



A spatiotemporal dynamic distributed solution to the MEG inverse problem

Camilo Lamus^{a,b,*}, Matti S. Hämäläinen^{c,d}, Simona Temereanca^{a,c},
Emery N. Brown^{a,b,d}, Patrick L. Purdon^{a,b}

^a Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

^b Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, USA

^c Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, USA

^d Harvard-MIT Division of Health Sciences and Technology, USA

ARTICLE INFO

Article history:

Received 2 September 2011

Revised 3 November 2011

Accepted 7 November 2011

Available online 30 November 2011

Keywords:

MEG/EEG

Source localization

Inverse problem

Dynamic spatiotemporal modeling

ABSTRACT

MEG/EEG are non-invasive imaging techniques that record brain activity with high temporal resolution. However, estimation of brain source currents from surface recordings requires solving an ill-conditioned inverse problem. Converging lines of evidence in neuroscience, from neuronal network models to resting-state imaging and neurophysiology, suggest that cortical activation is a distributed spatiotemporal dynamic process, supported by both local and long-distance neuroanatomic connections. Because spatiotemporal dynamics of this kind are central to brain physiology, inverse solutions could be improved by incorporating models of these dynamics. In this article, we present a model for cortical activity based on nearest-neighbor autoregression that incorporates local spatiotemporal interactions between distributed sources in a manner consistent with neurophysiology and neuroanatomy. We develop a dynamic Maximum a Posteriori Expectation-Maximization (dMAP-EM) source localization algorithm for estimation of cortical sources and model parameters based on the Kalman Filter, the Fixed Interval Smoother, and the EM algorithms. We apply the dMAP-EM algorithm to simulated experiments as well as to human experimental data. Furthermore, we derive expressions to relate our dynamic estimation formulas to those of standard static models, and show how dynamic methods optimally assimilate past and future data. Our results establish the feasibility of spatiotemporal dynamic estimation in large-scale distributed source spaces with several thousand source locations and hundreds of sensors, with resulting inverse solutions that provide substantial performance improvements over static methods.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) are non-invasive brain imaging techniques that provide high temporal resolution measurements of magnetic and electric fields at the scalp generated by the synchronous activation of neuronal populations. It has been estimated that a detectable signal can be recorded if as few as one in a thousand synapses become simultaneously active in an area of about 40 square millimeters of cortex (Hämäläinen et al., 1993). The exceptional time resolution of these techniques provides a unique window into the dynamics of neuronal process that cannot be obtained with other functional neuroimaging modalities, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), which measure brain activity indirectly through associated slow metabolic or cerebrovascular changes.

Localizing active regions in the brain from MEG and EEG data requires solving the neuromagnetic inverse problem, which consists

of estimating the cerebral current distribution underlying a time series of measurements at the scalp. The ill-posed nature of the electromagnetic inverse problem and the relatively large distance between the sensors and the sources limit the spatial resolution of MEG and EEG. For the inverse problem, solution of the corresponding forward problem in MEG/EEG, i.e., determining the measured magnetic and electric field at the scalp generated by a given distribution of neuronal currents, is a prerequisite. This problem can be solved by adopting a quasistatic approximation to Maxwell's equations, resulting in a linear map which relates the activity of arbitrarily located neuronal sources to the signals in a set of sensors (Hämäläinen et al., 1993; Moshier et al., 1999).

Two types of models have been proposed in the neuromagnetic inverse problem literature: equivalent current dipole models and distributed source models. Equivalent current dipole models are based on the assumption that a small set of current dipoles with unknown locations, amplitudes, and orientations can closely approximate the actual current distribution (Bertrand et al., 2001; Campi et al., 2008; Jun et al., 2005; Kiebel et al., 2008; Moshier et al., 1992; Somersalo et al., 2003; Uutela et al., 1998). Distributed source models, on the other hand, assume that the recorded activity results from the

* Corresponding author at: 77 Massachusetts Avenue Room 46-6057, Cambridge, MA 02139, USA.

E-mail address: lamus@mit.edu (C. Lamus).

activation of a spatial distribution of current dipoles with known locations (see Baillet et al. (2001) for a review). Most source localization methods assume that scalp measurements and their underlying sources are independent across time, and convenient probabilistic or computational prior constraints are imposed to obtain unique solutions. For example, inverse methods have been proposed that penalize current sources with large amplitude or power (Auranen et al., 2005; Hämäläinen and Ilmoniemi, 1994; Uutela et al., 1999; Van Veen et al., 1997), impose spatial smoothness (Pascual-Marqui et al., 1994), or favor focal estimates (Gorodnitsky et al., 1995). Furthermore, Bayesian methods have been used to obtain spatially sparse estimates, where components of the current source covariance are estimated directly (Mattout et al., 2006; Phillips et al., 2005) or additional hierarchical priors are assigned in order to compute posterior distributions (Nummenmaa et al., 2007; Sato et al., 2004; Wipf and Nagarajan, 2009). The assumption of temporal independence in all of these methods allows the inverse solution at each point in time to be computed individually, without regard for dynamics, treating the probability distribution of the underlying sources as static in time. While this approach is computationally convenient, it ignores the temporal structure observed in neural recordings at many different levels (Buzsáki and Draguhn, 2004), which could be used to improve inverse solutions.

Recent methods for source localization have incorporated temporal smoothness constraints as part of a general Bayesian framework. The approach taken in these methods is to specify an arbitrary prior distribution for the dipole sources in space and time, sometimes in terms of basis functions, with limited or space-time separable interactions in order to obtain simplified estimation algorithms (Baillet and Garnero, 1997; Bolstad et al., 2009; Daunizeau and Friston, 2007; Daunizeau et al., 2006; Friston et al., 2008; Greensite, 2003; Limpiti et al., 2009; Ou et al., 2009; Trujillo-Barreto et al., 2008; Zumer et al., 2008). Linear state-space models have also been used to model source current dynamics. However, in order to reduce computational complexity, these methods either apply spatially-independent approximations to their respective estimation algorithms (Galka et al., 2004; Yamashita et al., 2004) or a priori fix model specific parameters to avoid the problem of parameter estimation (Long et al., 2011). Overall, while these methods incorporate temporal structure in their models, they specify highly constrained spatiotemporal interactions, such as space-time separability or spatial independence, that may not accurately reflect dynamic relationships between different brain areas.

Converging lines of evidence from neurophysiology, biophysics, and neuroimaging illustrate that dynamic spatiotemporal interactions are a central feature of brain activity. Intracranial recordings in different species, including humans, exhibit strong spatial correlations during rest and task periods that persist up to distances of 10 mm along the cortical surface (Bullock et al., 1995; Destexhe et al., 1999; Leopold et al., 2003). These local spatial interactions are supported neuroanatomically by axonal collateral projections from pyramidal cells that spread laterally at distances of up to 10 mm (Nunez, 1995). Biophysical spatiotemporal dynamic models of neuronal networks at various levels of abstraction have been effective in reproducing properties of electromagnetic scalp recordings seen during both normal and disease states (David et al., 2005; Gross et al., 2001; Izhikevich and Edelman, 2008; Jirsa et al., 2002; Kim and Robinson, 2007; Robinson et al., 2005; Sotero et al., 2007; Wright et al., 2004). Furthermore, fMRI and PET studies have shown temporally coherent fluctuation in activation within widely distributed cortical networks during resting-state and experimentally administered task periods (Fox and Raichle, 2007; Fox et al., 2005; Gusnard and Raichle, 2001; Raichle et al., 2001).

In this article we present a new dynamic source localization method that models local spatiotemporal interactions between distributed cortical sources in a manner consistent with neurophysiology and

neuroanatomy, and then uses this model to estimate an inverse solution. Specifically, (1) we describe a model of the spatiotemporal dynamics based on nearest-neighbors multivariate autoregression along the cortical surface; (2) We develop an algorithm for dynamic estimation of cortical current sources and model parameters from MEG/EEG data based on the Kalman Filter, the Fixed Interval Smoother, and the Expectation–Maximization (EM) algorithms; (3) We derive expressions to relate our dynamic estimation formulas to those of standard static algorithms; (4) We apply our spatiotemporal dynamic method to simulated experiments of focal and distributed cortical activation as well as experimental data from a human subject. Portions of this work have been previously presented in Lamus et al. (2007).

Methods

Measurement model

In an MEG/EEG experiment, we obtain a temporal set of recordings from hundreds of sensors located above the scalp. The data are sampled by N_y sensors at times $\{\Delta t\}_{t=1}^T$, where Δ is the sampling interval and T is the number of measurements in time. Let $y_{n,t}$ denote the measurement at time t in sensor n , and define $y_t \equiv [y_{1,t}, y_{2,t}, \dots, y_{N_y,t}]'$ as the $N_y \times 1$ vector of measurements at all sensors at time t . We assume that the measurements were generated by N_x current dipole sources distributed on the cortical surface and oriented perpendicular to it. Let $x_{n,t}$ denote the source amplitude of the n th dipole at time t , and define $x_t \equiv [x_{1,t}, x_{2,t}, \dots, x_{N_x,t}]'$ as the $N_x \times 1$ vector of cortical source activity in all considered locations, or cortical state vector, at time t . Typically, $N_y \sim$ a few hundred, and $N_x \sim$ several thousand. The relationship between the measurement vector y_t and the cortical state vector x_t is given by the measurement equation,

$$y_t = Gx_t + v_t, \quad (1)$$

where G is the $N_y \times N_x$ lead field gain matrix computed using a quasi-static approximation of the Maxwell's equations (Hämäläinen et al., 1993), i.e., the solution of the forward problem, and v_t is a $N_y \times 1$ Gaussian white noise vector with zero mean covariance matrix C representing background instrumental and environmental noise independent from x_t for all time points. Since the orientation of the current generators of the electromagnetic field, i.e., the apical dendrites of pyramidal cells, is perpendicular to the cortical surface, the choice of fixing the current dipole orientation along this direction is justified (Hämäläinen et al., 1993). Nevertheless, our development can be easily extended to account for unconstrained source orientations.

Spatiotemporal dynamical source model

As we pointed in the Introduction section, evidence from neurophysiology studies, neuroanatomy, biophysics, and neuroimaging suggests that cortical activation is a distributed spatiotemporal dynamic process (Bullock et al., 1995; David et al., 2005; Destexhe et al., 1999; Fox and Raichle, 2007; Fox et al., 2005; Gross et al., 2001; Gusnard and Raichle, 2001; Izhikevich and Edelman, 2008; Jirsa et al., 2002; Kim and Robinson, 2007; Leopold et al., 2003; Nunez, 1995; Raichle et al., 2001; Robinson et al., 2005; Sotero et al., 2007; Wright et al., 2004). Because spatiotemporal dynamics of this kind are fundamental to brain physiology, inverse solutions could be greatly improved by incorporating models that approximate these dynamics.

One way to model local spatiotemporal connections of this type is to use a nearest-neighbor autoregressive model. In this autoregressive model, neuronal currents at a given point in time and space $x_{n,t}$ are a function of past neuronal currents within a small local neighborhood along the cortical surface and a small perturbation $w_{n,t}$ that

accounts for unknown factors affecting the evolution of cortical currents,

$$x_{n,t} = \lambda \left[\underbrace{a_n x_{n,t-1}}_{\text{Past activity}} + \underbrace{(1-a_n) \sum_{i \in N(n)} d_{n,i} x_{i,t-1}}_{\text{Past activity of neighbors}} \right] + \underbrace{w_{n,t}}_{\text{Unaccounted factors}}, \quad (2)$$

where $N(n)$ is the index set of nearest neighbors of the n th dipole source, a_n is a scalar between 0 and 1 that weighs the past neuronal currents between the n th dipole source and its set of nearest neighbors $N(n)$, and λ is a positive scalar used to guarantee the stability of the cortical state dynamics. The state noise $w_t \equiv [w_{1,t}, w_{2,t}, \dots, w_{N_x,t}]'$ is a $N_x \times 1$ Gaussian vector with zero mean and covariance matrix Q , independent from the measurement noise v_t . The weights $d_{n,i}$, which represent the individual influence of neighboring sources in the autoregression, are assumed to decay with the distance from the n th to the i th dipole source. A simple relation that reflects this modeling assumption is to make the weights inversely proportional to the distance between dipoles,

$$d_{n,i} \propto \frac{1}{\text{distance from } n\text{th to } i\text{th dipole}}, \quad (3)$$

where the proportionality constant is chosen to make the sum of the weights over the index i equal 1, allowing the power (prior variance) in every dipole $x_{n,t}$ to remain approximately constant over time.

Fig. 1 illustrates this nearest-neighbors autoregressive model. The left panel shows the cortical surface reconstructed from high-resolution magnetic resonance images (MRI) using *Freemurfer* (Dale et al., 1999; Fischl et al., 1999), where the caudal portion is represented by its triangulated mesh of dipoles. The zoomed-in panel in the right isolates a dipole source (central red dot) and its corresponding nearest-neighbor dipoles (green dots). At a given point in time, the activity of the central dipole (red dot) is a function of its past activity and the weighted activity of a small neighborhood of dipole sources (green dots), where the weights (black dashed arrows) are inversely proportional to the distance from the central dipole to its neighbor. We employ a cortical surface representation with $N_x = 5124$ dipole sources, with an average distance of 6.2 mm between nearest neighbors, yielding a model that is consistent with the local spatial properties of intracranial electrophysiology and neuroanatomy. In Appendix A, we analyze how modifications in our spatial model, such as those arising from choosing a different weighting scheme or a misspecification of the weights a_n and $d_{n,i}$ (Eq. (2)), influence the smoothness encoded a priori in the dipole covariance. There we show that as long as the modification in the spatial model is not too large, the smoothness modeled in the dipole sources is not dramatically altered.

We can readily rewrite Eq. (2) as the multivariate autoregressive model,

$$x_t = Fx_{t-1} + w_t, \quad (4)$$

where the state feedback matrix F encompasses the neighborhood interactions between the sources at time t in terms of the sources in the previous time step $t-1$:

$$(F)_{n,i} = \begin{cases} \lambda a_n & \text{if } n = i \\ \lambda (1-a_n) d_{n,i} & \text{if } i \in N(n), \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

and $(F)_{n,i}$ is the n th row of the i th column of F . The initial dipole source vector x_0 is assumed to be Gaussian with zero mean and covariance matrix Σ_0 , and independent of $\{w_t\}_{t=1}^T$.

