frontiers in Neuroscience

METHODS
published: 16 November 2017
doi: 10.3389/fnins.2017.00635

Spatio Temporal EEG Source Imaging with the Hierarchical Bayesian Elastic Net and Elitist Lasso Models

Deirel Paz-Linares[1,2], Mayrim Vega-Hernández[2], Pedro A. Rojas-López[1,2], Pedro A. Valdés-Hernández[2,3], Eduardo Martínez-Montes[2,4] and Pedro A. Valdés-Sosa[1,2*]*

[1] The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China, [2] Neuroinformatics Department, Cuban Neuroscience Center, Havana, Cuba, [3] Department of Biomedical Engineering, Florida International University, Miami, FL, United States, [4] Politecnico di Torino, Turin, Italy

OPEN ACCESS

Edited by:
Alexandre Gramfort,
Inria Saclay - Île-de-France Research Centre, France

Reviewed by:
Jean-Marc Lina,
École de Technologie Supérieure, Canada
Mathurin Massias,
Inria Saclay - Île-de-France Research Centre, France

***Correspondence:**
Eduardo Martínez-Montes
eduardo@cneuro.edu.cu
Pedro A. Valdés-Sosa
pedro.valdes@neuroinformatics-collaboratory.org

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 06 May 2017
Accepted: 01 November 2017
Published: 16 November 2017

Citation:
Paz-Linares D, Vega-Hernández M, Rojas-López PA, Valdés-Hernández PA, Martínez-Montes E and Valdés-Sosa PA (2017) Spatio Temporal EEG Source Imaging with the Hierarchical Bayesian Elastic Net and Elitist Lasso Models. Front. Neurosci. 11:635. doi: 10.3389/fnins.2017.00635

The estimation of EEG generating sources constitutes an Inverse Problem (IP) in Neuroscience. This is an ill-posed problem due to the non-uniqueness of the solution and regularization or prior information is needed to undertake Electrophysiology Source Imaging. Structured Sparsity priors can be attained through combinations of (L1 norm-based) and (L2 norm-based) constraints such as the Elastic Net (ENET) and Elitist Lasso (ELASSO) models. The former model is used to find solutions with a small number of smooth nonzero patches, while the latter imposes different degrees of sparsity simultaneously along different dimensions of the spatio-temporal matrix solutions. Both models have been addressed within the penalized regression approach, where the regularization parameters are selected heuristically, leading usually to non-optimal and computationally expensive solutions. The existing Bayesian formulation of ENET allows hyperparameter learning, but using the computationally intensive Monte Carlo/Expectation Maximization methods, which makes impractical its application to the EEG IP. While the ELASSO have not been considered before into the Bayesian context. In this work, we attempt to solve the EEG IP using a Bayesian framework for ENET and ELASSO models. We propose a Structured Sparse Bayesian Learning algorithm based on combining the Empirical Bayes and the iterative coordinate descent procedures to estimate both the parameters and hyperparameters. Using realistic simulations and avoiding the inverse crime we illustrate that our methods are able to recover complicated source setups more accurately and with a more robust estimation of the hyperparameters and behavior under different sparsity scenarios than classical LORETA, ENET and LASSO Fusion solutions. We also solve the EEG IP using data from a visual attention experiment, finding more interpretable neurophysiological patterns with our methods. The Matlab codes used in this work, including Simulations, Methods, Quality Measures and Visualization Routines are freely available in a public website.

Keywords: EEG source imaging, inverse problem, sparsity regularization, sparse Bayesian learning, empirical Bayes, elastic net, mixed norms, elitist lasso

INTRODUCTION

Electrophysiological Source Imaging (ESI) constitutes a relatively inexpensive and non-invasive approach to study neural activity with a high temporal resolution. ESI is a classic example of an Inverse Problem (usually referred to as EEG IP), given the little amount of data available as compared to the large number of parameters needed to model the spatially distributed whole brain activity this problem is ill-posed in the Hadamard sense (Hadamard, 1923). Its mathematical properties are in the first place determined by the forward model, i.e., the equation relating the electrical potential measured at the scalp ($\mathcal{V}$) and the originating Primary Current Density (PCD) $\mathcal{J}$ created by the electrical activity of large neuronal masses in every time t, which is a Type I Fredholm Integral Equation[1]:

$$\mathcal{V}(\mathbf{r}_e, t) = \int \mathcal{K}(r_e, r) \cdot \mathcal{J}(r, t)\, d\mathbf{r}^3 \tag{1}$$

Here, the kernel $\mathcal{K}(r_e, r)$ is called Electric Lead Field (ELF), and maps the space of brain generators r to the position of the measuring electrodes r_e. The ELF can be computed by using the quasistatic approximation of Maxwell equations and a model of the geometric and electrical properties of the head (Riera and Fuentes, 1998). The EEG IP is then defined as estimating the parameters $\mathcal{J}$ given measurements $\mathcal{V}$ and a known ELF. In practice, discretization of Equation (1) leads to a linear system:

$$V = KJ + \varepsilon \tag{2}$$

where the unknown parameters (J) constitute an $S \times T$ matrix (T is the number of time instants and S is the number of spatial generators, i.e., voxels) that represents the discretized PCD. V and ε are $N \times T$ matrices, representing the measured EEG and the measurement noise in N electrodes (generally no more than 128), usually distributed according to a standard system (Klem et al., 1999). It need be noted that the discretized ELF, K, is an $N \times S$ ill-conditioned matrix due to the large correlations among its columns. Since the region of potential brain generators is usually of thousands of voxels, then $N << S$ and the system of linear Equation (2) is highly underdetermined, i.e., it does not have a unique solution. This discretization and dimensionality reduction affects the accuracy of reconstructed solutions. Additionally, the forward model by itself consists in an oversimplification of the real generator space of EEG data and electromagnetic properties of the head. Therefore, some authors have suggested to test the inverse solutions with a different head model than the one used for simulations, in order to evaluate the methods in typical real data scenarios, when the individual MRI and head model is not available. This would allow the study to avoid committing the so-called "Inverse Crime" (Kaipio and Somersalo, 2004).

[1]For the mathematical notation used in this paper see Appendix G in Supplementary Materials.

TABLE 1 | Different models for Tikhonov Regularization and their corresponding penalty functions.

	Model	Penalty function
a)	Ridge	$\lVert J \rVert_2^2$
b)	LORETA	$\lVert \mathbb{L}J \rVert_2^2$
c)	LASSO	$\lVert J \rVert_1$
d)	LASSO Fusion	$\lVert \mathbb{L}J \rVert_1$
e)	MPLS	$\sum_k \lambda_k \psi_k(J)$
f)	Fused LASSO	$\lambda_1 \sum_{i=1}^{S-1} \lVert J_{i,\cdot} \rVert_1 + \lambda_2 \sum_{i=1}^{S-1} \lVert J_{i+1,\cdot} - J_{i,\cdot} \rVert_1$
g)	ENET	$\lambda_1 \lVert J \rVert_2 + \lambda_2 \lVert J \rVert_1$
h)	ENETL	$\lambda_1 \lVert \mathbb{L}J \rVert_2 + \lambda_2 \lVert \mathbb{L}J \rVert_1$
i)	Smooth LASSO	$\lambda_1 \lVert \mathbb{L}J \rVert_2 + \lambda_2 \lVert J \rVert_1$
j)	MXN	$\lVert J \rVert_{p,q} = \left(\sum_{t=1}^{T} \left(\sum_{i=1}^{S} \lvert J_{i,t} \rvert^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}$
k)	GLASSO	$\sum_{s=1}^{S} \left(\sum_{t=1}^{T} \lvert J_{i,t} \rvert^2 \right)^{\frac{1}{2}}$
l)	ELASSO	$\sum_{t=1}^{T} \left(\sum_{i=1}^{S} \lvert J_{i,t} \rvert \right)^2$

Sparsity Regularization with L1/L2 Norms Models: The Elastic Net and the Elitist Lasso

A well-known way to find a unique solution to the EEG IP is through Tikhonov Regularization (Tikhonov and Arsenin, 1977). It uses additional or prior information on the parameters, which is usually introduced in the form of mathematical constraints and it is closely related to Penalized Least Squares (PLS), which is generally expressed as:

$$\hat{J} = argmin_J \left\{ \lVert V - KJ \rVert_2^2 + \lambda P(\mathbb{L}J) \right\} \tag{3}$$

The first term in the cost function in (3) is the residual of the model and the second summarizes the imposed constraints (penalty terms). The regularization parameter λ controls the relative weight of the penalty function P. The matrix $\mathbb{L}$ adds information about the correlation structure of the parameter matrix. In the context of the EEG IP, it has taken the form of the identity matrix (independent PCD in each voxel) or the Laplacian operator (discrete second derivative) for requiring spatial smoothness of the PCD.

A well-known example of Tikhonov Regularization is the *Ridge* method (Hoerl and Kennard, 1970), where the penalty function is the L2 norm of the parameters, **Table 1a**. Particular cases of this method have given rise to several well-established inverse solutions such as Minimum-Norm, Weighted Minimum-Norm and Low Resolution Electromagnetic Tomography (LORETA), **Table 1b**. Their advantages and disadvantages have been thoroughly studied (Pascual-Marqui, 1999).

Sparse PCDs have also been searched through methods based on such as the Least Absolute Shrinkage Selection Operator (LASSO) (Tibshirani, 1996), **Table 1c**, and its version LASSO Fusion (Land and Friedman, 1996), **Table 1d**, which use the L1 norm as the penalty function.

In the last years, sparsity constraints based on L1/L2 norm have become very popular to achieve a high spatial resolution,

in reconstructing and differentiating sources of brain activity, associated to different cerebral functions/states. This constraint can also fit better with physiological knowledge about brain activity in particular experimental and real-life situations. However, its mathematical treatment is not straightforward and strongly depends on the models and algorithmic approaches.

A special example consists in algorithms performing smooth and sparse estimations in separate steps (Liu et al., 2005; Palmero-Soler et al., 2007) to explicit combining models as the sum of L1 and L2 penalty functions while using iterative algorithms to solve it (Nagarajan et al., 2006; Valdés-Sosa et al., 2009; Tian and Li, 2011).

These L1/L2 models have been included in a framework called Multiple Penalized Least Squares (MPLS) (Valdés-Sosa et al., 2006; Vega-Hernández et al., 2008), **Table 1e**, which consists in a generalization of formula (3) to multiples penalty terms with their corresponding lambdas.

The Elastic Net in the Context of Multiple Penalized Least Squares Models

A very promising approach in the context of MPLS is based on a flexible model, combining L1/L2 norms, called the Elastic Net (ENET) (Zou and Hastie, 2005), **Tables 1g,h.** Solving this model is of particular interest for incorporating together the advantages of the L2 norm (Ridge) and L1 norm (LASSO) family models. Other particular examples of MPLS' are the Fused LASSO (Tibshirani et al., 2005), **Table 1f**, based only in the L1 norm, and the smooth LASSO (Hebiri, 2008), **Table 1i**, combining L1 and L2 norms

For minimizing the PLS's (or its MPLS generalization) cost function in (3), considering models such as the Elastic Net, many algorithms exhibit advantages in the convergence time. Some Important examples are general modified Newton-Raphson algorithms, such as Local Quadratic Approximation (LQA) (Fan and Li, 2001) and Majorization-Minorization (MM) (Hunter and Li, 2005), without a considerable loss of speed in computations.

In a seemingly different approach combining L1/L2 norms, some authors have used the idea of structured sparse penalization based on *mixed-norms* (MXN) models (Kowalski and Torrésani, 2009a,b), **Table 1j**. In this context using the L1 norm of a vector whose elements are obtained as the L2 norms of other vectors is known as Group Lasso (GLASSO) (Yuan and Lin, 2006; Kowalski and Torrésani, 2009a,b), **Table 1k**. An important application of the GLASSO penalty is for example the Focal Vector Field reconstruction (Haufe et al., 2008) where sparsity is imposed on the amplitude of the PCD but keeping smoothness in the 3 spatial components (x, y, z) that defines the direction of this vector magnitude. The penalization function is then the L1 norm of the vector formed by the L2 norms of the PCD vector in each voxel.

The Elitist Lasso in the Context of Mixed-Norm Models

With the same goal, a model based on the L2 norm of a vector whose elements are obtained as the L1 norms of other vectors, has been called Elitist Lasso (ELASSO) (Kowalski and Torrésani, 2009a,b), **Table 1l**. This type of penalization was extended to the spatio-temporal context, consisting in the application of an L1 norm along the first dimension of the parameter matrix, and an L2 norm along the second dimension (Ou et al., 2009). This model presents similarities with respect to the MPLS models discussed above, particularly the ENET, therefore it is a suitable candidate among the MXN models for taking advantages of algorithmic solutions developed for those ones.

Although originally a second-order cone programming was used (Ou et al., 2009), it has been shown that these models for imposing structured sparsity can be estimated by a generalized shrinkage operator (Kowalski and Torrésani, 2009b). However, the regularization approach using MXN as a penalty function becomes a convex, non-differentiable, irrational, and non-separable (along columns or rows) optimization problem, which makes the inference process computationally very expensive.

More recently, efficient proximal operators and gradient-based based algorithms have been developed to compute a solution to the spatio-temporal EEG IP, by using special cases of mixed-norms (e.g., FISTA, Beck and Teboulle, 2009; Gramfort et al., 2012, 2013).

Limitations of the Regularization Approach

The estimated solutions using the algorithms discussed above, for the MPLS and MXN formulations of L1/L2 norms, become very sensitive to the set of constraints selected to regularize them and to the values of the regularization parameters. Within the state-of-the-art literature, usually the regularization parameters are selected ad hoc, or via heuristic information criteria, as Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Generalized Cross Validation (GCV). These procedures lead to compute the Inverse Solution for a large set of values of the regularization parameters, in order to evaluate the information criteria, thus being computationally expensive, in many cases non-optimal and usually leading to an inadequate balance of the constraints. Critical examples are the Multiple Shrinkage Problem of MPLS models, such as the ENET (Zou and Hastie, 2005; Vega-Hernández et al., 2008), and the selection of regularization parameters for the MXN models in Gramfort et al. (2012, 2013) and Gramfort et al. (2014).

Bayesian Inference and Sparse Bayesian Learning

As dealing with the regularization parameters estimation in the Classical Inference approach constitutes an important limitation, other approaches have been proposed within the Bayesian framework (Schmidt et al., 1999; Trujillo-Barreto et al., 2004; Wipf et al., 2006). The Bayesian Inference have emerged as an alternative formulation, which is more general and can incorporate models, equivalent to the Tikhonov Regularization.

