

# Quantifying the effect of demixing approaches on directed connectivity estimated between reconstructed EEG sources

Alessandra Anzolin<sup>1,2,3\*</sup>, Paolo Presti<sup>1,3\*</sup>, Frederik Van De Steen<sup>3</sup>, Laura Astolfi<sup>1,2</sup>, Stefan Haufe<sup>4</sup>, Daniele Marinazzo<sup>3,5</sup>

1. Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy

2. Neuroelectrical Imaging and Brain-Computer Interface Laboratory, Fondazione Santa Lucia IRCCS, Rome, Italy

3. Department of Data Analysis, Faculty of Psychology and Educational Sciences, University of Ghent, Ghent, Belgium

4. Institute of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany

5. The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China

\* equal contribution

Corresponding author: Daniele Marinazzo – Department of Data Analysis – Faculty of Psychology and Educational Sciences – Ghent University – 2 Henri Dunantlaan – B 9000 Ghent, Belgium

[daniele.marinazzo@ugent.be](mailto:daniele.marinazzo@ugent.be)

Tel. +32 9 2646375

Alessandra Anzolin ORCID 0000-0001-9377-5152

Paolo Presti ORCID 0000-0002-1274-538X

Frederik Van De Steen ORCID 0000-0001-5715-5992

Laura Astolfi ORCID 0000-0002-1025-7526

Stefan Haufe ORCID 0000-0003-1470-9195

Daniele Marinazzo ORCID 0000-0002-9803-0122

## Abstract

Electrical activity recorded on the scalp using electroencephalography (EEG) results from the mixing of signals originating from different regions of the brain as well as from artifactual sources. In order to investigate the role of distinct brain areas in a given experiment, the signal recorded on the sensors is typically projected back into the brain (source reconstruction) using algorithms that address the so-called EEG “inverse problem”. Once the activity of sources located inside of the brain has been reconstructed, it is often desirable to study the statistical dependencies among them, in particular to quantify directional dynamical interactions between brain areas. Unfortunately, even when performing source reconstruction, the superposition of signals that is due to the propagation of activity from sources to sensors cannot be completely undone, resulting in potentially biased estimates of directional functional connectivity. Here we perform a set of simulations involving interacting sources to quantify source connectivity estimation performance as a function of the location of the sources, their distance to each other, the noise level, the source reconstruction algorithm, and the connectivity estimator. The generated source activity was projected onto the scalp and projected back to the cortical level using two source reconstruction algorithms, Linearly Constrained Minimum Variance (LCMV) beamforming and ‘Exact’ Low-resolution Tomography (eLORETA). In source space, directed connectivity was estimated using Multi-Variate Granger Causality (MVGC) and Time-Reversed Granger Causality (TRGC), and compared with the imposed ground truth. Our results demonstrate that all considered factors significantly affect the connectivity estimation performance.

Keywords: Brain Connectivity; Source Reconstruction; Granger Causality; Modelling

## 1. Introduction

Understanding how the joint dynamics of separate brain regions gives rise to functions is a fascinating and challenging issue. Several techniques are continuously being developed to investigate these dynamics. EEG signals, due to their high temporal resolution and non-invasiveness, are often employed to investigate how brain activity is modulated in different tasks or conditions [1]–[5]. One of the main downsides of scalp EEG is the low spatial resolution resulting from head volume conduction [6], [7]. It is well known that the electrical activity measured at the sensors level is a mixture of the activity coming from all of the sources in the brain (in addition to contributions coming from outside of it). In other words, the geometry of the head and the presence of several tissues with different electrical properties between the cortex and the scalp distort the electric field generated by active neurons. Therefore, the electrodes cannot be seen as a proxy for brain regions. The high correlation between signals recorded from neighbouring electrodes at scalp level leads the majority of connectivity estimation algorithms to estimate inaccurate patterns including spurious links, and to taint results with poor interpretability. Therefore, making inferences on connectivity from the EEG signal is still not straightforward [8], [9]. In order to overcome or attenuate the volume conduction problem, several strategies and algorithms have been proposed to estimate source activities from multi-channel EEG recordings [10]. For example, simple spatial filters as the Laplacian can reduce the correlations among scalp-recorded channels induced by source mixing [6]. Another possibility is to apply Blind Source Separation (BSS) techniques that allow the separation of the data into underlying components representing the activity of potentially extended networks at the source level. Two algorithms specifically developed for Granger-causal

interactions assume that these components follow a multivariate autoregressive (MVAR) model with independent innovation noise [11], [12]. While such approaches allow one to reduce the volume conduction effect, the problem of the interpretability of the results is not completely addressed since directed dynamical influences are estimated between components and not on the cortical brain activity. Another important choice concerns the connectivity estimator. There are different kind of algorithms, which can be categorized based on whether they estimate directed [9], [13], [14] or undirected [6], [15]–[17] interactions, whether they assess bivariate [6], [15]–[17] or general multivariate [18], [19] relationships, and whether they are robust to artifacts of volume conduction [9], [16], [20], [21] or not [6], [15]. Since the behavior of undirected and bivariate metrics under volume conduction has already been extensively studied [14], [17], here we focus on the more challenging case of directed connectivity estimation within a low-dimensional multivariate system. Among the connectivity estimators that are applicable in this setting, worth noting is the class of multivariate estimators based on the concept of Wiener-Granger Causality (GC) [13]. These data-driven approaches are computationally simple and require no a priori assumption on the presence or absence of interactions between specific pairs of variables. For this reason we decided to focus on the classical time-domain measure Multi-Variate Granger Causality (MVGC) [18] and its Time-Reversed version (TRGC) that uses time-reversed data as surrogates for statistical testing to achieve a reduced sensitivity to artifacts of volume conduction [9], [19]. One of the possible solutions to achieve interpretable results is the reconstruction of brain sources prior to conducting connectivity estimation. To solve the ill-posed (as the number of sources is higher than the number of sensors) EEG inverse problem, several parametric and non-parametric methods were developed. Most of the available techniques (including minimum norm estimates and their generalizations, LORETA, VARETA, S-MAP, ST-MAP, Backus-Gilbert, LAURA, Shrinking LORETA, FOCUSS (SLF), SSLOFO and ALF for non-parametric methods and beamforming techniques, BESA, subspace techniques for parametric methods) are very well described in [22]. Other previous studies on real EEG data have already compared different inverse solutions [23], [24]. Hedrich et al. used the resolution matrix to compare source reconstruction techniques (MNE, dSPM, sLORETA, and cMEM) in terms of spatial resolution, demonstrating that MEG and hdEEG signals allow very accurate localization [24]. Among them, we decided to focus on two of the most commonly used: the ‘exact’ Low Resolution Tomography (eLORETA) [25], and the Linearly Constrained Minimum Variance (LCMV) Beamformer [26]. They are efficiently computable on long time series and make minimal assumptions. For these reasons they are most widely adopted/implemented in the currently available toolboxes. Moreover, they allowed the investigation of two different types of methods: beamformers and distributed inverses. Inverse approaches for extracting cortical waveforms and Granger-based estimators for connectivity measures can be combined to extract and investigate the human brain circuits, but an evaluation of the combined approach under different conditions is still necessary. In fact, differences between algorithms for the inverse problem solution and between connectivity estimators could result in differences in terms of the spurious connections they bring about. In this context the main objective of this study is to quantify the amount of spurious estimated links at source level in different conditions. Recently some solutions for reducing the wrong connections were proposed: orthogonalization methods as “leakage correction” [27] [28], [29], adaptive parcellation [31] [32], and inverse solution estimation in the frequency domain (BC-VARETA) [33]. In particular, although the reliability of the leakage correction was investigated in different works, other evidence demonstrated that such correction applied to inverse solution signals could produce false human connectomes [30] because it is unable to remove the

spurious secondary correlations (auto-correlation of the time series and the non-linear relationships such as power correlations).

The goal of the present study is to propose a basic yet comprehensive framework to quantify the effect of mixing and to further validate the performance of the different approaches such as those mentioned above, and more to come. We investigated the ability of source localization approaches and methods for brain connectivity estimation to provide accurate and reliable estimates as insensitive as possible to the spurious effects induced by the residual mixing of the signals at source level, and thereby allow one to interpret the obtained results in neurophysiological terms. In particular, the present study:

- Demonstrates the possibility of significantly reducing negative effects of volume conduction on connectivity estimates by employing TRGC as opposed to GC without time-reversal.
- Provides guidelines on how to correctly interpret results obtained from using the specific combination of methods tested in similar experimental scenarios (at different depth and relative position);
- Compares the effect of two of the most used source reconstruction algorithms, eLORETA and LCMV, on the accuracy of the source connectivity patterns.

We thus generated simulated data, mimicking brain source signals with an imposed connectivity pattern. The influence of volume conduction on connectivity estimates was investigated by assigning these simulated source signals to different anatomical location in the brain. The choice of the forward model used to describe the mapping from activations inside the brain to scalp potentials is the ‘New York Head’ [34]. The results of these simulations allow us to identify the best-performing combination of algorithms for the estimation of the brain activity and connectivity in different experimental conditions.

## 2. Methods

Over the past few decades, different techniques of source localization applied to EEG data were developed to provide a non-invasive yet localized estimate of brain activity [22]. Such techniques employ voltage measurements at various locations on the scalp to estimate the current sources inside the brain which best fit these data. Source localization techniques are based on the following generative model of EEG data:

$$\Phi(t) = \mathbf{L}J(t) + \varepsilon(t) \quad (1)$$

where  $\Phi(t) \in R^M$  is the EEG signal measured from  $M$  scalp locations at time  $t$ ,  $J(t) \in R^{3N}$  is the macroscopic primary current density to which we will refer to in terms of “activity” of  $N$  sources (with a 3D orientation in the space),  $\mathbf{L} \in R^{M \times 3N}$  is the leadfield matrix summarizing the propagation of the  $N$  active sources  $j$  to the EEG sensors, and  $\varepsilon(t) \in R^M$  is the measurement noise. The lead-field matrix  $\mathbf{L}$  contains information about the geometry and the conductivity of the majority of the tissues in the head (between the sensors and the sources) and can be obtained by *forward modeling*. The estimation of the sources  $J(t)$  from the measures  $\Phi(t)$  constitutes the source reconstruction and is also known as *inverse modeling*. The two modeling approaches will be described in detail below.

### 2.1 Forward Problem

The estimation of the scalp potentials resulting from the current distribution inside of the head (modelled by hypothetical dipoles) is called forward problem [22]. The model to solve it consists in a

set of assumptions and computational approximations of the geometry and electrical properties of the different tissues in the head. This model and Maxwell's equations are then used as a method to find the relationship between these magnitudes. The result is the scalp activity as a function of the current density (produced by neuronal generators) and describes how the electrical field spreads through the different layers of the head. The New York Head is an accurate finite element electrical model of the average adult human head [34]. It is based on a highly detailed nonlinear average of T1-weighted structural MR image of 152 adults provided by the International Consortium for Brain Mapping (ICBM) [35]. A detailed segmentation of this average image into six tissue types (scalp, skull, CSF, grey matter, white matter, air cavities) was performed at the native MRI resolution of 0.5 mm<sup>3</sup>. The suitability of this volume conductor model to serve as an approximation for individual heads was tested by comparison with additional BEMs and FEMs constructed for four subjects [34]. The model was evaluated for 231 electrode positions and for 75000 nodes of a mesh of the cortical surface. In the present work we used a lead field matrix  $L$  extracted from the original one, on a subset of 108 sensors.

## 2.2 Inverse Problem

The EEG inverse problem is defined as the estimation of brain sources from potentials measured on the scalp. It is an ill-posed problem because for all admissible output voltages, the solution is non-unique (since the number of active dipoles exceeds the number of sensors) and unstable because the solution is highly sensitive to small changes in the noisy data [22]. The accuracy of the source reconstruction is affected by a high number of factors including the head model errors, the source-modelling errors, and EEG noise (instrumental or biological) [36]. Several methods were developed that provide a solution to solve the inverse problem. In the present study, we focused on two methods: i) the Linearly Constrained Minimum Variance Beamformer (LCMV) and ii) the 'Exact' Low Resolution Tomography (eLORETA).

