



**HAL**  
open science

## Strategies for statistical thresholding of source localization maps in magnetoencephalography and estimating source extent

Kostiantyn Maksymenko, Bernard Giusiano, Nicolas Roehri, Christian G. Bénar, Jean-Michel Badier

► **To cite this version:**

Kostiantyn Maksymenko, Bernard Giusiano, Nicolas Roehri, Christian G. Bénar, Jean-Michel Badier. Strategies for statistical thresholding of source localization maps in magnetoencephalography and estimating source extent. *Journal of Neuroscience Methods*, 2017, 290, pp.95-104. 10.1016/j.jneumeth.2017.07.015 . hal-03577165

**HAL Id: hal-03577165**

**<https://hal.science/hal-03577165v1>**

Submitted on 16 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Strategies for statistical thresholding of source localization maps in magnetoencephalography and estimating source extent

Kostiantyn Maksymenko<sup>a,b</sup>, Bernard Giusiano<sup>a</sup>, Nicolas Roehri<sup>a</sup>, Christian-G. Bénar<sup>a,\*</sup>, Jean-Michel Badier<sup>a</sup>

<sup>a</sup>*Aix Marseille Univ, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France*

<sup>b</sup>*Project-Team Athena, INRIA Sophia Antipolis, France*

---

## Abstract

Background: Magnetoencephalography allows defining non-invasively the spatio-temporal activation of brain networks thanks to source localization algorithms. A major difficulty of MNE and Beamforming methods, two classically used techniques, is the definition of proper thresholds that allow deciding the extent of activated cortex.

New method: We investigated two strategies for computing a threshold, taking into account the difficult multiple comparison issue. The strategies were based either on parametric statistics (Bonferroni, FDR correction) or on empirical estimates (local FDR and a custom measure based on the survival function).

Results: We found thanks to the simulations that parametric methods based on the sole estimation of  $H_0$  (Bonferroni, FDR) performed poorly, in particular in high SNR situations. This is due to the spatial leakage originating from the source localization methods, which give a ‘blurred’ reconstruction of the patch extension : the higher the SNR, the more this effect is visible.

Comparison with existing methods: Adaptive methods such as local FDR or our proposed ‘concavity threshold’ performed better than Bonferroni or classical FDR. We present an application to real data originating from auditory stimulation in MEG.

Conclusion: In order to estimate source extent, adaptive strategies should be preferred to parametric statistics when dealing with ‘leaking’ source reconstruction algorithms.

*Keywords:* Magnetoencephalography, Source localization, statistical threshold, source extent

---

## 1. Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) can localize neural electrical activity based on noninvasive measurements of neuronal electromagnetic signals. Their excellent time resolution provide a unique window on the dynamics of human brain functions. However, the only way to localize the putative electric sources in the brain is through the solution of an ill-posed inverse problem, which can only be solved by introducing strong *a priori* assumptions on the generation of EEG and MEG signals (Baillet et al., 2001).

Many solutions for solving the inverse problem have been proposed in the literature (reviews in (Baillet et al., 2001; Michel et al., 2004)). Different classes of solutions exist, based on a limited number of dipolar sources (equivalent dipoles), on sources placed along the cortex or on a regular grid within the brain volume (distributed sources) or based on spatial filtering (beamforming).

A major difficulty for distributed sources and beamforming is to find a threshold that will determine the number of active regions and their extent. In particular, extracting the correct extent of active cortical sources is an important issue, in order to estimate which brain regions are involved in a given paradigm (after registration to an atlas), or in presurgical evaluation of epilepsy, where the clinicians have to decide on the amount of cortex to be resected (Rosenow and Lüders, 2001). Within a statistical framework, a hypothesis test can be performed for each brain source, which results in a heavy multiple comparison problem (of the order of several thousand sources). Thus, C. Grova and colleagues have shown that there can exist in some conditions a possible threshold, as the work is based on receiver operating characteristics that tests all possible thresholds (Grova et al., 2006), but does not evaluate a way to actually estimate the threshold in real-life situations (even though Otsu thresholding was proposed for visualization purposes).

---

\*Corresponding Author

*Email address:* christian.benar@univ-amu.fr (Christian-G. Bénar )

Several parametric and non-parametric methods have been proposed to take into account the multiple comparison problem in general (Goeman and Solari, 2014; Efron, 2005) and in neuroscience in particular (Nichols and Hayasaka, 2003; Genovese et al., 2002). However, most attention so far has been directed towards multiple comparison in the analysis of fMRI data, and much less towards EEG/MEG thresholding at the source level (Pantazis et al., 2005). Unlike fMRI, which directly measures every voxel in the brain, EEG/MEG data reconstruction is based on an inverse operator. Such operator can change the noise properties and produce a leakage effect, which reduces spatial resolution (Grave De Peralta Menendez et al., 1997). Random field theory and permutation methods were analyzed in (Pantazis et al., 2005), including a non-parametric framework for setting the threshold, but in practice the parametric methods are often reduced to simple thresholding according to p-values.

The objective of the current article is to propose and validate statistical strategies for thresholding source localization maps in MEG. We used two linear inverse methods - Minimum Norm Estimate and Beamforming (linearly constrained minimum variance). For different patch/SNR configurations we reconstructed source activity and applied different statistical thresholds in order to obtain the estimated patch location, which were compared with the simulated patch.

## 2. Materials and methods

In the following, we assume that the brain surface model is defined, with a description of every source position and orientation. In general, if orientation is not fixed, methods could be easily generalized using a three-dipole source representation.

### 2.1. Minimum Norm Estimate (MNE)

The observation model is:

$$x = As + n, \quad (1)$$

where  $x \in \mathbb{R}^N$  is an observation vector (EEG or MEG) at the fixed time moment,  $N$  is a number of sensors;  $s \in \mathbb{R}^M$  is the vector of source amplitudes,  $M$  is the number of sources;  $A \in \mathcal{M}_{N,M}$  is the gain matrix;  $n \in \mathbb{R}^N$  is the measurement noise.

The prior hypotheses are:  $n \sim \mathcal{N}(0, C)$ ,  $s \sim \mathcal{N}(0, R)$ , i.e. both  $n$  and  $s$  are normally distributed vectors with zero mean and covariance matrices  $C$  and  $R$  respectively. This leads to (Baillet et al., 2001):

$$\hat{s} = (A^T C^{-1} A + R^{-1})^{-1} A^T C^{-1} x = Wx. \quad (2)$$

It can be also shown that:

$$(A^T C^{-1} A + R^{-1})^{-1} A^T C^{-1} = R A^T (A R A^T + C)^{-1}. \quad (3)$$

It is more efficient to use the second expression because it requires the inversion of a matrix that is square in the number of sensors, compared to square in the number of sources (Liu et al., 2002). In practice, the *a priori* source covariance matrix is unknown and we can add a regularization parameter by writing  $R = \frac{R}{\lambda^2}$ . This results in

$$W = R A^T (A R A^T + \lambda^2 C)^{-1}. \quad (4)$$

It should be noted that after some manipulations (whitening and scaling) we can use the following approximation:

$$\lambda^2 \approx \frac{1}{SNR}, \quad (5)$$

where SNR is the (power) signal-to-noise ratio of the whitened data, but in practice we can only estimate this value of SNR (Hämäläinen, 2010).

