



**HAL**  
open science

## Bayesian EEG source localization using a structured sparsity prior

Facundo Hernan Costa, Hadj Batatia, Thomas Oberlin, Carlos d’Giano,  
Jean-Yves Tournernet

► **To cite this version:**

Facundo Hernan Costa, Hadj Batatia, Thomas Oberlin, Carlos d’Giano, Jean-Yves Tournernet. Bayesian EEG source localization using a structured sparsity prior. *NeuroImage*, 2017, Vol. 144 (Part. A), pp. 142-152. 10.1016/j.neuroimage.2016.08.064 . hal-01502837

**HAL Id: hal-01502837**

**<https://hal.science/hal-01502837>**

Submitted on 6 Apr 2017

**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.



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <http://oatao.univ-toulouse.fr/>  
Eprints ID : 17566

**To link to this article** : DOI:10.1016/j.neuroimage.2016.08.064

**To cite this version :**

Costa, Facundo and Batatia, Hadj and Oberlin, Thomas and D'Giano, Carlos and Tourneret, Jean-Yves *Bayesian EEG source localization using a structured sparsity prior*. (2017) NeuroImage, Vol. 144 (Part. A). pp. 142-152. ISSN 1053-8119

Any correspondence concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)

# Bayesian EEG source localization using a structured sparsity prior

Facundo Costa <sup>\*,a</sup>, Hadj Batatia <sup>a</sup>, Thomas Oberlin <sup>a</sup>, Carlos D'Giano <sup>b</sup>, Jean-Yves Tournier <sup>a</sup>

<sup>a</sup> University of Toulouse, INP/ENSEEIH - IRIT, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France

<sup>b</sup> FLENI Hospital, Buenos Aires, Argentina

## ABSTRACT

This paper deals with EEG source localization. The aim is to perform spatially coherent focal localization and recover temporal EEG waveforms, which can be useful in certain clinical applications. A new hierarchical Bayesian model is proposed with a multivariate Bernoulli Laplacian structured sparsity prior for brain activity. This distribution approximates a mixed  $\ell_{20}$  pseudo norm regularization in a Bayesian framework. A partially collapsed Gibbs sampler is proposed to draw samples asymptotically distributed according to the posterior of the proposed Bayesian model. The generated samples are used to estimate the brain activity and the model hyperparameters jointly in an unsupervised framework. Two different kinds of Metropolis–Hastings moves are introduced to accelerate the convergence of the Gibbs sampler. The first move is based on multiple dipole shifts within each MCMC chain, whereas the second exploits proposals associated with different MCMC chains. Experiments with focal synthetic data shows that the proposed algorithm is more robust and has a higher recovery rate than the weighted  $\ell_{21}$  mixed norm regularization. Using real data, the proposed algorithm finds sources that are spatially coherent with state of the art methods, namely a multiple sparse prior approach and the Champagne algorithm. In addition, the method estimates waveforms showing peaks at meaningful timestamps. This information can be valuable for activity spread characterization.

## Keywords:

EEG  
MCMC  
Inverse problem  
Source localization  
Structured-sparsity  
Hierarchical Bayesian model  
 $\ell_{20}$  norm regularization  
Medical imaging

## 1. Introduction

EEG source localization problem has attracted considerable attention in the literature resulting in a wide range of methods developed in the last years. These can be classified into two groups: (i) the dipole-fitting models that represent the brain activity as a small number of dipoles with unknown positions; and (ii) the distributed-source models that represent the brain activity as a large number of dipoles in fixed positions. Dipole-fitting models (Sommariva and Sorrentino, 2014; da Silva and Van Rotterdam, 1998) try to estimate the amplitudes, orientations and positions of a few dipoles that explain the measured data. Unfortunately, the corresponding estimators are very sensitive to the initial guess of the number of dipoles and their initial locations (Grech et al., 2008). On the other hand, the distributed-source methods model the brain activity using a large number of dipoles with fixed positions and try to estimate their amplitudes (Grech et al., 2008) by solving an ill-posed inverse problem. One of the most simple ways to solve this inverse problem is to use an  $\ell_2$  norm regularization as the minimum norm estimator (Pascual-

Marqui, 1999) or its variants Loreta (Pascual-Marqui et al., 1994) and sLoreta (Pascual-Marqui et al., 2002). However, these methods usually overestimate the active area size (Grech et al., 2008).

Sparsity constraints can remedy the overestimation issue when dealing with applications with discretely localized activity such as certain kinds of epilepsy (Berg et al., 2010). In distributed activity applications, promoting sparsity should provide spatially coherent localization even though it is unable to estimate the activity extension. To apply sparsity, ideally an  $\ell_0$  pseudo norm regularization (Candes, 2008) should be used. Unfortunately, this procedure is intractable in an optimization framework. As a consequence, the  $\ell_0$  pseudo norm is usually approximated by the  $\ell_1$  norm via convex relaxation (Uutela et al., 1999), even if the two regularizations do not always provide the same solution (Candes, 2008). In a previously reported work, we proposed to combine them in a Bayesian framework (Costa et al., 2015), using the  $\ell_0$  pseudo norm to locate the non-zero positions and the  $\ell_1$  norm to estimate their amplitudes. However the methods studied in Candes (2008), Uutela et al. (1999), and Costa et al. (2015) consider each time sample independently leading in some cases to unrealistic solutions (Gramfort et al., 2012).

To improve source localization, it is possible to make use of the temporal structure of the data. This can be done by considering sparse Bayesian learning using multiple measurement vectors (Zhang and Rao, 2011) or by using the STOUT (Castaño-Candamil et al., 2015) and dMAP-EM (Lamus et al., 2012) methods that apply

\* Corresponding author at: INP - ENSEEIHT Toulouse 2, rue Charles Camichel, B.P. 7122, 31071 Toulouse Cedex 7, France. Tel.: +33 05 34 32 22 52.

E-mail address: [facundo.costa@enseeiht.fr](mailto:facundo.costa@enseeiht.fr) (F. Costa).

physiological considerations to the source representation. It is also possible to model the time evolution of the dipole activity and estimate it using Kalman filtering (Galka et al., 2004; Long et al., 2011), particle filters (Somersalo et al., 2003; Sorrentino et al., 2013; Chen and Godsill, 2013) or by encouraging spatio-temporal structures by promoting structured sparsity (Huang and Zhang, 2010).

Structured sparsity has been shown to improve results in several applications including audio restoration (Kowalski et al., 2013), image analysis (Yu et al., 2012) and machine learning (Huang et al., 2011). Structured sparsity has also been applied to M/EEG source localization by Gramfort et al. by using the  $\ell_{21}$  mixed norm (Gramfort et al., 2012). This approach promotes sparsity among different dipoles (via the  $\ell_1$  portion of the norm) and groups all the time samples of the same dipole together, forcing them to be either jointly active or inactive (with the  $\ell_2$  norm portion). This work was reconsidered by the same authors yielding the iterative reweighted mixed norm estimator (Strohmeier et al., 2014) and the time–frequency mixed-norm estimator (Gramfort et al., 2013). However, all these methods require the manual tuning of the regularization parameters.

Several Bayesian methods have also been used to solve the inverse problem (Friston et al., 2008; Stahlhut et al., 2013; Wipf et al., 2010; Lucka et al., 2012). Friston et al. (2008) developed the multiple sparse priors (MSP) approach, in which they segment the brain into different pre-defined regions and promote all the dipoles in each region to be active or inactive jointly. In contrast, Wipf et al. developed the Champagne algorithm to promote activity to be concentrated on a sparse set of dipoles (Wipf et al., 2010). Lucka et al. (2012) studied a hierarchical Bayesian model (HBM) offering significant improvements over established methods such as MNE and sLoreta.

Similar to Wipf et al., this paper develops a new method encouraging sparse activity considering each dipole separately (Friston et al., 2008). The proposed method uses a multivariate Bernoulli Laplace prior (approximating the weighted  $\ell_{20}$  mixed norm) for the dipole amplitudes without assuming any additional prior information such as the amount or position of the active dipoles. Since the parameters of the proposed model cannot be computed with closed-form expressions, we investigate a Markov chain Monte Carlo sampling technique to draw samples that are asymptotically distributed according to the posterior of the proposed model. Then the brain activity, the model parameters and hyperparameters are jointly estimated in an unsupervised framework. In order to avoid the sampler to becoming stuck around local maxima, specific Metropolis–Hastings moves are introduced. These moves significantly accelerate the convergence speed of the proposed sampler. From the medical point of view, the proposed approach aims at providing the localization of the main sources of the brain activity to help making decisions when selecting candidate patients for recessive surgery, in the case of discretely localized epilepsy (Berg et al., 2010). In addition, considering several time samples simultaneously allows us to estimate the temporal waveforms of the activity. Estimating these waveforms can be useful in some clinical applications, such as the estimation of the spread patterns of the activity in epilepsy (Quintero-Rincón et al., 2016).

