Assessing brain effective connectivity from EEG – a simulation study

Stefan Haufe [a,b,*], Vadim V. Nikulin [c,d],

Klaus-Robert Müller [a,b,e] and 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, Korea*

[f] *Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany*

Abstract

Information flow between brain areas from EEG measurements is hard to estimate reliably due to the presence of noise, as well as due to volume conduction. We present results of a simulation study, in which we test the ability of popular measures of effective connectivity to detect neuronal interactions, as well as machine learning approaches to inverse source reconstruction and blind source separation to improve the estimation. We find that volume conduction limits the neurophysiological interpretability of sensor-space connectivity analyses and may even lead to conflicting

results. On the other hand, the estimation of connectivity on the basis of source activity, which has been estimated with some of the tested machine learning methods, yields correct and interpretable results. The success here crucially depends on the correctness of the source reconstruction, which in turn depends on the capability of the method to model multiple spatially-distinct interacting sources. As a novel methodological contribution, we propose to use time-reversed data as surrogates in order to cancel out artifactual confounders that are not related to true interaction, e. g., due to the volume conduction. We show that this novel approach improves the performance of Granger-causal connectivity measures. Finally, integrating the insights of our study, we provide a guidance for measuring information flow in EEG. Software and benchmark data is made available.

1 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

* Corresponding author.
  Email address: `stefan.haufe@tu-berlin.de` (Stefan Haufe).

2

inverse problem. In the EEG-based analysis of directed information flow between brain regions (i.e., *effective brain connectivity* 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, electroencephalographic recordings have been widely used in neuroscience to estimate brain effective connectivity (e.g., Kamiński et al., 1997; Babiloni et al., 2004; Astolfi et al., 2004; Babiloni et al., 2005; Silberstein, 2006; Srinivasan et al., 2007; Supp et al., 2007; Blinowska et al., 2010). Only recently, the fact that volume conduction has to be accounted for in EEG- and MEG-based brain connectivity studies has been seriously acknowledged (Nolte et al., 2004; Schlögl and Supp, 2006; Nolte et al., 2006, 2008; Gómez-Herrero et al., 2008; Schoffelen and Gross, 2009; Nolte and Müller, 2010).

In this paper, we present results of a series of simulation experiments in which we systematically assessed common measures of effective connectivity in terms of their ability to infer source interactions from pseudo-EEG recordings under various conditions and after various common preprocessings such as inverse source reconstruction and blind source separation. We here 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; an opinion that does however not seem to be the predominant one in the field. While Theiler and Prichard even postulate "the importance of avoiding (that's right - avoiding) 'real' data", since "Too many studies, we feel, have been corrupted by the dogma that a methodology is not tested unless it is tested on real data." (Theiler and Prichard, 1997), we would like to argue that both real and sim-

3

ulated data are in general suitable for evaluating a method. However, 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, 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 to attempt an assessment of a broad variety of methods on standardized 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 and noise mixing may easily mislead connectivity estimation depending on the type of measure used, while the success of source estimation 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 in the most instructive way, namely by visually demonstrating how certain properties of the data in combination with the characteristics of the various methods can potentially spoil connectivity analyses. The present study is thus primarily a qualitative one, in which we do not attempt to exhaustively chart the huge space of adjustable parameters. Rather, 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. Since space limitations allow us to present only our most important results graphically within the paper, we provide the remaining results in an extensive supplement, which can be downloaded along with the data and code used in our simulations at www.bbci.de.

As a methodological contribution we further introduce the concepts of weak and strong asymmetries in multivariate time series in the context of causal modeling. Weak asymmetries are by definition not related to time-lagged interactions. Nevertheless they frequently mislead certain types of causal analyses. 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 state-of-the-art measures of effective connectivity as well as methods for inverse source reconstruction and blind source separation in Section 2. Section 3 introduces a novel methodology for improving connectivity estimates by distinguishing between weak and strong data asymmetries. Section 4.1 presents a series of simulations, in which connectivity measures and preprocessing steps are evaluated within a common framework. The results are discussed in Section 5, before we reach conclusions in Section 6.

2   Methods

2.1   *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.

2.1.1 Granger causality based on model errors

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

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

where the *innovations* $\boldsymbol{\varepsilon}(t)$ 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 $\mathbf{x}(t)$. Notably, 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 $\mathbf{x}_{\{1,...,M\}} = \mathbf{x}$, as well as for the reduced set $\mathbf{x}_{\{1,...,M\}\backslash\{i\}}$ of available time series. Denoting the prediction errors of the full model by $\boldsymbol{\varepsilon}^{\text{full}}$ and those of the reduced model by $\boldsymbol{\varepsilon}^{\backslash i}$, 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. I.e.,

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

If only unidirectional flows are to be detected, or if one is only interested in the dominant direction of a bidirectional flow, it is reasonable to investigate

the pairwise *net flow*, which is obtained by *antisymmetrizing* the outcome of a connectivity measure M via $M_{i,j}^{\text{net}} = M_{i,j} - M_{j,i}$. A time series x_i is said to be the *net driver* of a time series x_j, if $M_{i,j}^{\text{net}}$ is significantly greater than zero. Analogously, x_j is the *net receiver* of x_i. Note that testing net connectivity scores is equivalent to performing a paired test for differences between $M_{i,j}$ and $M_{j,i}$.

2.1.2 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. Hence, a Granger-causal dependency can be inferred directly from the offdiagonal MVAR coefficients instead of prediction errors. This is done per frequency by *partial directed coherence* (PDC). Let $\tilde{B}(f)$ be the Fourier transform of the coefficients of a multivariate AR model fitted on all available time series, and $\bar{\tilde{B}}(f) = I_M - \tilde{B}(f)$, PDC is defined as

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

and is an estimate of the strength of the information flow from x_i to x_j at frequency f. Note that this is the squared absolute value of the complex-valued quantity introduced in Baccalá and Sameshima (2001). Unlike GC, PDC is not independent of the scale of the data.

2.1.3 The phase-slope index

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

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

where

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

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 $\text{CHY}_{i,j}(f)$ 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

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

From the Hermitian property of the cross-spectral matrices it follows that $\Psi_{i,j} = -\Psi_{j,i}$. Hence, PSI measures only net flows. Using the same property,

it can be shown that PSI exactly flips its sign when being applied to time-reversed data. Moreover, due to the use of the normalized cross-spectrum, PSI is invariant with respect to rescaling of the data.

2.2 A model of EEG data

The most general generative model of EEG data is given by

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

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* $\boldsymbol{\eta}(t)$ 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. The variables A and $\mathbf{s}$ are not identifiable given the observations $\mathbf{x}$ without imposing further assumptions.

2.3 Inverse source reconstruction

In reality, the matrix A describes a physical process, namely the propagation of the brain electric currents from the source regions to the EEG electrodes. If a physical model of the head exists, the mixing patterns of idealized brain sources (the columns of A) can be computed. This step is called *forward modeling. Inverse source reconstruction* is concerned with the estimation of $\mathbf{s}$ given

9

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 contraint, 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.

2.3.1 The weighted minimum-norm estimate

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

2.3.2 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 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 model 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.

2.3.3   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 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). The power of the source time courses has to be normalized using a noise covariance estimate in order to obtain an unbiased estimate of the spatial distribution of source power.

2.4   Blind source separation

Instead of using a possibly inaccurate forward model and carrying out its ill-posed inversion, the mixing matrix can also be estimated jointly with the

source time series in a completely data-driven way. This approach is called *blind source separation* (BSS).

2.4.1 Independent component analysis

The goal of *independent component analysis* (Comon, 1994; Bell and Sejnowski, 1995; Hyvärinen and Oja, 2000) is to search for "maximally independent" sources. Practical ICA implementations focus on minimizing certain aspects of dependence. Joint diagonalization of appropriate matrices is one popular way to achieve this. For example, the second order blind identification (SOBI) approach (Belouchrani et al., 1997) and the related temporal decorrelation source separation (TDSEP) approach (Ziehe and Müller, 1998) perform joint diagonalization of symmetrized empirical cross-covariance matrices related to multiple time lags. In the joint approximate diagonalization (JADE) approach (Cardoso and Souloumiac, 1996), slices of the fourth-order cumulant tensor are simultaneously diagonalized in order to extract source time series with distinct non-Gaussian distributions. Despite its independence assumption, various authors have used ICA in source synchronization and connectivity studies using the argument that ICA finds the "least dependent" sources (Beckmann et al., 2005; Meinecke et al., 2005; Astolfi et al., 2007), the coupling of which may still be analyzed.