This framework addresses the EEG IP in terms of finding the posterior *pdf* of the parameters (J) using the Bayes rule:

$$p\,(J|V,\alpha,\beta) \propto p\,(V|J,\beta)\,p\,(J|\alpha) \tag{4}$$

Where $p\,(V|\,J,\beta)$ is the likelihood, $p\,(J\,|\alpha)$ is the parameters' prior and (α,β) are hyperparameters. The prior plays a similar role as a penalty function in the classical framework. As a particular case, the *maximum a posteriori* estimate in this

formalism leads to a fully equivalent solution to the optimization problem in Equation (3), when the likelihood is a Gaussian (with mean KJ and unit variance) and the prior has the following exponential form (using $\alpha = \lambda$):

$$p\left(J|\alpha\right) = \frac{1}{z}e^{-\alpha P(\mathbb{L}J)} \tag{5}$$

An important advantage of the Bayesian framework is that in a second level of inference (MacKay, 2003) one can estimate the hyperparameters of interest, typically variances and/or precisions of the likelihood and the priors. One way to achieve this is by using the Empirical Bayes procedure, where the aim is to maximize the posterior of hyperparameters:

$$\left(\hat{\alpha}, \hat{\beta}\right) = argmax_{(\alpha,\beta)} p\left(\alpha, \beta | V\right) \tag{6}$$

This approach has been widely developed in the last years, especially with sparse priors, which has been called Sparse Bayesian Learning (SBL). An important example is the Relevance Vector Machine (RVM) (Tipping, 2001; Schmolck and Everson, 2007), consisting in estimating sparse parameter vectors (with a Gaussian prior) while learning the hyperparameters (with a Gamma prior). Although theoretically Bayesian Learning methods are more robust for computing the inverse solution, fast and efficient inference algorithms have been currently developed only for simple models (Tipping and Faul, 2003). Bayesian formulations of the ENET (Li and Lin, 2010) or multiple variants of LASSO in the MPLS family (Kyung et al., 2010) have been developed previously, by using scale Gaussian mixtures (Andrews and Mallows, 1974). Although scale Gaussian mixtures constitute a considerable improvement for computing these models still involves the use of computationally intensive Expectation-Maximization (EM) and/or Monte Carlo (MC) algorithms, which are impractical for the EEG IP due its high dimensionality.

Objectives

In this work we propose a new method based on an Empirical Bayes formalism to promote Structured Sparsity through two models based on mixtures of L1/L2 norms penalization functions (Laplace/Normal priors):

- The **Elastic Net** (ENET) (Zou and Hastie, 2005; Vega-Hernández et al., 2008).
- The **Elitist Lasso** (ELASSO) (Kowalski and Torrésani, 2009a,b).

Although we develop the theory for ENET and ELASSO models, the whole methodology can be generally applied to other types of Laplace/Normal priors. The algorithm can be considered as a Sparse Bayesian Learning method since it allows to estimate the regularization (hyper-) parameters that control sparsity and perform variable selection, while at the same time avoiding Monte Carlo estimations or EM type algorithms.

The aims of this study are to present the theory and an insightful validation of the methods as electrophysiological source localization methods with a large set of simulations and real EEG visual Event Related Potentials. We also compare the performance of these models with other well-established methods in the ERP synthetic data and discuss the theoretical and experimental differences. The Matlab codes used in this work, including Simulations, Methods, and Quality Measures are freely available upon request to the authors.

METHODS

Hierarchical Elastic Net Model

The Bayesian formulation of the ENET can be interpreted as a Laplace/Normal prior model[2], with associated hyperparameters (α_1, α_2) that are inversely related to the variances of the Normal and Laplace *pdfs*:

$$p\left(J\,|\alpha_1, \alpha_2\,\right) = \frac{1}{z}e^{-\sum_t\left(\alpha_{1,t}\|J_{\cdot,t}\|_2^2 + \alpha_{2,t}\|J_{\cdot,t}\|_1\right)} \tag{7}$$

Where $\alpha_{1,t}$ and $\alpha_{2,t}$ represents the tth element of the hyperparameters vectors (α_1, α_2), the notation $J_{\cdot,t}$ refers to the tth column of J, and z is the normalization constant.

Since the squared L2 norm and the L1 norm are separable as a sum of terms depending on the individual components, this prior introduces stochastic independence in the whole spatio-temporal map of the parameters, i.e., Equation (7) can be factorized over voxels and time points (indexed by i and t, respectively) with corresponding normalization constants $z_{i,t}$:

$$p\left(J\,|\alpha_1, \alpha_2\,\right) = \prod_{t,i} \frac{1}{z_{i,t}} e^{-\alpha_{1,t}J_{i,t}^2 - \alpha_{2,t}|J_{i,t}|} \tag{8}$$

This prior can be expressed as a scale Gaussian mixture (Andrews and Mallows, 1974), with a mixing *pdf* corresponding to a Truncated Gamma on the new hyperparameter γ. This formulation can be found in the Bayesian literature with different parametrizations (Kyung et al., 2010; Li and Lin, 2010). In our case, it is convenient to reorganize the variances and mixing distribution as detailed in the following Lemma.

Lemma 1 [Proof in Appendix A (Supplementary Materials)]: Let $J_{i,t}$ be the random variable distributing with pdf $\frac{1}{z_{i,t}}e^{-\alpha_{1,t}J_{i,t}^2}e^{-\alpha_{2,t}|J_{i,t}|}$, where $z_{i,t}$ is a normalization constant. Then the following equality holds:

$$\frac{1}{z_{i,t}}e^{-\alpha_{1,t}J_{it}^2 - \alpha_{2,t}|J_{i,t}|} = \int_0^{+\infty} N\left(J_{i,t}|0, \Lambda_{i,t}\right) TGa\left(\gamma_{i,t}\left|\frac{1}{2}, 1, \left(k_t, \infty\right)\right.\right) d\gamma_{i,t} \tag{9}$$

$$\Lambda_{i,t} = \frac{1}{2\alpha_{1,t}}\overline{\Lambda}_{i,t}, \quad \overline{\Lambda}_{i,t} = \left(1 - \frac{k_t}{\gamma_{i,t}}\right) \tag{10}$$

$$k_t = \frac{\alpha_{2,t}^2}{4\alpha_{1,t}} \tag{11}$$

Where $TGa\left(\gamma_{i,t}\left|\frac{1}{2}, 1, \left(k_t, \infty\right)\right.\right)$ is the Truncated Gamma pdf, with lower truncation limit k_t.

[2]Note that this mixture of Laplacian and Gaussian distributions is not the Normal-Laplace distribution described in Reed (2006).

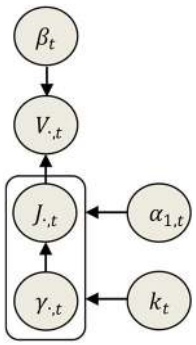

FIGURE 1 | Directed network describing the instantaneous relationship between data ($V_{\cdot,t}$), parameters ($J_{\cdot,t}$) and hyperparameters (β_t, $\gamma_{\cdot,t}$, $\alpha_{1,t}$, k_t) in the Elastic Net model.

With this particular formulation of the ENET prior we change the relevant hyperparameters from the old regularization parameters (α_1, α_2) to (α_1, k). The first one acts as the scale of variances of $J_{i,t}$ (Equation 10), and the second acts as the lower truncation limit of the *pdf* for $\gamma_{i,t}$. Although this makes more challenging the interpretation of the influence of the hyperparameters, it speeds the Empirical Bayes algorithm (described in section Structured SBL algorithm for the Elastic Net and Elitist Lasso models) and ensures its convergence. Given that k is now directly learned (without estimating α_2, independently of α_1), it is necessary to impose a convenient gamma prior for this hyperparameter. For α_1 we can just choose a non-informative prior. Applied to the EEG IP stated in Equation (2), this new Bayesian ENET model can be analytically described by the following set of *pdfs*, see also the schematic representation in **Figure 1**.

Likelihood:

$$V_{\cdot,t} \sim N\left(V_{\cdot,t} \middle| KJ_{\cdot,t}, \beta_t I\right) \tag{12}$$

Prior of parameters:

$$J_{\cdot,t} \sim N\left(J_{\cdot,t} \middle| 0, diag\left(\Lambda_{\cdot,t}\right)\right) \tag{13}$$

Prior of hyperparameters:

$$\gamma_{\cdot,t} \sim \prod_i TGa\left(\gamma_{i,t} \middle| \frac{1}{2}, 1, \left(k_t, \infty\right)\right) \tag{14}$$

$$\beta_t \sim \ non\text{-}informative, \ \log p(\beta_t) = const \tag{15}$$

$$\alpha_{1,t} \sim \ non\text{-}informative, \ \log p(\alpha_{1,t}) = const \tag{16}$$

$$k_t \sim Ga\left(\tau + 1, \upsilon\right), \ \tau > 0, \ \upsilon > 0 \tag{17}$$

Here β_t plays the role of the noise variance and τ and υ are the scale and shape parameters of the Gamma prior for k_t.

In the original Bayesian ENET model the parameter's prior (8) controls the degree of sparsity/smoothness through variable selection according to the values of α_1 and α_2 directly, i.e., when $\alpha_1 \ll \alpha_2$ we have that $p\left(J \middle| \alpha_1, \alpha_2\right) \approx La(J|0, \ \alpha_2)$, promoting sparse solutions, and when $\alpha_2 \ll \alpha_1$ we have that $p\left(J \middle| \alpha_1, \alpha_2\right) \approx N\left(J \middle| 0, \frac{1}{2\alpha_1}\right)$, which promotes smoothness for not-too-large values of α_1. In our formulation of the ENET model the variable selection is performed in a similar fashion in the hyperparameters level, but in this case when k is large ($\alpha_1 \ll \alpha_2$ for finite α_1) we have by Equation (14) that $\gamma \approx k$, which implies by Equation (10) that $\Lambda \rightarrow 0$, finally promoting sparsity; and when k is small ($\alpha_2 \ll \alpha_1$), $\Lambda \rightarrow \frac{1}{2\alpha_1}$ by Equation (10) so that in Equation (13) we have that $p\left(J \middle| \gamma, \alpha_1, k\right) \approx N\left(J \middle| 0, \frac{1}{2\alpha_1}\right)$.

Hierarchical Elitist Lasso Model

Here, we propose a Bayesian formulation of the Elitist Lasso model using the following multivariate *pdf*:

$$p\left(J\right) = \frac{1}{Z} e^{-\alpha \|J\|_{1,2}^2} \tag{18}$$

where α is a hyperparameter that controls simultaneously the level of spatial and temporal sparsity. This prior can be easily written for each time point separately by decomposing the mixed-norm as $\|J\|_{1,2}^2 = \sum_t \left\|J_{\cdot,t}\right\|_1^2$. However, the penalization over one spatial component $J_{i,t}$ cannot be separated from the remaining components $J_{j,t}$ $\left(j \neq i\right)$. We then propose to rearrange $\left\|J_{\cdot,t}\right\|_1^2$ in the form $\left(\delta_{i,t} + \left|J_{i,t}\right|\right)^2$, for any index i, where $\delta_{i,t} = \sum_{j \neq i} \left|J_{j,t}\right|$ is a magnitude carrying the dependence on the other components of the vector parameters. This allows us to derive a hierarchical Bayesian model for this formulation based on the theory of Markov Random Fields.

Definition 1: Let x be a random vector or random matrix, which has joint pdf of the form $p\left(x\right) = \left(\frac{1}{Z}\right) e^{-\mathcal{P}(x)}$, where $\mathcal{P}\left(x\right)$ can be decomposed as $\sum_{(i,j) \in \mathcal{J}} \mathcal{P}_{ij}\left(x_i, x_j\right)$; $\mathcal{J}$ is a set of pairs of index and Z is a normalization constant. Then we say that x is a "Pair-wise Markov Random Field" (pMRF), and the functions $\left\{\mathcal{P}_{ij}\left(x_i, x_j\right)\right\}$ are defined as potentials.

Definition 2: Let x be a random vector or random matrix. Let $x_{\mathcal{H}}$ denote a subset of elements of x, which has conditional joint pdf of the form $p\left(x_{\mathcal{H}} \middle| x_{\mathcal{H}^C}\right) = e^{-\alpha \mathcal{P}(x)} / Z$, where $x_{\mathcal{H}^C}$ is the complement of $x_{\mathcal{H}}$ in x, $\mathcal{P}\left(x\right)$ can be decomposed as $\sum_{(i,j) \in \mathcal{J}} \mathcal{P}_{ij}\left(x_i, x_j\right)$, $\mathcal{J} \subseteq \left\{\left(i,j\right) : x_i \bigvee x_j \in x_{\mathcal{H}}\right\}$, and Z is a normalization constant. Then we say that $\left(x_{\mathcal{H}}, x_{\mathcal{H}^C}\right)$ is a "Pair-wise Conditional Markov Random Field" (pCMRF). [Note that any couple of sets $\left(x_{\mathcal{H}}, x_{\mathcal{H}^C}\right)$ from a pMRF constitutes a pCMRF; see the proof of Lemma 2 (a) in Appendix B (Supplementary Materials)].

In our case, the random vector consisting in a column of the spatio-temporal matrix J constitutes a pMRF, since its prior can be written as:

$$p\left(J_{\cdot,t} \middle| \alpha\right) = \frac{1}{z} e^{-\alpha \left\|J_{\cdot,t}\right\|_1^2} \tag{19}$$

Where the exponent can be decomposed as:

$$\left\|J_{\cdot,t}\right\|_1^2 = \sum\nolimits_i J_{i,t}^2 + \sum\nolimits_{i>j} 2\left|J_{i,t}\right|\left|J_{j,t}\right|; \quad (i,j = 1, \ldots, S) \tag{20}$$

Hence the potentials of the Definition 1 are proportional to:

$$P_{ij} = \begin{cases} J_{i,t}^2 & , i = j \\ 2\left|J_{i,t}\right|\left|J_{j,t}\right| & , i \neq j \end{cases} \tag{21}$$

pMRFs underlies many models of the Statistical Physics approach (Kindermann and Snell, 1980) and also learning algorithms (Murphy, 2012). In the case of the model described by Equations (19–21), we can reformulate the prior in a hierarchical model at the parameters level. For this, we need to use some relevant properties posed in the following Lemma.

Lemma 2 [Proofs in Appendix B (Supplementary Materials)]: The following properties can be verified for the conditional probabilities in the pCMRF associated to the Elitist Lasso model of Equation (19).

(a)
$$p\left(J_{i,t}\middle|J_{i^{\mathcal{C}},t},\alpha\right)=p\left(J_{i,t}\middle|\delta_{i,t},\alpha\right)=\frac{1}{Z_{i,t}}e^{-\alpha J_{i,t}^2}e^{-2\alpha\delta_{i,t}\left|J_{i,t}\right|} \quad (22)$$

Where $\delta_{i,t}=\sum_{j\neq i}\left|J_{j,t}\right|$ and $Z_{i,t}$ is a normalization constant.

(b)
$$p\left(\delta_{\cdot,t}\right)=\frac{1}{\overline{Z}}e^{-\alpha\left\|W^{-1}\delta_{\cdot,t}\right\|_1^2}I_{\mathcal{R}_+^S}\left(\delta_{\cdot,t}\right) \quad (23)$$

Where $I_{\mathcal{R}_+^S}(\delta_{\cdot,t})$ is the indicator function of the region $\mathcal{R}_+^S=\left\{x\in\mathcal{R}^S:x_i\geq 0,i=1,\ldots,S\right\}$, the matrix W can be expressed as the subtraction of the identity matrix from a ones-matrix: $W=1_{S\times S}-\mathbb{I}_{S\times S}$, and $\overline{Z}$ is a normalization constant.