The paper is organized as follows: Section 2 presents the proposed Bayesian model. Section 3 introduces the partially collapsed Gibbs sampler used to generate samples distributed according to the posterior of this model and the Metropolis–Hastings moves that are used to accelerate the convergence of the sampler. Experimental results conducted for both synthetic and real data are presented in Section 4. Conclusions are finally reported in Section 5.

## 2. Proposed method

EEG source localization is an inverse problem consisting in estimating the brain activity of a patient from EEG measurements

taken from  $M$  electrodes during  $T$  time samples. In a distributed source model, the brain activity is represented by a finite number of dipoles located at fixed positions on the brain cortex. More precisely, we consider  $N$  dipoles located on the cortical surface and oriented orthogonally to it (see Hallez et al., 2007 for motivation). The EEG measurement matrix  $\mathbf{Y} \in \mathbb{R}^{M \times T}$  can be written as

$$\mathbf{Y} = \mathbf{H} \mathbf{X} + \mathbf{E} \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{N \times T}$  contains the dipole amplitudes,  $\mathbf{H} \in \mathbb{R}^{M \times N}$  is the lead-field matrix and  $\mathbf{E}$  is the additive noise.

### 2.1. Likelihood

It is very classical to assume that the noise samples are independent and identically distributed according to a Gaussian distribution (Grech et al., 2008). Note that when this assumption does not hold it is possible to estimate the noise covariance matrix from measurements that do not contain the signal of interest and use it to whiten the data (Maris, 2003). Denoting as  $\sigma_n^2$  the noise variance, the independence assumption leads to the likelihood

$$f(\mathbf{Y}|\theta) = \prod_{t=1}^T \mathcal{N}(\mathbf{y}^t | \mathbf{H}\mathbf{x}^t, \sigma_n^2 \mathbb{I}_M) \quad (2)$$

where  $\mathbb{I}_M$  is the identity matrix of size  $M$  and  $\theta = \{\mathbf{X}, \sigma_n^2\}$  contains the unknown parameters.

### 2.2. Prior distributions

#### 2.2.1. Brain activity $\mathbf{X}$

To promote structured sparsity of the source activity, we consider the weighted  $\ell_{20}$  mixed pseudo-norm

$$\|\mathbf{X}\|_{20} = \#\{i | \sqrt{v_i} \|\mathbf{x}_i\|_2 \neq 0\} \quad (3)$$

where  $v_i = \|\mathbf{h}^i\|_2$  is a weight introduced to compensate the depth-weighting effect (Grech et al., 2008; Uutela et al., 1999) and  $\#\mathcal{S}$  denotes the cardinal of the set  $\mathcal{S}$ . Since this prior leads to intractable computations, we propose to approximate it by a multivariate Laplace Bernoulli prior for each row of  $\mathbf{X}$ <sup>1</sup>

$$f(\mathbf{x}_i | z_i, \lambda) \propto \begin{cases} \delta(\mathbf{x}_i) & \text{if } z_i = 0 \\ \exp\left(-\frac{1}{\lambda} \sqrt{v_i} \|\mathbf{x}_i\|_2\right) & \text{if } z_i = 1 \end{cases} \quad (4)$$

where  $\propto$  means “proportional to”,  $\lambda$  is the parameter of the exponential distribution and  $\mathbf{z} \in \{0, 1\}^N$  is a vector indicating if the rows of  $\mathbf{X}$  are non-zero. To make the analysis easier we introduce the hyperparameter  $a = \frac{\sigma_n^2}{\lambda^2}$  leading to

$$f(\mathbf{x}_i | z_i, a, \sigma_n^2) \propto \begin{cases} \delta(\mathbf{x}_i) & \text{if } z_i = 0 \\ \exp\left(-\sqrt{\frac{v_i a}{\sigma_n^2}} \|\mathbf{x}_i\|_2\right) & \text{if } z_i = 1. \end{cases} \quad (5)$$

The elements  $z_i$  are then assigned a Bernoulli prior with parameter  $\omega \in [0, 1]$

$$f(z_i | \omega) = \mathcal{B}(z_i | \omega). \quad (6)$$

Note that the Dirac delta function  $\delta(\cdot)$  in the prior of  $\mathbf{x}_i$  promotes sparsity while the Laplace distribution regulates the amplitudes of the non-zero rows. The parameter  $\omega$  allows the importance of these two terms to be balanced. In particular,  $\omega = 0$  yields  $\mathbf{X} = 0$  whereas  $\omega = 1$  leads to the Bayesian formulation of the group-lasso (Yuan

<sup>1</sup> In this paper, we will denote as  $\mathbf{m}_i$  the  $i$ -th row of the matrix  $\mathbf{M}$  and as  $\mathbf{m}^j$  its  $j$ -th column.

and Lin, 2006). Unfortunately the prior (5) still leads to an intractable posterior. It is possible to fix this problem by introducing a latent variable vector  $\tau^2 \in (\mathbb{R}^+)^N$  as suggested in Raman et al. (2009). More precisely, we use the following gamma and Bernoulli–Gaussian priors for  $\tau_i^2$  and  $\mathbf{x}_i$

$$f(\tau_i^2|a) = \mathcal{G}\left(\tau_i^2 \left| \frac{T+1}{2}, \frac{v_i a}{2} \right.\right) \quad (7)$$

$$f(\mathbf{x}_i|z_i, \tau_i^2, \sigma_n^2) = \begin{cases} \delta(\mathbf{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}(\mathbf{x}_i|0, \sigma_n^2 \tau_i^2 \mathbb{1}_T) & \text{if } z_i = 1 \end{cases} \quad (8)$$

which yield the marginal distribution of  $\mathbf{x}_i$  defined in (5) (Raman et al., 2009).

### 2.3. Hyperparameter priors

The proposed method allows one to balance the importance between sparsity of the solution and fidelity to the measurements using two hyperparameters: (1)  $\omega$  that adjusts the proportion of non-zero rows and (2)  $a$  that controls the amplitudes of the non-zeros. The hyperparameter vector will be denoted as  $\phi = \{\omega, a\}$ . The corresponding hierarchy between the model parameters and hyperparameters is illustrated in Fig. 1. In contrast to the  $\ell_{21}$  mixed norm the proposed algorithm is able to estimate the model hyperparameters from the data by assigning hyperpriors to them following a so-called hierarchical Bayesian analysis. These hyperpriors, along with the prior of the noise variance  $\sigma_n^2$ , were chosen to be as non-informative as possible and can be found in Appendix A.

### 2.4. Posterior distribution

Using the previously described priors and hyperpriors, the posterior distribution of the proposed Bayesian model is

$$f(\theta, \mathbf{z}, \tau^2, \phi|\mathbf{Y}) \propto f(\mathbf{Y}|\theta)f(\theta|\mathbf{z}, \tau^2)f(\mathbf{z}, \tau^2|\phi)f(\phi) \quad (9)$$

where  $f(\mathbf{Y}|\theta)$  has been defined in (2) and

$$f(\theta|\mathbf{z}, \tau^2) \propto f(\sigma_n^2) \prod_{i=1}^N f(\mathbf{x}_i|z_i, \tau_i^2, \sigma_n^2) \quad (10)$$

$$f(\mathbf{z}, \tau^2|\phi) = \prod_{i=1}^N f(z_i|\omega)f(\tau_i^2|a) \quad (11)$$

$$f(\phi) = f(a|\alpha, \beta)f(\omega). \quad (12)$$

The posterior distribution (9) is intractable and does not allow us to derive closed-form expressions for the Bayesian estimators of the different parameters and hyperparameters. Thus we propose to draw samples from (9) and use them to estimate the brain activity jointly with the model hyperparameters. The following section provides more details about the sampling method investigated in this paper.

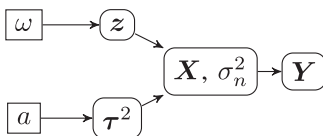


Fig. 1. Directed acyclic graph for the Bayesian model illustrating the dependencies between the model parameters and hyperparameters.

