Models and Algorithms of Brain Connectivity, Spatial Sparsity, and Temporal Dynamics for the MEG/EEG Inverse Problem

by

Camilo Lamus Garcia Herreros

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computational Neuroscience
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2015

Author .....
Department of Brain and Cognitive Sciences
August 20, 2015

Certified by .....
Patrick L. Purdon
Associate Professor of Anesthesia at Harvard Medical School
Associate Bioengineer at the Massachusetts General Hospital
Thesis Supervisor

Certified by .....
Emery N. Brown
Professor of Computational Neuroscience
Edward Hood Taplin Professor of Medical Engineering
Warren M. Zapol Professor of Anesthesia at Harvard Medical School
Thesis Supervisor

Accepted by .....
Matthew A. Wilson
Sherman Fairchild Professor of Neuroscience and Picower Scholar
Director of Graduate Education for Brain and Cognitive Sciences
Chairman, Department Committee on Graduate Theses

Models and Algorithms of Brain Connectivity, Spatial Sparsity, and Temporal Dynamics for the MEG/EEG Inverse Problem

by

Camilo Lamus Garcia Herreros

Submitted to the Department of Brain and Cognitive Sciences
on August 20, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computational Neuroscience

Abstract

Magnetoencephalography (MEG) and electroencephalography (EEG) are noninvasive functional neuroimaging techniques that provide high temporal resolution recordings of brain activity, offering a unique means to study fast neural dynamics in humans. Localizing the sources of brain activity from MEG/EEG is an ill-posed inverse problem, with no unique solution in the absence of additional information. In this dissertation I analyze how solutions to the MEG/EEG inverse problem can be improved by including information about temporal dynamics of brain activity and connectivity within and among brain regions. The contributions of my thesis are: 1) I develop a dynamic algorithm for source localization that uses local connectivity information and Empirical Bayes estimates to improve source localization performance (Chapter 1). This result led me to investigate the underlying theoretical principles that might explain the performance improvement observed in simulations and by analyzing experimental data. In my analysis, 2) I demonstrate theoretically how the inclusion of local connectivity information and basic source dynamics can greatly increase the number of sources that can be recovered from MEG/EEG data (Chapter 2). Finally, in order to include long distance connectivity information, 3) I develop a fast multi-scale dynamic source estimation algorithm based on the Subspace Pursuit and Kalman Filter algorithms that incorporates brain connectivity information derived from diffusion MRI (Chapter 3). Overall, I illustrate how dynamic models informed by neurophysiology and neuroanatomy can be used alongside advanced statistical and signal processing methods to greatly improve MEG/EEG source localization. More broadly, this work provides an example of how advanced modeling and algorithm development can be used to address difficult problems in neuroscience and neuroimaging.

Thesis Supervisor: Patrick L. Purdon
Title: Associate Professor of Anesthesia at Harvard Medical School

Associate Bioengineer at the Massachusetts General Hospital

Thesis Supervisor: Emery N. Brown
Title: Professor of Computational Neuroscience
Edward Hood Taplin Professor of Medical Engineering
Warren M. Zapol Professor of Anesthesia at Harvard Medical School

Acknowledgments

One thing that is better than a great Ph.D. advisor is a great Ph.D. thesis committee, a wonderful wife, and a fantastic dog.

I feel extremely fortunate to have had Patrick Purdon and Emery Brown as my advisors. They have provided me with the optimal balance between the freedom needed to explore ideas and the guidance necessary to achieve goals. Patrick's vision and contagious enthusiasm have been the guiding forces behind my research. I have been privileged to learn from his expertise in signal processing, brain imaging, and neurosciences. Our brainstorming and writing sessions have been some of the most enjoyable and enriching portions of my Ph.D. His work ethic and his unwavering determination to resolve every challenge have become my example to follow. Emery's expertise in statistics and neurosciences shaped the way in which I investigate and examine. I have been privileged to learn from him the statistical way to properly analyze data to solve research questions. His precise guidance has resonated with me throughout the years, and his keen enthusiasm and intellectual breadth continue to be a source of inspiration.

