The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1379–1408
https://doi.org/10.1214/19-AOAS1321

QUANTIFYING TIME-VARYING SOURCES IN MAGNETOENCEPHALOGRAPHY—A DISCRETE APPROACH

BY ZHIGANG YAO[1,*], ZENGYAN FAN[1,†], MASAHITO HAYASHI[2] AND WILLIAM F. EDDY[3]

[1]*Department of Statistics and Applied Probability, National University of Singapore,* [*]*zhigang.yao@nus.edu.sg;* [†]*stafz@nus.edu.sg*

[2]*Graduate School of Mathematics, Nagoya University, masahito@math.nagoya-u.ac.jp*

[3]*Department of Statistics, Carnegie Mellon University, bill@stat.cmu.edu*

We study the distribution of brain source from the most advanced brain imaging technique, Magnetoencephalography (MEG) which measures the magnetic fields outside of the human head produced by the electrical activity inside the brain. Common time-varying source localization methods assume the source current with a time-varying structure and solve the MEG inverse problem by mainly estimating the source moment parameters. These methods use the fact that the magnetic fields linearly depend on the moment parameters of the source and work well under the linear dynamic system. However, magnetic fields are known to be nonlinearly related to the location parameters of the source. The existing work on estimating the time-varying unknown location parameters is limited. We are motivated to investigate the source distribution for the location parameters based on a dynamic framework, where the posterior distribution of the source is computed in a closed form discretely. The new framework allows us not only to directly approximate the posterior distribution of the source current, where sequential sampling methods may suffer from slow convergence due to the large volume of measurement, but also to quantify the source distribution at any time point from the entire set of measurements reflecting the distribution of the source, rather than using only the measurements up to the time point of interest. Both a dynamic procedure and a switch procedure are pro- posed for the new discrete approach, balancing estimation accuracy and computational efficiency when multiple sources are present. In both simulation and real data, we illustrate that the new method is able to provide comprehensive insight into the time evolution of the sources at different stages of the MEG and EEG experiment.

1. Introduction. The human brain produces a wide range of bioelectromagnetic signals from various electrical impulses when activated. The signals produced by the neurons in the brain varies from 10s of femto-Tesla (fT) to 100s of fT which is approximately a billion times smaller than the Earth's magnetic field. Magnetoencephalography (MEG) is a noninvasive imaging technique that is able to detect the weak magnetic fields generated by the neuronal activity within the brain. The MEG recording is able to measure the magnetic fields caused by the neuronal activity inside of the brain based on the instrument that is placed close to the scalp. The Superconducting Quantum Inference Devices (SQUIDs) are sufficiently sensitive magnetometers that are used to measure the extremely subtle magnetic fields, and these magnetometers are fixed in a one-size-fits-all helmet. The extreme sensitivity of SQUIDs makes them ideal for studies. During the MEG scanning of SQUIDs, patients sit under the machine that operates in a magnetically shielded room, and they are restricted from moving. The latest optically pumped magnetometer (OMP) system (Boto et al. (2018)), which is equipped with

Received May 2019; revised October 2019.
Key words and phrases. MEG inverse problem, discrete posterior distribution, expectation-maximization, spatiotemporal model, source localization.

a customized helmet, allows free and natural movement, including head nodding, stretching, drinking and playing a ball game. With the excellent temporal resolution on a millisecond scale, the MEG has been applied to provide new insights into the neural basis of developmental disorders.

1.1. *MEG inverse problem.* In neuromagnetism the neuronal current $\mathbf{J}(r)$ at location r is divided into the primary current $\mathbf{J}^{\mathrm{p}}(r)$ and the volume current $\mathbf{J}^{\mathrm{v}}(r)$ (Hämäläinen et al. (1993)). Since the primary current $\mathbf{J}^{\mathrm{p}}(r)$ is related to the movement of ions and the volume current $\mathbf{J}^{\mathrm{v}}(r)$ does not build up any electric charge, the source of brain activity can be captured by finding the primary current $\mathbf{J}^{\mathrm{p}}(r)$. The primary current $\mathbf{J}^{\mathrm{p}}(r)$ can then be regarded as current dipoles and approximated by the summation of N current dipoles,

$$\mathbf{J}_n^{\mathrm{p}}(r) = Q_n \delta(r - r_n),$$

where $\delta(\cdot)$ is the Dirac delta function and Q_n is a charge dipole at location r_n, for $n = 1, \ldots, N$.

The forward problem in neuromagnetism focuses on calculating the magnetic field outside of the brain from a given primary current $\mathbf{J}_n^{\mathrm{p}}$ within the brain. Using the quasistatic approximation of the Maxwell's equations (Sarvas (1984)), the magnetic field generated by a primary current $\mathbf{J}_n^{\mathrm{p}}$ in the head is approximated by the Biot–Savart equation

$$\mathbf{B}_l(\mathbf{J}_n^{\mathrm{p}}) = \frac{\mu_0}{4\pi} \int_{\Omega} \frac{\mathbf{J}_n^{\mathrm{p}}(r') \times (r_l - r')}{|r_l - r'|^3} \, dv, \tag{1.1}$$

where r_l is the location at which the magnetic field is computed, r' refers to the source location within the relevant source region, μ_0 is the permittivity of free space and Ω is the head volume. Here, the integral in (1.1) as a function of r_l is the solution of Maxwell's equations with $\mathbf{B}_l(\mathbf{J}_n^{\mathrm{p}}) \to 0$ as $|r_l| \to \infty$. Furthermore, the magnetic field at location r_l of N current dipoles is the summation of $\mathbf{B}_l(\mathbf{J}_n^{\mathrm{p}})$ over each dipole $n = 1, \ldots, N$.

The MEG inverse problem is to infer the source current given the measured magnetic fields collected from the MEG experiment. However, the solution of the inverse problem is not unique since the measured magnetic fields could result from an infinite number of possible source currents. The solution is also highly sensitive to small changes in the noise of the measured data. Because of the nonuniqueness and instability, the general inverse problem is ill posed, and this fact makes the MEG inverse problem challenging to solve. Two types of models have been developed for the MEG inverse problem (Baillet, Mosher and Leahy (2001)), equivalent current dipole (ECD) models and distributed source models. The ECD models are based on the assumption that the locations of the current dipoles are unknown and have to be estimated. On the other hand, the distributed source models assume that the measured magnetic fields are generated from the current dipoles with known locations.

1.2. *Existing source localization methods.* The problem of source localization aims to get a meaningful structure of the source current for the inverse problem. In the literature two categories of methods focusing on addressing the challenging source localization were proposed. The first category assumes that the current source is static during the MEG scans which allows us to solve the inverse problem at each time point independently using the quasistatic approximation. Several existing methods were proposed to investigate the current source under the distributed source model and interpret the pattern from the observed magnetic fields. The L_2 norm and its variations were implemented to solve the distributed source current by using the regularization methods, including the minimum norm estimate (MNE) (Hämäläinen and Ilmoniemi (1994)), minimum current estimate (Uutela, Hämäläinen and

Somersalo (1999)), depth-weighted minimum norm estimate (Lin et al. (2006)) and low-resolution electromagnetic tomography algorithm (Pascual-Marqui, Michel and Lehmann (1994)). The MUltiple SIgnal Classification (MUSIC) algorithm (Mosher, Lewis and Leahy (1992)) is a subspace scanning method in which the solution is found by searching a single source current through the three-dimensional head volume entirely and projected to an estimate of the signal subspace. The beamforming methods assume that the source currents are uncorrelated, and the goal of the beamformers is to find a set of filter coefficients of the measured magnetic fields, subjected to some constraints. The Linearly Constrained Minimum Variance (LCMV) beamforming method in Veen, Joseph and Hecox (1992), which was first applied to the inverse problem, is operated by searching a selected region of the head volume to analyze the source current distribution subjected to the minimum variant constraint.

The second category of the methods on source localization incorporates the source current with a time-varying structure. By assuming the variability of source activity, it is able to investigate the source current at each time point t, $\mathbf{J}_t^{\mathrm{p}}$, and provide the temporal evolution of the source current during the MEG scans. The spatiotemporal regularization was utilized in Ou, Hämäläinen and Golland (2009) and improves the reconstruction accuracy of the distributed source current. In Long et al. (2011), a dynamic state-space model was proposed to model the movement of the sources, and the Kalman filter (KF) and fixed interval smoother (FIS) were used to solve the high-dimensional state estimation. The beamforming method with spatial and temporal effects was proposed to summarize the information of the sources during the voxel-based searching of the head volume; see Zhang and Liu (2015), Zhang and Su (2015). Recent work has addressed the source localization of time-varying currents as part of a Bayesian framework. In Baillet and Garnero (1997), a Bayesian approach with a maximum a posteriori (MAP) estimator of source activities was built in the distributed source model. The variational Bayesian learning algorithm was derived to reconstruct the distributed sources in the probabilistic generative model; see Fukushima et al. (2015), Trujillo-Barreto, Aubert-Vázquez and Penny (2008). In Lamus et al. (2012), the authors developed a dynamic Maximum Posterior Expectation-Maximization (dMAP-EM) source localization algorithm, based on the KF, FIS and EM algorithm, to obtain a spatiotemporal distributed solution for the source current. Two sequential important sampling (SIS) (Liu and Chen (1998)) based methods, the regular SIS method with rejection and improved SIS method with resampling, were developed in Yao and Eddy (2014) to address the source localization in the ECD models. These authors investigated the source localization by finding the marginal posterior distribution of the source current, given the measured magnetic fields, thus providing the variation of the location, orientation and strength of the source current at each time point.

The magnetic fields (1.1), generated from the current dipole Q_n, can be approximately represented by the Biot–Savart in a discrete matrix form,

$$\mathbf{B}_l\big(\mathbf{J}_n^{\mathrm{p}}\big) = \mathbf{L}_l(r_n) \cdot Q_n, \tag{1.2}$$

where $\mathbf{L}_l(r')$ is the lead field describing the magnetic fields generated by a unit dipole (with unit moment parameter) satisfying $\mathbf{B}_l(\mathbf{J}_n^{\mathrm{p}}) = \int_\Omega \mathbf{L}_l(r') \cdot \mathbf{J}_n^{\mathrm{p}}(r')\,dv$. In the distributed source models, the location parameter r_n of the source current Q_n is assumed to be known; thus, the lead field in (1.2) can be calculated from the forward model. In this case the magnetic fields linearly depend on the current dipole Q_n. Several approaches, such as the KF and FIS, were proposed to estimate the current dipole Q_n, and they work well under the linear dynamic system. However, the magnetic fields also nonlinearly depend on the unknown location parameter r_n; thus, the recovery of the location parameter usually involves nonlinear optimization, in which applying the KF would degrade the performance (Arulampalam et al. (2002)). The existing work on time-varying source current with unknown location parameter is limited, and this motivates us to develop new approaches to investigate the source distribution for the location parameter.

1.3. *Goal of this paper.* The goal of this paper is to invent a new Bayesian framework to find the posterior distribution of the source current $\mathbf{J}_t^{\mathrm{p}}$ at time point t, given the entire collection of measurements $\mathcal{Y}_T$, which consists of measurements $\mathbf{Y}_t$ for $1 \le t \le T$. The posterior distribution $p(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_T)$ for the source current $\mathbf{J}_t^{\mathrm{p}}$ can be interpreted as a solution for the MEG inverse problem. In contrast to the existing literature, our proposed methodology is novel based on the following two aspects. First, we develop a discrete approach for computing the discrete posterior distribution of the source current $\mathbf{J}_t^{\mathrm{p}}$, and the discrete posterior distribution is used to approximate the continuous posterior distribution $p(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_T)$. The SIS schemes in Yao and Eddy (2014) investigated the source distribution by numerically sampling the continuous posterior distribution. However, the posterior distribution in Yao and Eddy (2014) does not have an analytically tractable closed form, and the sampling procedure may suffer from slow convergence due to the high dimensionality of the measurements. In comparison with the sampling schemes, our method gives the discrete posterior distribution for the source current with a closed form and are able to calculate it directly, even when the dimensionality of the measurements is large. Second, we use the entire collection of measurements to investigate the source distribution $p(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_T)$ instead of the source distribution $p(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_t)$, using only the measurements up to the time point in which we are interested. The MEG allows for a real-time recording of the brain activity on a millisecond scale. For each time point t, the past measurements and the future measurements both reflect the trajectory of the time-varying source $\mathbf{J}_t^{\mathrm{p}}$, for $1 \le t \le T$. In contrast to previous related approaches (Baillet and Garnero (1997), Trujillo-Barreto, Aubert-Vázquez and Penny (2008), Yao and Eddy (2014)), we utilize the entire set of measurements to recover the location of the source current.

For the proposed discrete approach, we focus on the selected three-dimensional region of interest (ROI), and the ROI is subsequently discretized into K voxels $\{V_k\}_{k=1}^K$. Then, we calculate the discrete posterior distribution $\mathrm{P}(\mathbf{J}_t^{\mathrm{p}} \in V_k|\mathcal{Y}_T)$ of the source current $\mathbf{J}_t^{\mathrm{p}}$ at each time point t, for all $1 \le k \le K$. Figure 6 presents the posterior distribution for the location parameter of a single source at six selected time points, where the probabilities are highlighted in different colors. The discrete posterior distribution indicates that the source would appear in the voxel with a corresponding probability. The region with nonzero probabilities can be interpreted as the activated area of the source current. Thus, the discrete approach is able to provide the source distribution on time evolution during the MEG scanning.

In order to calculate the discrete posterior distribution, the EM algorithm with incomplete data is implemented to estimate the parameters in the source model. We further develop a switch procedure and dynamic procedure to implement the proposed discrete approach. The switch procedure is proposed to deal with the case involving multiple sources, and the dynamic procedure is developed to balance the estimation accuracy and computational efficiency when no available information on the ROI is provided. With the proposed approach, we will be able to investigate both MEG and EEG recordings that contain valuable time-sensitive information and shed light on the time evolution of the source localization.

1.4. *Outline of this paper.* In Section 2 we present the methodology of the discrete approach used to recover the source distribution. First, we utilize a dynamic spatiotemporal model to reformulate the source localization problem in Section 2.1. After which, we propose the discrete approach for the calculation of the posterior distribution for the location parameter of the source current in Section 2.2. An estimation procedure for the parameters, which are introduced by the source model, will be presented in Section 2.3. We further develop the switch procedure and dynamic procedure to implement the discrete approach for the calculation of the posterior distribution in Section 2.4 and Section 2.5. Simulation studies are described in Section 3. In Section 4 a MEG data application is presented. An extension to the EEG data is illustrated in Section 5. A short discussion and concluding remarks are given in Section 6.

2. Methodology.

2.1. *A dynamic spatiotemporal model.* In an MEG experiment, the observed magnetic fields are scanned at L sensors, and the data is recorded for a fixed time period, T milliseconds. Let $\mathbf{Y}_t = (Y_{t,1}, \ldots, Y_{t,L})^{\mathrm{T}}$ be the measurements of L sensors collected at time point t, $1 \le t \le T$, and allow $\mathcal{Y}_T = (\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ be the entire collection of measurements of the experiment. The magnetic field measured from the lth sensor at time point t is

$$Y_{t,l} = \mathbf{B}_l(\mathbf{J}_t^{\mathrm{p}}) + U_{t,l}, \quad 1 \le l \le L, 1 \le t \le T,$$

where $U_{t,l}$ denotes the noise of the measurements. Homogeneous conductor models for the human head are widely used for MEG studies (Sarvas (1984)). In this paper we consider a current source in a spherically symmetric conductor. The magnetic field $\mathbf{B}_l(\mathbf{J}_t^{\mathrm{p}})$ in (1.1) can be easily computed as

$$B_l(\mathbf{J}_t^{\mathrm{p}}) = \frac{\mu_0}{4\pi} \frac{\mathbf{q}_t \times (\mathbf{r}_l - \mathbf{p}_t) \cdot \mathbf{e}_z}{\|\mathbf{r}_l - \mathbf{p}_t\|^3}, \quad 1 \le l \le L, \tag{2.1}$$

where the source current $\mathbf{J}_t^{\mathrm{p}} = (\mathbf{p}_t^{\mathrm{T}}, \mathbf{q}_t^{\mathrm{T}})^{\mathrm{T}}$ is mathematically parameterized with location parameter $\mathbf{p}_t$ and moment parameter $\mathbf{q}_t$, $\mathbf{p}_t = (p_{t,1}, p_{t,2}, p_{t,3})^{\mathrm{T}}$ contains the location parameters in the three-dimensional head model, $\mathbf{q}_t = (q_{t,1}, q_{t,2}, q_{t,3})^{\mathrm{T}}$ contains the moments and strength of the source at time t, $\mathbf{r}_l$ is the location of the lth sensor placed on the scalp. Equation (2.1) may also apply to a dipole in a horizontally layered conductor. Note that MEG is sensitive only to the tangential component of the source current. This is because a source within the head pointing toward the sensor would produce a magnetic field parallel to the sensor which would, therefore, not be measured. We also remark that because the magnetometers measure only the z-component of the magnetic field, $\mathbf{e}_z = (0, 0, 1)$; a unit vector, is used to find the z-component of $\mathbf{B}_l(\mathbf{J}_t^{\mathrm{p}})$ (Hämäläinen et al. (1993)).

