Structural connectivity to reconstruct brain activation and effective connectivity between brain regions

Brahim Belaoucha, Théodore Papadopoulo

Structural connectivity to reconstruct brain activation and effective connectivity between brain regions

Brahim Belaoucha*, Théodore Papadopoulo

Inria, Université Côte d'Azur

ABSTRACT

Understanding how brain regions interact to perform a specific task is very challenging. EEG and MEG are two non-invasive imaging modalities that allow the measurement of brain activation with high temporal resolution. Several works in EEG/MEG source reconstruction show that estimating brain activation can be improved by considering spatio-temporal constraints but only few of them use structural information to do so. In this work, we present a source reconstruction algorithm that uses brain structural connectivity, estimated from diffusion MRI (dMRI), to constrain the EEG/MEG source reconstruction. Contrarily to most source reconstruction methods which reconstruct activation for each time instant, the proposed method estimates an initial reconstruction for the first time instants and a multivariate autoregressive model that explains the data in further time instants. This autoregressive model can be thought as an estimation of the effective connectivity between brain regions. We called this algorithm iterative Source and Dynamics reconstruction (iSDR). This paper presents the overall iSDR approach and how the proposed model is optimized to obtain both brain activation and brain region interactions. The accuracy of our method is demonstrated using synthetic data in which it shows a good capability to reconstruct both activation and connectivity. iSDR is also tested with real data obtained from [**dataset**] (face recognition task). The results are in phase with other works published with the same data and others that used different imaging modalities with the same task showing that the choice of using an autoregressive model gives relevant results.

Keywords: Autoregressive model, EEG/MEG source reconstruction, diffusion MRI, effective connectivity.

1 INTRODUCTION

There are three main techniques to reconstruct brain activity from EEG/MEG measurements. Dipole fitting assumes that the measurements are generated from a fixed number of sources (a few active regions). Its optimization is nonlinear and non-convex and computationally expensive even when considering a relatively small number of possible active regions. Scanning techniques avoid this complexity issue by scanning through regions of interests (ROIs) [**beamformer**, **music**, **Roninson-etal:99**, **Gross-etal:01**, **Mosher-etal:99**]. Most scanning techniques assume some statistical properties on sources and noise (asynchronous, uncorrelated). These properties are used to derive an optimization criterion that aims at identifying the activated sources. They are also sometimes used to estimate the "dimension" of the signal space in order to achieve a separation between the sources and noise spaces. Finally, distributed models use a significantly large number of sources distributed over the brain substrate or volume and estimate a kind of activation probability. Because the size of the sensor array of EEG/MEG is much smaller than the usual dimension of the source space, the distributed model is ill-posed and a prior has to be used to obtain a unique solution. Many priors lead to convex (and sometimes even linear) problems, which makes distributed models attractive. Distributed models were originally designed to work for reconstructing sources for a single time instant but were generalized to deal with reconstructing sources in a time window allowing to constraint sources not only spatially but also temporally [**hamalainen-sarvas:89**, **pascual-marqui-michel-etal:94**, **gramfort:hal-00690774**].

We denote the number of sensors by N_c, the number of sources by N_s and the number of time samples by T. The gain matrix (lead-field) $\mathbf{G}$ is a real matrix of size $N_c \times N_s$, obtained by solving the forward problem [**meijs-peters:87**, **Meijs1988**, **hamalainen-sarvas:89**, **yvert-bertrand-etal:95**, **dassios-kariotou:03**]. It allows us to relate cortical activation $\mathbf{J} \in \mathbb{R}^{N_s \times T}$ and the MEG/EEG measurements $\mathbf{M} \in \mathbb{R}^{N_c \times T}$ with:

$$\mathbf{M} = \mathbf{GJ} + \mathbf{E} \qquad \qquad \textbf{Equation } 1.$$

where $\mathbf{E} \in \mathbb{R}^{N_c \times T}$ is the measurement noise.

Setting a prior on the source space consists of assuming that the prior term $P(\mathbf{J})$ is small for a given source estimate $\mathbf{J}$. A good source estimate is obtained according to the following minimization:

$$\mathbf{J}^* = \underset{\mathbf{J}, \lambda}{\operatorname{argmin}}\, U(\mathbf{J}) = D(\mathbf{M}, \mathbf{J}) + \lambda P(\mathbf{J}) \, . \qquad \qquad \textbf{Equation } 2.$$

where λ is a positive number that controls the trade-off between the data fit term $D(\mathbf{M}, \mathbf{J})$ which quantifies how well the source estimates explains the measured signal and the regularity of the estimate. Although such methods can deal with temporal or spatio-temporal priors, very often only spatial priors are used because it is easier to consider independent time sample reconstruction as it makes the reconstruction faster. Very often also, it is assumed that sources are centered and follow a Gaussian distribution. In that case, an optimal source reconstruction can be estimated linearly source by minimizing the following cost function:

$$U(\mathbf{J}) = \|\mathbf{M} - \mathbf{GJ}\|_F^2 + \lambda \|\mathbf{WJ}\|_F^2 \qquad \qquad \textbf{Equation } 3.$$

where λ is a positive number used to balance D and P. This method is also known as Tikhonov regularization (Ridge Regression) [**Tikhonov**]. Several methods can be found in the literature that use this formulation, they differ mostly by the choice of the weighting matrix $\mathbf{W}$.

The minimum norm estimate (MNE) sets $\mathbf{W}$ to the identity matrix in Equation 3 [**hamalainen1984interpreting**]. It is known that MNE overestimates the extent of the activity regions due to underestimating the magnitude of active sources resulting from the use of l_2 norm. Due to its simple linear solution, this approach remains attractive. In the low resolution brain electromagnetic tomography (LORETA) reconstruction [**pascual-marqui-michel-etal:94**], $\mathbf{W}$ is set to $\mathbf{L}^T \mathbf{L}$ where $\mathbf{L}$ is the 3D-discretized Laplacian matrix of the source space. The ith row of $\mathbf{L}$ acts like a discrete differentiating operator by computing differences between the ith node of the source space and its direct neighbors. This method provides the solution with the maximum spatial smoothness. In the general case, when $\mathbf{W}$ is different from the identity matrix, we need to invert a big matrix of size $N_s \times N_s$ which is computationally expensive.

Some nonlinear approaches favor specific source configurations. In Minimum Current Estimate – MCE or LASSO – ([**MatsuuraMCE**]), few sources are active at each time sample. The solution is obtained by solving the following functional for each time instant independently [**MatsuuraMCE**, **phdGramfort**]:

$$J_\lambda = \underset{J}{\operatorname{argmin}}\, \frac{1}{2} \|\mathbf{M} - \mathbf{GJ}\|_2^2 + \lambda \|\mathbf{J}\|_1 \, , \lambda > 0.$$

This results in non-continuous source activation in both time and space, as it can be seen as promoting spatial sparsity independently at each time instant. It thus fails to recover smooth time courses of cortical sources. To overcome this limitation, structured sparsity can be used. In the Mixed Norm Estimate (MxNE) [**gramfort:hal-00690774**], the regularization term induces both spatial sparsity and temporal smoothness of sources by using a mixed norm

l_{21} instead of l_1 norm. $l_{21}-$norm is defined as:

$$\|\mathbf{J}\|_{21} = \sum_i \sqrt{\sum_j J_i(j)^2}, \qquad \text{Equation 4.}$$

where the sum over i is over sources (spatial position) and the sum over j is the sum over time samples i.e. the l_{21}-norm corresponds to the sum of the l_2-norm of individual source time courses. The solution for such a criterion can be found in [**phdGramfort**, **gramfort:hal-00690774**]. MxNE results in few active sources in a given time window. Like MNE, it underestimates the source activity due to the use of the l_2 norm and soft-threshold, which is used to solve the minimization, over the time courses [**irMxNE**].