I am tremendously grateful to my thesis committee members Matti Hämäläinen and John Gabrieli. Matti's unparalleled expertise in the physics and neurosciences of MEG and EEG has greatly helped my research. Throughout my Ph.D., he has remained a source of continuous feedback, greatly improving the quality of my research. I have learned from his ability to succinctly and clearly present complex ideas. His enthusiasm, sense of humor, and kindness, have brightened some of my darkest winter days. I am very grateful to John for broadening the neuroscience perspective of my research. I am much obliged to him for being so accessible, in particular for accommodatingly serving in my thesis committee.

I was privileged to simultaneously work in two labs, where I learned from fantastic lab mates. Specifically, I would like to thank Antonio Molins, Demba Ba, Gabriel Obregon, Ian Cajigas, Behtash Babadi, Francisco Flores, Pavitra Krishnaswamy, Zhe Chen, Wasim Malik, Anne Smith, and Mike Prerau. We had lively discussions about

statistics, math, science, and just about every other debatable subject.

I would like to thank my Mom, Dad, Daniel and Nana. Independently of time or place, their love and support have remained constant. I am most thankful to my wife, Alicia: the best thing about her is that she is the best. I am simply without words. Lastly, I would like to thank my dog, Simon. He had the uncanny ability to know precisely when I needed to step away from my work. When it was time for a break, he would take me for a walk in the park, and would remind me to look up at the sky, and to stop and smell the roses.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

A Spatiotemporal Dynamic Solution to the MEG Inverse Problem Using the Empirical Bayes Paradigm

1.1 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-posed 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.

1.2 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 [1]. 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 [1, 2].

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 [3, 4, 5, 6, 7, 8, 9]. Distributed source models, on the other hand, assume that the recorded activity results from the activation of a spatial distribution of current dipoles with known locations(see [10] 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 [11, 12, 13, 14], impose spatial smoothness [15], or favor focal estimates [16]. Furthermore, Bayesian methods have been used to obtain spatially sparse estimates, where components of the current source covariance are estimated directly [17, 18] or additional hierarchical priors are assigned in order to compute posterior distributions [19, 20, 21]. 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 [22], 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 [23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. Linear state-space models have also been used to model source cur-

rent dynamics. However, in order to reduce computational complexity, these methods either apply spatially-independent approximations to their respective estimation algorithms [33, 34] or *a priori* fix model specific parameters to avoid the problem of parameter estimation [35]. 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 [36, 37, 38]. These local spatial interactions are supported neuroanatomically by axonal collateral projections from pyramidal cells that spread laterally at distances of up to 10 mm [39]. 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 [40, 41, 42, 43, 44, 45, 46, 47]. 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 [48, 49, 50, 51].

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; and 4) We apply our spatiotemporal dynamic method to simulated ex-

periments of focal and distributed cortical activation as well as experimental data from a human subject.

1.3 Methods

1.3.1 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 sensors at times $\{\Delta t\}_{t=1}^{T}$, where Δ is the sampling interval and T is the number of measurements in time. Let $y_{i,t}$ denote the measurement at time t in sensor i, and define $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{n,t}]'$ as the $n \times 1$ vector of measurements at all sensors at time t. We assume that the measurements were generated by p current dipole sources distributed on the cortical surface and oriented perpendicular to it. Let $\beta_{i,t}$ denote the source amplitude of the ith dipole at time t, and define $\boldsymbol{\beta}_t = [\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{p,t}]'$ as the $p \times 1$ vector of cortical source activity in all considered locations, or cortical state vector, at time t. Typically, $n \sim$ a few hundred, and $p \sim$ several thousand. The relationship between the measurement vector $\mathbf{y}_t$ and the cortical state vector $\boldsymbol{\beta}_t$ is given by the measurement equation,