(c)
$$p\left(J_{i,t}\middle|\alpha\right)=\int p\left(J_{\cdot,t}\middle|\alpha\right)dJ_{i^{\mathcal{C}},t}=$$
$$\int\prod\nolimits_i p\left(J_{i,t}\middle|\delta_{i,t},\alpha\right)p\left(\delta_{\cdot,t}\middle|\alpha\right)dJ_{i^{\mathcal{C}},t}d\delta_{\cdot,t} \quad (24)$$

Using the properties (a,b) we show that taking the conditional *pdf* (22), instead of Equation (19), leads to a simpler model, separable in the parameters level. This implies considering $\delta_{i,t}$ as new hyperparameters, which inherit the multivariate behavior (spatial correlation structure) of J in the original model. As proved in Appendix B (Supplementary Materials), using a Dirac conditional distribution for this hyperparameters given the parameters, we can reformulate the non-separated model of Equation (19) into a hierarchical model that factorizes over different components of the parameters. Property (c) validates the use of the new model proposed in (a,b), showing that the marginal *pdf* of the parameters in the original model is identical to the marginal prior in the new one. Although this approximation is algebraically simpler, its effect might be related to other known approximations such as the mean field or a more general one in the Variational Bayes approach (Friston et al., 2007; Trujillo-Barreto et al., 2008). Also, it could be seen as a local ENET approximation of the ELASSO penalty, since the latter can then be explicitly re-written as a sum of an L2 and L1 norms of the parameters. Indeed, realizing that Equation (22) represents a combination of Normal and Laplace priors for $J_{i,t}$, a further transformation of the Bayesian mixed-norm model is possible, combining the results in Lemma 1 and Lemma 2 and dealing with α and $2\alpha\delta_{i,t}$ as hyperparameters, similarly to α_1 and α_2 in the ENET model. The resulting prior of parameters is given below, the analogous schematic representation is showed in **Figure 2**:

$$J_{\cdot,t}\sim N\left(J_{\cdot,t}\middle|0,diag\left(\Lambda_{\cdot,t}\right)\right) \quad (25)$$

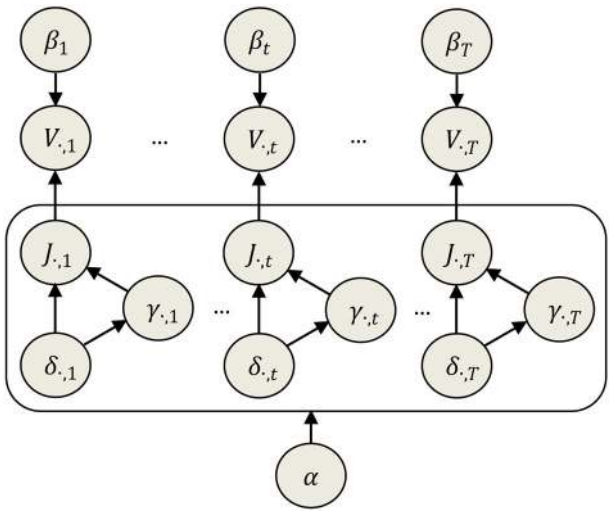

FIGURE 2 | Directed network describing the instantaneous relationship between data (V), parameters (J), and hyperparameters (β, γ, δ, α) in the Elitist Lasso model. Note that alpha affects all time points simultaneously, i.e., it is just one hyperparameter controlling the same level of sparsity for all time points.

$$\Lambda_{i,t}=\frac{1}{2\alpha}\overline{\Lambda}_{i,t},\qquad \overline{\Lambda}_{i,t}=\left(1-\frac{\alpha\delta_{i,t}^2}{\gamma_{i,t}}\right) \quad (26)$$

And the priors of hyperparameters are:

$$\gamma_{\cdot,t}\sim TGa\left(\gamma_{i,t}\middle|\frac{1}{2},1,\left(\alpha\delta_{i,t}^2,\infty\right)\right) \quad (27)$$

$$\delta_{\cdot,t}\sim\frac{1}{Z}e^{-\alpha\left\|W^{-1}\delta_{\cdot,t}\right\|_1^2}I_{\mathcal{R}_+^S}\left(\delta_{\cdot,t}\right) \quad (28)$$

$$\beta_t\sim\ non\text{-}informative,\ \ \log p(\beta_t)=const$$
$$\alpha\sim\ non\text{-}informative,\ \ \log p\left(\alpha\right)=const \quad (29)$$

Similarly to ENET, the matrix $diag\left(\Lambda_{\cdot,t}\right)$ can be interpreted as an effective prior variance of the parameters. However, here, the degree of sparsity in variable selection is twofold. On one hand, through the regularization parameter α, which is unique for the whole spatio-temporal map and imposes the same degree of sparsity to each column (time point) of parameters. On the other hand, the hyperparameters $\delta_{i,t}$ controls sparsity locally, so that when $\delta_{i,t}$ is large we will have $\gamma_{i,t}\approx\alpha\delta_{i,t}^2$, by Equation (27), and the effective prior variances will tend to zero $\Lambda_{i,t}\rightarrow 0$, by Equation (26), promoting higher degree of sparsity independently of α.

Structured SBL Algorithm for the Elastic Net and Elitist Lasso Models

For the Elastic Net and Elitist Lasso models the joint distribution of data, parameters, and hyperparameters in the spatio-temporal EEG IP admits the following factorization:

$$p\left(V,J,\Theta\right)=\prod\nolimits_t p\left(V_{\cdot,t},J_{\cdot,t},\Theta_t\right) \quad (30)$$

Where Θ represents the hyperparameters of both Elastic Net and Elitist Lasso: $\Theta_t = \begin{cases} \gamma_{\cdot,t}, \alpha_{1,t}, k_t, \beta_t\ ; & \text{ENET} \\ \gamma_{\cdot,t}, \delta_{\cdot,t}, \alpha, \beta_t\ ; & \text{ELASSO} \end{cases}$

Each factor in Equation (30) is the product of the likelihood, prior of parameters and priors of the hyperparameters:

$$p\left(V_{\cdot,t}, J_{\cdot,t}, \Theta_t\right) = N\left(V_{\cdot,t} \middle| KJ_{\cdot,t}, \beta_t I\right) N\left(J_{\cdot,t} \middle| 0, diag\left(\Lambda_{\cdot,t}\right)\right)$$

It is straightforward to derive an explicit formula for the maximum a posteriori estimate of the parameters J, from the data likelihood and its conjugated normal prior *pdf*, by using the result in formula [C1] of Appendix C (Supplementary Materials). The Empirical Bayes procedure then consists in minimizing the negative log-posterior of hyperparameters; see formula [C2] in Appendix C (Supplementary Materials).

$$\begin{aligned} \mathcal{L} &= -\log p\left(\Theta | V\right): \\ \mathcal{L} &= \frac{1}{2}\sum\nolimits_t \left\{ \log|\beta_t I| + \frac{1}{\beta_t}\left\| V_{\cdot,t} - K\mu_{\cdot,t} \right\|_2^2 + \log\left|diag\left(\Lambda_{\cdot,t}\right)\right| + \right. \\ &\left. + \mu_{\cdot,t}^{\mathcal{J}}\left(diag\left(\Lambda_{\cdot,t}\right)\right)^{-1}\mu_{\cdot,t} + \log\left|\overline{\Sigma}_t^{-1}\right| - 2\log p\left(\Theta_t\right) \right\} \end{aligned} \quad (31)$$

In Equation (31) $\mu_{\cdot,t}$ and Σ_t represents the corresponding instantaneous parameter's posterior mean and covariance. It is important to note that these estimates are equivalent to the classical Minimum Norm Estimates in the regularization approach, when the hyperparameters are known or comprised into one hyperparameter that can be estimated through heuristic methods (e.g., minimizing information criteria or the generalized cross-validation function) (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui, 2002; Sánchez-Bornot et al., 2008; Vega-Hernández et al., 2008). However, here, with the Empirical Bayes approach, the estimation of hyperparameters is carried out by optimizing the posterior *pdf* of hyperparameters in Equation (31) (also known as Type II Likelihood). As hyperparameters control the level of sparsity, with an iterative algorithm, even simple models can lead to solutions totally different to the MNE solutions (MacKay, 1992, 2003; Schmidt et al., 1999; Friston et al., 2002, 2008; Trujillo-Barreto et al., 2004; Wipf et al., 2006).

The Type-II Likelihood in Equation (31) will take different forms depending on the model choice (ENET or ELASSO), since they use different expressions for the effective prior variances, according to Equations (10) and (26). In both cases, the hyperparameters' priors follow from Equations (14) to (17) and (27) to (29), respectively:

ENET:

$$\begin{aligned} -\log p\left(\Theta_t\right) &= \sum\nolimits_i \left[\log\int_{k_t}^{\infty} Ga\left(x \middle| \frac{1}{2}, 1\right) dx - \log Ga\left(\gamma_{i,t} \middle| \frac{1}{2}, 1\right) \right] \\ &\quad - \log p\left(\alpha_{1,t}\right) - \log p\left(k_t\right) - \log p\left(\beta_t\right) \end{aligned}$$

ELASSO:

$$\begin{aligned} -\log p\left(\Theta_t\right) &= \sum\nolimits_i \left[\log\int_{\alpha\delta_{i,t}^2}^{\infty} Ga\left(x \middle| \frac{1}{2}, 1\right) dx - \log Ga\left(\gamma_{i,t} \middle| \frac{1}{2}, 1\right) \right] \\ &\quad + \log\overline{Z} + \alpha\left\| W^{-1}\delta_{\cdot,t} \right\|_1^2 - \log p\left(\alpha\right) - \log p\left(\beta_t\right) \end{aligned}$$

To minimize the non-convex function $\mathcal{L}$ we use an iterative algorithm based on the coordinate descent strategy (Tipping, 2001; Wipf and Nagarajan, 2009), regarding the arguments $\left\{\mu,\ \overline{\Lambda},\ \alpha, \alpha_1,\ k,\ \beta\right\}$, where we replaced γ by $\overline{\Lambda}$ in Equation (31) using Equations (10) and (26). We use the matrix Σ_t, which is a non-linear function of the hyperparameters, for the estimation of parameters and hyperparameters in the next step. In the particular case of the mixed-norm we do not minimize $\mathcal{L}$ over the hyperparameter δ, due to the non-differentiability, but instead we just update it by using its relationship with parameters given in Lemma 2, (a). Appendix D (Supplementary Materials) shows the update formulas -and their derivations- based on the matrix derivative of $\mathcal{L}$ with respect to hyperparameters in each model.

Hereinafter, these methods for structured sparse Bayesian learning (SSBL) will be called ENET-SSBL and ELASSO-SSBL. The pseudo codes for the corresponding algorithms are given in Appendix F (Supplementary Materials). For implementation details and a further comprehension of our algorithms check Appendix E (Supplementary Materials).

RESULTS

Simulation Study and Quality Measures

We evaluate the proposed methods by reconstructing the spatio-temporal PCD from simulated EEG data. We selected 3 sources, called Occipital "O," Motor "M" and Temporal Lobe "TL" patches, with different spatial extension with regard to the geodesic distance in the surface: 30, 20, and 10 mm radius, respectively, as shown in **Figure 3**, top panel. The simulation was set up in a cortical surface of 6003 possible generators extracted from the MNI template brain (http://www.bic.mni.mcgill.ca). We generated 4 different spatial configurations using these patches to explore situations with different hemispherical distribution of the three sources: Simulation 1 with all sources in the left (L) hemisphere (O-L, M-L, TL-L); Simulation 2 where the occipital source is changed to the right (R) hemisphere (O-R, M-L, TL-L); Simulation 3 changing the TL source to the right (O-L, M-L, TL-R); and Simulation 4 moving both O and TL to the right hemisphere (O-R, M-L, TL-R). Time courses of 1 s, using a 200 Hz sampling frequency (201 time points), were also simulated for each patch, as shown in **Figure 3**, bottom panel. Time courses for patches O and TL were slow waves of 1 and 3 Hz, respectively. The frequency ratio (1/3 Hz) was chosen in order to reach enough variability along time in the relative intensity of these sources. The time course for patch M corresponded to a narrow Gaussian pulse that appears at six different time points when the other two patches showed qualitatively different relative intensity. This course became sparse after thresholding it to set all time points with intensity below 0.1% of the maximum equal to 0. In general, this set up of temporal courses (one sparse and two smooth in time) for each source allows not only to test the ability of the inverse methods to correctly estimate the time course of activations but also to model many different combinations of number and extension of simulated sources, i.e., their effective sparsity. **Figure 3** shows that there are time points when all sources are active (e.g., $t1 = 0$ ms), others where only one (e.g.,

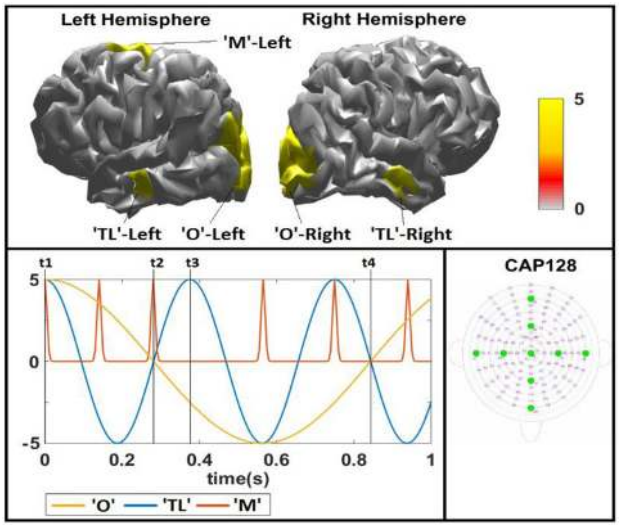

FIGURE 3 | Simulated spatio-temporal sources. **Top:** Left and right views of the simulated activation patches in generators space (red). Five simulated patches with different geodesic spatial extension (Occipital "O" 30 mm, Temporal Lobe "TL" 20 mm, Motor "M" 10 mm) in the Left and only Occipital and Temporal Lobe in the Right hemisphere. **Bottom:** Left: Simulated time evolution of sources, for the "O" (∼1 Hz cosine, yellow), the "TL" (∼3 Hz cosine, blue), and the "M" (narrow Gaussian pulse signal). We highlight some time points of interest (t1 = 0 ms, t2 = 280 ms, t3 = 375 ms, t4 = 845 ms) where relative intensity of sources lead to qualitatively different scenarios. Right: Spatial representation of the electrodes in the plane according to the 120–130 system.

t2 = 280 ms) or two (e.g., *t3* = 375 ms) are active or even those when none of them are active (e.g., around 845 ms). Different amplitudes in the time courses imply different amplitudes of corresponding active voxels in each source, effectively changing the extension of each source and the general sparsity of the map. With this type of simulation, we can explore the behavior of the performance of inverse solutions, and their estimated hyperparameters with the different levels of sparsity, directly by looking at them along time points. Note that this can also be considered a challenging non-ideal scenario for the methods proposed here, given the presence simulations with simultaneous sources of different extensions, also following different temporal courses.