REMARK 2.1. All the methodology developed in this paper is legitimately extendable to the EEG source analysis. Speaking of the extension to EEG, the potential field generated by the source currents is used to replace the magnetic field defined in (2.1). Similarly, the potential field generated by a source $\mathbf{J}_t^{\mathrm{p}}$ at the lth sensor is given by

$$H_l(\mathbf{J}_t^{\mathrm{p}}) = \frac{1}{4\pi\rho} \frac{\mathbf{q}_t \cdot (\mathbf{r}_l - \mathbf{p}_t)}{\|\mathbf{r}_l - \mathbf{p}_t\|^3}, \tag{2.2}$$

where ρ is the conductivity and $1 \le l \le L$.

Let $\mathbf{B}(\mathbf{J}_t^{\mathrm{p}}) = (B_1(\mathbf{J}_t^{\mathrm{p}}), \ldots, B_L(\mathbf{J}_t^{\mathrm{p}}))^{\mathrm{T}}$ be the collection of magnetic fields with source current $\mathbf{J}_t^{\mathrm{p}}$ generated at L sensors. Then, the general framework of the MEG measurement $\mathbf{Y}_t$ is given by

$$\mathbf{Y}_t = \mathbf{B}(\mathbf{J}_t^{\mathrm{p}}) + \mathbf{U}_t, \quad 1 \le t \le T, \tag{2.3}$$

where $\mathbf{U}_t$ contains environmental noise independent from $\mathbf{J}_t^{\mathrm{p}}$ and $\mathbf{U}_t = (U_{t,1}, \ldots, U_{t,L})^{\mathrm{T}}$ is Gaussian noise vector with zero mean and covariance matrix $\mathbf{V}$. For simplicity, we assume that the noises $U_{t,l}$, $1 \le l \le L$ are uncorrelated between every pair of sensors and homogeneous. In this case the covariance matrix $\mathbf{V}$ is a diagonal matrix, $\mathbf{V} = \mathrm{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_L^2)$.

Evidence of neurophysiology, biophysics and neuroimaging illustrates that cortical activation is a spatiotemporal dynamic process (Destexhe, Contreras and Steriade (1999)). One way to model the spatiotemporal connections of the source $\mathbf{J}_t^{\mathrm{p}}$ is to use a autoregressive model. In the first order vector autoregressive model, the evolution of the source $\mathbf{J}_t^{\mathrm{p}}$ is a function of the source at the previous time point,

$$\mathbf{J}_t^{\mathrm{p}} = A\mathbf{J}_{t-1}^{\mathrm{p}} + \mathbf{b} + \mathbf{Z}_t, \quad 2 \le t \le T, \tag{2.4}$$

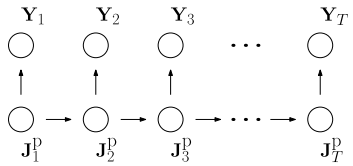


FIG. 1. *Illustration of the causal relationship between the MEG observations $\mathcal{Y}_T$ and source $\mathcal{J}_T^p$.*

where A is 6×6 coefficient matrix, $\mathbf{b}$ is 6×1 constant vector and the evolution noise $\mathbf{Z}_t$ is assumed to be Gaussian, $\mathbf{Z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_6^2)$. At the first time step we assume that $\mathbf{J}_1^{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0)$. The autoregressive model (2.4) is stable if the roots of $\Phi(x) = \det(\mathbf{I} - Ax) = 0$ lie outside the unit circle. Let $\mathcal{J}_T^{\mathrm{p}}$ be the collection of source $\mathbf{J}_t^{\mathrm{p}}$ for $1 \le t \le T$. The causal relationship between the MEG measurements $\mathcal{Y}_T$ and source $\mathcal{J}_T^{\mathrm{p}}$ is described in Figure 1.

Throughout the paper the framework is mainly based on a single source and can be generalized to the case with multiple sources. Since estimating the number of sources is not the focus of the paper, the number of sources N is determined by the method proposed in Yao et al. (2018). We extend $\mathbf{J}_t^{\mathrm{p}}$ to be the collection of N current sources, where $\mathbf{J}_t^{\mathrm{p}} = (\mathbf{J}_{t,1}^{\mathrm{p}\ \mathrm{T}}, \ldots, \mathbf{J}_{t,N}^{\mathrm{p}\ \mathrm{T}})^{\mathrm{T}}$, and $\mathbf{J}_{t,n}^{\mathrm{p}}$ is the nth source parameterized with $(\mathbf{p}_{t,n}^{\mathrm{T}}, \mathbf{q}_{t,n}^{\mathrm{T}})^{\mathrm{T}}$ for $1 \le n \le N$. Then, the magnetic field in (2.1) is generated from N sources and is given by $\mathbf{B}_l(\mathbf{J}_t^{\mathrm{p}}) = \sum_{n=1}^{N} \mathbf{B}_l(\mathbf{J}_{t,n}^{\mathrm{p}})$. We further assume that the N sources are uncorrelated, and each source $\mathbf{J}_{t,n}^{\mathrm{p}}$ is modeled with the first order autoregressive model,

$$\mathbf{J}_{t,n}^{\mathrm{p}} = A_n \mathbf{J}_{t-1,n}^{\mathrm{p}} + \mathbf{b}_n + \mathbf{Z}_{t,n}, \quad 2 \le t \le T, 1 \le n \le N,$$

where A_n is the coefficient matrix satisfying the stable condition and $\mathbf{b}_n$ is a constant vector, $\mathbf{J}_{1,n}^{\mathrm{p}} \sim \mathcal{N}(\boldsymbol{\mu}_{0,n}, \mathbf{\Sigma}_{0,n})$, $\mathbf{Z}_{t,n} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_n)$. Thus, the general framework of N uncorrelated sources $\mathbf{J}_t^{\mathrm{p}}$ is consistent with model (2.4), where the parameters $\boldsymbol{\mu}_0$, $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}$, A and $\mathbf{b}$ contain the information of N sources, correspondingly. To be precise, $A = \mathrm{diag}(A_1, \ldots, A_N)$, $\mathbf{b} = (\mathbf{b}_1^{\mathrm{T}}, \ldots, \mathbf{b}_N^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\mu}_0 = (\boldsymbol{\mu}_{0,1}^{\mathrm{T}}, \ldots, \boldsymbol{\mu}_{0,N}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{\Sigma}_0 = \mathrm{diag}(\mathbf{\Sigma}_{0,1}, \ldots, \mathbf{\Sigma}_{0,N})$ and $\mathbf{\Sigma} = \mathrm{diag}(\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_N)$.

Let $\mathbf{\Theta} = \{\boldsymbol{\mu}_0, \mathbf{\Sigma}_0, A, \mathbf{b}, \mathbf{\Sigma}, \mathbf{V}\}$ be the list of parameters introduced in the framework of measurements (2.3) and source model (2.4). From the causal relationship described in Figure 1, we note that the sequences of source $\mathcal{J}_T^{\mathrm{p}}$ and measurements $\mathcal{Y}_T$ have the following Markov properties:

1. $p(\mathbf{J}_t^{\mathrm{p}} | \mathcal{J}_{t-1}^{\mathrm{p}}, \mathbf{\Theta}) = p(\mathbf{J}_t^{\mathrm{p}} | \mathbf{J}_{t-1}^{\mathrm{p}}, \mathbf{\Theta})$, for $2 \le t \le T$.
2. $p(\mathbf{J}_t^{\mathrm{p}} | \mathbf{J}_{t-1}^{\mathrm{p}}, \mathcal{Y}_{t-1}, \mathbf{\Theta}) = p(\mathbf{J}_t^{\mathrm{p}} | \mathbf{J}_{t-1}^{\mathrm{p}}, \mathbf{\Theta})$, for $2 \le t \le T$.
3. $p(\mathbf{Y}_t | \mathcal{J}_T^{\mathrm{p}}, \mathbf{\Theta}) = p(\mathbf{Y}_t | \mathbf{J}_t^{\mathrm{p}}, \mathbf{\Theta})$, for $1 \le t \le T$.

Throughout the paper we use p as a genetic symbol for continuous probability distribution.

2.2. *A discrete approach.* To interpret MEG data, we aim to investigate the source distribution for location parameter $\mathbf{p}_t$ given the externally measured magnetic fields. Thus, we focus on computing the posterior distribution of $\mathbf{J}_t^{\mathrm{p}}$, given the entire measurements $\mathcal{Y}_T$, for $1 \le t \le T$. In this section we build a discrete approach to approximate the posterior distribution $p(\mathbf{J}_t^{\mathrm{p}} | \mathcal{Y}_T, \mathbf{\Theta})$, reformulate the discrete posterior distribution with a closed form and present the forward-backward algorithm to further compute it. The location parameter of interest is $\mathbf{p}_t$ of the source current $\mathbf{J}_t^{\mathrm{p}}$, for $1 \le t \le T$; thus, the moment and strength parameter $\mathbf{q}_t$ is fixed for all the time points throughout this paper.

With an available ROI the movement of location parameter $\mathbf{p}_t$ of source $\mathbf{J}_t^{\mathrm{p}}$ is assumed to be restricted within the ROI at all times. For the discrete approach, each dimension of

the ROI is discretized with mesh grid K_i, $i = 1, 2, 3$. The mesh grids are used to construct a sequence of voxels $\{V_k\}_{k=1}^K$ and approximate the three-dimensional ROI, where V_k is the kth voxel with its center $\mathbf{c}_k$ and $K = K_1 \cdot K_2 \cdot K_3$. To investigate the source current $\mathbf{J}_t^{\mathrm{p}}$, we introduce a corresponding set of binary indicator variables $v_{tk} \in \{0, 1\}$, where $k = 1, \ldots, K$, describing which of the K voxels the source $\mathbf{J}_t^{\mathrm{p}}$ is located in, so that if $\mathbf{J}_t^{\mathrm{p}} \in V_k$, then $v_{tk} = 1$ and $v_{tk'} = 0$ for $k' \neq k$. Applying the coding scheme, we can then approximate the continuous probabilities introduced in (2.3) and (2.4) using the corresponding discrete probabilities. For the continuous probability $p(\mathbf{J}_1^{\mathrm{p}}|\boldsymbol{\Theta})$, there exists $1 \leq k^* \leq K$, such that

$$\begin{aligned} p\big(\mathbf{J}_1^{\mathrm{p}}|\boldsymbol{\Theta}\big) &\approx \mathrm{P}\big(\mathbf{J}_1^{\mathrm{p}} \in V_{k^*}|\boldsymbol{\Theta}\big) = \prod_{k=1}^{K} \mathrm{P}\big(\mathbf{J}_1^{\mathrm{p}} \in V_k|\boldsymbol{\Theta}\big)^{v_{1k}} \\ &= \prod_{k=1}^{K} \mathrm{P}(v_{1k} = 1|\boldsymbol{\Theta})^{v_{1k}}. \end{aligned} \tag{2.5}$$

Similarly, we have the following approximations:

$$p\big(\mathbf{J}_t^{\mathrm{p}}|\mathbf{J}_{t-1}^{\mathrm{p}}, \boldsymbol{\Theta}\big) \approx \prod_{k=1}^{K}\prod_{l=1}^{K} \mathrm{P}(v_{tk} = 1|v_{t-1,l} = 1, \boldsymbol{\Theta})^{v_{tk}v_{t-1,l}} \tag{2.6}$$

for $2 \leq t \leq T$, and

$$p\big(\mathbf{Y}_t|\mathbf{J}_t^{\mathrm{p}}, \boldsymbol{\Theta}\big) \approx \prod_{k=1}^{K} \mathrm{P}(\mathbf{Y}_t|v_{tk} = 1, \boldsymbol{\Theta})^{v_{tk}} \tag{2.7}$$

for $1 \leq t \leq T$.

From Bayes' theorem, the posterior distribution is stated as

$$p\big(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}\big) = \frac{p(\mathbf{J}_t^{\mathrm{p}}, \mathcal{Y}_t|\boldsymbol{\Theta})p(\mathcal{Y}_{T\setminus t}|\mathbf{J}_t^{\mathrm{p}}, \boldsymbol{\Theta})}{p(\mathcal{Y}_t|\boldsymbol{\Theta})p(\mathcal{Y}_{T\setminus t}|\boldsymbol{\Theta})}, \quad 1 \leq t \leq T, \tag{2.8}$$

where $\mathcal{Y}_t$ is the collection of the measurements up to time point t and $\mathcal{Y}_{T\setminus t}$ contains the remaining measurements from time point $t+1$ to T. Under the discrete approach the posterior distribution in (2.8) can be approximated by

$$\begin{aligned} \mathrm{P}\big(\mathbf{J}_t^{\mathrm{p}} \in V_k|\mathcal{Y}_T, \boldsymbol{\Theta}\big) &= \mathrm{P}(v_{tk} = 1|\mathcal{Y}_T, \boldsymbol{\Theta}) \\ &= \frac{\mathrm{P}(v_{tk} = 1, \mathcal{Y}_t|\boldsymbol{\Theta})\mathrm{P}(\mathcal{Y}_{T\setminus t}|v_{tk} = 1, \boldsymbol{\Theta})}{p(\mathcal{Y}_t|\boldsymbol{\Theta})p(\mathcal{Y}_{T\setminus t}|\boldsymbol{\Theta})}, \end{aligned}$$

for $1 \leq k \leq K$ and $1 \leq t \leq T$. Let $\alpha_{tk}(\boldsymbol{\Theta}) = \mathrm{P}(v_{tk} = 1, \mathcal{Y}_t|\boldsymbol{\Theta})/p(\mathcal{Y}_t|\boldsymbol{\Theta})$, $\beta_{tk}(\boldsymbol{\Theta}) = \mathrm{P}(\mathcal{Y}_{T\setminus t}|v_{tk} = 1, \boldsymbol{\Theta})/p(\mathcal{Y}_{T\setminus t}|\boldsymbol{\Theta})$, and we have

$$\mathrm{P}\big(\mathbf{J}_t^{\mathrm{p}} \in V_k|\mathcal{Y}_T, \boldsymbol{\Theta}\big) = \alpha_{tk}(\boldsymbol{\Theta})\beta_{tk}(\boldsymbol{\Theta}). \tag{2.9}$$

Therefore, the calculation of the discrete posterior distribution (2.9) consists of two parts, a filtering procedure on $\alpha_{tk}(\boldsymbol{\Theta})$ using the past measurements $\mathcal{Y}_t$ and a smoothing procedure on $\beta_{tk}(\boldsymbol{\Theta})$ using the remaining measurements $\mathcal{Y}_{T\setminus t}$. The forward-backward algorithm (Rabiner (1989)) is implemented to compute $\alpha_{tk}(\boldsymbol{\Theta})$ and $\beta_{tk}(\boldsymbol{\Theta})$ in (2.9), and the details of the algorithm are summarized in Appendix A.

Each step of the forward-backward recursion involves multiplying by the probabilities in (2.5)–(2.7), and the computational complexity of calculating these probability matrices are $O(K)$, $O(K^2)$ and $O(L)$, correspondingly. Therefore, the overall cost of the discrete approach is of $O(LK^2T)$ in time. However, the entry m_{kl} of the $K \times K$ matrix for transition probability in (2.6) is proportional to $\exp(-(V_k - V_l)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(V_k - V_l))$. That is, $m_{kl} \neq 0$ over

TABLE 1
Comparison of computational complexity

Proposed discrete method	MNE	SIS
$O(LKT)$	$O(L^2KT)$	$O(LKT)$

some interval, and $m_{kl} = 0$ outside the interval. The complexity on calculating the $K \times K$ transition probability matrix can be reduced to $O(K)$ (Felzenszwalb, Huttenlocher and Kleinberg (2004)), and, thereby, the complexity of the discrete approach can be computed in linear time $O(LKT)$. In comparison with other existing methods, we summarize the computational complexity for three methods in Table 1. To make the complexity comparable, we consider a distributed source model with K known locations for implementing MNE, although the number of known locations usually takes value of several thousands. The complexity of obtaining MNE at one selected time point is $O(L^2K)$. In Table 1 we report the complexity for MNE at all time points. To implement the sampling scheme, we consider K samples at each time point to make the procedure comparable; computing the weight for each sample takes $O(L)$ in time. The overall SIS procedure takes time $O(LKT)$.