## 3. A partially collapsed Gibbs sampler

We investigate a partially collapsed Gibbs sampler that samples the variables  $z_i$  and  $\mathbf{x}_i$  jointly. If  $\mathbf{X}_{-i}$  denotes the matrix  $\mathbf{X}$  whose  $i$ th row has been replaced by zeros, the resulting sampling strategy is summarized in Algorithm 1. The corresponding conditional distributions are described in Appendix B.

**Algorithm 1.** Partially Collapsed Gibbs sampler.

```

Initialize  $\mathbf{X} = 0$  and  $\mathbf{z} = 0$ 
Sample  $a$  and  $\tau^2$  from their prior distributions
repeat
  Sample  $\sigma_n^2$  from  $f(\sigma_n^2|\mathbf{Y}, \mathbf{X}, \tau^2, \mathbf{z})$ 
  Sample  $\omega$  from  $f(\omega|\mathbf{z})$ 
  for  $i=1$  to  $N$  do
    Sample  $\tau_i^2$  from  $f(\tau_i^2|\mathbf{x}_i, \sigma_n^2, a, z_i)$ 
    Sample  $z_i$  from  $f(z_i|\mathbf{Y}, \mathbf{X}_{-i}, \sigma_n^2, \tau_i^2, \omega)$ 
    Sample  $\mathbf{x}_i$  from  $f(\mathbf{x}_i|z_i, \mathbf{Y}, \mathbf{X}_{-i}, \sigma_n^2, \tau_i^2)$ 
  end for
  Sample  $a$  from  $f(a|\tau^2)$ 
until convergence
  
```

### 3.1. Multiple dipole shift proposals

The partially collapsed Gibbs sampler summarized in Algorithm 1 may get stuck around local maxima of the variable  $\mathbf{z}$  from which it can be difficult to escape in a reasonable amount of iterations (examples illustrating this situation are shown in Costa et al., 2015). In order to bypass this problem, we introduce specific Metropolis–Hastings moves. These moves consist of proposing a new value of  $\mathbf{z}$  (referred to as “candidate”) after each sampling iteration. The candidate is then accepted or rejected with an appropriate acceptance rate according to the Metropolis–Hastings rule, which guarantees that the target distribution is preserved.

Before presenting the proposal scheme, it is interesting to mention that it was inspired by an idea developed in Bourguignon and Carfantan (2005). The authors of Bourguignon and Carfantan (2005) proposed to move a random non-zero element of a binary sequence to a random neighboring position after each iteration of the MCMC sampler. We have generalized their scheme by proposing to move a random subset of  $K$  estimated non-zeros simultaneously to random neighboring positions. According to experimental results (some of them described in Section 4), the simple choice  $K=2$  provides good results in most practical cases. Since there is a high correlation between the variables  $\tau^2$  and  $\mathbf{z}$ , it is convenient to update their values jointly. The resulting proposal is shown in Algorithm 2 where

$$f(\mathbf{z}_r, \tau_r^2|\mathbf{Y}, a, \sigma_n^2, \omega) \propto (1-\omega)^{C_0} \omega^{C_1} (\sigma_n^2)^{-\frac{T C_1}{2}} |\Sigma|^{-\frac{T}{2}} \quad (13)$$

$$\prod_{i \in \mathcal{I}_1} (\tau_i^2)^{-\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^T K_t}{2}\right) \prod_{i=1}^N \mathcal{G}\left(\tau_i^2 \left| \frac{T+1}{2}, \frac{v_i a}{2} \right.\right). \quad (14)$$

**Algorithm 2.** Multiple dipole shift proposal.

```

 $\bar{\mathbf{z}} = \mathbf{z}$ 
repeat  $K$  times
  Set  $\text{ind}_{\text{old}}$  to be the index of a random non-zero of  $\mathbf{z}$ 
  Set  $\mathbf{p} = [\text{ind}_{\text{old}}, \text{neigh}_\gamma(\text{ind}_{\text{old}})]$ 
  
```

Set  $\text{ind}_{\text{new}}$  to be a random element of  $\mathbf{p}$   
Set  $\bar{z}_{\text{ind}_{\text{old}}} = 0$  and  $\bar{z}_{\text{ind}_{\text{new}}} = 1$   
**end**  
Sample  $\bar{\mathbf{X}}$  from  $f(\bar{\mathbf{X}}|\bar{\mathbf{z}}, \mathbf{Y}, \sigma_n^2, \tau^2)$ .  
Sample  $\bar{\tau}^2$  from  $f(\bar{\tau}^2|\bar{\mathbf{X}}, \sigma_n^2, a, \bar{\mathbf{z}})$ .  
Set  $\{\mathbf{z}, \tau^2\} = \{\bar{\mathbf{z}}, \bar{\tau}^2\}$  with probability  $\min\left(\frac{f(\bar{\mathbf{z}}, \bar{\tau}^2|\cdot)}{f(\mathbf{z}, \tau^2|\cdot)}, 1\right)$   
Resample  $\mathbf{X}$  if the proposal was accepted

where  $\mathbf{r} = \{i: z_i \neq \bar{z}_i\}$ ,  $\mathbf{I}_k = \{i: z_i = k\}$  and  $C_k = \#\mathbf{I}_k$  for  $k = \{0, 1\}$  and

$$\Sigma^{-1} = \frac{1}{\sigma_n^2} \left[ (\mathbf{H}^h)^T \mathbf{H}^h + \text{diag}\left(\frac{1}{\tau^2}\right) \right] \quad (15)$$

$$\mu_t = - \frac{\Sigma(\mathbf{H}^h)^T (\mathbf{H}^h \mathbf{x}_{-r}^t - \mathbf{y}^t)}{\sigma_n^2} \quad (16)$$

$$K_t = \frac{(\mathbf{H}^h \mathbf{x}_{-r}^t - \mathbf{y}^t)^T (\mathbf{H}^h \mathbf{x}_{-r}^t - \mathbf{y}^t)}{\sigma_n^2} - \mu_t^T \Sigma^{-1} \mu_t. \quad (17)$$

Note that  $\mathbf{m}_{-s}$  denotes the vector  $\mathbf{m}$  whose rows belonging to  $\mathbf{s}$  have been removed,  $\mathbf{M}_{-s}$  is the matrix  $\mathbf{M}$  whose columns belonging to  $\mathbf{s}$  have been removed,  $\text{diag}(\mathbf{s})$  is the diagonal square matrix whose diagonal elements are the elements of  $\mathbf{s}$  and  $|\mathbf{M}|$  is the determinant of the matrix  $\mathbf{M}$ .

**Algorithm 2** also uses the following neighborhood definition

$$\text{neigh}_\gamma(i) \triangleq \{j \neq i \mid |\text{corr}(\mathbf{h}^i, \mathbf{h}^j)| \geq \gamma\} \quad (18)$$

where  $\text{corr}(\mathbf{v}_1, \mathbf{v}_2)$  is the correlation between vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ . The neighborhood size can be adjusted by setting  $\gamma \in [0, 1]$  ( $\gamma = 0$  corresponds to a neighborhood containing all the dipoles and  $\gamma = 1$  corresponds to an empty neighborhood). To maximize the moves efficiency, the value of  $\gamma$  has to be selected carefully. Experiments during this study have shown that a good compromise is obtained with  $\gamma = 0.8$ . A comparison of the results obtained with and without multiple dipole shift proposals can be found in [Costa et al. \(2015\)](#).

### 3.2. Inter-chain proposal

Another possibility to improve the convergence speed of the proposed partially collapsed Gibbs sampler is to run multiple MCMC chains in parallel and exchange some information between them. Several methods have already been explored to perform this “exchange of information”, including Metropolis-coupled MCMC ([Geyer, 1991](#)), Population MCMC ([Laskey and Myers, 2003](#)) and simulated tempering ([Geyer and Thompson, 1995](#); [Marinari and Parisi, 1992](#)). In this paper, we introduce inter-chain moves by proposing to exchange the values of  $\mathbf{z}$  and  $\tau^2$  between different chains. This exchange is accepted with the probability shown in [Algorithm 3](#). Note that “a between-chain exchange” is made after each iteration with probability  $p$  (adjusted to  $\frac{1}{1000}$  by cross validation) according to [Algorithm 3](#). A comparison of the results obtained with and without these inter-chain proposals can be found in [Costa et al. \(2015\)](#).