### 2.2. Beamforming (BF)

The data model is the same as for MNE but we will represent it in another way (Van Veen and Buckley, 1988; Van Drongelen et al., 1996):

$$x = \sum_{i=1}^M A(q_i) s(q_i) + n, \quad (6)$$

where  $q_i$  corresponds to the location of  $i$ -th source,  $s(q_i)$  is associated dipole amplitude, and  $A(q_i) \in \mathbb{R}^N$  the corresponding column of the leadfield matrix.

Every source amplitude is assumed to be a random variable with mean  $\bar{s}(q_i) = \mathbb{E}[s(q_i)]$  and variance  $R(q_i) = \mathbb{E}[(s(q_i) - \bar{s}(q_i))^2]$ . Moreover, we assume that all sources are uncorrelated, and that sensor-level noise is zero mean with covariance matrix  $Q$ . We can calculate the mean and the covariance matrix of the observed data vector  $x$ :

$$\bar{x} = \sum_{i=1}^M A(q_i) \bar{s}(q_i) \quad (7)$$

$$C(x) = \sum_{i=1}^M A(q_i) R(q_i) A^T(q_i) + Q \quad (8)$$

For every source  $q_i$  the objective is to construct an operator (spatial filter)  $W(q_i) \in \mathbb{R}^N$ , such as  $\hat{s}_i = W^T(q_i)x$ . An ideal spatial filter satisfies:

$$W^T(q_i)A(q_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}. \quad (9)$$

In that case we obtain:

$$\begin{aligned} \hat{s}(q_i) &= W^T(q_i)x \\ &= \sum_{j=1}^M W^T(q_i)A(q_j)s(q_j) + W^T(q_i)n \\ &= s(q_i) + W^T(q_i)n, \end{aligned} \quad (10)$$

In the absence of noise this would lead to perfect reconstruction of the source activity. But in the context of the EEG/MEG signals, when  $M > N$ , the perfect filter is not possible. The idea of linearly constrained minimum variance (LCMV) filtering ([Van Veen and Buckley, 1988](#)) is to find  $W(q_i)$  which minimizes the variance of the filter output while satisfying the constraint:

$$\begin{cases} \min_{W(q_i)} \text{Var}(\hat{s}_i), \\ W^T(q_i)A(q_i) = 1. \end{cases} \quad (11)$$

We obtain:

$$\begin{cases} \min_{W(q_i)} W^T(q_i)C(x)W(q_i), \\ W^T(q_i)A(q_i) = 1. \end{cases} \quad (12)$$

The solution of this optimization problem can be obtained using Lagrange multipliers. As the result, the optimal filter is:

$$W(q_i) = [A^T(q_i)C^{-1}(x)A(q_i)]^{-1}A^T(q_i)C^{-1}(x). \quad (13)$$

The source activity vector can be estimated as:  $\hat{s} = Wx$ , where each line of  $W$  is a vector  $W^T(q_i)$ .

### 2.3. Source model and gain matrix computation

The subject's MRI was segmented with the *Freesurfer* software. Head and source model were constructed thanks to the *BrainStorm* software ([Tadel et al., 2011](#)), based on the MRI of a real subject. Brain surface is represented by 15002 vertices (each vertex corresponding to a source). The left hemisphere was divided in 125 patches ([Figure 1](#)). Average patch area was  $8.2 \text{ cm}^2$ . The gain matrix was also calculated in *BrainStorm* with *OpenMEEG* ([Gramfort et al., 2010; Kybic et al., 2005](#)), based on these head and source models and on the position of 246 MEG sensors (magnetometers). Data were exported from *BrainStorm* to *Fieldtrip* ([Oostenveld et al., 2011](#)).

In a second step, and in order to test the impact of patch size, we varied the number of patches between 40 (resulting in an average area of  $25.5 \text{ cm}^2$ ) and 225 (average area of  $4.5 \text{ cm}^2$ ).

### 2.4. Signal simulation

We created realistic background signals based on a multivariate autoregressive (MVAR) model. This allowed obtaining signals with both spatial and temporal correlation resembling that of real signals. Thus, we computed the coefficients of a multivariate autoregressive model (MVAR) of order 40 on real MEG data of the same subject. Background noise was simulated using this MVAR(40) model with Gaussian white noise as input. The obtained signal was divided in 100 trials of simulated MEG signals on 246 sensors. Each trial contained 512 time bins which correspond to 1 second.



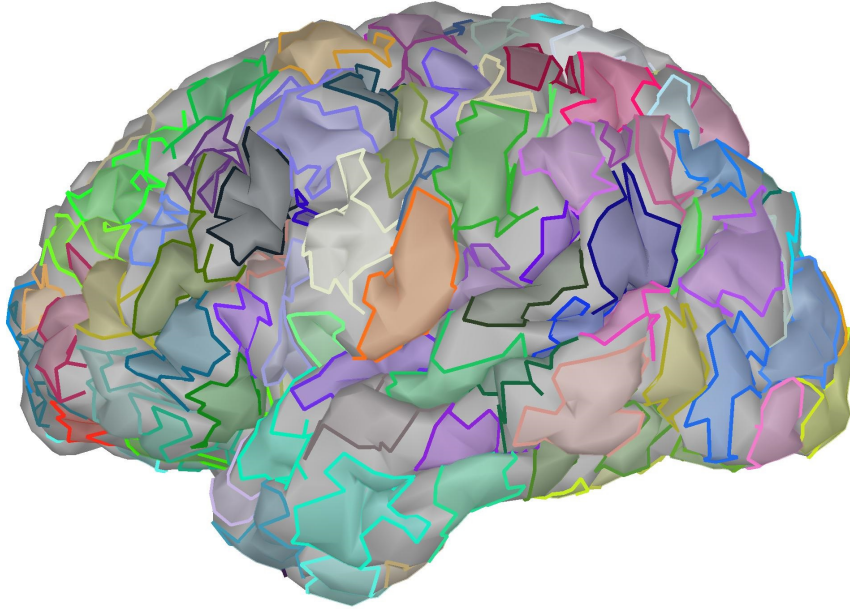


Figure 1: Atlas used for simulations. Only the left hemisphere is shown, with 125 patches.

The signal of interest (SOI) was modeled by a transient wave (Figure 2) (Grova et al., 2006). For each active patch, we summed the leadfield of dipoles in the patch (constrained to be orthogonal to the surface) and multiplied the resulting field by the SOI. We obtained the corresponding signals at channel level by multiplying the SOI with the gain matrix (i.e. we assumed that all dipoles in the patch had the same time course with the same amplitude). The SOI did not vary across trials. We obtained a complete simulation by adding background noise and SOI, taking into account the different SNRs.

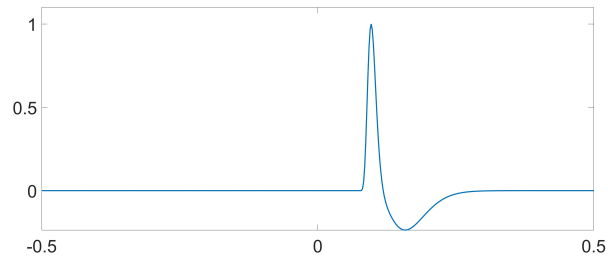


Figure 2: Signal of interest.

The SNR was calculated as a ratio of maximal amplitude of the SOI to the standard derivation of the noise averaged across trials. Noise standard derivation was calculated across all channels. Figure 3 presents examples of signals at different SNRs.