2.3. *Parameter estimation*. Under the models (2.3) and (2.4), the posterior distribution $p(\mathbf{J}_t^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta})$ depends on the parameter $\boldsymbol{\Theta}$. To obtain the discrete posterior distribution in (2.9), an estimate for the parameter $\boldsymbol{\Theta}$ needs to be determined. In this section we apply the EM algorithm (Dempster, Laird and Rubin (1977)) to find a MLE $\hat{\boldsymbol{\Theta}}$ of the parameter $\boldsymbol{\Theta}$ with incomplete MEG data $(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}})$, as we have no access to the collection of the source $\mathcal{J}_T^{\mathrm{p}}$ during the MEG scans. The optimization problem is defined as

$$\hat{\boldsymbol{\Theta}} = \operatorname*{argmax}_{\boldsymbol{\Theta}} \ell(\boldsymbol{\Theta}, \mathcal{Y}_T), \tag{2.10}$$

where $\ell(\boldsymbol{\Theta}, \mathcal{Y}_T)$ is the log-likelihood function of parameter $\boldsymbol{\Theta}$, given the entire set of measurements $\mathcal{Y}_T$. Under the discrete approach in Section 2.2, the unobserved source $\mathcal{J}_T^{\mathrm{p}}$ is assumed to be the discrete variable; thus, the log-likelihood function in (2.10) can be written into the following form:

$$\ell(\boldsymbol{\Theta}, \mathcal{Y}_T) = \log p(\mathcal{Y}_T|\boldsymbol{\Theta}) = \log \sum_{\mathcal{J}_T^{\mathrm{p}}} p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta}),$$

where

$$p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta}) = \prod_{t=1}^{T} p(\mathbf{Y}_t|\mathbf{J}_t^{\mathrm{p}}, \boldsymbol{\Theta}) \prod_{t'=2}^{T} p(\mathbf{J}_{t'}^{\mathrm{p}}|\mathbf{J}_{t'-1}^{\mathrm{p}}, \boldsymbol{\Theta}) p(\mathbf{J}_1^{\mathrm{p}}|\boldsymbol{\Theta}), \tag{2.11}$$

under the Markov properties of $\mathcal{Y}_T$ and $\mathcal{J}_T^{\mathrm{p}}$. Applying Jensen's inequality, we have

$$\begin{aligned}
\ell(\boldsymbol{\Theta}, \mathcal{Y}_T) &= \log \sum_{\mathcal{J}_T^{\mathrm{p}}} q(\mathcal{J}_T^{\mathrm{p}}) \frac{p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta})}{q(\mathcal{J}_T^{\mathrm{p}})} \\
&\geq \sum_{\mathcal{J}_T^{\mathrm{p}}} q(\mathcal{J}_T^{\mathrm{p}}) \log \frac{p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta})}{q(\mathcal{J}_T^{\mathrm{p}})} \qquad (2.12) \\
&= \sum_{\mathcal{J}_T^{\mathrm{p}}} q(\mathcal{J}_T^{\mathrm{p}}) \log p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta}) - \sum_{\mathcal{J}_T^{\mathrm{p}}} q(\mathcal{J}_T^{\mathrm{p}}) \log q(\mathcal{J}_T^{\mathrm{p}}) \\
&=: L(q(\mathcal{J}_T^{\mathrm{p}}), \boldsymbol{\Theta}, \mathcal{Y}_T), \qquad (2.13)
\end{aligned}$$

where $q(\mathcal{J}_T^{\mathrm{p}})$ is a probability distribution on the unobserved variables $\mathcal{J}_T^{\mathrm{p}}$. From (2.12) we have

$$\ell(\boldsymbol{\Theta}, \mathcal{Y}_T) \geq L\big(q(\mathcal{J}_T^{\mathrm{p}}), \boldsymbol{\Theta}, \mathcal{Y}_T\big) \tag{2.14}$$

for any probability distribution $q(\mathcal{J}_T^{\mathrm{p}})$. When $q(\mathcal{J}_T^{\mathrm{p}}) = \mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta})$, the equality in (2.12) holds,

$$\ell(\boldsymbol{\Theta}, \mathcal{Y}_T) = L\big(\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}), \boldsymbol{\Theta}, \mathcal{Y}_T\big) = \log p(\mathcal{Y}_T|\boldsymbol{\Theta}). \tag{2.15}$$

To maximize the log-likelihood function $\ell(\boldsymbol{\Theta}, \mathcal{Y}_T)$ in (2.10), we construct an alternating EM algorithm, maximizing $L(\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}), \boldsymbol{\Theta}, \mathcal{Y}_T)$ as defined in (2.13). The EM algorithm is summarized in Appendix B.

Under the following regularity conditions:

(C1) The parameter space Ω is an open set in the Euclidean space.

(C2) The density function $p(y|\boldsymbol{\Theta})$ of the measurements $\mathbf{Y}_t$, $1 \leq t \leq T$, is continuous in Ω and differentiable in the interior of Ω.

(C3) The level set $\Omega_{\boldsymbol{\Theta}_0} = \{\boldsymbol{\Theta} \in \Omega : \ell(\boldsymbol{\Theta}, \mathcal{Y}_T) \geq \ell(\boldsymbol{\Theta}_0, \mathcal{Y}_T)\}$ is compact for any $\ell(\boldsymbol{\Theta}_0, \mathcal{Y}_T) > -\infty$, and $\Omega_{\boldsymbol{\Theta}_0}$ is in the interior of Ω for any $\boldsymbol{\Theta}_0 \in \Omega$.

(C4). The distribution of the unobserved variables $\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta})$ has the same support for all $\boldsymbol{\Theta} \in \Omega$.

(C5) The function $Q(\boldsymbol{\Theta}'|\boldsymbol{\Theta})$ is continuous in both $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}'$ and differentiable in $\boldsymbol{\Theta}'$.

(C6) All the stationary points in $\mathcal{S}_\ell$ are isolated, where $\mathcal{S}_\ell$ denotes the stationary points of the log-likelihood function $\ell(\boldsymbol{\Theta}, \mathcal{Y}_T)$.

(C7) For all $\boldsymbol{\Theta} \in \mathcal{S}_\ell$, there exists a unique global maximum of $Q(\cdot|\boldsymbol{\Theta})$.

We prove the convergence of the EM sequence $\{\boldsymbol{\Theta}^{(j)}\}$ in Theorem 2.1 and the proof can be found in Appendix C.

THEOREM 2.1. *The iterative procedure of EM algorithm does not cause a decrease in the log-likelihood function* $\ell(\boldsymbol{\Theta}, \mathcal{Y}_T)$. *Furthermore, we assume that* (C1)–(C7) *hold. Then, for any starting value* $\{\boldsymbol{\Theta}^{(0)}\}$, *the EM sequence* $\{\boldsymbol{\Theta}^{(j)}\}$, $\boldsymbol{\Theta}^{(j)} \to \boldsymbol{\Theta}^*$, *when* $j \to \infty$, *for some stationary point* $\boldsymbol{\Theta}^* \in \mathcal{S}_\ell$.

2.4. *A switch procedure.* When we apply the discrete approach to the case with multiple sources, the possible states for N sources go to K^N, where K is the number of discrete voxels. In the Supplementary Material (Yao et al. (2020)), we compare the possible states for N sources with mesh grids in the discretization. Even with the mesh grids $K_i = 6$, $i = 1, 2, 3$, the calculation for the discrete posterior distribution needs to include more than 40,000 possible states which makes the computation procedure impossible. Thus, a method that enables a more achievable posterior calculation of multiple sources is desirable.

In this section we propose a switch procedure to reduce the computational complexity for the case with multiple sources. Applying the Bayes' rule, we have

$$\begin{aligned}
&\mathrm{P}\big(\mathbf{J}_{t,n}^{\mathrm{p}} \in V_{k_n} | \mathbf{J}_{t,n'}^{\mathrm{p}} \in V_{k_{n'}}, \mathcal{Y}_T, \boldsymbol{\Theta}, n' \neq n\big) \\
&\quad = \frac{\mathrm{P}(\mathbf{J}_{t,1}^{\mathrm{p}} \in V_{k_1}, \ldots, \mathbf{J}_{t,N}^{\mathrm{p}} \in V_{k_N} | \mathcal{Y}_T, \boldsymbol{\Theta})}{\mathrm{P}(\mathbf{J}_{t,n'}^{\mathrm{p}} \in V_{k_{n'}}, n' \neq n | \mathcal{Y}_T, \boldsymbol{\Theta})} \\
&\quad \propto \mathrm{P}\big(\mathbf{J}_{t,1}^{\mathrm{p}} \in V_{k_1}, \ldots, \mathbf{J}_{t,N}^{\mathrm{p}} \in V_{k_N} | \mathcal{Y}_T, \boldsymbol{\Theta}\big)
\end{aligned}$$

for $1 \leq n \leq N$. To reduce the computational complexity, the marginal posterior probability $\mathrm{P}(\mathbf{J}_{t,n}^{\mathrm{p}} \in V_{k_n} | \mathbf{J}_{t,n'}^{\mathrm{p}} \in V_{k_{n'}}, \mathcal{Y}_T, \boldsymbol{\Theta}, n' \neq n)$, $1 \leq n \leq N$, is used to approximate the posterior distribution of N sources, and we calculate the marginal posterior distribution of the

source $\mathbf{J}^{\mathrm{p}}_{t,n}$ by assuming that the states of other sources $\mathbf{J}^{\mathrm{p}}_{t,n'}$ are known. During the EM iterations, we let $\zeta_{t,k_n}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-1)}) = \mathrm{P}(v_{t,k_n} = 1 | \mathbf{J}^{\mathrm{p}}_{t,n'} \in V_{k_{n'}}, \mathcal{Y}_T, \boldsymbol{\Theta}_{\mathrm{s}}^{(j-1)}, n' \neq n)$ at the jth iteration, where $\{\boldsymbol{\Theta}_{\mathrm{s}}^{(j)}\}$ is the EM sequence with the switch procedure. For $1 \leq n \leq N$, we compute the marginal posterior probability by

$$
\begin{aligned}
\zeta_{t,k_n}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-1)}) = \mathrm{P}\Bigg(v_{t,k_n} = 1 \Big| \mathbf{J}^{\mathrm{p}}_{t,1} = \sum_{k_1=1}^{K} \zeta_{t,k_1}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-1)}) \mathbf{c}_{k_1}, \ldots, \\
\mathbf{J}^{\mathrm{p}}_{t,n-1} = \sum_{k_{n-1}=1}^{K} \zeta_{t,k_{n-1}}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-1)}) \mathbf{c}_{k_{n-1}}, \\
\mathbf{J}^{\mathrm{p}}_{t,n+1} = \sum_{k_{n+1}=1}^{K} \zeta_{t,k_{n+1}}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-2)}) \mathbf{c}_{k_{n+1}}, \ldots, \\
\mathbf{J}^{\mathrm{p}}_{t,N} = \sum_{k_N=1}^{K} \zeta_{t,k_N}(\boldsymbol{\Theta}_{\mathrm{s}}^{(j-2)}) \mathbf{c}_{k_N} \Bigg).
\end{aligned} \tag{2.16}
$$

When the EM iteration converges, we will use the estimated parameter $\hat{\boldsymbol{\Theta}}_{\mathrm{s}}$ to calculate the posterior distribution $\mathrm{P}(\mathbf{J}^{\mathrm{p}}_{t,1} \in V_{k_1}, \ldots, \mathbf{J}^{\mathrm{p}}_{t,N} \in V_{k_N} | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})$ which is approximated by the marginal posterior distribution $\mathrm{P}(\mathbf{J}^{\mathrm{p}}_{t,n} \in V_{k_n} | \mathbf{J}^{\mathrm{p}}_{t,n'} \in V_{k_{n'}}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}, n' \neq n)$ in the switch procedure.

The posterior distribution from the switch procedure is compared with the one from the nonswitch procedure in Theorem 2.2; the proof can be found in Appendix D.

THEOREM 2.2. *Under regularity conditions* (C1) *and* (C7), *there exists* $\varepsilon > 0$, *such that*

$$
\begin{aligned}
\Bigg| & \mathrm{P}(\boldsymbol{J}^{\mathrm{p}}_{t,n} \in V_{k_n} | \boldsymbol{J}^{\mathrm{p}}_{t,n'} \in V_{k'_n}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}, n' \neq n) \\
& - \sum_{k_{n'}=1}^{K} \mathrm{P}(\boldsymbol{J}^{\mathrm{p}}_{t,1} \in V_{k_1}, \ldots, \boldsymbol{J}^{\mathrm{p}}_{t,N} \in V_{k_N} | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \Bigg| \leq c\varepsilon,
\end{aligned}
$$

where $1 \leq n \leq N$, c *is a positive constant*, $\mathrm{P}(\boldsymbol{J}^{\mathrm{p}}_{t,n} \in V_{k_n} | \boldsymbol{J}^{\mathrm{p}}_{t,n'} \in V_{k'_n}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}, n' \neq n)$ *is the marginal posterior distribution of source* $\boldsymbol{J}^{\mathrm{p}}_{t,n}$ *obtained from the switch procedure*, $\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}$ *is the estimate obtained from the nonswitch procedure and* $\mathrm{P}(\boldsymbol{J}^{\mathrm{p}}_{t,1} \in V_{k_1}, \ldots, \boldsymbol{J}^{\mathrm{p}}_{t,N} \in V_{k_N} | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$ *is the posterior distribution of* N *sources from the nonswitch procedure*.

2.5. *A dynamic procedure*. When no information on the ROI is available, we have to discretize the whole head model. Implementing the discrete approach to the parameter estimation procedure, we note that the estimate depends on the discretization. Increasing mesh girds will improve the estimation accuracy of calculating the discrete posterior distribution of the source during the EM iterations. However, it will also increase the computational complexity. This motivates us to develop a dynamic procedure to implement the discrete approach in order to balance the estimation accuracy and computational complexity when no information on the ROI is available.

We assume that a true ROI, ROI_0, exists and is assumed to be restricted within the head model. The movement of the source current $\mathbf{J}^{\mathrm{p}}_t$ is assumed to be restricted within ROI_0 at all times. In the EM algorithm with the discrete approach, calculating the discrete posterior

distributions $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ and $\eta^{tk}_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})$ depends on the discretization in the E-step of the jth iteration. For the dynamic procedure we assume that a shrunken ROI, $\mathrm{ROI}^{(j)}$, exists and covers the true ROI. Utilizing the shrunken ROI, $\mathrm{ROI}^{(j)}$, with increased mesh grids, the dynamic procedure will improve the estimation accuracy of calculating the discrete posterior distributions $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ and $\eta^{tk}_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})$ in the E-step of the jth iteration. Then, the posterior distributions are used to update $\boldsymbol{\Theta}^{(j)}$, and the update will also be improved in the following M-step. By implementing the dynamic procedure, we expect to obtain an accurate estimate $\hat{\boldsymbol{\Theta}}_{\mathrm{d}}$ with a shrunken ROI.

In the dynamic procedure we calculate the intermediate posterior distribution $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ for the location parameter $\mathbf{p}_t$ based on the current ROI in the jth iteration. After which, we use the intermediate marginal posterior distribution $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ to obtain a shrunken ROI for the $(j+1)$th iteration. This new ROI is constructed by $\mathrm{ROI}^{(j+1)} = I_1^{(j+1)} \times I_2^{(j+1)} \times I_3^{(j+1)}$, where the one-dimensional interval is given by

$$I_i^{(j+1)} = \Big[\min_{1\le t\le T}\{\mu_{t,i}^{(j)} - 3\cdot\sigma_{t,i}^{(j)}\},\ \max_{1\le t\le T}\{\mu_{t,i}^{(j)} + 3\cdot\sigma_{t,i}^{(j)}\}\Big],$$

where $\mu_{t,i}^{(j)} = \sum_{k=1}^{K}\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})c_{k,i}$, $\sigma_{t,i}^{(j)} = \mathrm{sqrt}(\sum_{k=1}^{K}\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})(c_{k,i} - \mu_{t,i}^{(j)})^2)$, $c_{k,i}$ is the ith component of $\mathbf{c}_k$, and $i = 1, 2, 3$. With the shrunken ROI, we also increase the mesh grids K_i, $i = 1, 2, 3$, during the iterative procedures. When the dynamic EM algorithm converges, an estimate $\hat{\boldsymbol{\Theta}}_{\mathrm{d}}$ will be available with a shrunken ROI.

3. Simulation study.

3.1. *MEG data generation*. In the simulation study we considered a single sphere head model (centered at the origin with radius 10 cm) and simulated 102 magnetometers which are randomly placed on the upper part of the head. Movement of the current sources inside the head model was restricted. In order to focus on the location parameters of the sources, we fixed some parameters (initial distribution parameters $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, and noise parameters $\boldsymbol{\Sigma}$, $\mathbf{V}$) in the model (2.3) and (2.4). The total length of each simulation was 100 time points, and of interest at any time point is the discrete posterior distribution of the sources.

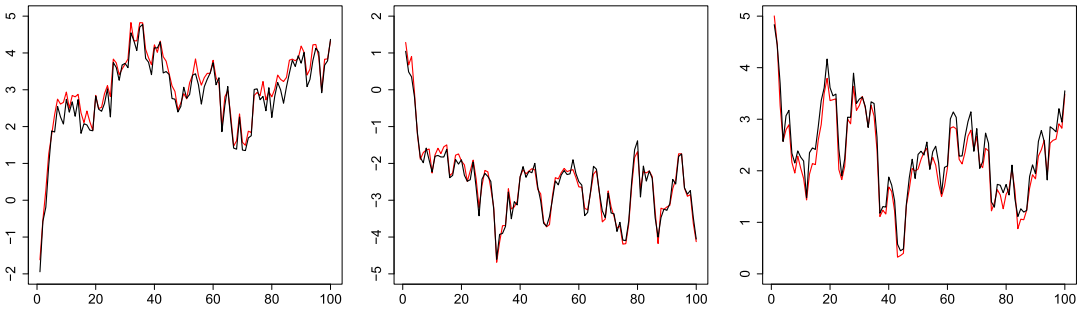
3.2. *Simulated Case* 1. In this example, we consider a single source which moves in *three* dimensions (x, y, z) in the head model. The parameters of the simulated source are summarized in Table 2.