2.4.2 Convolutive ICA

In the *blind deconvolution* setting

$$\mathbf{x}(t) = \sum_{p=0}^{P} A(p)\mathbf{s}(t-p) + \boldsymbol{\varepsilon}(t) , \qquad (8)$$

the observation sequence is modeled as a spatio-temporally-filtered version of source signals (Attias and Schreiner, 1998; Parra and Spence, 2000). The convolutive ICA with an auto-regressive inverse model approach (CICAAR, Dyrholm et al., 2007) estimates the parameters $A(p)$ under the assumption of independent non-Gaussian (e. g., hyperbolic secant distributed) source signals $\mathbf{s}(t)$. These "convolutive sources" have an interpretation as innovations of a source MVAR process (where "sources" refers to the instantaneously demixed time series as in (7)). Therefore, CICAAR can be used to assess EEG source effective connectivity (Haufe et al., 2010).

2.4.3   MVARICA

The multivariate autoregression plus independent component analysis (MVAR-ICA) approach by Gómez-Herrero et al. (2008) assumes the basic linear model for the EEG and a standard multivariate AR model for the sources. This implies that the EEG time series follows a multivariate AR model with linearly transformed coefficients and innovations. Thus, MVARICA involves fitting a sensor-space MVAR model using the ARFIT algorithm (Neumaier and Schneider, 2001). The remaining mutual dependence of the innovations is removed in a subsequent ICA step on the innovation time series, which is performed using efficient fast independent component analysis (EFICA, Koldovský et al., 2006). This leads to an estimate of the mixing matrix, which is used to obtain the underlying sources' time series as well as the source MVAR coefficients.

2.4.4 Sparsely-connected sources analysis

In sparsely-connected sources analysis (SCSA, Haufe et al. (2010)), the same model as in MVARICA is assumed. Under the assumptions that the source MVAR innovations are i.i.d. and subject to non-Gaussian distributions (as in CICAAR, the hyperbolic secant distribution is assumed), the source MVAR coefficients are identified along with the mixing matrix in a single estimation step. As a regularization, SCSA additionally sparsifies the connectivity graph using an $\ell_{1,2}$-norm penalty on the source MVAR coefficients.

3 Data asymmetries

Many measures of causal interaction (a. k. a. 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. In fact, methods to estimate causal relations 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 between 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

$$C(p) = \left\langle \left(\mathbf{x}(t) - \widehat{\mu}_{\mathbf{x}}\right) \left(\mathbf{x}(t-p) - \widehat{\mu}_{\mathbf{x}}\right)^{\top} \right\rangle , \qquad (9)$$

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 but can also be considered as 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, do exhibit a strong asymmetry. This strong asymmetry is still be observed when both channels after linearly mapping the source activity to EEG sensor space. However, the 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 estimate of causal drive is based. A potential remedy is to test against *surrogate* data, which are usually designed to share all the weak asymmetries of the data under study, but to lack the strong asymmetries. Thus, a statistical test for differences of the results obtained from both datasets should ideally reveal the strong asymmetries (i. e. time-lagged interactions). A standard choice for surrogate data is the original data, in which the temporal order has been randomly *permuted* (Kamiński et al., 2001).

As an alternative, we here suggest to compare the specific result of a causal analysis with the outcome of the method applied to *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 $\widetilde{C}(p)$, is given as

$$\widetilde{C}(p) = C(-p) = C^\top(p) \tag{10}$$

implying that time inversion inverts all strong asymmetries but none of the weak asymmetries.

Our approach is able to cancel out weak asymmetries in a similar way as permutation testing. However, it additionally amplifies strong asymmetries, which are present in the surrogates in the opposite direction. Consequently, empirical results indicate that time inversion testing achieves greater statistical power than permutation testing (see Haufe et al. (2012)).

4 Experiments

In the following we present a series of five experiments designed to assess state-of-the-art approaches to EEG-based 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 EEG comprising realistic effects of volume conduction and noise. The third experiment demonstrates the influence of the choice of the reference electrode and the SNR. The last two experiments assess the effectiveness of inverse source reconstruction and blind source separation techniques.

4.1 Experiment 1: two interacting sources

4.1.1 Setting

We first assess measures of effective connectivity on the source level, where no instantaneous mixing of signals due to volume conduction is present. The simulated system comprises two sources $s_{1/2}(t)$ following a stable bivariate AR process of order $P = 5$. We generate $T = 10\,000$ source samples. The AR coefficients are sampled independently from $\mathcal{N} \sim (0, 0.01)$. By setting the off-diagonal coefficients $B_{1,2}(p), 1 \leqslant p \leqslant P$ to zero, while all other coefficients remain nonzero, unidirectional flow from s_1 to s_2 is modeled. The innovations $\boldsymbol{\varepsilon}(t)$ of the source AR process are drawn from the univariate Gaussian distribution. We perform 100 repetitions of the experiment. For each repetition, a dataset comprising distinct innovation terms and source AR coefficients is

generated. Since in practice source time series cannot be observed, the present experiment mainly serves as a proof-of-concept for effective connectivity measures and a baseline for source demixing algorithms. We consider the noiseless case here, while the influence of noise has been studied in detail in the literature (Nolte et al., 2008, 2010).

We apply Granger causality, partial directed coherence and the phase-slope index to *normalized* the 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 Connectivity Analysis" toolbox (Seth, 2010)[2] is used to compute GC, and the MVARICA toolbox (Gómez-Herrero et al., 2008, code not anymore publicly available) 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 scores related to all frequency bins to obtain a global measure of interaction.

Statistical significance of the results is assessed in five ways. First, the 2×2 connectivity matrices are antisymmetrized, and the resulting net flow is tested for being significantly different from zero using a two-sided one-sample t-test. Second, testing against temporally permuted as well as time-reversed surrogate data is applied. Here, we use a two-sided paired t-test to assess whether the difference in estimated information flow according to each of the connectivity measures significantly differs from zero. Finally, antisymmetrization is combined with subsequent permutation and time inversion testing. In all

---

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

[2] http://www.informatics.sussex.ac.uk/users/anils/aks_code.htm

[3] http://www.gps.caltech.edu/~tapio/arfit

cases, t-scores are converted into z-scores via $z = \mathrm{cdf}_z^{-1}(\mathrm{cdf}_t(t, \nu))$, 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. Significant effects are indicated by z-scores with absolute values greater than 1.96 for tests involving antisymmetrization, and 2.24 for test not involving antisymmetrization (accounting for the simultaneous testing of two hypotheses using Bonferroni correction).

4.1.2 Results

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. In the three testing variants, in which antisymmetrization is used, the corresponding *negative* flow (i.e. inflow) from s_2 to s_1 is trivially also observed. If plain permutation testing (without antisymmetrization) 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 for PSI owing to the intrinsic antisymmetry of this measure. Interestingly, we observe highly significant ($z_{2,1} < 10$) negative flow from s_2 to s_1 for all three measures if time-reversed surrogates without antisymmetrization are used. Thus, all methods are capable of discovering the reverse flow from reversed time series, indicating in some sense their robustness to model violations (since the dynamics of time-reversed data *cannot* be necessarily described by the finite-order linear MVAR model used by GC and PDC), and hence demonstrating the versatility of the idea of time inversion testing.

4.2 Experiment 2: simulated EEG

4.2.1 Setting

We now consider simulated EEG data comprising effects of volume conduction, as well as noise. The artificial EEG signal is generated according to

$$\mathbf{x}(t) = \frac{(1-\gamma)\left(\frac{1}{2}\sum_{i=1}^{2}\frac{\mathbf{a}_i s_i(t)}{\|(S^\top)_i\|_2}\right)}{\|\operatorname{\mathbf{vec}}(S)\|_2} + \gamma\frac{\boldsymbol{\eta}(t)}{\|\operatorname{\mathbf{vec}}(E)\|_2}, \qquad (11)$$

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), $\boldsymbol{\eta}$ is noise and γ, $0 \leqslant \gamma \leqslant 1$ is a parameter that adjusts the SNR. Moreover, $E = (\boldsymbol{\eta}(1), \ldots, \boldsymbol{\eta}(T))$ 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 $\|\operatorname{\mathbf{vec}}(S)\|_2$ and $\|\operatorname{\mathbf{vec}}(E)\|_2$, 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 Figure 1. The current moment vectors of both dipoles are tangentially oriented, leading to bipolar field patterns. In general, 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

20

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 Figure 1.