### *Linearly Constrained Minimum Variance (LCMV)*

Linearly Constrained Minimum Variance filtering (LCMV) [26], [36] is a spatial filtering method that lets brain activity coming from a specific location pass (i.e. leave unaltered), while attenuating activity originating at other locations. The output of the filter is an estimate of the activity generated by neural populations within a restricted area of the brain. A map of the brain activity as a function of location is obtained by designing multiple spatial filters, each with a different pass location, and depicting the variance of the estimated neural activity over time as a function of pass location. This spatial filtering approach falls within the general category of beamforming. The source signal at each location in the brain consists of the three dipole moments, so that three spatial filters for each location are required. The  $N \times 3$  matrix  $\mathbf{W}(\mathbf{q}_0)$  represents the transfer function of the filter for the narrowband volume element  $Q_0$  centered in  $\mathbf{q}_0$ . The output of the filter  $J$  is the inner product of  $\mathbf{W}(\mathbf{q}_0)$  and the potentials due to the  $N$  active dipoles  $\Phi$ .

$$J = \mathbf{W}^T(\mathbf{q}_0)\Phi \quad (2)$$

Under ideal conditions, the transfer function of the filter has to satisfy two conditions:

$$\mathbf{W}^T(\mathbf{q}_0)L(\mathbf{q}_0) = \mathbf{I} \quad (3)$$

$$\mathbf{W}^T(\mathbf{q}_0)L(\mathbf{q}_s) = 0 \quad (4)$$

where  $\mathbf{q}_s$  is any position other than  $\mathbf{q}_0$  within the volume of the brain. As this cannot be achieved under general conditions, *eq. 4* is replaced by the condition that the variance of the filter output is minimal. The optimal filter that achieves this is given by:

$$\mathbf{W}(\mathbf{q}_0) = [\mathbf{L}^T(\mathbf{q}_0)\mathbf{C}_\phi^{-1}\mathbf{L}(\mathbf{q}_0)]^{-1} \mathbf{L}^T(\mathbf{q}_0)\mathbf{C}_\phi^{-1} \quad (5)$$

where

$$\mathbf{C}_\phi = E[\boldsymbol{\Phi}\boldsymbol{\Phi}^T]. \quad (6)$$

is an estimate of the sensor-space covariance. Hence, after some algebra, the estimated variance becomes:

$$\widehat{var}(\mathbf{q}_0) = tr \{[\mathbf{L}^T(\mathbf{q}_0)\mathbf{C}_\phi^{-1}\mathbf{L}(\mathbf{q}_0)]^{-1}\} \quad (7)$$

To localize the electrical activity of the brain sources, the variance of the LCMV filter output is evaluated as a function of location within the volume of the brain, normalized by the output of the same LCMV filter on reference (noise) data. As the simulated signal in the present study was supposed to represent the brain's general background (resting-state) activity, the only possible noise source was measurement noise, which was modelled as spatially uncorrelated sensor noise with identity covariance matrix. Regions of large relative variance are presumably active, while regions with small relative variance can be considered inactive. Nevertheless, in the present study we only considered the estimated source time series (filter output) to assess connectivity patterns between them. Factors that may influence the accuracy of the LCMV are:

- The pass-band of the filter, indicating the spatial resolution. The spatial extent of the pass-band depends on the transfer matrices  $\mathbf{L}(\mathbf{q})$ , which in turn depend on the number of electrodes, their distribution, and source location.
- The SNR, because of the variance minimization procedure used to determine the spatial filters.

### ***'Exact' Low Resolution Tomography (eLORETA)***

'Exact' Low Resolution Electromagnetic Tomography (eLORETA) [37] is a linear inverse method characterized by spatially smooth current density. In the most general case, linear solutions to the EEG inverse problem are of the following form [38]:

$$\tilde{\mathbf{J}}(\lambda) = argmin_{\mathbf{J}} (\|\mathbf{L}\mathbf{J} - \boldsymbol{\Phi}\|_2^2 + \lambda\mathbf{J}^T\mathbf{S}\mathbf{J}) \quad (8)$$

where  $\lambda$  represents the Tikhonov regularization parameter which can be estimated through the general cross validation approach [39], and where  $\mathbf{S}$  is a symmetric positive definite weight matrix. The idea of eLORETA is to find an appropriate  $\mathbf{S}$  matrix in *eq. 8* such that the solution has zero localization error for all single point sources in the brain [37]. These weights are obtained from the following expression:

$$\mathbf{S}_i = [\mathbf{L}_i^T(\mathbf{L}\mathbf{S}^{-1}\mathbf{L}^T + \lambda\mathbf{I}_M)^+ \mathbf{L}_i]^{1/2}, \quad (9)$$

where  $\mathbf{S}_i$  for  $i = 1, \dots, N$  (number of voxels) are the diagonal elements of the weight matrix  $\mathbf{S}$ ,  $\mathbf{L}_i \in R^{M \times 1}$  represents the  $i$ -th column of lead field matrix  $\mathbf{L}$ , and the symbol  $+$  refers to Moore-Penrose pseudoinverse. The solution to (*eq. 9*) can be found by iterating four steps. First, we have to initialize the diagonal matrix  $\mathbf{S}$  with  $S_i = 1$ , for  $i = 1, \dots, N$  and then compute:

$$\mathbf{C} = (\mathbf{L}\mathbf{S}^{-1}\mathbf{L}^T + \lambda\mathbf{I}_M)^+ . \quad (10)$$

Holding  $\mathbf{C}$  fixed, we compute new weights for all the dipoles  $i = 1, \dots, N$ :

$$\mathbf{S}_i = [\mathbf{L}_i^T \mathbf{C} \mathbf{L}_i]^{1/2} \quad (11)$$

and then we return to eq. 10 until convergence. Once the  $\mathbf{S}_i$  have been estimated, the eLORETA solution is given by the following expression:

$$[\mathbf{J}]_i = \mathbf{S}_i^{-1} \mathbf{L}_i^T (\mathbf{L}\mathbf{S}^{-1}\mathbf{L}^T + \lambda\mathbf{I}_M)^+ \boldsymbol{\Phi} . \quad (12)$$

It has been suggested that eLORETA solution achieves exact localization to single test point sources under ideal (no-noise) conditions. Even if this theoretical property does not hold anymore in the presence of two or more sources (thus, in any setting involving source interaction), the studies that have evaluated this method in realistic scenarios, have found that it outperforms other linear solutions [25].

### 2.3 Multivariate Directed Connectivity Estimation

#### *Multivariate Granger Causality (MVGC)*

The concept of Granger causality [13], [40] is based on the predictability of time series. Namely, if a time series  $\mathbf{Y}_2$  contains information that improves the predictability of future values of another time series  $\mathbf{Y}_1$  above and beyond what can be predicted on the basis of  $\mathbf{Y}_1$  alone, then  $\mathbf{Y}_2$  is said to Granger-cause  $\mathbf{Y}_1$ . In other words, if the prediction error decreases by adding the past values of  $\mathbf{Y}_2$  to a regression model for predicting  $\mathbf{Y}_1$ , we can say that  $\mathbf{Y}_2$  Granger-causes  $\mathbf{Y}_1$ . In the BIVAR (bi-variate vector-autoregressive) formulation, this notion is described as follows:

$$\begin{pmatrix} \mathbf{Y}_{1,t} \\ \mathbf{Y}_{2,t} \end{pmatrix} = \sum_{k=1}^p \begin{pmatrix} A_{11,k} & A_{12,k} \\ A_{21,k} & A_{22,k} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{1,t-k} \\ \mathbf{Y}_{2,t-k} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \quad (13)$$

$$\boldsymbol{\Sigma} \equiv \text{cov} \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (14)$$

where  $p$  is the optimal order of the model and  $\mathbf{A}$  is the matrix of autoregressive coefficients. At this point, one can perform a *full regression* (eq. 15), using both time series, and a *reduced regression* (eq. 16), using only the target time series:

$$\mathbf{Y}_{1,t} = \sum_{k=1}^p A_{11,k} \mathbf{Y}_{1,t-k} + \sum_{k=1}^p A_{12,k} \mathbf{Y}_{2,t-k} + e_{1,t} \quad (15)$$

$$\mathbf{Y}_{1,t} = \sum_{k=1}^p A'_{11,k} \mathbf{Y}_{1,t-k} + e'_{1,t} \quad (16)$$

In the full regression, the dependence of  $\mathbf{Y}_{1,t}$  on the past of  $\mathbf{Y}_2$ , in addition to its own past, is encapsulated in the coefficients  $A_{12,k}$ . There is no dependence between  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  if the coefficients are null for all lags  $k$ ,  $A_{12,1} = A_{12,2} = \dots = A_{12,p} = 0$ . Prediction error estimation is based on full and reduced regression residuals. In particular  $\boldsymbol{\Sigma}'_{11} \equiv \text{var}(e'_{1,n})$  is the residual variance in the case of reduced regression and  $\boldsymbol{\Sigma}_{11} \equiv \text{var}(e_{1,n})$  is the residual variance in the case of full regression. Pairwise time-domain Granger Causality (PGC) is defined as

$$f_{Y_2 \rightarrow Y_1} = \log \frac{\Sigma'_{11}}{\Sigma_{11}} \quad (17)$$

If a dynamical influence from  $Y_2$  to  $Y_1$  exists, the value of  $f_{Y_2 \rightarrow Y_1}$  is greater than zero. However, if there are joint dependencies between  $Y_1$  and  $Y_2$  and a third time series, e.g.  $Y_3$  then spurious Granger-causal influences may be reported using PCG. This issue can emerge, for example, when there is no direct influence  $Y_2 \rightarrow Y_1$  but there are (possibly lagged) dependencies of  $Y_1$  and  $Y_2$  on  $Y_3$ . To overcome this problem, Barnett and Seth propose a different way to compute GC, introducing the so called *Pairwise Conditional Granger Causality (PWCGC)*, which conditions out common dependencies between variables before estimating pairwise GC scores, provided such dependencies are present in the data [18]. The MVAR model is a generalization of BIVAR to three or more time series. In the case of three variables, the corresponding full regression (eq. 18) and reduced regressions (eq. 19) take the following forms:

$$Y_{1,t} = \sum_{k=1}^p A_{11,k} Y_{1,t-k} + \sum_{k=1}^p A_{12,k} Y_{2,t-k} + \sum_{k=1}^p A_{13,k} Y_{3,t-k} + e_{1,t} \quad (18)$$

$$Y_{1,t} = \sum_{k=1}^p A'_{11,k} Y_{1,t-k} + \sum_{k=1}^p A'_{13,k} Y_{3,t-k} + e'_{1,t} \quad (19)$$

$$f_{Y_2 \rightarrow Y_1 | Y_3} = \log \frac{\Sigma'_{11}}{\Sigma_{11}} \quad (20)$$

Here,  $f_{Y_2 \rightarrow Y_1 | Y_3}$  may be read as “the degree to which the past of  $Y_2$  helps to predict  $Y_1$ , over and above the degree to which  $Y_1$  is already predicted by its own past and the past of  $Y_3$ ”. In our simulation study, we are going to use this variant of MVGC. Additionally, it is worth noting that we use the state-space formulation of Granger causality, which eliminates the bias due to the fact that the reduced model is VARMA (Vector Auto Regressive Moving Average) and not VAR [41]. This state-space formulation can also prove convenient when trying to solve the source reconstruction and the connectivity between sources at the same time [42].

### ***Time Reversed Granger Causality (TRGC)***

Granger-causal estimators are prone to detect spurious influences not only in the presence of hidden common drivers but also in the presence of additive correlated noise [9], [14], [17], [19], [43]. Correlated noise is ubiquitous in EEG data, which are by their very nature linear mixtures of contributions from different sources. Since this mixing process cannot be fully undone using source imaging techniques, it poses a serious problem for EEG-based brain connectivity analysis using GC. To overcome the problem of spurious connectivity, Haufe et al. proposed time-reversal [9], [19]. The intuitive idea behind this approach is that, if connectivity is defined based on temporal delays, directed influence should be reduced (if not reversed) if the temporal order is reversed. This is often contrasted by the observation that two signals that are correlated but non-interacting often appear spuriously Granger-causal interacting no matter if GC is applied on the original or time-reversed data. If, however, GC estimates obtained on original and time-reversed data are contrasted with each other, the instantaneous influence of volume conduction can be removed, and the false detection of connectivity can be avoided. Here we define TRGC based on the *Granger-scores* defined in eq. (17) [14]. The residual covariance matrix of the time-reversed AR process (defined on time-reversed data  $\widetilde{Y}_1(t)$ ,  $\widetilde{Y}_2(t)$ ) is denoted by:



$$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{bmatrix} \quad (21)$$

for the full model, and by

$$\tilde{\Sigma}' = \begin{bmatrix} \tilde{\Sigma}'_{11} & \tilde{\Sigma}'_{12} \\ \tilde{\Sigma}'_{21} & \tilde{\Sigma}'_{22} \end{bmatrix} \quad (22)$$

for the reduced model. Consequently, the Granger score on time-reversed data is defined as

$$f_{\tilde{Y}_2 \rightarrow \tilde{Y}_1} = \log \left( \frac{\tilde{\Sigma}'_{11}}{\tilde{\Sigma}_{11}} \right) \quad (23)$$