Prior model for the parameters

The model defined in the [Measurement model section](#) and the [Spatiotemporal dynamical source model section](#) yields a probabilistic “forward model” for the dynamic generation of the electromagnetic scalp recording. The next step in the description of our model is to define prior densities for the unknown parameters $\{G, C, F, Q, \Sigma_0\}$. We should note that while in theory all the parameters in our model have uncertainties, some of them can be fixed a priori based on knowledge of the system under investigation. For example, the lead field matrix G can be computed using the boundary-element model based on high-resolution magnetic resonance images (MRI) (Hämäläinen and Sarvas, 1989) and the measurement noise covariance C can be reliably estimated from empty room recordings or pre-stimulus data. The covariance Σ_0 can be updated as described in the [Inference with a dynamic Maximum a Posteriori Expectation–Maximization algorithm \(dMAP-EM\) section](#). The state feedback matrix F is constructed as indicated in the [Spatiotemporal dynamical source model section](#) to incorporate nearest-neighbor interaction in order to model basic local intracortical dynamic connections. The selection of the values of λ and a_n in F (Eq. (5)) is discussed in the [Initialization of algorithms section](#).

We note that the process of fixing the lead field G and state feedback F matrices can be seen as the analog of selecting the matrix of covariates or regressors in a linear model, where the experimenter’s knowledge guides the selection of the regressor as means of testing a scientific hypothesis based on collected data. In our case, the regressors in F , which are fixed based on neurophysiological modeling considerations, determine how much the state x_t can be explained by its past x_{t-1} , and the unexplained portion is left to the state noise term w_t . In the same way, the regressors in G , obtained from geometrical and biophysical knowledge, separate the measured MEG y_t into

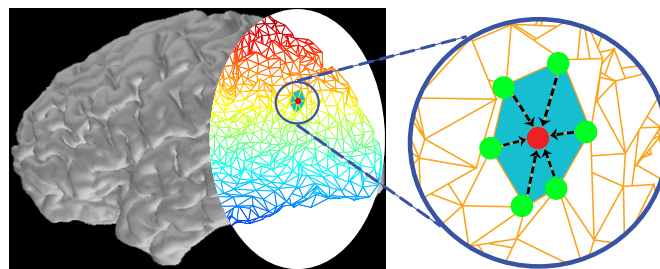


Fig. 1. Illustration of the spatiotemporal dynamic source model. The left panel shows the reconstructed cortical surface where the caudal portion of cortex is depicted as a triangulated mesh of dipole sources. The zoomed-in panel in the right shows a dipole source (central red dot) and its nearest-neighbor dipole sources (green dots). At a given point in time, the activity at the central dipole (red dot) is a function of its past activity and the weighted activity of a small neighborhood of dipole sources (green dots), plus a perturbation that represents unknown factors affecting the dipole source activity. The weights (black dashed arrows) are inversely proportional to the distance from the central dipole.

components representing the signals coming from the brain x_t , and the instrument and environmental noise v_t .

The remaining unknown parameters, which must be estimated from data, are the elements of the state noise covariance matrix Q . We assume that state noise covariance matrix Q is diagonal,

$$Q(\theta) \equiv \text{diag}(\theta), \quad (6)$$

where the parameter vector is $\theta = [\theta_1, \theta_2, \dots, \theta_{N_x}]'$, thus yielding a number of unknowns in the order of thousands ($N_x \approx 5000$ in our case).

A common probability model for the parameter vector θ is the conjugate prior distribution. In this case the conjugate prior follows an inverse gamma density:

$$p(\theta) = \prod_{n=1}^{N_x} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\theta_n}\right)^{\alpha+1} \exp\left(\frac{-\beta}{\theta_n}\right) \quad (7)$$

where the hyperparameters α and β are set such that the mean of this prior is consistent with the order of magnitude in the model units, and the variance is maximized yielding a non-informative prior (see Appendix B).

Inference with a dynamic Maximum a Posteriori Expectation–Maximization algorithm (dMAP-EM)

In order to localize active regions of cortex from scalp recordings, we have to derive estimates for the sequence of source amplitudes

$$\{x_t\}_{t=1}^T \equiv \{x_1, x_2, \dots, x_T\} \quad (8)$$

and parameters

$$\theta = [\theta_1, \theta_2, \dots, \theta_{N_x}]' \quad (9)$$

in the model. Specifically, for the parameters, we use the Maximum a Posteriori (MAP) estimate,

$$\hat{\theta}_{\text{MAP}} \equiv \underset{\theta}{\text{argmax}} p(\theta | \{y_t\}_{t=1}^T), \quad (10)$$

where $p(\theta | \{y_t\}_{t=1}^T)$ is the posterior density of the parameters θ conditioned on the full set of measurements

$$\{y_t\}_{t=1}^T \equiv \{y_1, y_2, \dots, y_T\}. \quad (11)$$

For the source amplitudes, we calculate the Empirical Bayes estimate, i.e., the conditional mean of the state vector given the full set of measurements $\{y_t\}_{t=1}^T$ and the estimate of the parameters in the model $\hat{\theta}_{\text{MAP}}$,

$$x_{t|T} \equiv E(x_t | \{y_t\}_{t=1}^T, \hat{\theta}_{\text{MAP}}), \quad (12)$$

where the notation in the subscript of $x_{t|T}$ indicates that we are conditioning on the measurements from time 1 until time T .

Once we obtain the MAP estimate of the parameters $\hat{\theta}_{\text{MAP}}$, the Empirical Bayes estimate of the source amplitudes (Eq. (12)) can be readily obtained using the well-known Kalman Filter (Kalman, 1960) and Fixed Interval Smoother algorithms (Rauch et al., 1965). However, the direct optimization required to compute $\hat{\theta}_{\text{MAP}}$ (Eq. (10)) would be computationally intractable. Therefore, we developed a dynamic Maximum a Posteriori Expectation–Maximization (dMAP-EM) algorithm to estimate source amplitudes and parameters in our model (Eqs. (10) and (12)) based on the work of Shumway and Stoffer (1982) and Green (1990). The Expectation–Maximization algorithm (Dempster et al., 1977) is a general iterative method to obtain Maximum Likelihood or Maximum a Posteriori estimates when the observed data can be regarded as incomplete. In our case we

treat the sequence of measurement and dipole sources, $\{y_t\}_{t=1}^T$, $\{x_t\}_{t=0}^T$ as the *complete data*, and iterate performing an E-step followed by an M-step until convergence is achieved. In each iteration of the algorithm, the expectations (E-step section) can be computed with the Kalman Filter (Kalman, 1960), Fixed Interval Smoother (Rauch et al., 1965), and lag-covariance (de Jong and Mackinnon, 1988) recursions, and the maximization (M-step section) can be obtained in closed form. In each iteration we can evaluate the posterior density of the parameters θ using the *innovations* form (Kitagawa and Gersch, 1996) (Evaluation of the log-posterior density section).

E-step

The dMAP-EM algorithm is initialized with the parameters $Q(\theta^{(0)})$ and $\Sigma^{(0)}$. In the r th iterate of the E-step, the algorithm computes the conditional expectation of the complete data log-likelihood $p(\{y_t\}_{t=1}^T, \{x_t\}_{t=0}^T | \theta)$, given the observed data $\{y_t\}_{t=1}^T$ and the previous estimates of the parameters $\theta^{(r-1)}$, with an added term for the log-prior density:

$$U(\theta | \theta^{(r-1)}) \equiv E[\log p(\{y_t\}_{t=1}^T, \{x_t\}_{t=0}^T | \theta) | \{y_t\}_{t=1}^T, \theta^{(r-1)}] + \log p(\theta). \quad (13)$$

We emphasize that the expectation in Eq. (13) is computed with respect to $p(\{x_t\}_{t=0}^T | \{y_t\}_{t=1}^T, \theta^{(r-1)})$, where we have conditioned on the full set of measurements and the parameter estimate of the previous iteration.

Before continuing with the algorithm we establish the following variables to simplify the notation. We define the conditional mean

$$x_{t|s}^{(r)} \equiv E(x_t | \{y_k\}_{k=1}^s, \theta^{(r-1)}), \quad (14)$$

the conditional covariance

$$P_{t|s}^{(r)} \equiv \text{Cov}(x_t | \{y_k\}_{k=1}^s, \theta^{(r-1)}), \quad (15)$$

and the conditional lag-covariance

$$P_{t,t-1|s}^{(r)} \equiv \text{Cov}(x_t, x_{t-1} | \{y_k\}_{k=1}^s, \theta^{(r-1)}), \quad (16)$$

where the subscript in Eqs. (14), (15), and (16) indicate that we are conditioning on the measurements from time 1 until time s , and the superscript r refers to the iteration number. For example, in Eq. (12), by setting s equal to the number of samples in time T we indicate that we are conditioning on the full set of measurements.

As shown in Appendix C, the conditional expectations in the function $U(\theta | \theta^{(r-1)})$ (Eq. (13)) can be obtained in closed form yielding

$$U(\theta | \theta^{(r-1)}) = -\frac{1}{2} \left\{ c_{N_x} + \log |\Sigma_0^{(r-1)}| + \text{tr} \left[\Sigma_0^{(r-1)-1} (P_{0|T}^{(r)} + x_{0|T}^{(r)} x_{0|T}^{(r)'}) \right] \right\} \quad (17)$$

$$-\frac{1}{2} \left\{ c_{N_x} T + T \log |Q| + \text{tr} [Q^{-1} A^{(r)}] \right\}$$

$$-\frac{1}{2} \left\{ c_{N_y} T + T \log |C| + \text{tr} [C^{-1} B^{(r)}] \right\}$$

$$+ \sum_{n=1}^{N_x} \left\{ \log \frac{\beta^\alpha}{\Gamma(\alpha)} - (\alpha + 1) \log \theta_n - \frac{\beta}{\theta_n} \right\},$$

where c_{N_x} and c_{N_y} are constants not depending on θ , and

$$A^{(r)} \equiv A_1^{(r)} - A_2^{(r)} F' - F A_2^{(r)'} + F A_3^{(r)} F' \quad (18)$$

$$B^{(r)} \equiv \sum_{t=1}^T \left[(y_t - G x_{t|T}^{(r)}) (y_t - G x_{t|T}^{(r)})' + G P_{t|T}^{(r)} G' \right], \quad (19)$$

where

$$\begin{aligned} A_1^{(r)} &\equiv \sum_{t=1}^T \left(P_{t|T}^{(r)} + x_{t|T}^{(r)} x_{t|T}^{(r)'} \right) \\ A_2^{(r)} &\equiv \sum_{t=1}^T \left(P_{t,t-1|T}^{(r)} + x_{t|T}^{(r)} x_{t-1|T}^{(r)'} \right) \\ A_3^{(r)} &\equiv \sum_{t=1}^T \left(P_{t-1|T}^{(r)} + x_{t-1|T}^{(r)} x_{t-1|T}^{(r)'} \right). \end{aligned} \quad (20)$$

The conditional expectations and covariances in Eq. (19) can be computed with the Kalman Filter (Kalman, 1960), Fixed Interval Smoother (Rauch et al., 1965), and lag-covariance (de Jong and Mackinnon, 1988) recursions.

The Kalman Filter, Fixed Interval Smoother, and lag-covariance recursions. We can use the forward and backward recursions (Kitagawa and Gersch, 1996) to compute the desired expectations and covariances. In the r th iteration we initialize the recursion with $x_{0|0}^{(r)} = 0$ and $P_{0|0}^{(r)} = \Sigma_0^{(r-1)}$. For the forward pass, $t = 1, 2, \dots, T$, we compute the predictions:

$$\begin{aligned} x_{t|t-1}^{(r)} &= F x_{t-1|t-1}^{(r)} \\ P_{t|t-1}^{(r)} &= F P_{t-1|t-1}^{(r)} F' + Q(\theta^{(r-1)}), \end{aligned} \quad (21)$$

and filtered estimates:

$$\begin{aligned} K_t &= P_{t|t-1}^{(r)} G' (G P_{t|t-1}^{(r)} G' + C)^{-1} \\ x_{t|t}^{(r)} &= x_{t|t-1}^{(r)} + K_t (y_t - G x_{t|t-1}^{(r)}) \\ P_{t|t}^{(r)} &= (I - K_t G) P_{t|t-1}^{(r)}, \end{aligned} \quad (22)$$

and for the backward pass, $t = T - 1, \dots, 0$, we find the smoothed estimates:

$$\begin{aligned} J_t &= P_{t|t}^{(r)} F' P_{t+1|t}^{(r)-1} \\ x_{t|T}^{(r)} &= x_{t|t}^{(r)} + J_t (x_{t+1|T}^{(r)} - x_{t+1|t}^{(r)}) \\ P_{t|T}^{(r)} &= P_{t|t}^{(r)} + J_t (P_{t+1|T}^{(r)} - P_{t+1|t}^{(r)}) J_t'. \end{aligned} \quad (23)$$

The lag-covariances can be obtained using the covariance smoothing algorithm (de Jong and Mackinnon, 1988):

$$P_{t,t-1|T}^{(r)} = P_{t|T}^{(r)} J_{t-1}' \quad (24)$$

We should note that the conditional mean and covariance that we ultimately use for source localization (Eq. (12)) correspond to the Fixed Interval Smoother estimate (FIS, Eq. (23)) obtained on the final iteration of the MAP-EM algorithm. The FIS provides an estimate of the state based on the full set of measurements, and results in improved performance in terms of reduced error covariance compared to the Kalman Filter (KF), which uses a causal subset of the data (Anderson and Moore, 2005).