In the simulation we applied the EM algorithm with the dynamic procedure to estimate the parameter $\boldsymbol{\Theta} = \{A, \mathbf{b}\}$ in the single source model. We chose the upper-head model as the

TABLE 2

Source simulation in Case 1: *The location parameter* $\boldsymbol{p}_t$ *of the source is expressed in terms of Cartesian coordinates* (x (*cm*), y (*cm*), z (*cm*)) *and is allowed to vary. The moments and strength parameter* $\boldsymbol{q}_t$ *is fixed during simulations*

$\boldsymbol{\mu}_0 = (\mathbf{p}_0^{\mathrm{T}}, \mathbf{q}_0^{\mathrm{T}})^{\mathrm{T}}$	$(-2, 1, 5, 3, 3, 3)^{\mathrm{T}}$
$\boldsymbol{\Sigma}_0 = \mathrm{diag}(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,6}^2)$	$\mathrm{diag}(0.0225, 0.0225, 0.0225, 10^{-4}, 10^{-4}, 10^{-4})$
$A = \mathrm{diag}(a_1, a_2, \ldots, a_6)$	$\mathrm{diag}(0.75, 0.8, 0.9, 1, 1, 1)$
$\mathbf{b} = (b_1, b_2, \ldots, b_6)^{\mathrm{T}}$	$(0.75, -0.5, 0.25, 0, 0, 0)^{\mathrm{T}}$
$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_6^2)$	$\mathrm{diag}\ (0.25, 0.25, 0.25, 10^{-4}, 10^{-4}, 10^{-4})$
$\mathbf{V} = \mathrm{diag}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_{102}^2)$	$\mathrm{diag}(6.25 * 10^{-5}, 6.25 * 10^{-5}, \ldots, 6.25 * 10^{-5})$
Number of time points	100



(a) x dimension (b) y dimension (c) z dimension

FIG. 2. *Marginal posterior means for location parameter* $\boldsymbol{p}_t = (p_{t1}, p_{t2}, p_{t3})^{\mathrm{T}}$ *of a single source with dynamic procedure in* 100 *time points. The simulated location parameters are plotted in a black line, and the estimated posterior means are plotted in a red line.*

initial ROI and let the initial mesh grids $K_i^{(1)} = 10$, $i = 1, 2, 3$. Since we applied the discrete approach to the simulated data, we started the dynamic EM algorithm by discretizing the initial ROI with the initial mesh grids. In the jth iteration, the mesh grid was increased by $K_i^{(j+1)} = K_i^{(j)} + 1$, $i = 1, 2, 3$, and the ROI was shrunk for the following iteration. The iteration procedure was performed until the conditional expectation in (B.1) converged, and we obtained the MLEs $\hat{\boldsymbol{\Theta}}_{\mathrm{d}} = \{\hat{A}_{\mathrm{d}}, \hat{\mathbf{b}}_{\mathrm{d}}\}$, where

$$\hat{A}_{\mathrm{d}} = \begin{pmatrix} \hat{A}_{\mathrm{d}}^* & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \end{pmatrix}, \qquad \hat{A}_{\mathrm{d}}^* = \begin{pmatrix} 0.6969 & 0.0260 & -0.0024 \\ 0.0479 & 0.8352 & 0.0382 \\ -0.0207 & 0.0029 & 0.9035 \end{pmatrix},$$

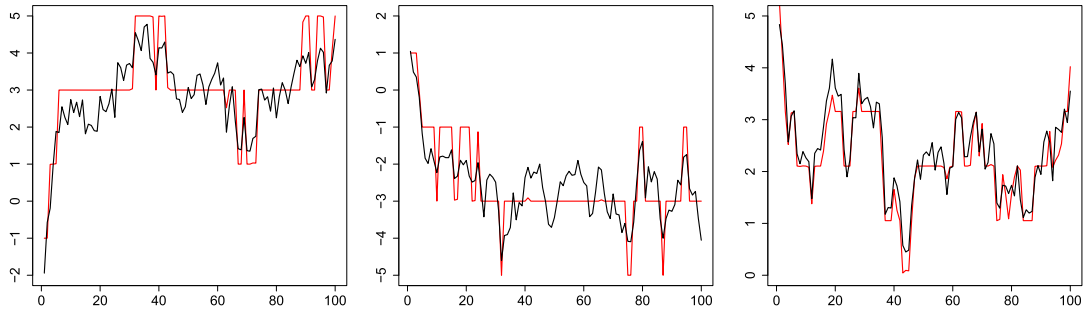
and $\hat{\mathbf{b}}_{\mathrm{d}} = (0.9173, -0.6657, 0.3194, 3.0032, 2.9536, 3.0488)^{\mathrm{T}}$. When we obtained the estimated parameter, we also obtained a shrunken ROI $[-1.7707, 3.1606]$ cm $\times$ $[-3.0992, 1.2701]$ cm $\times$ $[1.4235, 4.7374]$ cm. Utilizing the estimated parameter and the shrunken ROI, we computed the discrete posterior distribution $\mathrm{P}(\mathbf{J}_t^{\mathrm{p}} | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{d}})$. To visualize the discrete posterior distribution, we plot the marginal posterior means for location parameter $\mathbf{p}_t$ of the source in Figure 2.

We also compared the numerical result from the dynamic procedure with the result from the nondynamic EM procedure. For the nondynamic EM algorithm we implemented the discrete approach by discretizing the initial ROI with initial mesh grids throughout the simulations. In this case we obtained the estimates

$$\hat{A}_{\mathrm{nd}} = \begin{pmatrix} \hat{A}_{\mathrm{nd}}^* & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \end{pmatrix}, \qquad \hat{A}_{\mathrm{nd}}^* = \begin{pmatrix} 0.7280 & -0.0018 & 0.0052 \\ -0.0019 & 0.7888 & 0.0208 \\ -0.0689 & -0.0258 & 0.8988 \end{pmatrix},$$

and $\hat{\mathbf{b}}_{\mathrm{nd}} = (0.7980, -0.6292, 0.4197, 3.0032, 2.9536, 3.0488)^{\mathrm{T}}$. The posterior means for the location parameter of the source are plotted in Figure 3. To compare the estimation accuracy from the two procedures, we repeated the simulation four times and computed the estimation error in maximum norm. From Table 3 we note that the estimates from the dynamic EM procedure are more accurate than the nondynamic EM procedure.

3.3. *Simulated Case* 2. In addition to Case 1, a case of two sources was performed. In the simulation the two sources were assumed to be uncorrelated and were allowed to move in *three* dimensions (x, y, z) in the head model. The parameters of the simulated sources are summarized in Table 4.



(a) x dimension (b) y dimension (c) z dimension

FIG. 3. *Marginal posterior means for location parameter* $\boldsymbol{p}_t = (p_{t1}, p_{t2}, p_{t3})^{\mathrm{T}}$ *of a single source without dynamic procedure in* 100 *time points. The simulated location parameters are plotted in a black line, and the estimated posterior means are plotted in a red line.*

To deal with the simulated data with two sources, the dynamic procedure and switch procedure were applied to implement the EM algorithm for estimating the parameters $\boldsymbol{\Theta} = \{A, \mathbf{b}\}$ in the source model. When the EM algorithm converged, we obtained the estimates

$$\hat{A} = \begin{pmatrix} \hat{A}_1 & \mathbf{0}_{6\times 6} \\ \mathbf{0}_{6\times 6} & \hat{A}_2 \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^{\mathrm{T}}, \hat{\mathbf{b}}_2^{\mathrm{T}})^{\mathrm{T}},$$

where

$$\hat{A}_i = \begin{pmatrix} \hat{A}_i^* & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \end{pmatrix}, \quad i = 1, 2,$$

$$\hat{A}_1^* = \begin{pmatrix} 0.2660 & 0.0942 & -0.1494 \\ 0.0283 & 0.5690 & 0.2668 \\ 0.0884 & 0.0570 & 0.8590 \end{pmatrix}, \qquad \hat{A}_2^* = \begin{pmatrix} 0.2501 & -0.0450 & 0.0243 \\ -0.0953 & 0.7234 & 0.0916 \\ -0.0481 & 0.0255 & 0.8550 \end{pmatrix},$$

$$\hat{\mathbf{b}}_1 = (3.6058, -2.5997, 0.2934, 2.9535, 3.0124, 2.9188)^{\mathrm{T}}$$

and

$$\hat{\mathbf{b}}_2 = (2.5589, -0.9985, 0.6659, 3.0527, 3.0327, 3.0541)^{\mathrm{T}}.$$

The discrete posterior distributions of the sources were calculated using the estimated parameter $\hat{\boldsymbol{\Theta}}$, and the posterior means for the location parameter of the two sources are plotted in Figure 4.

To compare the result with the nondynamic procedure, we performed the simulation with the nondynamic EM algorithm and switch procedure. The comparison of the estimation accuracy between the dynamic EM algorithm with switch procedure and nondynamic EM algorithm with switch procedure is shown in Table 3. We also performed the simulation with a

TABLE 3

Comparison of mean errors for MLEs from the EM algorithm with different procedures. Standard deviations computed based on four repetitions are shown in parentheses

	Case 1		Case 2	
	Dynamic EM	Nondynamic EM	Dynamic EM with switch procedure	Nondynamic EM with switch procedure
$\|\hat{A} - A\|_{\max}$	0.1293 (0.0078)	0.1642 (0.0128)	0.3346 (0.0577)	0.4074 (0.0578)
$\|\hat{\mathbf{b}} - \mathbf{b}\|_{\max}$	0.2095 (0.0257)	0.3110 (0.1165)	1.7887 (0.1788)	2.3282 (0.4476)

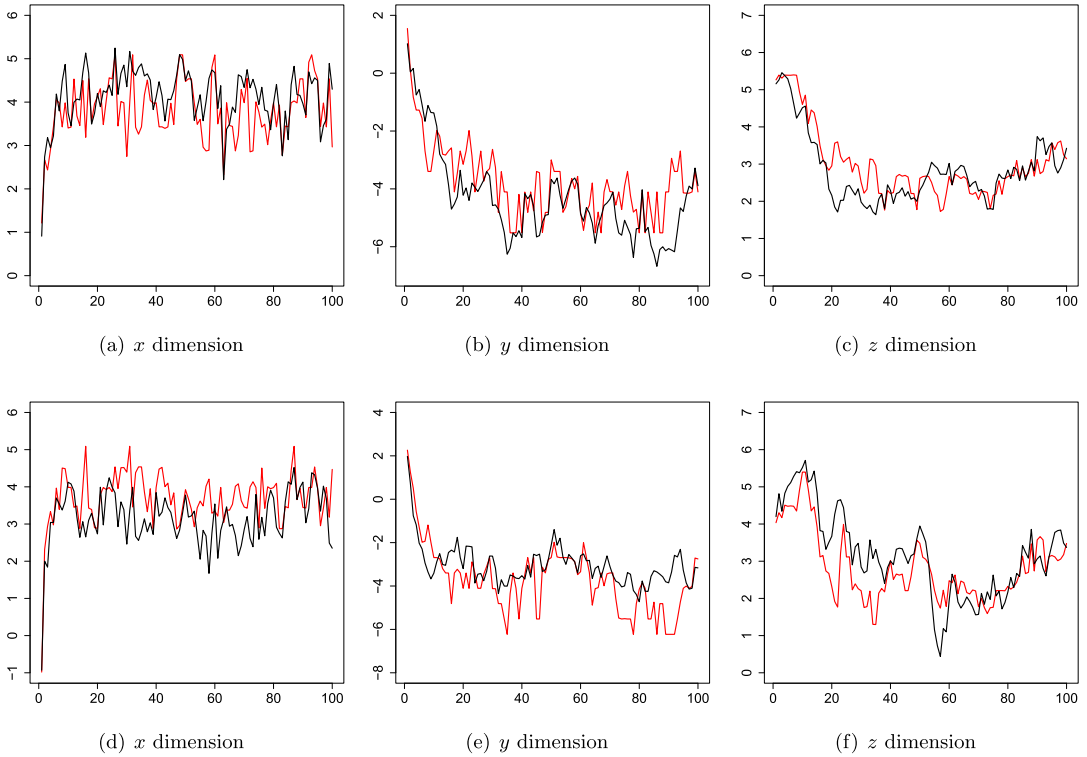
TABLE 4

Source simulation in Case 2: *The location parameter* $\boldsymbol{p}_t$ *of the sources is expressed in terms of Cartesian coordinates* (x (*cm*), y (*cm*), z (*cm*)) *and is allowed to vary. The moments and strength parameter* $\boldsymbol{q}_t$ *is fixed during simulations*

$\boldsymbol{\mu}_{0,1} = (\mathbf{p}_{0,1}^{\mathrm{T}}, \mathbf{q}_{0,1}^{\mathrm{T}})^{\mathrm{T}}$	$(1, 1, 5, 3, 3, 3)^{\mathrm{T}}$
$\boldsymbol{\mu}_{0,2} = (\mathbf{p}_{0,2}^{\mathrm{T}}, \mathbf{q}_{0,2}^{\mathrm{T}})^{\mathrm{T}}$	$(-1, 2, 4, 3, 3, 3)^{\mathrm{T}}$
$\boldsymbol{\Sigma}_{0,1} = \mathrm{diag}(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,6}^2)$	$\mathrm{diag}(0.01, 0.01, 0.01, 10^{-4}, 10^{-4}, 10^{-4})$
$\boldsymbol{\Sigma}_{0,2} = \mathrm{diag}(\sigma_{0,1}^2, \sigma_{0,2}^2, \ldots, \sigma_{0,6}^2)$	$\mathrm{diag}(0.01, 0.01, 0.01, 10^{-4}, 10^{-4}, 10^{-4})$
$A_1 = \mathrm{diag}(a_{1,1}, a_{1,2}, \ldots, a_{1,6})$	$\mathrm{diag}(0.5, 0.8, 0.9, 1, 1, 1)$
$A_2 = \mathrm{diag}(a_{2,1}, a_{2,2}, \ldots, a_{2,6})$	$\mathrm{diag}(0.45, 0.75, 0.85, 1, 1, 1)$
$\mathbf{b}_1 = (b_{1,1}, b_{1,2}, \ldots, b_{1,6})^{\mathrm{T}}$	$(2, -1, 0.25, 0, 0, 0)^{\mathrm{T}}$
$\mathbf{b}_2 = (b_{2,1}, b_{2,2}, \ldots, b_{2,6})^{\mathrm{T}}$	$(1.8, -0.8, 0.5, 0, 0, 0)^{\mathrm{T}}$
$\boldsymbol{\Sigma}_1 = \mathrm{diag}(\sigma_{1,1}^2, \sigma_{1,2}^2, \ldots, \sigma_{1,6}^2)$	$\mathrm{diag}(0.25, 0.25, 0.09, 10^{-4}, 10^{-4}, 10^{-4})$
$\boldsymbol{\Sigma}_2 = \mathrm{diag}(\sigma_{2,1}^2, \sigma_{2,2}^2, \ldots, \sigma_{2,6}^2)$	$\mathrm{diag}(0.25, 0.25, 0.09, 10^{-4}, 10^{-4}, 10^{-4})$
$\mathbf{V} = \mathrm{diag}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_{102}^2)$	$\mathrm{diag}(6.25 * 10^{-5}, 6.25 * 10^{-5}, \ldots, 6.25 * 10^{-5})$
Number of time points	100

nondynamic, nonswitch procedure for comparison in the Supplementary Material (Yao et al. (2020)).

REMARK 3.1. The dynamic switch EM algorithm was started with initial ROI $[-10, 10]$ cm $\times$ $[-10, 10]$ cm $\times$ $[0, 10]$ cm and mesh grids $K_i = 10$, $i = 1, 2, 3$. The ROI



(a) x dimension (b) y dimension (c) z dimension

(d) x dimension (e) y dimension (f) z dimension

FIG. 4. *Marginal posterior means for location parameter* $\boldsymbol{p}_t = (p_{t1}, p_{t2}, p_{t3})^{\mathrm{T}}$ *of two sources with the dynamic and switch procedure in* 100 *time points. Top row*: *results for source* 1; *bottom row*: *results for source* 2. *The simulated location parameters are plotted in a black line, and the estimated posterior means are plotted a in red line.*

was shrunk, and the mesh grids were increased by 1 during iterations. The nondynamic switch EM algorithm was implemented with the initial ROI and mesh grids throughout all the iterations.

4. Real data application 1. The first real data analysis reports the source localization for the Brain-Controlled Interfaces (BCI) data collected at the Center for Advanced Brian Magnetic Source Imaging (CABMSI) at the Presbyterian University Hospital in Pittsburgh. The data consists of MEG scans of 102 magnetometers recorded at 37,000 milliseconds (ms). During the experiment the subjects performed a two-dimensional center-out task using wrist movement. In the imagined movement task, the subjects were first asked to imagine that they were performing the center-out movement using their wrist. During the overt movement task, subjects controlled a 2D cursor using their wrist to perform the center-out task. Each trial started after the subject held the cursor in the center for a holding period, followed by a target onset. In order for a trial to be considered successful, the subject needed to move the cursor to the target and hold it there for the duration of the holding period.