**Algorithm 3.** Inter-chain proposals.

Define a vector  $\mathbf{c} = \{1, 2, \dots, L\}$  where  $L$  is the number of chains  
**for**  $i = \{1, 2, \dots, L\}$   
Choose (and remove) a random element from  $\mathbf{c}$  and denote it

by  $k$

Denote as  $\{\bar{\mathbf{z}}_k, \bar{\tau}_k^2\}$  the sampled values of  $\{\mathbf{z}, \tau^2\}$  of MCMC chain number  $\#k$

For the chain  $\#i$  set  $\{\mathbf{z}_i, \tau_i^2\} = \{\bar{\mathbf{z}}_k, \bar{\tau}_k^2\}$  with probability  $\frac{f(\bar{\mathbf{z}}_k, \bar{\tau}_k^2|\cdot)}{f(\mathbf{z}, \tau^2|\cdot)}$

Resample  $\mathbf{X}$  if the proposal has been accepted

**end**

### 3.3. Estimators

The point estimators used in this study are defined as

$$\hat{\mathbf{z}} \triangleq \arg \max_{\mathbf{z} \in \{0,1\}^N} (\#\mathcal{M}(\bar{\mathbf{z}})) \quad (19)$$

$$\hat{p} \triangleq \frac{1}{\#\mathcal{M}(\hat{\mathbf{z}})} \sum_{m \in \mathcal{M}(\hat{\mathbf{z}})} p^{(m)} \quad (20)$$

where  $\mathcal{M}(\bar{\mathbf{z}})$  is the set of iteration numbers  $m$  for which  $\mathbf{z}^{(m)} = \bar{\mathbf{z}}$  after the burn-in period and  $p^{(m)}$  is the  $m$ -th sample of  $p \in \{> \mathbf{X}, a, \sigma_n^2, \omega, \tau^2\}$ . Thus the estimator  $\hat{\mathbf{z}}$  in (19) is the maximum a posteriori estimator of  $\hat{\mathbf{z}}$  whereas the estimator used for all the other sampled variables in (20) is the minimum mean square error (MMSE) estimator.

It is interesting to note that the proposed method does not only provide point-estimators as the methods based on the  $\ell_{21}$  mixed norm. For instance, in some cases different values of  $\mathbf{z}$  can have a significant posterior probability. In this case the sampler may oscillate between different values of  $\mathbf{z}$  that usually differ by minor variations. In these cases, the proposed sampling method is able to identify several possible solutions (each of them corresponding to a different value of  $\mathbf{z}$ ) with their corresponding probabilities.

## 4. Experimental results

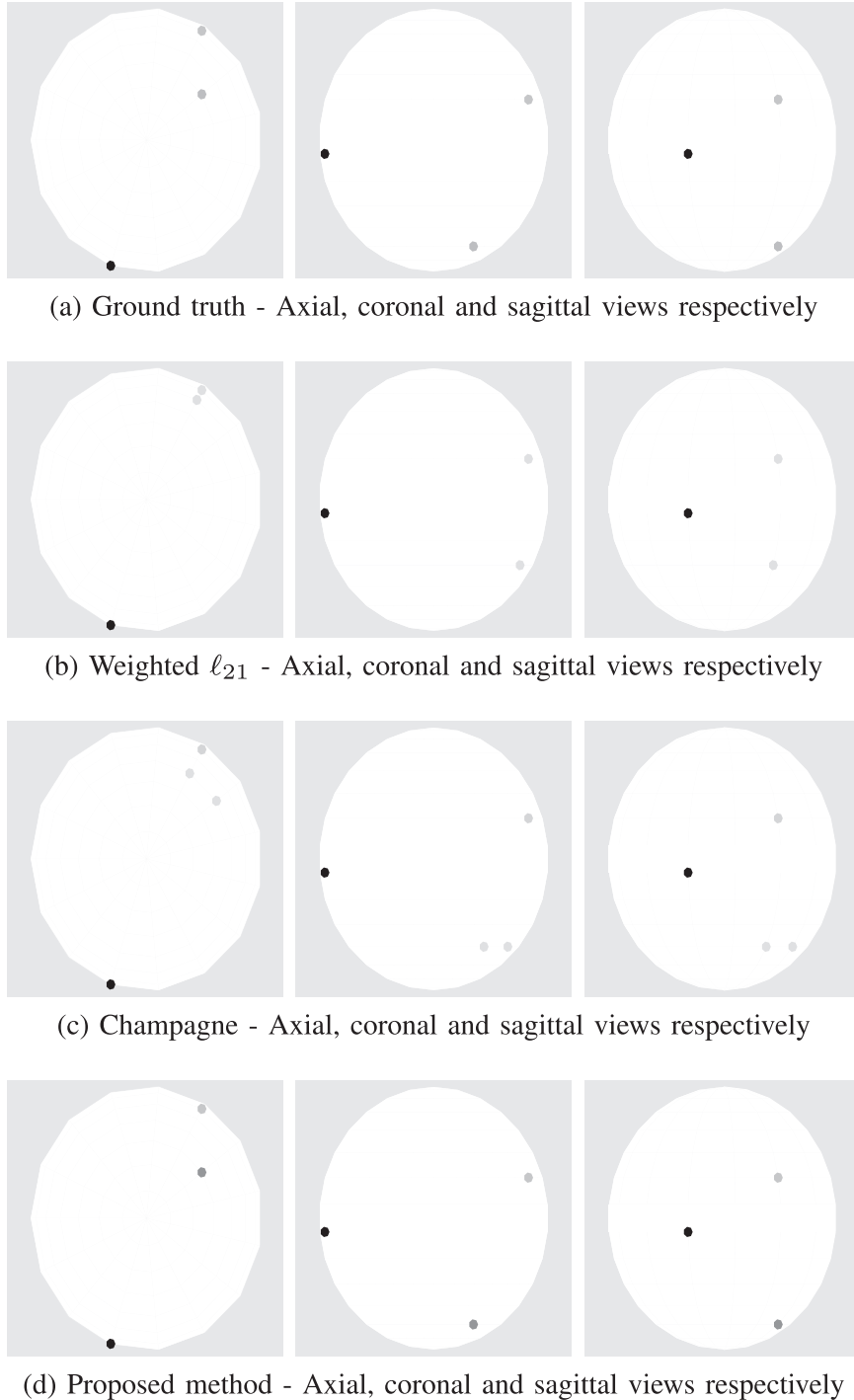
### 4.1. Synthetic data

Synthetic data are first considered to compare the  $\ell_{21}$  mixed norm approach, the Champagne model and the proposed method using a 212-dipole Stok three-sphere head model ([Stok, 1986](#)) with 41 electrodes. Two kinds of activations are considered: (1) three dipoles with low SNR and (2) multiple dipoles with high SNR. Note that additional experiments are available in the associated technical report ([Costa et al., 2015](#)).

#### 4.1.1. Three-dipoles with low SNR

Three dipoles were assigned damped sinusoidal excitations with frequencies varying between 5 and 20 Hz. These excitations were 500 ms long (a period corresponding to a stationary dipole activity) and sampled at 200 Hz. Different levels of noise were used to compare the performance of the different methods. The parameters of the proposed multiple dipole shift proposal were set to  $K=2$ ,  $\gamma = 0.8$  and  $C=8$  MCMC chains were run in parallel. The potential scale reduction factor (PSRF) ([Gelman and Rubin, 1992](#)) was used to assess the convergence of the proposed method. After running a fixed number of 10,000 iterations, the PSRFs of all the sampled variables were computed and we checked that these values were below 1.2 as recommended in ([Gelman et al., 1995](#), p. 332). For the  $\ell_{21}$  mixed norm approach, the value of the regularization parameter  $\lambda$  was chosen using cross-validation.