The noise terms $\boldsymbol{\eta}(t) = \boldsymbol{\eta}^{\text{sensor}} + A^{\text{biol.}}\boldsymbol{\eta}^{\text{biol.}}$ are composed of *sensor noise* $\boldsymbol{\eta}^{\text{sensor}}$, 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 $\eta_1^{\text{biol.}}(t), \ldots, \eta_{10}^{\text{biol.}}(t)$ of which are generated using random stable univariate AR models of order $P^{\text{biol.}} = 10$, and are mixed by a spread matrix $A^{\text{biol.}}$ obtained representing ten randomly placed dipoles with random current moment vectors. Sensor and biological noises are scaled to contribute equally to the overall noise $\boldsymbol{\eta}$.

We generate 100 artificial EEG datasets using the source time series introduced in Experiment 1. The underlying source dipoles are kept 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 (e. g., Kamiński et al., 1997). As in Section 4.1, statistical significance of each entry of the sensor-space connectivity matrices estimated by GC, PDC and PSI is assessed using antisymmetrization, permutation testing, time inversion testing, and combinations of antisymmetrization and permutation/time inversion testing.

[Insert Figure 1 around here]

21

4.2.2 Results

The connectivity graphs estimated by GC, PDC and PSI using five different variants of statistical testing are visualized in Figure 2 using *head-in-head plots* (Nolte et al., 2004, 2008). Each head-in-head plot is composed of 19 small circles representing the human scalp. These are arranged within one large scalp plot according to the positions of the 19 electrodes of the original 10-20 electrode placement system. Each of the small scalp plots thereby shows the estimated interaction of the respective electrode to all 58 other electrodes, where red and yellow colors ($z > 0$) stand for information outflow and blue and cyan colors ($z < 0$) stand for information inflow. Bonferroni correction is used to account for multiple testing. The Bonferroni factor is 551, which is the number of electrode pairs ($19 \cdot (59 - 1)/2$) visualized. The Bonferroni-corrected significance threshold ($z = 4.0$) is indicated by a thin black line in the colorbar, while the z-score corresponding to an uncorrected p-value ($z = 2.0$) is indicated by a thick black line.

The phase-slope index correctly reveals information flow from the left to the right hemisphere regardless of the statistical testing procedure used (note that the three variants not involving permutation testing yield numerically identical results owing to the antisymmetry of this method). The observed connectivity matrices resemble the true field patterns of the underlying sources in that almost all electrodes in the right hemisphere are estimated to receive information from almost all electrodes in the left hemisphere, while 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. Generally (assuming that driving and receiving sources are similarly strong and noise sources contribute equally much to all sensors), PSI roughly varies

22

with $\log(|a_{a,i}||a_{b,j}|/|a_{b,i}||a_{a,j}|)$, which we call the *driver-receiver ratio* related to a pair of electrodes (i, j) and a pair of driving and receiving sources (a, b).

Granger causality yields rather noisy connectivity patterns in all five testing cases with only very few significant connections between electrodes in general. For plain permutation testing (Figure 2 (a)), most electrodes are estimated to send information to their closest neighbors. This is explained by the fact that neighboring sensors (due to volume conduction) pick up highly correlated signals, which (due to independent noise contributions) improve mutual predictability. However, this effect does not relate to the simulated underlying flow from the left to the right hemisphere. In fact, in conflict with the driving source being located in the left hemisphere, GC detects a number of significant flows to originate from electrodes in the right hemisphere. For the remaining four testing protocols there are few significant connections, which generally exhibit little structure. The clearest picture is observed for the combination of antisymmetrization and time inversion testing (Figure 2 (e)). Here, predominantly interhemispheric flows passing from two electrodes in the left hemisphere to two electrodes in the right hemisphere are detected, indicating that time inversion testing reveals strong asymmetries more successfully than permutation testing.

Partial directed coherence behaves similar as GC unless plain antisymmetrization (without the use of further surrogate data) is used. Here, highly significant gross symmetric bilateral information exchange from electrodes with higher SNR to electrodes with lower SNR is observed. Thus, the corresponding plot in Figure 2 (c) is a good example of how weak asymmetries in the data can spoil causal estimates. Moreover, PDC depends on the *scale* of the data, which is a further confounder for weak asymmetries. Interestingly, the application of

23

PDC on non-normalized data suggests that information flows from low-SNR to high-SNR channels, i. e., exactly in the opposite direction (see supplement). The effects of SNR and scale are however ruled out by testing against appropriate surrogate data.

In summary, only PSI robustly detects the simulated general interhemispheric flow in this example regardless of the testing protocol used. For GC and PDC, depending on the protocol, there can be spurious connectivity or almost no significant connectivity at all. Thus, our results signify that the presence of source mixing and noise poses more serious challenges for some connectivity measures than the analysis of unmixed signals as performed in the previous experiment. Moreover, all results obtained in sensor space might easily lead to wrong conclusions regarding the number, size and location of the brain areas involved in the interaction. For example, due to the bipolar structure of the field patterns considered here, one might incorrectly infer the presence of four sources (two sending and two receiving ones).

[Insert Figure 2 around here]

4.3   Experiment 3: influence of reference electrode and SNR

4.3.1   Setting

We inferred previously that PSI designates two electrodes as driver and receiver depending on the driver-receiver ratio. However, this ratio depends on how much signal the reference electrode picks up from the underlying driving and receiving sources. In this experiment, we demonstrate how a change of reference can drastically affect this ratio and hence influences connectivity

estimation. To this end, we rereference the data used in Experiment 2 by subtracting the activity of the P3 electrode. This choice is particularly harmful here (and hence suitable for demonstration purposes), since one of the sources (the driver) is most strongly expressed in P3.

Another confounder that might affect the interpretation of sensor-space connectivity maps is the signal-to-noise ratio. Here we present results for two additional SNRs of $\gamma = 0.25$ and $\gamma = 0.75$ using the standard nose-referenced dataset. In all experiments the data are normalized and subjected to effective connectivity analysis by means of GC, PDC and PSI with subsequent statistical testing using antisymmetrization and time inversion.

4.3.2 Results

Figure 3 (a) shows sensor-space connectivity maps obtained from rereferenced data. The PSI estimate differs substantially from what is obtained using nose-referenced data (see Figure 2) in that the general left-to-right flow is not so obvious anymore. Most importantly, regions in the right hemisphere appear as drivers of regions in the left hemisphere, which is in conflict with the simulated direction of information flow. At the same time, no significant interaction is estimated at all for the left posterior region, since, after rereferenciation, all activity in the region around P3 vanishes (see Figure 1), which causes global changes in the driver-receiver ratio. Granger causality and partial directed coherence are affected by these changes in a similar way as PSI, although their limited statistical power makes these changes harder to notice.

Results obtained using different SNRs are depicted in Figure 3 (b) and (c). Granger causality and partial directed coherence do not change much when

the SNR is decreased to $\gamma = 0.25$, i.e., the few significant connections remain. This is different for PSI, which exhibits much less significant connections than before. Importantly, the upper lobes of the two bipolar field patterns do not reach significance anymore, which might lead to different conclusions regarding the number and location of the underlying interacting sources compared to the case $\gamma = 0.5$. Analogously, for $\gamma = 0.75$ one might erroneously come to the conclusion that much larger areas within both hemispheres are interacting than for $\gamma = 0.5$ and $\gamma = 0.25$, although in fact the simulated sources in all cases are just point sources without spatial extent. Interestingly, the estimation using GC and PDC does not improve with higher SNR in terms of significance. Rather, most of the significant connections seen for lower SNRs vanish.

Generally speaking, our results demonstrate that sensor-space connectivity maps greatly depend on the reference electrode used (which is an arbitrary choice) and the SNR of the data (which cannot be controlled for). A problematic influence of the choice of the reference has been acknowledged previously in the literature (Nunez et al., 1997; Marzetti et al., 2007). As a potential solution to the problem, the data can be transformed into a reference-free representation, e.g., by computing the scalp Laplacian (Kayser and Tenke, 2006), or by performing genuine source estimation using inverse source reconstruction or blind source separation techniques. While the former approach is not suitable for other than superficial radially-oriented sources (results on the current data are provided in the supplement nonetheless), the latter approaches are more general and therefore considered in the following two sections.

[Insert Figure 3 around here]

26

4.4 Experiment 4: inverse source reconstruction

4.4.1 Setting

Working on solutions of the EEG inverse problem allows one to study brain interaction directly in terms of the estimated signal-generating brain structures, which are – as our previous results indicate – otherwise hard to infer from sensor-space connectivity maps. A number of studies have therefore investigated effective connectivity on source estimates obtained from linear inverse imaging (Babiloni et al., 2005; Astolfi et al., 2006; Gow et al., 2008) or beamforming (Martino et al., 2011; Brookes et al., 2011; Wibral et al., 2011). Here, we compare distributed inverse imaging according to WMN and S-FLEX with beamforming via LCMV as preprocessings for EEG-based source connectivity analysis. The source reconstruction is conducted in the same head model in which the data were simulated. The interior of the whole brain shell is partitioned into $N = 2\,142$ voxels of $10\,\mathrm{mm}$ 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, these details play a minor role regarding the evaluation of inverse methods and source connectivity estimation in general.

