Electromagnetic Source Imaging via Bayesian Modeling With Smoothness in Spatial and Temporal Domains

Jiwen Liang, Zhu Liang Yu, Member, IEEE, Zhenghui Gu, Member, IEEE, and Yuanqing Li, Fellow, IEEE

Abstract—Accurate reconstruction of cortical activation from electroencephalography and magnetoencephalography (EEG/MEG) is a long-standing challenge because of the inherently ill-posed inverse problem. In this paper, a novel algorithm under the empirical Bayesian framework, source imaging with smoothness in spatial and temporal domains (SI-SST), is proposed to address this issue. In SI-SST, current sources are decomposed into the product of spatial smoothing kernel, sparseness encoding coefficients, and temporal basis functions (TBFs). Further smoothness is integrated in the temporal domain with the employment of an underlying autoregressive model. Because sparseness encoding coefficients are constructed depending on overlapped clusters over cortex in this model, we derived a novel update rule based on fixed-point criterion instead of the convexity based approach which becomes invalid in this scenario. Entire variables and hyper parameters are updated alternatively in the variational inference procedure. SI-SST was assessed by multiple metrics with both simulated and experimental datasets. In practice, SI-SST had the superior reconstruction performance in both spatial extents and temporal profiles compared to the benchmarks.

Index Terms—E/MEG source imaging, spatial-temporal smoothness, variational inference, empirical Bayesian.

I. INTRODUCTION

Electroencephalography (EEG) and magnetoencephalography (MEG) are widely used noninvasive techniques to explore human brain function with millisecond temporal resolution. Reconstruction of the cortical activation from the E/MEG recordings has a great importance in the field of neuroscience research and clinical applications [1], [2]. The problem mentioned above can be addressed by solving the forward and inverse models successively. The inverse problem aims at estimating current sources on cortical surface based on E/MEG recordings and the lead field matrix calculated in forward problem. A wide variety of source imaging algorithms for the inverse problem can be categorized into dipole fitting and distributed source imaging. In dipole fitting, it is essential to estimate the brain activity with a few equivalent current dipoles (ECDs) [3], [4]. Because of the preference for acquiring sparse results, dipole fitting methods would obtain poor imaging performance when sources are activated with sophisticated configuration.

In distributed source imaging models, the sources are located on voxels discretized over the cortical surface. Unfortunately, it involves solving an inherently ill-posed linear inverse problem since the source space dimension (∼O(10⁴)) is always enormously larger than the sensor space’s (∼O(10²)). Therefore, prior assumptions are essential to acquire meaningful estimates in the neurophysiology context. In the minimum-norm estimate (MNE) [5], the introduced L₂-norm penalty of current sources is interpreted as underlying energies which are expected as small as possible. Low resolution brain electromagnetic tomography (LORETA) [6] and standard LORETA [7] are variants of MNE as well, in which the spatial smoothness are integrated through discrete Laplacian operator. The aforementioned algorithms can be framed into the Bayesian scheme with the L₂-norm penalty interpreted as the negative log likelihood. Moreover, under empirical Bayesian framework [8], which is also referred to as sparse Bayesian learning (SBL), hyper parameters introduced in the regularization constraints can be updated with other coefficients simultaneously. Champagne [9] is one of the representative SBL algorithms which is constructed with relaxed orientation for each voxel. It outperforms several benchmarks in various source configurations.

Spatial smoothing kernel is widely used for the recovery of extended neural activation, with which current sources are converted into a sparse latent space. In Smooth Champagne [10], smoothing kernel is constructed directly based on the distances between voxels. On the contrary, smoothing kernel in BEStES [11] is formulated implicitly with the Markov Random Field (MRF) model [12]. Additionally, spatial Gaussian functions [13] and cortical patch bases [14] have been proposed with the terminology of spatial basis functions (SBFs), which play the equivalent role with smoothing kernel.
The source imaging algorithms mentioned above emphasize the spatial constraints but pay little attention to correlations in the temporal domain. Consideration of these correlations benefits to both time course reconstruction and source localization thanks to the intrinsic properties of the neural activation. It is convenient and efficient to integrate time-related information with the usage of temporal basis functions (TBFs). Current sources are restricted in a subspace spanned by the TBFs with much less dimensions than the number of time points. TBFs can be either predefined, such as wavelet dictionaries [15] and Gabor atoms [16], or estimated in data-driven manners like singular value decomposition (SVD) [17] and stimulus-evoked source imaging based on spatio-temporal basis function (SVD) [18]. Furthermore, to obviate the incompatibility between chosen TBFs and subsequent algorithms, source imaging based on spatio-temporal basis function (SI-STBF) [19] has been proposed where TBFs are allowed to be updated with other coefficients simultaneously under the empirical Bayesian framework.

Motivated by the prior studies, we propose a novel source imaging algorithm based on a Bayesian probabilistic model where the smoothness is integrated in both spatial and temporal domains. Entire hyper parameters can be updated in a data-driven fashion. The main contributions and creative points of this paper are twofold. First, further temporal smoothness is enforced via an underlying autoregressive model. TBFs are represented as the sum of self average at previous time points and residual error variables. The number of TBFs can be adaptively determined by pruning irrelevant components under the empirical Bayesian framework. Second, we propose a new update rule based on fixed-point criterion for scale parameters in sparseness encoding coefficients. The novel update manner maintains applicable when the typically used convexity based approach becomes invalid.

The rest of this paper is organized as follows. In Section II, we derive the probabilistic Bayesian model and illustrate the corresponding algorithm in detail. In Section III, we expound the simulation protocol and performance metrics. Section IV demonstrates the source imaging results with the comparison to benchmarks for both simulated and experimental E/MEG datasets, followed by some brief discussion in Section V and conclusions in Section VI.

For notation, upper and lower case bold face letters denote matrices and vectors respectively. Scalars are represented by roman lowercase variables. $\| \cdot \|_F$ denotes the Frobenius norm of a matrix. $x_i$ is the $i$th column of matrix $X$. $I_m$ indicates the identity matrix and $I_m$ denotes a $m \times 1$ vector with all elements being unit. $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian distribution of $x$ with the mean $\mu$ and covariance $\Sigma$.