The goal of our analysis is to investigate the dynamics of the possible existing sources in the BCI data. Previous work on this data focused on estimating the distribution of the source when the number of sources is assumed to be known (Yao and Eddy (2014)). In this section we consider the data after movement onset from all the magnetometers and, mainly, focus on the time varying characteristics of the source location with a dynamically estimated number of sources. We adopt the viewpoint developed in Yao et al. (2018) about the changing number of sources that might exist in the BCI data and exploit the distribution of sources using the estimated number of sources at different stages of the experiment. Throughout the analysis, we assume a unit moment for all possible sources for simplicity. A single sphere head model, with its center (1.07, 0.74, 1.65), radius 10.5 and measured in centimeters (cm), was constructed based on the magnetometer positions and head shape information from the BCI data.

Specifically, we have applied the proposed discrete approach to the BCI data through two subanalyses (short time frame and long time frame): (1a) With the estimated number of sources introduced in Yao et al. (2018), we investigate the source distribution in space and time within two selected time windows after the movement onset; (1b) as a previous study has suggested that there are still some active sources present after the movement, we contrast the behavior of the sources for the same selected time windows with no noise estimation; (2) the source estimation for a longer time frame is also reported.

4.1. *Activity for short time frame.*

1. In 1a the sources with the estimated number obtained by the Fourier transform were investigated for two selected time windows [12,000, 12,099] ms and [20,000, 20,099] ms, where time window [12,000, 12,099] ms was analyzed for the time varying characteristics of two sources and the time window [20,000, 20,099] ms was selected for the analysis with three sources. In 1b we considered the same time windows for the source investigation with the estimated number of sources without noise estimation, which suggested only one and two sources, respectively.

2. The Matlab toolbox "fieldtrip" was used to obtain an MNE by searching the entire head. We set the area around the MNE to be the 3D ROI.

3. To implement the proposed discrete approach, we chose mesh grids K_i, $i = 1, 2, 3$, for x, y, z dimensions of the ROI, respectively. Then, the ROI was subsequently discretized into $K_1 \cdot K_2 \cdot K_3$ voxels.

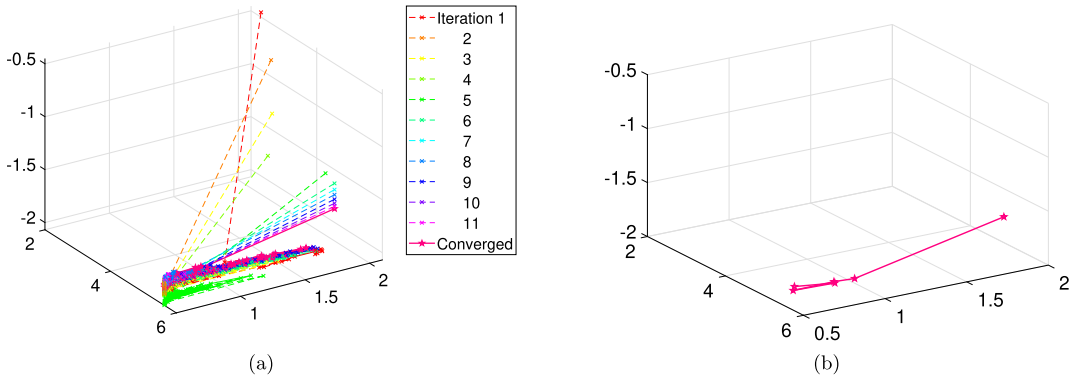


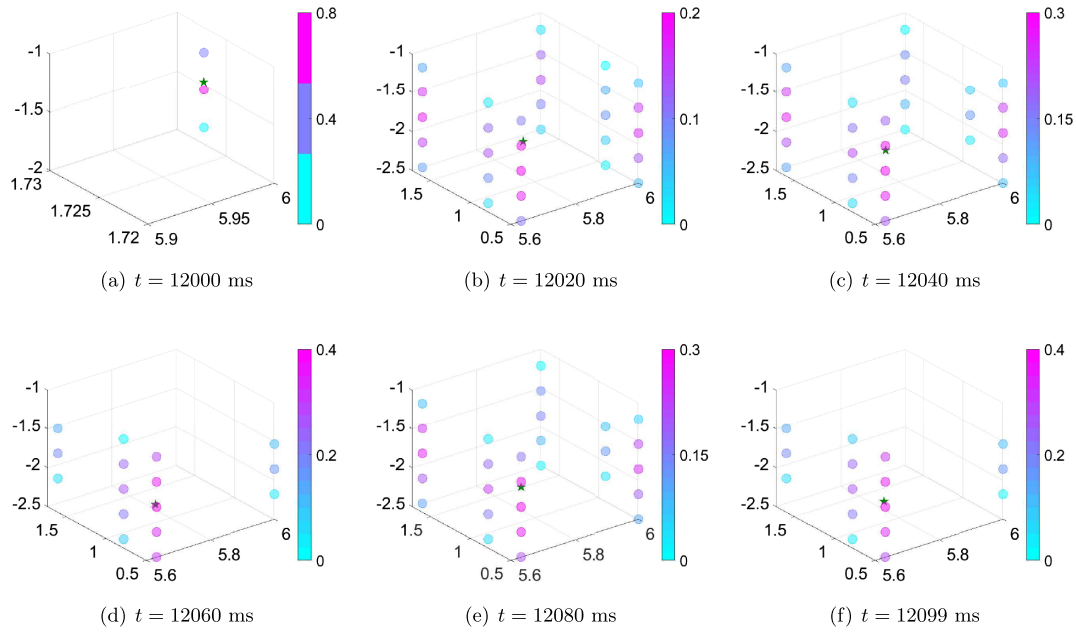
FIG. 5. *Trajectory of posterior means for location parameter* (x, y, z) *of a single source in the time window* [12,000, 12,099] *ms.* (*a*) *Trajectories of a single source during EM iterations.* (*b*) *Trajectory of the converged source is highlighted at six selected time points* 12,000, 12020, 12040, 12060, 12080 *and* 12,099 *ms.*

4. Motivated by the MNE, we manually set the parameters $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$ in the source model, and, therefore, the initial stage of the source was constrained in the small neighbourhood of the MNE. To let the movement of the sources within the ROI all the time, the covariance matrix $\boldsymbol{\Sigma}$ in the source model was also manually set from the ROI that we chose in 2. The noise estimation of the selected data was obtained using the Fourier transform.

5. We used the estimates $\hat{A}$ and $\hat{\mathbf{b}}$ from the converged EM algorithm to calculate the discrete posterior probability distribution of the location parameter (x, y, z) at each time point.

To illustrate the results of the analysis, we now explain the sources distribution (single-source case in 1b and two-source case in 1a) for the time window [12,000, 12,099] ms. For the single-source case the trajectories of posterior means at each time point during EM iterations are shown in Figure 5(a), and the posterior means of the converged source are highlighted in Figure 5(b) at six selected time points. To visualize the time varying characteristics, the target posterior distribution for the location parameters with nonzero probabilities in the 3D ROI are provided. Figure 6 shows the dynamics of the target posterior distribution at the same selected time points. As shown in Figure 6, the posterior means at the selected time points are highlighted using green stars which vary from $(6, 1.73, -1.47)$ cm at the first time point $t = 12{,}000$ ms to $(5.66, 0.53, -1.81)$ cm at the last time point $t = 12{,}099$ ms. Figure 7 represents the marginal posterior distribution at three selected time points from which we can see the dynamics for the location parameter in each dimension. The dynamics of the two-source case for the same time window have been summarized in Figure 8 to Figure 10 with similar interpretation. The results (two-source case in 1b and three-source case in 1a) for the time window [20,000, 20,099] ms can be found in the Supplementary Material (Yao et al. (2020)).

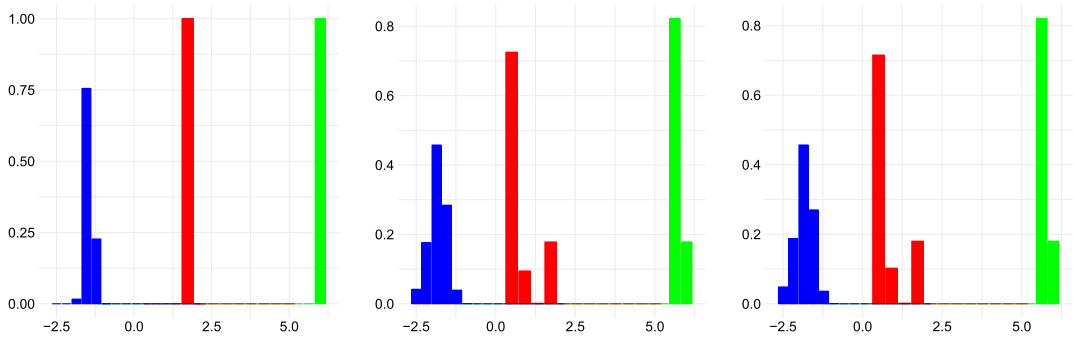
4.2. *Activity for long time frame.* To investigate the effect of window length in source distribution for the BCI data, we carried out the same analysis for a sequence of two time windows starting at the same time point that was previously studied, but differing in length. In this case, each time window was selected with four different lengths (200, 300, 400 and 500 ms, respectively). The distributions of the two sources estimated by the Fourier transform were investigated for four selected time windows [12,000, $12k^*$99] ms, where $k^* = 1, 2, 3, 4$. We chose the ROI and discretized it, similar to the manner that it was done in the previous section, but the noise estimation of the selected data was obtained by the Fourier transform using the measurements within the first 100 ms. Although the data covariance matrix in noise estimation changes over the length of the window, we have observed that it is necessary to



(a) $t = 12000$ ms (b) $t = 12020$ ms (c) $t = 12040$ ms

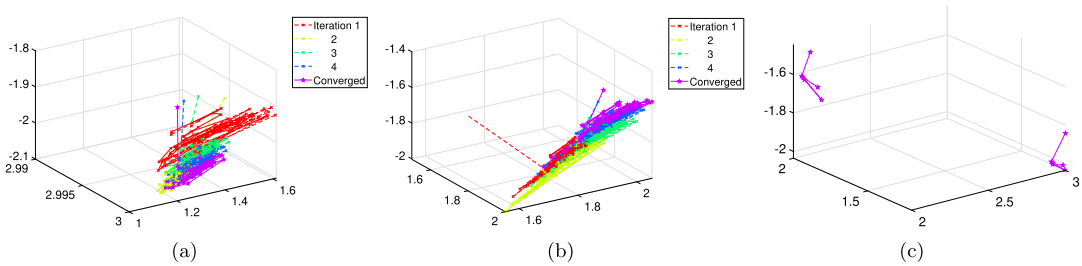
(d) $t = 12060$ ms (e) $t = 12080$ ms (f) $t = 12099$ ms

FIG. 6. *Posterior distribution for location parameter* (x, y, z) *of a single source in the time window* [12,000, 12,099] *ms. Green star: posterior mean for location parameter at the selected time point.*



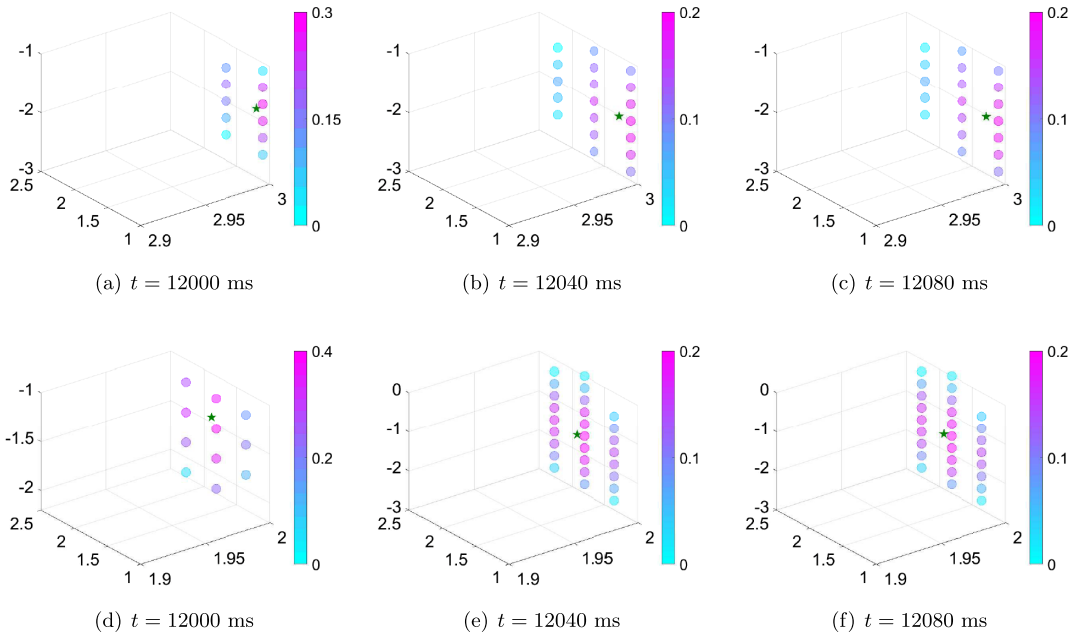
(a) $t = 12000$ ms (b) $t = 12040$ ms (c) $t = 12080$ ms

FIG. 7. *Marginal posterior distribution for location parameter* (x, y, z) *of a single source in the time window* [12,000, 12,099] *ms. Green bar: marginal posterior distribution for parameter* x; *red bar: marginal posterior distribution for parameter* y; *blue bar: marginal posterior distribution for parameter* z.



(a) (b) (c)

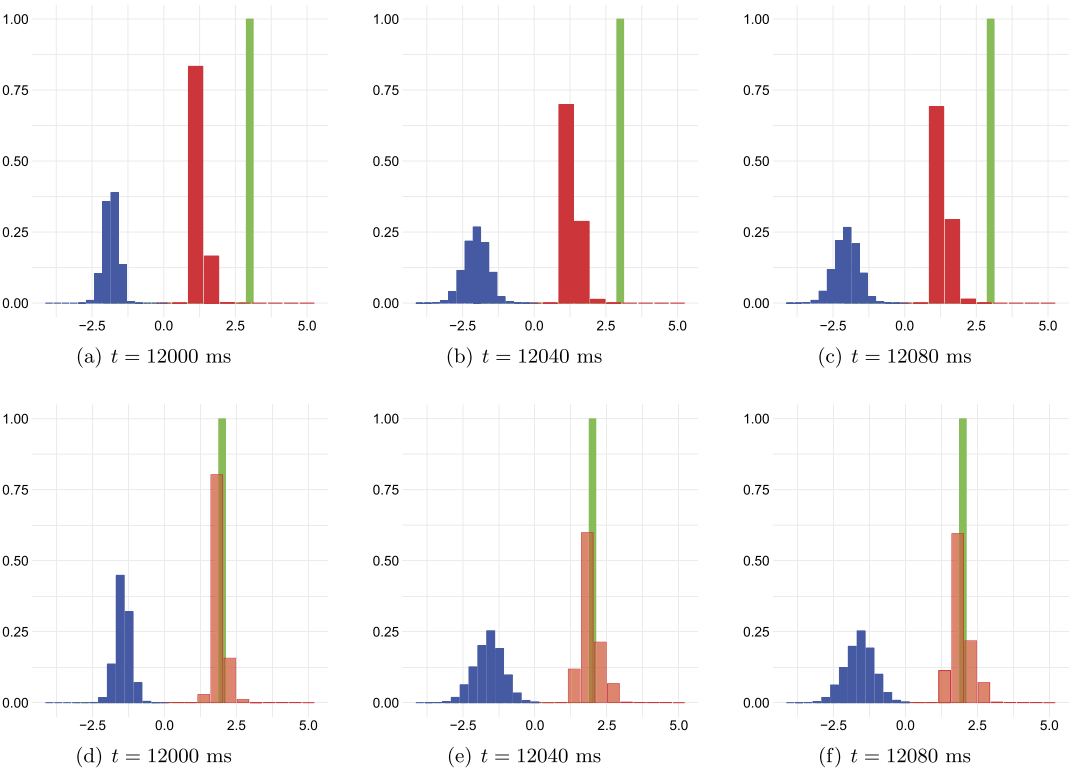
FIG. 8. *Trajectory of posterior means for location parameter* (x, y, z) *of two sources in the time window* [12,000, 12,099] *ms.* (a) *and* (b) *Trajectories of the two sources during EM iterations.* (c) *Trajectories of the two converged sources are highlighted at six selected time points* 12,000, 12,020, 12,040, 12,060, 12,080 *and* 12,099 *ms.*



(a) $t = 12000$ ms (b) $t = 12040$ ms (c) $t = 12080$ ms

(d) $t = 12000$ ms (e) $t = 12040$ ms (f) $t = 12080$ ms

FIG. 9. *Posterior distribution for location parameter* (x, y, z) *of two sources in the time window* [12,000, 12,099] *ms. Top row*: *results for source* 1; *bottom row*: *results for source* 2. *Green star*: *posterior mean for location parameter at the selected time point.*



(a) $t = 12000$ ms (b) $t = 12040$ ms (c) $t = 12080$ ms

(d) $t = 12000$ ms (e) $t = 12040$ ms (f) $t = 12080$ ms

FIG. 10. *Marginal posterior distribution for location parameter* (x, y, z) *of two sources in the time window* [12,000, 12,099] *ms. Top row*: *results for source* 1; *bottom row*: *results for source* 2. *Green bar*: *marginal posterior distribution for parameter* x; *red bar*: *marginal posterior distribution for parameter* y; *blue bar*: *marginal posterior distribution for parameter* z.