In Appendix D, we show that the Kalman Filter and Fixed Interval Smoother estimates can be interpreted as the solution to a penalized least squares problem, analogous to that of the well-known L_2 minimum-norm estimate (MNE) (Hämäläinen and Ilmoniemi, 1994). Viewed from this perspective, we can readily see that the FIS, KF, and MNE source localization methods are structurally similar, but with an important difference in how the prior mean for the source activity is represented at a given time. The KF and FIS optimally update their prior means by assimilating data from the past, and, in addition, from the future measurements in the case of FIS. In contrast,

the MNE assumes that the prior mean is zero at all times and favors source values close to zero.

M-step

The M-step in the r th iterate is achieved by maximizing with respect to the parameters θ the function $U(\theta|\theta^{(r-1)})$ computed in the E-step (Eq. (17)):

$$\theta^{(r)} \equiv \underset{\theta}{\operatorname{argmax}} U(\theta|\theta^{(r-1)}). \quad (25)$$

It is easy to see that the maxima is achieved at

$$\theta_n^{(r)} = \frac{(A^{(r)})_{n,n} + 2\beta}{T + 2(\alpha + 1)}, \quad (26)$$

where $(A^{(r)})_{n,n}$ is the n th row of the n th column of $A^{(r)}$ (Eq. (18)), and α and β are the hyperparameters defined in Eq. (7).

Evaluation of the log-posterior density

The dMAP-EM algorithm iterates for $r = 1, 2, \dots, r_o$ performing E-steps and M-steps until the logarithm of the posterior density of the parameters

$$\log p(\theta|\{y_t\}_{t=1}^T) = \underbrace{\log p(\{y_t\}_{t=1}^T|\theta)}_{\text{Log-likelihood}} + \underbrace{\log p(\theta)}_{\text{Log-prior}} - \underbrace{\log p(\{y_t\}_{t=1}^T)}_{\text{Log-evidence}} \quad (27)$$

evaluated at $\theta = \theta^{(r)}$ reaches an asymptote at some iteration r_o .

We can evaluate the logarithm of the likelihood, i.e., the first term in Eq. (27), using the innovations form (Kitagawa and Gersch, 1996),

$$\begin{aligned} \log p(\{y_t\}_{t=1}^T|\theta^{(r)}) &= -\frac{N_y T}{2} \log(2\pi) + \frac{1}{2} \sum_{t=1}^T \log |G P_{t|t-1}^{(r)} G' + C| \\ &\quad + \frac{1}{2} \sum_{t=1}^T (y_t - G x_{t|t-1}^{(r)})' (G P_{t|t-1}^{(r)} G' + C)^{-1} (y_t - G x_{t|t-1}^{(r)}), \end{aligned} \quad (28)$$

and the log-prior is obtained is obtained from Eq. (7):

$$\log p(\theta^{(r)}) = \sum_{n=1}^{N_s} \left\{ \log \frac{\beta^\alpha}{\Gamma(\alpha)} - (\alpha + 1) \log \theta_n^{(r)} - \frac{\beta}{\theta_n^{(r)}} \right\}. \quad (29)$$

Since the evidence in the data, $p(\{y_t\}_{t=1}^T)$, is a constant not depending on θ , we do not need to compute it in any iteration.

Summary of the dMAP-EM algorithm

The algorithm is initialized with parameters $\theta^{(0)}$ and $\Sigma_0^{(0)}$. For the r th iteration, we set the state noise covariance $Q(\theta^{(r-1)}) = \text{diag}(\theta^{(r-1)})$ and initial state covariance $P_{0|0}^{(r)} = \Sigma_0^{(r-1)}$. We then perform an E-step by running the Kalman Filter, Fixed Interval Smoother, and lag-covariance recursion (E-step section), and perform an M-step (Eq. (26)) to update the parameters $\theta^{(r)}$. At each iteration we can update Σ_0 with the heuristic $\Sigma_0^{(r)} = P_{0|T}^{(r-1)}$. The algorithm iterates for $r = 1, 2, \dots, r_o$, performing an E-step followed by an M-step until the log-posterior density evaluated at $\theta^{(r_o)}$ converges (Evaluation of the log-posterior density section). The Maximum a Posteriori (MAP) estimate of the parameters is then $\hat{\theta}_{\text{MAP}} = \theta^{(r_o)}$, and the Empirical Bayes estimate of the sources amplitudes is $x_{t|T} = x_{t|T}^{(\hat{\theta})}$. Fig. 2 summarizes the computations performed in our dMAP-EM algorithm.

Initialize parameters $\theta^{(0)}$ and $\Sigma_0^{(0)}$.
 Iterate until convergence at some iteration r_o ($r = 1, 2, \dots, r_o$).
 E-step.

1. Assign $Q(\theta^{(r-1)}) = \text{diag}(\theta^{(r-1)})$ and $P_{0|0}^{(r)} = \Sigma_0^{(r-1)}$.
2. For $t = 1, 2, \dots, T$ use the Kalman Filter (Eqs. 21 and 22) and Fixed Interval Smoother (Eq. 23) recursion to find $x_{t|T}^{(r)}$ and $P_{t|T}^{(r)}$.
3. Compute the lag-covariance $P_{t,t-1|T}^{(r)}$ (Eq. 24).
4. Find $A^{(r)}$ (Eq. 18) using Equation 20 and the estimates $x_{t|T}^{(r)}$, $P_{t|T}^{(r)}$, and $P_{t,t-1|T}^{(r)}$.

M-Step.

1. Update $\theta^{(r)}$ using $A^{(r)}$ and Equation 26.
2. Assign $\Sigma_0^{(r)} = P_{0|T}^{(r-1)}$.

Convergence check.

1. Compute $p(\theta^{(r)} | \{y_t\}_{t=1}^T)$ using Equations 27, 28, and 29.
2. If $p(\theta^{(r)} | \{y_t\}_{t=1}^T)$ converged, then $r_o = r$ and the estimates are $\hat{\theta}_{MAP} = \theta^{(r_o)}$ and $x_{t|T} = x_{t|T}^{(r_o)}$. If not, got back to E-step part 1.

Fig. 2. The dMAP-EM algorithm.

Initialization of algorithms

The state feedback matrix F was constructed as indicated in the Spatiotemporal dynamical source model section to incorporate nearest-neighbor interaction in order to model basic local intracortical connections. To evaluate the weights $d_{n,i}$ (Eqs. (3) and (5)), the distance between dipoles was obtained from the triangular tessellation of the cortical surface. The value of a_n in Eq. (5) was set to 0.51 to indicate that the source amplitude in the present time is approximately equally weighted between the past activation in the same dipole and its neighborhood, and the choice $0.5 < a_n \leq 1$ assures that F is nonsingular.¹ The value of λ in Eq. (5) was set to 0.95: this constrains the modulus of the largest eigenvalue of F to be strictly less than 1, and guarantees stability in the source model (Eq. (4)). We should note that although it would be interesting to estimate the parameters $\{a_n\}_{n=1}^{N_x}$ and λ , this would add ~5000 degrees of freedom to the estimation task, making this option unfeasible.

The measurement noise covariance C was estimated as the sample covariance from empty room recordings. To initialize the algorithms

we set the source covariance at time zero ($\Sigma_0^{(0)}$) as a multiple of the identity, $\Sigma_0^{(0)} = \sigma_x^2 I$, where

$$\sigma_x^2 \equiv \text{SNR} \frac{N_y}{\text{tr}(C^{-1/2} G G' C^{-1/2})}, \quad (30)$$

thus approximating the power signal-to-noise ratio (SNR) of the measurements. The state noise covariance $Q(\theta^{(0)})$ was initialized as a multiple of the identity, $Q(\theta^{(0)}) = \sigma_w^2 I$, where σ_w^2 was set to a tenth of σ_x^2 to establish a starting value in scale with the data SNR.

Data analysis

Design of simulation studies

We employed simulation studies to compare the source localization performance of four methods: (1) The L_2 minimum-norm estimate (MNE, $x_i^{(MNE)}$ in Eq. (D.6)) (Hämäläinen and Ilmoniemi, 1994); (2) An extension of MNE, which we call static Maximum a Posteriori Expectation–Maximization (sMAP-EM), where the variance of each dipole source is estimated similarly to that shown in Friston et al. (2008) and Wipf and Nagarajan (2009); (3) The FIS without the estimation of the parameters θ ($x_{t|T}^{(0)}$ in Eq. (23), i.e., the smoothed estimate in the first iteration of our algorithm); (4) And our dMAP-EM algorithm ($x_{t|T}^{(r)}$ in Eq. (23), i.e., the smoothed estimate in the last iteration of our algorithm).

We should note that sMAP-EM uses a static source model and the inverse gamma prior (Eq. (7)) for the source variances. Therefore, the sMAP-EM is similar to our method, but with the state feedback matrix

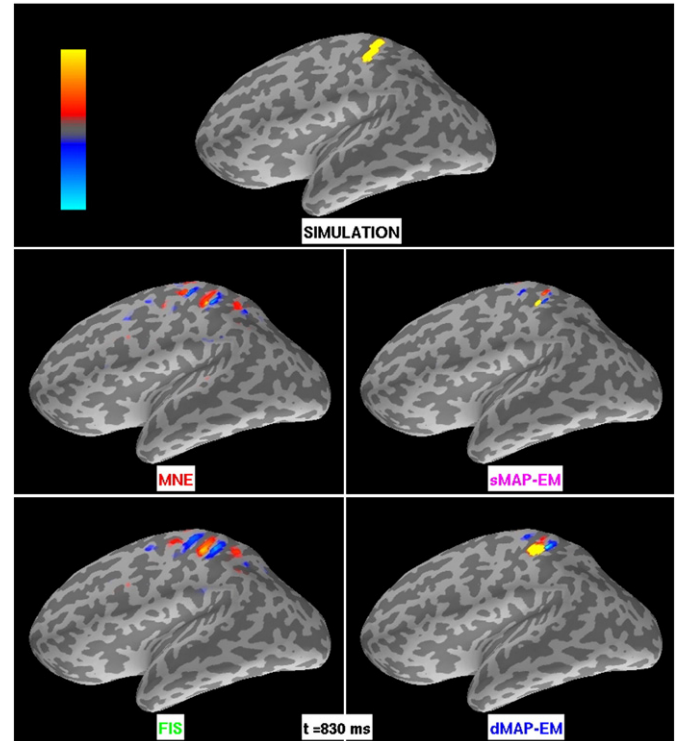


Fig. 3. Large cortical patch simulation results for MNE, sMAP-EM, FIS, and dMAP-EM methods. The intensity maps represent the amplitude of simulated and estimated sources. The colorbar's maximum (bright yellow) and minimum (bright blue) corresponds to ± 3 nA for all methods. The top panel shows the simulated activity. The center left panel shows the estimates obtained with MNE, while the bottom left panel shows the FIS estimates. Both of these methods yielded estimates that extended far beyond the simulated active region. The sMAP-EM estimates (center right panel) show a spatial extent significantly smaller than the true activation. The bottom right panel shows the dMAP-EM estimates, whose spatial extent closely matches the simulated active region.

¹ Since we can write the feedback matrix as $F = \lambda a \left(I - \frac{a-1}{a} D \right)$, where $(D)_{i,j} = d_{i,j}$ and D is a stochastic matrix (i.e., its row sum equals 1), then $\frac{1}{\lambda a} \left(\sum_{k=0}^{\infty} \left(\frac{a-1}{a} D \right)^k \right) \lambda a \left(I - \frac{a-1}{a} D \right) = I$ and F is invertible, if $\left\| \frac{a-1}{a} D \right\|_2 = \left| \frac{a-1}{a} \right| \|D\|_2 < 1$. Since D is a stochastic matrix, the modulus of its eigenvalues is less or equal than 1, and $\|D\|_2 \leq 1$. It follows that $a > 0.5$ is sufficient to make F nonsingular.

F set to zero, i.e., $F = 0$. It provides a baseline to compare how the matrix F , as specified to model local intracortical connections, can be used as covariates or regressors to explain the source activity x_t in terms of its past x_{t-1} .

We constructed two simulated data sets with active regions of different sizes to compare the performance of these methods in cases where the activity is distributed across a large area, and where it is highly focal:

- *Large patch*: we selected a large active region within primary somatosensory cortex (Fig. 3 top panel);
- *Small patch*: we chose a small active region over primary auditory cortex (Fig. 5 top panel).

In order to avoid committing an “inverse crime”, where simulated data come from the same source space and dynamic system used in estimation, we simulated cortical activation on a highly discretized mesh with $\sim 150,000$ dipole sources in each hemisphere and a temporal generating model differing from that of our autoregressive model. The time course of the sources on each active patch was simulated as a 10-Hz sinusoidal oscillation over a period of one second in order to emulate a realistic MEG experiment, $\sin(2\pi \cdot 10 \cdot \Delta t)$, where the sampling frequency $1/\Delta$ was 200 Hz thus yielding 200 time samples.

The lead field matrix was computed with the MNE software package (Hämäläinen and Sarvas, 1989) using a single-compartment boundary-element model (BEM) based on high-resolution MRIs processed with *Freesurfer* (Dale et al., 1999; Fischl et al., 1999) from a human subject. The measurement equation (Eq. (1)) was then used to obtain the simulated MEG recordings, where the measurement noise was set to achieve a power signal-to-noise ratio (SNR) of 5, a

value typical for MEG measurements, with signal amplitudes scaled uniformly across the active regions to achieve this SNR.

Results of simulation studies

We compared dipole source estimates and their 95% Bayesian credibility (confidence) intervals (CIs) from the static MNE, sMAP-EM, FIS, and dMAP-EM algorithms, using the simulated measurements described in the Design of simulation studies section. The 95% CIs are obtained from the diagonal elements of the posterior covariance matrix of each method. We should note that these CIs are Empirical Bayes estimates since we are also conditioning on the model parameters θ . Specifically, the 95% CIs of the FIS dipole source estimates are equal to ± 2 times the diagonal elements of posterior covariance in the first EM iteration $P_{1|T}^{(1)}$ (Eq. (23)), while the CIs of the dMAP-EM source estimates equal ± 2 times the diagonal elements of $P_{t|T}^{(T)}$, i.e. the posterior covariance of the last EM iteration. Similarly, the CIs for the MNE and sMAP-EM estimates can be obtained by using the well-known formula for the posterior covariance in a jointly Gaussian model, or equivalently by setting $F = 0$ in our formulation.