The ELF was computed for 120 sensors disposed on the scalp (selection of Biosemi 128 electrodes cap), using the BEM method for the individual head model (volume conduction), with 4 layers: scalp, outer skull, inner skull, gray matter (Valdés-Hernández et al., 2009). The spatio-temporal scalp electric potential (EEG) was computed as the product of the ELF and the simulated PCD, corrupted by adding biological noise in the sources, following an autoregressive process with spectral peak between 8 and 12 Hz (Alpha rhythm). We also add sensors noise as a Gaussian white process, such that both sources and electrodes noise were adjusted to obtained a signal-to-noise ratio of 6 dB. We generated 100 trials for each of the 4 simulated spatio-temporal sources by adding 100 samples of the biological and sensors noise in order to avoid a biased interpretation of the results due to a good choice of noise by chance. To avoid the "Inverse Crime" which recognizes that usually EEG source localization is done using standard (approximate) electrodes positions, head geometry, tissue conductivities, etc., instead of the real individual head model, some authors have proposed to use an ELF with decreased spatial resolution of the generators space for computing the solutions (Kaipio and Somersalo, 2004). Remarkably, we use a more challenging approach consisting in computing the solutions using an ELF obtained from a different subject (i.e., changing the locations of the generators corresponding to similar anatomical areas and also the geometry of the whole head model).

For comparison purposes, we computed the LORETA, ENETL and LASSO Fusion solutions using classical regularization approach (Hunter and Li, 2005; Sánchez-Bornot et al., 2008; Vega-Hernández et al., 2008). Computations were done by using the Elastic Net linear regression software package GLMNET (Qian et al., 2013; https://web.stanford.edu/~hastie/glmnet_matlab/). The regularization parameters for these three methods were computed as those minimizing the GCV function in a suitable interval made up of 100 values. All these inverse solutions used a graph Laplacian matrix, which introduces prior information about neighbors' structure in the generators space. All codes used in this work, including simulations, inverse solutions, quality measures and visualization of the results are freely available upon request to the authors.

For a quantitative evaluation of the solutions, we computed some quality measures based on the comparison with the true simulated one. There are two distances: the Dipole Localization Error, DLE, which measure how far are the estimated local maxima from the simulated ones and the more general Earth Mover's Distance between simulated and estimated solutions (Grova et al., 2006; Haufe et al., 2008; Molins et al., 2008). The lower these measures are, the better the match between estimated and simulated PCD. Two other performance measures offer were derived from a Receiver Operating Characteristic (ROC) analysis: the "area under the curve" (AUC), which is a measure of how good is the estimated solution for resembling the simulated one using any threshold, and the source retrieval index (F1-score) which measures a compromise between the fraction of simulated sources correctly estimated as active (recall or sensitivity) and the fraction of estimated sources that were simulated as active (precision) (Bradley et al., 2016; Hansen, 2017). Importantly, we did not threshold the estimated solutions for computing any of the distance measures. ROC analysis finds the AUC using the binarized simulated solution (nonzero sources will be set to 1) and the non-thresholded solutions. None of the inverse methods tested here incorporates any thresholding, thus making all voxels to have nonzero values. In order to compute the F1-score, we allow the ROC analysis to select the optimal operating point of the ROC curve (using *perfcurve* function in Matlab) and use the corresponding threshold to compute the recall (sensitivity) and precision. F1-score is the geometric mean between these two measures. This means that the methods will be compared according to their best performance individually. Here, we report AUC and F1 as percentages where the highest 100% values

are achieved only by a perfect reconstruction of the simulated sources.

Bayesian Elastic Net and Elitist Lasso Solutions

The inverse solution was computed with the ENET-SSBL and ELASSO-SSBL algorithms for the 100 trials of each simulated configuration. A typical trial was selected as the one ranking in the place 50 (median) after ordering all trials according to their best overall performance using all quality measures (sum of all rankings). **Figure 4** shows the simulated and estimated spatial maps by both methods for the selected typical trials of each simulation, in the three time points of interest: *t1*, when all sources are similarly active; *t2* when the M source is maximum and the other two are close to zero and *t3*, when only the largest (O) and the smallest (TL) patches were nonzero.

Figure 5 shows the time evolution of the average activation of those voxels belonging to each patch, from both ENET-SSBL and ELASSO-SSBL solutions obtained in each simulated configuration (rows). These time courses are normalized between −1 and 1 for better visualization. Note that, despite being one trial in the median of the performance distribution, they resemble the original simulated time courses but without achieving the same level of smoothness for TL and O, or sparsity for M.

Finally, we computed the quality measures for each solution obtained using the two proposed methods from each of the 100 trials of each spatio-temporal simulation. **Table 2** shows the average and standard deviation values across the 100 trials for all measures in the three time points of interest. We highlighted in bold the better values for each comparison. It is easy to see that even if it looks that ELASSO-SSBL gives better values in the majority of comparisons, the mean values of the measures are closer than the sum of their standard deviations, suggesting that both methods have similar performance according to these measures.

Comparison of Loreta, Elastic Net, Lasso Fusion and SSBL Solutions

Figure 6 shows the simulated and estimated solutions using LORETA, ENETL, LASSO-Fusion, ENET-SSBL and ELASSO-SSBL in the same style as in **Figure 4**. In this case, the average and standard deviation maps across 100 trials are shown only for the first simulated configuration. The std maps are normalized by the maximum of the matching mean map, to better show those regions with high variability, which usually correspond to the regular appearance of spurious sources. In general, the proposed methods show the lowest variability and smaller regions of high variance -unrelated to true sources- than all other tested methods (especially in the sparsest *t2*).

Figure 7 shows the corresponding time evolution of the mean (curves) and std (error bars) of activations of voxels belonging to the different patches. In this case, we left the curves to show the scale of estimated solutions, in order to compare among different methods. Note that the estimation of the M patch is consistently biased with all methods, although the five narrow Gaussian pulses are clearly distinguished.

Consequently, we computed the quality measures EMD, DLE, AUC and F1-Score for all trials of Simulation 1. **Figure 8** shows the time evolution of the mean value of each quality measure for all methods. The sensitivity to the sparsity level in different time points is evident for all methods and all measures, especially showing the worst behavior around those time points where the M patch becomes active. However, it is clear that our two proposed methods offer the best average quality measures among all methods at all time points.

The mean and std values of all quality measures for the three time points *t1*, *t2* and *t3* in Simulation 1 are shown in **Table 3A**. Analogously to **Table 2**, we highlighted in bold the best performance for each row. In the average values, it is clearer that the ELASSO-SSBL achieves the best results overall, showing similar values as ENET-SSBL, but noticeably better than those of the other solutions. In order to test if these results were not driven by the particular selection of the spatial simulated patches, we ran 100 simulated trials with three similar patches but with randomized positions and extensions, without overlapping. **Table 3B** shows the corresponding mean and std values of the quality measures across these pseudo-random repetitions. It is easy to check that most of the values are very similar (within the std window) to those obtained in our particular simulated configuration, except for the F1-score which showed a much lower performance for all methods in this case. For each of these performance measures, we carried out non-parametric rank tests (Wilcoxon Rank-Sum Test, Gibbons and Chakraborti, 2011) between each pair of source localization methods, using all 100 trials of every time point of every of the four simulations. We then counted how many times (out of the 804 simulations, 201 time points × 4 simulations) each method was significantly better than another. Results are shown in **Table 4**. In this case, the two methods proposed here are almost always significantly better than any of the other methods in more than 50% of cases for all quality measures (except for the AUC comparison with ENETL).

Estimated Hyperparameters and Sensitivity to Sparsity Levels

One key point of this study is to test the ability of the proposed algorithm to correctly learn adequate values of the hyperparameters, and at the same time, check if the estimated hyperparameters effectively control the sparsity level of the solutions. In **Figure 9** we show the time evolution of the average (curves) and standard deviation (error bars) of the logarithm of the two hyperparameters ENET-SSBL, the two regularization parameters of ENETL and the regularization parameters of Lasso Fussion and LORETA. The regularization parameters were found automatically by cross-validation performed by the GLMNET toolbox. It is interesting to see that only the learned hyperparameters of ENET-SSBL, which are optimized for each time separately, closely follows the level of sparsity of the simulated solution (red curve); LORETA shows changes of the variance of the parameter, but LASSO Fusion and ENETL give very similar values for all time points (i.e., different levels of sparsity). The level of sparsity was computed as 1 minus the L1 norm of the solution in each time point, normalized by the maximum value along time. For visualization purposes, the curves was adjusted to fit in the same range of the hyperparameters/regularization parameters shown in each panel.

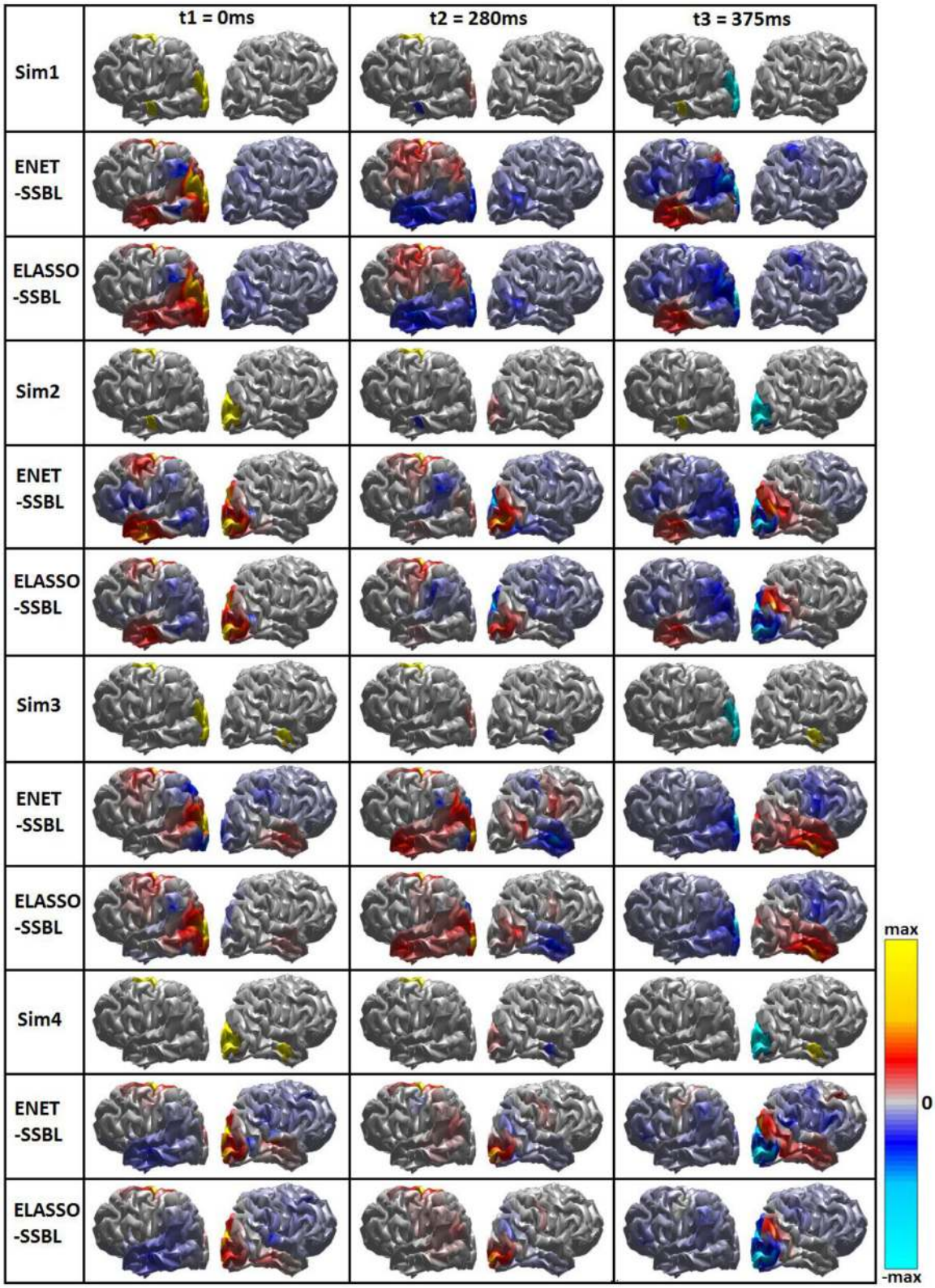

FIGURE 4 | Simulated and estimated spatial maps with SSBL algorithm for the two proposed models ENET and ELASSO. Inverse solutions for a typical trial of each of the 4 simulated configurations (Sim1, Sim2, Sim3, and Sim4) at three time points with different sparsity. Bipolar color map shows the solution within a window of ± the maximum absolute value, gray color corresponds to zero values. A logarithmic scale has been set for a better visualization.

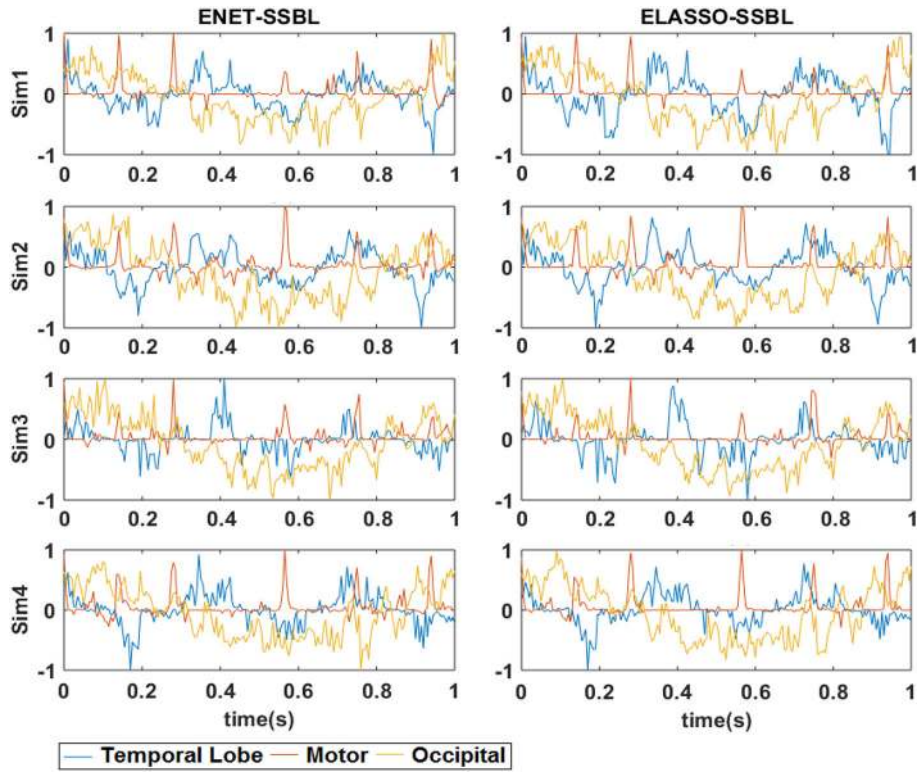

FIGURE 5 | Estimated temporal courses with SSBL algorithm for the two proposed models ENET and ELASSO. The curves show the normalized temporal courses of the average activation across all voxels belonging to each of the three simulated spatial patches: Occipital (yellow), Temporal Lobe (blue), and Motor (red). These correspond to the same typical trials as **Figure 4**, for each of the 4 simulated configurations (Sim1, Sim2, Sim3, and Sim4).

TABLE 2 | Mean and standard deviation of the quality measures Earth Mover's Distance (EMD), Distance Localization Error (DLE), Area Under receiver operating Curve (AUC), and F1-Score (F1), across 100 trials, at time points t1 = 0 ms, t2 = 280 ms, and t3 = 375 ms of Simulations 1, 2, 3, and 4, for both algorithms ENET-SSBL and ELASSO-SSBL.