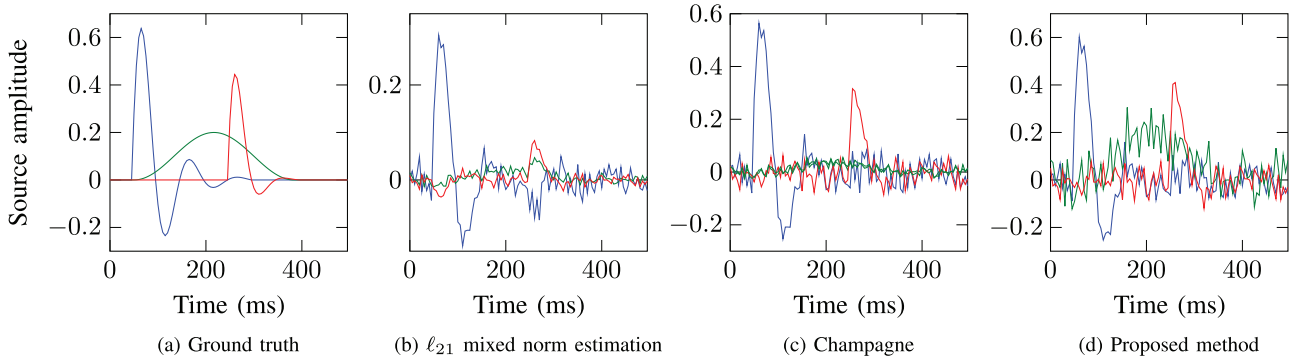
For high values of the input SNR ( $\geq 20$  dB), the results obtained with all methods are almost always identical to the ground truth. However, for lower values of SNR the proposed method outperforms



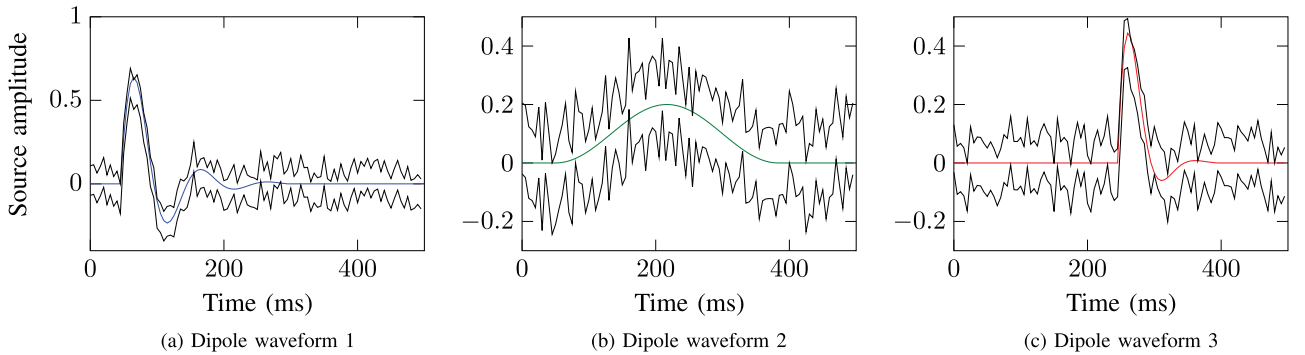
**Fig. 2.** Results for synthetic data with three active dipoles and low SNR: Comparison between the ground truth positions of the active dipoles with the positions estimated by different algorithms. The proposed method is the only one to find all dipoles in the correct places (a) Ground truth - Axial, coronal and sagittal views respectively, (b) Weighted  $\ell_{21}$  - Axial, coronal and sagittal views respectively, (c) Champagne - Axial, coronal and sagittal views respectively, (d) Proposed method - Axial, coronal and sagittal views respectively.

the other two. The estimated dipole locations associated with SNR = -3 dB are shown in Fig. 2 and the corresponding estimated waveforms in Fig. 3. Fig. 4 shows the ground truth of the three waveforms compared to the  $\mu \pm 2\sigma$  boundaries estimated by the proposed method. Note that only the dipoles with highest activity are displayed for the  $\ell_{21}$  approach. The approach based on the  $\ell_{21}$  norm manages to recover only two of the three non-zero activities at the correct positions and seems to underestimate considerably the amplitude of this activity. This is a known problem caused by approximating the  $\ell_0$  pseudo-norm by the  $\ell_1$  norm. In comparison, the

Champagne method spreads the activity of some of the active dipoles to its neighbors. The proposed algorithm oscillates between several values of  $z$  (specified in Table 1). However, the most probable value of  $z$  found by the algorithm is the correct one whereas the other most likely values of  $z$  have one of the non-zeros moved to a close neighbor. Finally, the histograms of the hyperparameters generated by the proposed Gibbs sampler are displayed in Fig. 5, showing a good agreement with the actual values of the parameters  $\omega$  and  $\sigma_n^2$  and allowing the parameter  $a$  associated with the latent variables  $\tau_i^2$  to be estimated.



**Fig. 3.** Results for synthetic data with three active dipoles and low SNR: Comparison between the ground truth and the waveforms estimated by different algorithms. The estimates obtained with the proposed method are closer to the ground truth (a) Ground truth, (b)  $\ell_{21}$  mixed norm estimation, (c) Champagne, (d) Proposed method.



**Fig. 4.** Results for synthetic data with three active dipoles and low SNR: Ground truth activation waveforms compared with  $\mu \pm 2\sigma$  boundaries estimated by the proposed method. The ground truth waveforms are within the estimated boundaries (a) Dipole waveform 1, (b) Dipole waveform 2, (c) Dipole waveform 3.

**Table 1**

Results for synthetic data with three active dipoles and low SNR: modes explored by the proposed algorithm. Positions 1, 2 and 3 correspond to the non-zero elements of the ground truth, showing that the ground truth mode is the most explored mode after convergence.

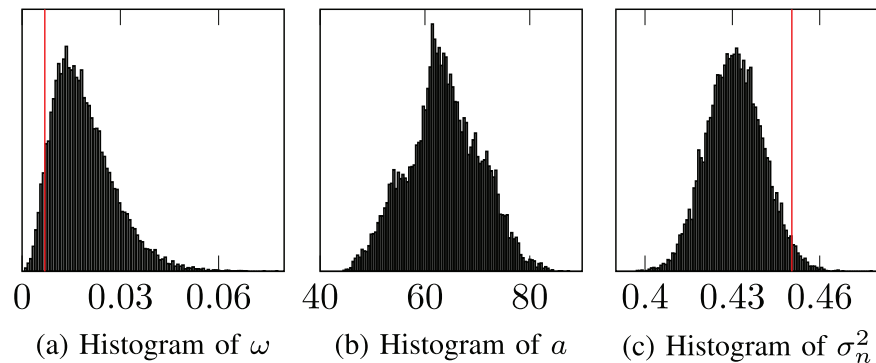
Active non-zeros	Percentage of samples
1 2 3	43
1 2 4	22
1 2 5	11
1 2 6	7
1 2 7	6
Others	11

To conclude, the proposed method improves the EEG source localization thanks to the use of a Laplace Bernoulli prior. Moreover, the use of an MCMC method makes it possible to recover different sets of source locations with their respective probabilities.

#### 4.1.2. Multiple dipoles

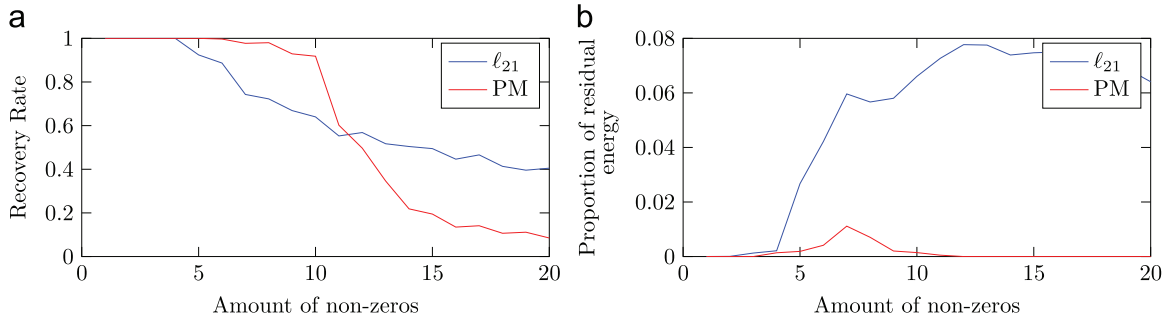
In each simulation of this section,  $P$  dipoles were activated with damped sinusoidal waves whose frequencies vary between 5 and 20 Hz. The activations were sampled at 200 Hz and scaled in amplitude so that each of them produced the same energy in the measurements. Fifty different sets of localizations were used for the active dipole positions for each value of  $P = 1, \dots, 20$ , resulting in a total of 1000 experiments. Noise was added to the measurements to obtain  $\text{SNR} = 30$  dB. For the  $\ell_{21}$  mixed norm regularization the regularization parameter was set according to the uncertainty principle which consists in finding a solution  $\hat{\mathbf{X}}$  such that  $\|\mathbf{H}\hat{\mathbf{X}} - \mathbf{Y}\| \approx \|\mathbf{H}\mathbf{X} - \mathbf{Y}\|$  (Morozov, 1966).