Finally, it is recommended [19] to transform the original GC scores in eq. (17) and eq. (23) into *net* GC scores

$$f_{Y_2 \rightarrow Y_1}^{net} = f_{Y_2 \rightarrow Y_1} - f_{Y_1 \rightarrow Y_2} \quad (24)$$

and

$$f_{\tilde{Y}_2 \rightarrow \tilde{Y}_1}^{net} = f_{\tilde{Y}_2 \rightarrow \tilde{Y}_1} - f_{\tilde{Y}_1 \rightarrow \tilde{Y}_2} \quad (25)$$

One way to test whether a Granger-causal influence can be explained by a trivial linear mixture of independent sources is to check whether the directionality of GC flips when time-reversing the signals. This approach is called conjunction-based time-reversed GC [19]. A connection can be regarded as significant (not caused by source mixing) according to conjunction-based TRGC if the GC scores for original and reversed data have opposing directions and are both significant:

$$f_{Y_2 \rightarrow Y_1} > 0 \wedge f_{\tilde{Y}_1 \rightarrow \tilde{Y}_2} > 0 \quad (26)$$

This is the definition adopted in the present paper. When net GC scores are used, the sign of the net GC score is required to flip for time-reversed data in comparison to the original data, and the conjunction-based TRGC criterion becomes  $f_{Y_2 \rightarrow Y_1}^{net} > 0 \wedge f_{\tilde{Y}_2 \rightarrow \tilde{Y}_1}^{net} < 0$ . Simulations have shown that TRGC leads to a reduced number of false connections, compared to original GC and its variants. Theoretical work presented in [19] has moreover shown that :

- The application of time reversal to any connectivity measures that is based on second order statistics - which, besides GC and pairwise-conditional GC also includes its direct extension to frequency domain (spectral GC) and the popular frequency-domain measures partial directed coherence (PDC) and directed transfer function (DTF), among others - prevents the spurious detection of connectivity on mixtures of independent sources that would otherwise be highly likely.
- The application of time reversal to Granger causality (that is, the use of TRGC) is guaranteed to always yield the correct direction of interaction for systems that do not contain causal loops and are noise-free.

## 2.4 Statistical assessment of significant connections

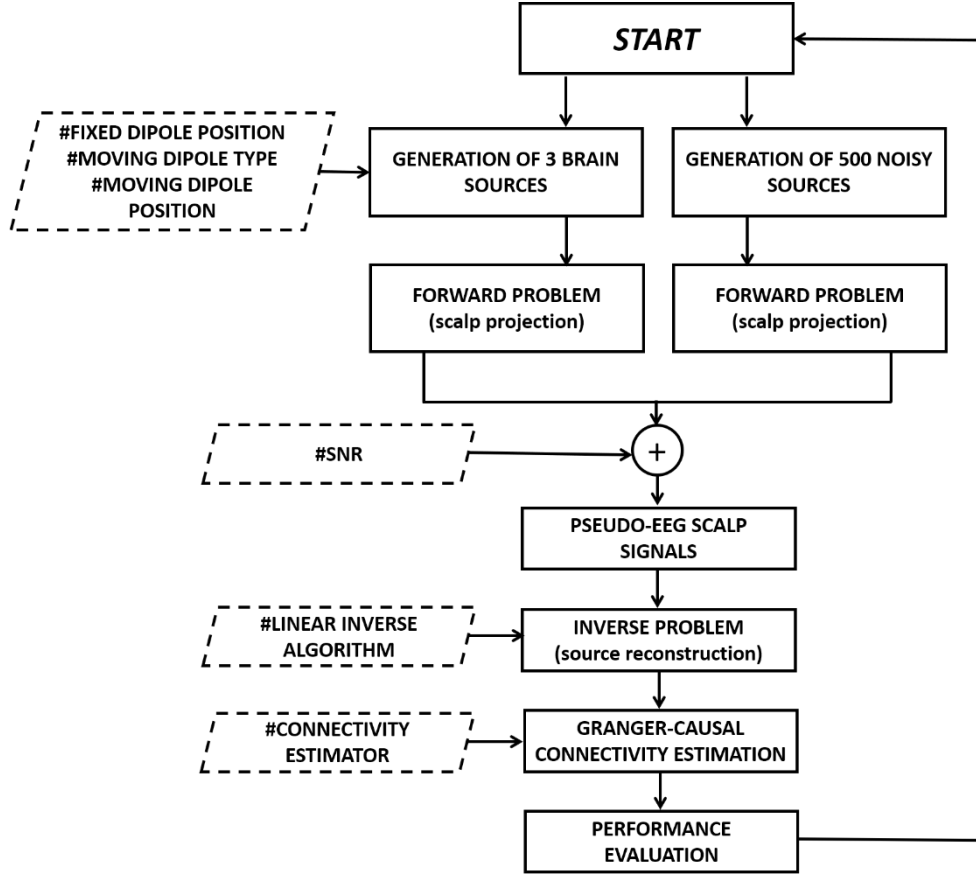
The standard way to assess the statistical significance of Granger scores is a likelihood ratio test, which can be derived from large-sample theory [44]. If  $\dim(X_1) = n_{x1}$ ,  $\dim(X_2) = n_{x2}$  and  $\dim(X_3) = n_{x3}$  (with  $n_{x1} + n_{x2} + n_{x3} = n$ ) then the difference in the number of parameters between the full model and the nested reduced model (see *eq. 19*) is just  $d \equiv p_{n_{x1} n_{x2}}$ . Thus, under the null hypothesis of zero Granger-causal influence, the GC estimator scaled by sample size,  $(m - p) F_{X_2 \rightarrow X_1 | X_3}(u)$ , has an asymptotic  $\chi^2$  distribution (where  $m$  is the order of integration and  $p$  the order of the model). Under the alternative hypothesis, the scaled estimator has an asymptotic noncentral -  $\chi^2(d; \nu)$  distribution, with non-centrality parameter  $\nu = (m - p) F_{X_2 \rightarrow X_1 | X_3}(u)$  equal to the scaled actual influence. Note that this standard statistical test, when applied to the conventional GC score, is capable to distinguish actually present GC from results obtained due to random signal fluctuations in the absence of GC. It is *not* capable of distinguishing actual GC effects that are due to genuine time delayed interaction from actual GC effects that are solely due to source mixing (or, more generally, additive mixed noise) in the absence of genuine time-delayed interaction. To test for the latter, time-reversed GC needs to be assessed. This is done here by evaluating the significance of the GC scores obtained on time-reversed data using the same statistical approach, and by evaluating the conjunction based TRGC criterion (26). In this work, we used an alpha level of 0.05, FDR corrected [45], and the corresponding p value is taken as a threshold to binarize the connectivity graph.

## 2.5 Simulation Framework

The simulation study developed for investigating the effects of the volume conduction on connectivity estimation accuracy and reliability is composed by the following main steps:

- Generation of brain signals with an imposed connectivity pattern
- Forward modeling
- Inverse modeling
- Connectivity estimation
- Performance evaluation

An overview of the simulation framework, with all of the considered factors, is shown in *fig.1*.



**Figure 1** - Block diagram illustrating the main steps of the simulation framework.

### 2.5.1 Simulated time series generation

Pseudo-EEG data were generated employing the toolbox developed by Haufe and Ewald, described in detail in [46]. We simulated three time series of 1000 samples and one connection. In particular, the source dynamic was generated by using the following multivariate autoregressive (MVAR) model with order 2:

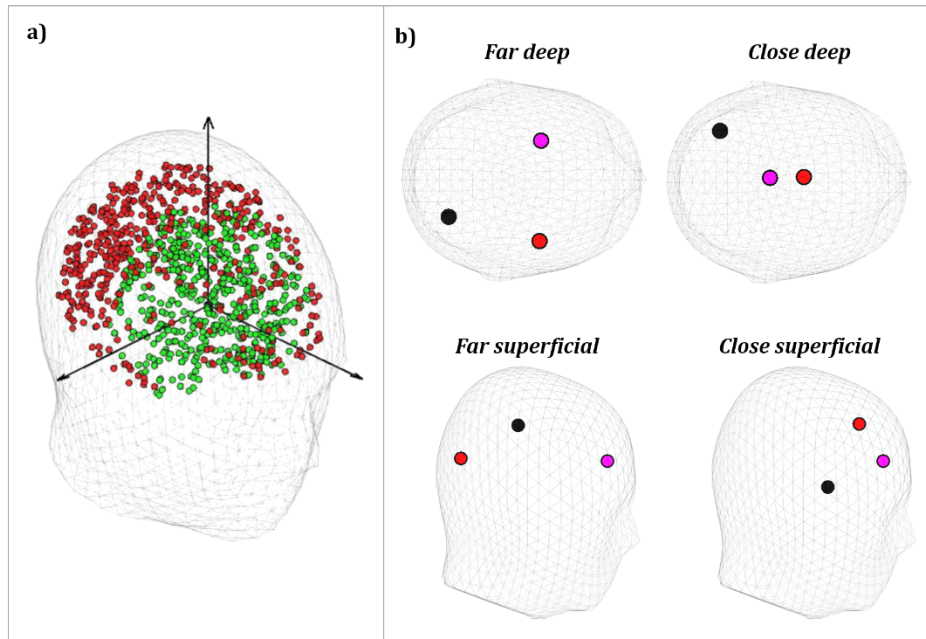
$$\begin{bmatrix} y_s(t) \\ y_r(t) \\ y_n(t) \end{bmatrix} = \sum_{p=1}^P \begin{bmatrix} a_{11}(p) & 0 & 0 \\ a_{21}(p) & a_{22}(p) & 0 \\ 0 & 0 & a_{33}(p) \end{bmatrix} \begin{bmatrix} y_s(t-p) \\ y_r(t-p) \\ y_n(t-p) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ e_3(t) \end{bmatrix} \quad (27)$$

where  $y_s(t)$ ,  $y_r(t)$  and  $y_n(t)$  are the *Sender*, *Receiver*, and *Non-Interacting dipole*, to respectively indicate the driving dipole, the receiving dipole, and the independent dipole. Each of them represents an active source contributing to the simulated EEG scalp potentials;  $a_{ij}(p)$  with  $i, j \in \{1, 2, 3\}$  and  $p \in \{1, \dots, P\}$  are the autoregressive coefficients;  $a_{21}$  is the coupling strength imposed from the *Sender* to the *Receiver*. Both the autoregressive components and the off-diagonal element were randomly chosen within the range  $[0.3 \ 1]$ .  $e_i(t)$  with  $i \in \{1, 2, 3\}$  represents uncorrelated standard normal distributed realizations of an innovation process. We also generated 500 mutually statistically independent sources of pink “brain noise”, generated by scaling the amplitude spectrum of random white Gaussian noise with the factor  $1/f$  using the Fourier transform and its inverse. These 500 signals are meant to represent the brain

background activity. The selected 1000 time points followed a discarded transient of previous 1000 points to ensure stability and independence from the initial conditions.

### 2.5.2 Simulated time series location

Brain activity was modelled with 1006 electric equivalent dipoles, equally distributed within the brain. Using the New York Head model, we obtained the dipole positions by subsampling the 75000 MNI coordinates available in the ICBM152 model. In panel a of *fig. 2* we showed all of the 1006 possible dipole locations.



**Figure 2** - Panel a) shows the 1006 locations uniformly distributed in the brain in which the activity was modelled. Green circles represent the 500 “deep” dipoles; the remaining red ones represents “superficial” dipoles. Panel b) presents the four conditions for the two fixed active dipoles, which are the red (Sender) and the purple (Receiver) one. The black circle represents the Non-Interacting dipole (noise).

For each simulation, we fixed the position of the Sender and Receiver on four possible configurations (represented in *fig.2b*) defined from the combination of the following factors:

- Depth of the dipoles: “superficial” (distance from the origin  $>6.5$  cm) or “deep” (distance from the origin  $<6$ cm);
- Distance between the dipoles: “far” (relative distance  $>8$ cm) or “close” (relative distance  $<5$ cm).

According with [34], the origin was located at the anterior commissure. The third dipole, the Non-Interactive one, that we will call *moving dipole*, changes its location at each iteration on the remaining 1004 positions. Two different cases were analysed by changing the moving dipole. In the first case, it is the Non-Interacting dipole, thus the connection is fixed. In the second case, the moving dipole is the Receiver; thus, the relative locations of Sender and Receiver varies across repetitions. The 500 additional noisy elements were randomly distributed within the brain. With regard to the sources direction, we assumed that dipoles are oriented perpendicularly to the cortical surface following the simulation framework proposed by Haufe and Ewald [46].