		Simulation 1		Simulation 2		Simulation 3		Simulation 4	
		ENET-SSBL	ELASSO-SSBL	ENET-SSBL	ELASSO-SSBL	ENET-SSBL	ELASSO-SSBL	ENET-SSBL	ELASSO-SSBL
EMD	t1	176 ± 50.1	**170.6 ± 38**	156 ± 58.1	**105.3 ± 49.1**	**186.8 ± 46.5**	192 ± 40.3	138.8 ± 54.2	**90.9 ± 26.1**
	t2	**124.1 ± 42.7**	125.2 ± 46.9	136.4 ± 69.2	**110.7 ± 52.7**	**122.8 ± 55.1**	124.5 ± 55.8	136.8 ± 73.9	**111.6 ± 45.8**
	t3	143.6 ± 55.9	**142.2 ± 56.3**	146.9 ± 68.6	**118.6 ± 52**	**143.6 ± 55.8**	149.3 ± 57.1	153.6 ± 82.2	**117.7 ± 61**
DLE	t1	1.77 ± 0.43	**1.48 ± 0.42**	2.78 ± 0.6	**2.73 ± 0.71**	2.06 ± 0.78	**2.03 ± 0.86**	2.81 ± 0.64	**3.24 ± 0.59**
	t2	3.03 ± 0.86	**2.92 ± 0.93**	3.77 ± 0.97	**3.76 ± 1.09**	3.45 ± 0.98	**3.19 ± 0.89**	4.2 ± 1.09	**4.13 ± 1.09**
	t3	2.09 ± 0.99	**1.9 ± 1.12**	3.21 ± 1.08	**2.83 ± 1.34**	2.57 ± 1.32	**2.41 ± 1.32**	3.96 ± 1.39	**3.85 ± 1.69**
AUC	t1	93.9 ± 1	**94.1 ± 0.7**	91 ± 1.4	**91.4 ± 1.2**	92.8 ± 1.4	**93.1 ± 1.4**	91.5 ± 1.1	**91.4 ± 1.3**
	t2	90.1 ± 2.9	**90.2 ± 3.5**	**86.6 ± 5.6**	85.8 ± 5.9	88.6 ± 4.1	**89 ± 3.5**	85.9 ± 4.2	**85.3 ± 4.9**
	t3	93.7 ± 4.3	**93.9 ± 3.9**	**90.2 ± 4.5**	89.9 ± 6.1	93.6 ± 3.6	**93.7 ± 4.3**	90.7 ± 3.8	**90.6 ± 5.4**
F1	t1	61.5 ± 10.2	**67.6 ± 7.3**	35.2 ± 14.2	**38.1 ± 14.1**	60.7 ± 9.8	**67.3 ± 6.5**	37.3 ± 14	**40.2 ± 12.3**
	t2	37.8 ± 16.8	**40 ± 19.4**	**28.3 ± 17.5**	25.4 ± 18	37.5 ± 17.5	**38.8 ± 20**	**26.2 ± 17.5**	24.4 ± 18.5
	t3	50.7 ± 20.1	**50.8 ± 25.1**	34.1 ± 18	**36.4 ± 19.5**	54 ± 22.7	**56.4 ± 23.8**	35.6 ± 17	**37.4 ± 17.6**

Best results between ENET-SSBL and ELASSO-SSBL algorithms are highlighted in bold, while the best results among all simulations (along rows) are underlined.

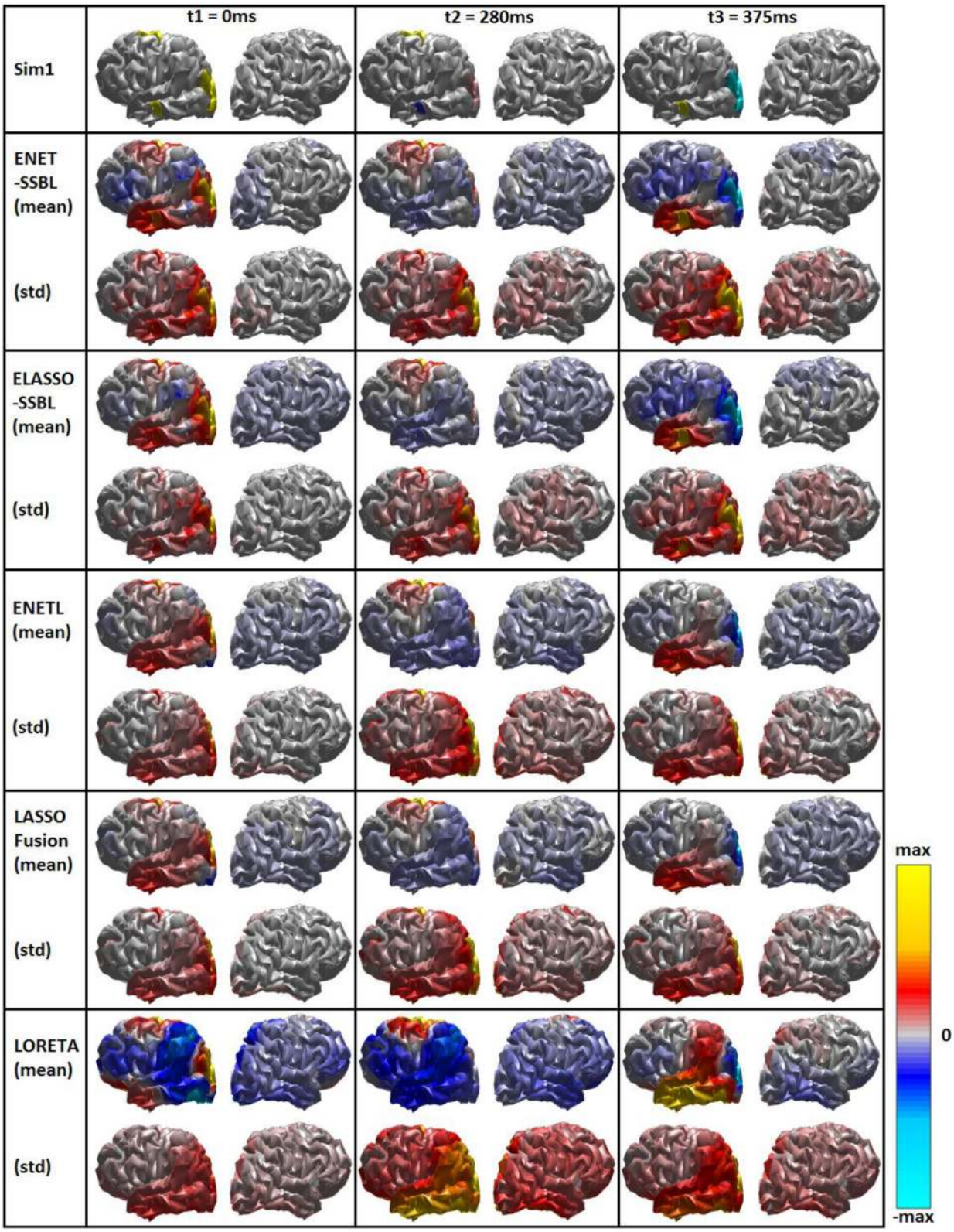

FIGURE 6 | Simulated (first row) and estimated spatial maps with the different inverse methods. Each rectangle shows the mean (Upper part) and the standard deviation (Lower part) maps across 100 trials for the first simulated configuration (Sim1), at the three time points of interest corresponding to different sparsity scenarios. Bipolar color map shows the solution within a window of ± the maximum absolute value, gray color corresponds to zero values. A logarithmic scale has been set for a better visualization.

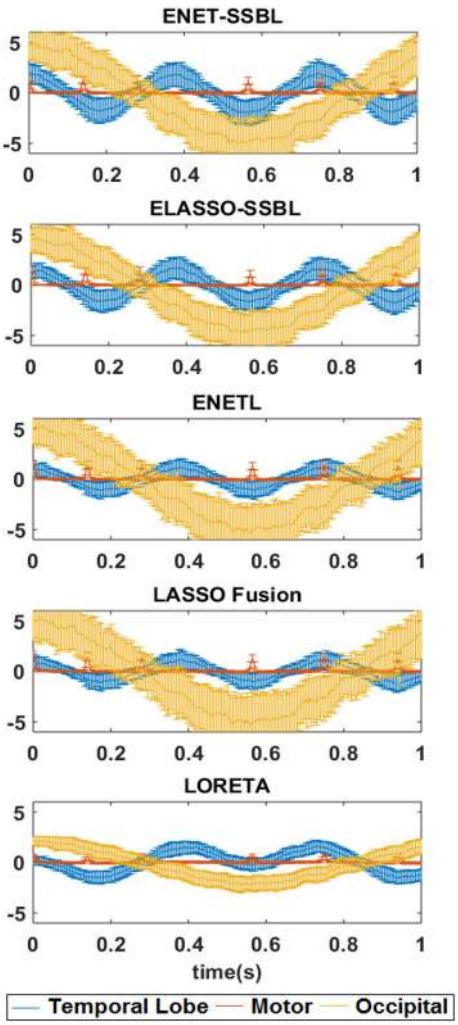

FIGURE 7 | Estimated temporal courses with the different source localization methods for the first simulated configuration. The temporal courses correspond to the average activation across all voxels belonging to each of the three simulated spatial patches: Occipital (yellow), Temporal Lobe (blue) and Motor (red). The mean curves and standard deviation (error bars) across all trials are shown (without normalization).

In the case of ELASSO-SSBL, the corresponding estimated values for the single hyperparameter was 5.18 ± 0.07. Typically, the SSBL algorithm for both methods reached convergence in around 15 iterations, 3.2 s (ENET-SSBL) and 4.2 s (ELASSO-SSBL) for obtaining a single inverse solution, in an Intel Core i5 CPU, at 1.70GHz and 4GB of RAM memory.

We finally compute the real sparsity level of estimated solutions with the different methods and compare it with that of the simulated PCD in Simulation 1. **Figure 10** shows the time evolution of the average L1 norms across the 100 trials of solutions for the different methods, together with that of the simulated solution (dashed curve). In this case, the L1 norm measures the level of activation, since it is usually minimized for sparse solutions, (i.e., it is an inverse measure of sparsity) and allows an easier relative comparison with the activation level of the simulated sources without normalization. Interestingly, all methods are able to resemble the change in sparsity of the simulation along time, although the ELASSO-SSBL solutions showed consistently closer values to the true ones among all other methods.

Real Data Study of a Visual Attention Experiment

EEG recordings using 30 electrodes (sampling frequency 128 Hz) were gathered from right-handed healthy subjects in a visual attention experiment devoted to study attentional modulation of early visual evoked potential, as explained in (Makeig et al., 1999, 2004; Makeig, 2002). Briefly, a sequence of geometric figures is presented as stimulus where the fixation point is defined as a cross and the subjects are requested to discriminate when a specific configuration appears in the attended location, by means of the physical action of pressing a button. The ethical clearance of this experiment was guaranteed by the Office of Naval Research according to the authors in the original paper (Makeig et al., 1999). The EEG recorded the brain activity for 25 to 75 repetitions of the stimulus, in 80 subjects, during a 3 s-long time window (1 s pre-stimulus and 2 s post-stimulus). Then, it is averaged over repetitions to cancel the background oscillatory activity. **Figure 11** shows the 30 electrode's voltage time series (384 time points) of the Grand Average (average across subjects) Visual Evoked Potential, marking the stimulus onset (A) and the two time instants selected to compute the inverse solutions. These corresponded to the global maximum negative peak (B: 281 ms post-stimulus) and global maximum positive peak (C: 430 ms post-stimulus). The median reaction times (i.e., when the subjects pressed a button) was about 350 ms.

To compute the ELF, we used an MNI standard brain and head model, defining a grid of 3244 generators (voxels) within the volume of the gray matter, the brainstem and the thalamus. The solutions were obtained by using the three orthogonal components of the Lead Field (three degrees of freedom), and also the same for the Laplacian matrix. The final sources maps were found as the L2 norm of each voxel of the estimated PCD. The ENET-SSBL and ELASSO-SSBL inverse solutions were compared with LORETA, one of the most used and well-studied solution. The results are visualized with the Neuronic Tomographic Viewer (http://www.neuronicsa.com/), using the maximum intensity projection view (i.e., projecting brain activations to three orthogonal planes). **Figures 12**, **13** show in color scale the solutions estimated with LORETA, ENET-SSBL and ELASSO-SSBL for the maximum negative and positive peak potentials (B and C time points, as marked in **Figure 11**), respectively.

Solutions for the ENETL and for the LASSO Fusion models were also estimated with the MM algorithm. **Figure 14** is a reproduction of Figure 3 in Vega-Hernández et al. (2008), and shows the ENETL solution with a set of regularization parameters, where λ is chosen via GCV for different fixed values of α (0.1; 0.01; 0.001). These values correspond to different ratios between regularization parameters ($\alpha_2/\alpha_1 = (1-\alpha)/\alpha$) from 9 (low sparsity) to 999 (high sparsity).

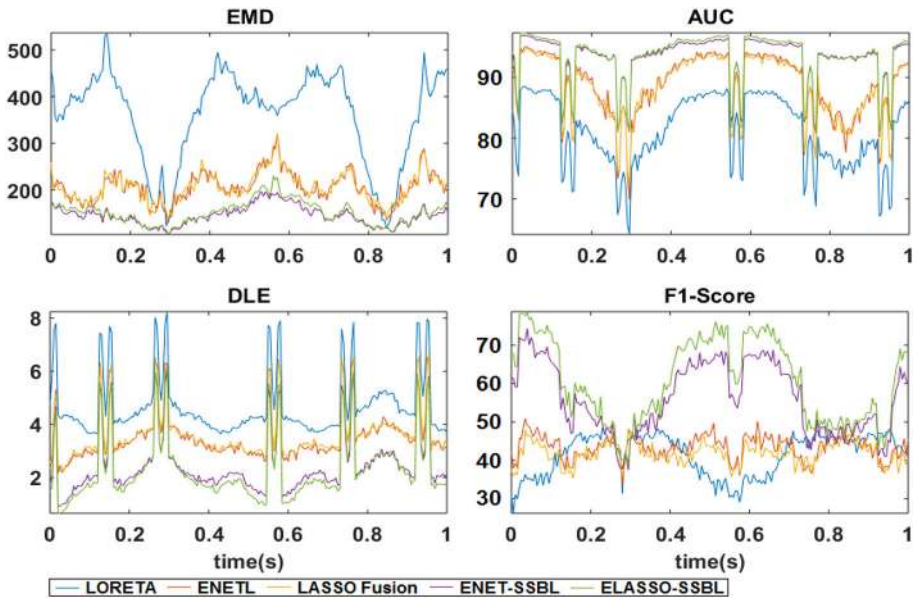

FIGURE 8 | Mean curves along time of the quality measures Earth Mover's Distance (EMD), Distance Localization Error (DLE), area under receiver operating curve (AUC), and F1-Score, across 100 trials of Simulation 1 from estimated solutions with LORETA, ENETL, LASSO Fusion, ENET-SSBL, and ELASSO-SSBL.