### 2.5. Source localization

In order to reconstruct the source activity from signals at the sensor level, we used two methods: Minimum Norm Estimate (MNE) and Beamforming (BF) (see section 2). These methods were applied to the signal averaged across trials. The continuous information across single trials was used for estimation of the observation covariance matrix, the full time window in the case of BF and the baseline period for MNE (i.e. only the noise). For MNE, the source covariance matrix was assumed to be the identity. By construction of the MNE method, the additive noise is assumed to be normally distributed. As the background noise originates from AR process and because of averaging across many trials we assume that noise is Gaussian.

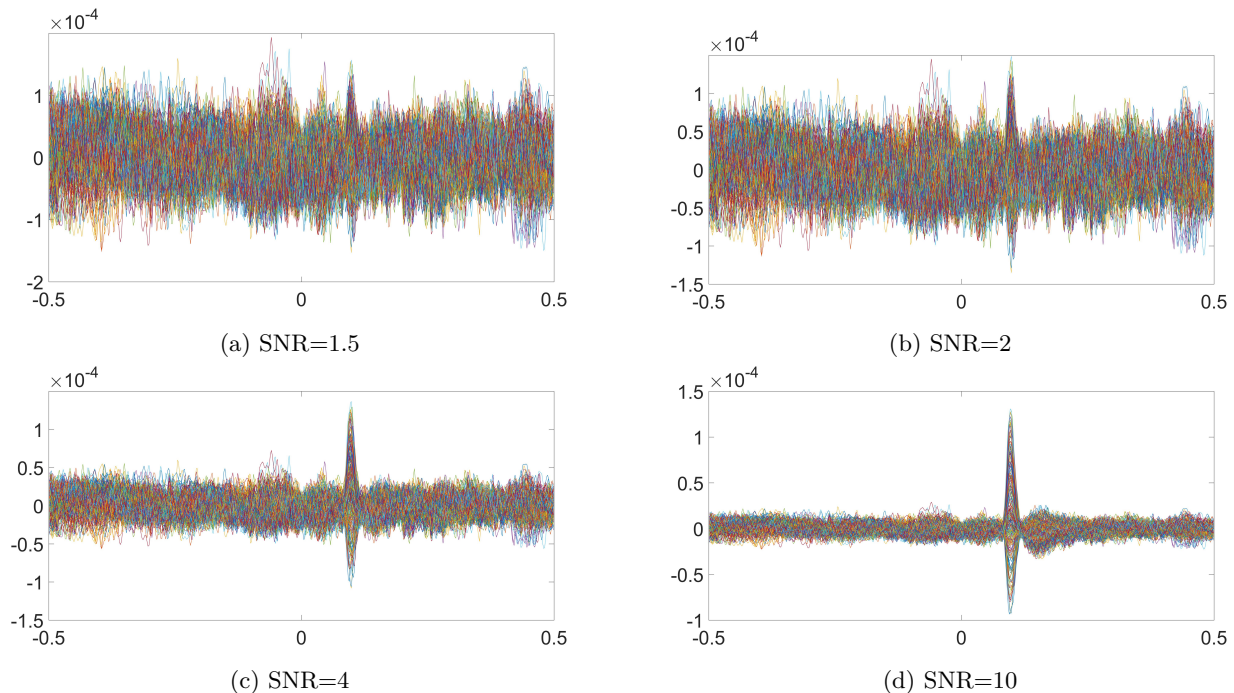


Figure 3: Example of average signals for different SNRs.

## 2.6. Statistical threshold

After application of the inverse operators, we calculated a z-score at source level, i.e. we centered and scaled the time course of each source relative to the baseline. It allows us to set all sources to the same scale of noise, and to correct for depth biases. An uncorrected p-value was computed thanks to the *norminv* Matlab function. To counteract the problem of multiple comparisons we applied the following methods.

### 2.6.1. Bonferroni correction

The simplest and most conservative approach to the multiple comparison problem is the Bonferroni correction. This approach assumes that the  $n$  tests are independent, which results in a threshold on p-values equal to  $\alpha/n$ , with  $\alpha$  desired significance level (Goeman and Solari, 2014; Abdi, 2007). This correction controls the probability of type I error, i.e. the probability of obtaining at least one false detection across all tests. In our case,  $n$  was chosen as the number of source multiplied by the number of time bins.

### 2.6.2. False Discovery Rate

The False Discovery Rate (FDR) is the proportion of incorrect rejections of the null hypothesis (i.e., false detections) among all detections. It is therefore less strict than the Bonferroni correction that tries to avoid to have even one false positive (Family-wise error rate). The FDR computation operates on the ordered p-values, and finds all p-values below a theoretical line (Benjamini and Hochberg, 1995; Nichols and Hayasaka, 2003; Genovese et al., 2002).

Both Bonferroni and FDR are parametric, i.e. they assume that the probability distribution of the noise (i.e., under the null hypothesis  $H_0$ ) is known. The other important property of these methods is that the threshold depends only on noise characteristics, meaning in our case that threshold does not change among the SNR value.

### 2.6.3. Local False Discovery Rates

(Efron, 2005) In this empirical Bayesian method, we assume that every source belongs one of two classes  $H_0$  (noise) or  $H_1$  (SOI) with probability  $p$  and  $(1-p)$  respectively. Each class has its own probability density function, Gaussian for  $H_0$  and arbitrary for  $H_1$ , and the complete distribution is  $H_{tot}=H_0+H_1$ . We also assume that the SOI is distributed around high values while the noise is centered in zero. So we can assume that there is some interval of values around zero that is generated mostly by  $H_0$  sources, where one can estimate empirically the properties of  $H_0$  density. The local FDR threshold is defined as the value for which the proportion of false positives  $H_0/H_{tot}$  passes below a certain level  $q$ .

The threshold is 'local' as it corresponds to one point of the histogram, whereas classical FDR is based on the integral of the probability density function between the threshold and infinity. Local FDR was introduced in the field of neuroimaging by Valdes-Sosa et al (Valdés-Sosa et al., 2005).

In order to estimate the  $H_0$  density, we used a mixture of three normal distributions. We used only data near the peak of the SOI (20 ms window), for better balance between  $H_1$  and  $H_0$  and for reducing computing time. Parameters of the method are thus the model of  $H_0$  and the interval on which to estimate this model.

