocchi, R., Stam, C.J., Rossini, P.M., 2004. Abnormal fronto-parietal coupling of brain rhythms in mild Alzheimer's disease: a multicentric EEG study. Eur. J. Neurosci. 19, 2583–2590.

Babiloni, F., Cincotti, F., Babiloni, C., Carducci, F., Mattia, D., Astolfi, L., Basilisco, A., Rossini, P.M., Ding, L., Ni, Y., Cheng, J., Christine, K., Sweeney, J., He, B., 2005. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. Neuroimage 24, 118–131.

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., 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 18, 14–30.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process. Mag. 25, 41–56.

Blinowska, K., Kus, R., Kaminski, M., Janiszewska, J., 2010. Transmission of brain activity during cognitive task. Brain Topogr. 23, 205–213.

Brockwell, P.J., Davis, R.A., 1998. Time Series: Theory and Methods (Springer Series in Statistics). Springer.

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.

Chatrian, G.E., Lettich, E., Nelson, P.L., Apr 1988. Modified nomenclature for the "10%" electrode system. J. Clin. Neurophysiol. 5 (2), 183–186.

Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. 2, 56–78.

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

Gosset, W.S., 1908. The probable error of a mean. Biometrika 6, 1–25 (originally published under the pseudonym "Student").

Gow, D.W., Segawa, J.A., Ahlfors, S.P., Lin, F.H., 2008. Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. Neuroimage 43, 614–623.

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

Grave de Peralta-Menendez, R., Gonzalez-Andino, S.L., 1998. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. IEEE Trans. Biomed. Eng. 45, 440–448.

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., 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., 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., Nolte, G., Müller, K.-R., Krämer, N., 2010a. Sparse causal discovery in multivariate time series. In: Guyon, I., Janzing, D., Schölkopf, B. (Eds.), Causality: Objectives and Assessment: JMLR W&CP, vol. 6, pp. 97–106.

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

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., Nikulin, V.V., Nolte, G., 2012. 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. : Lecture Notes in Computer Science, vol. 7191. Springer, Berlin/ Heidelberg, pp. 25–33.

Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A., Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22, 324–333.

Horwitz, B., 2003. The elusive concept of brain connectivity. Neuroimage 19, 466–470.

Ioannides, A.A., Bolton, J.P.R., Clarke, C.J.S., 1990. Continous probabilistic solutions to the biomagnetic inverse problem. Inverse Prob. 6, 523–542.

Jeffs, B., Leahy, R.M., Singh, M., 1987. An evaluation of methods for neuromagnetic image reconstruction. IEEE Trans. Biomed. Eng. 34, 713–723.

Jirsa, V., McIntosh, A., 2007. Handbook of brain connectivity. Understanding complex systems. Springer.

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.

Kamiński, M., Blinowska, K., Szelenberger, W., 1997. Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness. Electroencephalogr. Clin. Neurophysiol. 102, 216–227.

Kamiński, M., Ding, M., Truccolo, W.A., Bressler, S.L., 2001. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. Biol. Cybern. 85, 145–157.

Kayser, J., Tenke, C.E., 2006. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin. Neurophysiol. 117, 348–368.

Marinazzo, D., Pellicoro, M., Stramaglia, S., 2008. Kernel method for nonlinear Granger Causality. Phys. Rev. Lett. 100, 144103.

Martino, J., Honma, S.M., Findlay, A.M., Guggisberg, A.G., Owen, J.P., Kirsch, H.E., Berger, M.S., Nagarajan, S.S., Mar 2011. Resting functional connectivity in patients with brain tumors in eloquent areas. Ann. Neurol. 69 (3), 521–532.

Marzetti, L., Nolte, G., Perrucci, M.G., Romani, G.L., Del Gratta, C., 2007. The use of standardized infinity reference in EEG coherency studies. Neuroimage 36, 48–63.

Matsuura, K., Okabe, Y., 1995. Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans. Biomed. Eng. 42, 608–615.

Neumaier, A., Schneider, T., 2001. Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Trans. Math. Softw. 27, 27–57.

Nolte, G., Dassios, G., 2005. Analytic expansion of the EEG lead field for realistic volume conductors. Phys. Med. Biol. 50, 3807–3823.

Nolte, G., Müller, K.-R., 2010. Localizing and estimating causal relations of interacting brain rhythms. Front. Hum. Neurosci. 4, 209.

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., Meinecke, F.C., Ziehe, A., Müller, K.-R., 2006. Identifying interactions in mixed and noisy complex systems. Phys. Rev. E 73, 051913.

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.

Nolte, G., Ziehe, A., Krämer, N., Popescu, F., Müller, K.-R., 2010. Comparison of Granger causality and phase slope index. In: Guyon, I., Janzing, D., Schölkopf, B. (Eds.), Causality: Objectives and Assessment: JMLR W&CP, vol. 6, pp. 267–276.

Nunez, P., Srinivasan, R., 2006. Electric fields of the brain: the neurophysics of EEG. Oxford University Press.

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.

Nunez, P.L., Silberstein, R.B., Shi, Z., Carpenter, M.R., Srinivasan, R., Tucker, D.M., Doran, S.M., Cadusch, P.J., Wijesinghe, R.S., 1999. EEG coherency II: experimental comparisons of multiple measures. Clin. Neurophysiol. 110, 469–486.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25, 230–242.

Rubner, Y., Tomasi, C., Guibas, L.J., 2000. The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision 40, 99–121.

Schlögl, A., Supp, G., 2006. Analyzing event-related EEG data with multivariate autoregressive parameters. Prog. Brain Res. 159, 135–147.

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

Seth, A.K., 2010. A MATLAB toolbox for Granger causal connectivity analysis. J. Neurosci. Methods 186, 262–273.

Silberstein, R.B., 2006. Dynamic sculpting of brain functional connectivity and mental rotation aptitude. Prog. Brain Res. 159, 63–76.

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.

Srinivasan, R., Winter, W.R., Ding, J., Nunez, P.L., 2007. EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. J. Neurosci. Methods 166, 41–52.

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.

Theiler, J., Prichard, D., 1997. Using 'surrogate surrogate data' to calibrate the actual rate of false positives in tests for nonlinearity in time series. Fields Inst. Commun. 11, 99–113.
Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D., 1992. Testing for nonlinearity in time series: the method of surrogate data. Phys. D: Nonl. Phenomena 58 (1–4), 77–94.
Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: influence, causality and biophysical modeling. NeuroImage 58 (2), 339–361.
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.
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.
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.