For each simulation, the estimated activity was defined as the  $P$  dipoles that had the highest value of  $\left\| \frac{x_i^2}{v_i^2} \right\|$ . The other dipoles were considered as residuals. We define the recovery rate as the proportion of active dipoles in the ground truth that are also present in the estimated activity. The average recovery rates of the



**Fig. 5.** Results for synthetic data with three active dipoles with low SNR: histograms of the hyperparameters by the proposed method after convergence. The actual values of  $\omega$  and  $\sigma_n^2$  are marked with a red vertical line, showing close values to the ones sampled by the algorithm (a) Histogram of  $\omega$ , (b) Histogram of  $a$ , (c) Histogram of  $\sigma_n^2$ .





**Fig. 6.** Results for synthetic data with multiple active dipoles: Performance as a function of the amount of active dipoles ( $P$ ).

proposed method and the  $\ell_{21}$  mixed norm approach are presented in Fig. 6a as a function of  $P$ . For  $P \leq 10$ , the proposed algorithm detects the non-zeros with an accuracy higher than 90% which drops to 60.2% for  $P=11$  and 49.7% for  $P=12$ . This drop of the recovery rate when a large number of non-zeros is present in the ground truth is well known, since the possible amount of non-zeros to recover correctly is limited by the operator span (Candes, 2008). For comparison, the  $\ell_{21}$  mixed norm regularization recovers up to  $P=5$  non-zeros with an accuracy higher than 90% and its recovery rate decreases slowly to reach 64% for  $P=10$ . Note that the proposed method has a higher recovery rate than the  $\ell_{21}$  approach for  $P \leq 11$ . Beyond this point, the poor performance of both methods prevents them from being used in real applications.

It is also interesting to analyze how much activity is present in the residual non-zeros. Thus, we define the proportion of residual energy as the amount of energy contained in the measurements generated by the residual non-zeros with respect to the total energy in the measurements. Fig. 6b shows the value of the residual energy obtained for both algorithms as a function of  $P$ . The  $\ell_{21}$  approach has up to 7.7% of the activity detected in residual non-zeros whereas the proposed algorithm never exceeds 1.1%, confirming its good sparsity properties.

## 4.2. Real data

Two real data sets are considered in this section. The first data set corresponds to the auditory evoked responses to left ear pure tone stimulus while the second one consists of the evoked responses to facial stimulus. The results of the proposed method are compared with the weighted  $\ell_{21}$  mixed norm (Gramfort et al., 2012), the Champagne model (Wipf et al., 2010) and the method investigated in Friston et al. (2008) based on multiple sparse priors.

### 4.2.1. Auditory evoked responses

The default data set of the MNE software (Gramfort et al., 2014, 2013) is used in this section. It consists of the evoked response to left-ear auditory pure-tone stimulus using a realistic BEM (Boundary element method) head model sampled with 60 EEG electrodes and 306 MEG sensors. The head model contains 1.844 dipoles located on the cortex with orientations that are normal to the brain surface. Two channels that had technical artifacts were ignored. The data was sampled at 600 Hz. The samples were low-pass filtered at 40 Hz and downsampled to 150 Hz. The noise covariance matrix was estimated from 200 ms of the data preceding each stimulus and was used to whiten the measurements. Fifty-one epochs were averaged to calculate the measurements  $\mathbf{Y}$ . The activity of the source dipoles was estimated jointly for the period from 0 ms to 500 ms after the stimulus. It is expected to find the brain activity primarily focused on the auditory cortices that are located close to the ears in both hemispheres of the brain (Gramfort et al., 2012).

The uncertainty principle was used to adjust the hyperparameter of the  $\ell_{21}$  mixed norm leading to having activity distributed all over the brain as shown in Fig. 7(a). By manually adjusting the

hyperparameter to produce a sparser result, it is possible to obtain a solution that has activity in the auditory cortices as shown in Fig. 7(b). In contrast, the proposed algorithm estimates its hyperparameters automatically and finds most of the activity in the auditory cortices without requiring any manual adjustment. The MSP method also finds the activity in both auditory cortices whereas the Champagne model finds an active patch on one of them. In addition, Fig. 8 shows the estimated waveforms using the  $\ell_{21}$  mixed norm and the proposed method. As we can see, they both have sharp peaks between 80 and 100 ms after the application of the stimulus, as expected in the response to an auditory stimulus (Gramfort et al., 2012).

To summarize, the proposed method finds the EEG source activity in areas that are spatially coherent with those found by the MSP and the Champagne methods. The main difference between the results is that the proposed method estimates the brain activity to be in only a few dipoles whereas the other algorithms estimate its extent. This is due to the sparsity-promoting prior that focuses the brain activity on the most important sources. Note that more details about the experiment are available in Costa et al. (2015).

### 4.2.2. Facial evoked responses

In a second step, data acquired from a face perception study where the subject was required to evaluate the symmetry of a mixed set of faces and scrambled faces was used, one of the default data sets of the SPM software.<sup>2</sup> Faces were presented during 600 ms every 3600 ms. The measurements were obtained by the electrodes of a 128-channel ActiveTwo system with a sampling frequency of 2048 Hz. The measurements were downsampled to 200 Hz and, after artifact rejection, 299 epochs corresponding to the non-scrambled faces were averaged and low-pass filtered to 40 Hz. A T1 MRI scan was then downsampled to generate a 8196 dipole head model.

The estimated activities are shown in Fig. 9. Using the proposed method, one can note that the activity is localized close to the fusiform region in the occipital lobe (Kanwisher et al., 1997). This is in good agreement with the results obtained by the MSP and Champagne algorithms. Again, the difference is related to the focalization of the activity among a reduced number of dipoles.

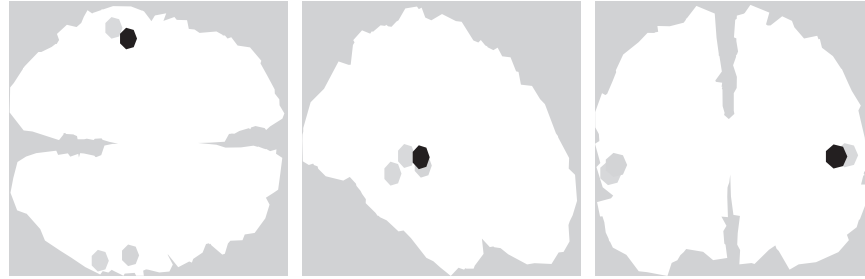
## 4.3. Computational cost

It is important to note that the price paid for the proposed method, while having several advantages over the  $\ell_{21}$  mixed norm approach, is its higher computational complexity. This problem is typical with MCMC methods when compared to optimization techniques. More precisely, the low SNR three-dipole experiment was processed in 6 s using a modern Xeon CPU E3-1240 @ 3.4 GHz processor (and a Matlab implementation with MEX files written in C) against 104 ms for the  $\ell_{21}$  mixed norm approach. However, it is interesting to note that the  $\ell_{21}$  norm approach requires running

<sup>2</sup> The SPM software is freely available at <http://www.fil.ion.ucl.ac.uk/spm>.



(a) Weighted  $\ell_{21}$  norm - Uncertainty principle for parameter  $\lambda$



(b) Weighted  $\ell_{21}$  norm - Manual adjustment of parameter  $\lambda$



(c) Proposed method

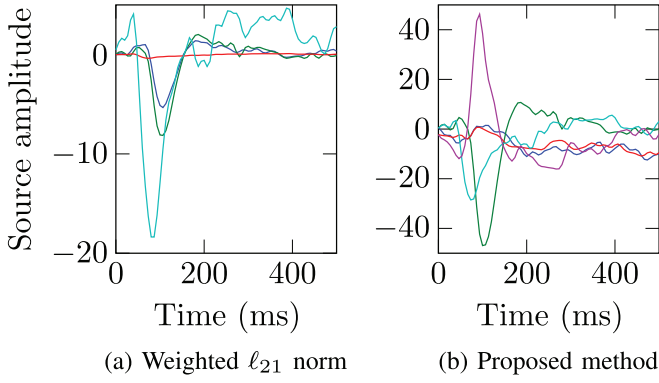


(d) MSP algorithm



(e) Champagne

**Fig. 7.** Results for real data auditory evoked responses: Active dipole positions estimated by different algorithms. The proposed method finds the activity focused in both auditory cortices as the MSP algorithm does. The Champagne model finds the activity in one of the auditory cortices (a) Weighted  $\ell_{21}$  norm - Uncertainty principle for parameter  $\lambda$ , (b) Weighted  $\ell_{21}$  norm - Manual adjustment of parameter  $\lambda$ , (c) Proposed method, (d) MSP algorithm, (e) Champagne.