Fig. 3 shows the spatial extent of the estimated activity obtained from the large patch simulation. In particular, these intensity maps represent the amplitude of dipole currents $x_{n,t}$ at a particular time, as opposed to (scaled or normalized) statistical maps. The MNE extends beyond the simulated area in primary sensory cortex, overlapping pre-central gyrus, parietal cortex, and a small active area in the temporal lobe. The sMAP-EM estimates were highly focal in comparison to the extent of the simulated active region. The FIS estimates were similar to the MNEs, but with better coverage of the simulated

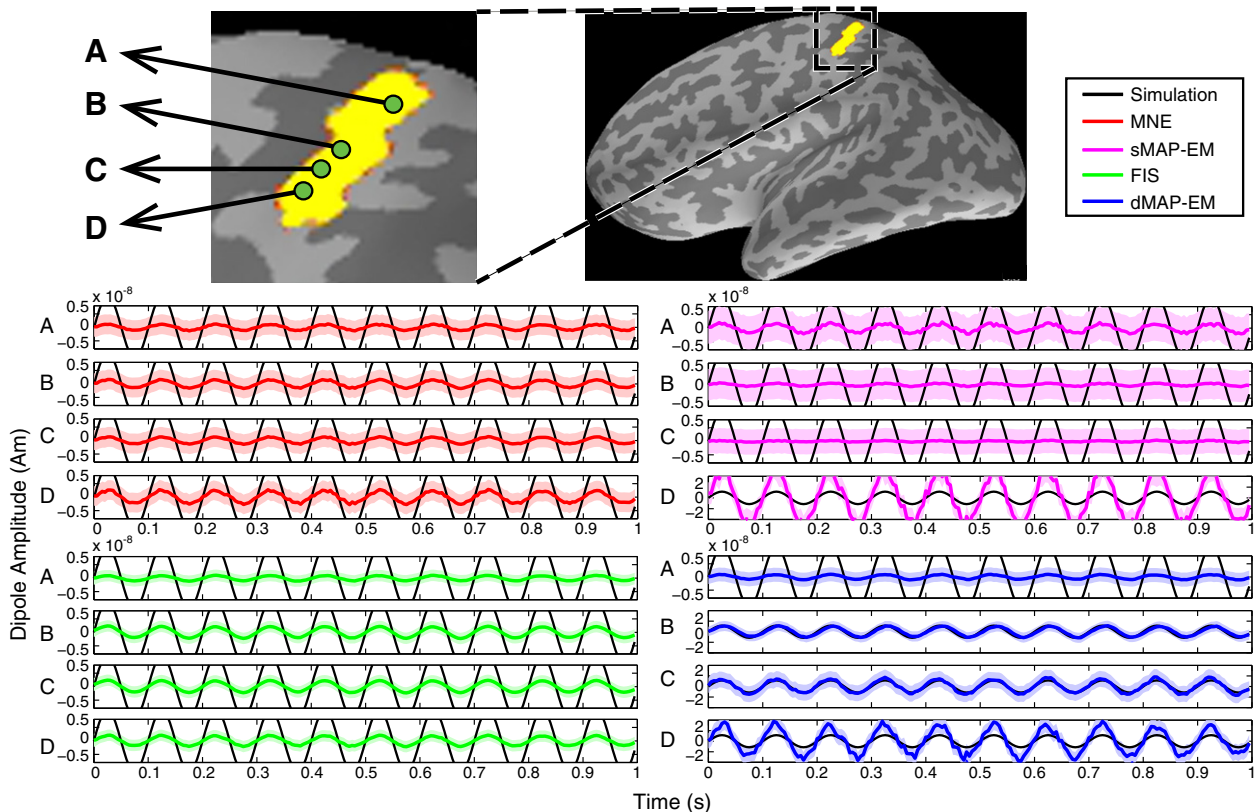


Fig. 4. Time course estimates for the large patch simulation. The upper panel shows a zoomed-in view of the simulated active area, where green dots represent the selected dipoles labeled A, B, C, and D. The black lines represent simulated sources, while the colored lines represent estimated sources with 95% CIs (colored shading). The center left panel shows the estimated time course of the MNE method in red, and the bottom left panels show the FIS estimates in green. These methods showed poor tracking performance and could not recover the amplitude of the simulated dipole sources. The center right panel shows the sMAP-EM estimates in magenta. This method either underestimated or overestimated the true source amplitude, and showed very large CIs. The bottom right panel shows the estimated time courses for the dMAP-EM method. For 3 of the 4 sources shown (B, C, and D), the dMAP-EM method tracks the simulated time course very closely, significantly better than MNE, FIS, or sMAP-EM, while showing small CIs.

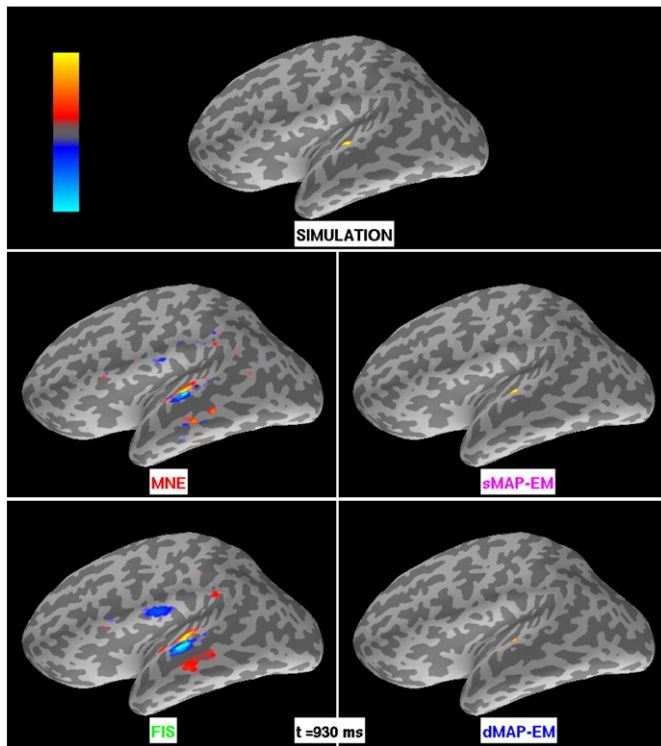


Fig. 5. Small cortical patch simulation results for MNE, sMAP-EM, FIS, and dMAP-EM methods. The intensity maps represent the amplitude of simulated and estimated sources. The colorbar's maximum (bright yellow) and minimum (bright blue) corresponds to ± 3.5 nA for MNE and FIS, and ± 70 nA for the simulated data, sMAP-EM, and dMAP-EM method. The top panel shows the simulated activity. The center left and bottom left panels show the estimates obtained with the MNE and FIS methods, respectively. Both methods yield spatially distributed estimates that extend far beyond the true underlying focal active patch. The center right and bottom right panels show the estimates obtained with the sMAP-EM and dMAP-EM methods, respectively. These algorithms yielded focal estimates that closely match the spatial extent of the simulated data.

active area. The dMAP-EM method yielded dipole source estimates whose spatial extent closely matched the simulated active region, as observed with the yellow and red hues over primary somatosensory cortex.

Fig. 4 compares the temporal tracking performance for representative dipole sources located inside the active area for the large patch simulation. The MNE time series greatly underestimates the amplitude of the true simulated time series and presents wide CIs. The sMAP-EM either underestimates (sub-panels B and C) or overestimates (sub-panel D) the source amplitude and presents larger CIs than the other methods. Similar to MNE, FIS time series estimates also fail to track the true underlying signal accurately, however, they exhibit smaller CIs. For most dipole sources (sub-panels B, C, and D) dMAP-EM closely tracks the true underlying time series while maintaining small CIs. We present in Supplementary information 1 a video with the results of this simulation study that visualizes dynamically the spatial localization and temporal tracking performance.

The estimation results from MNE, sMAP-EM, FIS, and dMAP-EM for the small patch simulation are shown in Fig. 5. Again, the intensity maps represent the amplitude of the source estimates, and not a statistical map. In MNE and FIS the estimates extended beyond the active area to cover regions in the inferior temporal lobe, parietal cortex, and the inferior frontal areas. In contrast, dipole source estimates obtained with sMAP-EM and dMAP-EM are focal and closely match the spatial extent of the active simulated area.

Fig. 6 compares the estimated and simulated time courses for representative dipole sources in the small patch simulation. As shown in

the upper left panel, the dipoles labeled A, C and D are outside the active region, while the dipole labeled B is inside the active region. Neither MNE nor FIS are able to accurately track the true underlying active time series (sub-panel B), and both methods produce spurious activation in locations outside the focal active patch (sub-panels A, C, and D). However, the FIS method shows smaller CIs. The sMAP-EM method tracks the active source (sub-panel B), although it slightly overestimates it. In the inactive locations (sub-panels A, C, and D) the sMAP-EM shows small but noisy estimates with very large CIs. The dMAP-EM method accurately tracks the time series for the active dipole source (sub-panel B), while correctly showing small, near-zero activity outside the focal active patch (sub-panels A, C, and D). Furthermore, the dMAP-EM shows small CIs in all cases. We present in the Supplementary information 2 a video with the results of this simulation study.

Error analyses

In this section we evaluate the source localization accuracy of our dMAP-EM methodology in comparison to MNE, sMAP-EM, and FIS. Specifically, we are interested in evaluating the sensitivity and specificity of these source localization methods, as well as the correlation between the simulated sources and their estimates. To do this, we computed receiver operating characteristic (ROC) curves (See Moon and Stirling (2000), or Darvas et al. (2004) for an application in the MEG/EEG inverse problem) and root mean square errors (RMSE) of sources estimates in the simulation studies described in Design of simulation studies and Results of simulation studies sections.

Receiver operating characteristic (ROC) curves

The ROC technique is used to evaluate the performance of binary detection systems, where the null hypothesis H_0 generally represents the absence of an underlying signal while the alternative hypothesis H_A denotes its presence. This analysis is done by determining the relation between the detection probability,

$$p_D \equiv \Pr(\text{reject } H_0 | H_A \text{ is true}), \quad (31)$$

and the false alarm probability,

$$p_{FA} \equiv \Pr(\text{reject } H_0 | H_0 \text{ is true}), \quad (32)$$

as we vary a threshold for the hypothesis test, where in practice, the probabilities are estimated by proportions from simulation studies.

For this analysis we consider a source signal to be inactive (H_0 is true) when the simulated n th dipole source at time t equals zero ($x_{n,t}^{(SIM)} = 0$), while a source signal is considered active (H_A is true) when $x_{n,t}^{(SIM)} \neq 0$. For a given estimation method, we define $\hat{x}_{n,t}$ as the n th dipole source estimate at time t . Furthermore, we reject the null hypothesis H_0 to indicate that the source estimate is active with respect to a test threshold $c > 0$ if $|\hat{x}_{n,t}| > c$. Consequently, for a specific threshold c , an estimate of the detection probability \hat{p}_D is given by the fraction of events where we correctly detected an active source, i.e., when the dipole source estimate was considered active ($|\hat{x}_{n,t}| > c$) given that the underlying true source was active ($x_{n,t}^{(SIM)} \neq 0$). Similarly, the estimate of false alarm probability \hat{p}_{FA} is given by the proportion of events where we considered the source estimate to be active ($|\hat{x}_{n,t}| > c$) but the underlying true source was inactive ($x_{n,t}^{(SIM)} = 0$) and made a “false alarm” (See Appendix E for details).

We computed ROC curves for the large and small patch simulation studies for MNE (red), sMAP-EM (magenta), FIS (green), and dMAP-EM (blue) estimates. The results from the large and small patch simulation studies are shown in the sub-panels A and B of Fig. 7, respectively. The ROC curves for dMAP-EM showed a superior source detection accuracy in both simulation studies. In the large patch study, the dMAP-EM achieved a $\sim 90\%$ detection of true active sources

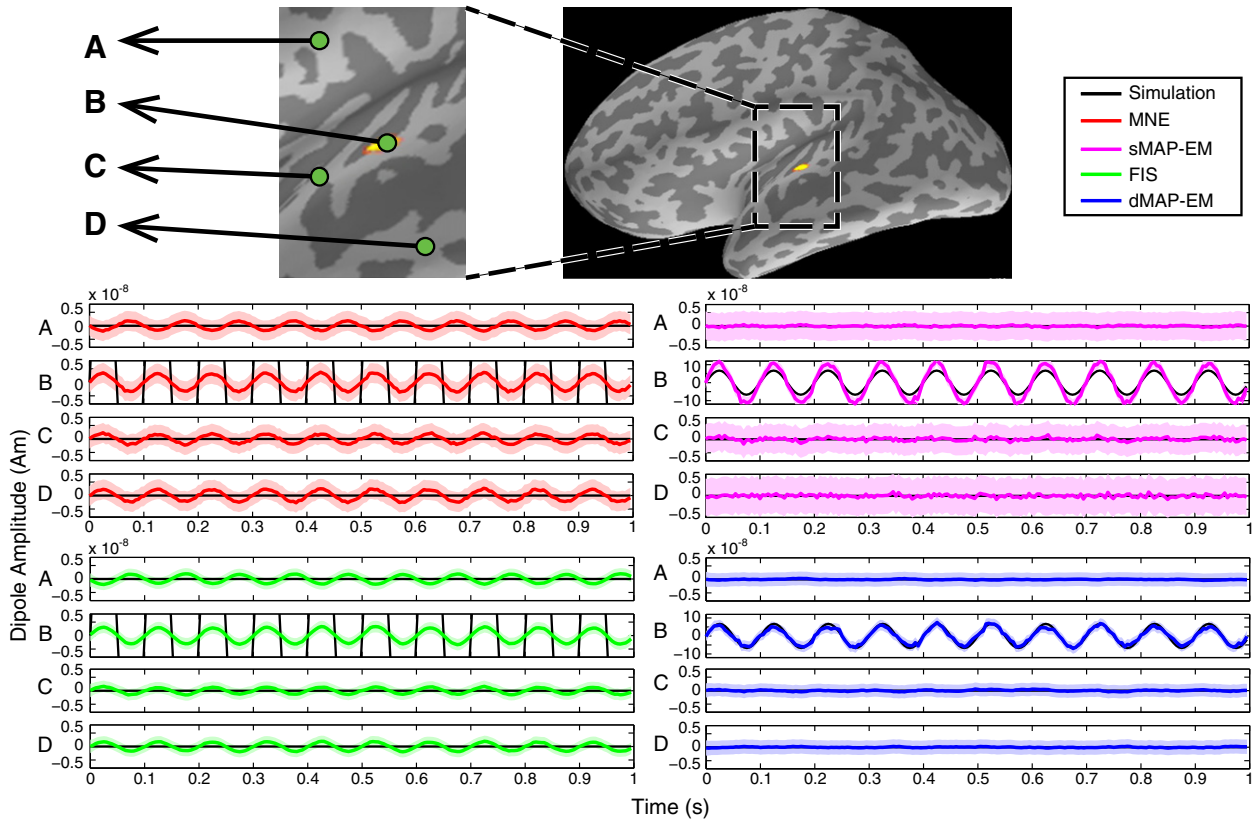


Fig. 6. Time course of estimation results for the small patch simulation. The upper shows a zoomed-in view of the simulated cortical area where green dots represent selected dipoles labeled A, B, C, and D. The black lines represent simulated sources, while the colored lines represent estimated sources with 95% CIs (colored shading). The center left panel shows the estimated time courses for the MNE method in red, and the lower left panels show the FIS estimates in green. These methods showed poor tracking performance for the focal active source (sub-panel B), and produced spurious activity in inactive simulated regions (sub-panels A, C, and D). The center right panel shows the sMAP-EM estimates, which tracks the active source (sub-panel B), but presents noisy estimates with very large CIs in the inactive sources (sub-panels A, C, and D). The lower right panel shows the estimated time courses of the dMAP-EM method, which shows accurate tracking performance for the focal active source (sub-panel B), and near-zero activity outside the active area (sub-panels A, C, and D), with very small CIs.