### 2.5.3 Pseudo-EEG signal generation.

After the signals generation, the time series representing both the source activity and the noise were projected onto 108 EEG electrodes defined by the New York Head model described in section 2.1 (forward problem). Then, they were summed according to following equation:

$$\mathbf{x}^{brain}(t) = \gamma * \frac{\mathbf{x}^{act_s}(t)}{\|\mathbf{x}^{act_s}(t)\|_F} + (1 - \gamma) * \frac{\mathbf{x}^{noise_s}(t)}{\|\mathbf{x}^{noise_s}(t)\|_F} \quad (28)$$

The signal-to-noise (SNR) parameter  $\gamma$  is drawn from the interval [0.1, 0.9]. In particular, in the present simulation study we selected  $\gamma$  equal to 0.5, 0.7 and 0.9 to simulate low, medium, and high SNR respectively. All of the variables in the formula refer to the signals at the scalp sensors level. In particular,  $\mathbf{x}^{act_s}$  and  $\mathbf{x}^{noise_s}$  are the projections of the active sources signals and of the brain noise sources activity respectively, thus their dimensions are [108 (number of sensors) x 1000 (signal length)].  $\|\mathbf{x}(t)\|_F$  is the Frobenius norm of the multivariate time series  $\mathbf{x}(t)$  (the square-root of the sum of the squared activity across time and space). The specific SNR definition used for this study represents the brain noise level and allows deep sources to have the same strength as shallow sources. It would make the discovery of deep sources more difficult but, on the other hand, as SNR is computed not per-source, but for all three sources together, this imbalance is mitigated. Other SNR definitions would allow for a fairer view of the difference in discovery between superficial and deep sources, but other biases would remain: as we only use a fixed set of 4 locations, the results would also depend on the correlation of the scalp patterns of these four particular sources. For the purpose of this study and given the fact that the separation between noise and signal in the context of EEG source reconstruction is not perfectly defined, we prefer to use this recipe. Finally, in order to simulate the measurement noise, spatially and temporally uncorrelated signals are added to  $\mathbf{x}^{brain}(t)$  with an imposed  $\gamma$  equal to 0.9. The overall pseudo-EEG data is defined from the following equation:

$$\mathbf{x}(t) = 0.9 * \frac{\mathbf{x}^{brain}(t)}{\|\mathbf{x}^{brain}(t)\|_F} + 0.1 * \frac{\mathbf{x}^{noise}(t)}{\|\mathbf{x}^{noise}(t)\|_F} \quad (29)$$

where  $\mathbf{x}^{noise}$  is the white uncorrelated noise.

### 2.5.4 Source reconstruction and directed connectivity estimation

The simulated pseudo-EEG signal was projected onto the cortical surface using two different inverse solutions: LCMV and eLORETA. The regularization parameter to be set in the eLORETA algorithm was chosen by means of a cross-validation approach equal to 0.01. In cortical source space, directed connectivity was estimated with MVGC and TRGC at the location of the three simulated active dipoles, and the statistical significance of the estimated connections was assessed.

### 2.5.5 Performance parameters

The quantitative evaluation of the connectivity estimation was performed by means of three parameters: the False Positive Rate (FPR), the False Negative Rate (FNR), and the Area Under ROC Curve (AUC). Such parameters were computed by comparing the estimated connectivity pattern with the imposed ground-truth. A false positive (FP) is an estimated (statistically significant) connection that is not present in the simulated data, while a true negative (TN) is an absent simulated connection that is

correctly estimated as being absent. The FPR (see eq. 30) is the number of false positives normalized by the number of absent connections. The FPR is thus defined as in the follows:

$$FPR = \frac{FP}{FP+TN} . \quad (30)$$

The FNR quantifies the percentage of missed (not statistically significant) connections (referred to as false negatives, FN) that are present in the simulated data relative to the total number of actually present simulated connections. The latter number is given as the sum of false negatives and true positives (TP, referring to actually present connections that are also estimated to be present). The FNR is thus defined as follows:

$$FNR = \frac{FN}{FN+TP} \quad (31)$$

In this study, the total number of possible connections is six (2 possible directions for three distinct pairs of variables). As only one interaction was modelled, FN+TP equals one, while the number of absent connections (FP + TN) is equal to five.

The AUC is a measure of binary classification accuracy, which is applied here to the problem of distinguishing between interacting and Non-Interacting signals. It takes into account both the FPR and FNR across the entire range of all possible thresholds for the connectivity measure; therefore, it is independent of a specific significance threshold. The AUC was derived from the Wilcoxon-Mann-Whitney statistic test (also known as the Wilcoxon rank-sum test statistic) [47].

## 2.6 Statistical Analysis

In order to statistically evaluate the accuracy of the employed algorithms in reconstructing the sources activity and estimating brain networks, a four-way ANalysis Of VAriance (ANOVA) was computed. The main within factors were:

- the fixed dipoles position (DIP\_POS) with 4 levels: Close Deep, Close Superficial, Far Deep, Far Superficial;
- the adopted inverse methods (L\_INV\_METH) with 2 levels: eLORETA, LCMV;
- the connectivity estimator (EST\_TYPE) with 2 levels: MVGC, TRGC;
- the signal-to-noise ratio (SNR) defined by 3 levels of  $\gamma$ : 0.5, 0.7, 0.9 which in the next sections will be identified as “low”, “medium” and “high” value of SNR.

The dependent variables were the three introduced performance parameters (FPR, FNR and AUC) averaged on the 1,004 possible location of the moving dipole. The simulation was repeated 100 times for each experimental condition. Additionally, a post hoc analysis was performed in order to highlight the significant comparisons between the various level of the included factors and their interaction, using Tukey's range test.

## 2.7 Topographical visualization of the results

As described in the previous paragraph, the ANOVA investigates the performance parameters averaged for more than one thousand possible locations of the moving dipole. In order to obtain a detailed overview on the variations of the accuracy of the estimation as a function of the position of the moving dipole, we averaged the performance parameters across the 100 repetitions and reported the obtained

results in 3D brain maps. The color of each one of the 1004 dipoles codes for the value of the FPR. We do not report the maps obtained for the false negatives because their amount is always very low (less than 5%). With the aim to summarize the complex information contained in the brain maps, we also calculated the FPR as function of the Euclidean distance between Sender and Receiver as well as between Sender and Non-Interacting dipole for each SNR level, inverse approach, and connectivity estimator. Each point in the figure (for each SNR level and linear inverse method) is the average of the FPR computed when the moving dipole is placed in a specific range of distance from the Sender (i.e. all the moving dipole positions 2 to 3 cm far from the Sender). The position of the fixed dipole (either Receiver or Non-Interacting dipole) in these analyses were *far* and *superficial* to explore the fluctuations of the performance parameters also in the best experimental condition.

### 3. RESULTS

#### 3.1 Statistical assessment of recovery of simulated connections

The results of the four-way ANOVA computed separately for the three performance parameters are reported in Table I. A four-way ANOVA consists of fifteen separate multiple tests (four main effects, six two-way interactions, four three-way interactions, and one four-way interaction). Therefore, a correction for multiple comparisons (Bonferroni-Holm) was performed and the post-hoc tests are available for consultation at <https://zenodo.org/record/1157196#.WmZ1K6jiY2w> (open-access repository).

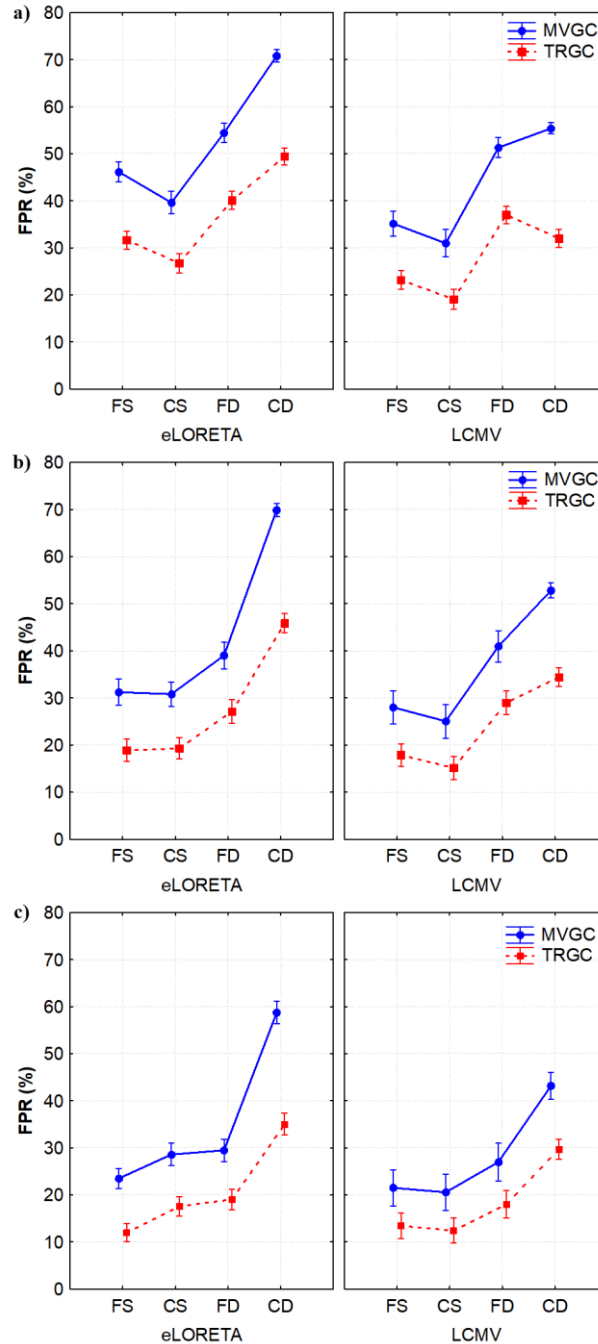
<i>Factors</i>	<i>FPR</i>	<i>FNR</i>	<i>AUC</i>
L_INV METH (1,99)	<b>85.69</b> $p=1*10^{-5}$	<b>70.04</b> $p=1*10^{-5}$	<b>39.66</b> $p=1*10^{-5}$
EST_TYPE (1,99)	<b>1694.6</b> $p=1*10^{-5}$	<b>5.04</b> $p=0.027$	<b>1616.6</b> $p=1*10^{-5}$
SNR (2,198)	<b>303.87</b> $p=1*10^{-5}$	<b>67.94</b> $p=1*10^{-5}$	<b>502.81</b> $p=1*10^{-5}$
DIP_POS (3,297)	<b>881.5</b> $p=1*10^{-5}$	<b>60.97</b> $p=1*10^{-5}$	<b>1213.41</b> $p=1*10^{-5}$
L_INV METH x EST_TYPE (1,99)	<b>16.93</b> $p=0.8*10^{-4}$	<b>7.83</b> $p=0.006$	<b>13.64</b> $p=0.3*10^{-4}$
L_INV METH x SNR (2,198)	<b>30.31</b> $p=1*10^{-5}$	<b>104.005</b> $p=1*10^{-5}$	<b>1.41</b> $p=0.25$
EST_TYPE x SNR (2,198)	<b>88.09</b> $p=1*10^{-5}$	<b>2.91</b> $p=0.056$	<b>84.52</b> $p=1*10^{-5}$
DIP_POS x L_INV METH (3,297)	<b>77.42</b> $p=1*10^{-5}$	<b>88.14</b> $p=1*10^{-5}$	<b>12.99</b> $p=1*10^{-5}$
EST_TYPE x DIP_POS (3,297)	<b>270.61</b> $p=1*10^{-5}$	<b>2.79</b> $p=0.04$	<b>228.33</b> $p=1*10^{-5}$
DIP_POS x SNR (6,594)	<b>80.26</b> $p=1*10^{-5}$	<b>92.6</b> $p=1*10^{-5}$	<b>57.88</b> $p=1*10^{-5}$
L_INV METH*EST_TYPE*SNR (2,198)	<b>42.52</b> $p=1*10^{-5}$	<b>6.35</b> $p=0.002$	<b>26.63</b> $p=1*10^{-5}$
L_INV METH*EST_TYPE*DIP_POS (3,297)	<b>37.22</b> $p=1*10^{-5}$	<b>3.14</b> $p=0.025$	<b>37.07</b> $p=1*10^{-5}$
L_INV METH*SNR*DIP_POS (6,594)	<b>13.42</b> $p=1*10^{-5}$	<b>78.3</b> $p=1*10^{-5}$	<b>47.3</b> $p=1*10^{-5}$
EST_TYPE*SNR*DIP_POS (6,594)	<b>2.83</b> $p=0.01$	<b>0.38</b> $p=0.89$	<b>2.48</b> $p=0.022$
L_INV METH x EST_TYPE x SNR x DIP_POS (6,594)	<b>67.86</b> $p=1*10^{-5}$	<b>5.38</b> $p=0.2*10^{-4}$	<b>73.68</b> $p=1*10^{-5}$

**Table I** - Results of the four-way ANOVA (*F* values and *p* values) computed considering as dependent variables *FPR*, *FNR*, and *AUC* and as within main factors the type of inverse algorithm (*L\_INV METH*), the connectivity estimator (*EST\_TYPE*), the SNR, and the position of the fixed dipoles (*DIP\_POS*). In the column "Factors", the degrees of freedom are also reported.