All of the above methods do not use any structural information to constrain the source space. In [**passingham-stephan-etal:0 Sporns2004418**, **tomassini-jbabdi-etal:07**, **finger-etal:16**, **osher-etal:16**], authors show that structural connectivity (connections between neurons through the white matter) can be an indicator of functional homogeneity of cortical regions. Diffusion magnetic resonance imaging (dMRI) is a non-invasive modality that allows the access to brain structural connectivity [**einstein:05**, **einstein:56**, **stejskal-tanner:65**, **basser-mattiello-etal:94a**, **fdt**]. As it is known that EEG/MEG can only "see" activations corresponding to coherent sources spreading across a few cm^2 of cortex, it seems interesting and justified – to some extent – to integrate the knowledge of the functional regions as provided by dMRI in the EEG/MEG reconstruction problem.

As stated above, only few works use diffusion MRI in conjunction with EEG/MEG measurements to reconstruct the brain activity. dMRI was first used as a pure spatial prior in the MEG/EEG inverse problem. In [**Hammond-etal:12**, **hammond-etal:13**], dMRI is used to create a network of connected brain regions. Assuming that connected regions should have similar activities, they look for a solution of the EEG source reconstruction problem that is smooth across this network. In [**philippe:13**], dMRI is used again to parcellate the cortex into functional regions. These regions are then used to spatially constrain the source reconstructions to promote solutions that are either constant or smooth in functional regions. In [**Knosche20137**], the authors use regions defined by dMRI to define LORETA patch-based algorithm i.e. smooth sources inside patches. The structural connectivity between brain regions was also used to reconstruct brain activity by imposing a stronger penalty for regions with weak anatomical connections [**PinedaPardo2014765**]. More elaborate time models have also been designed. In [**galka-okito-etal:04**], the authors modeled the brain activity at each source location i by:

$$\mathbf{J}_t(i) = a_i \mathbf{J}_{t-1}(i) + b_i \sum_{j \neq i} \mathbf{J}_{t-1}(j) + \varepsilon_t$$

where ε_i is noise at the source location i and the sum is over the direct neighbors of i. In their method, there are only two parameters per source (a_i, b_i) to be estimated which reduces the dynamics that their model can explain. In this model, for a given time window, only $3 \times N_s$ parameters $(\mathbf{J}_0(i), a_i, b_i)$, $i = 1..N_s$ need to be estimated instead of the $T \times N_s$ parameters that would be estimated with standard spatio-temporal models. This model was simplified in [**long-purdon-etal:06**], by simply assuming $a_i = 1$ and $b_i = \frac{1}{n_i}$, where n_i the number of neighbors of node i, so the solution for a given time window only requires estimating N_s parameters (as in the purely spatial problem for one time instant). In [**lamus-etal:12**], the same $3 \times N_s$ need to be estimated but the activities of neighboring sources are weighted by their distances to source i.

Another nonlinear model is considered in [**giralso-dendekker-etal:10**]. They modeled the brain activity as follows:

$$\mathbf{J}_t = (a_1 \mathbf{I} + b_1 \mathbf{L}) \mathbf{J}_{t-1} + a_2 \mathbf{J}_{t-1}^2 + a_3 \mathbf{J}_{t-1}^3 + a_4 \mathbf{J}_{t-2} + \varepsilon_t$$

where L is a Laplacian operator applied to neighboring sources. Although this nonlinear model could explain more complex dynamics compared to the previous mentioned work, it uses only $5 + 2 \times N_s$ parameters (J_0, J_1 and a_1, b_1, a_2, a_3, a_4) to explain source dynamics.

Finally, only few works use MAR models combined with dMRI information to constrain sources' dynamics. [**fukushima-yama**, **fukushima-yamashita-etal:13**, **Fukushima2015408**] also start by extracting a network of connected brain regions, but in addition to the spatial constraints, the authors use this network to temporally constrain the sources' dynamics by assuming a multivariate autoregressive model (MAR) whose elements are constrained by anatomical connections obtained from dMRI. The MAR model is defined by:

$$\mathbf{J}_t = \sum_{k=1}^{K} \mathbf{A}_k \mathbf{J}_{t-\Delta_k} + \varepsilon_t \,,$$

where $\Delta_{1:K}$ contains all the considered time lags. The only possible nonzero elements of the MAR matrices $\mathbf{A}_k$, $k = 1..K$ are the elements corresponding to anatomical connections between a pair of sources/regions extracted from dMRI. The time-lag, in their method, for each pair of sources/regions is fixed by the length of the pathways between them. Consequently, nonzero MAR coefficients never overlap between matrices, which reduces the degree of freedom of the model [**Fukushima2015408**]. A dynamic hierarchical variational Bayesian (dhVB) method is used to estimate both the sources magnitudes and their interactions. Their method is time-consuming (days for realistic head models) and has several parameters to tune. Unfortunately, the code of dhVB is not published and cannot be used to compare iSDR to dhVB.

Our approach (presented in the next section), follows a similar model. The model has thus more degrees of freedom compared to [**long-purdon-etal:06**, **giralso-dendekker-etal:10**, **lamus-etal:12**], which allows us to express more complex source configurations. Furthermore, in addition to anatomical connections obtained from dMRI (like in the work of [**fukushima-yamashita-etal:13**]), we also assume that neighboring cortical regions are connected. Part of this work was presented in [**belaoucha-etal-embs:16**, **belaoucha:tel-01534876**].

2 METHODS AND PROCEDURES

2.1 *The iSDR method*

In our method, we assume that sources' dynamics follow a stable MAR model in the time window of interest $[1, T]$ whose elements are constrained by neighboring and anatomical connected regions.

$$\mathbf{J}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{J}_{t-i} + \varepsilon_\mathbf{t} \,,$$

where the matrices $\mathbf{A}_i$, $i \in [1, p]$ contain non-zero elements only for neighbor sources or sources connected through white matter.

This formulation provides more degree of freedom in the model compared to [**Fukushima2015408**] as it allows two brain regions/sources to interact at p different time lags which allows us to explain complex brain activations.

The T time samples in the window of interest are numbered in the range $[1, T]$. We rewrite the MxNE functional by considering only the EEG/MEG samples in the range $[p+1, T]$ (as the ones in the range $[1, p]$ depend – through the MAR model – on samples that are outside the window of interest) and relate them to sources magnitudes in the range $[1, T-1]$. The rewritten functional takes the form:

$$U(\mathbf{J}) = \|\mathbf{M}_v - \mathbf{G}_d \mathbf{J}_v\|_2^2 + \lambda \|\mathbf{J}\|_{21} \qquad\qquad \textbf{Equation } 5.$$

where

- $\mathbf{M}_v = \text{vec}(\mathbf{M})$, $\mathbf{M} \in \mathbb{R}^{N_c \times (T-p)}$, represents the measurements between $p+1$ and T.