DISCUSSION

Analysis of the Algorithms ENET-SSBL and ELASSO-SSBL

The Bayesian formulation of the ENET model proposed here is similar to Li and Lin (2010), and allows variable selection using the second level of inference for hyperparameters learning. Distinctively, we reformulate the hyperparameters as a scale factor (α_1) of the parameters variances and the truncation coefficient (k) of the Truncated Gamma *pdf* that intervenes in variable selection. This allows the application of the Empirical Bayes procedure to estimate parameters and hyperparameters within an iterative coordinate descent algorithm, similar to the original RVM (Tipping, 2001; Wipf and Nagarajan, 2009). With this result, we avoid the Double Shrinkage Problem of the classical ENETL (Zou and Hastie, 2005; Vega-Hernández et al., 2008) and the use of computationally intensive MC/EM algorithms (Kyung et al., 2010; Li and Lin, 2010). Differently from the original RVM, here the hyperparameters (α_1, k) controls the global degree of sparsity, while γ acts in variable selection over voxels individually. A possible extension of our algorithm could be derived from assuming different values of the hyperparameters over individual voxels or groups of voxels (region of interest), to impose a variable degree of spatial sparsity.

Our Bayesian formulation of the ELASSO model (Kowalski and Torrésani, 2009b) pursues a similar constraint of spatial sparsity and temporal smoothness in the parameters matrix as previous works (Haufe et al., 2008; Ou et al., 2009; Gramfort et al., 2012). However, this model represents a non-separable Markov Random Field in the spatial dimension of parameters (Kindermann and Snell, 1980; Murphy, 2012) that can be reorganized into a Normal/Laplace hierarchical model at the parameters level. This is done through the introduction of a new hyperparameter (δ) which implicitly impose the spatial correlation structure among the parameters. This approximation allows to perform variable selection and learning of the hyperparameters by means of Empirical Bayes in ELASSO model, which has not been previously reported in the literature. Interestingly, this approximation might be related to others that has been used in the Bayesian context -such as those in Variational Bayes and Mean Field approximation (Friston et al., 2007; Trujillo-Barreto et al., 2008)—while it could also be seen as a local ENET approximation of the ELASSO penalty. This interesting relationship and its possible algorithmic relevance should be explored in future studies.

One interesting theoretical result in this work is that both ENET and ELASSO models can be expressed as particular cases of combined Normal/Laplace priors. Therefore, our algorithmic approach might be useful for many other models (including Group LASSO, Smooth LASSO, combinations of ENET and ELASSO, and others). Some of them have been tackled in the Bayesian Approach, but only using MC/EM algorithms.

Analysis of Simulations

We tested the performance of the proposed methods in a simulated scenario where activations were not completely sparse in space nor completely smooth in time, thus challenging the assumptions of the models and exploring their capacity to adapt in non-ideal conditions. We also use a completely different brain and head model (based on the Lead Field of different subjects) for simulating the data and reconstructing the PCD, avoiding the inverse crime in more unfavorable conditions than the reported before in the literature. Finally, we have added noise in both the level of sources and sensors, as have been suggested in the

TABLE 3 | Mean and standard deviation across trials of the quality measures Earth Movers Distance (EMD), Distance Localization Error (DLE), Area Under receiver operating Curve (AUC), and F1-Score (F1) at the time points t1 = 0 ms, t2 = 280 ms, and t3 = 375 ms for all studied methods.

		ENET-SSBL	ELASSO-SSBL	ENETL	LASSO Fusion	LORETA
(A) RESULTS ACROSS THE 100 TRIALS OF SIMULATION 1						
EMD	t1	176 ± 50.1	**170.6 ± 38**	249.2 ± 139.2	259.2 ± 131.2	458.3 ± 86.8
	t2	**124.1 ± 42.7**	125.2 ± 46.9	165.8 ± 137.9	170.3 ± 124	252.8 ± 84.4
	t3	143.6 ± 55.9	**142.2 ± 56.3**	231.1 ± 183.3	235.4 ± 158.2	360.3 ± 144.8
DLE	t1	1.77 ± 0.43	**1.48 ± 0.42**	2.34 ± 0.73	2.51 ± 0.63	4.74 ± 0.81
	t2	3.03 ± 0.86	**2.92 ± 0.93**	3.67 ± 1.34	3.83 ± 1.32	4.9 ± 1.06
	t3	2.09 ± 0.99	**1.9 ± 1.12**	3.37 ± 1.2	3.37 ± 1.14	4.33 ± 1.08
AUC	t1	93.9 ± 1	**94.1 ± 0.7**	92.2 ± 2.1	91.4 ± 2	84.6 ± 2.8
	t2	90.1 ± 2.9	**90.2 ± 3.5**	85.3 ± 8	85.4 ± 7.7	74.5 ± 9.2
	t3	93.7 ± 4.3	**93.9 ± 3.9**	88.3 ± 9.9	88 ± 10.1	78.8 ± 12
F1	t1	61.5 ± 10.2	**67.6 ± 7.3**	42.2 ± 14.3	37.5 ± 14.1	30.8 ± 13.5
	t2	37.8 ± 16.8	**40 ± 19.4**	33.5 ± 18.5	36 ± 16.7	31.3 ± 20.9
	t3	50.7 ± 20.1	**50.8 ± 25.1**	42.4 ± 14.7	42.9 ± 14.3	46.8 ± 10.1
(B) RESULTS ACROSS THE 100 TRIALS OF RANDOMIZED SIMULATION						
EMD	t1	158.2 ± 72.7	**146.5 ± 54.5**	240.6 ± 139.3	239.5 ± 142.4	369.8 ± 133.8
	t2	116 ± 58	**109.8 ± 52.1**	161.6 ± 98.4	160.4 ± 95.4	207.4 ± 105.1
	t3	105.6 ± 61.8	**104.8 ± 58.2**	161.9 ± 101.4	167.1 ± 102.1	181.8 ± 82.6
DLE	t1	2.72 ± 1.23	**2.53 ± 1.17**	3.87 ± 1.63	4.07 ± 1.65	5.07 ± 2.21
	t2	4.12 ± 1.86	**3.93 ± 1.47**	5.61 ± 2.03	5.44 ± 2.03	6.24 ± 2.29
	t3	**3.52 ± 1.82**	3.53 ± 1.86	5.51 ± 2.72	5.65 ± 2.72	5.55 ± 3.33
AUC	t1	93.2 ± 4.4	**93.5 ± 4.3**	92.4 ± 4	92.3 ± 4	86.8 ± 6.8
	t2	86.7 ± 10.2	**87.8 ± 7.2**	83.7 ± 10.6	84 ± 10.8	77.1 ± 11.9
	t3	**92.5 ± 5.6**	92.4 ± 6	89.6 ± 10	89.9 ± 10.4	85.2 ± 14.3
F1	t1	48.2 ± 15.4	**49.4 ± 15.2**	45.9 ± 15.8	45.7 ± 15.8	45.8 ± 13.2
	t2	37.5 ± 16	36.8 ± 16.8	38.7 ± 17	39.9 ± 16.2	**42.8 ± 14.2**
	t3	44.1 ± 18.5	44.5 ± 18	46.6 ± 13.1	45.4 ± 14.5	**48.4 ± 13.9**

Best results among algorithms (along rows) are highlighted in bold.

literature (Haufe et al., 2013; Haufe and Ewald, 2016). As there is no ground truth for the EEG IP, we cannot know if this option is a more realistic situation or more challenging than typical adding of only sensors white noise. Similarly, it is not easy to use real EEG background signals as only "noise" (Kobayashi et al., 2003; Stephen et al., 2003), since it might carry information about other active sources that can obscure the important results about the ability to recover controlled real sources. In any case, using both sources of noise is not modeled directly by any of the methods studied here, thus becoming an additional challenge for the reconstruction. We also here always reported results obtained from 100 repetitions with different noise in every spatio-temporal configuration, for avoiding bias in interpretation due to a random "good" choice of the simulated noise.

In the analysis of our simulations, on one hand, ENET-SSBL was able to recover spatially patch-wise smooth solutions (**Figures 4**, **6**), by accordingly tuning the degree of sparsity through the values of the hyperparameters, consistently with the theory (**Figure 9**). We also found a monotonous convergence pattern of the target likelihood function $\mathcal{L}$ for this solution in a single trial. The estimated values of hyperparameters were sensitive to the time-varying degree of sparsity, so that the values of the hyperparameters (α_1 and α_2) behaved consistently with the theory (**Figure 9**). They are higher when the degree of sparsity is higher. Noticeably, in 400 simulations (100 trials with SNR of 6dB for 4 different combinations of the three sources across hemispheres), the learning procedure of the proposed algorithm showed to be very robust, showing low standard deviation values in all time points. This is a very promising property -given that the EEG IP is also a very ill-conditioned problem- and it should thus be more thoroughly explored in future studies. On the other hand, the ELASSO-SSBL was also consistent with the theory, showing a monotonous convergence pattern. In **Figure 4**, the maps of ENET-SSBL and ELASSO-SSBL obtained for a typical trial (ranking in the median of the quality measures) are very similar, with the latter being

TABLE 4 | Statistical comparison of quality measures among all inverse methods, across all 100 trials at every time point for the 4 simulated configurations, using the non-parametric rank test.

	ENET-SSBL	ELASSO-SSBL	ENETL	LASSO Fusion	LORETA	ENET-SSBL	ELASSO-SSBL	ENETL	LASSO Fusion	LORETA
	EMD					**DLE**				
ENET-SSBL	–	2.1	69.8	69.7	90.2	–	1.5	87.1	88.9	97.5
ELASSO-SSBL	**49.8**	–	**77.61**	**79.9**	**92.2**	**39.7**	**–**	**90.4**	**91.7**	**99.3**
ENETL	2.6	0.1	–	0.1	85.3	1.9	0.4	–	4.6	86.7
LASSO Fusion	2.0	0.2	0.3	–	84.8	2.1	0.4	0	–	82.3
LORETA	1.4	0.1	0	0.1	–	0	0	0	0.124	–
	AUC					**F1-Score**				
ENET-SSBL	–	3.2	81.6	82.1	100	–	0	45.5	51.0	45.1
ELASSO-SSBL	**41.2**	–	**85.7**	**85.7**	**100**	**46.6**	**–**	**58.5**	**66.7**	**57.3**
ENETL	1.6	0.1	–	6.1	100	21.0	14.9	–	13.7	26.7
LASSO Fusion	1.1	0.2	0.2	–	100	14.2	9.0	0	–	19.2
LORETA	0	0	0	0	–	36.7	29.6	44.2	51.5	–

Each cell shows the percentage in which the method in the row significantly outperforms the method in the column across all 804 (201 × 4) cases. Highest percentages along each column are highlighted in bold to show easily show the best algorithm.

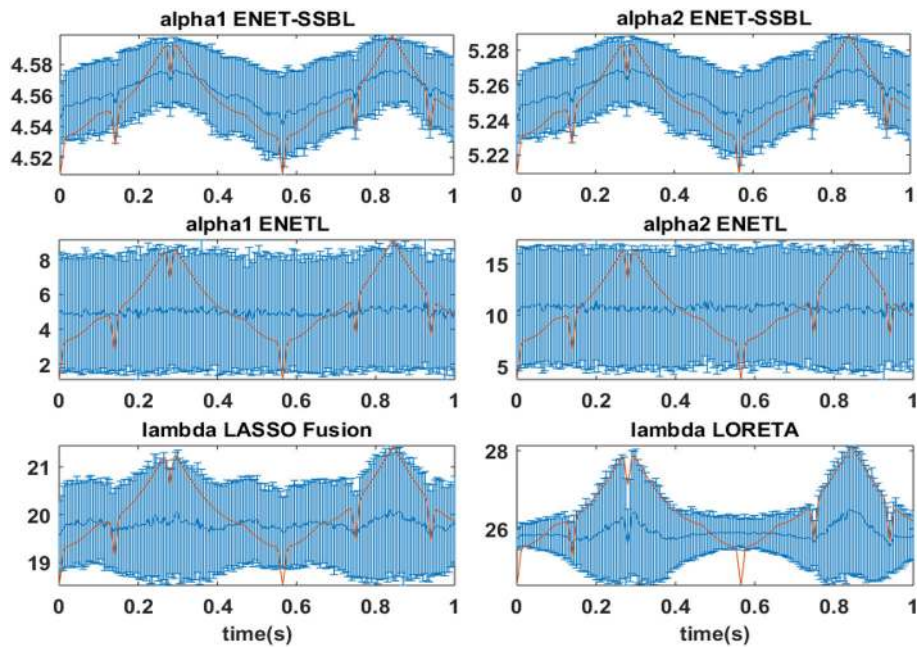

FIGURE 9 | Hyperparameters estimates with the different source localization methods for the four simulated configuration. Blue curves represent the mean temporal behavior and standard deviation (error bars) across 400 trials of the logarithmic values of estimated hyperparameters with the different methods. These values were found with the ENET-SSBL algorithms and GLMNET toolbox by generalized cross validation. The red trace corresponds to the normalized sparsity level along time, i.e., the percentage of non-active voxels in the simulations.

slightly sparser in some of the simulations. The average time courses of the reconstructed sources in voxels belonging to simulated patches were similar to the original simulated ones, but much noisier (**Figure 5**). The occipital source seemed to be easier to recover, while the motor one was mostly absent, and it was very difficult to estimate its amplitude with either method.

The proposed methods were also compared in average with those estimated with the known methods LORETA, ENETL and LASSO Fusion (Hunter and Li, 2005; Sánchez-Bornot et al., 2008; Vega-Hernández et al., 2008). This was illustrated for the first simulated configuration, which is challenging as all three sources are active in the same hemisphere (**Figure 6**). LORETA showed the smoothest average maps, also with higher variance

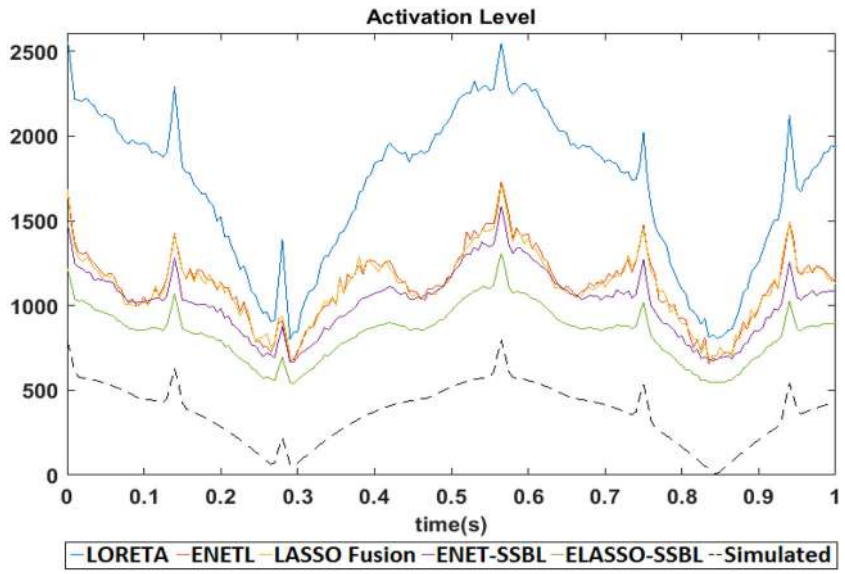

FIGURE 10 | Temporal behavior of the average activation level (i.e., L1-norm) of simulated (dashed) and estimated solutions with the different methods. These curves are obtained from averaging all 100 trials of Simulation 1. We use the L1-norm as a measure of activation (since sparsity is related to its minimization), therefore those curves better resembling the curve of the simulated solution can be interpreted to represent solutions with a more similar level of sparsity to the sparsity of true sources.