use early observations for stable noise estimation. The posterior distributions for the two sources were obtained for location parameter (x, y, z) at each time point in the selected time windows [12,000, 12k^*99] ms, $k^* = 1, 2, 3, 4$. To illustrate the results from the four selected time windows, we compared the source distributions with the result from the previous analysis from the time window [12,000, 12,099] ms. Since we considered five time windows with different lengths, we denote $\mathcal{Y}_{[12,000,12k^*99]}$ as the measurements in the time window [12,000, 12k^*99] ms and $\hat{\mathbf{\Theta}}_{[12,000,12k^*99]}$ as the estimate obtained from the measurements $\mathcal{Y}_{[12,000,12k^*99]}$, for $k^* = 0, 1, \ldots, 4$. For $12{,}000 \le t \le 12{,}099$, we focused on the posterior distributions $\mathrm{P}(\mathbf{J}_t^{\mathrm{p}} | \mathcal{Y}_{[12,000,12k^*99]}, \hat{\mathbf{\Theta}}_{[12,000,12k^*99]})$ for $k^* = 0, 1, \ldots, 4$. We found that the estimated posterior means of the two sources changed as we increased k^* from 0 to 4. Figure 11 exhibits such changes in source 1 for only selected time points. The corresponding plot for source 2 can be found in the Supplementary Material (Yao et al. (2020)). On one hand, this result is expected because the selection on time windows should have affected the parameter estimation in EM and the smoothing procedure of the posterior calculation; on the other hand, this serves as an example that the estimation of the source distribution based on more data could result in a significant difference from the short frame. Moreover, we also noted that the changes in the posterior means were subtle in circumstances where k^* was increased from 0 to 3, as shown in the Supplementary Material (Yao et al. (2020)). However, when k^* was greater than 3, the marginal posterior means was largely different from those with k^* less than 3. This difference seems be caused by a temporary effect. This result matches well with our observation that the filtering result $\mathrm{P}(\mathbf{J}_t^{\mathrm{p}} | \mathcal{Y}_{[12,000,12,099]}, \hat{\mathbf{\Theta}}_{[12,000,12,099]})$ might not be a good estimate of the distribution $\mathrm{P}(\mathbf{J}_t^{\mathrm{p}} | \mathcal{Y}_{[12,000,12k^*99]}, \hat{\mathbf{\Theta}}_{[12,000,12k^*99]})$ when k^* is bigger than 3. A similar phenomenon was observed for the time window [20.000, 20k^*99] at $k^* = 3$, where k^* was increased from 0 to 4 with no noise estimation. The details can be found in the Supplementary Material (Yao et al. (2020)). We did not find the temporary effect for one-source case in [12,000, 12k^*99] ms and three-source case in [20,000, 20k^*99] ms, for all $k^* = 0, 1, \ldots, 4$. All other analyses are in the Supplementary Material (Yao et al. (2020)).

5. Real data application 2. For the second real data application, we extend the proposed methodology to a set of EEG recordings under spatial working memory (SWM) task. The SWM reflected in brain activities is often related to relevant brain networks. Thus, source localization studies using the EEG recordings should provide insight into how the brain temporally and spatially responds to different SWM loads.

The EEG recordings are collected from a participant under three memory load conditions which consist of 26, 27 and 29 trials, respectively. For each trial it contains three phases: encoding, retention and probing. During the encoding phase, it began with a cross in the center of the screen. Depending on the load condition, 1, 3 or 5 white dots was/were presented sequentially on the screen. During the retention phase, a fixation cross was then displayed, followed by a red dot presented for the probing phase. The participant was required to indicate whether the red dot appeared at a previously occupied location. The EEG data are processed and down sampled to 250 Hz. For each trial of EEG recording for the retention phase, it consists of a baseline duration, time before event onset, lasting for 200 ms and an event duration lasting for the following 3992 ms.

To investigate the association between EEG responses and brain network in the SWM task, we identify the source distribution at each time step during the event onset period for the retention phase. Since the SWM task often observes sustained negative activity during retention (Liu et al. (2018)), three ROIs associated with fMRI-based deactivation pattern in higher capability group are used for source searching in our study. For each memory load condition, the source distribution is obtained with the estimated number of sources trial by

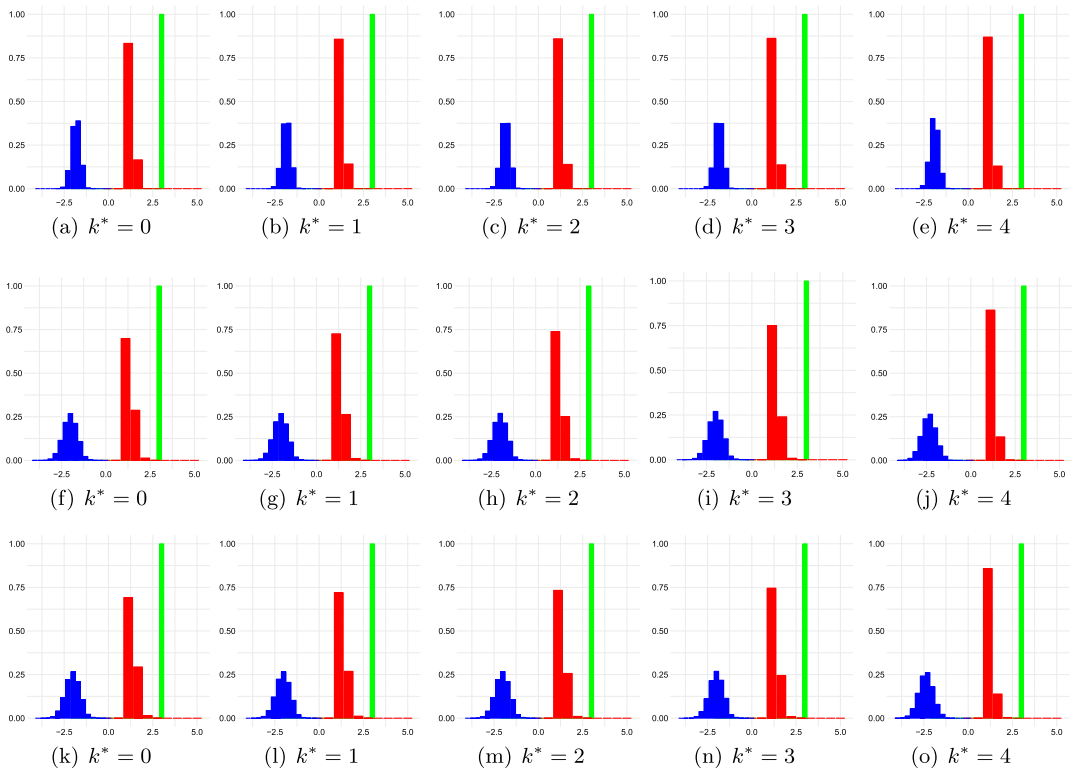


FIG. 11. *Marginal posterior distribution for location parameter* (x, y, z) *of source* 1 *in the time window* [12,000, 12k*99] *ms, where* $k^* = 0, 1, 2, 3, 4$. *Top row*: $t = 12{,}000$ *ms*; *middle row*: $t = 12{,}040$ *ms*; *bottom row*: $t = 12{,}080$ *ms*. *Green bar*: *marginal posterior distribution for parameter* x; *red bar*: *marginal posterior distribution for parameter* y; *blue bar*: *marginal posterior distribution for parameter* z.

trial. In Figure 12 we plot the source location with significant posterior probability in the three highlighted ROIs (red areas) within the human cortex (grey dots). To distinct brain network under different load conditions, the source distribution from four trials (with *two* to *five* estimated sources) for each load condition are illustrated column by column. With two or three estimated sources (the first two rows in Figure 12), the source distributions under load 1 and load 5 share some similarity that the two/three sources are concentrated in the rightmost ROI, while the source distribution under load 3 spread out to the other two ROIs. When more sources are presented, the distribution pattern change. In the last row of Figure 12, it shows the source distribution with five estimated sources. The patterns of the source distributions under load 3 and load 5 are similar, with most of the sources distributed in the rightmost ROI. However, none of the sources distributed in the leftmost ROI under load 1 at the selected time point.

The construction of EEG sources also shed light on exploring the brain's functional organization. In this case we searched the sources among the 400 parcels within the whole brain with parcellating the cerebral cortex into seven functionally consistent subregions. In Figure 13 we highlighted the networks with significant posterior probabilities of one selected source under three memory loads. From the the pattern of functional connections between 14 subregions, the Limbic network (5 and 12 in vertical axis of Figure 13) is one of the commonly activated regions under *three* memory loads. As we increase the memory condition, the pattern of the activated regions seems to be more concentrated.

6. Conclusion. The quantification of the source current in a time-varying source model for the MEG data is still a practically urgent problem. Due to the nonlinearity of the mea-

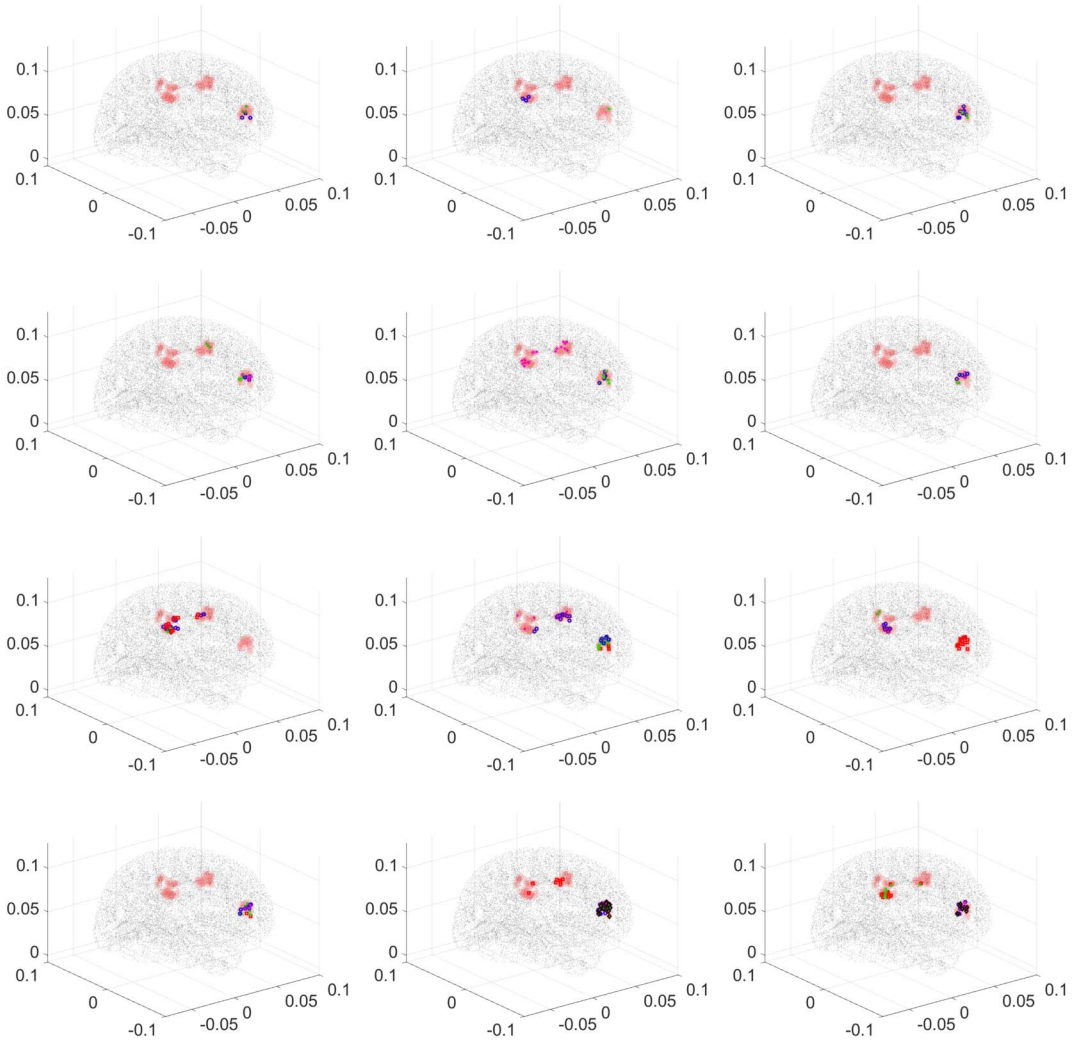
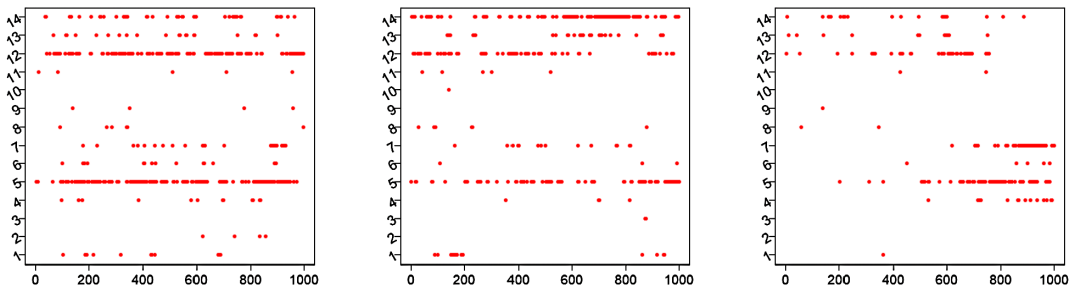


FIG. 12. *Posterior distribution for location parameter* (x, y, z) *of estimated number of sources at selected time step* $t = 600$ *ms during the retention phase under three load conditions. Column* 1: *load condition* 1; *column* 2: *load condition* 3; *column* 3: *load condition* 5. *Source* 1: *green* $\times$; *source* 2: *blue* $\circ$; *source* 3: *magenta* $+$; *source* 4: *red* $\triangle$; *source* 5: *black* $\diamond$. *Grey dot*: *cortex*; *red area*: *ROI*.



(a) load 1 (b) load 3 (c) load 5

FIG. 13. *Significant source distributions among the* 400 *parcellations in seven networks. The horizontal axis represents time points. The vertical axis represents the seven networks in each hemisphere. From* 1 *to* 7, *the networks are Visual*, *Somatomotor*, *Dorsal Attention*, *Ventral Attention*, *Limpic*, *Frontoparietal*, *Default in the left hemispher*; *From* 8 *to* 14, *they are the seven networks in the right hemisphere.*

surement with respect to the source location parameter, effective methods for dealing with this problem are lacking. Common regularization-based methods mainly focus on estimating the source moment, while other time-varying methods attempt to find estimates of the source location using measurements up to the time point of interest. The latter is encouraged for improvement, as using the entire set of measurements is more justifiable for estimating the time-varying source. Meanwhile, these methods are often restricted to a predetermined number of sources. In MEG a reliable estimation of the number of sources present in the data, or perhaps at different stages of the experiment, is quite crucial for the whole problem.

With the goal of proposing a framework that allows the flexible estimation of the evolution of the source in MEG, we introduced a discrete approach for calculating the posterior distribution of the source. We emphasized the importance of directly calculating the posterior distribution of the source through a discrete model, and this method differs in the fact that it only samples the intractable continuous target distribution. In this respect this discrete approach has improved the posterior distribution by providing the probability of the possible sources present in the brain rather than using some approximated samples (Yao and Eddy (2014)). In both the single-source case and mutiple-source case, the proposed approach was seen to be more reliable in estimating the source distribution.

We show the suboptimality of the EM algorithm in estimating the model parameters within the discrete source model. The performance of our proposed approach was examined in some simulated examples with varying model setups. For the single-source case, the combined procedure seemed to well capture the true time evolution of the sources. For the multisource model, we have proposed a dynamic procedure and a switch procedure for estimation accuracy and computational efficiency. The proposed approach gave rather satisfactory and consistent results.

In our analysis we adopted the results of the estimated number of sources in the real MEG/EEG data application from Yao et al. (2018) and implemented the source estimation dynamically with and without noise estimation. As suggested in Yao et al. (2018), we did not estimate the number of sources for the BCI data from the entire set of data at once but rather attempted to estimate it from some selected windows. We found the existence of a temporal effect at several selected time points in this data. In fact, we noted that the source distribution estimated from the proposed method and the sequential sampling method in Yao and Eddy (2014) did not differ much in the beginning but started to diverge as more measurements were included. In addition, this phenomenon was only found to be significant for two-source cases in the BCI data. However, the investigation for the EEG data in the SWM task was conducted with the entire measurements after event onset in retention.

To summarize, a reliable estimation of the source distribution depends on a reliable estimation of the number of the sources and the source localization algorithm. With our method, we explore the use of combined approaches in a more advanced form to further examine the evolution of sources.