All individual factors have a significant effect on the FPR, FNR, and AUC (while this is not the case for all the interactions). In the following, we show a graphical depiction of the means of the four-way interaction factor (L\_INV METH x EST\_TYPE x SNR x DIP\_POS) for each investigated performance measure.

### False Positive Rate

Figure 3 shows means obtained for the FPR for different levels of SNR ( $\gamma$ ) and dipole positions when specific algorithms for the inverse solution and connectivity estimation are employed.



**Figure 3** - Means associated with the four-way interaction factor (L\_INV METH x EST\_TYPE x SNR x DIP\_POS) of the ANOVA performed on the FPR. Each panel corresponds to a specific value of the SNR parameter  $\gamma$ : 0.5 (panel a), 0.7 (panel b), 0.9 (panel c). For each panel, there are two graphs associated with the two different inverse solutions: eLORETA on the left and LCMV on the right. X-axes report the four levels of the factor DIP\_POS: far/superficial (FS), Close/Superficial (CS), Far/Deep (FD) and Close/Deep (CD). Colours code for the different connectivity estimators and whiskers represent 95% confidence intervals.



These graphs show how the two different inverse methods and the location of the fixed dipoles influence the amount of false positive connections when the estimation is performed with different connectivity estimation algorithms for different levels of SNR.

**Connectivity Estimator:** As expected, we found that the amount of false positive decreases when the connectivity pattern is extracted by means of TRGC. FPR associated with the TRGC is significantly lower (Tukey test) with respect to the other method independently of the dipoles position, the SNR level, and to the inverse algorithm (see all of the subplots).

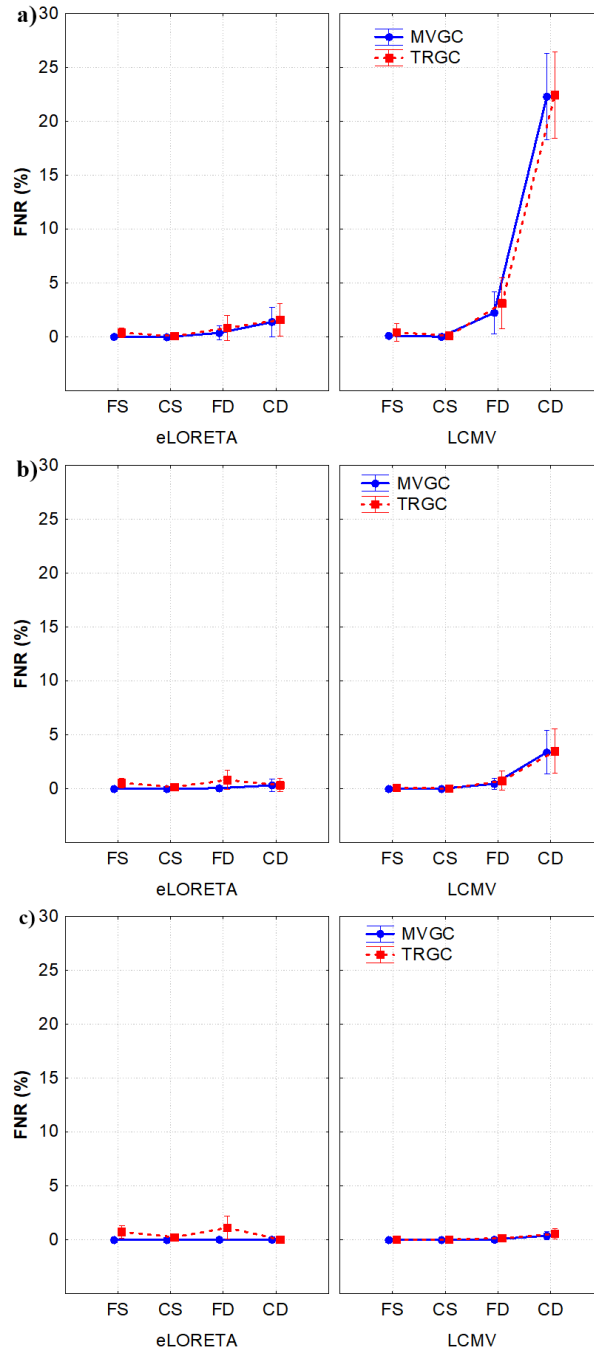
**Inverse Algorithm:** For each panel, we can compare the performance associated with the different inverse solutions comparing the two subplots. Regardless of the SNR, the LCMV algorithm (on the right) for source reconstruction has globally better performance than eLORETA (on the left) for all of the three SNR values. The post hoc analysis reveals a significant increase of the FPR for eLORETA, compared to LCMV, in all of the considered conditions of SNR, dipoles position, and connectivity estimator. Only in the most advantageous configuration, when  $\gamma$  is equal to 0.9, indicating high SNR, and the linked dipoles are in the *Far/Superficial* configuration, such difference is not significant. In the worst case, corresponding to the *Close/Deep* configuration, the FPR is considerably high, especially for the eLORETA reconstruction, where it exceeds 70%, and the difference between the performances of the two inverse methods appears to be emphasized.

**Fixed dipoles position:** It is worth to note how performance critically depends on the position of the fixed dipoles. Independently of the employed inverse algorithm and connectivity estimator, the ANOVA suggested that when they are located deep in the brain the number of false positives significantly increases. When the fixed dipoles are superficial and  $\gamma$  is equal 0.7, the relative distance (close/far) does not have a significant influence on the FPR. For the highest SNR levels, the ANOVA highlighted a significant increase of the FPR comparing the “superficial” conditions with the “deep” conditions and, in particular for eLORETA, suggesting that the optimal configuration for the source reconstruction is given by far and superficial dipoles. For the very low SNR value of 0.5, for all the L\_INV METH and EST\_TYPE levels, the statistical test revealed a significant decrease of FPR in the *Close/Superficial* case relative to the *Far/Superficial* case.

**SNR:** In all considered conditions, the test indicates a significant improvement of performance when the simulated SNR is higher. In more in detail, when the SNR level is 0.9, the amount of false positives is less than 30% in all cases except for the *Deep/Close* condition. The analysis of FPR suggests that the best combination of factors is given, for all of the considered SNR levels, by: i) dipoles located superficial in the brain and not too close; ii) LCMV as algorithm for the inverse problem solution, and iii) TRGC as connectivity estimator. Only in this case the percentage of false positives reached low values (around 10% for SNR equal to 0.9).

### ***False Negative Rate***

The graphs in Figure 4 depict the means of the four-way interaction factor (L\_INV METH x EST\_TYPE x SNR x DIP\_POS) obtained for the FNR index.



**Figure 4** - Means associated with the four-way interaction factor ( $L\_INV\_METH \times EST\_TYPE \times SNR \times DIP\_POS$ ) of the ANOVA performed on the FNR. Each panel corresponds to a specific  $\gamma$  level: 0.5 (panel a), 0.7 (panel b), 0.9 (panel c). For each panel, there are two graphs associated with the two different inverse solutions: eLORETA on the left and LCMV on the right. X-axes always show the four levels of the factor  $DIP\_POS$ : far/superficial (FS), Close/Superficial (CS), Far/Deep (FD) and Close/Deep (CD). Colours code for the different connectivity estimators and whiskers represent 95% confidence intervals.

The percentage of false negatives is less than 5% in all simulated cases, except for the lowest SNR level ( $\gamma$  equal to 0.5) when LCMV is employed. In the easier condition with a higher signal to noise ratio and interacting dipoles that are not deep and close at the same time, the FNR is around 1% regardless of the chosen connectivity estimator.

**Connectivity Estimator:** The factor  $EST\_TYPE$  does not have a significant effect on the FNR index independently of all the other factors (SNR value, type of algorithm chosen for the source reconstruction, and connectivity estimation): the difference between the connectivity estimators never exceeds 1%.

Also, the slight increase of false negatives (see eLORETA) associated with the TRGC is not statistically significant in this case.

**Inverse Algorithm:** The percentage of FN obtained with the two inverse methods is strictly linked to the dipoles' position. Results reported in panel a) show that for  $\gamma$  equal to 0.5, FNR significantly increases for LCMV only when the dipoles are located deep in the brain (accounting for an increase of 20% in the *Close/Deep* condition). Panel b) shows a similar but attenuated trend for  $\gamma$  equal to 0.7 (increase of less than 5% in the *Close/Deep* condition). As shown in panel c), there are no significant differences between LCMV and eLORETA for the highest SNR value.

**Fixed dipoles position:** The factor DIP\_POS is significant for low and medium SNR values and L\_INV METH corresponding to LCMV. In such conditions, for deep dipoles, the FNR is significantly higher, regardless the connectivity estimator. Moreover, focusing on the deep locations, there is a significant increase of the false negatives when the dipoles are close compared to when they are further away.

**SNR:** The signal-to-noise ratio associated with the three levels of the factor SNR significantly influences the presence of false negatives only when the inverse problem is solved by the LCMV algorithm. This is particularly the case for the condition *Close/Deep*, in which the FNR decreases from 20% when  $\gamma$  is equal to 0.5 (panel a) to 4% when  $\gamma$  is equal to 0.7 (panel b), and to 1% for the highest SNR level (panel c).

This suggests that the amount of false negatives is independent of the algorithm employed for solving the inverse problem and for the connectivity estimation when the dipoles are located superficially in the brain. In case of poor signal quality (low SNR), the FNR is considerable when the sources to be reconstructed are located deep in the brain. Algorithms that are more prone to missing connections are LCMV for source reconstruction and TRGC as connectivity estimation.

## AUC

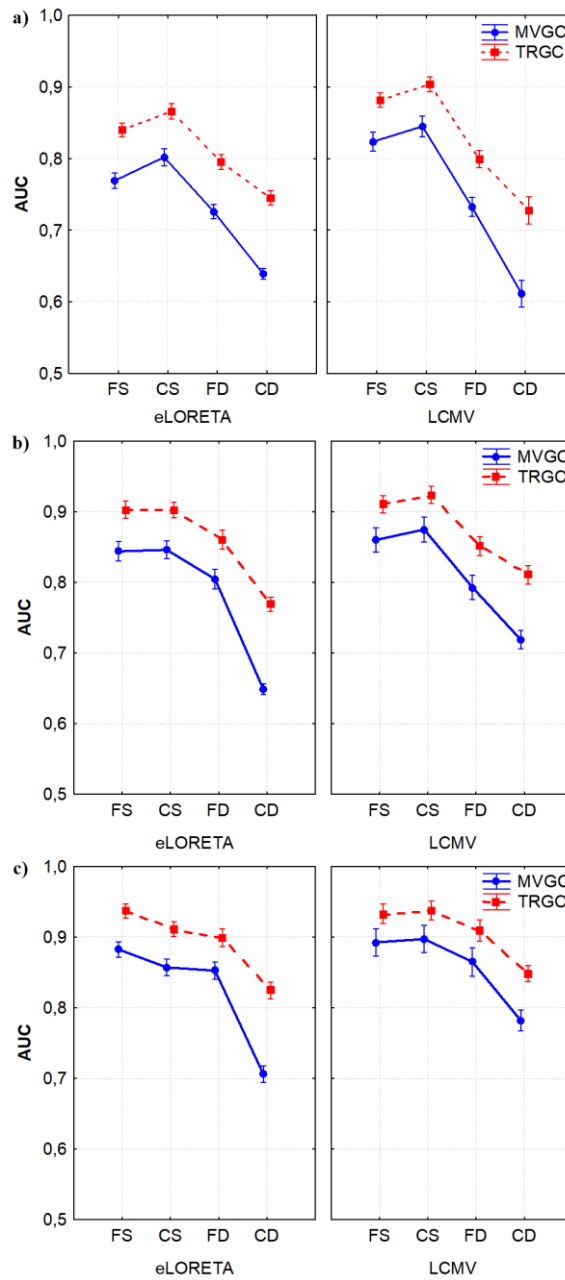
The graphs in Figure 5 depict the means of the four-way interaction factor (L\_INV METH x EST\_TYPE x SNR x DIP\_POS) obtained for the AUC parameter. The AUC index hereby summarizes the effect of the four considered factors on the accuracy of the estimation in terms of false positives and false negatives, providing a unifying measure of the discriminability of actually present and non-existent connections.

**Connectivity Estimator:** As expected from the previous results concerning the FPR trend, the accuracy of the estimation considerably increases when performed by means of TRGC. The increase of the performances associated with TRGC is statistically significant and amounts to about 10%.