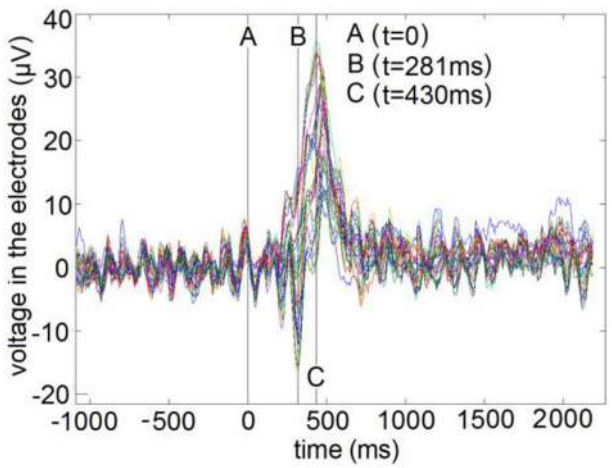

FIGURE 11 | Time series of the Visual Evoked Potential for all electrodes. A, Reference time point ($t = 0$ ms) of stimulus onset. B, Global negative maximum potential ($t = 281$ ms). C, Global positive maximum potential ($t = 430$ ms).

in all regions not involving real activations, which suggests that spurious sources are regularly shown by this solution. In the case of ENETL and LASSO Fusion, the solutions were able to recover the large patches (O and M) but missed the smaller patch "TL" in the left hemisphere. ENET-SSBL and ELASSO-SSBL did recover all sources better, although showing a higher variance except in the sparser situation occurring at *t2*. The non-normalized average time courses for the voxels belonging to the three simulated patches (**Figure 7**) showed again that the amplitude of the occipital patch (the largest in extension) is easier to recover, while the other two smaller patches are systematically underestimated by all methods. As expected, LORETA was the solution that offered the more biased estimation, although with the lowest variance in all patches.

A more general comparison in performance of the solutions was done using the quality measures in **Figure 8**, showing that ENET-SSBL and ELASSO-SSBL had smaller distances (EMD and DLE) and higher performance measures (AUC and F1-score) than ENETL, LASSO Fusion and LORETA. Quantitatively, ELASSO-SSBL and ENET-SSBL have similar results, but they both outperforms LORETA, ENETL, and LASSO Fusion, consistently along all different sparsity scenarios (**Table 3**). ELASSO-SSBL gives the best results for all time points in almost all measures (highlighted in bold in **Table 3A**). In the case of randomized simulations (**Table 3B**), LORETA showed the highest F1 score for those time points where only one or two sources were active with small amplitude. As these values are computed using the optimal threshold for each solution, this can be due to very low optimal thresholds that allow a very high recall for LORETA thanks to its smooth behavior (Bradley et al., 2016). However, these differences are not likely to be significant since the standard deviation of F1-score is very high. The significant non-parametric comparisons between all pair of methods according to all quality measures (**Table 4**) showed that ELASSO-SSBL is better than all other regularization methods in more than the 50% of all simulation cases. Particularly, DLE and AUC values of ELASSO-SSBL are significantly better than those of the other regularization methods (ENETL, LASSO Fusion, LORETA) in more than 85%, while it is significantly worst in less than 1% of the cases.

If we compare the two proposed models, we will find that in general ELASSO-SSBL performed slightly better than

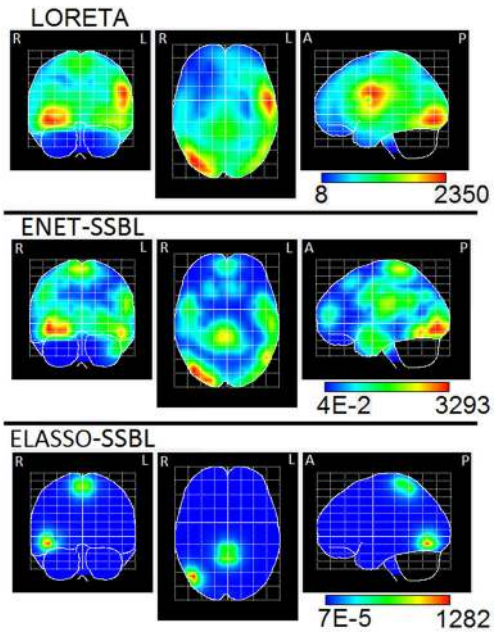

FIGURE 12 | Maximum intensity projection of the PCD estimated with LORETA, ENET-SSBL, and ELASSO-SSBL, at the global negative maximum potential (time B in **Figure 9**). The three orthogonal planes are the coronal (left), axial (center) and sagittal (right) views. R, L, A, P stand for Right, Left, Anterior, Posterior, respectively.

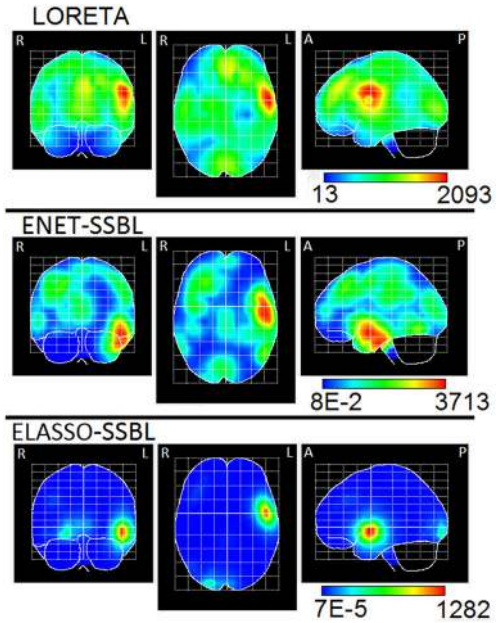

FIGURE 13 | Maximum intensity projection of the PCD estimated with LORETA, ENET-SSBL, and ELASSO-SSBL, at the global positive maximum potential (time C in **Figure 9**). The three orthogonal planes are the coronal (left), axial (center) and sagittal (right) views. R, L, A, P stand for Right, Left, Anterior, Posterior, respectively.

ENET-SSBL (**Figure 8** and **Tables 2**, **3**), which might be due to a better recover of the spatial extension of the simulated patches. However, this difference is significant in about 40% of the simulation cases (**Table 4**), which means that there are many scenarios where they give similar results. This might be interpreted as a higher performance of the spatio-temporal assumptions in ELASSO-SSBL, which is able to adapt to an adequate degree of sparsity by means of incorporating information of the spatial patches across all time points into the estimates of the regularization parameter.

Finally, computation times were similar in ENET-SSBL and ELASSO-SSBL, but the latter is more memory demanding because the estimation of the hyperparameter depends on parameters of the whole spatio-temporal map. Convergence was reached around 30 iterations of explicit estimators for a computational complexity of $O\left(N^2\right)$ (N is the number of electrodes). Although we did not use the MCMC algorithm for our simulations, other studies in the literature report that these methods take about 10 or 100 thousand iterations to converge, (Jun et al., 2005, 2008; Nummenmaa et al., 2007; Li and Lin, 2010). The cost of MCMC in the EEG Inverse Problem is determined by the Gibbs Sampler of the parameters posterior distribution, that amounts to $O\left(S^2\right)$ (S is the number of spatial generators).

The contrast between the superiority of ELASSO-SSBL in these general results (**Table 4**) and the less evident superiority in previous figures can be explained by the use of more challenging situations and time points for presenting the first results. That, together with the use of a completely different ELF to avoid the "inverse crime," the use of both physiological and sensors noise sources and of simulations that do not follow the assumptions of our models, would suggest that the Empirical Bayes algorithm for the models proposed here is very robust to many of the typical factors that influence practical ESI, thus becoming a promising method to realistic experimental situations.

This robustness might be closely related to the efficacy in adequately estimating the hyperparameters of the model, which controls the levels of spatial sparsity and temporal smoothness. **Figure 9** shows that the hyperparameters learned by ENET-SSBL closely follows the level of sparsity of the simulated solutions. The estimated regularization parameters with LORETA (**Figure 9**) also showed some correlation with the level of sparsity, especially for its variance, but smaller than the correlation showed by ENET-SSBL. Regularization parameters for ENETL and LASSO Fusion did not show sensitivity to the changes in sparsity, which could be explained by a wrong selection of the appropriate range for these parameters in the heuristic GCV procedure used for its estimation. Importantly, all methods showed in average a very similar pattern of the level of activation of their solutions to that of the simulated sources along time points (**Figure 10**). This would contradict the logical suggestion that a correct learning of the hyperparameters (i.e., following the sparsity of real sources) leads to estimate solutions with similar sparsity. In this case, we found that the level of activation for ENETL, LASSO Fusion and ENET-SSBL were similar in average. However, ELASSO-SSBL showed the smallest L1-norm of activated sources (highest degree

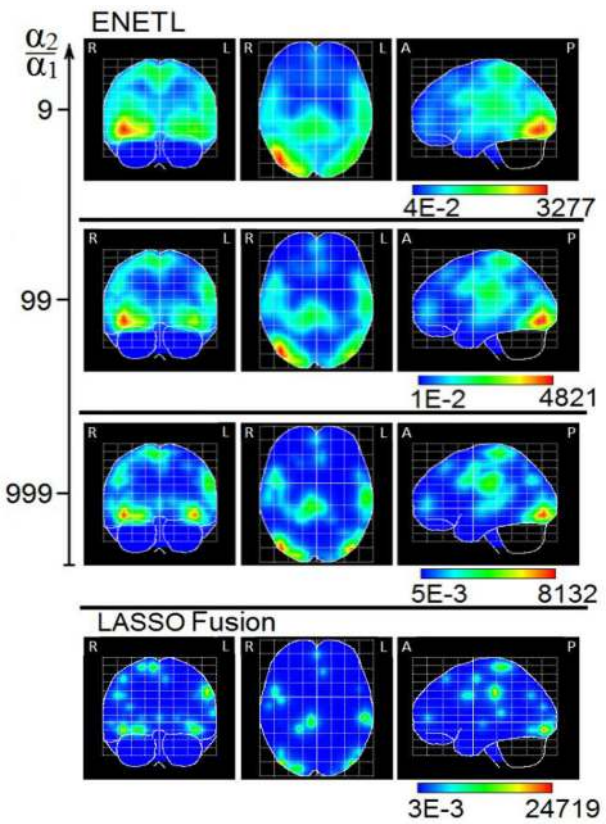

FIGURE 14 | Maximum intensity projection of the PCD estimated with ENETL (using different ratio between regularization parameters) and LASSO Fusion at the global negative maximum potential (time B in **Figure 9**). The three orthogonal planes are the coronal (left), axial (center) and sagittal (right) views. R, L, A, P stand for Right, Left, Anterior, Posterior, respectively. This figure is a reproduction of results shown in **Figure 3** in Vega-Hernández et al. (2008).

of sparsity) along all-time points and it was also the closest to the real level of sparsity of the true simulation, resembling its shape along all time points. This result is very interesting as this method uses only one hyperparameter to control the level of sparsity for all time points simultaneously. As expected, LORETA showed the lowest sparsity along space/time, which usually conveys the cost of introducing spurious sources.

Analysis of Real Data

In the analysis of real data, the maximum of the estimated solution with the three methods (LORETA, ENET-SSBL and ELASSO-SSBL) in the negative peak (time B in **Figure 11**), is located in the occipital area of the right hemisphere (Brodmann area 19), that corresponds to the visual cortex (**Figure 12**). As expected, LORETA is the most disperse solution, such that secondary activities appear not only in occipital areas but also in the superior temporal cortex. ENET-SSBL showed similar solutions than LORETA, but less disperse, such that the secondary activities in temporal, frontal and centro-parietal areas can be distinguished among them. Contrary to LORETA, the secondary activation in the temporal cortex was not the most intense, but the centro-parietal one in the motor cortex (Brodmann area 4). The solution with the ELASSO-SSBL method was the sparsest, showing only the main source in the right visual area and a secondary centro-parietal activation (motor cortex). Results from ENET-SSBL and ELASSO-SSBL are more consistent with the neurophysiology of visual attention where pre-motor activations in the occipital visual areas explain the processing of visual information, while secondary activations in the motor cortex are also expected related to the physical response (Hillyard and Anllo-Vento, 1998; Di Russo et al., 2001).

A great similarity between ENET-SSBL and ENETL solutions can be seen in the case when the ratio between regularization parameters in the latter is 99, while the hyperparameters are learned in the former (see **Figures 12**, **14**). Although preliminary, this suggests that the learning strategy might be an effective way to find optimal intermediate levels of sparsity in real scenarios. It would also be very useful since learning the hyperparameter would offer a direct and justified way of estimating its optimal value without the need of using heuristic methods and heavy statistical post-processing as in Vega-Hernández et al. (2008). This is especially important in the case of ENET, since it contains two regularization parameters leading to a bi-dimensional search space and the corresponding high computational burden.

On the other hand, the ELASSO-SSBL solution distinguishes from the very sparse but scattered LASSO Fusion solution (**Figure 14**). Encouragingly, this suggests that with ELASSO-SSBL the degree of sparsity can be adapted to show a few small activated regions instead of many point sources, which is indeed more realistic in cognitive experiments.

The inverse solutions in the later positive peak (time C in **Figure 11**) showed similar patterns as those computed in the negative peak with occipital and parieto-temporal activations (**Figure 13**). Again, this is in agreement with research on the neurological foundations of the visual evoked potentials, showing that late positive potentials can be related to processing of other cognitive aspects of visual information and attention, as well as post-motor responses, whose sources are likely to be found in the superior anterior temporal lobe (Makeig et al., 1999, 2004; Bonner and Price, 2013). However, in this case the three methods did not agree in the location of the main source. Yet very spread, LORETA solution showed the maximum activity in the superior temporal cortex (Brodmann area 43) and secondary activations in occipital, frontal and parietal areas, being difficult to decide which is relevantly separated from the main activation. Differently, ENET-SSBL and ELASSO-SSBL found the maximum activity in the anterior temporal lobe (Brodmann area 21). ENET-SSBL showed again a similar solution as LORETA with a better trade-off between sparsity and smoothness, which would allow to identify different sources. The ELASSO-SSBL solution is the sparsest again, clearly separating two PCD sources. The maximum activity was located in the superior gyrus of the anterior temporal lobe and the secondary activation in the occipital area. Again, sources were not scattered focal activations but well-localized smooth patches of activations. It seems that the ELASSO-SSBL applied to just one time instant (i.e., without temporal information) does not behave simply like the LASSO

model (Laplace prior), but it is able to find intermediate levels of spatial group sparsity. More comprehensive studies are needed to completely characterize the performance of this model with the proposed sparse Bayesian Learning algorithm.