We use own implementations of WMN, S-FLEX and LCMV for transforming the pseudo-EEG measurements into source time series. The regularization parameter of the WMN estimate 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 characterized by nonzero estimated 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 (disregarding the noise dipoles) with the estimated current densities (Haufe et al., 2008).

In order to reduce the dimensionality in source space for 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 nearest (in the Euclidean sense) EEG electrode. This enables us to present the results using the familiar head-in-head plots. Dipoles that are further than $5\,\mathrm{cm}$ away from any electrode are not assigned to any region. The source activity within each ROI is averaged separately for each spatial dimension to yield a $3M$-dimensional time series. Since this time series has at most rank M, a small amount of noise is added in order to establish full rank. The resulting time series are normalized and subjected to effective connectivity analysis using GC, PDC and PSI. This yields $3M \times 3M$ connectivity matrices consisting of 3×3 blocks $R(i, j)$, which describe the interactions between the i-th and the j-th dipole in all three spatial dimensions. We define the *total flow* from the i-th to the j-th voxel as the sum over all entries of $R(i, j)$. This operation yields an $M \times M$ matrix, which is antisymmetrized and tested entrywise for being significantly different from a corresponding matrix obtained from time-reversed source time courses.

4.4.2   Results

The results of the inverse source reconstructions are depicted in the upper panels of Figure 4 (a), (b) and (c) as heat maps showing estimated dipole moment vector amplitudes averaged over time instants and repetitions (and normalized by the estimated noise amplitude in case of LCMV). The plot is overlaid with arrows representing the true interacting dipolar sources. The estimated effective connectivity between ROIs is depicted in the corresponding

lower panels as 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 estimated interactions between electrodes but interactions between source space regions-of-interest below these electrodes. Since we simulated the true signal source dipoles to lie exactly below C3 and C4, it is hence sufficient to assess whether the flow from (the region below) C3 to (the region below) C4 is present when evaluating source-connectivity head-in-head plots. All other connections are incorrect by definition and indicate mislocalization and (as a result) spurious connectivity. Since information flow is only meaningful if it is estimated between active regions, the information about the strength of the source activity is encoded in the visualization by means of the alpha (transparency) value. Here, the ROI with maximal strength is drawn with full opacity, while 10 % opacity is used for the ROI with minimal strength. Transparency values for ROIs in between are assigned using a monotonous sigmoidal nonlinearity.

The source activity estimated by WMN is spread over the entire brain. The true sources are not well separated, as the source amplitude exhibits only one local maximum, which is in between the two true dipoles. In contrast, S-FLEX estimates two dominant focal patches of activity, each of which is close to one of the two simulated interacting sources (although located slightly too deep). The localization performance of LCMV lies in between the other two. As with WMN, almost the whole brain is covered by estimated active sources. However, there are two distinguished local maxima close to the simulated interacting sources. The superior localization performance of S-FLEX compared to WMN and LCMV is empirically evidenced by lower EMD scores (see Table 1). Note that the seemingly poor localization performance of some of the methods is

explained by the fact that we localize single trial data here, which are characterized by low SNR and a large number of active noise sources in the brain. Thus the conditions under which source reconstruction is performed here are much worse than in standard event-related potentials studies, in which pre-averaging can be used to effectively cancel out noise contributions. Notably, an increase in SNR to $\gamma = 0.75$ improves the localization result of all three methods not only in terms of EMD scores, but also visibly (see also supplement).

The use of PSI on source time courses reveals the underlying interhemispheric left-to-right flow for all three inverse source reconstruction methods. In combination with S-FLEX, this flow is constrained to the true generating brain areas, namely the regions below C3 and C4, which have been correctly identified by S-FLEX. The LCMV beamformer estimates broader active areas, in which some of the activity of the simulated interacting sources seems to leak. Consequently, the left-to-right flow is observed also between area closer to neighboring electrodes of C3 and C4 then to these electrodes themselves. A similar situation is observed for WMN. Here, however, an additional occipital active region appears to transmit information from the left central area to the right central area, which is incorrect. Granger causality and partial directed coherence do not detect significant information flow on WMNE and LCMV source estimates. On the other hand, both perform similar as PSI on the S-FLEX sources, i. e., correctly estimate highly significant flow between the brain areas under C3 and C4. Notably, this result is observed for all statistical testing protocols except plain permutation testing, which incorrectly indicates bilateral flow between the areas below C3 and C4 (see supplement).

Thus, our results indicate that S-FLEX is more successful in demixing the

31

	WMN	S-FLEX	LCMV
$\gamma = 0.25$	56.0 ± 0.4	$\mathbf{27.5 \pm 0.7}$	51.1 ± 0.3
$\gamma = 0.5$	47.8 ± 0.2	$\mathbf{8.8 \pm 0.5}$	46.4 ± 0.3
$\gamma = 0.75$	36.3 ± 0.2	$\mathbf{0.4 \pm 0.1}$	42.6 ± 0.2

Table 1

 Earth-mover's distance (EMD) scores calculated between the two simulated dipo-lar signal sources and the current distributions estimated according to the weighted minimum-norm (WMN) and sparse basis field expansions (S-FLEX) inverse imaging estimates, as well as the linearly-constrained minimum-variance (LCMV) beamform-ing estimate for three different signal-to-noise ratios. Lower values indicate better source reconstruction. Entries marked in bold indicate superior performance. sources than WMN and LCMV, with the consequence that Granger-causal connectivity estimation on the sources is much improved.

[Insert Figure 4 around here]

4.5 Experiment 5: blind source separation preprocessing

4.5.1 Setting

An alternative to source reconstruction based on a physical model is blind source separation (BSS), which amounts to demixing the sensor signals solely based on statistical assumptions. The aim of this experiment is to demonstrate the application of popular BSS techniques to the problem of source connectiv-ity analysis, and to show how sensitive all methods are with respect to certain properties of the data such as the distribution of the sources/innovations. We

consider SCSA, MVARICA, CICAAR, TDSEP and JADE.

We apply all BSS methods to normalized time series associated with the strongest PCA components. Here we choose five components, which ensures that the two interacting source components are well-represented in the PCA subspace. Implementations of SCSA, MVARICA, CICAAR, TDSEP and JADE are obtained from the respective authors of Haufe et al. (2010), Gómez-Herrero et al. (2008), Dyrholm et al. (2007)[4], Ziehe and Müller (1998)[5] and Cardoso and Souloumiac (1996)[6]. The joint diagonalization in TDSEP is carried out using fast Frobenius-norm joint diagonalization (FFDIAG, Ziehe et al., 2004)[7]. The number of time lags in TDSEP is chosen to be $\tau = 100$, while it is selected from $\{1, \ldots, 9\}$ for SCSA, MVARICA and CICAAR using the Bayesian information criterion (BIC, Schwarz (1978)). For SCSA, P is selected from $\{1, \ldots, 9\}$ by evaluating the BIC criterion for the solution of the unregularized problem. Using the selected P, the regularization parameter of this method is determined by 5-fold cross-validation. A final SCSA fit using optimal values for P and λ is then performed on the complete time series for each dataset. We use the default parameters for JADE.

The result of applying BSS techniques is a decomposition of the data into five components consisting of a field pattern and a time series. The time series are normalized and analyzed using GC, PDC and PSI. Since only the outer product of a field pattern and a one-dimensional source time series defines a BSS component, there is a degree of freedom regarding the sign and scale

---

[4] http://www.machlea.com/mads/cicaar/index.html

[5] http://user.cs.tu-berlin.de/~ziehe/code/

[6] http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m

[7] http://user.cs.tu-berlin.de/~ziehe/code/ffdiag_pack.zip

of both factors (although normalization of the time series can be used to overcome the scale invariance). Moreover, the order of the BSS components is arbitrary. We therefore have to use the Kuhn-Munkres algorithm (Kuhn, 1955; Munkres, 1957) to match BSS components obtained in the various repetitions of the experiment. We use a modification of the implementation by Tichavský and Koldovský (2004)[8] employing the goodness-of-fit (GOF) score to measure distances between patterns (Haufe et al., 2010).