II. METHODS

A. Probabilistic Generative Model

With the help of Maxwell’s equation [20], the E/MEG recordings can be approximately expressed as the linear mapping of current sources

$$B = LS + \varepsilon$$  \hspace{1cm} (1)

where $B = [b_1, b_2, \ldots, b_T] \in \mathbb{R}^{P \times T}$ is the E/MEG measurement with $P$ sensors and $T$ time points, $b_t$ represents a snapshot of measured signal at time point $t$. $S \in \mathbb{R}^{D \times T}$ denotes the current sources and $D$ is the number of discretized voxels over cortical surface. $L \in \mathbb{R}^{P \times D}$ is the lead field matrix whose columns reflect the sensors response induced by the unit current in a certain candidate location. The orientation of each source is restricted perpendicularly to the cortical surface. $\varepsilon \in \mathbb{R}^{P \times T}$ indicates the measurement noise with independently and identically Gaussian distribution $\mathcal{N}(\varepsilon_t | 0, \mathbf{I})$ at each time point. The model can be easily whitened via SVD of the covariance as $\Sigma_\varepsilon = U \Lambda U^T$. Multiplying by $\Lambda^{-1/2} U^T$ from the left, (1) is converted into $\tilde{B} = \tilde{L} S + \tilde{\varepsilon}$, where $\tilde{B} = \Lambda^{-1/2} U^T B$, $\tilde{L} = \Lambda^{-1/2} U^T L$ and $\tilde{\varepsilon}_t \sim \mathcal{N}(\tilde{\varepsilon}_t | 0, \mathbf{I})$. Therefore, we assume the model has been whitened and the tildes are dropped in the remainder of this paper without loss of generality.

1) Temporal Modeling: In SI-STBF [19], current sources $S$ are constructed as the linear combination of $K$ learnable temporal basis functions (TBFs)

$$S = W \Phi = \sum_{k=1}^{K} w_k \phi_k^T$$  \hspace{1cm} (2)

where $\Phi = [\phi_1, \phi_2, \ldots, \phi_K]^T \in \mathbb{R}^{K \times T}$, $\phi_k$ indicates the $k$th TBF with $T$ time points. $W$ is composed of weighting coefficients vector $w_k$ correlated to the $k$th TBF. Both $\Phi$ and $W$ are updated during the optimization procedure in SI-STBF.

In this paper, we inherit the factorization in (2) and impose further smoothness in the TBFs through exploiting an underlying autoregressive (AR) model. Specifically, the amplitude at time point $t$ in the $k$th TBF, denoted by $\phi_{kt}$, is modeled as a noisy average of previous amplitudes with the width $\gamma$

$$\phi_{kt} = \frac{1}{T} \sum_{\gamma=1}^{T} \phi_{k\gamma} + \epsilon_{kt}$$  \hspace{1cm} (3)

where $\epsilon_{kt}$ denotes the residual error variable. (3) can be expressed collectively as follows:

$$\phi_k = M \xi_k$$  \hspace{1cm} (4)

$$\Phi = \Xi M^T$$  \hspace{1cm} (5)

where $\Xi = [\xi_1, \xi_2, \ldots, \xi_K]^T \in \mathbb{R}^{K \times T}$ and $M = (I - N)^{-1} \in \mathbb{R}^{T \times T}$ with $N$ satisfied

$$N^T_j = \begin{cases} \frac{1}{T} & 0 < j - i < \gamma \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (6)

where $N^T_j$ denotes the $(i, j)^{th}$ element of $N$. Rather than enforcing the prior distribution in TBFs directly, we assume $\xi_k \sim \mathcal{N}(\xi_k | 0, \Theta_k^{-1})$ implying the constraints for $\phi_k$ based on the linear relationship between them.

2) Spatial Modeling: To reconstruct the current sources with spatial coherence and distributed activation, weighting coefficients $W$ is factorized into the product of smoothing kernel $A$ and sparseness encoding coefficients $\Theta$ as

$$W = A \Theta$$  \hspace{1cm} (7)

where $\Theta = [\theta_1, \theta_2, \ldots, \theta_K] \in \mathbb{R}^{D \times K}$. With the employment of (7), weighting coefficients are transformed into the
latent space, in which the sparseness can be readily imposed. Specifically, \( A \) is defined as a Green’s function [21] based on a graph Laplacian over the cortical surface, and \( \Theta \) is constructed depending on the overlapped clusters of dipoles in the following manner:

\[
\theta_k = \sum_i \zeta_{i,k} \tag{8}
\]

\( \zeta_{i,k} \) indicates the coefficients vector corresponding to the \( i \)th cluster. The zero mean Gaussian distribution with learnable scale \( \gamma_i \) is imposed to \( \zeta_{i,k} \) as follows:

\[
p(\zeta_{i,k}|\gamma_i) = \mathcal{N}\left(\zeta_{i,k}|0, \gamma_i^{-1}I_{|I_i|}\right) \tag{9}
\]

where \( I_{|I_i|} \) is a diagonal matrix with 1 in the position of the \( i \)th cluster and 0 otherwise. Hence, the \( d \)th component of \( \zeta_{i,k} \) is enforced to zero if the \( d \)th dipole is not included in the \( i \)th cluster. For convenience, \( G_{i} \) is introduced to denote the dipole indexes in the \( i \)th cluster. The variables \( \zeta_{i,k} \) with distinct cluster indexes \( i \) are assumed mutually statistically independent. Because of the linear relationship in (8), \( \theta_k \) implicitly follows the distribution as:

\[
p(\theta_k|\gamma) = \mathcal{N}\left(\theta_k|0, \Gamma^{-1}\right)
\]

\[
\Gamma^{-1} = \begin{bmatrix}
\sum_{i \in I_d} \gamma_i^{-1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sum_{i \in I_D} \gamma_i^{-1}
\end{bmatrix} \tag{10}
\]

where \( I_d \) denotes the cluster indexes with \( I_d = \{i|d \in G_i\} \). The detailed construction procedure of smoothing kernel \( A \) and clusters \( \{G_i\} \) is presented in Section II-C.

### B. Full Imaging Algorithm

Substituting (5) and (7) into (2), current sources are represented as below

\[
S = A \Theta \Xi M^T \tag{11}
\]

According to (1), the generative model is converted into the following form

\[
B = LA \Theta \Xi M^T + \epsilon = F \Theta \Xi M^T + \epsilon \tag{12}
\]

where \( F = LA \). Both \( \theta_k \) and \( \xi_k \) are assumed statistically independent with different \( k \). Therefore, the full probabilistic model is constructed as follows

\[
p(B|\Theta, \Xi) \propto \exp\left(-\frac{1}{2}\|B - F \Theta \Xi M^T\|^2_F\right)
\]

\[
p(\Xi|\alpha) = \prod_{k=1}^{K} N\left(\xi_k|0, \alpha_k^{-1}I\right) \propto \prod_{k=1}^{K} \exp\left(-\frac{\alpha_k}{2} \xi_k^T \xi_k\right)
\]

\[
p(\Theta|\gamma) = \prod_{k=1}^{K} N\left(\theta_k|0, \Gamma^{-1}\right) \propto \prod_{k=1}^{K} \exp\left(-\frac{1}{2} \theta_k^T \Gamma \theta_k\right) \tag{13}
\]

The full probabilistic generative model is presented graphically in Fig. 1.