CONCLUSIONS

In this work we introduced a Bayesian formulation of Structured Sparsity regularization models combining L1/L2 norm constraints, for solving the EEG IP. In particular, we developed ENET and ELASSO models that have been previously addressed with the classical statistical framework, which presents practical limitations for selecting optimal values for one or more regularization parameters that are critical for correctly estimate solutions. We have used the Empirical Bayes approach for deriving a Sparse Bayesian Learning algorithm that allows both the estimation of parameters and the learning of hyperparameters by means of iterative algorithms. Using simulations, we found that our methods ENET-SSBL and ELASSO-SSBL are able to recover complex sources more accurately and with a more robust behavior under different sparsity scenarios than the classical ENETL, LASSO-Fusion and LORETA inverse solutions. Importantly, we have used a large set of simulations that do not cope with the theoretical assumptions of the models, and explore different sparsity scenarios. The quality measures used in this work were also found from the non-thresholded solutions to avoid the influence of arbitrary post-processing. Proper methods for a statistically founded thresholding and a more careful evaluation of the Bayesian estimator will be the subject of future developments with these models. In a real EEG study with a visual attention experiment, our methods localized electrical sources of early negative and late positive components of the Visual Evoked Potential that are more interpretable from the neurophysiological point of view, as compared with other known methods such as LORETA, ENETL and LASSO Fusion. Theoretically, the principles behind the proposed algorithms can be applied to other models based on combined Normal/Laplace priors and other types of inverse problems such as image recovery. Other possible extensions to deal with the vector PCD field (i.e., estimating the three spatial components in each voxel) or with other multidimensional data, such as time-frequency and space-time-frequency EEG tensors, should also be considered as new lines of research.

ETHICS STATEMENT

The data used in this paper was taken from Makeig et al. (1999). It is available at the website http://sccn.ucsd.edu/eeglab/download/eeglab_data.set

AUTHOR CONTRIBUTIONS

DP-L: Mathematical formulation of the Elastic Net and Elitist Lasso probabilistic models. Empirical Bayes formulation and inference formulas. Derivation and implementation of the algorithms ENET-SSBL and ELASSO-SSBL. Simulations and Real Data study. Analysis and interpretation of the results. Writing the paper. MV-H: Review, mathematical analysis and implementation of the Elastic Net in the context of Tikhonov Regularization. Review and mathematical analysis of the Bayesian Elastic Net. PR-L: Computational issues of the ENET-SSBL and ELASSO-SSBL algorithms. PV-H: Building the substrate for specific subject EEG forward model, quantitative analysis of inverse solutions. PV-S: Introduction in our work the theory MPLS models and Sparsity Constraints. Advising in the theory of the EEG Inverse Problem and software platforms. Advising in the mathematical issues of the Elastic Net as an MPLS model in the EEG Inverse Problem. Providing Simulations and Real Data. Analysis and Discussion of the Results. EM-M: Review on the EEG IP theory. Advising on Sparsity Regularization techniques in the context of Tikhonov Regularization and Sparse Bayesian Learning. Guide in the mathematical issues of Bayesian models and Empirical Bayes. Proposal of the Bayesian Analysis of the Elitits Lasso. Guide in developing the algorithms ENET-SSBL and ELASSO-SSBL. Analysis of the Results.

ACKNOWLEDGMENTS

This study was funded by the Grant No. 61673090 from the National Nature Science Foundation of China. Authors would like to especially acknowledge the collaboration in this work from Lidice Galán-García, Alberto Taboada-Crispi, Esin Karahan, Eduardo Gonzalez-Moreira and Marlis Ontivero-Ortega.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2017.00635/full#supplementary-material

REFERENCES

Andrews, D. F., and Mallows, C. L. (1974). Scale mixtures of normal distributions. *J. R. Stat. Soc. Ser. B* 36, 99–102.

Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2010). Bayesian compressive sensing using laplace priors. *IEEE Trans. Image process.* 19, 53–63. doi: 10.1109/TIP.2009.2032894

Beck, A., and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2, 183–202. doi: 10.1137/080716542

Bonner, M. F., and Price, A. R. (2013). Where is the anterior temporal lobe and what does it do? *J. Neurosci.* 33, 4213–4215. doi: 10.1523/JNEUROSCI.0041-13.2013

Bradley, A., Yao, J., Dewald, J., and Richter, C. P. (2016). Evaluation of electroencephalography source localization algorithms with multiple cortical sources. *PLoS ONE* 11:e0147266. doi: 10.1371/journal.pone.0147266

Di Russo, F., Martínez, A., Sereno, I. M., Pitzalis, S., and Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.* 15, 95–111. doi: 10.1002/hbm.10010

Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96, 1348–1360. doi: 10.1198/016214501753382273

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., et al. (2008). Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39, 1104–1120. doi: 10.1016/j.neuroimage.2007.09.048

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. *Neuroimage* 16, 465–483. doi: 10.1006/nimg.2002.1090

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234. doi: 10.1016/j.neuroimage.2006.08.035

Gibbons, J. D., and Chakraborti, S. (2011). "Nonparametric statistical inference," in *International Encyclopedia of Statistical Science*, ed M. Lovric (Berlin; Heidelberg: Springer), 977–979.

Gramfort, A., Kowalski, M., and Hämaläinen, M. (2012). Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* 57, 1937–1961. doi: 10.1088/0031-9155/57/7/1937

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M. S., and Kowalski, M. (2013). Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations. *Neuroimage* 70, 410–422. doi: 10.1016/j.neuroimage.2012.12.051

Grova, C., Daunizeau, J., Lina, J. M., Bénar, C. G., Benali, H., and 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

Hadamard, J. (1923). *Lecture on the Cauchy Problem in Linear Partial Differential Equations*. New Haven, CT: Yale University Press.

Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comp.* 32, 35–42. doi: 10.1007/BF02512476

Hansen, S. T. (2017). *EEG Based Inference of Spatio-Temporal Brain Dynamics.* Doctoral dissertation, Technical University of Denmark (DTU).

Haufe, S., and Ewald, A. (2016). A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain Topogr.* 1–18. doi: 10.1007/s10548-016-0498-y

Haufe, S., Nikulin, V. V., Müller, K. R., and Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: a simulation study. *Neuroimage* 64, 120–133. doi: 10.1016/j.neuroimage.2012.09.036

Haufe, S., Nikulin, V. V., Ziehe, A., Müller, K. R., and Nolte, G. (2008). Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage* 42, 726–738. doi: 10.1016/j.neuroimage.2008.04.246

Hebiri, M. (2008). Regularization with the smooth-lasso procedure. *arXiv preprint* arXiv 2008; 0803.0668.

Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 95, 781–787. doi: 10.1073/pnas.95.3.781

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67. doi: 10.1080/00401706.1970.10488634

Hunter, D. R., and Li, R. (2005). Variable selection using MM algorithms. *Ann. Stat.* 33, 1617–1642. doi: 10.1214/009053605000000200

Jun, S. C., George, J. S., Kim, W., Paré-Blagoev, J., Plis, S., Ranken, D. M., et al. (2008). Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC. *Neuroimage* 40, 1581–1594. doi: 10.1016/j.neuroimage.2007.12.029

Jun, S. C., George, J. S., Paré-Blagoev, J., Plis, S. M., Ranken, D. M., Schmidt, D. M., et al. (2005). Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data. *Neuroimage* 28, 84–98. doi: 10.1016/j.neuroimage.2005.06.003

Kaipio, J. P., and Somersalo, E. (2004). *Statistical and Computational Invese Problems*. New York, NY: Springer.

Kindermann, R., and Snell, J. L. (1980). Markov random fields and their applications. *Am. Math. Soc. Contemp. Math.* 1, 271–4132. doi: 10.1090/conm/001

Klem, G. H., Luders, H. O., Jasper, H. H., and Elger, C. (1999). Recommendations for the practice of clinical neurophysiology. Guidelines of the international federation of clinical physiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 1–304.

Kobayashi, K., Yoshinaga, H., Oka, M., Ohtsuka, Y., and Gotman, J. (2003). A simulation study of the error in dipole source localization for EEG spikes with a realistic head model. *Clin. Neurophysiol.* 114, 1069–1078. doi: 10.1016/S1388-2457(03)00064-6

Kowalski, M., and Torrésani, B. (2009a). Sparsity and persistence: mixed norms provide simple signals models with dependent coefficients. *Signal Image Video Process.* 3, 251–264. doi: 10.1007/s11760-008-0076-1

Kowalski, M., and Torrésani, B. (2009b). "Structured sparsity: from mixed norms to structured shrinkage," in *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations* (Saint Malo).

Kyung, M., Gilly, J., Ghoshz, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 5, 369–412. doi: 10.1214/10-BA607

Land, S., and Friedman, J. (1996). *Variable Fusion: A New Method of Adaptive Signal Regression,* Technical Report. Stanford, CA: Department of Statistics, Stanford University.

Li, Q., and Lin, N. (2010). The Bayesian elastic net. *Bayesian Anal.* 5, 151–170. doi: 10.1214/10-BA506

Liu, H., Schimpf, P. H., Dong, G., Gao, X., Yang, F., and Gao, S. (2005). Standardized shrinking LORETA-FOCUSS (SSLOFO): a new algorithm for spatio-temporal EEG source reconstruction. *Biomed. Eng. IEEE Trans.* 52, 1681–1691. doi: 10.1109/TBME.2005.855720

MacKay, D. J. (1992). Bayesian interpolation. *Neural Comput.* 4, 415–447. doi: 10.1162/neco.1992.4.3.415

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Magnus, J. R., and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, UK: Wiley.

Makeig, S. (2002). Response: event-related brain dynamics – unifying brain electrophysiology. *Trends Neurosci.* 25:390. doi: 10.1016/S0166-2236(02)02198-7

Makeig, S., Delorme, A., Westerfield, M., Jung, T. P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol.* 2, 747–762. doi: 10.1371/journal.pbio.0020176

Makeig, S., Westerfield, M., Jung, T. P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *J. Neurosci.* 19, 2665–2680.

Molins, A., Stufflebeam, S. M., Brown, E. N., and Hämäläinen, M. S. (2008). Quantification of the benefit from integrating MEG and EEG data in minimum *l*2-norm estimation. *Neuroimage* 42, 1069–1077. doi: 10.1016/j.neuroimage.2008.05.064

Murphy, K. P. (2012). *Machine Learning a Probabilistic Perspective*. Cambridge: MIT Press.

Nagarajan, S. S., Portniaguine, O., Hwang, D., Johnson, C., and Sekihara, K. (2006). Controlled support MEG imaging. *NeuroImage* 33, 878–885. doi: 10.1016/j.neuroimage.2006.07.023

Nummenmaa, A., Auranen, T., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Sams, M., et al. (2007). Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *Neuroimage* 35, 669–685. doi: 10.1016/j.neuroimage.2006.05.001

Ou, W., Hämäläinen, M., and Golland, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* 44, 932–946. doi: 10.1016/j.neuroimage.2008.05.063

Palmero-Soler, E., Dolan, K., Hadamschek, V., and Tass, P. A. (2007). swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys. Med. Biol.* 52, 1783–1800. doi: 10.1088/0031-9155/52/7/002

Park, T., and Casella, G. (1999). The Bayesian Lasso. *J. Am. Stat. Assoc.* 103, 681–686.

Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *Int. J. Biol.* 1, 75–86.

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.

Qian, J., Hastie, T., Friedman, J., Tibshirani, R., and Simon, N. (2013). *Glmnet for Matlab*. Available online at: https://web.stanford.edu/~hastie/glmnet_matlab/

Reed, W. J. (2006). "The normal-Laplace distribution and its relatives," in *Advances in Distribution Theory, Order Statistics, and Inference*. eds N. Balakrishnan, J. M. Sarabia, and E. Castillo (Boston, MA: Birkhäuser), 61–74.

Riera, J. J., and Fuentes, M. E. (1998). Electric lead field for a piecewise homogeneous volume conductor model of the head. *Biomed. Eng. IEEE Trans.* 45, 746–753.

Sánchez-Bornot, J. M., Martínez-Montes, E., Lage-Castellanos, A., Vega-Hernández, M., and Valdés-Sosa, P. A. (2008). Uncovering sparse brain effective connectivity: a voxel-based approach using penalized regression. *Stat. Sin.* 8, 1501–1518.

Schmidt, D. M., George, J. S., and Wood, C. C. (1999). Bayesian inference applied to the biomagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212. doi: 10.1002/(SICI)1097-0193(1999)7:3<195::AID-HBM4>3.0.CO;2-F

Schmolck, A., and Everson, R. (2007). Smooth relevance vector machine: a smoothness prior extension of the RVM. *Mach. Learn.* 68, 107–135. doi: 10.1007/s10994-007-5012-z

Stephen, J. M., Aine, C. J., Ranken, D., Hudson, D., and Shih, J. J. (2003). Multidipole analysis of simulated epileptic spikes with real background activity. *J. Clini. Neurophysiol.* 20, 1–16. doi: 10.1097/00004691-200302000-00001

Tian, T. S., and Li, Z. (2011). A spatio-temporal solution for the EEG/MEG inverse problem using group penalization methods. *Stat. Interface* 4, 521–533. doi: 10.4310/SII.2011.v4.n4.a10

Tibshirani, R. (1996). Regression and Shrinkage via the LASSO. *J. R. Stat. Soc. Ser. B* 58, 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67, 91–108. doi: 10.1111/j.1467-9868.2005.00490.x

Tikhonov, A. N., and Arsenin, V. Y. (1977). *Solution to Ill Posed Problems.* Washington, DC: V. H. Winston, OR

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* 1, 211–244.

Tipping, M. E., and Faul, A. C. (2003). "Fast marginal likelihood maximization for sparse bayesian models," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics* (Key West, FL), 3–6.

Trujillo-Barreto, N. J., Aubert-Vázquez, E., and Penny, W. D. (2008). Bayesian M/EEG source reconstruction with spatio-temporal priors. *Neuroimage* 39, 318–335. doi: 10.1016/j.neuroimage.2007.07.062

Trujillo-Barreto, N. J., Aubert-Vázquez, E., and Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *Neuroimage* 21, 1300–1319. doi: 10.1016/j.neuroimage.2003.11.008

Valdés-Hernández, P. A., von Ellenrieder, N., Ojeda-Gonzalez, A., Kochen, S., Alemán-Gómez, Y., Muravchik, C., et al. (2009). Approximate average head models for EEG source imaging. *J. Neurosci. Methods* 185, 125–132. doi: 10.1016/j.jneumeth.2009.09.005

Valdés-Sosa, P. A., Bornot-Sánchez, J. M., Vega-Hernández, M., Melie-García, L., Lage-Castellanos, A., and Canales-Rodrìguez, E. (2006). "Granger causality on spatial manifolds: applications to neuroimaging," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, eds B. Schelter, M. Winterhalder, and J. Timmer (Weinheim: Wiley), 514.

Valdés-Sosa, P. A., Vega-Hernández, M., Sánchez-Bornot, J. M., Martínez-Montes, E., and Bobes, M. A. (2009). EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *Hum. Brain Mapp.* 30, 1898–1910. doi: 10.1002/hbm.20784

Vega-Hernández, M., Martínez-Montez, E., Sánchez-Bornot, J. M., Lage-Castellanos, A., and Valdés-Sosa, P. A. (2008). Penalized least squares methods for solving the EEG inverse problem. *Stat. Sin.* 18, 1535–1551.

Wipf, D., and Nagarajan, S. (2009). A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage* 44, 974–966. doi: 10.1016/j.neuroimage.2008.02.059

Wipf, D. P., Ramirez, R. R., Palmer, J. A., Makeig, S., and Rao, B. D. (2006). *Automatic Relevance Determination for Source Localization with MEG and EEG Data*, Technical Report, San Diego, CA: University of California.

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MM and handling Editor declared their shared affiliation.

Copyright © 2017 Paz-Linares, Vega-Hernández, Rojas-López, Valdés-Hernández, Martínez-Montes and Valdés-Sosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.