($\hat{p}_D \approx 0.9$) with as few as $\sim 2\%$ false alarms ($\hat{p}_{FA} \approx 0.02$), while the other methods required at least $\sim 40\%$ false alarms to achieve the same detection accuracy. Similarly, in the small patch simulation, the dMAP-EM achieved a $\sim 95\%$ detection of true active sources with as few as $\sim 2\%$ false alarms, whereas the other methods required at least $\sim 40\%$ false alarms to detect $\sim 95\%$ of the true active sources. Furthermore, the area under the ROC curve (see tables in sub-panels A and B of Fig. 7), which is a comprehensive measure of the detection accuracy, was greater in the dMAP-EM method than in the other methods, thus indicating a significant improvement in the source localization accuracy of our method.

Root mean square errors (RMSE)

To further characterize the accuracy of each method, we evaluated the deviation of the dipole source estimates $\hat{x}_{n,t}$ in relation to the true simulated sources $x_{n,t}^{(SIM)}$ in absolute terms by computing root mean square errors (RMSE) in each simulation study. Specifically, for the MNE, FIS, sMAP-EM, and dMAP-EM methods, we computed the RMSE of each dipole source ($n = 1, 2, \dots, N_x$) over the length in time ($T = 200$) of the simulation,

$$RMSE_n \equiv \sqrt{\frac{\sum_{t=1}^T (\hat{x}_{n,t} - x_{n,t}^{(SIM)})^2}{T}}, \quad (33)$$

and calculated summary statistics of these errors. We separated the summary statistics for the RMSEs of dipole sources inside and outside the simulated active region to disentangle the origin of the estimation errors. Given that the number of dipoles inside the active patch was

small, we computed scatter plots and the sample mean of the RMSE of sources in this region (Fig. 7, sub-panels C and D). Since the number of dipole sources located outside the active patch was relatively large, we computed box-plot summaries of the RMSE of sources from this region, displaying 0.01, 0.25, 0.5, 0.75, and 0.99 approximate quantiles (Fig. 7, sub-panels E and F). We should note that in box-plots the approximate 0.75 and 0.25 quantiles are represented by the bottom (thin line) and top (thick colored line) of the box, while the 0.5 quantile (median) is denoted by the colored dashed line across the box. The approximate 0.99 and 0.01 quantiles are given by the top and bottom “whiskers”, i.e., the horizontal black lines, and the gray crosses outside whiskers are considered outliers.

Sub-panels C and D of Fig. 7 show the scatter plots and the average (dashed lines) RMSE of dipole sources inside the active region in the large and small patch simulation studies, respectively. In both simulation scenarios, the dMAP-EM estimates showed a significant improvement in the average RMSE compared to the other methods. In the large patch simulation, the dMAP-EM method yielded an average RMSE of 7 nAm, which represents a reduction of $\sim 5.4\%$, $\sim 2.7\%$, and $\sim 23\%$ in the average of errors relative to MNE, FIS, and sMAP-EM methods, respectively. Similarly, the dMAP-EM method resulted in an average RMSE of 26 nAm in the small patch simulation, showing a reduction of $\sim 42\%$ relative to MNE and FIS, and $\sim 33\%$ relative to the sMAP-EM errors. In summary, the average RMSE of sources in active regions is significantly reduced in the dMAP-EM method for both large and small patch simulation scenarios.

Sub-panels E and F of Fig. 7 show the RMSE box-plots of dipole sources located outside the active region in the large and small patch-simulations, respectively. In both simulated scenarios, our dMAP-EM

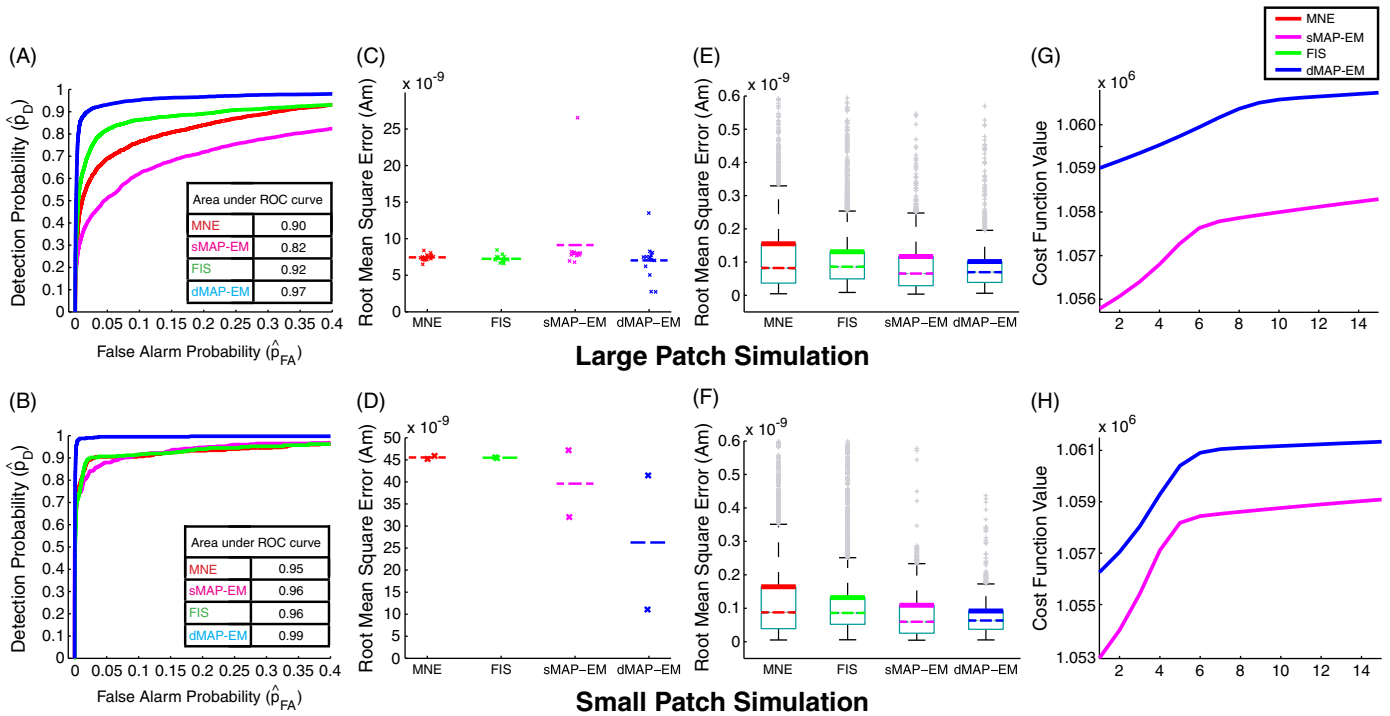


Fig. 7. ROC curves, RMSE, and convergence of algorithms. ROC curves and area under the ROC curves from the large patch simulation (sub-panel A) and the small patch simulation (sub-panel B) show that the dMAP-EM method outperforms MNE, FIS, and sMAP-EM methods as it detects more active dipole sources while making significantly fewer false alarms. Sub-panels C and D show scatter plots of the RMSE of dipole sources inside the active region in the large patch and small path simulations, respectively. As indicated by the average RMSE (dashed lines), the dMAP-EM method significantly reduces the estimation error compared to the other methods. Sub-panels E and F show box-plot summaries of the RMSE of dipole sources outside the active region from the large patch and small patch simulations, respectively. The dMAP-EM method significantly reduces the 0.99 quantiles (top horizontal black line or “whisker”) and 0.75 quantiles (thick colored line at top of “box”) of the RMSEs in relation to the other methods. Sub-panels G and H show the convergence of the cost function optimized by the dMAP-EM and sMAP-EM algorithms in the large patch and small patch simulation studies, respectively. In all cases, the cost function reaches a plateau in less than 15 iterations.

method yielded the smallest quantiles among the analyzed methods. The reduction in RMSE relative to MNE, FIS, and sMAP-EM are shown in Table 1. The dMAP-EM method improved the 0.99 and 0.75 RMSE quantiles by 9% to 51%, showing that the dMAP-EM estimates significantly and robustly improved source localization accuracy.

Convergence and computational requirements

The convergence of the dMAP-EM algorithm was assessed by tracking the logarithm of the posterior density (Eq. (27)),

$$\text{cost}(\theta^{(r)}) \equiv \log p(\{y\}_{t=1}^T | \theta^{(r)}) + \log p(\theta^{(r)}), \quad (34)$$

Table 1

Percentage reduction of the RMSE quantiles for dipole sources outside the active region under dMAP-EM, compared to MNE, FIS, and sMAP-EM.

	Quantile		0.5	0.75	0.99
	Method				
Large patch	dMAP-EM (nAm)		0.06	0.10	0.19
	Reduction with respect to:	MNE	25%	33%	40%
		FIS	25%	23%	24%
		sMAP-EM	0%	9%	20%
Small patch	dMAP-EM (nAm)		0.06	0.09	0.17
	Reduction with respect to:	MNE	25%	43%	51%
		FIS	25%	30%	32%
		sMAP-EM	0%	10%	26%

omitting the log-evidence term, which does change during EM iteration. Sub-panels G and H of Fig. 7 show the convergence, i.e., the cost evaluated at each iteration, of the dMAP-EM and sMAP-EM algorithms for the large patch and small patch simulation studies, respectively. In both simulation scenarios, the cost function reaches a plateau in less than 15 iterations.

The runtime of the dMAP-EM algorithm per EM iteration is effectively $O(TN_x^3)$ (Mendel, 1971), assuming that the number of sensors N_y is fixed and much smaller than the number of dipole sources N_x ($N_y \ll N_x$), and T is the number of measurements in time. The algorithm was implemented in Matlab (The MathWorks, Natick, MA) and run on a dual 6-core Linux workstation at 2.67 GHz with 24 GB RAM. In our analyses the number of dipole sources was $N_x = 5124$, the number of sensor was $N_y = 204$, and the number of measurements in time was $T = 200$. Since we did not attempt any kind of model reduction procedure, the algorithm yielded a computation time of ~ 2.5 h per EM iteration.

Analysis of experimental data from human subjects

We applied the MNE, FIS, sMAP-EM, and dMAP-EM algorithms to *mu*-rhythm MEG data from a human subject. The *mu*-rhythm originates from motor and somatosensory cortices, and consists of oscillations with 10 and 20 Hz components. Data were collected using a 306-channel Neuromag Vectorview MEG system at Massachusetts General Hospital. The subject was instructed to rest with eyes open during the recording. We recorded 12 min of data at a sampling frequency of 601 Hz with a bandwidth of 0.1 to 200 Hz, and later down-sampled to 200.3 Hz. The data were visually inspected to select 1 s of strong *mu*-rhythm activity, as evidenced by its characteristic “ μ ” shape, for subsequent analyses. In Fig. 8, we show the results of the three methods where the intensity maps represent the amplitude of

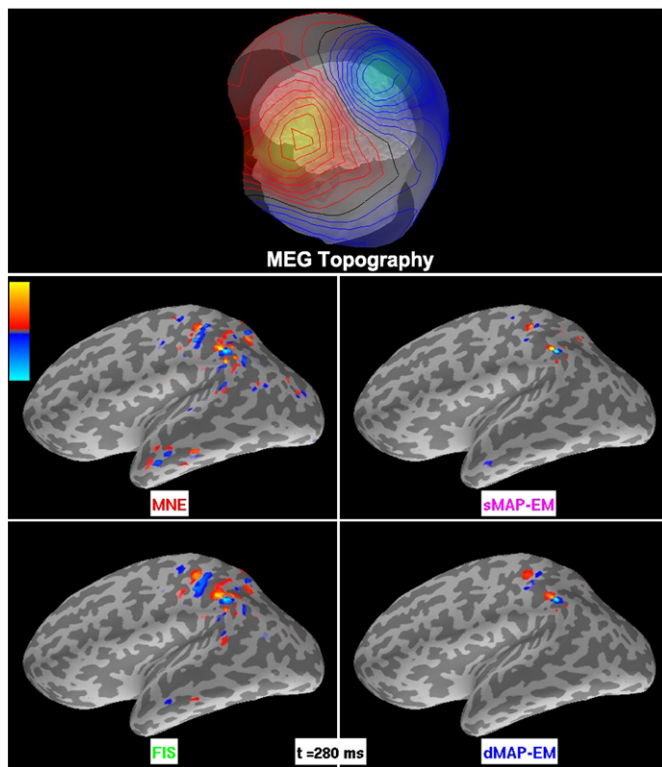


Fig. 8. Analysis of human MEG μ -rhythm data. The top panel shows the topography of the magnetic field component normal to the sensor surface, with isocontour lines indicating a field change of 100 fT. In the remaining panels, the intensity maps represent the amplitude of source estimates. The colorbar's maximum (bright yellow) and minimum (bright blue) corresponds to ± 1.8 nAm for MNE and FIS, and ± 9 nAm for the sMAP-EM, and dMAP-EM methods. The center and bottom left panels show the estimates obtained with the MNE and FIS methods, respectively. The estimated activity appears broad and spatially distributed, covering many different cortical regions. Similar to the simulated scenarios, the sMAP-EM algorithm (center right panel) yielded highly focal estimates. The bottom right panel shows the dMAP-EM estimates. This method resulted in more compact estimates covering primary somatosensory cortex and other parietal regions.

source estimates. This figure also includes the topography of the magnetic field component normal to the sensor surface at the same time instant (Hämäläinen and Ilmoniemi, 1994; Numminen et al., 1995). Similar to Figs. 3 and 5 in the simulated scenario, the MNE and FIS produced broad, spatially distributed estimates with activity covering primary somatosensory and motor regions as well as parietal, occipital, and temporal areas. The sMAP-EM method yielded highly focal estimates that appear spatially irregular. The dMAP-EM estimate presents a more compact spatial extent that covers primary somatosensory and parietal areas.