We obtain a global alignment of BSS components using the field patterns estimated in the first repetition of the experiment as a template, to which patterns obtained in later repetitions are transformed. That is, starting from the second repetition, the optimal pairing between the newly-obtained and the template field patterns is computed. Having found the pairing, the new field patterns are permuted and scaled to approximate the template as well as possible. The source connectivity graphs estimated by GC, PDC and PSI are also permuted accordingly. To measure the success of each BSS estimation, we compute the average goodness-of-fit score between each of the true field patterns and its best matching estimated counterpart. Finally, statistical testing using anti-symmetrization and time-reversed surrogates is employed to assess whether there is significant information flow between the estimated source time series.

To highlight the requirement of non-Gaussianity of the source MVAR innovations for blind source separation approaches (Comon, 1994; Dyrholm et al., 2007; Haufe et al., 2010), the analysis is repeated on problem instances in which the innovations are drawn from the super-Gaussian Laplace distribution as well as the sub-Gaussian uniform distribution instead of the Gaussian

---

[8] http://si.utia.cas.cz/downloadPT.htm

distribution.

4.5.2   Results

Figures 5 and 6 depict the source field patterns (averaged across repetitions) estimated by SCSA, MVARICA, CICAAR, TDSEP and JADE, as well as the connectivity estimates calculated from the associated source time series for Gaussian- and Laplace-distributed innovation terms. The estimated connectivity graphs are presented as matrices, in which red and yellow colors ($z > 0$) in the intersection of the i-th row and the j-th column denote that there is net information flow from the i-th to the j-th source, while blue and cyan colors ($z < 0$) denote the opposite case. Since there are ten possible pairwise interactions between five source time series, the z-score corresponding to a Bonferroni-corrected p-value of p = 0.05 is given by z = 2.8.

For Gaussian-distributed source innovations, the MVAR-coupled (correlated) sources model defined by Eq. (11) is not identifiable. It is hence expected that all methods must fail to systematically identify the underlying interacting sources as separate components. Interestingly, this is not the case for TDSEP, which does recover the sources. However, the goodness-of-fit is far from the optimal value of 1.0 even for high signal-to-noise ratios (see Table 2). Nevertheless, the flow from the left-hemisphere source to the right-hemisphere source is significant for all effective connectivity measures. while there is no other significant flow. In contrast to TDSEP, the insufficiently-demixed sources obtained from all other BSS approaches either lead to the detection of a large number interactions (as for SCSA, CICAAR and MVARICA), or to no estimated interaction at all (as for JADE).

For super-Gaussian (Laplace-distributed) innovations, all approaches but JADE recover the interacting sources, although there are huge differences in the separation accuracy as measured by GOF scores. SCSA and CICAAR, which are specialized approaches explicitly modeling super-Gaussian innovations, achieve highest GOF scores, followed by MVARICA and TDSEP. Since MVAR-ICA is capable of dealing with general non-Gaussian distributed innovation sequences, it outperforms all other methods for uniformly-distributed (sub-Gaussian) innovations. Graphical results for the uniform case are presented in the supplement.

To summarize, independent component analysis according to JADE did not succeed to reconstruct MVAR-coupled sources in our simulations. Although making the same assumption of independence of the sources, TDSEP is able to reconstruct MVAR-coupled sources even for Gaussian innovations to some degree. The success of TDSEP might be explained by the simple structure of our simulations, in which the "least-dependent" sources (according to TD-SEP) seem to coincide with the true interacting sources, even though these are coupled through the MVAR model. For non-Gaussian data, specialized approaches explicitly assuming MVAR-coupled sources are more successful. Compared to maximum-likelihood approaches like SCSA and CICAAR, which require explicit specification of the underlying distribution, MVARICA is more flexible, but potentially less powerful. To overcome the current confinement to super-Gaussian innovations, SCSA and CICAAR may be extended in the future to support general non-Gaussian distributions, where the "degree of non-Gaussianity" (i.e., the kurtosis) of the target distribution is inferred from the data along with the model parameters (Dyrholm et al., 2007).

[Insert Figure 5 around here]

	SCSA	CICAAR	MVARICA	TDSEP	JADE
$\gamma = 0.25$	0.27 ± 0.02	0.27 ± 0.03	0.24 ± 0.02	$\mathbf{0.39 \pm 0.02}$	0.27 ± 0.02
$\gamma = 0.5$	0.43 ± 0.03	0.53 ± 0.02	0.56 ± 0.02	$\mathbf{0.73 \pm 0.02}$	0.42 ± 0.02
$\gamma = 0.75$	0.61 ± 0.03	0.66 ± 0.02	0.69 ± 0.02	$\mathbf{0.76 \pm 0.02}$	0.62 ± 0.02

(a) Gaussian distributed innovations

	SCSA	CICAAR	MVARICA	TDSEP	JADE
$\gamma = 0.25$	$\mathbf{0.55 \pm 0.02}$	$\mathbf{0.53 \pm 0.02}$	0.46 ± 0.02	0.43 ± 0.02	0.25 ± 0.02
$\gamma = 0.5$	$\mathbf{0.94 \pm 0.01}$	0.92 ± 0.01	0.90 ± 0.01	0.71 ± 0.03	0.50 ± 0.02
$\gamma = 0.75$	$\mathbf{0.99 \pm 0.01}$	$\mathbf{0.98 \pm 0.01}$	0.92 ± 0.01	0.74 ± 0.02	0.51 ± 0.02

(b) Laplace distributed innovations

	SCSA	CICAAR	MVARICA	TDSEP	JADE
$\gamma = 0.25$	0.25 ± 0.01	0.30 ± 0.01	$\mathbf{0.56 \pm 0.02}$	0.45 ± 0.02	0.31 ± 0.02
$\gamma = 0.5$	0.43 ± 0.01	0.44 ± 0.01	$\mathbf{0.95 \pm 0.01}$	0.56 ± 0.03	0.51 ± 0.02
$\gamma = 0.75$	0.55 ± 0.01	0.56 ± 0.01	$\mathbf{0.98 \pm 0.01}$	0.60 ± 0.03	0.57 ± 0.02

(c) Uniformly distributed innovations

Table 2

Goodness-of-fit (GOF) scores calculated between the EEG field patterns of the simulated sources in the left and right hemispheres and their best matching counterparts estimated by sparsely-connected sources analysis (SCSA), combined multivariate autoregressive estimation and independent component analysis (MVARICA), convolutive independent component analysis via inverse autoregression (CICAAR), temporal decorrelation source separation (TDSEP), and joint approximate diagonalization (JADE) for three different signal-to-noise ratios, and for three different source MVAR innovation distributions. Larger values indicate better source separation. Entries marked in bold indicate superior performance.

[Insert Figure 6 around here]

5   Discussion

5.1   Estimation biases caused by weak data asymmetries

We have pointed out that connectivity measures based on Granger's predictability argument are prone to be dominated by weak data asymmetries such as different signal-to-noise ratios or different scalings. To counteract that, we proposed a new strategy based on time inversion for testing the robustness of causal analyses. Since our method amplifies strong asymmetries in the data, while suppressing weak asymmetries in a similar way as permutation testing, it increases the sensitivity of Granger-causal methods compared to permutation testing. A further strategy for rejecting false positives based on the evaluation of the asymmetry-symmetry ratio is introduced in Haufe et al. (2012).

5.2   Sensor-space analyses

Our study illustrates that the estimation of brain interaction from EEG measurements is challenging. We tested three widely-used measures of effective connectivity. All of them estimate the correct direction of information flow when being applied directly to source time series, but only PSI is able to indicate the flow from the corresponding artificial EEG, unless our novel testing approach involving time inversion is used. In addition to that, sensor-space connectivity analyses were shown to depend on the choice of the reference elec-

38

trode, since this choice affects the relative strengths with which driving and receiving sources are expressed in the sensors. A further difficulty of sensor-space analyses is that the signal-to-noise ratio affects the extent of the scalp regions between which significant interactions can be estimated. Finally, if the underlying interacting sources have more complex spread patterns, if their field patterns more heavily overlap and/or if there are more than two interacting sources, sensor-space connectivity maps may become even harder to comprehend.

5.3 Source connectivity analyses

We assessed various source reconstruction algorithms, which have been used previously in EEG-based synchronization and connectivity studies as preprocessing steps for subsequent connectivity estimation. The diffuse spatial distribution of the current density estimated by WMN and other linear inverses (Grave de Peralta-Menendez and Gonzalez-Andino, 1998; Haufe et al., 2008) prevents sufficient spatial separation of the interacting sources. The same appears to hold for the LCMV beamformer, the solution of which is also linear in the observations. More importantly, beamformers have been reported to be unable to deal with strongly correlated source signals (Van Veen et al., 1997). Since MVAR-coupling of the sources as simulated here also induces (weak) instantaneous correlations, this issue might have affected the performance of LCMV in the present study. Nonlinear distributed inverse imaging according to S-FLEX achieves a better spatial separation of the sources and at the same time a better demixing of the source signals than the application of WMN or LCMV. In our simulation, this was the key to correct connectivity esti-

mation even by Granger-causal measures. The performance of blind source separation techniques was demonstrated to crucially depend on the (type of) non-Gaussianity of the source innovations.