Estimation of posterior \( p(\Theta, \Xi|B) \) is adequate to reconstruct current sources \( S \) from the E/MEG measurements \( B \).

Nevertheless, it is intractable owing to the coupling between \( \Theta \) and \( \Xi \) in the generative model. Instead, approximate posterior \( q(\Theta, \Xi) \) is estimated under the variational Bayesian (VB) framework [22], [23], which can be obtained by maximizing the free energy \( \mathcal{F} \) with the definition as

\[
\mathcal{F} = \langle \log p(B, \Theta, \Xi) \rangle - \langle \log q(\Theta, \Xi) \rangle \tag{14}
\]

where \( \langle f(x) \rangle_{q(x)} \) denotes the expectation of \( f(x) \) under the distribution \( q(x) \). Mean field assumption is widely used to establish the VB algorithms, in which the variational posterior is decomposed into mutually independent components. Each ingredient is updated alternatively until the free energy can be hardly promoted or another stopping criterion is fulfilled. In this work, we assume that \( q(\Theta, \Xi) \) is factorized over groups in the following manner

\[
q(\Theta, \Xi) = \prod_{k=1}^{K} q(\theta_k) q(\xi_k) \tag{15}
\]

The VB algorithm can be framed into empirical Bayesian scheme with its update procedure partitioned into two alternate steps as expectation-maximization (EM) algorithm [24], [25]. In the VB-E step, \( \mathcal{F} \) is promoted w.r.t the approximate posterior, given the hyper parameters \( \alpha, \gamma \). In the VB-M step, free energy is maximized w.r.t the hyper parameters with fixed \( q(\theta_k) \) and \( q(\xi_k) \) for each \( k \).

1) **VB-E Step**: In the VB-E step, to derive the update rule for each ingredient \( q(\theta_j) \) in the variational posterior, \( \mathcal{F} \) is maximized with fixed \( q(\theta_j), j \neq k, q(\Xi) \) as well as entires hyper parameters. The update rule is obtained as

\[
q(\theta_k) = \mathcal{N}(\bar{\theta}_k, \Sigma_{\bar{\theta}_k})
\]

\[
\bar{\theta}_k = \Sigma_{\bar{\theta}_k} F^T \left( BM(\xi_k)q(\Xi) - F \sum_{j \neq k} \langle \xi_j M^T M \xi_k \rangle q(\Xi)\theta_j \right) + \Gamma
\]

\[
\Sigma_{\bar{\theta}_k}^{-1} = F^T F(\xi_k M^T M \xi_k) q(\Xi) + \Gamma \tag{16}
\]
The detail derivation is presented in the supplemental document. Similarly, the update for \(q(\xi_k)\) is given as
\[
q(\xi_k) = N(\xi_k | \bar{\xi}_k, \Sigma_{\xi_k})
\]
\[
\bar{\xi}_k = \Sigma_{\xi_k} M^\top (B^\top F(\theta_k) q(\Theta) - M \sum_{j \neq k} (\theta_j^\top F(\theta_k) q(\Theta) \bar{\xi}_j))
\]
\[
\Sigma_{\xi_k}^{-1} = M^\top M (\theta_k^\top F(\theta_k) q(\Theta) + \alpha_k I)
\]

2) VB-M Step: In the VB-M step, \(\alpha, \gamma\) are optimized by promoting \(F\) with fixed \(q(\Theta, \Xi)\). Specifically, the EM-based update for \(\alpha_k\) is obtained by maximizing \((\log p(\xi_k | \alpha_k))_{q(\xi_k)}\) with fixed \(q(\xi_k)\) as below
\[
\alpha_k = \arg \max_{\alpha_k} (\log p(\xi_k | \alpha_k))_{q(\xi_k)} = \frac{T}{\langle \xi_k^T \xi_k \rangle_{q(\xi_k)}}
\]

To avoid the prohibitively slow convergence of EM-based update for \(\gamma\) owing to high-resolution lead field matrix, a convexity based approach has been widely used to facilitate the optimization [8], [9], [26]. It differs from the EM-based update manner based on the fixed-point criterion by taking the change of \(F\) w.r.t \(\gamma\) and rearranging terms [8], [27]. The update formulas for \(\gamma\) are as follows
\[
\gamma_i^{(\text{new})} = \frac{\mu_i}{\rho_i}
\]
\[
\mu_i = \sum_{k=1}^{K} \text{trace} \left( F_{[i]}^T (F^{-1} F^T + I / x_k)^{-1} F_{[i]} \right)
\]
\[
\rho_i = \sum_{k=1}^{K} \theta_k^T \Gamma I_{[i]} \tilde{\theta}_k
\]
\[
x_k = \langle \xi_k^T M^T M \xi_k \rangle_{q(\Xi)}
\]

Unfortunately, because the term \(\theta_k^T \Gamma \tilde{\theta}_k\) in (19) is no longer linear to \(\gamma_i\), we are not allowed to derive a convex-based update rule for \(\gamma\) even though \(\log |\Gamma| + \log |\Sigma_{\theta k}|\) is still convex to \(\gamma_i^{-1}\). To remedy this issue, we propose a novel update manner based on the fixed-point criterion by taking the gradient of \(F\) w.r.t \(\gamma\) and rearranging terms [8], [27]. The update formulas for \(\gamma\) are as follows
\[
\gamma_i^{(\text{new})} = \gamma_i^{(\text{old})} \frac{\mu_i}{\rho_i}
\]

The expectation terms in the above update equations are computed as follows
\[
\langle \xi_i^T M^T M \xi_j \rangle_{q(\Xi)} = \bar{\xi}_i^T M^T \bar{\xi}_j + \delta_{ij} \text{trace} (M^T M \Sigma_i)
\]

\[
\langle \theta_i^T F^T \theta_j \rangle_{q(\Theta)} = \bar{\theta}_i^T F^T \bar{\theta}_j + \delta_{ij} \text{trace} (F^T F \Sigma_i)
\]
where \(\delta_{ij}\) is equal to 1 if \(i = j\) or 0 otherwise. \(\langle \xi_i^T \xi_j \rangle_{q(\Xi)}\) is the generalization of (21) with \(M = I\).

In summary, the variational posterior \(q(\theta_k), q(\xi_k)\) as well as the hyper parameters \(\alpha, \gamma\) are updated alternately with (16) (17) (18) in each epoch until the relative change of \(F\) is less than specific threshold. The procedure is encapsulated in Algorithm 1, which is referred to as SI-SST (Source Imaging with Smoothness on Spatial and Temporal domain).