**Inverse Algorithm:** On average, the difference between LCMV and eLORETA is not significant, but there are combinations of the factors for which either of the two performed better. The main discrimination is given by the linked dipoles position. When the sources are located deep in the brain (especially if they are also close), the accuracy of the connectivity estimation appears significantly higher when LCMV is employed to reconstruct the brain activity. Once again, the only exception is the low SNR setting, in which this relationship is reversed because LCMV is more sensitive to the SNR level compared to the eLORETA algorithm, which shows more stable performance (for example, in the *close/deep* condition, the range of variability is less than 0.1 for eLORETA and more than 0.2 for LCMV).

**Fixed dipoles position:** Independent of the employed inverse algorithm and connectivity estimator, the accuracy of the estimation significantly decreases when the linked dipoles are located deep in the brain. For higher SNR levels, the ANOVA highlights a performance degradation in terms of the AUC dropping from 90% (*Far/Superficial*) to 70% (*Close/Deep*).

**SNR:** As expected, the performance significantly improves in all considered conditions when the simulated SNR is high. More specifically, when the SNR level is 0.9, the accuracy is higher than 85% for eLORETA and higher than 90% for LCMV in all the cases except for the *Deep/Close* condition.

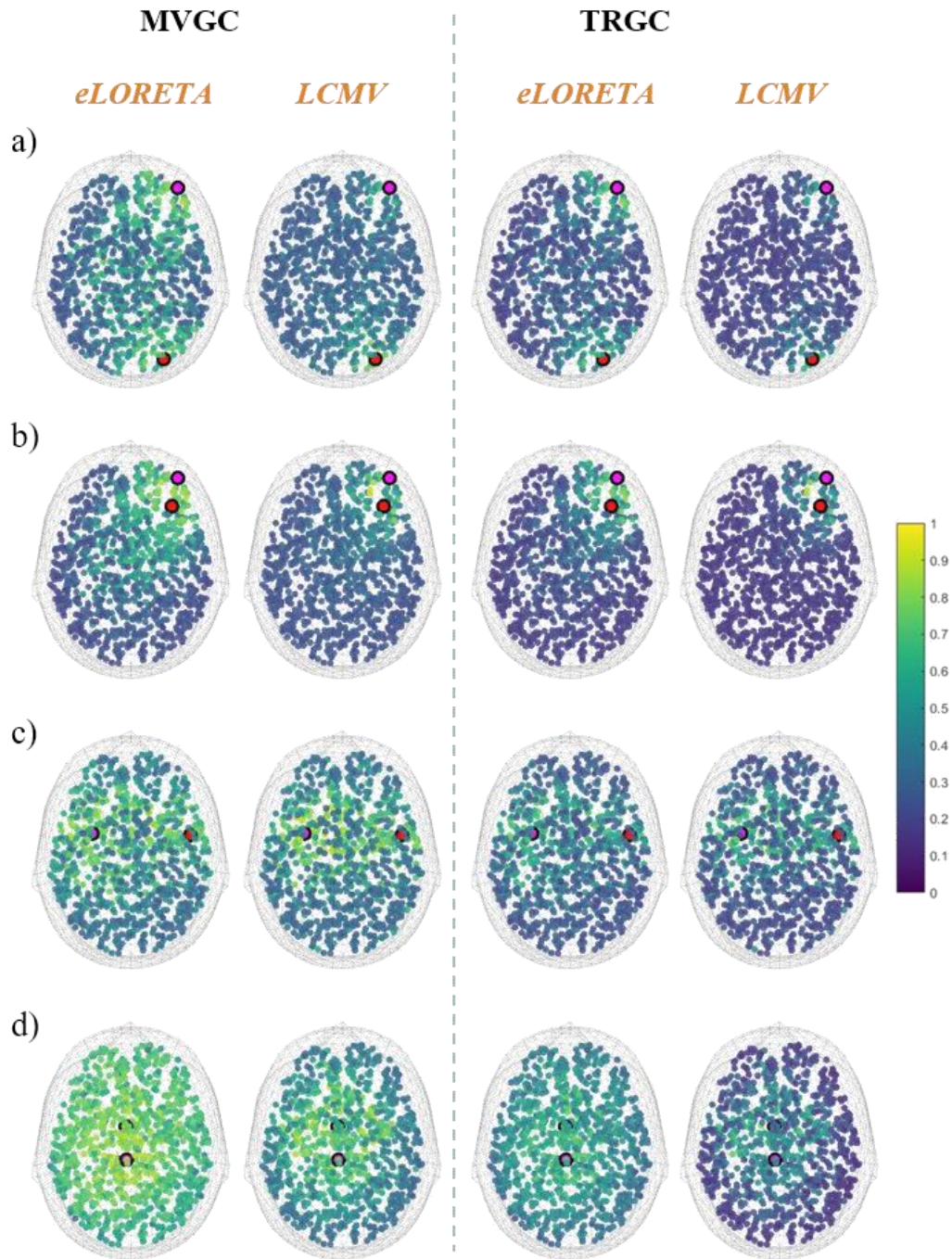


**Figure 5** - Means associated with the four-way interaction factor ( $L\_INV\_METH \times EST\_TYPE \times SNR \times DIP\_POS$ ) of the ANOVA performed on the AUC. Each panel corresponds to a specific  $\gamma$  level: 0.5 (panel a), 0.7 (panel b), 0.9 (panel c). In each panel, there are two graphs associated with the two different inverse solution: LCMV on the left and eLORETA on the right. X-axes report the four levels of the factor  $DIP\_POS$ : far/superficial (FS), Close/Superficial (CS), Far/Deep (FD) and Close/Deep (CD). Colours code for the different connectivity estimators and whiskers represent 95% confidence intervals.

## 3.2 Brain maps

### *Non-Interacting dipole position*

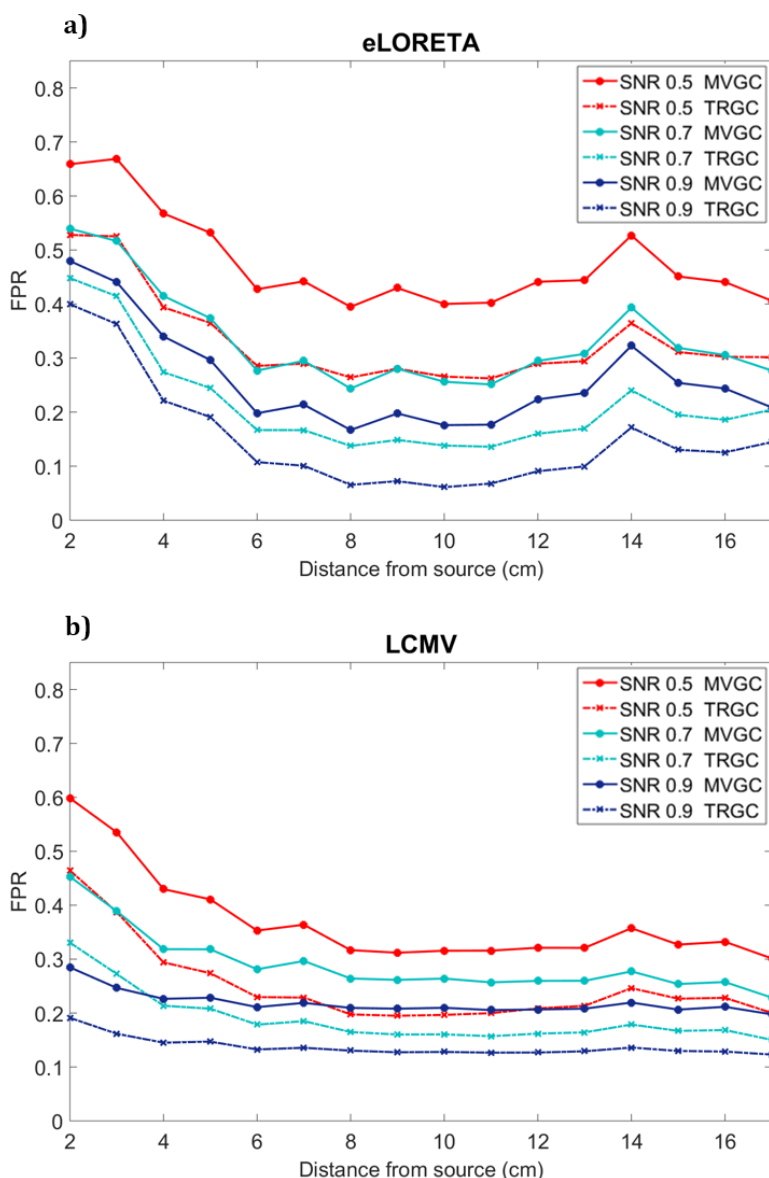
As mentioned before, in each simulated condition, the moving dipole changes its position over 1004 locations equally distributed in the brain. In order to map the performance of the two connectivity estimators for each investigated source reconstruction algorithm and each position of the fixed dipoles, MVGC and TRGC were computed considering all 1004 possible configurations of the network. Since each simulation was iterated 100 times, we were able to obtain an average performance value. Only the maps depicting the FPR are reported because of the greater sensitivity of this indicator to the factors considered in the analysis. It is also possible to consult the same brain maps obtained for the AUC in the supplementary material; they show the same behaviour of the FPR but in a smaller range of variability. For each choice of fixed dipoles position and inverse method we report transparent axial views of the head and the value assumed by the performance parameter (coded by its color) in the position of the moving dipole associated with that measure. Figure 6 reports the results obtained for the lowest SNR level, when  $\gamma$  is equal to 0.5.



**Figure 6** – Spatial distribution of the FPR in the Moving Non-Interacting Dipole condition for low SNR ( $\gamma = 0.5$ ). The Sender (red circle) and Receiver (purple circle) of the interaction are represented in the Far-Superficial (a), Close Superficial (b), Far-Deep (c), and Close-Deep (d) conditions. The other points represent the mean value of the FPR across 100 iterations (coded by the colour bar on the right side) when the third active dipole (the Non-Interacting one) is moved across the brain. The first two columns refer to the classical GC (MVGC algorithm); the last two to the TRGC. For each column, results obtained with eLORETA and LCMV are reported next to each other.

The percentage of false positives depends on the distance of the Non-Interacting dipole from the two fixed ones. The most relevant result is that when the fixed dipoles are located deep in the brain and close to each other (panel d), high FPR values are spread across the whole brain, and reach 100% in the vicinity of the Sender and Receiver. Only TRGC combined with the LCMV algorithm mitigates this effect, which is then limited to the configurations in which the Non-Interacting dipole is close to the other two. Panels a), b), and c) clearly show a strong increase of the FPR when the Non-Interacting dipole is located in the areas close to the Receiver or to the Sender. Similar maps displaying the results obtained for  $\gamma$

equal to 0.7 and 0.9 are reported in the supplementary material. These results confirm the trends commented for the previous maps but with globally better performances. The FPR considerably increases around the fixed dipoles. This phenomenon is focal when LCMV is employed and more spread-out if eLORETA is combined with the MVGC estimator. Again, when the fixed dipoles are located deep in the brain and close to each other, high FPR values are spread across the whole brain and reach 100% in the vicinity of Receiver and Sender. Maps associated with all the other fixed dipoles positions show that when the sources included in the model are far one from the other, the best performance is obtained with eLORETA. In more detail, figure 7 shows the value of the FPR as function of the distance of the moving dipole from the Sender for all SNR values, inverse algorithms, and connectivity estimators in the case of far and superficial interactive sources.



**Figure 7** - FPR as function of the distance (in cm) of the Non-Interacting Moving Dipole from the Sender of the interaction for the two inverse reconstruction algorithms, eLORETA and LCMV. The different SNR levels are drawn in red (high), light blue (medium) and dark blue (low). The circle marker codes for MVGC and the cross for the TRGC.