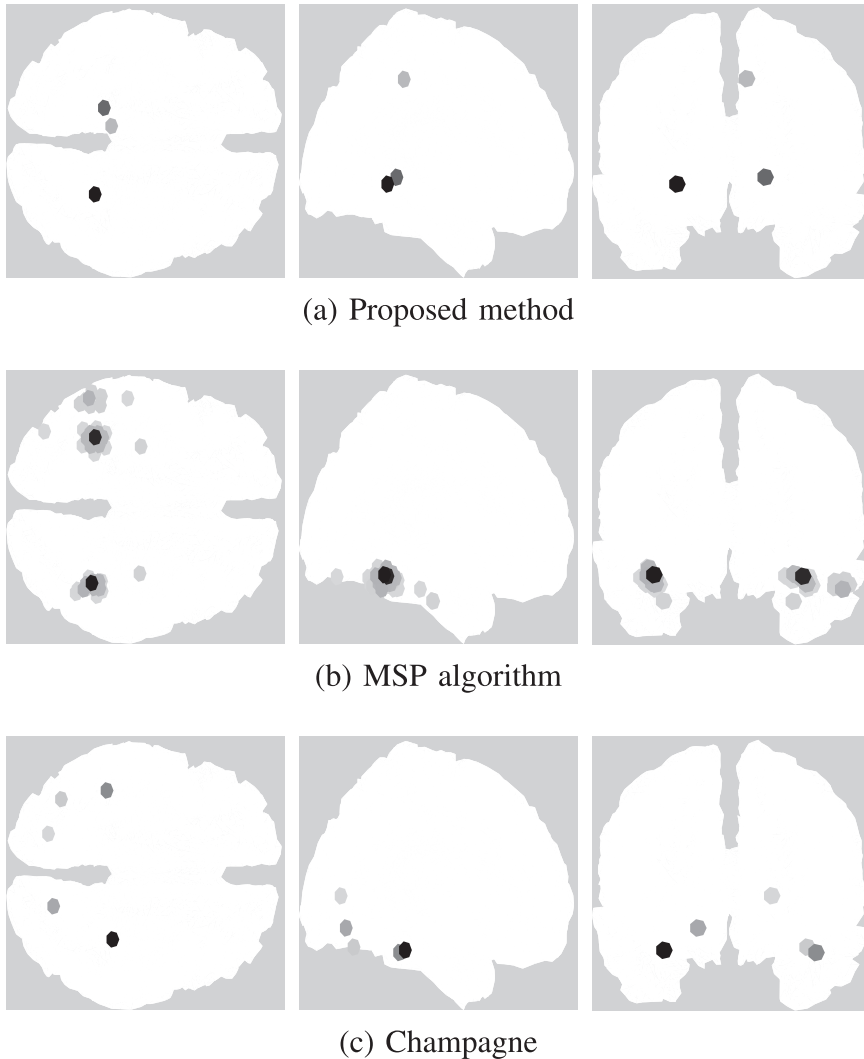
**Fig. 8.** Estimated waveforms for real data auditory evoked responses: Measurements and estimated activation waveforms. Both algorithms find the peak activity around 90ms after the stimulus was applied as expected (a) Weighted  $\ell_{21}$  norm, (b) Proposed method.

the algorithm multiple times to adjust the regularization parameter.

## 5. Conclusion

We presented a Bayesian mathematical model for sparse EEG

reconstruction that approximates the  $\ell_{20}$  mixed norm in a Bayesian framework by a multivariate Bernoulli Laplacian prior. A partially collapsed Gibbs sampler was used to sample from the target posterior distribution. We introduced multiple dipole shift proposals within each MCMC chain and exchange moves between different chains to improve the convergence speed. Using the generated samples, the source activity was estimated jointly with the model hyperparameters in an unsupervised framework. The proposed method was compared with the  $\ell_{21}$  mixed norm, the Champagne algorithm and a method based on multiple sparse priors in a wide variety of situations including several multi-dipole synthetic activations and two different real data sets. Using synthetic data sets, the proposed algorithm presented several advantages including better recovery of dipole locations and waveforms in low SNR conditions, the capacity of correctly detecting a higher amount of non-zeros, providing sparser solutions and avoiding underestimation of the activation amplitude. Using the real data sets, the proposed algorithm finds activity in locations that are spatially coherent with those found by the MSP and the Champagne algorithms. Finally, the possibility of providing several solutions with their corresponding probabilities is interesting. Future work will be devoted to a generalization of the proposed model to cases where the head model is not precisely known. Possible future options are to extend the current work to run



**Fig. 9.** Results for facial evoked responses: Active dipole positions estimated by different algorithms. The proposed method finds the activity in locations that are compatible with the ones estimated by the Champagne and MSP algorithms (a) Proposed method, (b) MSP algorithm, (c) Champagne.

models for estimating the spread patterns of the activity in epilepsy.

## Acknowledgement

This work is part of the DynBrain project funded by STIC-AM-SUD in collaboration with ITBA (Buenos Aires) and the FLENI hospital (Buenos Aires).

The authors would also like to thank Professor Steve McLaughlin from Heriot Watt University for improving the readability of the paper.

## Appendix A. Parameter priors

In this appendix the priors that were used for the variance of the noise  $\sigma_n^2$  and the hyperparameters  $a$  and  $\omega$  are detailed.

### A.1. Noise variance activity $\sigma_n^2$

The noise variance is assigned a Jeffrey's prior

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} \mathbf{1}_{\mathbb{R}^+}(\sigma_n^2) \quad (21)$$

where  $\mathbf{1}_{\mathbb{R}^+}(\xi) = 1$  if  $\xi \in \mathbb{R}^+$  and 0 otherwise. This choice is very classical when no information about a scale parameter is available (see Casella and Robert, 1999 for details).

### A.2. Hyperprior of $a$

A conjugate gamma prior is assigned to  $a$

$$f(a|\alpha, \beta) = \mathcal{G}(a|\alpha, \beta) \quad (22)$$

with  $\alpha = \beta = 1$ . These values of  $\alpha$  and  $\beta$  yield a vague hyperprior for  $a$ . The conjugacy of this hyperprior will make the analysis easier.

### A.3. Hyperprior of $\omega$

A uniform prior on  $[0, 1]$  is used for  $\omega$

$$f(\omega) = \mathcal{U}_{[0,1]}(\omega) \quad (23)$$

reflecting the absence of knowledge for this hyperparameter.

## Appendix B. Conditional distributions

The conditional distributions of the model parameters used in Algorithm 1 are detailed below.

### B.1. Conditional distribution of $\tau_i^2$

The conditional distribution of  $\tau_i^2$  is a gamma ( $\mathcal{G}$ ) or a generalized inverse Gaussian ( $\mathcal{GIG}$ ) distribution depending on the value of  $z_i$ . More precisely

$$f(\tau_i^2|\mathbf{x}_i, \sigma_n^2, a, z_i) = \begin{cases} \mathcal{G}\left(\tau_i^2 \mid \frac{T+1}{2}, \frac{v_i a}{2}\right) & \text{if } z_i = 0 \\ \mathcal{GIG}\left(\tau_i^2 \mid \frac{1}{2}, v_i a, \frac{\|\mathbf{x}_i\|^2}{\sigma_n^2}\right) & \text{if } z_i = 1. \end{cases} \quad (24)$$

### B.2. Conditional distribution of $\mathbf{x}_i$

The conditional distribution of the  $i$ th row of  $\mathbf{X}$  is

$$f(\mathbf{x}_i|z_i, \mathbf{Y}, \mathbf{X}_{-i}, \sigma_n^2, \tau_i^2) = \begin{cases} \delta(\mathbf{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_i, \sigma_i^2) & \text{if } z_i = 1 \end{cases} \quad (25)$$

with

$$\boldsymbol{\mu}_i = \frac{\sigma_i^2 \mathbf{h}_i^T (\mathbf{Y} - \mathbf{H}\mathbf{X}_{-i})}{\sigma_n^2}, \quad \sigma_i^2 = \frac{\sigma_n^2 \tau_i^2}{1 + \tau_i^2 \mathbf{h}_i^T \mathbf{h}_i}. \quad (26)$$

### B.3. Conditional distribution of $z_i$

The conditional distribution of  $z_i$  is a Bernoulli distribution

$$f(z_i|\mathbf{Y}, \mathbf{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) = \mathcal{B}\left(z_i \mid 1, \frac{k_1}{k_0 + k_1}\right) \quad (27)$$

with

$$k_0 = 1 - \omega, \quad k_1 = \omega \left(\frac{\sigma_n^2 \tau_i^2}{\sigma_i^2}\right)^{-\frac{T}{2}} \exp\left(\frac{\|\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right). \quad (28)$$