Algorithm 1 The SI-SST Algorithm

Input:
- Measurement \(B\), lead field matrix \(L\), spatial smoothing kernel \(A\), temporal smooth matrix \(M\), cortical clusters \([G_i]\) and tolerance \(\delta\)

Output:
- Current sources estimation \(S\)

1: Initialize encoding coefficients \(\Theta\), TBFs \(\Phi\) and the hyper parameters \(\gamma, \alpha\), calculate \(\Xi\) with \(\Xi = (M^{-1} \Phi^T)^T\)

2: repeat
3: update the variational posterior distribution of \(\theta_k, \xi_k\) with (16) and (17) respectively
4: update the hyper parameters \(\alpha, \gamma\) using (18) and (20) respectively
5: compute the free energy \(F\) based on (19)
6: until relative change of \(F < \delta\)
7: \(S = A \Theta \Xi M^\top\), where \(\Theta\) and \(\Xi\) are means of the variational posterior \(q(\Theta)\) and \(q(\Xi)\) respectively

C. Construction of Clusters and Smoothing Kernel

In this work, spatial smoothing kernel \(A\) is derived from the spatial coherence prior [25] and the clusters \([G_i]\) are constructed by the data drive parcellization (DDP) of the cortical surface [28].

1) Spatial Smoothness Constraint: We construct the smoothing kernel based on a graph Laplacian \(\tilde{A}\) over the geodesic surface [25], [28]. \(\tilde{A} \in \mathbb{R}^{D \times D}\) is derived from the adjacency matrix \(A \in \mathbb{R}^{D \times D}\), which reflects the face connection of cortical surface \((A_{ij} = 1\) if dipole \(i\) and \(j\) are neighboring and \(A_{ij} = 0\) otherwise)
\[
\tilde{A}_{ij} = \begin{cases} -\sum_{d=1}^{D} A_{id} & i = j \\ A_{ij} & i \neq j \end{cases}
\]

The smoothing kernel is defined as the Green’s function [21] with positive scale \(\sigma\) and it can be approximated by the first several terms in the Taylor expansion as:
\[
A = \exp(\sigma \tilde{A}) \approx \sum_{k=0}^{8} \frac{\sigma^k}{k!} \tilde{A}^k
\]
Notably, the approximation can be interpreted as weighted average of different orders of graph Laplacian $\tilde{A}$. Meanwhile, $\sigma$ controls the degree of spatial smoothness of the smoothing kernel $A$. We set $\sigma = 0.6$ recommended by [25].

2) DDP of the Cortical Surface: DDP consists of a pre-localization procedure called Multivariate Source Pre-localization (MSP) [29] and a subsequent region growing process. MSP estimates a probability like coefficient which measures the contribution of each dipole to the E/MEG measurements. The dipole with highest MSP score is selected as the seed and eliminated from candidate set. Region is grown around this seed by clustering the neighbor remaining as the seed and eliminated from candidate set. Region is grown around this seed by clustering the neighbor remaining as the seed and eliminated from candidate set. Region is grown around this seed by clustering the neighbor remaining as the seed and eliminated from candidate set. Region is grown around this seed by clustering the neighbor remaining as the seed and eliminated from candidate set.

D. Implementational Details

In this section, we discuss some implementational details concerning the initialization and computational cost of SI-SST in practice.

1) Initialization and Hyper Parameters: Because the iterative update is adopted in SI-SST, the initial values of parameters and hyperparameters are necessary and essential to the entire algorithm. In our simulations, encoding coefficients $\Theta$ is initialized to $0$ while the TBFs $\Phi$ is initialized based on the SVD of E/MEG measurements $B$. Specifically, the decomposition $B = U \Lambda V^\top$ is conducted by SVD and the first $K$ components of $V^\top$ with largest singular values are chosen as the initial value of $\Phi$. We set $K = 7$ in this work. The initial hyper parameters is chosen as $y = 1_{d_t} \cdot \alpha = 1_K$. It is worth noting that the final numbers of clusters $\{G_i\}$ and TBFs $\Phi$ are automatically determined by the hyper parameters $y$ and $\alpha$ after convergence. More concretely, if $\gamma_i$ is large enough (e.g., $\gamma_i > 10^6$), the corresponding cluster $G_i$ is pruned during the optimization because it has negligible influence in the computation and can be interpreted as irrelevant to the measurements. Similarly, if $a_k$ is large enough (e.g., $a_k > 10^4$), the associated ingredients $\theta_k$ and $\xi_k$ are eliminated from VB-EM procedure.

2) Stopping Criterion and Computation Requirements: Throughout our implementation, iterative update is terminated if the relative change in free energy $\mathcal{F}$ is smaller than a specific tolerance value $\delta$. In this work, $\delta$ is chosen as $10^{-8}$. For $P = 62$, $T = 501$, $D = 6001$, the number of parcels $[G_i]$ derived by DDP is typically around 1600 which equals the dimensions of $y$. Using a standard PC (Corei7-10700 CPU 2.9 GHz and 16 GB RAM), the convergence time is approximately 40 seconds. To facilitate reproducibility, the MATLAB code for the implementation of SI-SST is available at the following URL: https://github.com/LiangJiwen1994/Source-Imaging.git.

III. SIMULATION PROTOCOL

A. Benchmark Source Imaging Algorithms

In this paper, SI-SST is tested against SI-STBF [19] for the similar modeling in both spatial and temporal domain. Additionally, SI-SST is compared with the widely used L2 norm based method, LORETA [6], as well as the sparse-constrained algorithm based on empirical Bayesian learning, SBL [8]. Furthermore, VB-GLM [15] and STTONNICA [30] are chosen as the benchmarks for the usage of both spatial and temporal constraints. The detailed description of the benchmark methods is presented in the supplementary document.

B. Validation Dataset

The cortical surface used for numerical simulations is obtained by downsampling the default surface of MNI/ICBM152 anatomy distributed with Brainstorm [31]. The downsampled cortical mesh consists of 6001 dipoles. The lead-field matrix is acquired via OpenMEEG [32] software package embedded in Brainstorm, based on the sensor configuration in 64-channel Neuroscan Quik-cap system and three layers head surface obtained from Boundary Element Method (BEM).

To generate spatial extended sources for Monte-Carlo experiments, we firstly select the seed dipoles on cortical mesh randomly and then absorb neighbor elements into the patches iteratively until the desired area of activated region is achieved. To reduce the variance of assessments caused by the diverse patches positions, 100 Monte-Carlo experiments were carried out such that the majority of the cortical surface is covered by the generated extended patches. Dipoles are activated with Gaussian-damped sinusoidal time courses which is defined as $s(t) = \sin(2\pi f t) \exp \left(-\left(\frac{\tau}{\omega}\right)^2\right)$, where $f$, $\tau$ and $\omega$ denote the frequency, phase and damping rate respectively. The time courses are partitioned into pre- and post-stimulus periods with 250 time points per phase. Sources are considered to be active during the post-stimulus period. Simulated sources are transformed into simulated measurements with the lead field matrix and the Gaussian noise in line with (1). The noise is integrated in specific signal-to-noise ratio (SNR) with the definition as $\text{SNR} = 20 \log_{10} \frac{S}{\sqrt{\text{MSE}}}$.