- $\mathbf{J}_v = \text{vec}(\mathbf{J})$, $\mathbf{J} \in \mathbb{R}^{N_s \times (T-1)}$ contains the sources' activity between the first time sample and $T-1$.

- $\mathbf{G}_d \in \mathbb{R}^{N_c(T-p) \times N_s(T-1)}$ is a spatio-temporal lead-field that will be further described below.

$vec(.)$ is the vectorization operator that transforms a matrix into a vector by stacking all it columns. It is worth mentioning that while the first p measurements are not part of the criterion 5, the first p source values are not neglected as they are predicted from the future MEG/EEG measurement $t > p$. The matrix $\mathbf{G}_d$ is a block matrix of the form:

$$\mathbf{G_d} = \begin{bmatrix} \mathbf{G}_p & \mathbf{G}_{p-1} & \cdots & \mathbf{G}_2 & \mathbf{G}_1 & & & \\ & \mathbf{G}_p & \mathbf{G}_{p-1} & \cdots & \mathbf{G}_2 & \mathbf{G}_1 & & \\ & & \ddots & \ddots & & & \ddots & \ddots \\ & & & \mathbf{G}_p & \mathbf{G}_{p-1} & \cdots & \mathbf{G}_2 & \mathbf{G}_1 \end{bmatrix}, \qquad \text{Equation 6.}$$

where $\mathbf{G}_i = \mathbf{G}\mathbf{A}_i$, $i \in [1, p]$, with $\mathbf{G}$ being the standard lead-field matrix. Our approach can be seen as an extension to the work of [**gramfort:hal-00690774**] in which, in addition to spatial sparsity, we constrain the dynamics of sources interactions. Thus activated regions are constrained to be spatially sparse in a time window of T samples.

The data fidelity term in Equation 5 includes errors coming both from the MAR model and source reconstruction. This functional can be seen as an extension to MxNE [**gramfort:hal-00690774, irMxNE**]. The major improvement of our method with compared to MxNE and irMxNE [**irMxNE**] is that the interactions between sources are included in the functional. In our work, the weights are not only in the diagonals of A_i matrices like in irMxNE but also contains nonzero elements that correspond to the neighbors or anatomical connections. These weights are estimated from the following functional:

$$V(\mathbf{A}) = \|\mathbf{J}_v^* - \mathbf{Q}'\mathbf{A}_v'\|_2^2 \qquad \text{Equation 7.}$$

where $\mathbf{A}_v = vec(\mathbf{A})$ is the vectorial form of $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1, \cdots, \mathbf{A}_p \end{bmatrix}$. $\mathbf{A}_v$ is a vector of size pN_s^2. $\mathbf{A}_v'$ is the sub-vector of the non-zero components of $\mathbf{A}_v$ and can be written $\mathbf{A}_v' = \mathbf{S_A}\mathbf{A}_v$. $\mathbf{S_A}$ is a selection matrix of size $K_A \times pN_s^2$, where K_A is the total number of the non-zero coefficients in the various $\mathbf{A}_i$ $i \in [1, p]$. $\mathbf{S_A}$ is a rectangular matrix where each row contains exactly one single non-zero coefficient equal to 1. $\mathbf{J}_v^*$ is the vectorial form of sources' magnitudes in the range $[p+1, T]$. $\mathbf{Q}$ is a $N_s(T-p) \times pN_s^2$ matrix:

$$\mathbf{Q} = \begin{bmatrix} \mathrm{D}(\mathbf{J}_p) & \mathrm{D}(\mathbf{J}_{p-1}) & \cdots & \mathrm{D}(\mathbf{J}_1) \\ \mathrm{D}(\mathbf{J}_{p+1}) & \mathrm{D}(\mathbf{J}_p) & \cdots & \mathrm{D}(\mathbf{J}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{D}(\mathbf{J}_{T-1}) & \mathrm{D}(\mathbf{J}_{T-2}) & \cdots & \mathrm{D}(\mathbf{J}_{T-p}) \end{bmatrix} \qquad \text{Equation 8.}$$

where $\mathrm{D}(\mathbf{J}_k)$ is the $N_s \times N_s^2$ matrix:

$$\mathrm{D}(\mathbf{J}_k) = \begin{bmatrix} \mathbf{J}_k^T & 0 & \cdots & 0 \\ 0 & \mathbf{J}_k^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{J}_k^T \end{bmatrix}$$

$\mathbf{Q}' = \mathbf{Q}\mathbf{S_A}^T$ is the submatrix of $\mathbf{Q}$ of size $N_s(T-p) \times K_A$, which contains only the columns corresponding to the non-zero entries of $\mathbf{A}_v$.

The Maximum Likelihood (ML) solution minimizing $V(\mathbf{A})$ for the MAR coefficients $\mathbf{A}$ is [**Weisberg-etall:05**] (Ordinary Least Square solution):

$$\mathbf{A}_v = \mathbf{S_A}^T \mathbf{A}'_v = \mathbf{S_A}^T (\mathbf{Q}'^T \mathbf{Q}')^{-1} \mathbf{Q}'^T \mathbf{J}^*_v \qquad \text{Equation 9.}$$

The full optimization problem of simultaneously minimizing $U(\mathbf{J})$ and $V(\mathbf{A})$ over $\mathbf{J}_v$ and $\mathbf{A}$ is non-convex, but can be solved by solving a sequence of weighted convex optimization problem (MxNE) with weights being defined based on the previous sources' estimates by minimizing the functional V.

$$\begin{cases} U(\mathbf{J}^k) = \left\| \mathbf{M}_v - \mathbf{G_d}^{k-1} \mathbf{J}^k_v \right\|_2^2 + \lambda \left\| \mathbf{J}^k \right\|_{21} \\ V(\mathbf{A}^k) = \left\| \mathbf{J}^{*k}_v - \mathbf{Q}'^k \mathbf{A}'^k_v \right\|_2^2 \end{cases} \qquad \text{Equation 10.}$$

where $\mathbf{G}_d^k$ is $\mathbf{G}_d$ (Equation 6) obtained from the kth estimate of the MAR model elements and $\mathbf{Q}'^k$ is obtained from the kth iteration of the sources magnitudes using Equation 8. iSDR consists in iterating between two steps. We call S-step (MxNE) the part of iSDR that consists in optimizing U. The A-step consists in estimating the MAR elements by minimizing V from the result of the S-step (i.e. with $\mathbf{J}^*_v$ and $\mathbf{Q}'$ fixed).

S-step:

It estimates the source activation for a given MAR model i.e solving the modified MxNE problem. Block coordinate descent is used to find the solution [**gramfort:hal-00690774**, **belaoucha-etal-embs:16**, **belaoucha:tel-01534876**].

A-step:

It consists first in constructing the matrix $\mathbf{Q}'$ and the vector $\mathbf{J}^*_v$ from the output of the S-step i.e. $\mathbf{J}^k$. Then, it estimates the MAR elements using Equation 9 and each MAR elements of a source is normalized by its maximum absolute value (this allows to avoid the scale ambiguity between $\mathbf{A}_i$ and $\mathbf{J}$). The MAR elements are saved in the vector $\mathbf{A}^k_v$. There is no need to re-scale the sources correspondingly as those will be re-estimated during the next S-step, which will take account of the re-scaling.