5.4 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 only 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 is 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. Note that, in contrast to the standard view on innovation sequences as consisting of pure noise, an interpretation as genuine brain activity justifies

to model them using non-Gaussian distributions. Moreover, considering that local brain activity is often spike-like, a super-Gaussian distribution seems to be more appropriate than a sub-Gaussian distribution.

5.5 Benchmarking

We would like to propose our data as a benchmark dataset for the evaluation of EEG-based connectivity estimation algorithms. This could help to establish a best practice procedure of such evaluations also including open data and open code.

Clearly, 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 datasets proposed here. However, it is rather challenging to accomodate the various approaches including sensor-space, inverse source reconstruction and 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

41

here quantified source estimation accuracy using the earth mover's distance in the inverse source reconstruction setting, and the goodness-of-fit in the BSS setting. A meaningful quantitative evaluation of the subsequent connectivity estimation is however only possible if the estimated sources can be matched to the true sources one-to-one, which is not always easily possible.

6   Conclusions

To sum up, we have studied simulated EEG data generated from a simple computational model of brain interaction. We have demonstrated that the interpretation of sensor-space EEG connectivity analyses is hard, if not even error-prone. In order to obtain interpretable results, it is necessary 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. Otherwise, source reconstruction will be poor, and so will be the subsequent connectivity analysis. Inverse source reconstruction methods that are suitable for source connectivity analysis should be able to spatially separate multiple distinct sources, while being applicable to entire EEG time series. Blind source separation techniques should explicitly or implicitly model the expected interactions between the sources. As a general practical rule we suggest that any methodology should be tested on appropriately simulated data before being applied to real data.

Acknowledgement

References

Astolfi, L., Bakardjian, H., Cincotti, F., Mattia, D., Marciani, M. G., De Vico Fallani, F., Colosimo, A., Salinari, S., Miwakeichi, F., Yamaguchi, Y., Martinez, P., Cichocki, A., Tocci, A., Babiloni, F., 2007. Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients. Brain Topogr 19, 107–123.

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., 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.

Attias, H., Schreiner, C. E., 1998. Blind source separation and deconvolution: the dynamic component analysis algorithm. Neural Comput 10, 1373–1424.

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 Proc Mag 18, 14–30.

Beckmann, C. F., DeLuca, M., Devlin, J. T., Smith, S. M., 2005. Investigations into resting-state connectivity using independent component analysis. Philos Trans Roy Soc B 360, 1001–1013.

Bell, A. J., Sejnowski, T. J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7, 1129–1159.

Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., Moulines, E., 1997. A blind source separation technique using second-order statistics. IEEE Trans Signal Proces 45, 434–444.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc 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.

Cardoso, J.-F., Souloumiac, A., 1996. Jacobi angles for simultaneous diagonalization. SIAM J Matrix Anal Appl 17, 161–164.

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

Comon, P., 1994. Independent component analysis, a new concept? Signal Processing 36, 287–314.

Dyrholm, M., Makeig, S., Hansen, L. K., 2007. Model selection for convolutive ICA with an application to spatiotemporal analysis of EEG. Neural Comput 19, 934–955.

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., 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., 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. 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.

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.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430.

Ioannides, A. A., Bolton, J. P. R., Clarke, C. J. S., 1990. Continous probabilistic solutions to the biomagnetic inverse problem. Inverse Probl 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., 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.

Koldovský, Z., Tichavský, P., Oja, E., 2006. Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound. IEEE Trans Neural Netw 17, 1265–1277.

Kuhn, H. W., 1955. The Hungarian method for the assignment problem. Nav Res Log 2, 83–97.

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.

Meinecke, F. C., Ziehe, A., Kurths, J., Müller, K.-R., 2005. Measuring Phase Synchronization of Superimposed Signals. Phys Rev Lett (8), 084102.

Munkres, J., 1957. Algorithms for the assignment and transportation problems. J Soc Indust Appl Math 5, 32–38.

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

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

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

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. Vol. 6 of JMLR W&CP. pp. 267–276.

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

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.

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.

Parra, L., Spence, C., 2000. Convolutive blind source separation of non-stationary sources,. IEEE Trans Speech Audio Proc 8, 320–327.

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.

Schwarz, G., 1978. Estimating the dimension of a model. Ann Stat 6, 461–464.

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.

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 Comm 11, 99–113.

Tichavský, P., Koldovský, Z., 2004. Optimal pairing of signal components separated by blind techniques. IEEE Signal Proc Let 11, 119–122.

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: Influence, causality and biophysical modeling. NeuroImage (in press).

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.

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.

Ziehe, A., Laskov, P., Nolte, G., Müller, K.-R., 2004. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. JMLR 5, 777–800.

Ziehe, A., Müller, K.-R., 1998. TDSEP–an efficient algorithm for blind separation using time structure. In: Niklasson, L. F. (Ed.), Proc. Int. Conf. on Artificial Neural Networks (ICANN '98). Springer, pp. 675–680.

Fig. 1 *(following page)*. 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. Also shown are field patterns after changing the reference to the P3 electrode (P3REF PAT 1/2).
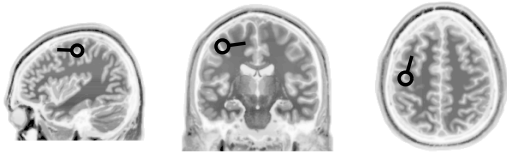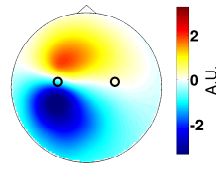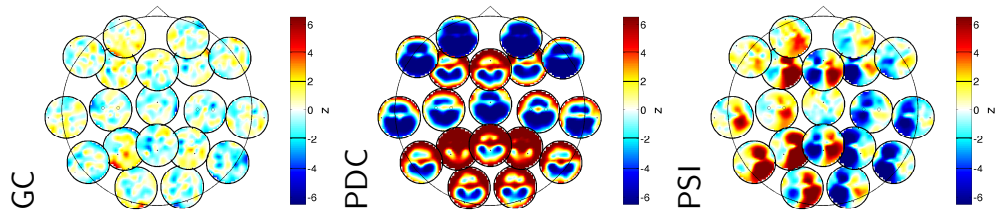
SOURCE 1   PAT 1   P3REF PAT 1
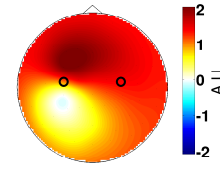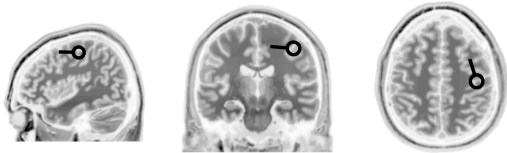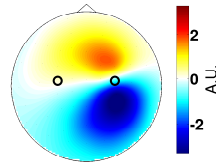
SOURCE 2   PAT 2   P3REF PAT 2
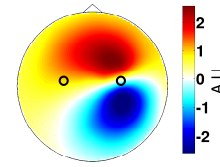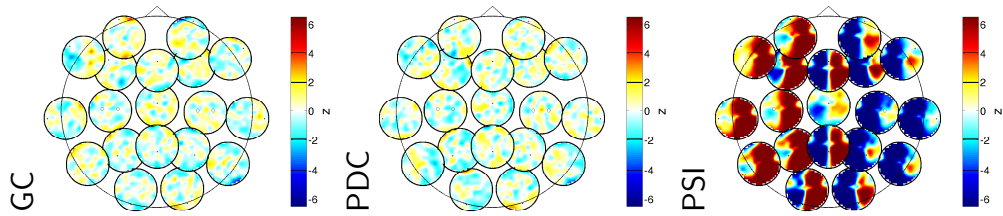
Fig. 2 *(following page)*. Comparison of effective connectivity of simulated EEG as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) using five different statistical testing strategies. Two source dipoles with tangential orientations are modeled $3\,\mathrm{cm}$ below the C3/4 electrodes. Information flow from the left (C3) to the right (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.
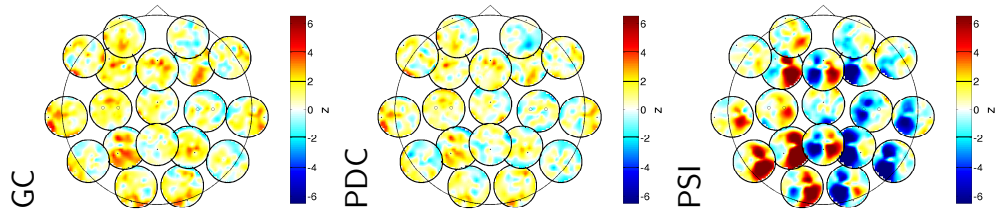
(a) Permutation testing
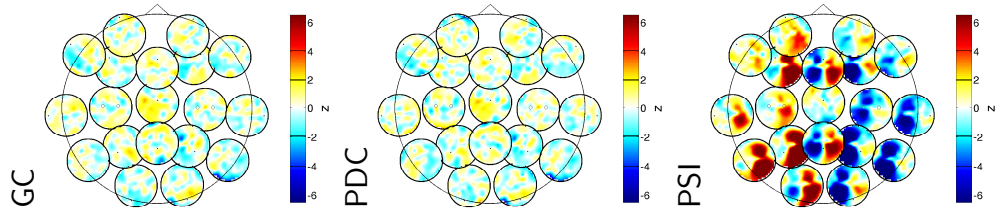
(b) Time inversion testing