In order to investigate the influence of multiple factors on the proposed algorithm, following scenarios are under consideration:

1) Single patch with various extents. SI-SST is tested with the compared methods under the configuration of one patch source with the following extents: 0.5cm$^2$, 2cm$^2$, 6cm$^2$, 10cm$^2$, 18cm$^2$, and 32cm$^2$. SNR is set 5 dB.
C. Validation Metrics

In this work, four validation metrics are utilized to acquire the quantitatively and comprehensively evaluation of the source imaging performance. The first metric is the area under the receiver operating characteristic curve (AUC) [23], which is an effective assessment for sensitivity and specificity of the detection. The next metric is the spatial dispersion (SD), which measures the spatial blurriness of the estimated sources [33], [34]. To evaluate the correctness of the localization, the distance of localization error (DLE) [33], [34] is introduced. Moreover, shape error (SE), which is the square error between the normalized estimated and simulated sources, is utilized to evaluate the accuracy of shape fitness of reconstructed time courses. Finally, mean correlation coefficient (MCC) measures the temporal fitness in another aspect, which represents the averaged correlation coefficient between the mean simulated and reconstructed times courses among active regions. For good source imaging performance, the large AUC and MCC as well as small SD, DLE, and SE values are desired. The detailed computations of the aforementioned metrics are presented in the supplemental document. The significance of the entire results is assessed by Kruskal-Wallis test, which plays the role as the non-parametric counterpart of one-way analysis of variance. If a test statistic from the Kruskal-Wallis test is significant, Wilcoxon rank sum tests are subsequently performed in SI-SST against each of the other compared methods to determine if SI-SST yields significantly better estimates.

IV. RESULTS

For clear visualization of the source imaging results at specific time point, the thresholded absolute value of the current sources, which is normalized with unit maxima, is shown on the cortical surface in this paper. Additionally, the thresholds are determined upon the level of background activity based on Otsu’s method [28], [35], [36].

A. Results of Simulated Data

1) Influence of Spatial Extent: Fig. 2 depicts the performance metrics of algorithms with varying extents. As shown in the plots, LORETA and VB-GLM acquire upgraded reconstruction performance with increasing AUCs ($p < 0.01$) and decreasing DLEs ($p < 0.01$), SDs ($p < 0.01$), and SEs ($p < 0.01$) as the spatial extents increase. On the contrary, SBL attains degraded performance with increasing extents. STTONNICA is not capable to acquire stable AUC values but it achieves clearly descending values of DLE ($p < 0.01$), SD ($p < 0.01$) and SE ($p < 0.01$). Notably, instead of being monotonic as the active patches enlarge, AUC values of SI-SST increase first and then descend. At the mean time, it acquires the values of SD, DLE, and SE in the converse tendency against AUC. In summary, SI-SST is appropriate to reconstruct the sources within a specific range of extents. Additionally, SI-STBF performs similarly as SI-SST owing to analogous spatial modeling. Regarding the MCC, SI-SST and VB-GLM acquire the stable temporal fitness and have superior performance against the compared methods with varying spatial extents. LORETA and SBL, both of which are modeling without the usage of TBFs, result in the relatively low MCCs and wild oscillations in statistics. Moreover, SI-SST outperforms all benchmarks with the exception in some extreme scenarios. Specifically, it attains the highest AUC ($p < 0.01$), MCC ($p < 0.01$ except VB-GLM) and lowest DLE ($p < 0.01$), SD ($p < 0.01$), and SE ($p < 0.05$) values among all the compared methods with extents from $2cm^2$ to $18cm^2$.

To more concretely illustrate how the source imaging capabilities of different algorithms are influenced by spatial extents, one realization of simulations with the imaging results is depicted in Fig. 3. Two active sources are located on postcentral gyrus and middle temporal area respectively with individual time course as shown in Fig. 4(a). LORETA and VB-GLM always reconstruct the sources with diffuse extents but SBL acquire focal imaging results invariably. Additionally, STTONNICA is difficult to acquire stable and reliable imaging results with varying extents sources. In line with Fig. 2, SI-SST fails to obtain perfect source imaging in extremely focal or diffuse sources activation but it reconstruct the sources with most plausible extents and accurate localizations among the compared methods. SI-STBF has the analogous performance with SI-SST but obtains sightly worse imaging results. Fig. 4(b) demonstrates the time courses reconstruction of each algorithms when sources are activated with the area of $10cm^2$. The correlation coefficients between the simulated and estimated time courses are depicted in each plots. SI-SST
Fig. 3. Source imaging results of different methods for varying spatial extents with two simultaneous simulated sources. The sources are activated in distinct temporal patterns. The rescaled absolute value of current activities is presented at the peak of one of the source. The SNR is 5 dB.

Fig. 4. The simulated and reconstructed time course for the two simultaneous sources. (a) Temporal profile of simulated sources. (b) The reconstructed (mean) time courses of different methods are presented in the sub-figures. The correlation coefficients between the true time courses and the estimated mean time courses are located in the legends.

produces the most accurate temporal profile in both postcentral gyrus and middle temporal area among all the compared methods.

2) Influence of SNR: The performance metrics of different algorithms with various SNRs are presented in Fig. 5. The source imaging performance of all the algorithms except STTONNICA is significantly influenced by the SNRs. As the SNRs increase, they acquire increasing AUCs ($p < 0.01$), MCCs ($p < 0.01$) and descending values of SD ($p < 0.01$), DLE ($p < 0.01$), and SE ($p < 0.01$). SI-SST is relatively robust to the SNRs and outperforms all the benchmarks. When noise is integrated with huge intensity (i.e., SNR is less than 0 dB), SI-SST acquires significantly superior performance in multiple criteria while the advantage gradually vanishes as SNRs enlarge.

To further illustrate the variation on sources reconstruction capabilities of different algorithms caused by varying SNRs, one simulation with the imaging results are presented in Fig. 6. Two sources are located on postcentral gyrus and middle temporal area resembling to the one in Fig. 3. Nevertheless, the sources are activated with the identical temporal profiles in this realization. Consistent with the demonstration in Fig. 5, all the methods except STTONNICA can reconstruct the sources with more accurate extents and locations with the decreased intensity of noise. Owing to the identical time courses of two active sources in this example, SI-SST is empirically capable to deal with the situation where different sources are activated with highly correlated temporal patterns.