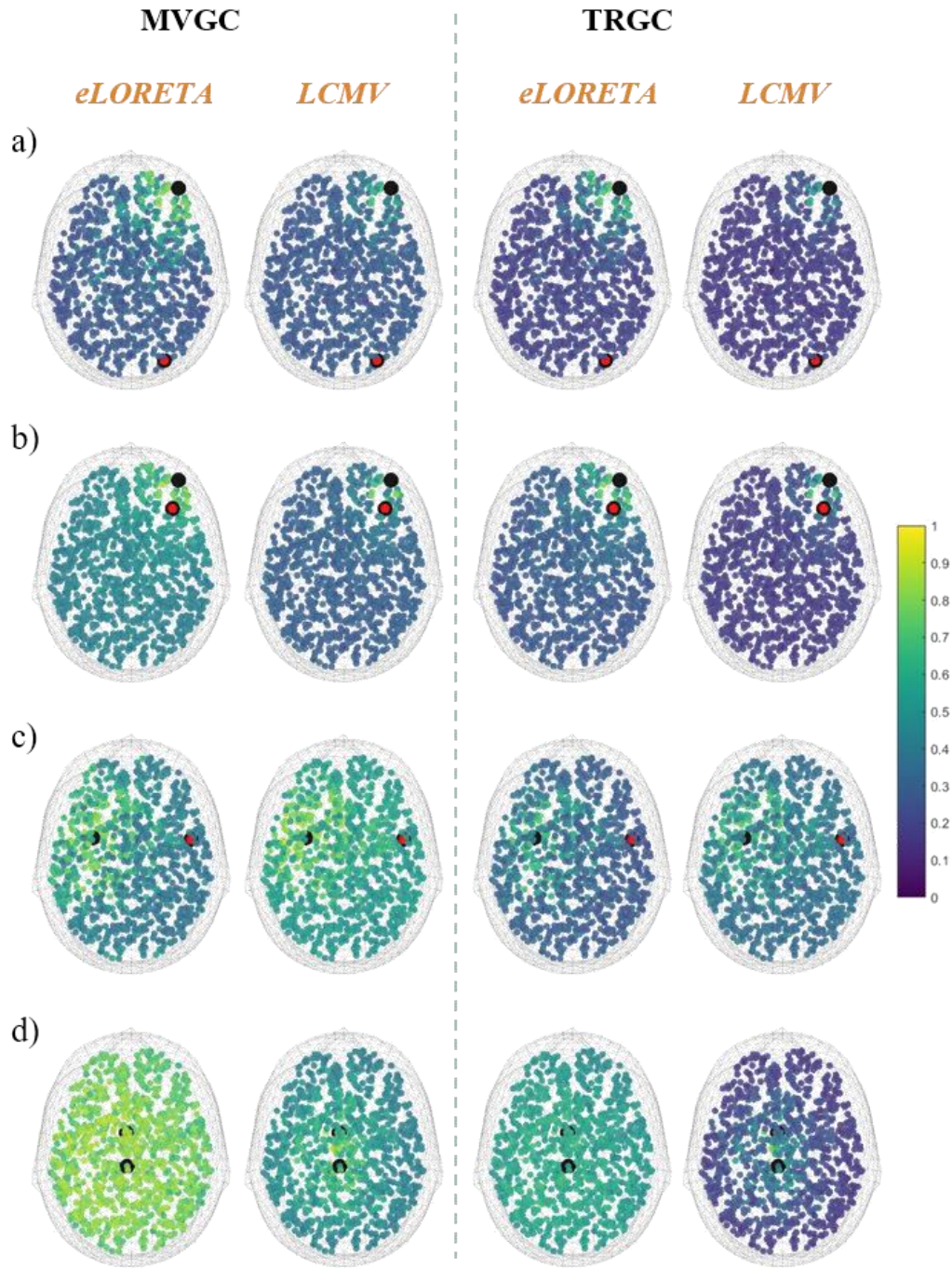
The results suggest that TRGC performs better than MVGC regardless of the distance of the moving dipole from the Sender. LCMV source reconstruction is less sensitive to the distance between dipoles.

For example, in panel a), an increase of the FPR 14 cm away from the Sender is noticeable. This second peak is related to the presence of the second interacting dipole (Receiver). When the LCMV algorithm is employed such increase is much less evident. The trends are similar for all  $\gamma$  levels, although higher FPRs are observed for lower SNRs. For high SNR, the best performance is achieved with eLORETA when the moving dipole is far from the other two.

### ***Interactive dipole position***

The last analysis was performed using a fixed location for the Non-Interacting and Sender dipoles, placing the Receiver dipole at different positions. The results are in line with the previous ones. Figure 8 depicts topographical maps for the low SNR level (in the supplementary material for medium and high SNR levels), while figure 9 depicts FPR as a function of the distance between Receiver and Sender.





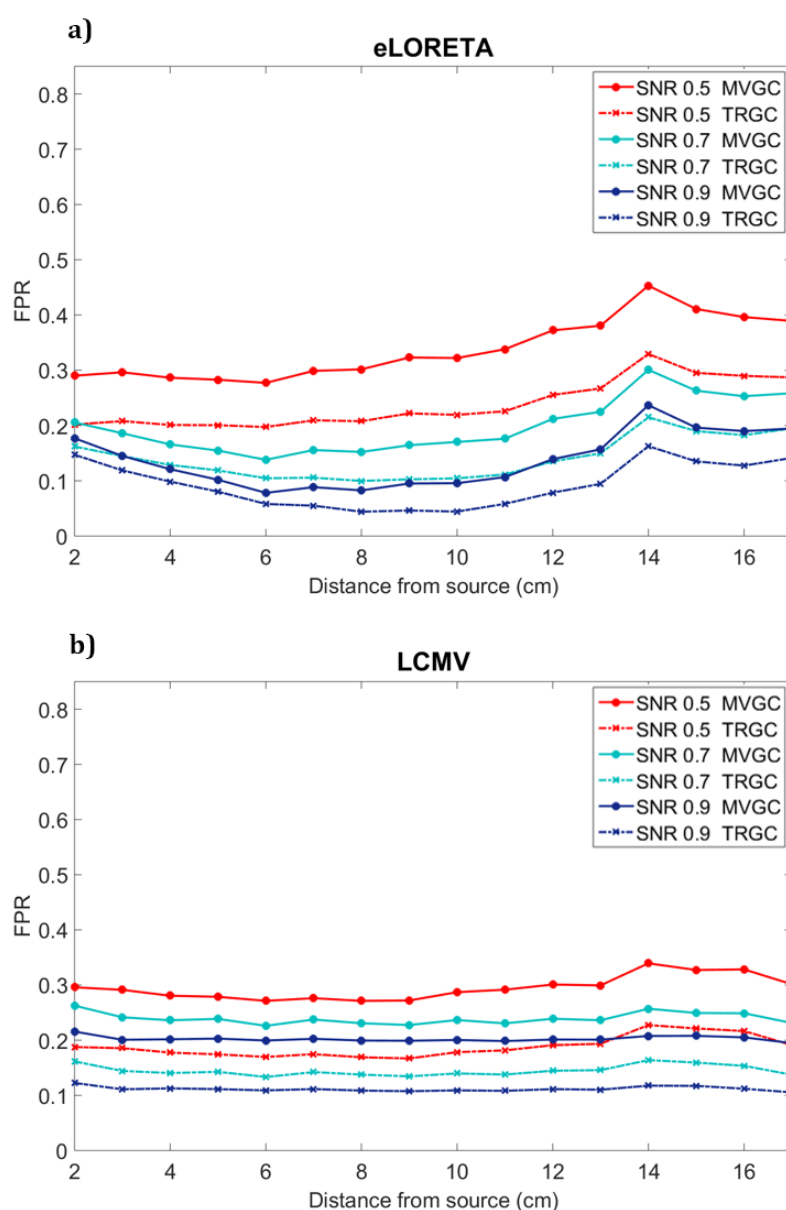
**Figure 8** - Spatial distribution of the FPR in the Moving Receiver Dipole condition for low SNR ( $\gamma = 0.5$ ). The Sender (red circle) and Non-Interacting dipole (black circle) are represented in the Far-Superficial (a), Close Superficial (b), Far-Deep (c), and Close-Deep (d) conditions. The other points represent the mean value of the FPR across 100 iterations (coded by the colour bar on the right side) when the third active dipole (the Receiver) is moved throughout the brain. The first two columns refer to the classical GC (MVGC algorithm); the last two to the TRGC. For each column, results obtained with eLORETA and LCMV are reported next to each other.

When the fixed dipoles are close (panels b and d), a high percentage of false positives appear throughout the brain, in particular for low and medium level of SNR. All other results are in line with the results reported above:

- the increase of the  $\gamma$  value corresponds to a decrease in the number of false positives independently of all other factors;
- on average, LCMV performed better than eLORETA. This advantage is predominantly due to an increased robustness w.r.t. the position of the nodes;

- TRGC provided more accurate connectivity estimates than MVGC;
- when the involved dipoles are far away from another, eLORETA leads to more accurate connectivity estimation, and the difference between the classical MVGC and TRGC is less pronounced than in the other conditions;
- with the dipoles in the *Far/Superficial* configuration, and  $\gamma$  equal to 0.9, the percentage of false positives is less than 10% for all the inverse solutions and connectivity algorithms;
- with the dipoles in the *Close/Deep* configuration, the percentage of false positives reaches 100% regardless of the SNR value.

Fig. 9 shows the FPR as function of the distance between Sender and Receiver for all SNR values, inverse algorithms, and connectivity estimators, when the fixed dipoles are in the Far/Superficial configuration.



**Figure 9.** FPR as function of the distance (in cm) of the moving Receiver from the Sender for the two inverse reconstruction algorithms, eLORETA and LCMV. The different SNR levels are drawn in red (high), light blue (medium), and dark blue (low). The circle marker codes for MVGC and the cross for the TRGC.

The first result is that the mean value of the FPR is lower than for the Non-Interacting moving dipole condition. Also, in this case, TRGC performed better than MVGC regardless of the distance of the moving Receiver dipole from the Sender. In panel a) it is possible to notice an increase of the FPR when the Receiver dipole is 14cm away from the Sender dipole (this being the position of the Non-Interacting dipole). Trends are similar for all the  $\gamma$  levels, where, generally, decreases in SNR are associated with increases in FPR. For high SNR, the best performance is achieved using eLORETA when the moving dipole is far away from the other two (FPR around 5%).

#### **4. Discussion and Conclusion**

It is well established that neuroelectrical measures recorded on the scalp need to be projected back into the brain in order to be able to infer where these signals have been generated. Unfortunately, even with state-of-the-art localization of the brain sources underlying the measured signals, directed dynamical influences between these reconstructed sources do not always reflect the ground truth. This issue has been anticipated [8]–[10], thoroughly analysed [48] and mitigation strategies were proposed [32], [49]. In the present simulation study, we focused on directional connectivity measures and quantified the extent to which the estimation of influences between reconstructed sources is possible. We employed an analysis framework combining source localization approaches and brain connectivity estimators with the goal of identifying those analysis pipelines that are least affected by the presence of head volume conduction and, therefore, provide the most accurate and reliable connectivity estimates. Several settings were simulated corresponding to both advantageous and disadvantageous conditions for the following brain connectivity estimation. To this end, we modulated the depth of the sources, the reciprocal distance between sources, and the SNR (brain noise level). Conscious of the fact that more than three sources are likely involved in a more realistic simulation of brain activity, making the issue even thornier, here we wanted to present a minimal set of situations allowing for a first glimpse of the possible modulations of the bias in connectivity estimates, and the extent to which some proposed solutions allow their mitigation. Importantly, it should be noted that we evaluate sources at known locations and show that even in this case a bias can be present. Estimating the source locations is more realistic but also makes the objective performance evaluation much more difficult. In other words, this study does not help in selecting the sources or the ROIs for the subsequent connectivity estimation (which is another critical issue) but helps in the definition of some criteria to be taken into account during this choice: the depth of the ROIs and their reciprocal distance. Not surprisingly, a convenient condition we identified is the presence of far and superficial dipoles in combination with a high SNR; in contrast, a disadvantageous condition is given by the presence of close and deep sources with a low SNR level. Our simulations suggest that all considered factors show a significant influence on the estimation quality and, consequently, their combination has a considerable impact on the connectivity estimation performance. LCMV source reconstruction appears to be more sensitive to the SNR value, while eLORETA achieves similar performance regardless of the SNR. In general, LCMV showed better performance than eLORETA but, when the simulated sources were assigned to distant locations, the eLORETA performance is similar to or better than the performance of LCMV. In other words, there are combinations of the factors for which either of the two performed better.

In agreement with the theoretical hypothesis, we demonstrated that the TRGC algorithm provides a better estimation of the directed statistical dependencies between sources than classical MVGC. Indeed, the percentage of spurious connections decreased significantly, and the overall detection of connectivity as measured by the AUC increased significantly in all considered experimental conditions when TRGC was used instead of MVGC. At the same time, the percentage of missed connections as measured by the FNR increased slightly, but still remained close to zero. As expected, we found that closer and deeper active sources decreased the obtained performance. Thus, a dependence between the dipoles position and the accuracy of the estimates was found and mapped. This is a clear effect of the volume conduction, since, when two sources are close to each other, they generate a highly mixed signal on the scalp, which compromises the correct estimation even after inverse source reconstruction. On the other hand, when the sources are far away from each other, they are less affected by volume conduction, leading to a better quality of the connectivity estimation. Summarizing, the analysis of the FPR and AUC indices suggests that the optimal combination of factors is given by: i) dipoles located superficial in the brain and not too close; ii) LCMV and eLORETA algorithms show similar performance when the active sources are far and superficial but LCMV is more accurate when the sources are located deep in the brain (when SNR is not too low); iii) TRGC as connectivity estimator. Without providing an indication of the specific ROIs to choose, this study can still provide some guidance on certain issues, such as the minimum distance for which it makes sense to include two distinct sources in the model as well as the selection of source reconstruction and connectivity estimation algorithms that promise to provide the most reliable and physiologically interpretable description of brain networks based on EEG data. In the future, it would be interesting to test other important factors that could influence the quality of the source reconstruction and, thus, the estimation accuracy such as the number of interactive dipoles, their orientation, the dependency from the coupling strength and the inclusion of a differently defined SNR allowing deep source to have a different strength with respect to the superficial ones.