Improving the convergence speed

If a source s, at a given iteration k, is found to be inactive, there exists no causality between this source and the remaining sources. To reduce the computation time, the A-step removes the corresponding MAR weights of this source and consequently the corresponding columns of $\mathbf{Q}'$. We also remove the contribution of inactive sources into the measurement i.e. the rows of $\mathbf{Q}'$ and $\mathbf{J}^*_v$ that correspond to inactive sources[1]. Consequently, if source s is inactivated at a given iteration k, it will remain inactive for all subsequent iterations. This also means that the sparsity of sources promoted by the regularization term of MxNE also promotes sparsity in $\mathbf{A}$.

Initialization

The algorithm starts with an S-step. Thus the first source estimation depends on the initialization of the MAR model. The effect of different initialization of A_i is outside the scope of this paper[2]. In this work, we initialize the MAR model in such a way to obtain an initial S-step solution close to the MxNE estimate. This is done by initializing $\mathbf{A}_1$ to the identity matrix and the remaining $\mathbf{A}_i$, $i \in [2, p]$ to zero matrix.

Stopping criteria

---

[1] While this can be seen as updating the matrix $\mathbf{S_A}$ and introducing another selection matrix on sources, it is better achieved by simply removing rows and columns in the matrices

[2] The interested reader is referred to [**belaoucha:tel-01534876**]

iSDR includes three stopping criteria. The first is when a maximum number of iterations is reached. The second tests if there is no significant change in the iSDR estimates between two successive iterations. The last stopping criterion compares the size of active sources obtained in iteration k and $k-1$. If they are the same, the algorithm stops.

2.2 *Experimental data*

We used the real dataset described in Wakeman et al. [**dataset**]. MEG/EEG data were simultaneously recorded during a face recognition task where a subject is shown famous, unknown or scrambled faces. The dataset also contains dMRI and T1 images. The stimuli (photos) were projected onto a screen (black background with a white fixation cross in the center) approximately 1.3 m in front of the participants. The start of the trial was indicated by the appearance of the fixation cross for a random duration. A gray-scale photograph was then projected onto the screen for a random duration between 800 and 1000 ms. The interval between two successive face projections is 1700 ms [**dataset**]. Both MEG (Elekta Neuromag Vectorview 306 system, 102 magnetometers, and 204 gradiometers) and EEG (70 electrodes conforming to the extended 10-10 system) were measured in a light magnetically shielded room. The data was presented to the subjects in six runs of 7.5 min each. In this work, we have used the data of the eleven subjects that have T1 and dMRI images.

2.3 *Data pre-processing*

Environmental noise was removed using Signal Space Separation (SSS) [**Taulu-etal:05**]. The EEG and MEG data were filtered using a low-pass filter with a cutoff frequency of 35 Hz. This frequency was chosen because some studies reported the presence of gamma-band in complex visual stimuli [**Lachaux2005491**]. We averaged the data of each condition (Familiar, Unfamiliar, Scrambled). We focus on the brain regions responsible for face perception and recognition. To do so, we subtracted the average measurement with famous pictures to that of scrambled faces.

Down-sampling

In [**Horowitz2015**], the authors show a correlation between the axon diameter and conduction velocity in the human brain. They have found the ratio between axon diameter and the axon conduction velocity to be 8.75 m/s for each μm, which coincide with what was obtained before in [**Aboitiz1992143**]. This results in speeds between 10 and 40 m/s for the majority of fibers. The average in this case is 25 m/s. The length of bundles can range between 50mm for U-shape connections to 100 mm for long connections. Thus the time needed to travel from the source to the target region is between 2 to 4 ms. This is why we decided to down-sample the measurements to 367 Hz (1 sample every 3ms).

MEG/EEG forward model

The T1 MRI images were processed with BrainStorm [**brainstorm**] to extract the geometrical model of the head. The forward EEG/MEG model was obtained using OpenMEEG [**Gramfort2010**] from BrainStorm considering five surfaces: scalp, inner and outer skull and white and grey cortical surfaces.

Cortical surface parcellation

The cortical surfaces of all participants were parcellated into regions using the MNN approach and *Tanimoto* similarity measure with parameter $T = 800$ [**belaoucha-clerc-etal:16**] using structural connectivity obtained by fiber tracking. It results in cortical regions of around $100mm^2$ [**belaoucha:tel-01534876**], which is about the minimum size of detectable activations [**RevModPhys_65_413**]. The parcellation algorithm resulted in an average of 592.36 regions with a standard deviation (SD) equal to 19.75. In this work, we assume a constant activation per cortical region. This allows us to greatly reduce the dimension of the source space from the number of nodes of the

cortical mesh ($\tilde{1}0k$ typically used in the distributed sources model) to the number of cortical regions.

Fiber tracking

In the literature, two types of local fiber tracking can be found. Deterministic tracking follows the direction of fibers at each image voxel [**basser-pajevic-etal:00**]. In this model the tracking error is cumulative. Probabilistic tracking thus introduces a stochastic dimension in the estimation of fiber direction at each image voxel [**Smith-Jenkinson-eta behrens-johansen-berg-etal:07**]. Consequently, it is less affected by error in the local fiber direction. This is why we used the latter method as implemented in FSL [**Smith-Jenkinson-etal:04**] to estimate the structural connectivity between brain regions.

2.4 *Choice of regularization parameter value*

In [**belaoucha:tel-01534876**, **belaoucha-etal-embs:16**], λ is obtained using a cross-validation technique [**arlot2010**]. But because this is very time consuming when using the MxNE based solver, we decided to use the following approach, which gives similar results.

We want to consider the complexity of the obtained MAR model as measured by p. For low values of p, more sources are detected to be active which yields a high number of effective connections to be estimated. For bigger values of p, fewer regions are detected to be active hence fewer parameters need to be estimated, which results in higher error in explaining the EEG/MEG measurements. We thus need to find a middle ground between explaining well the measurements and minimizing the complexity of MAR model. To do so, we consider the following criterion:

$$CR = \frac{\|\mathbf{M} - \mathbf{M}_e\|_2}{\|\mathbf{M}\|_2} + \sqrt{p}\frac{N_A}{K_A}$$

Equation 11.

where $\mathbf{M}_e$ is the reconstructed measurement, N_A is the number of estimated non-zero MAR elements and K_A is the total number of allowed connections. The first term is increases with λ whereas the second term is negatively correlated to λ. The optimal λ for each EEG/MEG measurements is the value that corresponds to smallest CR.

3 **RESULTS AND DISCUSSION**

We start testing iSDR by using synthetic data and compare its results to the ones obtained from 2 other EEG/MEG source reconstruction methods (MxNE and Lasso). In the second part, we use the real data obtained from [**dataset**]. The iSDR reconstruction is compared to the results obtained by other works using the same dataset and to other results based on different datasets but using the same task i.e. face recognition.

3.1 *Synthetic data*

In this section, we use synthetic data to investigate the accuracy of iSDR method. We compare its reconstruction accuracy with those of two other sparse approaches, MxNE and Lasso. It is clear that the comparison is somewhat unfair to those methods as the model used to generate the simulated data is the one used in the iSDR approach. On the other hand, it should be remembered that iSDR explains the data with much less parameters than MxNe or Lasso. Elements of the lead-field matrix (10×400) are randomly generated from a zero-mean normal distribution with a unit standard deviation. Each column, contribution of a source to the measurement, is then normalized to give the same importance to the different sources. For putative connexions, we randomly connect five sources to each source.