APPENDIX A

The forward-backward algorithm (Rabiner (1989)) is an efficient inference algorithm which computes the posterior distribution of the hidden state variables given the entire set of measurements in two passes. The forward-backward algorithm for computing the discrete posterior distribution in (2.9) is summarized in Table 5 and Table 6. In Table 5 we start the forward recursion from the first time point $t = 1$ and compute the filtering posterior distribution $\alpha_{1k}(\boldsymbol{\Theta})$ given $\boldsymbol{\Theta}$, for $1 \leq k \leq K$. After which, we compute the filtering posterior distribution $\alpha_{tk}(\boldsymbol{\Theta})$ using the previous filtering posterior distribution $\alpha_{t-1,l}(\boldsymbol{\Theta})$, for $t = 2, \ldots, T$. Table 6 illustrates the backward procedure to calculate the smoothing posterior distribution

TABLE 5
Forward procedure of the forward-backward algorithm

Aim: Calculation of $\alpha_{tk}(\boldsymbol{\Theta})$, $1 \le t \le T$, $1 \le k \le K$.

Input: Parameter $\boldsymbol{\Theta}$, and discretization $\{V_k\}_{k=1}^K$.

1. Compute $\mathrm{P}(v_{1k} = 1|\boldsymbol{\Theta})$, $\mathrm{P}(v_{tk} = 1|v_{t-1,l} = 1, \boldsymbol{\Theta})$, $2 \le t \le T$, and $\mathrm{P}(\mathbf{Y}_t|v_{tk} = 1, \boldsymbol{\Theta})$, $1 \le t \le T$.
2. Compute
$$\begin{aligned} c_1(\boldsymbol{\Theta}) &:= p(\mathbf{Y}_1|\boldsymbol{\Theta}) \approx \textstyle\sum_{k=1}^K \mathrm{P}(\mathbf{Y}_1|v_{1k} = 1, \boldsymbol{\Theta})\mathrm{P}(v_{1k} = 1|\boldsymbol{\Theta}), \\ \alpha_{1k}(\boldsymbol{\Theta}) &= \mathrm{P}(\mathbf{Y}_1|v_{1k} = 1, \boldsymbol{\Theta})\mathrm{P}(v_{1k} = 1|\boldsymbol{\Theta})/c_1(\boldsymbol{\Theta}) \end{aligned},$$
for $1 \le k \le K$.
3. For $t = 2, \ldots, T$, compute $\alpha_{tk}(\boldsymbol{\Theta})$ by using $\alpha_{t-1,l}(\boldsymbol{\Theta})$,
$$\begin{aligned} c_t(\boldsymbol{\Theta}) &:= p(\mathbf{Y}_t|\mathcal{Y}_{t-1}, \boldsymbol{\Theta}) \\ &\approx \textstyle\sum_{k=1}^K \{\mathrm{P}(\mathbf{Y}_t|v_{tk} = 1, \boldsymbol{\Theta}) \sum_{l=1}^K \mathrm{P}(v_{tk} = 1|v_{t-1,l} = 1, \boldsymbol{\Theta})\alpha_{t-1,l}(\boldsymbol{\Theta})\}, \\ \alpha_{tk}(\boldsymbol{\Theta}) &= 1/c_t(\boldsymbol{\Theta})\mathrm{P}(\mathbf{Y}_t|v_{tk} = 1, \boldsymbol{\Theta}) \textstyle\sum_{l=1}^K \mathrm{P}(v_{tk} = 1|v_{t-1,l} = 1, \boldsymbol{\Theta})\alpha_{t-1,l}(\boldsymbol{\Theta}) \end{aligned}$$
for $1 \le k \le K$.

Output: $\{\alpha_{tk}(\boldsymbol{\Theta})\}_{t=1,k=1}^{T,K}$ and $\{c_t(\boldsymbol{\Theta})\}_{t=1}^T$.

$\beta_{tk}(\boldsymbol{\Theta})$. We start the calculation from the last time point $t = T$ and initialize $\beta_{Tk}(\boldsymbol{\Theta}) = 1$ for $1 \le k \le K$. Then, we calculate the smoothing posterior distribution $\beta_{tk}(\boldsymbol{\Theta})$ using the smoothing posterior distribution $\beta_{t+1,l}(\boldsymbol{\Theta})$, for $t = T - 1, \ldots, 1$. From the output of the forward-backward algorithm, the discrete posterior distribution of the source current $\mathbf{J}_t^{\mathrm{p}}$ is given in (2.9).

APPENDIX B

We start the EM algorithm from a reasonable initialization $\boldsymbol{\Theta}^{(0)}$. Let $\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)}) := L(\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}), \boldsymbol{\Theta}, \mathcal{Y}_T)$, $j = 1, 2, \ldots$. For the following iterations, we compute the posterior probability $\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(J-1)})$ to maximize $\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)})$ in the E-step. It is noted that we cannot treat the summation over $\mathcal{J}_T^{\mathrm{p}}$ in (2.13) at each time point independently. Since the expectation of the binary variable v_{tk} is just the probability that it takes the value 1, we have

$$\mathrm{P}\big(v_{tk} = 1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}\big) = \mathrm{E}\big(v_{tk}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}\big) = \sum_{\mathcal{J}_T^{\mathrm{p}}} \mathrm{P}\big(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}\big)v_{tk}$$

TABLE 6
Backward procedure of the forward-backward algorithm

Aim: Calculation of $\beta_{tk}(\boldsymbol{\Theta})$, $1 \le t \le T$, $1 \le k \le K$.

Input: Parameter $\boldsymbol{\Theta}$, discretization $\{V_k\}_{k=1}^K$, and $\{c_t(\boldsymbol{\Theta})\}_{t=1}^T$ from the forward procedure.

1. Compute $\mathrm{P}(v_{1k} = 1|\boldsymbol{\Theta})$, $\mathrm{P}(v_{tk} = 1|v_{t-1,l} = 1, \boldsymbol{\Theta})$, $2 \le t \le T$, and $\mathrm{P}(\mathbf{Y}_t|v_{tk} = 1, \boldsymbol{\Theta})$, $1 \le t \le T$.
2. Initialize $\beta_{Tk}(\boldsymbol{\Theta}) = 1$, for $1 \le k \le K$.
3. For $t = T - 1, \ldots, 1$, compute $\beta_{tk}(\boldsymbol{\Theta})$ by using $\beta_{t+1,l}(\boldsymbol{\Theta})$ and $c_{t+1}(\boldsymbol{\Theta})$,
$$\beta_{tk}(\boldsymbol{\Theta}) = 1/c_{t+1}(\boldsymbol{\Theta}) \textstyle\sum_{l=1}^K \beta_{t+1,l}(\boldsymbol{\Theta})\mathrm{P}(\mathbf{Y}_{t+1}|v_{t+1,l} = 1, \boldsymbol{\Theta})\mathrm{P}(v_{t+1,l} = 1|v_{tk} = 1, \boldsymbol{\Theta})$$
for $1 \le k \le K$.

Output: $\{\beta_{tk}(\boldsymbol{\Theta})\}_{t=1,k=1}^{T,K}$.

for $1 \le t \le T$, $1 \le k \le K$, and

$$\begin{aligned}\mathrm{P}(v_{t-1,l}=1, v_{tk}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) &= \mathrm{E}(v_{t-1,l}v_{tk}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) \\ &= \sum_{\mathcal{J}_T^{\mathrm{p}}} \mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) v_{t-1,l}v_{tk}\end{aligned}$$

for $2 \le t \le T$, $1 \le k \le K$. For the first term of $\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)})$ defined in (2.13), we have

$$\begin{aligned}&\sum_{\mathcal{J}_T^{\mathrm{p}}} \mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) \log p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta}) =: Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)}) \\ &\quad\approx \sum_{t=1}^{T}\sum_{k=1}^{K} \mathrm{P}(v_{tk}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) \log \mathrm{P}(\mathbf{Y}_t|v_{tk}=1, \boldsymbol{\Theta}) \\ &\qquad+ \sum_{t'=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K} \mathrm{P}(v_{t'-1,l}=1, v_{t'k}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) \\ &\qquad\cdot \log \mathrm{P}(v_{t'k}=1|v_{t'-1,l}=1, \boldsymbol{\Theta}) \\ &\qquad+ \sum_{k=1}^{K} \mathrm{P}(v_{1k}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}) \log \mathrm{P}(v_{1k}=1|\boldsymbol{\Theta}),\end{aligned} \tag{B.1}$$

where the complete likelihood function $p(\mathcal{Y}_T, \mathcal{J}_T^{\mathrm{p}}|\boldsymbol{\Theta})$ is approximated by the discrete distributions (2.5)–(2.7). For the jth iteration, we let $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)}) := \mathrm{P}(v_{tk}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)})$ be the intermediate discrete posterior distribution and $\eta_{t-1,l}^{t,k}(\boldsymbol{\Theta}^{(j-1)}) := \mathrm{P}(v_{t-1,l}=1, v_{tk}=1|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)})$ be the intermediate discrete joint posterior distribution. Instead of targeting the posterior distribution $\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)})$, we compute the intermediate posterior distributions $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ and $\eta_{t-1,l}^{t,k}(\boldsymbol{\Theta}^{(j-1)})$ during the E-step at the jth iteration. To be precise, we obtain $\alpha_{tk}(\boldsymbol{\Theta}^{(j-1)})$ and $\beta_{tk}(\boldsymbol{\Theta}^{(j-1)})$ from the forward-backward algorithm given the intermediate estimate $\boldsymbol{\Theta}^{(j-1)}$. Then, we have

$$\xi_{tk}(\boldsymbol{\Theta}^{(j-1)}) = \alpha_{tk}(\boldsymbol{\Theta}^{(j-1)})\beta_{tk}(\boldsymbol{\Theta}^{(j-1)})$$

for $1 \le t \le T$, and

$$\begin{aligned}\eta_{t-1,l}^{t,k}(\boldsymbol{\Theta}^{(j-1)}) &= \alpha_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\beta_{tk}(\boldsymbol{\Theta}^{(j-1)})\mathrm{P}(\mathbf{Y}_t|v_{tk}=1, \boldsymbol{\Theta}^{(j-1)}) \\ &\quad\cdot \mathrm{P}(v_{tk}=1|v_{t-1,l}=1, \boldsymbol{\Theta}^{(j-1)})/c_t(\boldsymbol{\Theta}^{(j-1)}),\end{aligned}$$

where $2 \le t \le T$, $1 \le k \le K$.

In the M-step of the jth iteration, we update $\boldsymbol{\Theta}^{(j)}$, in which the intermediate posterior distributions $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$ and $\eta_{t-1,l}^{t,k}(\boldsymbol{\Theta}^{(j-1)})$ are treated as constant. The maximization of $\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)})$ is equivalent to maximize (B.1), and it gives the closed form of the updates $\boldsymbol{\Theta}^{(j)}$; see Table 7.

We perform the E-step and M-step until the function $\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)})$ defined in (2.13) converges. The EM algorithm is summarized in Table 8.

APPENDIX C

PROOF OF THEOREM 2.1. In the jth iteration of the EM algorithm, we are given the estimate $\boldsymbol{\Theta}^{(j-1)}$ from the previous iteration. From (2.15) we have

$$\ell(\boldsymbol{\Theta}^{(j-1)}, \mathcal{Y}_T) = L(\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}), \boldsymbol{\Theta}^{(j-1)}, \mathcal{Y}_T). \tag{C.1}$$

TABLE 7
Closed form for updates of estimated parameters at each iteration in the EM algorithm

Update $\boldsymbol{\Theta}^{(j)} = \{\boldsymbol{\mu}_0^{(j)}, \boldsymbol{\Sigma}_0^{(j)}, A^{(j)}, \mathbf{b}^{(j)}, \boldsymbol{\Sigma}^{(j)}, \mathbf{V}^{(j)}\}$, given discretization $\{V_k\}_{k=1}^K$, intermediate posterior distributions $\{\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})\}_{t=1,k=1}^{T,K}$ and $\{\eta_{t-1,l}^{t,k}(\boldsymbol{\Theta}^{(j-1)})\}_{t=2,k,l=1}^{T,K}$, and constant moment parameter $\mathbf{q}$.

$$
\begin{aligned}
&\mathbf{d}_k := (\mathbf{c}_k^{\mathrm{T}}, \mathbf{q}^{\mathrm{T}})^{\mathrm{T}},\\
&\boldsymbol{\mu}_0^{(j)} = \textstyle\sum_{k=1}^K \xi_{1k}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_k,\\
&\boldsymbol{\Sigma}_0^{(j)} = \textstyle\sum_{k=1}^K \xi_{1k}(\boldsymbol{\Theta}^{(j-1)})(\mathbf{d}_k - \boldsymbol{\mu}_0^{(j)})(\mathbf{d}_k - \boldsymbol{\mu}_0^{(j)})^{\mathrm{T}},\\
&A^{(j)} = [(\textstyle\sum_{t=2}^T \sum_{k=1}^K \xi_{tk}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_k)(\sum_{t=2}^T \sum_{l=1}^K \xi_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_l^{\mathrm{T}}) - (T-1)\\
&\qquad \cdot (\textstyle\sum_{t=2}^T \sum_{k=1}^K \sum_{l=1}^K \eta_{t-1,l}^{tk}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_k\mathbf{d}_l^{\mathrm{T}})] \cdot [(\textstyle\sum_{t=2}^T \sum_{l=1}^K \xi_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_l)\\
&\qquad \cdot (\textstyle\sum_{t=2}^T \sum_{l=1}^K \xi_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_l^{\mathrm{T}}) - (T-1)(\sum_{t=2}^T \sum_{l=1}^K \xi_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_l\mathbf{d}_l^{\mathrm{T}})]^{-1},\\
&\mathbf{b}^{(j)} = (\textstyle\sum_{t=2}^T \sum_{k=1}^K \xi_{tk}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_k - A^{(j)} \sum_{t=2}^T \sum_{l=1}^K \xi_{t-1,l}(\boldsymbol{\Theta}^{(j-1)})\mathbf{d}_l)/(T-1),\\
&\boldsymbol{\Sigma}^{(j)} = \tfrac{1}{T-1}(\textstyle\sum_{t=2}^T \sum_{k=1}^K \sum_{l=1}^K \eta_{t-1,l}^{tk}(\boldsymbol{\Theta}^{(j-1)})(\mathbf{d}_k - A^{(j)}\mathbf{d}_l - \mathbf{b}^{(j)})(\mathbf{d}_k - A^{(j)}\mathbf{d}_l - \mathbf{b}^{(j)})^{\mathrm{T}}),\\
&\mathbf{V}^{(j)} = (\textstyle\sum_{t=1}^T \sum_{k=1}^K \xi_{tk}(\boldsymbol{\Theta}^{(j-1)})(\mathbf{Y}_t - \mathbf{B}(\mathbf{d}_k))(\mathbf{Y}_t - \mathbf{B}(\mathbf{d}_k))^{\mathrm{T}})/T.
\end{aligned}
$$

The update $\boldsymbol{\Theta}^{(j)}$ is obtained by maximizing the function $L(\mathrm{P}(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}), \boldsymbol{\Theta}, \mathcal{Y}_T)$; thus, we have

$$
\begin{aligned}
L\big(\mathrm{P}\big(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}\big), \boldsymbol{\Theta}^{(j-1)}, \mathcal{Y}_T\big) &\le L\big(\mathrm{P}\big(\mathcal{J}_T^{\mathrm{p}}|\mathcal{Y}_T, \boldsymbol{\Theta}^{(j-1)}\big), \boldsymbol{\Theta}^{(j)}, \mathcal{Y}_T\big)\\
&\le \ell\big(\boldsymbol{\Theta}^{(j)}, \mathcal{Y}_T\big),
\end{aligned}
\tag{C.2}
$$

where the second inequality comes from Jensen's inequality (2.14). From (C.1) and (C.2), we have that the EM sequence $\{\boldsymbol{\Theta}^{(j)}\}$ does not cause a decrease in the log-likelihood function. Under the regularity conditions, we refer the proof of Theorem 3 in Vaida (2005) for the convergence of the EM sequence. □

APPENDIX D

PROOF OF THEOREM 2.2. We first consider the case with two sources. From the EM algorithm with the switch procedure and the nonswitch procedure, there exists $\delta > 0$, such that $\|\hat{\boldsymbol{\Theta}}_{\mathrm{s}} - \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}\| \le \delta$.