#### 2.6.4. Concavity threshold

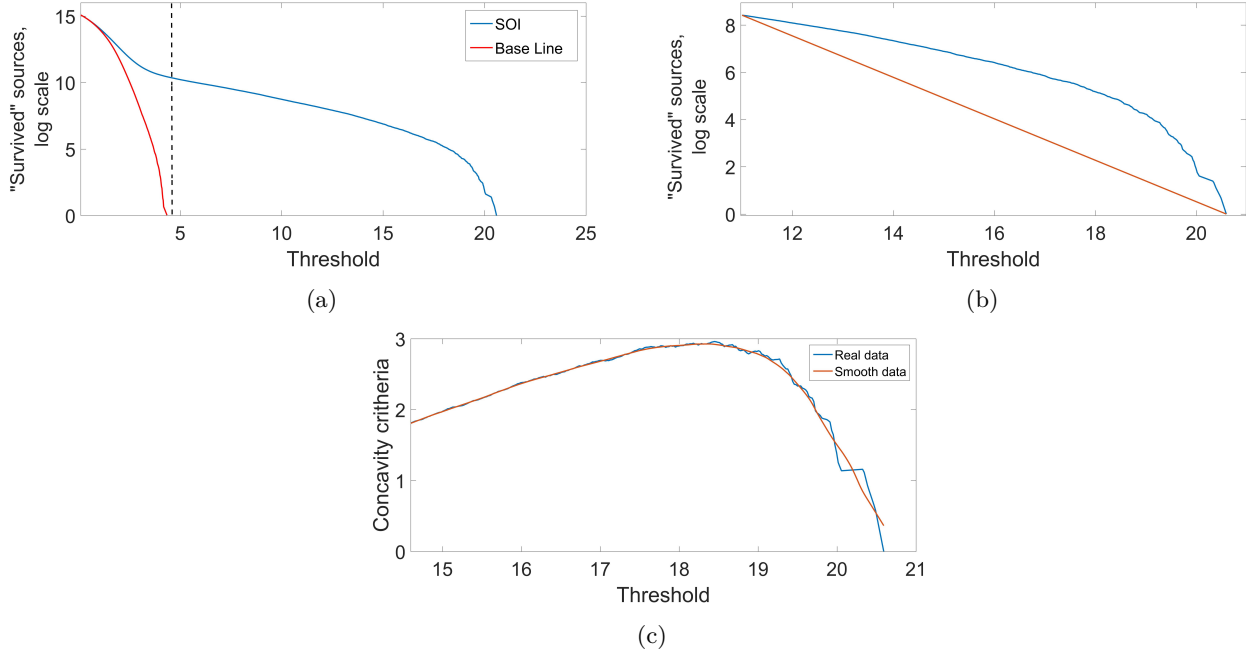


Figure 4: Computation of concavity threshold. a) Survival function, i.e. number of 'surviving sources' as a function of threshold, in log scale. Dashed line: higher value in the background. b) Concave part of the function (i.e., part above the higher value in background). A support line is drawn between the first and the last values (in red). c) Plot of the distance between the survival function and the support line. The concavity threshold is chosen as the value that maximizes this distance.

The principle here is to analyze the empirical survival function of sources, i.e. the number of detected sources according to the threshold value. It is more useful to represent this in a logarithmic scale. The first thing that we can notice is the difference between these function of baseline and SOI period (Figure 4 - (a)). We only analyzed values greater than the maximal baseline. We can notice that this part of survival function is concave. Furthermore, there is a moment when the slopes increases, before the function finally reaches zero (Figure 4 - (b)). The hypothesis is that this inflexion could be a good threshold for detecting active patches. We measured the concavity of this part of survival function by a simple difference with the support line, and the threshold value is the value that maximizes this difference (Figure 4 - (c)).

#### 2.7. Performance quantification

We quantified the accuracy of reconstructed patches (i.e., after thresholding) based on several measures.

The first was the F-measure (Powers, 2011). This measure considers both precision  $p$  and recall  $r$  of the test, with  $p$  number of true positive sources divided by the number of all detected sources (true positives + false positives), and  $r$  number of true positive sources divided by the number of actually activated sources (true positives + false negatives). We can observe that  $p$  is simply related to the (classical) FDR:  $p=1-\text{FDR}$ . The F value is a summary measure (harmonic mean of precision and recall):

$$F = 2 \times \frac{p \cdot r}{p + r}$$

This F-measure permits to handle in a straightforward manner the fact that there are many more negative elements (inactive brain sources) than positive ones (active brain sources), which would render a Receiver Operating Characteristic (ROC) analysis less optimal (however see (Grova et al., 2006) for a proposed solution to this issue).

A limitation of the F-measure is that it does not take into consideration the spatial properties of the localization result. This means that false detections located near to the simulated patch will be considered equally than false detections located in remote areas. Thus, we used a measure based on the Euclidean distance between the sources, the Dipole Localization Error (DLE) defined as (Becker et al., 2014), (Yao and Dewald, 2005):

$$DLE = \frac{1}{2Q} \sum_{k \in I} \min_{l \in \hat{I}} \|r_k - r_l\| + \frac{1}{2\hat{Q}} \sum_{l \in \hat{I}} \min_{k \in I} \|r_k - r_l\|, \quad (14)$$

where  $I$  and  $\hat{I}$  denote respectively the original and estimated indexes of active sources,  $Q$  and  $\hat{Q}$  are the numbers of original and estimated active sources.  $r_k$  denotes the position of the  $k$ -th source dipole in space and  $\|\cdot\|$  is an Euclidean norm in  $\mathbb{R}^3$ . DLE is measured in centimeters.

We also computed an “optimal threshold” – i.e., the threshold that maximizes localization performance. Obviously, this threshold cannot be implemented in real situations, but it shows the best-case properties of a source localization algorithm and gives a reference to which shows other thresholds can be compared. We chose the F-measure to compute the optimal threshold as it is very computationally expensive to compute it for DLE measure.

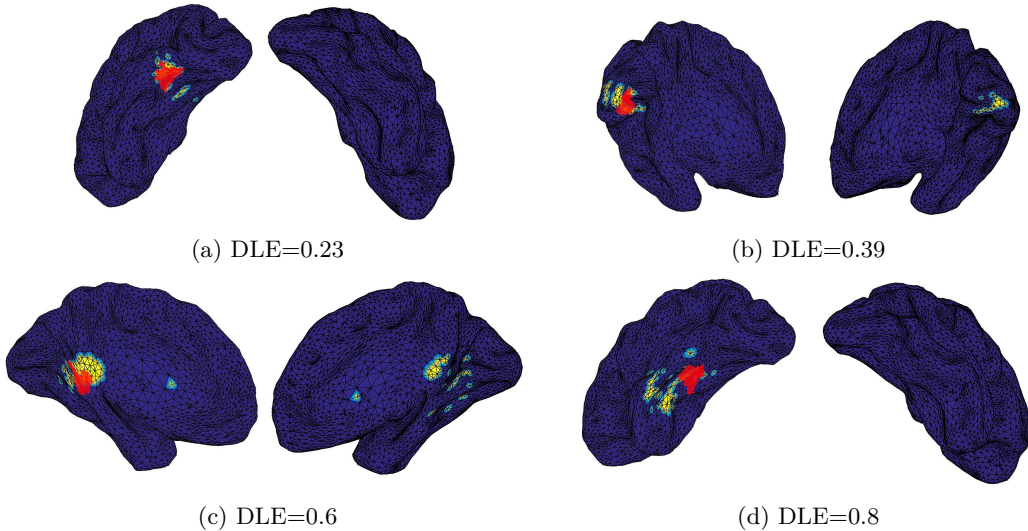


Figure 5: Examples of source detection according DLE value. In red is the simulated patch, in yellow - detected sources. DLE units - cm.

We illustrate in Figure 5 examples of DLE measures. High DLE value correspond to remote false detected sources. In the fourth subplot (DLE=0.8) one can see that detected patch is largely offset. Qualitatively, results with DLE value above 0.8 can be considered as performing poorly.

### 3. Results

For each inverse problem method (MNE and BF) we simulated 125 patches for the following SNR values: 0.25, 0.5, 1, 1.5, 2, 4, 10. In the case of MNE we also tried different values of the hyperparameter lambda: 0.02, 0.2, 2, 20. In total we had 4375 different parameter configurations.