We use the following simple model:

$$\begin{cases} J_t(S1) & = & 0.9J_{t-1}(S1) - 0.15J_{t-1}(S2) \\ J_t(S2) & = & 0.25J_{t-1}(S1) + 0.97J_{t-1}(S2) \end{cases}$$

with $J_0(S1) = 10$ and $J_0(S2) = 0$ where $S1$ and $S2$ are two connected sources. We then select randomly two such connected sources 100 times using the previous MAR model. This temporal model is transformed in measurements by applying the lead-field and additive noise is added (at the sensor level). This procedure was repeated for different signal-to-noise ratio (SNR). Results are presented in Figure 1. The optimal regularization parameter λ for both MxNE and Lasso was found using a 3-fold cross validation technique. For iSDR, we use the technique explained earlier.

(a) Reconstruction accuracy   **(b)** Source focality   **(c)** S_0 estimation
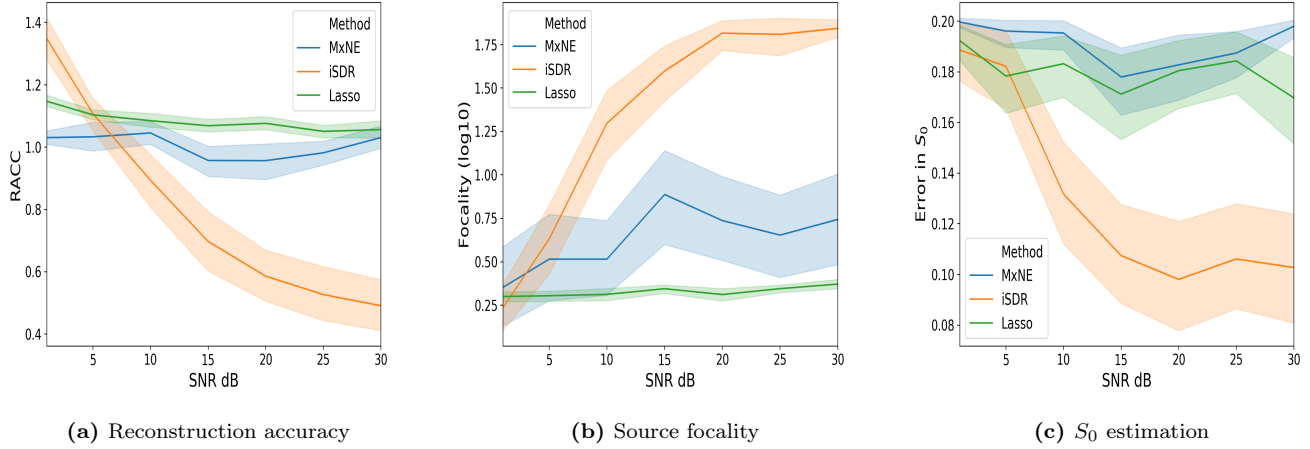
Figure 1. Results of reconstruction using synthetic data. Mean and 95% CI of (a) RACC (b) source focality in log10 scale and (c) Euclidean distance between the estimated initial and the estimated values of S1 and S2.

In Figure 1a, we show the reconstruction accuracy (RACC) for different SNR levels and reconstruction methods. It is computed as follows:

$$RACC = \frac{\|\mathbf{J}_{gt} - \mathbf{J}_e\|_2}{\|\mathbf{J}_{gt}\|_2}$$

Equation 12.

where $\mathbf{J}_{gt}$ and $\mathbf{J}_e$ are the simulated and reconstructed source activation respectively.

iSDR outperforms both MxNE and Lasso as soon as the SNR is bigger than $6dB$. At lower SNR levels, MxNE gives slightly better reconstructions. This comes from the A-step as the MAR model elements are estimated from a noisy source activation estimation. Low SNR leads to inaccurate estimation of MAR elements and thus in inaccurate reconstruction of source activation in the next step.

Figure 1b shows the focality of the reconstructed source space which is computed as the intersection between the estimated and real active sources in the time window divided by the number of the reconstructed active sources. The value is then converted to a percentage and we compute its log10 value. A perfect value is $log10(100) = 2$. The value $log10(50) = 1.7$ corresponds to the case where the method reconstructed two active sources when only one was truly active. Lasso gives the worst results ($\approx 0\%$). This is mainly due to a large number of detected active sources over the whole time window as Lasso assumes independent time samples. iSDR gives better focal reconstructions compared to MxNE.

In any MAR model, the time course depends not only on the eigenvalues of the model but also on the initial values of sources. That is why we also investigated the accuracy of the estimated initial values of S1 and S2 by computing the mean Euclidean distance between the estimated and real values (see Figure 1c). iSDR gives a better estimation of the initial source activation compared to both MxNE and Lasso. This is due to the estimation of the dynamics which helps in the source reconstruction of the initial values. It can also be shown that iSDR can accurately reconstruct the effective connectivity between brain regions [**belaoucha-etal-embs:16**, **belaoucha:tel-01534876**].
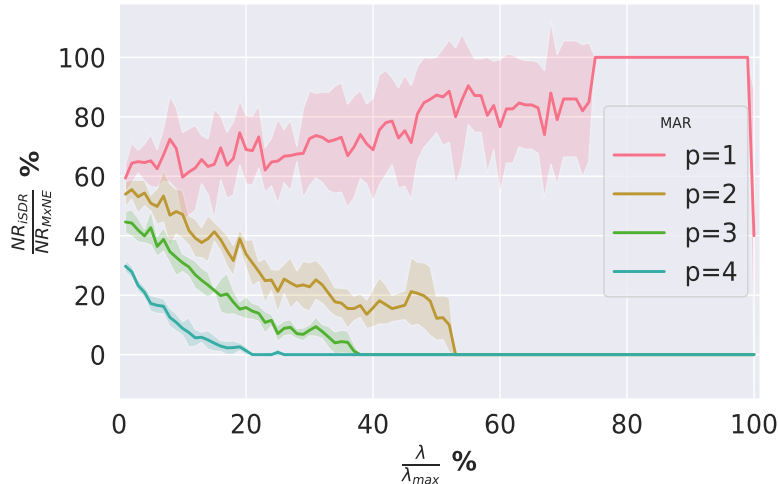
Figure 2. Mean and standard deviation of the relative number of active regions, compared to MxNE, detected by iSDR for different MAR models and λ values. λ_{max} is the smallest λ value that results to empty active set.

3.2 Real data

Effect of MAR order

In this part, we want to investigate the effect of increasing the MAR model order p on the number of reconstructed active regions. The real dataset from [**dataset**] is used while varying the regularization parameter between 1 to 100% of λ_{max} (the smallest λ that results in an empty active set). Four MAR models were tested: $p \in \{1, 2, 3, 4\}$. The results are shown in Figure 2. The number of active regions is inversely proportional to both λ and MAR order. It can also be seen that λ_{max} decreases with p. Furthermore, it can be seen that the behaviour on the iSDR model is more consistent for $p \in [2, 4]$ than for $p = 1$ where the fraction of number of sources gradually increases with λ before brutally falling to zero because the value of λ is equal to λ_{max}. Generally speaking, iSDR explains the data with less activated regions than MxNE (except for $p = 1$ and $\frac{\lambda}{\lambda_{max}} > 75\%$ where iSDR has the same active regions as MxNE).