The time course estimates and their 95% Bayesian credibility intervals (CIs) are shown in Fig. 9. The 95% CIs (light colored areas) are obtained as described in the Results of simulation studies section. The upper panel shows a zoomed-in view of the cortical activity map obtained with the MNE method, where green dots represent the selected dipoles in primary motor (A), primary somatosensory (B), and parietal association areas (C and D). We note that these particular areas were selected as they have been reported to generate the μ -rhythm signals (Hari and Salmelin, 1997). Both MNE and FIS methods yielded estimates with smaller amplitudes, however, the CIs for the FIS method were smaller. In sMAP-EM estimates, the amplitude of the dipoles estimates A, B, and C was small with wider CIs in comparison to the dMAP-EM estimates. The dMAP-EM method yielded estimates with higher-amplitude oscillations, especially in dipoles B, C, and D, where the estimate follows the stereotypical μ -shape that characterizes these data, and presented smaller CIs than

those of the MNE and sMAP-EM methods. In the Supplementary information 3 we present a video of the results of this analysis.

Discussion

Intracranial electrophysiological recordings have shown that on a local scale, brain activity is spatially and temporally correlated (Bullock et al., 1995; Destexhe et al., 1999; Leopold et al., 2003; Nunez, 1995). Similarly, non-invasive fMRI and PET studies have shown that brain activation is temporally coherent in spatially distributed networks (Fox and Raichle, 2007; Fox et al., 2005; Gusnard and Raichle, 2001; Raichle et al., 2001). On the modeling side, biophysical spatiotemporal dynamic models of neuronal networks have been able to simulate electromagnetic scalp signals similar to those seen in recordings during normal and disease states (David et al., 2005; Gross et al., 2001; Izhikevich and Edelman, 2008; Jirsa et al., 2002; Kim and Robinson, 2007; Robinson et al., 2005; Sotero et al., 2007; Wright et al., 2004). We incorporated these insights to probabilistically model cortical activation as a distributed spatiotemporal dynamic process, and used this model as the basis for an inverse solution. This probabilistic model acts as a soft a priori constraint on the evolution of spatiotemporal cortical trajectories, allowing the recorded data to update our belief on this trajectory at each moment in time. In this model of cortical activity, the nearest-neighbor interactions are intended primarily to model spatial dependencies observed with intracranial electrophysiology. However, as shown by the large and small patch simulation results, the flexibility of this soft constraint results in accurate estimates of both extended and focal brain activity.

Spatiotemporal MEG/EEG source localization algorithms (Baillet and Garnero, 1997; Bolstad et al., 2009; Daunizeau and Friston, 2007; Daunizeau et al., 2006; Friston et al., 2008; Greensite, 2003; Limpi et al., 2009; Ou et al., 2009; Trujillo-Barreto et al., 2008; Zumer et al., 2008) have been previously developed with the aim of obtaining temporally smooth estimates by imposing computationally convenient prior constraints on dipole sources. While these methods provide a way of constraining the temporal evolution of inverse solutions, the relationship between prior constraints and underlying physiology is unclear. In contrast, our spatiotemporal dynamic model is structured to represent well-known local cortical biophysics, neuroanatomy, and electrophysiology, and can be developed further to account for more complex brain dynamics and spatial interactions. Other authors have proposed state-space models in the EEG inverse problem with volumetric Laplacian spatial interactions, using recursive least-squares (Yamashita et al., 2004) or an approximate version of the Kalman filter (Galka et al., 2004). In contrast, our approach establishes dynamic relationships along the cortical surface to represent local cortical interactions observed in physiological studies, and uses the full-dimensional Kalman filter and Fixed-Interval Smoother in conjunction with an EM algorithm to provide statistically optimal estimates of dipole source activity as well as dipole-specific model parameters. While these calculations are high dimensional and computationally demanding, we chose this approach in order to place the model in a spatial and temporal scale consistent with the biophysics of cortical interactions.

To understand the dynamic algorithms in relation to previously developed methods, we re-expressed the Kalman Filter and Fixed Interval Smoother estimates in a form analogous to that of the well-known static MNE (see Appendix D). This analysis showed that the dynamic methods have a structure that is very similar to MNE, with the Kalman Filter and Fixed Interval Smoother prediction covariance matrices playing the same role as the prior source covariance, or regularization matrix, in MNE. However, there is a critical difference in how the methods account for the prior mean source activity at a given moment in time. MNE assumes that this prior mean is zero at all times, while the Kalman filter and Fixed Interval Smoother

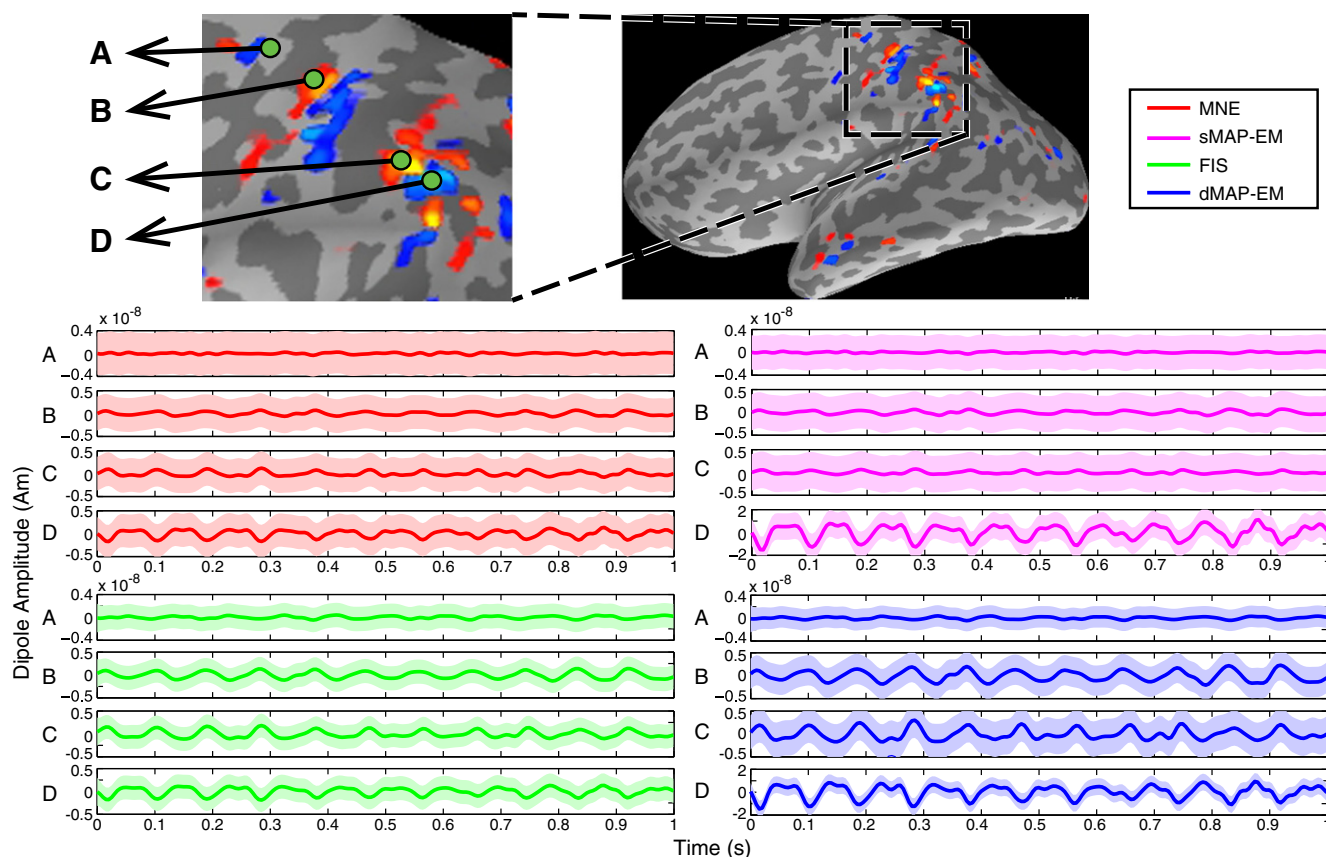


Fig. 9. Time course of estimation results for human MEG μ -rhythm data. The upper panel shows a zoomed-in view of the cortical activity map obtained with the MNE method, where the green dots represent dipoles labeled A, B, C, and D. The colored lines are estimated sources with 95% CIs represented by light colored areas. The center left panel shows the estimated time course of the MNE method in red, and the bottom left panels show the FIS estimates in green. These methods produced estimates with smaller amplitudes. The center right panel shows the sMAP-EM estimates in magenta, which yielded large CIs. The dMAP-EM method (blue) yielded estimates with higher-amplitude oscillations and smaller CIs, with the stereotyped μ -shape that characterizes these data (dipoles B, C, and D).

optimally update their prior means by assimilating data from the past, as well as the future in the case of FIS. In this way, as pointed out by Desai (2005) and Long et al. (2011), MNE and other similarly structured static algorithms impose a tendency towards zero source activity at every moment in time by ignoring estimates computed for other time points. Any reasonable model approximating spatio-temporal brain dynamics, employed within this dynamic estimation framework, would avoid this tendency towards zero, and would improve performance by enabling information from past and future estimates to help infer the cortical state at a particular point in time.

We designed simulation studies to compare source localization performance of the static MNE, sMAP-EM, FIS, and dMAP-EM estimates. The simulations were constructed to assess algorithm performance on both distributed and focal cortical activity. We simulated MEG data with a highly discretized cortical mesh and a deterministic temporal signal for source activity outside our autoregressive model class to avoid committing an “inverse crime”. In both simulations, i.e., with either a large or small active patch, the dMAP-EM method outperformed the MNE, sMAP-EM, and FIS methods in terms of spatial localization accuracy, temporal tracking of the simulated time series, the posterior error covariance, as well as RMSE and ROC analyses. These results suggest that the joint estimation of model parameters and source localization in a spatiotemporal dynamic model, as performed by the dMAP-EM algorithm, can significantly improve inverse solutions. Furthermore, the fact that dMAP-EM algorithm provides more accurate dipole source estimates than FIS, which does not estimate model parameters, indicates that parameter estimation within the dynamic model is critically important in this inverse problem. We also applied these methods to human MEG μ -rhythm data,

and obtained results similar to the simulated scenarios: The dMAP-EM method yielded distributed yet spatially compact estimates with pronounced time series amplitude in areas of cortex consistent with previous μ -rhythm studies, while the MNE and FIS produced spatially spread estimates with small amplitudes, and the sMAP-EM yielded highly focal and spatially irregular estimates.

Recent work by Wipf and Nagarajan (2009) has taken on the important task of characterizing the many seemingly dissimilar static methods described in the MEG inverse literature within a unified statistical framework. An important conclusion of that work is that most static MEG inverse methods can be viewed as solutions to a covariance selection problem, allowing fast covariance selection algorithms from statistics to be applied directly to the MEG inverse problem. Framing the inverse problem in terms of covariance selection, however, leaves open the question of how one should specify the form of that covariance, particularly in the presence of spatiotemporal phenomena. Dynamic modeling and estimation provides a framework to integrate mechanisms and empiricism from biophysics and neurophysiology into solutions for the MEG inverse problem. In this view, the solution to the inverse problem becomes one of specifying and identifying biophysical and dynamical models that closely approximate the underlying neurophysiological system. The spatial and temporal covariance structure then emerges from the second-order statistics inherent in the spatiotemporal dynamic model. If these models can be phrased in the appropriate statistical framework, fast and efficient algorithms with well-known properties can be applied to compute inverse solutions and parameter estimates.

Sophisticated dynamic models have been studied previously in the context of MEG/EEG data (David et al., 2006; Kiebel et al.,

2009). These Dynamic Causal Models (DCMs) use simplified spatial representations, such as equivalent current dipoles, to place greater emphasis on temporal modeling while retaining computational tractability. In contrast, the focus of our work has been to explore how spatiotemporal dynamic models in the distributed source framework, and inspired by underlying neurophysiology, can be used to improve source localization. This approach necessitates a higher-dimensional spatial model that can represent local cortical dynamics on a spatial scale consistent with neurophysiological recordings, balanced by a relatively simple temporal model to make computations tractable. In the long run, we envision an approach where more complex spatiotemporal models will be used to solve the MEG/EEG inverse problem. These models could include long-distance connectivity information derived from diffusion tensor imaging (DTI) and nonlinear interactions, while preserving a realistic spatial scale to represent brain activity.