In this section, we present only results for the DLE measure, as results for the F-measure are very similar and lead to the same conclusions (not shown).

#### 3.1. Performance of optimal threshold

The analysis of the optimal threshold gives the potential of each method, as it represents the best possible result in terms of F-measure. The main conclusions of these results (Figure 6) are:

- Optimal lambda value does not depend on SNR for MNE.

- Performance increase according to SNR for both MNE and BF. This seems trivial but we will see in the following that this may not always be the case.
- MNE and BF show, in average, the same performance.
- Starting from SNR = 1 the major part of patches shows a good performance ( $DLE < 1$ )

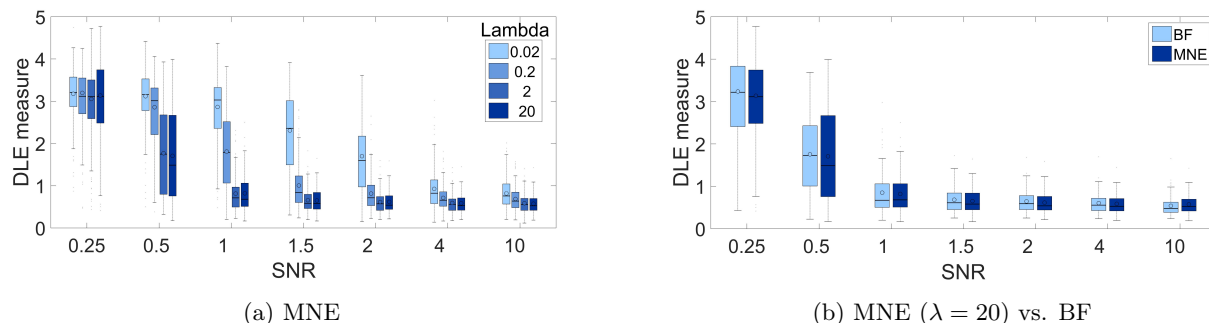


Figure 6: Performance of optimal threshold: (a) MNE performance according to SNR and lambda values. (b) Comparison of MNE and BF according to SNR. We chose  $\lambda = 20$  for MNE. DLE units - cm.

### 3.2. Bonferroni and FDR performance

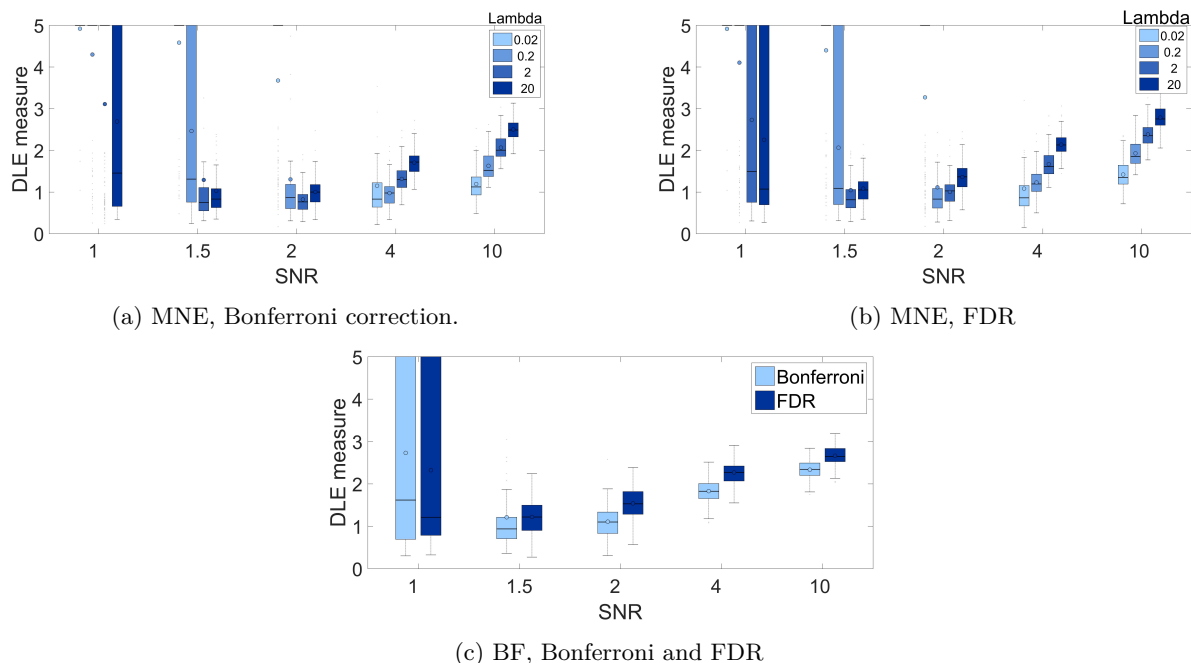


Figure 7: Bonferroni and FDR thresholding performance. DLE units - cm.

Performance of Bonferroni and FDR thresholds is presented in (Figure 7). We can observe two main properties:

- Performance decreases with higher SNR (for both MNE and BF).
- Optimal lambda for MNE depends on SNR.

In contrast to the optimal threshold, Bonferroni and FDR performance decreases with higher SNR. The best performance of these methods is at the 1.5-2 SNR level. Moreover, the best lambda value varies according to the SNR.

The reason why performance decreases with SNR is the spatial leakage effect of inverse algorithms. This leakage has been described in several other studies, and originate from the large point spread function of source reconstruction algorithms (Wens, 2015; Wens et al., 2015). As discussed above, Bonferroni and FDR thresholds do not depend on SNR, while source leakage does: the higher the SNR, the more 'leaked' sources there are, resulting in a higher number of false detections.

### 3.3. Performance of Local FDR and Concavity thresholds

These thresholds are both 'empirical' as they are determined from the data.

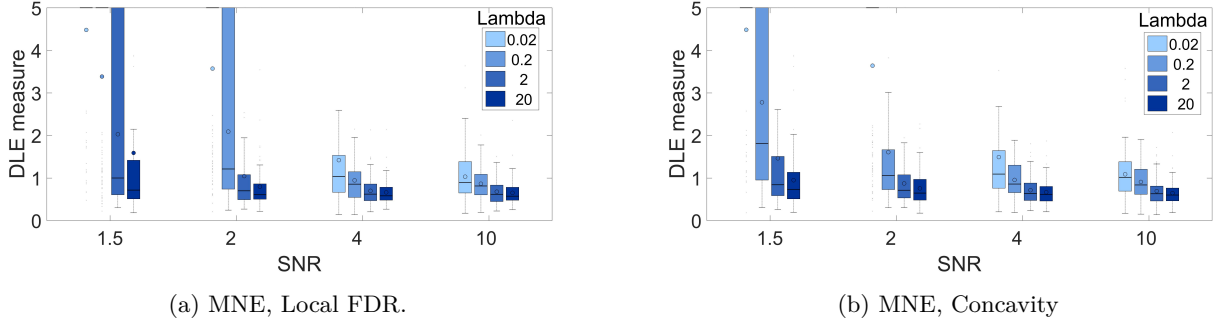


Figure 8: Local FDR and Concavity thresholds performance for MNE. DLE units - cm.