When increasing the regularization parameter, the number of active regions is decreased which is expected due to the soft-threshold function in the MxNE solver. When increasing the MAR order, a sparser solution is obtained which can explain the EEG/MEG data because it can represent more complex time courses. iSDR gives sparser solution compared to MxNE even with the same λ, this is due to the step value of MxNE which is smaller than its value when no MAR model is assumed which makes the solution of iSDR sparser than MxNE.

Optimal values for λ

The mean and 95% confidence interval, computed across subjects, of CR (Equation 11) for the different MAR models ($p = 1 \ldots 4$) and λ values are presented in Figure 3. An optimal value of the regularization can be found using this approach which corresponds to λ that results in the minimum value of CR.

We can see that on average the optimal regularization parameter is getting bigger when increasing the MAR order. Using a same regularization parameter, we can see from Figure 2 that the number of reconstructed active regions/sources is inversely proportional to the MAR order. The higher is the MAR order, the higher is also the complexity of the reconstructed time course activation and the less is the number of reconstructed activation needed to explain the EEG/MEG data.
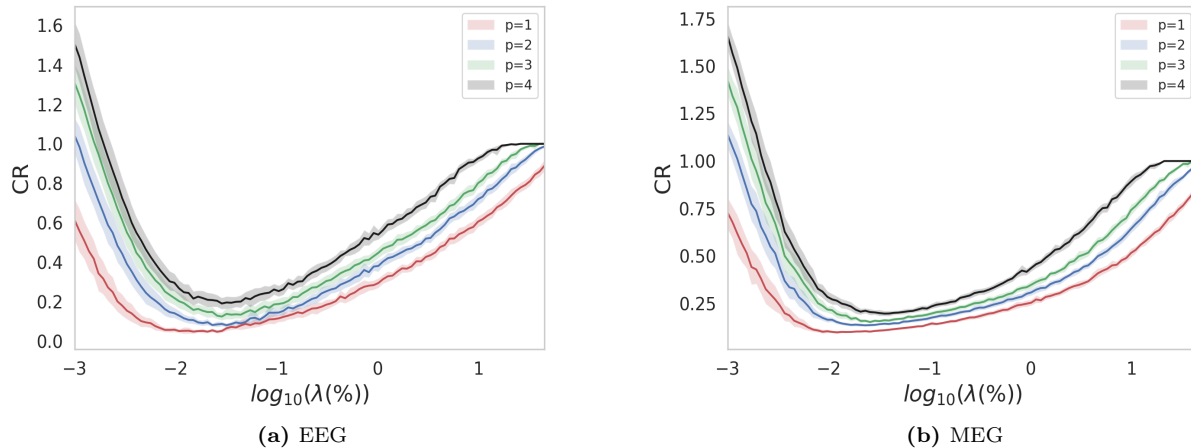
Median activation

(a) EEG   **(b)** MEG

Figure 3. Mean and 95% confidence interval (CI) of CR, over the 11 subjects, computed using Equation 11 for different λ values as percentage value of λ_{max} and MAR order (p) with (a) EEG and (b) MEG data.

iSDR is applied with different MAR models ($p \in \{1, 2, 3, 4\}$) to the MEG and EEG data of the dataset [**dataset**]. To date, although no definite clear picture has emerged, several imaging studies have suggested that the right Fusiform area (FG), right temporal pole (TP) and posterior superior temporal sulcus (STS) might play a role in face recognition [**Kanwisher4302**, **Simpson15088**]. We show the median time course, with different MAR orders, of the reconstructed right FG region activation in Figures 4a and 4b obtained from EEG and MEG respectively. We use the median so that the displayed magnitude is not biased towards subjects with high magnitudes. As a reference, Figure 5 shows the source reconstruction in the same region using the MNE method.

Reconstructions using EEG and MEG show a negative peak at the FG around 200ms which matches what can be found in the literature [**halgren-etal:00**, **journal.pone.0071408**, **Simpson15088**, **dataset**, **Henson01072003**, **TRUJILLOBARRETO2008318**, **Halgren-etal**]. The magnitude of the FG peak when $p = 1$ is low compared to the remaining MAR orders. This is due to the simplicity of the model which could not reconstruct accurately the time course in the FG region. Because of this simplicity, iSDR reconstructs a high number of active regions which in turn exhibit reduced magnitudes due to the l_2-norm used in the MxNE method. The time courses obtained for $p > 1$ are comparable to what has is shown in [**Simpson15088**]. A second peak at around 300ms is observed in the MEG reconstruction with $p = 4$ which may be due to the high MAR order which will explain better the subject activation variability. This peak is similarly present only in the MEG reconstruction using MNE approach. It is interesting to note that in [**Barbeau2008SpatioTD**] the author shows, using intracranial EEG, that the average activation of the FG has two peaks separated by 90ms. This coincides with our findings with Figure 4b for $p = 4$ where there are indeed two peaks at 200ms and just after 300ms. This is also found in [**journal.pone.0071408**] for a comparison task of face to body pictures. Nevertheless, studying what is the optimal MAR order for reconstructing source activation for a given task, while being an interesting question, is not the objective of this paper.

Grand average over the subjects

To get an idea about the grand average activation across subjects, we computed the mapping between a reference subject chosen randomly and the remaining ten subjects using optimal transport [**ferradans2013regularized**, **gallardo2018solving**]. Because iSDR uses a soft-threshold function in the S-step (MxNE solver), it underestimates the magnitude of activation. Thus a bias correction step is applied beforehand using the method explained in [**belaoucha:tel-01534876**]. Then we computed the power of the median time courses obtained with iSDR and $p = 2$ in the whole time window across subjects in each cortical region. The resulting brain activation map ob-
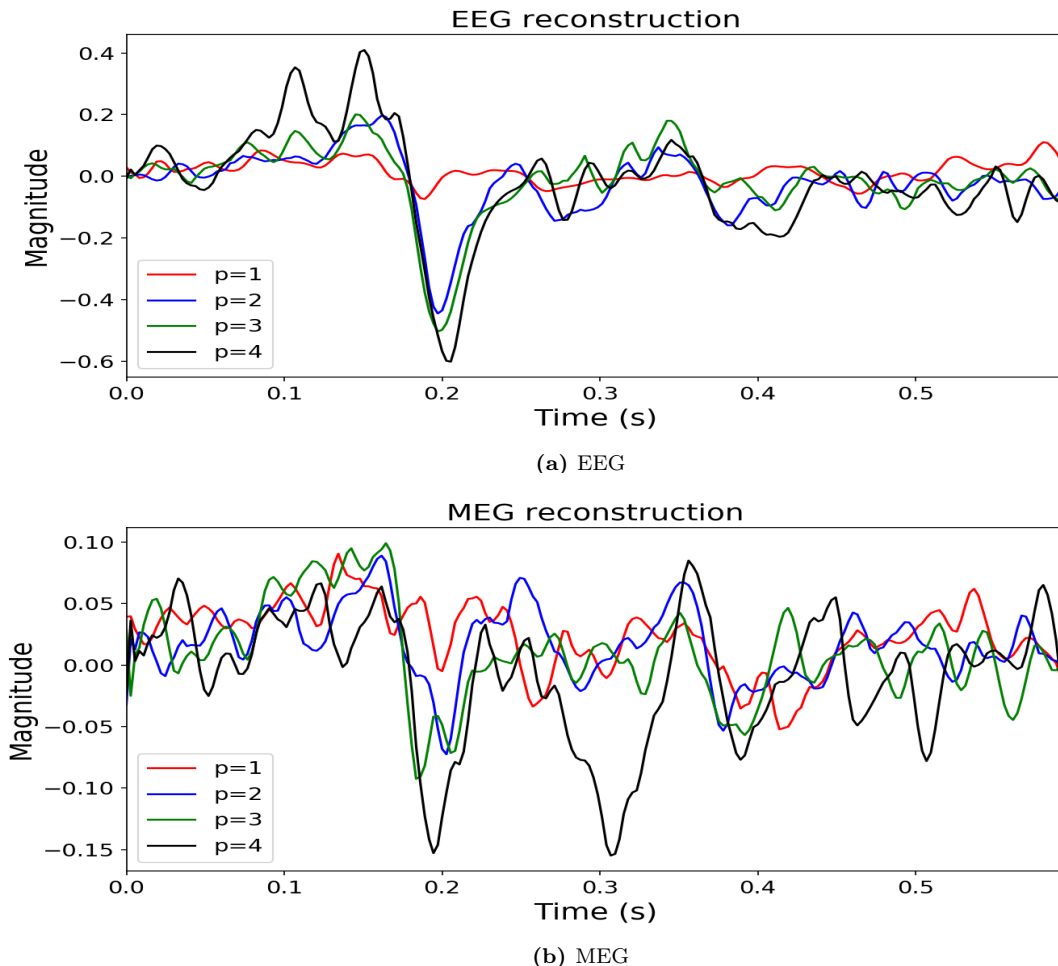
(a) EEG