B. Results of Real Data

1) Results With Median Nerve Stimulation MEG Data: The MEG data were acquired at the Massachusetts General Hospital in 2005 with a Neuromag Vectorview 306 system (102 magnetometers and 204 planar gradiometers) based on the experiments of median nerve stimulation. The dataset was composed of MEG measurements recorded at 1793 Hz of one healthy subject from stimulations at both arms and it was available from Brainstorm download page. In terms of our analysis, we used merely the recordings during the left arm stimulation. The pre-stimulus window is -100 ms to 0 ms and the post-stimulus window was chosen as 0 ms to 300 ms, where 0 ms indicated the stimulus onset. The averaged respond across 301 trials was used in the subsequent procedure. Noise covariance was estimated from the pre-stimulus record of both arms. The overlapping spheres model [37], which was efficient for MEG data without loss of accuracy, was used to acquire the lead field matrix according to the subject’s individual anatomy with its cortical surface downsampled with 6001 voxels.

As expounded in the prior studies [38], a sophisticated region on the cortical surface is activated under the median nerve stimulation, including the contralateral primary somatosensory cortex (cSI), bilateral secondary somatosensory cortices (SII), and posterior parietal cortex (PPC). Fig. 7 depicts the average time courses for the four aforementioned cortical regions reconstructed by distinct algorithms. As shown in the plots, SI-SST, SI-STBF, and VB-GLM result in smoother time courses reconstruction than LORETA and SBL. Furthermore, SI-SST yields significantly smoother profile of reconstructed time courses than SI-STBF, even though they are modeled in similar spatial pattern. Notably, the peak for cSI at around 21 ms, which appears in reconstructed time courses of most methods, disappears in the reconstruction of SI-SST due to the imposed temporal smoothness. The peak is eliminated from the time course in SI-SST because it dose not acquire enough interpretation from the measurements to struggle from the further temporal constraint introduced by the underlying AR model. Overall, the reconstructed activation of

1 https://neuroimage.usc.edu/bst/download.php
Fig. 5. Performance metrics for various levels of SNRs. The figures are results of 100 Monte-Carlo simulations. Data are shown as the medians with 95% confidence intervals.

Fig. 6. Source imaging results of different methods for varying SNRs with two simultaneous simulated sources. The two sources are activated with 6cm² extents in identical temporal pattern. The rescaled absolute value of current activities at the peak of sources are presented in the imaging results.

Fig. 7. The averaged time courses of sources located at contralateral primary somatosensory cortex (cSI), bilateral secondary somatosensory cortices (SII) and posterior parietal cortex (PPC), which are estimated by different methods for MEG recordings under the median nerve stimulation.

2) Results With Face Processing Experiment MEG Data: This data is obtained from the OpenNeuro database. Its accession number is ds000117. The dataset was composed of simultaneous MEG/EEG recordings from 16 participants performing a simple visual recognition task from presentations of famous, unfamiliar and scrambled faces. The experimental design and multiple modalities data acquirement were expounded in [40]. Only the MEG data were under consideration in our analysis, which were recorded at 1100 Hz with an Elekta-Neuromag VectorView 306 system as well. The averaged response of MEG measurement obtained from the subject 01 during the famous face presentation epoch was used for the subsequent analysis. The pre-stimulus window was chosen as -100 ms to 0 ms and the post-stimulus window was 0 ms to 300 ms (0 ms means stimulus onset). The lead field matrix was acquired with Brainstorm based on overlapping sphere model with the cortical surface composed of 6000 vertices. The noise covariance was estimated from the empty room recordings.

Fig. 8. Source imaging results for MEG recordings of median nerve stimulation. The reconstructions of different methods are demonstrated respectively from the left to right columns. The upper two rows show the results at 35 ms after stimulus onset and the lower two rows present the results at 85 ms post-stimulus. The imaging are presented as the rescaled absolute value of sources at each time point.

SI-SST has the temporal pattern in agreement with the prior studies [17].

The source imaging results of all the algorithms at 35 ms and 85 ms are depicted in Fig. 8. As reflected in the plots, all the methods are capable to locate the sources in cSI at 35 ms. LORETA and VB-GLM produce too diffuse estimations. On the contrary, SBL reconstructs the activation with extremely focal extents. It is a challenge to locate the activation at iSII area at 85 ms while only SI-SST, SI-STBF, and STTONNICA produce the source imaging with correct locations and plausible extents. Compared to SI-STBF, SI-SST presents more compact imaging result without outlier patches in agreement with the prior studies [38], [39].
by different methods. For SI-SST, the estimated sources are activated with the peaks in both occipital and fusiform areas at around 100 ms. Additionally, the activations near the right fusiform gyrus have another peak at around 170 ms. With the usage of underlying AR model, SI-SST achieves the smoother reconstructed time courses compared to the benchmarks. Fig. 10 presents the source imaging results of all the algorithms at 100 ms and 170 ms under the famous faces stimulus. As shown in the plots, SI-SST successfully locates the neural activation at the occipital area at 100 ms and reconstructs the sources at bilateral fusiform at 170 ms. Most methods are capable to reconstruct the occipital activation and the right fusiform gyrus sources. However, SBL and SI-STBF fail to locate the source on the left ventral surface of fusiform while LORETA and VB-GLM produce diffuse reconstructions which cover many functionally irrelevant regions. In sum, SI-SST reconstructs the consistent results with the previous reports [15], [41] on both the time courses and activation regions.

V. DISCUSSION

Because of the inherent ill-posedness of the E/MEG source imaging problem, prior constraints are of great importance to acquire a meaningful solution in the context of neurophysiology. It has been widely accepted by previous studies [42], [43] that cortical activations enjoy the spatial coherence and local homogeneity. Multiple L2-norm based methods have been proposed according to these properties. Specifically, current sources are assumed to be activated with similar amplitudes in MNE [5]. Spatial smoothness is imposed in LORETA [6] with the usage of discrete Laplacian operator which represents the local correlation between dipoles. The L2-norm based methods have preference for diffused estimates [28], [34] in line with our numerical experiments. On the contrary, algorithms with sparseness enforced are appropriate for the focal sources reconstruction while they achieve the poor imaging performance with extended activations. Champagne [9] is a representative algorithm in which the spatial sparseness induced with the usage of individual covariance components for dipoles.