Conclusions and future work

In this work we (1) developed a distributed stochastic dynamic model based on a nearest-neighbors autoregression on the cortical surface to represent spatiotemporal cortical dynamics, (2) derived the dMAP-EM algorithm for optimal dynamic estimation of cortical current sources and model parameters from MEG/EEG data based on the Kalman Filter, Fixed Interval Smoother, and Expectation–Maximization (EM) algorithms, (3) developed expressions to relate our dynamic estimation method to standard static algorithms, and (4) applied the spatiotemporal dynamic method to simulated experiments of focal and distributed cortical activation as well as to human experimental data. The results showed that our dMAP-EM method outperforms MNE, sMAP-EM, and FIS methods in terms of spatial localization accuracy, temporal tracking, posterior error covariance, and RMSE and ROC measures.

Our results demonstrate the feasibility of spatiotemporal dynamic estimation in distributed source spaces with several thousand dipoles and hundreds of sensors, resulting in inverse solutions with substantial performance improvements over static methods. Our analysis of known cortical biophysics, models (static vs. dynamic), source estimates, and error in localization revealed clear reasons why one would expect the dynamic methods to perform better than the static MNE. In future work, we will develop new techniques to improve computational performance by means of model reduction or high-performance computing, and will incorporate more realistic neurophysiological models within this dynamic modeling framework.

Supplementary materials related to this article can be found online at [doi:10.1016/j.neuroimage.2011.11.020](https://doi.org/10.1016/j.neuroimage.2011.11.020)

Appendix A. Robustness of source model against variations in the feedback matrix F

In this appendix we analyze how modifications in our spatial model (F) influence the prior source covariance, and show that the resulting smoothness encoded a priori in the dynamic source model is robust against misspecification of the F matrix. To do this, we first take a modified state model (Eq. (4)) where the feedback matrix $\tilde{F} = F + \Delta_F$ has been perturbed by Δ_F . Then, we consider the prior source covariance of the original model Σ and that of the modified model $\tilde{\Sigma}$, when they have reached equilibrium (steady-state), i.e., $\lim_{t \rightarrow \infty} \text{Cov}(x_t)$. At last, we derive an upper bound for the matrix 2-norm of the difference between the equilibrium covariances $\Delta_\Sigma = \tilde{\Sigma} - \Sigma$, as a function of the perturbation of the spatial model Δ_F .

We assume that both F and \tilde{F} yield stable dynamical systems (i.e., the modulus of their largest eigenvalue is strictly less than 1). From the stability condition, the equilibrium prior state covariance of the original model (Σ) and that of the modified model ($\tilde{\Sigma}$) correspond

to the respective (unique) solutions of the discrete Lyapunov equations (Brockwell and Davis, 2006):

$$\begin{aligned}\Sigma &= F\Sigma F' + Q \\ \tilde{\Sigma} &= \tilde{F}\tilde{\Sigma}\tilde{F}' + Q.\end{aligned}\quad (\text{A.1})$$

We now express the difference in the equilibrium covariances Δ_Σ by subtracting the equalities above to obtain:

$$\Delta_\Sigma = F\Delta_\Sigma F' + \tilde{F}\Delta_\Sigma \tilde{F}' + (F\tilde{\Sigma}\Delta_F' + \Delta_F\tilde{\Sigma}\Delta_F'). \quad (\text{A.2})$$

Now we take the matrix induced 2-norm in Eq. (A.1), and apply repeatedly the triangle inequality and the sub-multiplicative property of induced norms:

$$\|\Delta_\Sigma\|_2 \leq \|F\|_2^2 \|\Delta_\Sigma\|_2 + 2\|F\|_2 \|\tilde{\Sigma}\|_2 \|\Delta_F\|_2 + \|\tilde{\Sigma}\|_2 \|\Delta_F\|_2^2. \quad (\text{A.3})$$

We rearrange terms in Eq. (A.3) to obtain:

$$\|\Delta_\Sigma\|_2 \leq \frac{\|\tilde{\Sigma}\|_2 (2\|F\|_2 + \|\Delta_F\|_2)}{1 - \|F\|_2^2} \cdot \|\Delta_F\|_2. \quad (\text{A.4})$$

The fact that both feedback matrices are stable is equivalent to $\|F\|_2, \|\tilde{F}\|_2 < 1$. Using the inverse triangle inequality and setting $F = \tilde{F} - \Delta_F$, we obtain $|\|\tilde{F}\|_2 - \|\Delta_F\|_2| \leq \|F\|_2 < 1$. This inequality implies that $\|\Delta_F\|_2 < 2$. We plug the previous result and the fact that $\|F\|_2 < 1$ in the inequality above (Eq. (A.4)), and find that the matrix 2-norm of the difference between the equilibrium covariances Δ_Σ is bounded above by a constant multiplied by the 2-norm of the perturbation in the feedback matrix Δ_F :

$$\|\Delta_\Sigma\|_2 \leq c \cdot \|\Delta_F\|_2, \quad (\text{A.5})$$

where the constant $c = \frac{4\|\tilde{\Sigma}\|_2}{1 - \|F\|_2^2}$.

From Eq. (A.4) we can conclude that as long as the misspecification of the feedback matrix Δ_F is small, the smoothness modeled a priori in the dipole sources is not dramatically altered.

Appendix B. A non-informative prior for the state noise covariance $Q = \text{diag}(\theta)$

In this section we present a derivation of the non-informative prior for the parameters θ . Our task is to set the hyperparameters α and β such that the prior probability density of θ_n (Eq. (7)) is non-informative by making it approximately flat. This can be achieved by giving a large variance to the prior while fixing the mean to a value consistent with the order of magnitude in the model. We assume the values of the hyperparameter α is greater than 2 in order to obtain analytic expression for the mean ($E[\theta_n] = \beta/(\alpha + 1)$) and variance ($\text{Var}(\theta_n) = \beta^2/[(\alpha - 1)^2(\alpha - 2)]$). First, we set the prior mean to the order of magnitude in our model units, i.e., $E[\theta_n] = 10^{-18}$, and since we want the prior variance $\text{Var}(\theta_n)$ to be large, we set it in a way such that it is much larger than the mean, $\text{Var}(\theta_n) \gg E[\theta_n]$. The relationship between the variance and the mean is $\text{Var}(\theta_n) = E(\theta_n)^2/(\alpha - 2)$, which implies that:

$$\alpha = E(\theta_n)^2 / \text{Var}(\theta_n) + 2 \approx 2. \quad (\text{B.1})$$

Plugging α into the equation for the mean yields:

$$\beta = E(\theta_n) \left(E(\theta_n)^2 / \text{Var}(\theta_n) + 1 \right) \approx E(\theta_n) = 10^{-18}. \quad (\text{B.2})$$

Therefore, setting the hyperparameters to $\alpha = 2 + \delta$ and $\beta = 10^{-18}$, where $\delta \ll \beta$, provides a non-informative prior with a mean that is consistent with our model.

Appendix C. Derivation of $U(\theta|\theta^{(r-1)})$ in the E-step

From Eq. (13) we have that

$$U(\theta|\theta^{(r-1)}) \equiv \mathbb{E} \left[\log p \left(\{y_t\}_{t=1}^T, \{x_t\}_{t=0}^T \middle| \theta \right) \middle| \{y_t\}_{t=1}^T, \theta^{(r-1)} \right] + \log p(\theta), \quad (\text{C.1})$$

where the complete data log-likelihood is derived from the measurement (Eq. (1)) and source (Eq. (4)) models:

$$\log p \left(\{y_t\}_{t=1}^T, \{x_t\}_{t=0}^T \middle| \theta \right) = \sum_{t=1}^T \log p(y_t|x_t, \theta) + \sum_{t=1}^T \log p(x_t|x_{t-1}, \theta) + \log p(x_0|\theta), \quad (\text{C.2})$$

with,

$$\begin{aligned} \log p(x_0|\theta) &= -\frac{1}{2} \{c_{N_x} + \log|\Sigma_0| + x_0' \Sigma_0^{-1} x_0\} \\ \log p(x_t|x_{t-1}, \theta) &= -\frac{1}{2} \{c_{N_x} + \log|Q| + (x_t - Fx_{t-1})' Q^{-1} (x_t - Fx_{t-1})\} \\ \log p(y_t|x_t, \theta) &= -\frac{1}{2} \{c_{N_y} + \log|C| + (y_t - Gx_t)' C^{-1} (y_t - Gx_t)\}, \end{aligned} \quad (\text{C.3})$$

where c_{N_x} and c_{N_y} are constants not depending on θ .

We apply Eq. (C.3) to the complete data log-likelihood (Eq. (C.2)) and compute its expectation with respect to $p(\{x_t\}_{t=0}^T | \{y_t\}_{t=1}^T, \theta^{(r-1)})$, where we have conditioned on the full set of measurements and the parameter estimate of the previous iteration (Shumway and Stoffer, 1982). We then add the logarithm of the prior density of θ and obtain Eq. (17).

Appendix D. Relationships between the dynamic and static estimators

In this appendix we present an algebraic analysis of the Kalman Filter (KF) and Fixed Interval Smoother (FIS) estimates that illustrates their relationship to the L_2 minimum-norm estimate (MNE) (Hämäläinen and Ilmoniemi, 1994). We emphasize that in this analysis the parameters θ in the model are assumed to be fixed. Specifically, we set aside the so-called problem of identification or learning of parameters and focus on comparing and contrasting the functional form of the source amplitude estimates given by these methods.

We note that while it has been well-established that, on one side, the MNE can be seen as a static Maximum a Posteriori estimate of the current dipole sources where the measurements are assumed to be independent in time (Hämäläinen and Ilmoniemi, 1994; Wipf and Nagarajan, 2009), and on the other, the KF and FIS algorithms are implementations of a Maximum a Posteriori estimation problem in a linear Gaussian state-space model (Kitagawa and Gersch, 1996; Roweis and Ghahramani, 1999), our contribution in this section is: (1) To show how the KF, and especially the FIS, are solutions to particular penalized least square problems structurally similar to the L_2 minimum-norm cost function, where the penalty term reflects how the information of past $\{y_k\}_{k=1}^{t-1}$ and future $\{y_k\}_{k=t+1}^T$ measurements is optimally accounted by the estimate; and (2) to present the formulas for the KF and FIS estimates in a way that parallel to those of the well-known MNE equation as an attempt to further introduce these dynamic estimation techniques in the broader neuroimaging community.

To facilitate the notation, in this section we will assume that all densities are conditioned by a set of parameters and use the notation $p(\cdot) \equiv p(\cdot|\theta)$. We begin by recalling that the MNE assumes the probability density of the source amplitude vector $p(x_t)$ to be Gaussian with mean $\mu_t = \mu^{(MNE)} = 0$ and covariance, or regularization matrix, $P_t = \Sigma^{(MNE)}$. Therefore, the MNE maximizes the posterior

density (Hämäläinen and Ilmoniemi, 1994),

$$p(x_t|y_t) \propto \underbrace{p(y_t|x_t)}_{\text{Likelihood}} \underbrace{p(x_t)}_{\text{"Prior"}}, \quad (\text{D.1})$$

where the likelihood $p(y_t|x_t)$ is Gaussian with mean Gx_t and covariance C (Eq. (1)). We must emphasize that the MNE “prior” density $p(x_t)$ does not contain any information about the measurements.

A similar interpretation can be given to the Kalman Filter estimate of x_t given data up to time t , as shown in Kitagawa and Gersch (1996); Roweis and Ghahramani (1999). In this case, the “prior” corresponds to the conditional density of x_t given data up to time $t-1$, $p(x_t|\{y_k\}_{k=1}^{t-1})$, which is Gaussian with mean $\mu_t = x_{t|t-1}$ and covariance $P_t = P_{t|t-1}$ given by the prediction step of Kalman Filter recursions (Eq. (21)). Then the Kalman Filter computes the Maximum a Posteriori estimate of x_t given data up to time t by maximizing the posterior density:

$$p(x_t|\{y_k\}_{k=1}^t) \propto \underbrace{p(y_t|x_t)}_{\text{Likelihood}} \underbrace{p(x_t|\{y_k\}_{k=1}^{t-1})}_{\text{"Prior"}}. \quad (\text{D.2})$$

We highlight that the Kalman Filter “prior” density $p(x_t|\{y_k\}_{k=1}^{t-1})$ contains information only from past measurements $\{y_k\}_{k=1}^{t-1}$.

The Fixed Interval Smoother estimate can be interpreted similarly, where the “prior” density corresponds to the conditional density of x_t given previous data up to time $t-1$ and future data after time $t+1$, $p(x_t|\{y_k\}_{k=1}^{t-1}, \{y_k\}_{k=t+1}^T)$. To the authors' knowledge, this particular interpretation of the FIS has not been reported in the literature. Again, the “prior” density is Gaussian, and its mean $\mu_t = x_{t|T}$ and covariance $P_t = P_{t|T}$ can be obtained by standard methods with computationally costly algebraic manipulation. In this case, the notation of the subscript $t|T$ reflects that we are conditioning on all data except the immediate data point y_t . Therefore, the Fixed Interval Smoother estimate maximizes the posterior density of the state given all measurements:

$$p(x_t|\{y_k\}_{k=1}^T) \propto \underbrace{p(y_t|x_t)}_{\text{Likelihood}} \underbrace{p(x_t|\{y_k\}_{k=1}^{t-1}, \{y_k\}_{k=t+1}^T)}_{\text{"Prior"}}. \quad (\text{D.3})$$

We emphasize that the Fixed Interval Smoother “prior” density $p(x_t|\{y_k\}_{k=1}^{t-1}, \{y_k\}_{k=t+1}^T)$ includes information from both past $\{y_k\}_{k=1}^{t-1}$ and future measurements $\{y_k\}_{k=t+1}^T$.