We present the analysis of Local FDR and Concavity thresholds performance for MNE (Figure 8). We can observe the following:

- Performance increases according to SNR (for both MNE and BF).
- Optimal lambda for MNE does not depends of SNR.
- Starting from SNR = 2 the major part of patches present a good performance ( $DLE < 1$ ). For SNR = 1.5 there are still many patches with good results.

Thus, in general, we have the same advantages that we noted for optimal threshold, but - importantly - these methods can be used in practice. We can finally compare the performance of these methods with those of the optimal threshold (Figure 9).

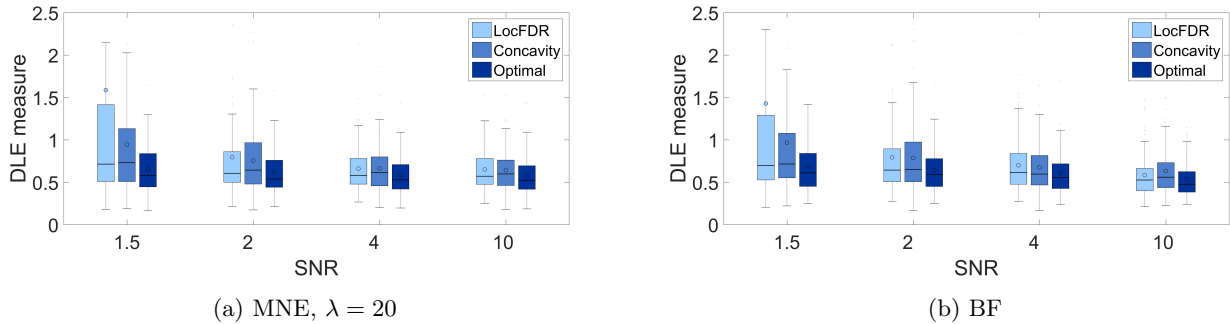


Figure 9: Local FDR, Concavity and Optimal thresholding performance for MNE and BF. DLE units - cm.

We see that Local FDR and Concavity thresholds perform well for both MNE and BF, with performance close to optimal, starting from SNR = 2. We notice that MNE and BF show the same performance on average. We also had the same conclusion for Local FDR and Concavity thresholds comparison. The question remains though of their respective performance for a given patch.

We compared Local FDR and Concavity thresholds across patches (see Figure 10). As we can see, in both figures points vary significantly around the  $y = x$  line. This means that inverse operators and thresholding methods can show different performance for every particular Patch-SNR configuration, even if on average they are similar. So, in practice, it seems interesting to apply both methods.





Figure 10: a) MNE and BF performance by patches (DLE measure). Every point in the figure corresponds to one Patch-SNR pair.  $\lambda = 20$  for MNE. b) Local FDR and Concavity thresholds performance across patches (DLE measure). Every point in the figure corresponds to one Patch-SNR-Inverse operator configuration.  $\lambda = 20$  for MNE. DLE units - cm.

### 3.4. Varying the patch sizes

In order to test the impact of patch size, we applied the beamforming method to simulations build on parcellations with varying number of patches (from 40 to 255, see section 2.3). Results of the DLE measure are shown in Figure 11. One can observe on Figure 11 that the performance of Bonferroni, FDR and local FDR decreases with decreasing patch sizes, but remains above 1 on average. The DLE measure for the concavity threshold stays stable and low up to a size of  $8.2\text{cm}^2$  (corresponding to a parcellation with 125 patches) and increases only for small patches ( $4.5\text{cm}^2$ ); for this fine parcellation, the best method is the local FDR.

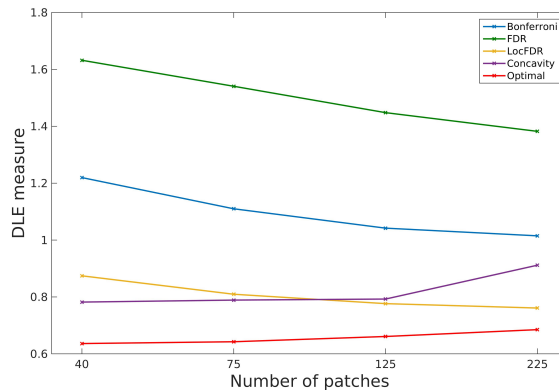


Figure 11: Average DLE measure for different patch sizes and different thresholds applied to the beamforming results.

### 3.5. Test on real data

We tested our methods on real MEG data arising from an auditory protocol (presentation of 'BA' and 'PA' sounds, (Trebuchon-Da Fonseca et al., 2005)). We took 100 trials with a 200 ms baseline and 300 ms of SOI. Figure 12 shows the real signal of all sensors averaged across trials. SNR of obtained signal corresponds to value  $\text{SNR} = 2$  in our simulations.

As can be seen in Figure 13, the Bonferroni correction detects many sources which are likely to be false positives. In contrast, the concavity threshold seems more adequate as results are more focused on putative primary auditory regions. We obtained similar results with BF and Local FDR.

We can notice, interestingly, that the survival functions of real data that we used for concavity threshold calculation have similar form to that of simulations. (Figure 14)

## 4. Discussion

Distributed sources solutions and spatial filters can be used for extracting peaks of activity in a continuous (2D or 3D) 'landscape' of values. In order to define the actual extent of activated cortex, as well as the number of significant sources, one needs to threshold source amplitudes. Grova and colleagues (Grova et al., 2006) have



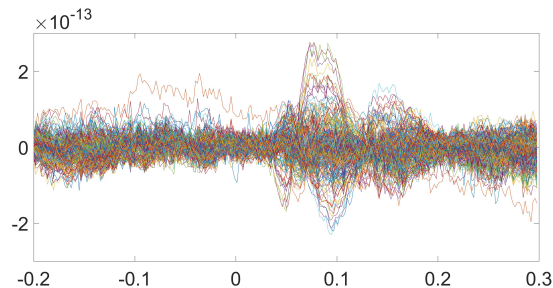


Figure 12: Averaged real signal.

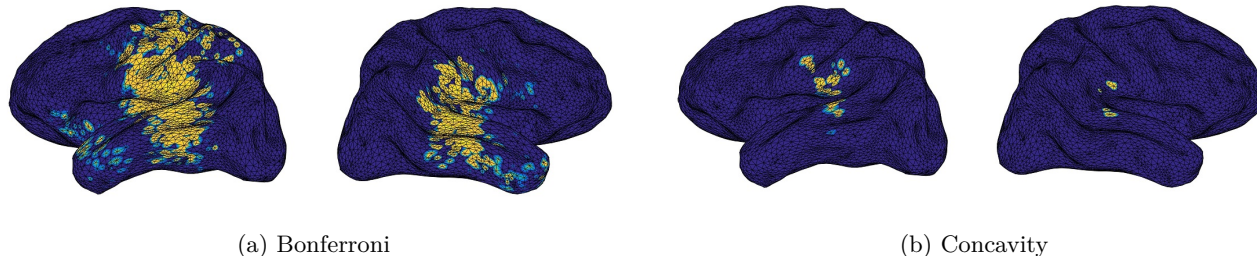


Figure 13: Detected source during all time of activity using MNE inverse method. a) with Bonferroni threshold. b) with Concavity threshold.

used ROC analysis, which proves that there can exist adequate thresholds, but does not provide a way for defining statistically the threshold (though the Otsu method (Otsu, 1979) was proposed for visualisation purposes). The objective of this article was thus to study the potential performance of different thresholding strategies, within a simulated framework where there is an actual ‘gold standard’.