Smoothing kernel is typically used constraints in the spatial domain, through which the brain sources are transformed into a latent space with sparseness. In Smooth Champagne [10], the kernel is defined depending on the distances between voxels. Additionally, to represent the local spatial correlation over cortical surface, discrete Laplacian operator is employed in STTONNICA [30], VB-GLM [15], and BESITES [11]. Furthermore, in SI-SST, smoothing kernel is derived from a Green’s function [21] based on the graph Laplacian operator. Notably, SI-STBF [19] has analogous configuration in spatial domain to SI-SST, but whose smoothing kernel is indirectly constructed through the covariance components.

Parameters sharing in sparseness encoding coefficients also makes a great difference to the source imaging results. For example, a global scale is employed in MNE and LORETA, with an implicit assumption that all the current sources have similar amplitudes. On the contrary, dipoles are equipped with individual parameters in SBL, which readily acquires focal sources reconstruction. Additionally, SI-SST has the sparseness encoding coefficients derived from DDP [28] over the cortical surface. Moreover, owing to the overlapped clusters generated from DDP, a novel update rule is proposed in SI-SST while the convexity based approach becomes invalid. It should be emphasized that smoothing kernel and parameters sharing need be regarded as an unified system instead of independent ingredients. They collaboratively determine the source imaging performance of the entire model. As shown in Fig. 2, distinct algorithms are appropriate to reconstruct the sources with different extents. Specifically, SBL has perfect performance in focal source reconstruction while LORETA is suitable to dual with extremely extended activation. In contrast, SI-SST achieves the best reconstructed results in a specific range of extents and has the degraded performance with extremely focal and extended sources. Besides, SI-STBF has the similar pattern in spatial modeling with SI-SST. However, it obtains slightly worse recovery results owing to the difference in parameters sharing. Furthermore, in Smooth Champagne, spatial smoothing kernel is defined with great smoothness degree and parameters sharing is constructed based on small amounts of non-overlapping clusters of the voxels. As shown in the supplemental material, compared to Smooth Champagne, SI-SST acquired superior performance in source localization and time courses reconstruction except that sources were activated with huge spatial extents. Therefore, smoothing kernel and parameters configuration should be modified simultaneously to achieve the satisfactory sources imaging performance within a range of extents as wide as possible.
In addition to manually fine tuning the smoothing kernel, learnable parameters can be introduced such that the algorithm becomes adaptive to source imaging tasks with a wide range of extents. Specifically, in BESTIES, a parameter integrated in MRF model, which controls the smoothness degree, is updated with the other coefficients simultaneously under the empirical Bayesian framework. Analogously, in STTONNICA [30], a smoothness related parameter is also integrated in the smoothing kernel. However, the optimization of them are both computationally inefficient. More concretely, the former involves calculating the inverse of a matrix with great dimensions in every epoch, while the latter utilizes grid-search for hyper parameters update. How to construct the flexible smoothing kernel with learnable parameters, which can be updated in an efficient fashion, still remains a challenge.

Apart from the prior constraints in the spatial domain, the employment of underlying temporal correlation is effective to promote the localization accuracy and robustness against the noise. Temporal basis functions (TBFs) have been proven effective in source reconstruction by the prior studies [18]. With the usage of TBFs, SI-SST, SI-STBF and VB-GLM are more likely to reconstruct smoother time courses compared to LORETA and SBL as depicted in Fig. 4. In both SI-STBF and SI-SST, TBFs are updated by variational Bayesian inference. The irrelevant ingredients with little interpretation of measurements are pruned automatically thanks to the automatic relevance determination (ARD) mechanism [44]. Additionally, because of the further temporal smoothness imposed with an underlying autoregressive model, compared to SI-STBF, SI-SST not only results in smoother time courses reconstruction reflected in Fig. 7 and Fig. 9, but also becomes more robust against the noise as shown in Fig. 5. As discussed in the supplemental document with three different scenarios of examples, SI-SST is capable to maintain the essential high frequency information of sources from severely corrupted measurements. Moreover, SI-SST acquire the smoothest PSD curve among the compared methods such that the redundant frequency components are pruned.

The proposed algorithm can be further promoted in multiple aspects. First, learnable parameters are integrated into the smoothing kernel as in STTONNICA [30] and BESTIES [11], such that the smoothness degree is allowed to be adaptively modified. However, effective update rule should be proposed collaboratively to obviate the expensive computational cost. Furthermore, the cross-clusters effect can be taken into account. The generative model in (8) is generalized by introducing the cross clusters terms. Alternatively, they are directly integrated into the covariance components in (10). Additionally, temporal constraints can be constructed in other manners. For instance, Gaussian distributions can be imposed to the second derivative of distinct TBFs, which can be represented as the variation model in [12]. The redundant components are pruned automatically with the help of ARD mechanism.

VI. CONCLUSION

The current study presented an E/MEG source imaging algorithm termed SI-SST, in which the smoothness is imposed in both spatial and temporal domain, via the employment of a spatial smoothing kernel and an underlying autoregressive model respectively. Sparseness encoding coefficients are constructed based on data driven parcellization of the cortex with overlapped clusters. Because the widely used convex approximation becomes invalid in this situation, we proposed a novel update rule derived from the fixed-point criterion. Entire variables and parameters in the model are updated simultaneously in a data-driven fashion under empirical Bayesian framework. SI-SST was assessed by multiple validation metrics with both simulated and experimental datasets. Numerical simulations showed that, compared to the benchmark methods, SI-SST had the superior imaging performance in a wide range of extents, and was more robust against diverse intensities of noise.