In the nonswitch procedure, we calculate the posterior distribution of two sources $\mathrm{P}(\mathbf{J}_{t,1}^{\mathrm{p}} \in V_{k_1}, \mathbf{J}_{t,2}^{\mathrm{p}} \in V_{k_2}|\mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$ and can further calculate the marginal posterior distribution

$$
\mathrm{P}\big(\mathbf{J}_{t,n}^{\mathrm{p}} \in V_{k_n}|\mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}\big) = \sum_{k_{n'}=1}^K \mathrm{P}\big(\mathbf{J}_{t,1}^{\mathrm{p}} \in V_{k_1}, \mathbf{J}_{t,2}^{\mathrm{p}} \in V_{k_2}|\mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}\big),
$$

TABLE 8
EM algorithm

Initialize $\boldsymbol{\Theta}^{(0)}$. For $j = 1, 2, \ldots,$

1. E-step. Calculate the posterior distribution $\xi_{tk}(\boldsymbol{\Theta}^{(j-1)})$, for $1 \le t \le T$, $1 \le k \le K$, and the joint posterior distribution $\eta_{t-1,l}^{tk}(\boldsymbol{\Theta}^{(j-1)})$, for $2 \le t \le T$, $1 \le k, l \le K$.
2. M-step. Maximize the expected complete data log-likelihood defined in (B.1), $\boldsymbol{\Theta}^{(j)} = \mathrm{argmax}_{\boldsymbol{\Theta}}\, Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(j-1)})$.

where $n = 1, 2$ and $n' \neq n$. In the switch procedure, we focus on calculating the marginal posterior distribution $\mathrm{P}(\mathbf{J}^{\mathrm{p}}_{t,n} \in V_{k_n} | \mathbf{J}^{\mathrm{p}}_{t,n'} \in V_{k_{n'}}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})$, for $n = 1, 2$, $n' \neq n$. For the first source current $\mathbf{J}^{\mathrm{p}}_{t,1}$, we aim to calculate

$$
\begin{aligned}
&\left| \mathrm{P}(v_{tk_1} = 1 | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \mathrm{P}(v_{tk_1} = 1, v_{tk_2} = 1 | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \right| \\
&\quad = \left| \frac{\mathrm{P}(v_{tk_1} = 1, \mathcal{Y}_t | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \mathrm{P}(\mathcal{Y}_{T\setminus t} | v_{tk_1} = 1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})}{p(\mathcal{Y}_t | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) p(\mathcal{Y}_{T\setminus t} | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})} \right. \\
&\qquad \left. - \sum_{k_2=1}^{K} \frac{\mathrm{P}(v_{tk_1} = 1, v_{tk_2} = 1, \mathcal{Y}_t | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \mathrm{P}(\mathcal{Y}_{T\setminus t} | v_{tk_1} = 1, v_{tk_2} = 1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})}{p(\mathcal{Y}_t | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) p(\mathcal{Y}_{T\setminus t} | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})} \right| \\
&\quad = \left| \alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \right|,
\end{aligned}
\tag{D.1}
$$

where we let $\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) := \mathrm{P}(v_{tk_1} = 1, \mathcal{Y}_t | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) / p(\mathcal{Y}_t | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})$ and $\beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) := \mathrm{P}(\mathcal{Y}_{T\setminus t} | v_{tk_1} = 1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) / p(\mathcal{Y}_{T\setminus t} | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})$ in the switch procedure, and $\alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) := \mathrm{P}(v_{tk_1} = 1, v_{tk_2} = 1, \mathcal{Y}_t | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) / p(\mathcal{Y}_t | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$ and $\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) := \mathrm{P}(\mathcal{Y}_{T\setminus t} | v_{tk_1} = 1, v_{tk_2} = 1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) / p(\mathcal{Y}_{T\setminus t} | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$ in the nonswitch procedure. In order to bound (D.1), we analyze α's and β's from two procedures, recursively. We suppose the same ROI and mesh grids for the discretization $\{V_k\}_{k=1}^{K}$.

First, we analyze α's in the forward recursion. For $t = 1$, we have

$$
\alpha^{\mathrm{s}}_{1k_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) = \frac{1}{c^{\mathrm{s}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})} \mathrm{P}(\mathbf{Y}_1 | v_{1k_1} = 1, \mathbf{J}^{\mathrm{p}}_{1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \, \mathrm{P}(v_{1k_1} = 1 | \mathbf{J}^{\mathrm{p}}_{1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}),
$$

where $c^{\mathrm{s}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) = p(\mathbf{Y}_1 | \mathbf{J}^{\mathrm{p}}_{1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}})$ in the switch procedure for $1 \leq k_1 \leq K$, and

$$
\alpha^{\mathrm{ns}}_{1,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) = \frac{1}{c^{\mathrm{ns}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})} \mathrm{P}(\mathbf{Y}_1 | v_{1k_1} = 1, v_{1k_2} = 1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \mathrm{P}(v_{1k_1} = 1, v_{1k_2} = 1 | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}),
$$

where $c^{\mathrm{ns}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) = p(\mathbf{Y}_1 | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$ in the nonswitch procedure for $1 \leq k_1, k_2 \leq K$. From regularity conditions (C2) and (C3), there exists positive constants c and C, such that $c \leq c^{\mathrm{s}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}), c^{\mathrm{ns}}_1(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \leq C$, and we have

$$
\begin{aligned}
&\left| \alpha^{\mathrm{s}}_{1k_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \alpha^{\mathrm{ns}}_{1,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \right| \\
&\quad \leq c \left| \mathrm{P}(v_{1k_1} = 1 | \mathbf{J}^{\mathrm{p}}_{1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \mathrm{P}(v_{1k_1} = 1, v_{1k_2} = 1 | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \right|,
\end{aligned}
\tag{D.2}
$$

where c is a positive constant. Throughout the proof, we use c as a genetic symbol for positive constant. Since $\|\hat{\boldsymbol{\Theta}}_{\mathrm{s}} - \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}\| \leq \delta$, there exist $\varepsilon > 0$, such that

$$
\left| \mathrm{P}(v_{1k_1} = 1 | \mathbf{J}^{\mathrm{p}}_{1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \mathrm{P}(v_{1k_1} = 1, v_{1k_2} = 1 | \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \right| \leq \varepsilon / K^T,
\tag{D.3}
$$

given that $\|\mathbf{c}_{k_2^*} - \sum_{k_2=1}^{K}\sum_{k_1=1}^{K} \mathrm{P}(v_{1k_1}=1, v_{1k_2}=1|\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\mathbf{c}_{k_2}\| \le \epsilon$, where $\mathbf{c}_{k_2^*}$ is the location of the second source assumed in the switch procedure. From (D.2) and (D.3), we have

$$\left|\alpha^{\mathrm{s}}_{1k_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \alpha^{\mathrm{ns}}_{1,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right| \le c\varepsilon \quad \text{for } 1 \le k_1 \le K. \tag{D.4}$$

From item 3 in Table 5, we have

$$\begin{aligned}
&\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \\
&\quad = \frac{1}{c^{\mathrm{s}}_t(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})}\mathrm{P}\big(\mathbf{Y}_t|v_{tk_1}=1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big) \\
&\qquad \times \sum_{l_1=1}^{K} \mathrm{P}\big(v_{tk_1}=1|v_{t-1,l_1}=1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big)\alpha^{\mathrm{s}}_{t-1,l_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})
\end{aligned}$$

and

$$\begin{aligned}
&\alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \\
&\quad = \frac{1}{c^{\mathrm{ns}}_t(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})}\mathrm{P}(\mathbf{Y}_t|v_{tk_1}=1, v_{tk_2}=1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \\
&\qquad \times \sum_{l_1=1}^{K}\sum_{l_2=1}^{K} \mathrm{P}(v_{tk_1}=1, v_{tk_2}=1|v_{t-1,l_1}=1, v_{t-1,l_2}=1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\alpha^{\mathrm{ns}}_{t-1,l_1,l_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}),
\end{aligned}$$

for $t = 2, \ldots, T$. Therefore, we have

$$\begin{aligned}
&\left|\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2}^{K} \alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right| \\
&\quad \overset{(a)}{\le} c\sum_{l_1=1}^{K}\Bigg|\mathrm{P}\big(v_{tk_1}=1|v_{t-1,l_1}=1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big)\alpha^{\mathrm{s}}_{t-1,l_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \\
&\qquad - \sum_{k_2=1}^{K}\sum_{l_2=1}^{K} \mathrm{P}(v_{tk_1}=1, v_{tk_2}=1|v_{t-1,l_1}=1, v_{t-1,l_2}=1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \\
&\qquad \cdot \alpha^{\mathrm{ns}}_{t-1,l_1,l_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\Bigg| \\
&\quad \le c\sum_{l_1=1}^{K}\big|\mathrm{P}\big(v_{tk_1}=1|v_{t-1,l_1}=1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big)\big| \cdot \Bigg|\alpha^{\mathrm{s}}_{t-1,l_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) \\
&\qquad - \sum_{l_2=1}^{K} \alpha^{\mathrm{ns}}_{t-1,l_1,l_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\Bigg| + c\sum_{l_1=1}^{K}\Bigg|\sum_{l_2=1}^{K} \alpha^{\mathrm{ns}}_{t-1,l_1,l_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\Bigg| \\
&\qquad \cdot \Bigg|\mathrm{P}\big(v_{tk_1}=1|v_{t-1,l_1}=1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big) \\
&\qquad - \sum_{k_2=1}^{K} \mathrm{P}(v_{tk_1}=1, v_{tk_2}=1|v_{t-1,l_1}=1, v_{t-1,l_2}=1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\Bigg| \\
&\quad \overset{(b)}{\le} c\varepsilon/K^{T-t} \le c\varepsilon,
\end{aligned} \tag{D.5}$$

where (a) follows from (D.2), (b) follows from (D.3) and (D.4), $2 \le t \le T$, and $1 \le k_1 \le K$. Second, we analyze β's in the backward recursion. For $t = T$, we initialize $\beta^{\mathrm{s}}_{Tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) = \beta^{\mathrm{ns}}_{T,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) = 1$, for $1 \le k_1, k_2 \le K$. Thus,

$$\left|\beta^{\mathrm{s}}_{Tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \beta^{\mathrm{ns}}_{T,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right| = 0 \le \varepsilon / K^T. \tag{D.6}$$

From item 3 in Table 6, we have

$$\beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) = \frac{1}{c^{\mathrm{s}}_{t+1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})} \sum_{l_1=1}^{K} \beta^{\mathrm{s}}_{t+1,l_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})\, \mathrm{P}\big(\mathbf{Y}_{t+1} | v_{t+1,l_1} = 1, \mathbf{J}^{\mathrm{p}}_{t+1,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big)$$
$$\cdot \mathrm{P}\big(v_{t+1,l_1} = 1 | v_{tk_1} = 1, \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big)$$

and

$$\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) = \frac{1}{c^{\mathrm{ns}}_{t+1}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})} \sum_{l_1=1}^{K} \sum_{l_2=1}^{K} \beta^{\mathrm{ns}}_{t+1,l_1,l_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \mathrm{P}(\mathbf{Y}_{t+1} | v_{t+1,l_1} = 1, v_{t+1,l_2} = 1,$$
$$\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) \mathrm{P}(v_{t+1,l_1} = 1, v_{t+1,l_2} = 1 | v_{tk_1} = 1, v_{tk_2} = 1, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}),$$

for $t = T-1, \ldots, 1$. Using the same derivation in the forward recursion of (D.5), we have

$$\left|\beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right| \le c\varepsilon / K^{T-t+1}. \tag{D.7}$$

Thus, we have

$$\left|\mathrm{P}\big(v_{tk_1} = 1 | \mathbf{J}^{\mathrm{p}}_{t,2} \in V_{k_2}, \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{s}}\big) - \sum_{k_2=1}^{K} \mathrm{P}\big(v_{tk_1} = 1, v_{tk_2} = 1 | \mathcal{Y}_T, \hat{\boldsymbol{\Theta}}_{\mathrm{ns}}\big)\right|$$
$$\overset{(a)}{=} \Big|\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})\beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}}) + \alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}})\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})$$
$$- \sum_{k_2=1}^{K} \alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\Big|$$
$$\le |\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}))| \cdot \left|\beta^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right| + \left|\alpha^{\mathrm{s}}_{tk_1}(\hat{\boldsymbol{\Theta}}_{\mathrm{s}}) - \sum_{k_2=1}^{K} \alpha^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right|$$
$$\cdot \left|\beta^{\mathrm{ns}}_{t,k_1,k_2}(\hat{\boldsymbol{\Theta}}_{\mathrm{ns}})\right|,$$
$$\overset{(b)}{\le} c\varepsilon,$$

where (a) follows from (D.1), (b) follows from (D.4), (D.5), (D.6) and (D.7), for $1 \le t \le T$, $1 \le k_1 \le K$. We can obtain the same result for the second source $\mathbf{J}^{\mathrm{p}}_{t,2}$. Without loss of generality, we can extend the result to the case with N sources, where $N > 2$. □

Acknowledgments. The authors thank Professor Rob Kass and his collaborators for sharing their BCI data. The authors would also like to thank collaborator Professor Helen Zhou at the Duke-NUS medical school for the spatial working memory (SWM) experiments.

Z. Yao and Z. Fan's research is supported by the MOE Grants Tier 1 R-155-000-196-114 and Tier 2 R-155-000-184-112 at the National University of Singapore.

M. Hayashi's research is supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (A) 17H01280, (B) 16KT0017, and Kayamori Foundation of Informational Science Advancement.

SUPPLEMENTARY MATERIAL

Supplement to "Quantifying time-varying sources in magnetoencephalography—A discrete approach" (DOI: 10.1214/19-AOAS1321SUPP; .pdf). We include all materials omitted from the main text.

REFERENCES

ARULAMPALAM, M. S., MASKELL, S., GORDON, N. and CLAPP, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50** 174–188.

BAILLET, S. and GARNERO, L. (1997). A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* **44** 374–385.

BAILLET, S., MOSHER, J. C. and LEAHY, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* **18** 14–30.

BOTO, E., HOLMES, N., LEGGETT, J., ROBERTS, G., SHAH, V., MEYER, S. S., MUÑOZ, L. D., MULLINGER, K. J., TIERNEY, T. M. et al. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature* **555** 657.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DESTEXHE, A., CONTRERAS, D. and STERIADE, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.* **19** 4595–4608.

FELZENSZWALB, P. F., HUTTENLOCHER, D. P. and KLEINBERG, J. M. (2004). Fast algorithms for large-state-space HMMs with applications to web usage analysis. In *Advances in Neural Information Processing Systems* 409–416.

FUKUSHIMA, M., YAMASHITA, O., KNÖSCHE, T. R. and SATO, M. (2015). MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. *NeuroImage* **105** 408–427.

HÄMÄLÄINEN, M. S. and ILMONIEMI, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Med. Biol. Eng. Comput.* **32** 35–42. https://doi.org/10.1007/bf02512476

HÄMÄLÄINEN, M. S., HARI, R., ILMONIEMI, R. J., KNUUTILA, J. and LOUNASMAA, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Phys.* **65** 413–497. https://doi.org/10.1103/revmodphys.65.413

LAMUS, C., HÄMÄLÄINEN, M. S., TEMEREANCA, S., BROWN, E. N. and PURDON, P. L. (2012). A spatiotemporal dynamic distributed solution to the MEG inverse problem. *NeuroImage* **63** 894–909.

LIN, F.-H., WITZEL, T., AHLFORS, S. P., STUFFLEBEAM, S. M., BELLIVEAU, J. W. and HÄMÄLÄINEN, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* **31** 160–171.

LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198 https://doi.org/10.2307/2669847

LIU, S., POH, J.-H., KOH, H. L., NG, K. K., LOKE, Y. M., LIM, J. K. W., CHONG, J. S. X. and ZHOU, J. (2018). Carrying the past to the future: Distinct brain networks underlie individual differences in human spatial working memory capacity. *NeuroImage* **176** 1–10.

LONG, C. J., PURDON, P. L., TEMEREANCA, S., DESAI, N. U., HÄMÄLÄINEN, M. S. and BROWN, E. N. (2011). State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing. *Ann. Appl. Stat.* **5** 1207–1228. MR2849772 https://doi.org/10.1214/11-AOAS483

MOSHER, J. C., LEWIS, P. S. and LEAHY, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* **39** 541–557.

OU, W., HÄMÄLÄINEN, M. S. and GOLLAND, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* **44** 932–946.

PASCUAL-MARQUI, R. D., MICHEL, C. M. and LEHMANN, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* **18** 49–65. https://doi.org/10.1016/0167-8760(84)90014-x

RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77** 257–286.

SARVAS, J. (1984). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* **32** 11–12.

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.

UUTELA, K., HÄMÄLÄINEN, M. and SOMERSALO, E. (1999). Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10** 173–180.

VAIDA, F. (2005). Parameter convergence for EM and MM algorithms. *Statist. Sinica* **15** 831–840. MR2233916

VEEN, B., JOSEPH, J. and HECOX, K. (1992). Localization of intra-cerebral sources of electrical activity via linearly constrained minimum variance spatial filtering. In *Proceedings of IEEE Workshop on Statistical Signal and Array Processing* 1 526–529.

YAO, Z. and EDDY, W. F. (2014). A statistical approach to the inverse problem in magnetoencephalography. *Ann. Appl. Stat.* **8** 1119–1144. MR3262548 https://doi.org/10.1214/14-AOAS716

YAO, Z., ZHANG, Y., BAI, Z. and EDDY, W. F. (2018). Estimating the number of sources in magnetoencephalography using spiked population eigenvalues. *J. Amer. Statist. Assoc.* **113** 505–518. MR3832204 https://doi.org/10.1080/01621459.2017.1341411

YAO, Z., FAN, Z., HAYASHI, M. and EDDY, W. F. (2020). Supplement to "Quantifying time-varying sources in magnetoencephalography—A discrete approach." https://doi.org/10.1214/19-AOAS1321SUPP

ZHANG, J. and LIU, C. (2015). On linearly constrained minimum variance beamforming. *J. Mach. Learn. Res.* **16** 2099–2145. MR3450503 https://doi.org/10.1631/jzus.a1400270

ZHANG, J. and SU, L. (2015). Temporal autocorrelation-based beamforming with MEG neuroimaging data. *J. Amer. Statist. Assoc.* **110** 1375–1388. MR3449033 https://doi.org/10.1080/01621459.2015.1054488