### B.4. Conditional distribution of $a$

The conditional distribution of  $a|\tau^2$  is the following gamma distribution

$$f(a|\tau^2) = \mathcal{G}\left(a \mid \frac{N(T+1)}{2} + \alpha, \frac{\sum_{i=1}^N v_i \tau_i^2}{2} + \beta\right). \quad (29)$$

### B.5. Conditional distribution of $\sigma_n^2$

The distribution of  $\sigma_n^2|\mathbf{Y}, \mathbf{X}, \tau^2, \mathbf{z}$  is the following inverse gamma ( $\mathcal{IG}$ ) distribution

$$\mathcal{IG}\left(\sigma_n^2 \mid \frac{(M + \|\mathbf{z}\|_0)T}{2}, \frac{1}{2} \left[ \|\mathbf{H}\mathbf{X} - \mathbf{Y}\|^2 + \sum_{i=1}^N \frac{\|\mathbf{x}_i\|^2}{\tau_i^2} \right]\right). \quad (30)$$

### B.6. Conditional distribution of $\omega$

Finally,  $\omega|\mathbf{z}$  has the following beta distribution

$$f(\omega|\mathbf{z}) = \mathcal{B}e(\omega|1 + \|\mathbf{z}\|_0, 1 + N - \|\mathbf{z}\|_0). \quad (31)$$

## References

- Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., et al., 2010. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia* 51 (4), 676-685.
- Bourguignon, S., Carfantan, H., Jul. 2005. Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data. In: Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP), Bordeaux, France.
- Candes, E.J., 2008. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci.* 346 (9), 589-592.
- Casella, G., Robert, C.P., 1999. Monte Carlo Statistical Methods. Springer-Verlag, New York.
- Castaño-Candamil, S., Höhne, J., Martínez-Vargas, J.-D., An, X.-W., Castellanos-Domínguez, G., Haufe, S., 2015. Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints. *NeuroImage* 118, 598-612.
- Chen, X., Godsil, S., May 2013. Multiple dipolar sources localization for MEG using

- Bayesian particle filtering. In: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP), Vancouver, Canada.
- Costa, F., Batatia, H., Oberlin, T., Tourneret, J.-Y., 2015. Bayesian Structured Sparsity Priors for EEG Source Localization Technical Report. University of Toulouse, ENSEIHT, Technical Report. [Online]. Available: <http://arxiv.org/abs/1509.04576>.
- Costa, F., Batatia, H., Chaari, L., Tourneret, J.-Y., 2015. Sparse EEG source localization using Bernoulli Laplacian priors. *IEEE Trans. Biomed. Eng.* 62 (12), 2888–2898.
- da Silva, F.L., Van Rotterdam A., Biophysical aspects of EEG and magnetoencephalogram generation. In: *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. Williams & Wilkins, Baltimore.
- Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage* 39 (3), 1104–1120.
- Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdés-Sosa, P., 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage* 23 (2), 435–453.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7 (4), 457–511.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman & Hall, London, UK.
- Geyer, C.J., Thompson, E.A., 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Soc.* 90 (431), 909–920.
- Geyer, C.J., Oct. 1991. Markov chain Monte Carlo maximum likelihood. In: *Proceedings of the 23rd Symposium on Interface Computer Science and Statistics*, Seattle, USA.
- Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* 57 (7), 1937.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al., 2013. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7 (267), 1–13.
- 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.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. *NeuroImage* 86, 446–460.
- Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., Vanrumste, B., 2008. Review on solving the inverse problem in EEG source analysis. *J. Neuroeng. Rehabil.* 4, 5–25.
- Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K.P., Fabri, S.G., Van Huffel, S., et al., 2007. Review on solving the forward problem in EEG source analysis. *J. Neuroeng. Rehabil.* 4, 46–75.
- Huang, J., Zhang, T., 2010. The benefit of group sparsity. *Ann. Stat.* 38 (August (4)), 1978–2004.
- Huang, J., Zhang, T., Metaxas, D., 2011. Learning with structured sparsity. *J. Mach. Learn. Res.* 12, 3371–3412.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17 (11), 4302–4311.
- Kowalski, M., Siedenburg, K., Dorfner, M., 2013. Social sparsity! Neighborhood systems enrich structured shrinkage operators. *IEEE Trans. Signal Process.* 61 (10), 2498–2511.
- Lamus, C., Hämäläinen, M.S., Temereanca, S., Brown, E.N., Purdon, P.L., 2012. A spatiotemporal dynamic distributed solution to the MEG inverse problem. *NeuroImage* 63 (2), 894–909.
- Laskey, K.B., Myers, J.W., 2003. Population Markov chain Monte Carlo. *Mach. Learn.* 50, 175–196.
- Long, C.J., Purdon, P.L., Temereanca, S., Desai, N.U., Hämäläinen, M.S., Brown, E.N., 2011. State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing. *Ann. Appl. Stat.* 5 (2B), 1207–1228.
- Lucka, F., Pursiainen, S., Burger, M., Wolters, C., 2012. Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: depth localization and source separation for focal primary currents. *NeuroImage* 61 (april (4)), 1364–1382 [Online]. Available: [/2012/LPBW12](https://doi.org/10.1016/j.neuroimage.2012.04.041).
- Marinari, E., Parisi, G., 1992. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* 19 (6), 451–458.
- Maris, E., 2003. A resampling method for estimating the signal subspace of spatio-temporal EEG/MEG data. *IEEE Trans. Biomed. Eng.* 50 (8), 935–949.
- Morozov, V.A., 1966. On the solution of functional equations by the method of regularization. *Sov. Math. Dokl.* 7, 414–417.
- Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18 (1), 49–65.
- Pascual-Marqui, R., et al., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Findings Exp. Clin. Pharmacol.* 24D, 5–12.
- Pascual-Marqui, R.D., 1999. Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.* 1 (1), 75–86.
- Quintero-Rincón, A., Pereyra, M., DGiano, C., Batatia, H., Risk, M., 2016. A new algorithm for epilepsy seizure onset detection and spread estimation from eeg signals. In: *Journal of Physics: Conference Series*, vol. 705, no. 1. IOP Publishing, San Nicolás de los Arroyos, Argentina, p. 012032.
- Raman, S., Fuchs, T.J., Wild, P.J., Dahl, E., Roth, V., Jun, 2009. The Bayesian group-lasso for analyzing contingency tables." In: *Proceedings of the 26th ACM Annual International Conference on Machine Learning (ICML)*, Montreal, Quebec.
- Somersalo, E., Voutilainen, A., Kaipio, J., 2003. Non-stationary magnetoencephalography by Bayesian filtering of dipole models. *Inv. Probl.* 19 (5), 1047–1063.
- Sommariva, S., Sorrentino, A., 2014. Sequential Monte Carlo samplers for semi-linear inverse problems and application to magnetoencephalography. *Inv. Probl.* 30 (11), 114 020–114 043.
- Sorrentino, A., Johansen, A.M., Aston, J.A., Nichols, T.E., Kendall, W.S., et al., 2013. Dynamic filtering of static dipoles in magnetoencephalography. *Ann. Appl. Stat.* 7 (2), 955–988.
- Stahlhut C., Attias, H.T., Sekihara, K., Wipf, D., Hansen, L.K., Nagarajan, S.S., 2013. A hierarchical Bayesian M/EEG imaging method correcting for incomplete spatio-temporal priors. In: *Proceedings of the IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, San Francisco, USA, April.
- Stok, C.J., 1986. *The Inverse Problem in EEG and MEG with Application to Visual Evoked Responses*. Ph.D. dissertation, University of Twente, Enschede, The Netherlands.
- Strohmeier, D., Haueisen, J., Gramfort, A., Improved MEG/EEG source localization with reweighted mixed-norms. 2014. In: *Proceedings of the 4th International Workshop on Pattern Recognition in Neuroimaging 2014 (PRNI 2014)*. IEEE, Tübingen, Germany, pp. 1–4.
- Uutela, K., Hämäläinen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* 10 (2), 173–180.
- Wipf, D.P., Owen, J.P., Attias, H.T., Sekihara, K., Nagarajan, S.S., 2010. Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. *NeuroImage* 49 (1), 641–655.
- Yu, G., Sapiro, G., Mallat, S., 2012. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *IEEE, Tübingen, Germany, Trans. Image Process.* 21 (5), 2481–2499.
- Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc.* 68 (1), 49–67.
- Zhang, Z., Rao, B.D., 2011. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Topics. Signal Process.* 5 (5), 912–926.