REFERENCES

[1] S. Baillet, J. C. Mosher, and R. M. Leahy, “Electromagnetic brain mapping,” IEEE Signal Process. Mag., vol. 18, no. 6, pp. 14–30, Nov. 2001.
[2] J. Wu, S. Nagarajan, and Z. Chen, “Bayesian machine learning: EEG/MEG signal processing measurements,” IEEE Signal Process. Mag., vol. 33, no. 1, pp. 14–36, Jan. 2016.
[3] B. He, T. Musha, Y. Okamoto, S. Homma, Y. Nakajima, and T. Sato, “Electric dipole tracing in the brain by means of the boundary element method and its accuracy,” IEEE Trans. Biomed. Eng., vol. BE-34, no. 6, pp. 406–414, Jun. 1987.
[4] J. C. Mosher, P. S. Lewis, and R. M. Leahy, “Multiple dipole modeling and localization from spatio-temporal MEG data,” IEEE Trans. Biomed. Eng., vol. 39, no. 6, pp. 541–557, Jun. 1992.
[5] M. S. Hämäläinen and R. J. Ilmoniemi, “Interpreting magnetic fields of the brain: Minimum norm estimates,” Med. Biol. Eng. Comput., vol. 32, no. 1, pp. 35–42, Jan. 1994.
[6] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, “Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain,” Int. J. Psychophysiol., vol. 18, no. 1, pp. 49–65, Oct. 1994.
[7] R. D. Pascual-Marqui, “Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details,” Methods Findings Exp. Clin. Pharmacol., vol. 24, pp. 5–12, Feb. 2002.
[8] D. Wipf and S. Nagarajan, “A unified Bayesian framework for MEG/EEG source imaging,” NeuroImage, vol. 44, no. 3, pp. 947–966, 2009.
[9] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, “Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG,” NeuroImage, vol. 49, no. 1, pp. 641–655, Jun. 2010.
[10] C. Cai, M. Divakar, D. Chen, K. Sekihara, and S. S. Nagarajan, “Robust empirical Bayesian reconstruction of distributed sources for electromagnetic brain imaging,” IEEE Trans. Med. Imag., vol. 39, no. 3, pp. 567–577, Mar. 2020.
[11] K. Liu, Z. L. Yu, W. Wu, Z. Gu, Y. Li, and S. Nagarajan, “Bayesian electromagnetic spatio-temporal imaging of extended sources with Markov Random field and temporal basis expansion,” NeuroImage, vol. 139, pp. 385–404, Oct. 2016.
[12] J. Daunizeau, J. Mattout, D. Clonda, B. Goulard, H. Benali, and J.-M. Lina, “Bayesian spatio-temporal approach for EEG source reconstruction: Conciliating ECD and distributed models,” IEEE Trans. Bio-med. Eng., vol. 53, no. 3, pp. 503–516, Mar. 2006.
[13] S. Haufe et al., “Large-scale EEG/MEG source localization with spatial flexibility,” NeuroImage, vol. 54, no. 2, pp. 851–859, Jan. 2011.
[14] T. Limbitti, B. D. V. Veen, and R. T. Wakai, “Cortical patch basis model for spatially extended neural activity,” IEEE Trans. Bio-med. Eng., vol. 53, no. 9, pp. 1740–1754, Sep. 2006.
[15] N. J. Trujillo-Barreto, E. Aubert-Vázquez, and W. D. Penny, “Bayesian M/EEG source reconstruction with spatio-temporal priors,” NeuroImage, vol. 39, no. 1, pp. 318–335, Jan. 2008.
[16] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Häätäläinen, and M. Kowalski, “Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations,” NeuroImage, vol. 70, no. 15, pp. 410–422, Apr. 2013.
A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood estimation and Bayesian inference for compound decision problems,” *Statist. Med.*, vol. 26, no. 21, pp. 3886–3910, Sep. 2007.

K. Liu et al., “Bayesian electromagnetic spatio-temporal imaging of extended sources based on matrix factorization,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2457–2469, Sep. 2019.

H. Hallez et al., “Review on solving the forward problem in EEG source analysis,” *J. Neuroeng. Rehabil.*, vol. 4, no. 1, p. 46, 2007.

J. D. Lépez, V. Litvak, J. Espinosa, K. Friston, and G. R. Barnes, “Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM,” *NeuroImage*, vol. 84, pp. 476–487, Jan. 2014.

C. M. Bishop, *Pattern Recognition and Machine Learning*, New York, NY, USA: Springer, 2006.

K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.

K. Friston et al., “Multiple sparse priors for the M/EEG inverse problem,” *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, Feb. 2008.

C. Stahlhut, H. T. Attias, K. Sekihara, D. Wipf, L. K. Hansen, and S. S. Nagarajan, “A hierarchical Bayesian M/EEG imaging method correcting for incomplete spatio-temporal priors,” in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 560–563.

D. J. C. MacKay, “Bayesian interpolation,” *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.

R. A. Chowdhury, J. M. Lina, E. Kobayashi, and C. Grova, “MEG source localization of spatially extended generators of epileptic activity: Comparing entropic and hierarchical Bayesian approaches,” *PLoS ONE*, vol. 8, no. 2, Feb. 2013, Art. no. e55969.

J. Mattout, M. Pelegrinissac, L. Garnero, and H. Benali, “Multivariate source prelocalization (MSP): Use of functionally informed basis functions for better conditioning the MEG inverse problem,” *NeuroImage*, vol. 26, no. 2, pp. 356–373, Jun. 2005.

P. A. Valdés-Sosa, M. Vega-Hernández, J. M. Sánchez-Bornot, E. Martínez-Montes, and M. A. Bobes, “EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis,” *Hum. Brain Mapping*, vol. 30, no. 6, pp. 1898–1910, Jun. 2009.

F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, “Brainstorm: A user-friendly application for MEG/EEG analysis,” *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–13, 2011.

A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc, “OpenMEEG: Opensource software for quasi-stationary electroencephalography,” *BioMed. Eng. OnLine*, vol. 9, p. 45, Sep. 2010.

W.-T. Chang, A. Nummenmaja, J.-C. Hsieh, and F.-H. Lin, “Spatially sparse source cluster modeling by compressive neuromagnetic tomography,” *NeuroImage*, vol. 53, no. 1, pp. 146–160, Oct. 2010.

M. Zhu, W. Zhang, D. L. Dickens, and L. Ding, “Reconstructing spatially extended brain sources via enforcing multiple transform sparseness,” *NeuroImage*, vol. 86, pp. 280–293, Feb. 2014.

N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMCS-9, no. 1, pp. 62–66, Feb. 1979.

C. Grova, J. Daunizeau, J.-M. Lina, C. Bénar, H. Benali, and J. Gotman, “Evaluation of EEG localization methods using realistic simulations of interictal spikes,” *NeuroImage*, vol. 29, no. 3, pp. 734–753, Feb. 2006.

M. Huang, J. C. Mosher, and R. Leahy, “A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG,” *Phys. Med. Biol.*, vol. 44, no. 2, pp. 423–440, 1999.

R. Hari and N. Forss, “Magnetoencephalography in the study of human somatosensory cortical processing,” *Phil. Trans. Roy. Soc. London B, Biol. Sci.*, vol. 354, no. 1387, pp. 1145–1154, 1999.

N. Forss et al., “Activation of the human posterior parietal cortex by median nerve stimulation,” *Exp. Brain Res.*, vol. 99, no. 2, pp. 309–315, May 1994.

D. G. Wakeman and R. N. Henson, “A multi-subject, multi-modal human neuroimaging dataset,” *Sci. Data*, vol. 2, no. 1, Dec. 2015, Art. no. 150001.

J. P. Owen, D. P. Wipf, H. T. Attias, K. Sekihara, and S. S. Nagarajan, “Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data,” *NeuroImage*, vol. 60, no. 1, pp. 305–323, Mar. 2012.

M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuttila, and O. V. Lounasmaa, “Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain,” *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413–497, Apr. 1993.

A. Destexhe, D. Contreras, and M. Steriade, “Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states,” *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 19, no. 11, pp. 4595–4608, 1999.

M. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jan. 2001.