(b) MEG

Figure 4. The group median activation using iSDR in the right FG computed over the eleven subjects using (a) EEG and (b) MEG data with different MAR orders.

tained from EEG and MEG can be found in Figure 6.

Both EEG and MEG gives comparable active regions. But the FG and inferior frontal activation are more pronounced in the MEG reconstruction. Temporal poles, posterior and superior temporal gyrus are found to be active in EEG and MEG reconstructions which match what was found in [**wakeman2011**] which uses the same dataset.

In Figure 7 and 8, we show the recovered interactions between brain regions, defined by the Mindboggle atlas [**klein2005mind**] computed as a weighted average of the absolute value of the MAR coefficients (with $p = 2$) over the eleven subjects. The weights are chosen as the energy of the corresponding region activation divided by the maximum active region energy. This allows a better visualization of the source activation importance.

Effective interactions are found both within and between brain hemispheres. These were also found in the work of [**Fukushima2015408**]. The right FG interacts with the following areas:

- Left and right cuneus area

- Right entorhinal (close to temporal pole)
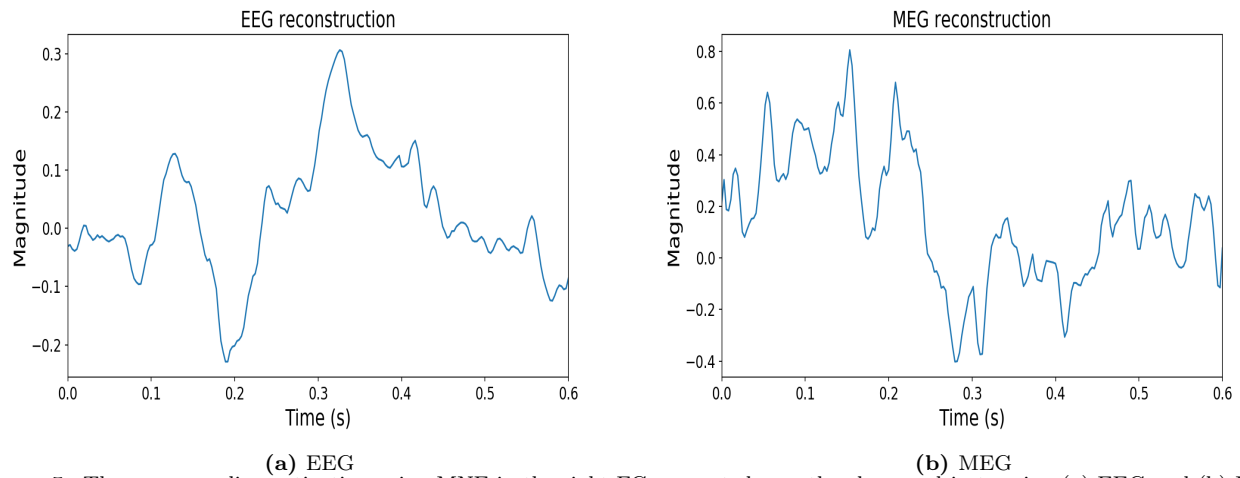
- Right inferior temporal and parietal

(a) EEG
(b) MEG

Figure 5. The group median activation using MNE in the right FG computed over the eleven subjects using (a) EEG and (b) MEG data.
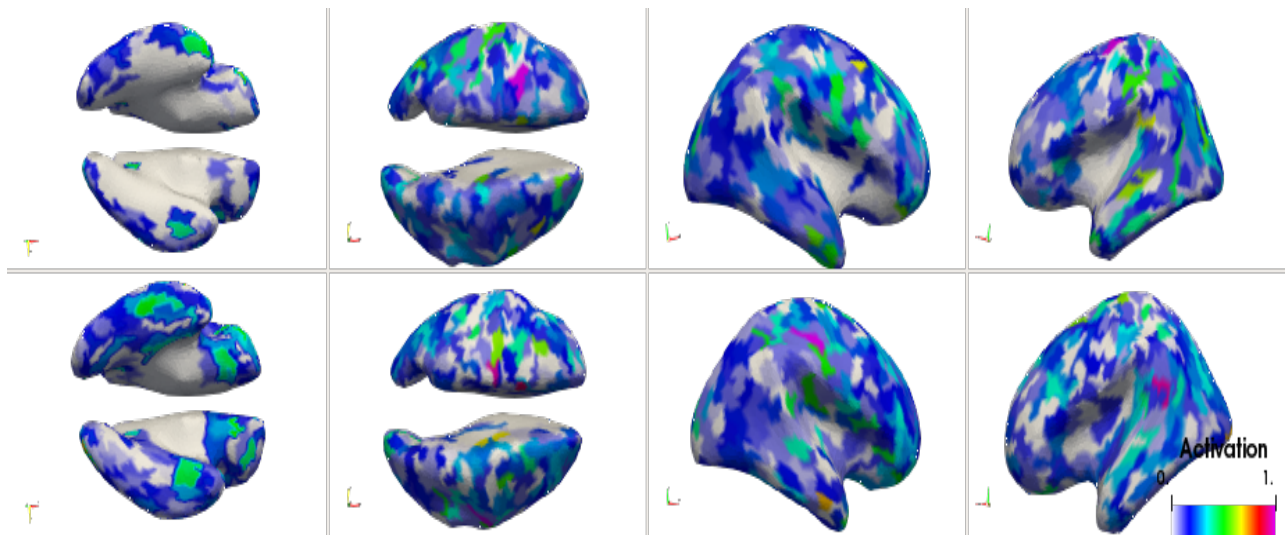

Figure 6. Group average activation of the eleven subjects obtained from iSDR with MAR model order equal to 2. We show the four brain views of the normalized activation obtained from EEG (upper row) and MEG (last row).