We agree with [30], [48], advocating for the application of measures for which promises and pitfalls are known, and which integrated knowledge of how neural activity in the whole brain as well as external (physiological or artifactual) activity contribute to the signals that we record on the scalp. In this regard, it should be noted that connectivity estimates can only at most be as focal as the reconstructed source current densities they are derived from, and we know that common inverse methods lead to very blurry results. To distinguish correctly-identified connections from connections that are observed in the vicinity of the true interacting sources due to blurry inverse solutions, a data-driven clustering in the space of brain-wide pairwise connectivities, as recently proposed in [49], may be a viable option, which may be preferable to a reduction of the source space to the level of static ROIs. Many other solutions for reducing the spurious links detection have been proposed and are currently being studied, as the leakage correction procedures [27]–[30], [32] and the adaptive parcellation [31], [33]. Unfortunately, the perfect solution to address the spurious links problem still does not exist, since a transformation (e.g. parametrized by a linear demixing matrix) designed to remove a single type of dependency does not (in general) automatically remove all other possible dependencies (e.g. non-linear or lagged). For these reasons we decided to investigate, apart from the classical GC, an approach (the TRGC) for which theoretical guarantees exist, and we demonstrate that these theoretical results lead to improvements in practice not only when estimating connectivity on sensors [9] but also in source space. The proposed framework will hopefully be useful for further validation of existing unmixing tools and upcoming ones combined with the causal connectivity estimates.

However, it must be kept in mind that – although of importance – the main problem in EEG-based brain connectivity analysis is not the spatial blur of correctly identified connections but the emergence of spurious connectivity as a result of observing mixtures of signals even at the level of reconstructed sources. This problem can only be addressed by using appropriate and validated connectivity measures that are robust to volume conduction effects by construction.

## Code and data availability

The code necessary to reproduce these simulations is available at:

[https://github.com/paolop21/simulation\\_source\\_connectivity](https://github.com/paolop21/simulation_source_connectivity).

Simulations were run in parallel on the HPC platform of the University of Ghent, with a running time depending on the available nodes on the network. An iteration of the simulation on a Dell Inspiron 5567, intel i7, 16 GB RAM takes 2,4 seconds. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Super- computer Center), funded by Ghent University, FWO and the Flemish Government – department EWI.

The results of the simulations and the structures necessary to run the code are available at:

<https://zenodo.org/record/1157196#.WmZ1K6jiY2w>

## 5. References

- [1] K. J. Friston, “Functional and effective connectivity in neuroimaging: A synthesis,” *Hum. Brain Mapp.*, vol. 2, no. 1–2, pp. 56–78, Jan. 1994.
- [2] B. Horwitz, “The elusive concept of brain connectivity,” *NeuroImage*, vol. 19, no. 2, pp. 466–470, Jun. 2003.
- [3] Astolfi L., Cincotti F, Mattia D, Salinari S, Babiloni C, Basilisco A, Rossini PM, Ding L, Ni Y, He B, Marciani MG, Babiloni F., “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*, vol. 22, no. 10, pp. 1457–1470, Dec. 2004.
- [4] Babiloni F, Cincotti F, Babiloni C, Carducci F, Mattia D, Astolfi L, Basilisco A, Rossini PM, Ding L, Ni Y, Cheng J, Christine K, Sweeney J, He B., “Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function,” *NeuroImage*, vol. 24, no. 1, pp. 118–131, Jan. 2005.
- [5] K. J. Blinowska, “Review of the methods of determination of directed connectivity from multichannel data,” *Med. Biol. Eng. Comput.*, vol. 49, no. 5, pp. 521–529, May 2011.
- [6] Nunez PL, Srinivasan R, Westdorp AF, Wijesinghe RS, Tucker DM, Silberstein RB, Cadusch PJ., “EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 5, pp. 499–515, Nov. 1997.
- [7] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.
- [8] F. V. de Steen, L. Faes, E. Karahan, J. Songsiri, P. A. Valdes-Sosa, and D. Marinazzo, “Critical Comments on EEG Sensor Space Dynamical Connectivity Analysis,” *Brain Topogr.*, pp. 1–12, Nov. 2016.
- [9] S. Haufe, V. V. Nikulin, K.-R. Müller, and G. Nolte, “A critical assessment of connectivity measures for EEG data: A simulation study,” *NeuroImage*, vol. 64, pp. 120–133, Jan. 2013.
- [10] C. Brunner, M. Billinger, M. Seeber, T. R. Mullen, and S. Makeig, “Volume Conduction Influences Scalp-Based Connectivity Estimates,” *Front. Comput. Neurosci.*, vol. 10, Nov. 2016.

- [11] G. Gómez-Herrero, M. Atienza, K. Egiazarian, and J. L. Cantero, "Measuring directional coupling between EEG sources," *NeuroImage*, vol. 43, no. 3, pp. 497–508, Nov. 2008.
- [12] S. Haufe, R. Tomioka, G. Nolte, K. R. Müller, and M. Kawanabe, "Modeling Sparse Connectivity Between Underlying Brain Sources for EEG/MEG," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1954–1963, Aug. 2010.
- [13] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
- [14] Nolte G, Ziehe A, Nikulin VV, Schlögl A, Krämer N, Brismar T, Müller KR., "Robustly Estimating the Flow Direction of Information in Complex Physical Systems," *Phys. Rev. Lett.*, vol. 100, no. 23, p. 234101, Jun. 2008.
- [15] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Hum. Brain Mapp.*, vol. 8, no. 4, pp. 194–208, 1999.
- [16] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2292–2307, Oct. 2004.
- [17] Vinck M, Huurdeman L, Bosman CA, Fries P, Battaglia FP, Pennartz CM, Tiesinga PH5, "How to detect the Granger-causal flow direction in the presence of additive noise?," *NeuroImage*, vol. 108, no. Supplement C, pp. 301–318, Mar. 2015.
- [18] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference," *J. Neurosci. Methods*, vol. 223, pp. 50–68, Feb. 2014.
- [19] I. Winkler, D. Panknin, D. Bartz, K. R. Müller, and S. Haufe, "Validity of Time Reversal for Testing Granger Causality," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2746–2760, Jun. 2016.
- [20] M. Vinck, R. Oostenveld, M. van Wingerden, F. Battaglia, and C. M. A. Pennartz, "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias," *NeuroImage*, vol. 55, no. 4, pp. 1548–1565, Apr. 2011.
- [21] G. Nolte, A. Ziehe, N. Krämer, F. Popescu, and K.-R. Müller, "Comparison of Granger Causality and Phase Slope Index," in *Causality: Objectives and Assessment*, 2010, pp. 267–276.
- [22] Grech R, Cassar T, Muscat J, Camilleri KP, Fabri SG, Zervakis M, Xanthopoulos P, Sakkalis V, Vanrumste B, "Review on solving the inverse problem in EEG source analysis," *J. Neuroengineering Rehabil.*, vol. 5, p. 25, Nov. 2008.
- [23] K. Mahjoory, V. V. Nikulin, L. Botrel, K. Linkenkaer-Hansen, M. M. Fato, and S. Haufe, "Consistency of EEG source localization and connectivity estimates," *NeuroImage*, vol. 152, pp. 590–601, May 2017.
- [24] T. Hedrich, G. Pellegrino, E. Kobayashi, J. M. Lina, and C. Grova, "Comparison of the spatial resolution of source imaging techniques in high-density EEG and MEG," *NeuroImage*, vol. 157, pp. 531–544, Aug. 2017.
- [25] R. D. Pascual-Marqui *et al.*, "Assessing interactions in the brain with exact low-resolution electromagnetic tomography," *Philos. Transact. A Math. Phys. Eng. Sci.*, vol. 369, no. 1952, pp. 3768–3784, Oct. 2011.
- [26] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–880, Sep. 1997.
- [27] J. F. Hipp, D. J. Hawellek, M. Corbetta, M. Siegel, and A. K. Engel, "Large-scale cortical correlation structure of spontaneous oscillatory activity," *Nat. Neurosci.*, vol. 15, no. 6, pp. 884–890, Jun. 2012.
- [28] G. L. Colclough, M. J. Brookes, S. M. Smith, and M. W. Woolrich, "A symmetric multivariate leakage correction for MEG connectomes," *NeuroImage*, vol. 117, pp. 439–448, Aug. 2015.
- [29] M. J. Brookes, M. W. Woolrich, and G. R. Barnes, "Measuring functional connectivity in MEG: a multivariate approach insensitive to linear source leakage," *NeuroImage*, vol. 63, no. 2, pp. 910–920, Nov. 2012.
- [30] R. Pascual-Marqui, R. J. Biscay, J. Bosch-Bayard, P. L. Faber, T. Kinoshita, K. Kochi, P. Milz, K. Nishida, M. Yoshimura, "Innovations orthogonalization: a solution to the major pitfalls of EEG/MEG 'leakage correction,'" *bioRxiv*, p. 178657, Aug. 2017.
- [31] S.-R. Farahibozorg, R. N. Henson, and O. Hauk, "Adaptive cortical parcellations for source reconstructed EEG/MEG connectomes," *NeuroImage*, vol. 169, pp. 23–45, Apr. 2018.

- [32] R. J. Biscay, J. F. Bosch-Bayard, and R. D. Pascual-Marqui, "Unmixing EEG Inverse Solutions Based on Brain Segmentation," *Front. Neurosci.*, vol. 12, 2018.
- [33] E. Gonzalez-Moreira, D. Paz-Linares, E. Martinez-Montes, P. Valdes-Hernandez, J. Bosch-Bayard, M. L. Bringas-Vega, P. Valdes-Sosa, "Populational Super-Resolution Sparse M/EEG Sources and Connectivity Estimation," *bioRxiv*, p. 346569, Jun. 2018.
- [34] Y. Huang, L. C. Parra, and S. Haufe, "The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting," *NeuroImage*, vol. 140, pp. 150–162, Oct. 2016.
- [35] Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313–327, Jan. 2011.
- [36] K. Whittingstall, G. Stroink, L. Gates, J. Connolly, and A. Finley, "Effects of dipole position, orientation and noise on the accuracy of EEG source localization," *Biomed. Eng. OnLine*, vol. 2, p. 14, 2003.
- [37] R. D. Pascual-Marqui, "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization," *ArXiv07103341 Math-Ph Physicsphysics Q-Bio*, Oct. 2007.
- [38] J. B. Bell, "Review of Solutions of Ill-Posed Problems," *Math. Comput.*, vol. 32, no. 144, pp. 1320–1322, 1978.
- [39] A. N. Tikhonov and V. I. Arsenin, *Solutions of ill-posed problems*. Winston, 1977.
- [40] J. F. Geweke, "Measures of Conditional Linear Dependence and Feedback Between Time Series," *J. Am. Stat. Assoc.*, vol. 79, no. 388, pp. 907–915, 1984.
- [41] L. Faes, S. Stramaglia, and D. Marinazzo, "On the interpretability and computational reliability of frequency-domain Granger causality," *F1000Research*, vol. 6, p. 1710, Sep. 2017.
- [42] B. L. P. Cheung, B. A. Riedner, G. Tononi, and B. D. Van Veen, "Estimation of cortical connectivity from EEG using state-space models," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2122–2134, Sep. 2010.
- [43] S. Haufe, V. V. Nikulin, and G. Nolte, "Alleviating the Influence of Weak Data Asymmetries on Granger-Causal Analyses," in *Latent Variable Analysis and Signal Separation*, 2012, pp. 25–33.
- [44] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Ann. Math. Stat.*, vol. 9, no. 1, pp. 60–62, Mar. 1938.
- [45] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.
- [46] S. Haufe and A. Ewald, "A Simulation Framework for Benchmarking EEG-Based Brain Connectivity Estimation Methodologies," *Brain Topogr.*, pp. 1–18, Jun. 2016.
- [47] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.
- [48] J. M. Palva, Wang SH, Palva S., Zhigalov A., Monto S., Brookes MJ., Schoffelen JM, Jerbi K., "Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures," *NeuroImage*, vol. 173, pp. 632–643, Jun. 2018.
- [49] S. H. Wang, M. Lobier, F. Siebenhühner, T. Puolivali, S. Palva, and J. M. Palva, "Hyperedge bundling: A practical solution to spurious interactions in MEG/EEG source connectivity analyses," *bioRxiv*, p. 219311, Nov. 2017.