(c) Antisymmetrization
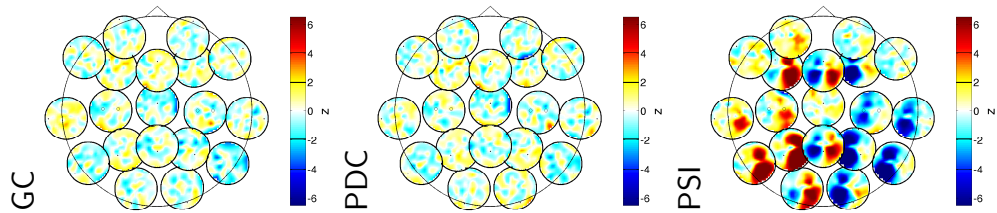
(d) Antisymmetrization + permutation testing
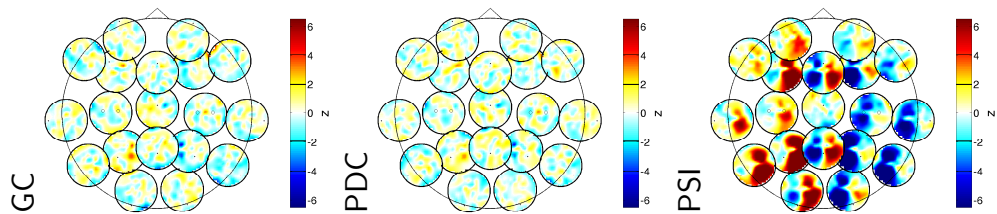
(e) Antisymmetrization + time inversion testing

Fig. 3 *(following page)*. Comparison of effective connectivity of simulated EEG as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) for re-referenced EEG data, as well as for two additional signal-to-noise ratios. Two source dipoles with tangential orientations are modeled 3 cm below the C3/4 electrodes. Information flow from the left (C3) to the right (C4) source is modeled by means of a bivariate AR model. The simulated EEG is super-imposed by non-interacting biological and sensor noise (SNR = 1). The significance of estimated interactions is assessed using antisymmetrization and time inversion testing, 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.
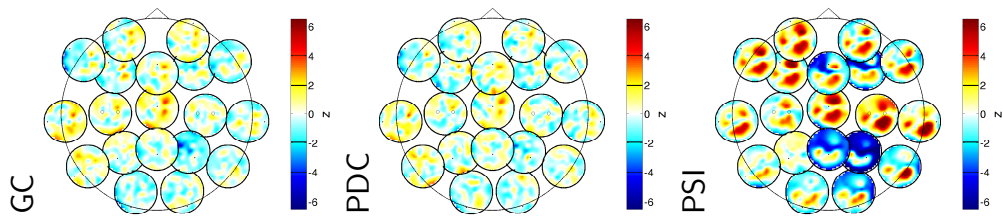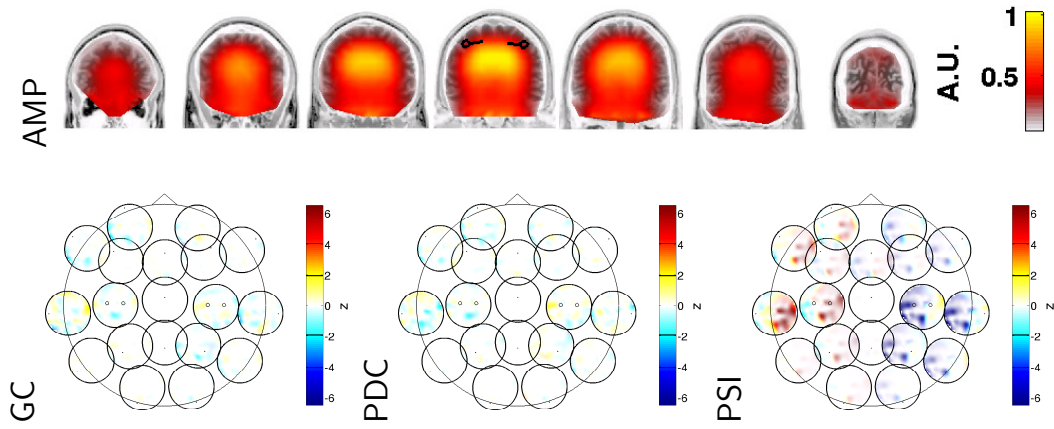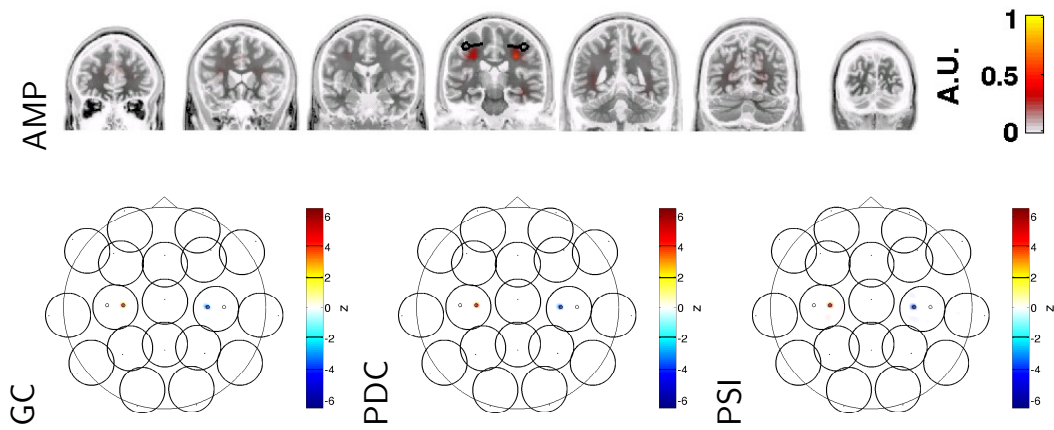
(a) P3 referenced
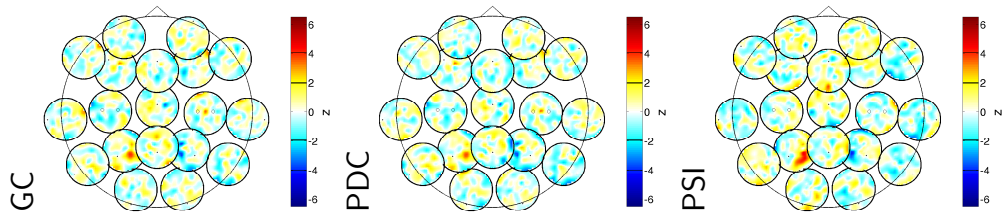
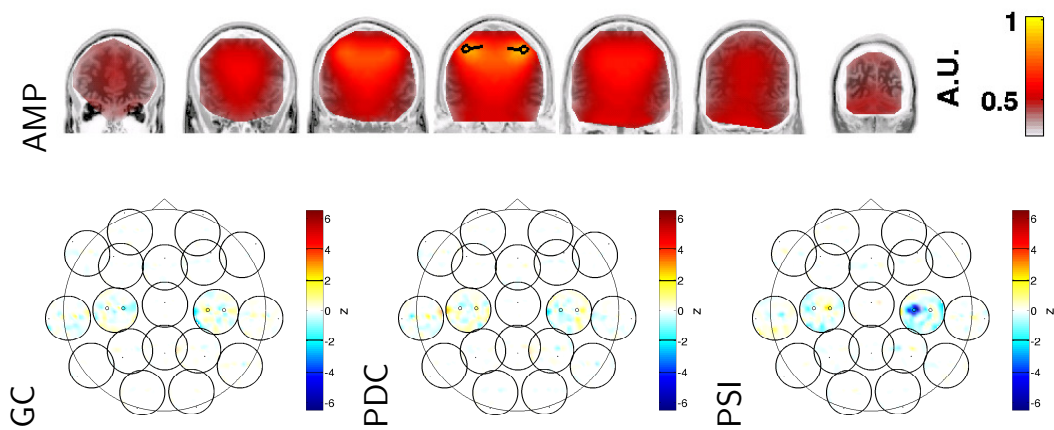(b) Low SNR ($\gamma = 0.25$)