We can see now that once we have available the “prior” densities' mean and covariance for each method, obtaining the Maximum a Posteriori estimates for the MNE, KF, and FIS by maximizing Eqs. (D.1), (D.2), and (D.3), respectively, corresponds to apply a penalized least-squares technique where the data fit term corresponds to the likelihood term and the penalty term is related to the “prior” density. Specifically, for each method, the penalized least squares function to minimize is

$$\arg \min_{x_t} \underbrace{\|y_t - Gx_t\|_{C^{-1}}^2}_{\text{Data fit}} + \underbrace{\|x_t - \mu_t\|_{P_t^{-1}}^2}_{\text{Penalty}}, \quad (\text{D.4})$$

where μ_t and P_t are the “prior” mean and covariance as defined above for each method. We should note at this point that Eq. (D.4) establishes a parallel between L_2 minimum-norm cost function, and the Kalman Filter and Fixed Interval Smoother estimates, where the data fit term is identical in these methods, but the penalty varies to indicate how and whether past and future measurements should influence the estimate.

The minimizer of Eq. (D.4) is given by

$$\hat{x}_t(\mu_t, P_t) = \mu_t + P_t G' (G P_t G' + C)^{-1} (y_t - G \mu_t). \quad (D.5)$$

Now we can simply replace the corresponding “prior” mean and covariance for each method to obtain the MNE ($x_t^{(MNE)}$), the Kalman Filter estimate ($x_{t|t}$) and the Fixed Interval Smoother estimate ($x_{t|T}$):

$$\begin{aligned} x_t^{(MNE)} &= 0 + \Sigma^{(MNE)} G' (G \Sigma^{(MNE)} G' + C)^{-1} (y_t - 0) \\ x_{t|t} &= x_{t|t-1} + P_{t|t-1} G' (G P_{t|t-1} G' + C)^{-1} (y_t - G x_{t|t-1}) \\ x_{t|T} &= \underbrace{x_{t|T \setminus t}}_{\text{Prior mean}} + \underbrace{P_{t|T \setminus t} G' (G P_{t|T \setminus t} G' + C)^{-1}}_{\text{Gain}} \underbrace{(y_t - G x_{t|T \setminus t})}_{\text{Residual}}. \end{aligned} \quad (D.6)$$

Eq. (D.6) allows us to compare and contrast the algebraic forms of these estimates. We should note that while line 1 is the well-known MNE formula and line 2 is in fact identical to the Kalman Filter algorithm (Eq. (21)), line 3 corresponds to a novel derivation of the Fixed Interval Smoother estimate that allows us to highlight similarities and differences between these estimates. We can see that each method builds a prediction of x_t using the respective “prior” mean, and then updates this prediction using the measurement at the same time point y_t by computing residuals. In the case of MNE, the prediction discards any observed data and assumes it is zero. The Kalman Filter is an improvement over MNE since it builds a prediction based on previous data $\{y_k\}_{k=1}^{t-1}$. Finally, the Fixed Interval Smoother goes further to build a prediction based on previous data $\{y_k\}_{k=1}^{t-1}$ as well as future data $\{y_k\}_{k=t+1}^T$. For all methods, MNE, KF, and FIS, once the prediction is made, the estimate is obtained by updating the prediction using the immediate data point y_t by adding a term that is proportional to the residuals.

Appendix E. Computation of ROC curves

In this appendix we present the computations that define the estimates of the detection (\hat{p}_D) and false alarm (\hat{p}_{FA}) probabilities. These quantities, which depend on the detection threshold c , are given by

$$\begin{aligned} \hat{p}_D(c) &= \frac{\sum_{n=1}^{N_x} \sum_{t=1}^T \text{indic}(|\hat{x}_{n,t}| > c) \cdot \text{indic}(x_{n,t}^{(SIM)} \neq 0)}{\sum_{n=1}^{N_x} \sum_{t=1}^T \text{indic}(x_{n,t}^{(SIM)} \neq 0)} \\ \hat{p}_{FA}(c) &= \frac{\sum_{n=1}^{N_x} \sum_{t=1}^T \text{indic}(|\hat{x}_{n,t}| > c) \cdot \text{indic}(x_{n,t}^{(SIM)} = 0)}{\sum_{n=1}^{N_x} \sum_{t=1}^T \text{indic}(x_{n,t}^{(SIM)} = 0)}, \end{aligned} \quad (E.1)$$

where $\text{indic}(\cdot)$ in Eq. (E.1) is the indicator function on a set S :

$$\text{indic}(x \in S) \equiv \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}. \quad (E.2)$$

We recall that $N_x \approx 5000$ represents the number of dipole sources and $T = 200$ is the number of time samples.

References

- Anderson, B.D.O., Moore, J.B., 2005. Optimal Filtering. Dover Publications, Mineola, NY.
- Auranen, T., Nummenmaa, A., Hämäläinen, M.S., Jääskeläinen, I.P., Lampinen, J., Vehtari, A., Sams, M., 2005. Bayesian analysis of the neuromagnetic inverse problem with $l(p)$ -norm priors. Neuroimage 26, 870–884.
- Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomic-functional priors in the EEG/MEG inverse problem. IEEE Trans. Biomed. Eng. 44, 374–385.
- Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 18, 14–30.
- Bertrand, C., Ohmi, M., Suzuki, R., Kado, H., 2001. A probabilistic solution to the MEG inverse problem via MCMC methods: the reversible jump and parallel tempering algorithms. IEEE Trans. Biomed. Eng. 48, 533–542.

- Bolstad, A., Van Veen, B.D., Nowak, R., 2009. Space-time event sparse penalization for magneto-/electroencephalography. Neuroimage 46, 1066–1081.
- Brockwell, P.J., Davis, R.A., 2006. Time series: theory and methods, 2nd edition. Springer, New York.
- Bullock, T.H., McClune, M.C., Achimowicz, J.Z., Iragui-Madoz, V.J., Duckrow, R.B., Spencer, S.S., 1995. Temporal fluctuations in coherence of brain waves. Proc. Natl. Acad. Sci. U. S. A. 92, 11568–11572.
- Buzsaki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929.
- Campi, C., Pascarella, A., Sorrentino, A., Piana, M., 2008. A Rao-Blackwellized particle filter for magnetoencephalography. Inverse Probl. 24, 025023.
- Dale, A.M., Fischl, B.R., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.
- Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., Leahy, R.M., 2004. Mapping human brain function with MEG and EEG: methods and validation. Neuroimage 23 (Suppl. 1), S289–S299.
- Daunizeau, J., Friston, K.J., 2007. A mesostate-space model for EEG and MEG. Neuroimage 38, 67–81.
- Daunizeau, J., Mattout, J., Clonda, D., Goulard, B., Benali, H., Lina, J.M., 2006. Bayesian spatio-temporal approach for EEG source reconstruction: conciliating ECD and distributed models. IEEE Trans. Biomed. Eng. 53, 503–516.
- David, O., Harrison, L.M., Friston, K.J., 2005. Modelling event-related responses in the brain. Neuroimage 25, 756–770.
- David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage 30, 1255–1272.
- de Jong, P., Mackinnon, M.J., 1988. Covariances for smoothed estimates in state space models. Biometrika 75, 601–602.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B Stat. Methodol. 39, 1–38.
- Desai, N.U., 2005. Source localization for magnetoencephalography by spatio-temporal Kalman filtering and Fixed-Interval smoothing. Master's thesis. Massachusetts Institute of Technology.
- Destexhe, A., Contreras, D., 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.
- Fischl, B.R., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.
- Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711.
- Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, Maurizio, Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. U. S. A. 102, 9673–9678.
- Friston, K.J., Harrison, L.M., Daunizeau, J., Kiebel, S.J., Phillips, C., Trujillo-Barreto, N.J., Henson, R.N., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39, 1104–1120.
- Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdes-Sosa, P.A., 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. Neuroimage 23, 435–453.
- Gorodnitsky, I.F., George, J.S., Rao, B.D., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalogr. Clin. Neurophysiol. 95, 231–251.
- Green, P.J., 1990. On use of the EM algorithm for penalized likelihood estimation. J. R. Stat. Soc. B Stat. Methodol. 52, 443–452.
- Greensite, F., 2003. The temporal prior in bioelectromagnetic source imaging problems. IEEE Trans. Biomed. Eng. 50, 1152–1159.
- Gross, J., Kujala, J., Hämäläinen, M.S., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A. 98, 694–699.
- Gusnard, D.A., Raichle, M.E., 2001. Searching for a baseline: functional imaging and the resting human brain. Nat. Rev. Neurosci. 2, 685–694.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.
- Hämäläinen, M.S., Sarvas, J., 1989. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE Trans. Biomed. Eng. 36, 165–171.
- Hämäläinen, M.S., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497.
- Hari, R., Salmelin, R., 1997. Human cortical oscillations: a neuromagnetic view through the skull. Trends Neurosci. 20, 44–49.
- Izhikevich, E.M., Edelman, G.M., 2008. Large-scale model of mammalian thalamocortical systems. Proc. Natl. Acad. Sci. U. S. A. 105, 3593–3598.
- Jirsa, V.K., Jantzen, K.J., Fuchs, A., Scott Kelso, J.A., 2002. Spatiotemporal forward solution of the EEG and MEG using network modeling. IEEE Trans. Med. Imaging 21, 493–504.
- Jun, S.C., George, J.S., Paré-Blagoev, J., Plis, S.M., Ranken, D.M., Schmidt, D.M., Wood, C.C., 2005. Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data. Neuroimage 28, 84–98.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 82, 35–45.
- Kiebel, S.J., Daunizeau, J., Phillips, C., Friston, K.J., 2008. Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. Neuroimage 39, 728–741.
- Kiebel, S.J., Garrido, M.I., Moran, R., Chen, C.C., Friston, K.J., 2009. Dynamic causal modeling for EEG and MEG. Hum. Brain Mapp. 30, 1866–1876.
- Kim, J.W., Robinson, P.A., 2007. Compact dynamical model of brain activity. Phys. Rev. E 75, 031907.

- Kitagawa, G., Gersch, W., 1996. Smoothness priors analysis of time series. Springer, New York.
- Lamus, C., Long, C.J., Hämäläinen, M.S., Brown, E.N., Purdon, P.L., 2007. Parameter estimation and dynamic source localization for the magnetoencephalography (MEG) inverse problem. 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington, D.C., USA, pp. 1092–1095.
- Leopold, D.A., Murayama, Y., Logothetis, N.K., 2003. Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. *Cereb. Cortex* 13, 422–433.
- Limpiti, T., Van Veen, B.D., Attias, H.T., Nagarajan, S.S., 2009. A spatiotemporal framework for estimating trial-to-trial amplitude variation in event-related MEG/EEG. *IEEE Trans. Biomed. Eng.* 56, 633–645.
- Long, C.J., Purdon, P.L., Temereanca, S., Desai, N.U., Hämäläinen, M.S., Brown, E.N., 2011. State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing. *Ann. Appl. Stat.* 5, 1207–1228.
- Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. *Neuroimage* 30, 753–767.
- Mendel, J.M., 1971. Computational requirements for a discrete Kalman Filter. *IEEE Trans. Autom. Control* AC16, 748–758.
- Moon, T.K., Stirling, W.C., 2000. Mathematical methods and algorithms for signal processing. Prentice Hall, Upper Saddle River, NJ.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39, 541–557.
- Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. *IEEE Trans. Biomed. Eng.* 46, 245–259.
- Nummenmaa, A., Auranen, T., Hämäläinen, M.S., Jääskeläinen, I.P., Lampinen, J., Sams, M., Vehtari, A., 2007. Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *Neuroimage* 35, 669–685.
- Numminen, J., Ahlfors, S.P., Ilmoniemi, R.J., Montonen, J., Nenonen, J., 1995. Transformation of multichannel magnetocardiographic signals to standard grid form. *IEEE Trans. Biomed. Eng.* 42, 72–78.
- Nunez, P.L., 1995. Neocortical dynamics and human EEG rhythms. Oxford University Press, USA.
- Ou, W., Hämäläinen, M.S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG inverse solver. *Neuroimage* 44, 932–946.
- Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65.
- Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J., 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. *Neuroimage* 24, 997–1011.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682.
- Rauch, H.E., Tung, F., Striebel, C.T., 1965. Maximum likelihood estimates of linear dynamic systems. *AIJAA J.* 3, 1445–1450.
- Robinson, P.A., Rennie, C.J., Rowe, D.L., O'Connor, S.C., Gordon, E., 2005. Multiscale brain modelling. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1043–1050.
- Roweis, S., Ghahramani, Z., 1999. A unifying review of linear Gaussian models. *Neural Comput.* 11, 305–345.
- Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage* 23, 806–826.
- Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* 3, 253–264.
- Somersalo, E., Voutilainen, A., Kaipio, J.P., 2003. Non-stationary magnetoencephalography by Bayesian filtering of dipole models. *Inverse Probl.* 19, 1047–1063.
- Sotero, R.C., Trujillo-Barreto, N.J., Iturria-Medina, Y., Carbonell, F., Jimenez, J.C., 2007. Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput.* 19, 478–512.
- Trujillo-Barreto, N.J., Aubert-Vázquez, E., Penny, W.D., 2008. Bayesian M/EEG source reconstruction with spatio-temporal priors. *Neuroimage* 39, 318–335.
- Uutela, K., Hämäläinen, M.S., Salmelin, R., 1998. Global optimization in the localization of neuromagnetic sources. *IEEE Trans. Biomed. Eng.* 45, 716–723.
- Uutela, K., Hämäläinen, M.S., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. *Neuroimage* 10, 173–180.
- Van Veen, B.D., Van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880.
- Wipf, D.P., Nagarajan, S.S., 2009. A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage* 44, 947–966.
- Wright, J.J., Rennie, C.J., Lees, G.J., Robinson, P.A., Bourke, P.D., Chapman, C.L., Gordon, E., Rowe, D.L., 2004. Simulated electrocortical activity at microscopic, mesoscopic and global scales. *Int. J. Bifurcation Chaos* 14, 853–872.
- Yamashita, O., Galka, A., Ozaki, T., Biscay, R., Valdes-Sosa, P.A., 2004. Recursive penalized least squares solution for dynamical inverse problems of EEG generation. *Hum. Brain Mapp.* 21, 221–235.
- Zumer, J.M., Attias, H.T., Sekihara, K., Nagarajan, S.S., 2008. Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data. *Neuroimage* 41, 924–940.