- Right lateral occipital and lingual

- Right middle and superior temporal

- Right superior parietal

Up to our knowledge, there is no clear face recognition network that was obtained from EEG/MEG or other imaging modalities. It is thus difficult to confirm whether all or some of the regions listed above play a role in the face recognition task which is outside the scope of this paper. As mentioned earlier, the number of active regions depends on the regularization parameter and MAR order, so different values will result in different active regions hence the effective relationship between brain regions.

Timings and code availability

Unfortunately, we could not compare our finding with the work of Fukushima et al. [**Fukushima2015408**] which also uses a similar MAR model as the code is not published. Their approach takes few days to recover source esti-
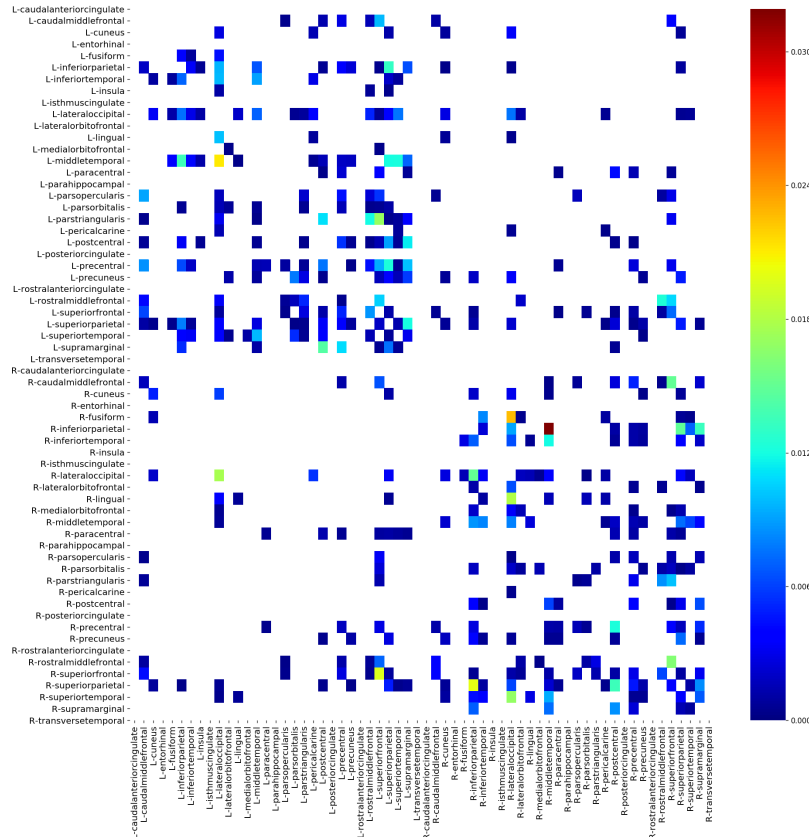
Figure 7. Weighted average of absolute MAR weights, over subjects and A_i's. This corresponds to the effective connectivity between the cortical regions recovered by iSDR with p=2 from EEG data. Weights less than 1% of maximum weight are neglected. Diagonal weights are also removed.

mates whereas iSDR requires only a few minutes with an Intel(R) Core(TM) i5-6200U 4-CPU running at 2.30GHz. A C++ implementation of iSDR is available at https://github.com/BBELAOUCHA/iSDR_p.

4 CONCLUSION AND PERSPECTIVES

In this paper, we have presented a way of reconstructing brain activation and effective connectivity between the brain regions using an extension of the MxNE solver. We have constrained the sources to follow a MAR model of order p. Using the l_{21} norm of MxNE, we obtain a structurally sparse source space as well as a sparse MAR model. The ability of the method to correctly estimate the MAR model is shown with synthetic data for reasonable noise levels. The approach is also quite fast compared to similar competing methods. While the accuracy comparison with MxNE and Lasso with synthetic data is clearly biased as the generated data was using the same MAR model that iSDR uses, we also show that iSDR can obtain reconstruction of quality similar to other works using real data of a face recognition task with much less parameters. The order and the optimal regularization parameter are task-dependent. Further work will investigate the use of information criteria method to select the right parameters of the iSDR reconstruction. In this work, only neighboring regions and regions that are anatomically connected can have non-zero interactions. The same framework can be used to restrict these interactions to known priors or to test various effective network hypotheses. In this work, MEG and EEG data were used separately. Another possible extension of this work is to jointly explain MEG and EEG data with a unique set of activated regions and MAR model.
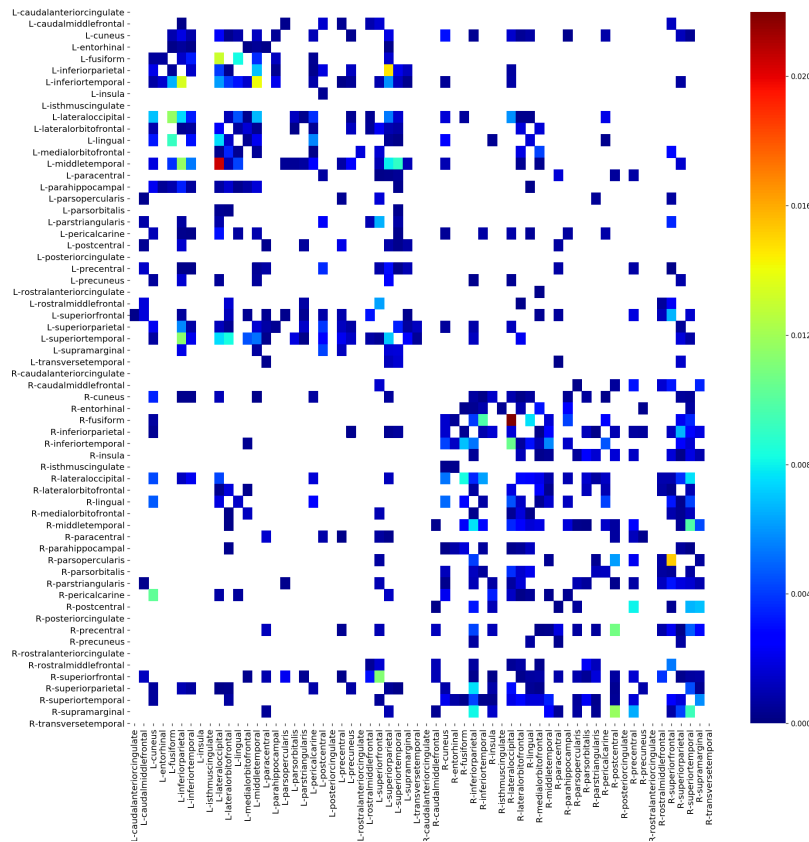
Figure 8. Weighted average of absolute MAR weights, over subjects and A_i's. This corresponds to the effective connectivity between the cortical regions recovered by iSDR with p=2 from MEG data. Weights less than 1% of maximum weight are neglected. Also we neglect the diagonal weights.

Representing the source dynamics by a linear autoregressive model certainly introduces errors to the estimated source interactions because it is a simplification of the non-linear complex source interactions that happen in the brain. Extending our model to include a non-linear source interactions to embrace a wider range of dynamic phenomena and time courses is also an interesting area.

5 ACKNOWLEDGEMENTS