A first strategy is to use statistical thresholds, where the sources that significantly differ from an estimate of the noise level are kept. We have shown here that these strategies do not give optimal results in high values of signal to noise ratio. Indeed, the source estimations that we used give a blurred view of the actual sources. This phenomenon of source ‘leakage’ is currently attracting much interest when it comes to estimating connectivity patterns, as it results in spurious high values of correlation among sources (Brookes et al., 2012). In fact, it can be assumed that statistical thresholding in the case of leaking algorithms such as L2 minimum norm will give satisfying results only in ‘average’ cases of SNRs, where the noise level just compensates for the leakage effect. An alternative to parametric estimates is to use non-parametric methods. Thus, a permutation method was proposed in (Maris and Oostenveld, 2007), where  $H_0$  (no difference between two conditions) is estimated by shuffling the labels of the conditions and taking the maximum value at each iteration. Such methods allows controlling for the family-wise error rate with empirical estimation of the  $H_0$  distribution, but should in principle also be affected by the leakage effect - which produces values above the noise level when the signal of interest has high SNR.

A second strategy is to use a more ‘adaptive’ way to define the threshold, taking into account the full distribution of values (i.e., under both  $H_0$  and  $H_1$ ). This is in line with (Grova et al., 2006), who used the Otsu method (Otsu, 1979), a classical image segmentation algorithm based on the hypothesis of a mixture of two Gaussian distributions (the threshold is the value that best separates the two distributions). We tested the local FDR method (Efron, 2005), an empirical Bayes method that does assume a known distribution type for  $H_1$  (at the expense of assuming that the number of points under  $H_0$  is much larger than those under  $H_1$ ). We also introduced another empirical method, the ‘concavity threshold’, based on the shape of the survival function of the source values. They show a better performance, close to that of the optimal threshold.

As an alternative source reconstruction method, it could be useful to study the performance of patch-based source localization that explicitly models source extent (Schmidt et al., 1999; Birot et al., 2011; Limpiti et al., 2006; Hillebrand and Barnes, 2011), as well as sparse reconstruction methods for which thresholding is directly integrated into the algorithm (Becker et al., 2014; Gramfort et al., 2013; Sohrabpour et al., 2016; Zhu et al., 2014; Haufe et al., 2011; Chowdhury et al., 2013). Another solution could be to decrease the number of sources or to add spatial constraints (Philippe et al., 2013; Belaoucha et al., 2016)

Another very important issue in thresholding is the determination of the number of activated regions. More work is necessary for assessing the capacity of the different thresholding methods in reovering the correct number

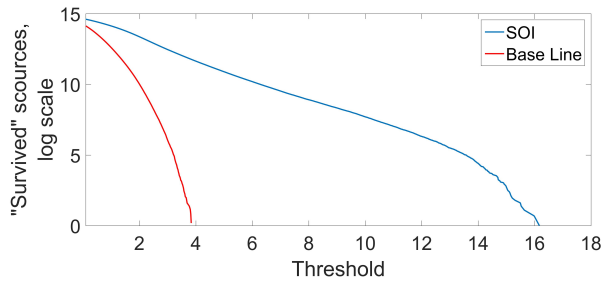


Figure 14: Survival function of real data in logarithmic scale after MNE source reconstruction.  $\lambda = 200$ .

of sources in a distributed source model, as was performed for the dipolar model (Supek and Aine, 1993). In summary, after analysis of different inverse operators and statistical thresholds, and - importantly - in the context of our limited simulations, we can conclude that:

- MNE and BF inversions have in average the same performance in source localization.
- Parametric thresholds based in noise properties, such as Bonferroni and FDR, are not efficient because of spatial leakage.
- Local FDR threshold have a good performance, but require parameters to tune and justify (model of  $H_0$  and extent of central part on which to estimate this model).
- Our proposed Concavity threshold has also shown a good performance, but it still lacks a theoretical explanation.
- Both Local FDR and Concavity show a result that are close to the optimal threshold.

Thus, statistical thresholding should be used with caution in the context of MNE and BF. Adaptive methods should be preferred, that are more robust to spatial leakage. Alternatively, methods with intrinsic thresholding based on sparsity constraints should be investigated.

## 5. Acknowledgements

CGB thanks C Grova for useful discussions. Part of this work was funded by a joint Agence Nationale de la Recherche (ANR) and Direction Générale de l'Offre de Santé (DGOS) under grant 'VIBRATIONS' ANR 13 PRTS 0011 01. This work has been carried out within the FHU EPINEXT with the support of the A\*MIDEX project (ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program managed by the French National Research Agency (ANR).

## References

- Abdi, H.H., 2007. The Bonferonni and Sidak Corrections for Multiple Comparisons. Encyclopedia of Measurement and Statistics. 1, 1–9. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.8747&rep=rep1&type=pdf>, doi:10.4135/9781412952644, arXiv:arXiv:1011.1669v3.
- Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. Signal Processing Magazine, IEEE 18, 14–30.
- Becker, H., Albera, L., Comon, P., Gribonval, R., Merlet, I., 2014. Fast, variation-based methods for the analysis of extended brain sources. European Signal Processing Conference , 41–45.
- Belaoucha, B., Kachouane, M., Papadopoulo, T., 2016. Multivariate autoregressive model constrained by anatomical connectivity to reconstruct focal sources, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), p. 4067–4070. doi:10.1109/EMBC.2016.7591620.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. doi:10.2307/2346101, arXiv:95/57289.