(c) High SNR ($\gamma = 0.75$)

Fig. 4 *(following page)*. Comparison of effective connectivity of simulated EEG as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) on source estimates obtained from three inverse source reconstruction algorithms. Two source dipoles with tangential orientations are modeled 3 cm below the C3/4 electrodes. Information flow from the left (C3) to the right (C4) source is modeled by means of a bivariate AR process. The simulated EEG is superimposed by non-interacting biological and sensor noise (SNR = 1). Upper panels: average source strength (estimated dipole amplitude) per voxel. Lower panels: Estimated connectivity between regions-of-interest in source space, which are defined based on the nearest EEG electrode. The significance of estimated interactions between regions is assessed using antisymmetrization and time inversion testing, 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.

(a) Weighted minimum-norm (WMN) inverse imaging

(b) Sparse basis field expansions (S-FLEX) inverse imaging

(c) Linearly-constrained minimum variance (LCMV) beamforming

Fig. 5 *(following page)*. Comparison of effective connectivity of simulated EEG ("non-Gaussian" case) as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) on source estimates obtained from blind source separation (BSS) according to sparsely-connected sources analysis (SCSA), combined multivariate autoregressive estimation and independent component analysis (MVARICA), convolutive independent analysis via inverse autoregression (CICAAR), temporal decorrelation source separation (TDSEP) and joint approximate diagonalization (JADE). Two source dipoles with tangential orientations are modeled $3\,\mathrm{cm}$ below the C3/4 electrodes. Information flow from the left (C3) to the right (C4) source is modeled by means of a bivariate AR process with *super-Gaussian* (Laplace-distributed) innovations. The simulated EEG is superimposed by non-interacting biological and sensor noise (SNR = 1). BSS techniques are applied to the five strongest principal components. The significance of estimated interactions between demixed signals is assessed using antisymmetrization and time inversion testing, measured in terms of z-scores and visualized as matrices, where entries with red and yellow colors (z > 0) stand for information outflow and entries with blue and cyan colors (z < 0) stand for information inflow of the source marked in the respective row. 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.
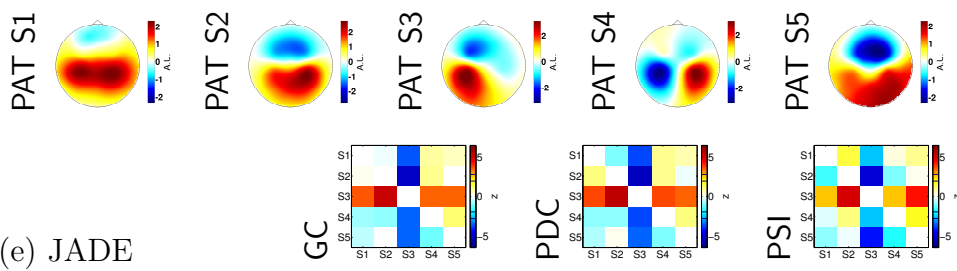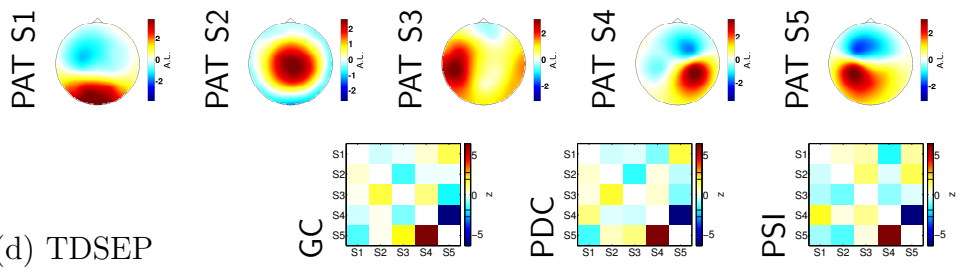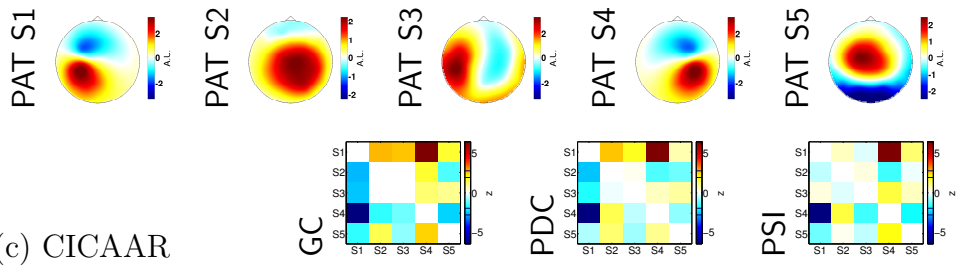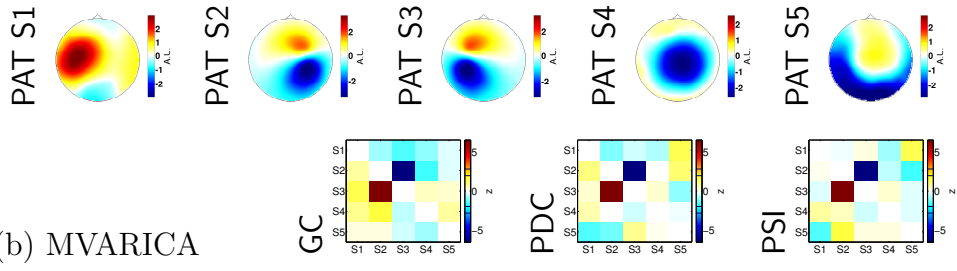
(a) SCSA

(b) MVARICA
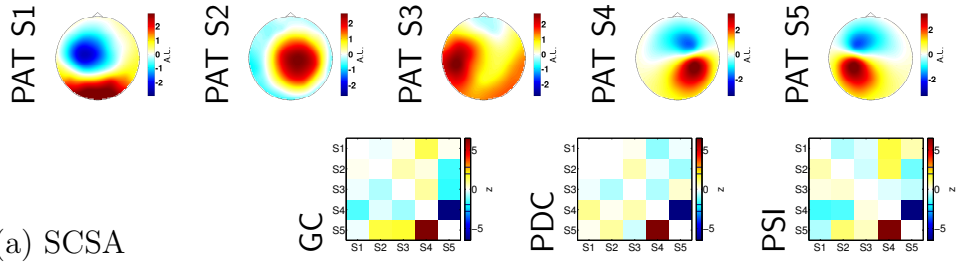
(c) CICAAR

(d) TDSEP

(e) JADE

60

Fig. 6 *(following page)*. Comparison of effective connectivity of simulated EEG ("Gaussian" case) as estimated by Granger causality (GC), partial directed coherence (PDC) and the phase-slope index (PSI) on source estimates obtained from blind source separation (BSS) according to sparsely-connected sources analysis (SCSA), combined multivariate autoregressive estimation and independent component analysis (MVARICA), convolutive independent analysis via inverse autoregression (CICAAR), temporal decorrelation source separation (TDSEP) and joint approximate diagonalization (JADE). Two source dipoles with tangential orientations are modeled 3 cm below the C3/4 electrodes. Information flow from the left (C3) to the right (C4) source is modeled by means of a bivariate AR process with *Gaussian-distributed* innovations. The simulated EEG is superimposed by non-interacting biological and sensor noise (SNR = 1). BSS techniques are applied to the five strongest principal components. The significance of estimated interactions between demixed signals is assessed using antisymmetrization and time inversion testing, measured in terms of z-scores and visualized as matrices, where entries with red and yellow colors (z > 0) stand for information outflow and entries with blue and cyan colors (z < 0) stand for information inflow of the source marked in the respective row. 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.

(a) SCSA

(b) MVARICA

(c) CICAAR

(d) TDSEP

(e) JADE