- Biot, G., Albera, L., Wendling, F., Merlet, I., 2011. Localization of extended brain sources from eeg/meg: the exo-music approach. *Neuroimage* 56, 102–13. URL: <https://www.ncbi.nlm.nih.gov/pubmed/21276860>, doi:10.1016/j.neuroimage.2011.01.054.
- Brookes, M.J., Woolrich, M.W., Barnes, G.R., 2012. Measuring functional connectivity in meg: a multivariate approach insensitive to linear source leakage. *Neuroimage* 63, 910–20. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22484306>, doi:10.1016/j.neuroimage.2012.03.048.
- Chowdhury, R.A., Lina, J.M., Kobayashi, E., Grova, C., 2013. Meg source localization of spatially extended generators of epileptic activity: comparing entropic and hierarchical bayesian approaches. *PLoS One* 8, e55969. URL: <https://www.ncbi.nlm.nih.gov/pubmed/23418485>, doi:10.1371/journal.pone.0055969.
- Efron, B., 2005. Local false discovery rates. *Discovery* , 63–79 URL: <http://statistics.stanford.edu/~ckirby/techreports/BIO/BIO234.pdf>, doi:10.1198/016214507000000941.
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1053811901910377>, doi:10.1006/nimg.2001.1037, doi:10.1006/nimg.2001.1037, doi:10.1006/nimg.2001.1037.
- Goeman, J.J., Solari, A., 2014. Multiple hypothesis testing in genomics. *Statistics in Medicine* 33, 1946–1978. doi:10.1002/sim.6082, arXiv:arXiv:1306.1646v1.
- Gramfort, A., Papadopoulos, T., Olivi, E., Clerc, M., 2010. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical engineering online* 9, 45. doi:10.1186/1475-925X-9-45.
- Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M.S., Kowalski, M., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage* 70, 410–422. URL: <http://dx.doi.org/10.1016/j.neuroimage.2012.12.051>, doi:10.1016/j.neuroimage.2012.12.051.
- Grave De Peralta Menendez, R., Hauk, O., Andino, S.G., Vogt, H., Michel, C., 1997. Linear inverse solutions with optimal resolution kernels applied to electromagnetic tomography. *Human Brain Mapping* 5, 454–467. doi:10.1002/(SICI)1097-0193(1997)5:6<454::AID-HBM6>3.0.CO;2-2.
- Grova, C., Daunizeau, J., Lina, J.M., Bénar, C.G., Benali, H., Gotman, J., 2006. Evaluation of EEG localization methods using realistic simulations of interictal spikes. *NeuroImage* 29, 734–753. doi:10.1016/j.neuroimage.2005.08.053.
- Hämäläinen, M., 2010. MNE Software User’s guide. MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging 16. URL: <http://nmr.mgh.harvard.edu/meg/manuals/MNE-manual-2.5.pdf>.
- Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., Müller, K.R., 2011. Large-scale EEG/MEG source localization with spatial flexibility. *NeuroImage* 54, 851–859.
- Hillebrand, A., Barnes, G.R., 2011. Practical constraints on estimation of source extent with meg beamformers. *Neuroimage* 54, 2732–40. URL: <https://www.ncbi.nlm.nih.gov/pubmed/20969964>, doi:10.1016/j.neuroimage.2010.10.036.
- Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., Papadopoulos, T., 2005. A common formalism for the integral formulations of the forward EEG problem. *IEEE Transactions on Medical Imaging* 24, 12–28. doi:10.1109/TMI.2004.837363.
- Limpiti, T., Veen, B.D.V., Wakai, R.T., 2006. Cortical patch basis model for spatially extended neural activity. *IEEE Transactions on Biomedical Engineering* 53, 1740–1754. doi:10.1109/TBME.2006.873743.
- Liu, A.K., Dale, A.M., Belliveau, J.W., 2002. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping* 16, 47–62. doi:10.1002/hbm.10024.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of eeg- and meg-data. *J Neurosci Meth* 164, 177–90. URL: <http://www.ncbi.nlm.nih.gov/pubmed/17517438>, doi:10.1016/j.jneumeth.2007.03.024.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave De Peralta, R., 2004. EEG source imaging. *Clinical Neurophysiology* 115, 2195–2222. doi:10.1016/j.clinph.2004.06.001.

- Nichols, T., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research* 12, 419–446. doi:[10.1191/0962280203sm341ra](https://doi.org/10.1191/0962280203sm341ra).
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* 2011. doi:[10.1155/2011/156869](https://doi.org/10.1155/2011/156869), [arXiv:156869](https://arxiv.org/abs/156869).
- Otsu, N., 1979. A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics* 9, 62–66. Minimize inter class variance.
- Pantazis, D., Nichols, T.E., Baillet, S., Leahy, R.M., 2005. A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *NeuroImage* 25, 383–394. doi:[10.1016/j.neuroimage.2004.09.040](https://doi.org/10.1016/j.neuroimage.2004.09.040).
- Philippe, A.C., Clerc, M., Papadopoulo, T., Deriche, R., 2013. Cortex parcellation via diffusion data as prior knowledge for the MEG inverse problem, in: *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, San Francisco, United States. pp. 994–997. URL: <https://hal.inria.fr/hal-00858019>, doi:[10.1109/ISBI.2013.6556644](https://doi.org/10.1109/ISBI.2013.6556644).
- Powers, D., 2011. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2, 37 – 63.
- Rosenow, F., Lüders, H., 2001. Presurgical evaluation of epilepsy. *Brain* 124, 1683–700. URL: <http://www.ncbi.nlm.nih.gov/pubmed/11522572>.
- Schmidt, D.M., George, J., Wood, C., 1999. A threshold selection method from gray level histograms. *Hum. Brain Mapp.* 7, 195–212.
- Sohrabbour, A., Lu, Y., Worrell, G., He, B., 2016. Imaging brain source extent from EEG/MEG by means of an iteratively reweighted edge sparsity minimization (IRES) strategy. *NeuroImage* URL: <http://dx.doi.org/10.1016/j.neuroimage.2016.05.064>.
- Supek, S., Aine, C.J., 1993. Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution. *IEEE Transactions on Biomedical Engineering* 40, 529–540. doi:[10.1109/10.237672](https://doi.org/10.1109/10.237672).
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience* doi:[10.1155/2011/879716](https://doi.org/10.1155/2011/879716), [arXiv:879716](https://arxiv.org/abs/879716).
- Trebuchon-Da Fonseca, A., Giraud, K., Badier, J.M., Chauvel, P., Liegeois-Chauvel, C., 2005. Hemispheric lateralization of voice onset time (vot) comparison between depth and scalp eeg recordings. *Neuroimage* 27, 1–14. doi:[10.1016/j.neuroimage.2004.12.064](https://doi.org/10.1016/j.neuroimage.2004.12.064).
- Valdés-Sosa, P.A., Sanchez-Bornot, J.M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., Canales-Rodríguez, E., Sa, J.M., Valde, P.A., Bosch-Bayard, J., Melie-garci, L., Canales-rodri, E., 2005. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 360, 969–81. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1854937&tool=pmcentrez&rendertype=abstract> $\backslash$ ~~delimiter"026E30F\$nh~~<http://www.ncbi.nlm.nih.gov/pubmed/16087441>, doi:[10.1098/rstb.2005.1654](https://doi.org/10.1098/rstb.2005.1654).
- Van Drongelen, W., Yuchtman, M., Van Veen, B.D., Van Huffelen, A.C., 1996. A spatial filtering technique to detect and localize multiple sources in the brain. *Brain Topography* 9, 39–49. doi:[10.1007/BF01191641](https://doi.org/10.1007/BF01191641).
- Van Veen, B.D., Buckley, K.M., 1988. Beamforming: a versatile approach to spatial filtering. URL: <http://dx.doi.org/10.1109/53.665> $\backslash$ ~~delimiter"026E30F\$npapers2://publication/doi/10.1109/53.665~~, doi:[10.1109/53.665](https://doi.org/10.1109/53.665).
- Wens, V., 2015. Investigating complex networks with inverse models: Analytical aspects of spatial leakage and connectivity estimation. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 91. doi:[10.1103/PhysRevE.91.012823](https://doi.org/10.1103/PhysRevE.91.012823), [arXiv:1401.0444](https://arxiv.org/abs/1401.0444).

- Wens, V., Marty, B., Mary, A., Bourguignon, M., Op de Beeck, M., Goldman, S., Van Bogaert, P., Peigneux, P., De Tiège, X., 2015. A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. *Human Brain Mapping* 36, 4604–4621. doi:[10.1002/hbm.22943](https://doi.org/10.1002/hbm.22943).
- Yao, J., Dewald, J.P.A., 2005. Evaluation of different cortical source localization methods using simulated and experimental EEG data. *NeuroImage* 25, 369–382. doi:[10.1016/j.neuroimage.2004.11.036](https://doi.org/10.1016/j.neuroimage.2004.11.036).
- Zhu, M., Zhang, W., Dickens, D.L., Ding, L., 2014. Reconstructing spatially extended brain sources via enforcing multiple transform sparseness. *NeuroImage* 86, 280–293.