$$\mathbf{y}_t = \mathbf{X}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \tag{1.1}$$

where $\mathbf{X}$ is the $n \times p$ lead field gain matrix computed using a quasistatic approximation of the Maxwell's equations [1], i.e., the solution of the forward problem, and $\boldsymbol{\varepsilon}_t$ is a $n \times 1$ Gaussian white noise vector with zero mean covariance matrix equal to the identity matrix $\mathbf{I}$ independent from $\boldsymbol{\beta}_t$ for all time points. In Equation (1.1) we assumed that the model has been spatially whitened, i.e., that the original raw data model $\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}\boldsymbol{\beta}_t + \tilde{\boldsymbol{\varepsilon}}_t$ has been premutiplied by the inverse of a matrix square root of the covariance of $\tilde{\boldsymbol{\varepsilon}}_t$. 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 [1]. Nevertheless, our development can be easily extended to account for unconstrained source orientations.

1.3.2 Spatiotemporal dynamical source model

As we pointed in the Introduction Section 1.2, evidence from neurophysiology studies, neuroanatomy, biophysics, and neuroimaging suggests that cortical activation is a distributed spatiotemporal dynamic process [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. 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 $\beta_{i,t}$ are a function of past neuronal currents within a small local neighborhood along the cortical surface and a small perturbation $\omega_{i,t}$ that accounts for unknown factors affecting the evolution of cortical currents,

$$\beta_{i,t} = \phi[\underbrace{f_{i,i}\beta_{i,t-1}}_{\text{Past activity}} + \underbrace{\sum_{j\in\mathcal{N}(i)} f_{i,j}\beta_{j,t-1}}_{\text{Past activity of neighbors}}] + \underbrace{(1-\phi^2)^{1/2}\omega_{i,t}}_{\text{Unaccounted factors}}, \tag{1.2}$$

where $\mathcal{N}(i)$ is the index set of nearest neighbors of the ith dipole source, and ϕ $(0 \leq \phi < 1)$ is a scalar that both represent the strengh of the history dependence in the dynamics and also guarantee the stability of the cortical state dynamics. The state noise $\boldsymbol{\omega}_t = [\omega_{1,t}, \omega_{2,t}, \ldots, \omega_{p,t}]$ is a $p \times 1$ Gaussian vector with zero mean and covariance matrix $\mathbf{Q}$, independent from the measurement noise $\boldsymbol{\varepsilon}_t$. The weights $f_{i,j}$, which represent the individual influence of neighboring sources in the autoregression, are assumed to decay with the distance from the ith to the jth dipole source. A simple relation that reflects this modeling assumption is to make the weights inversely proportional to the distance between dipoles,

$$f_{i,j} \propto \frac{1}{\text{distance from } i\text{th to } j\text{th dipole}}, \tag{1.3}$$

where the proportionality constant is chosen such that the contribution of the neighbors to the dynamics of the ith source equals its self contribution, while the total contribution is equal to one: $\sum_{j\in\mathcal{N}(i)} f_{i,j} = f_{i,i}$ and $\sum_{j\in\mathcal{N}(i)} f_{i,j} + f_{i,i} = 1$. This allows the power (prior variance) in every dipole $\beta_{i,t}$ to remain approximately constant over time.

Figure 1-1 illustrates this nearest-neighbors autoregressive model. The left panel shows the cortical surface reconstructed from high-resolution magnetic resonance images (MRI) using *Freesurfer* [52, 53], 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 $p = 5124$ dipoles 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 1.7.1, we analyze how modifications in our spatial model, such us those arising from choosing a different weighting scheme or a misspecification of the weights $f_{i,j}$ (Eq. 1.2), 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 Equation (1.2) as the multivariate autoregressive model,

$$\boldsymbol{\beta}_t = \phi\mathbf{F}\boldsymbol{\beta}_{t-1} + (1-\phi^2)^{1/2}\boldsymbol{\omega}_t, \tag{1.4}$$

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

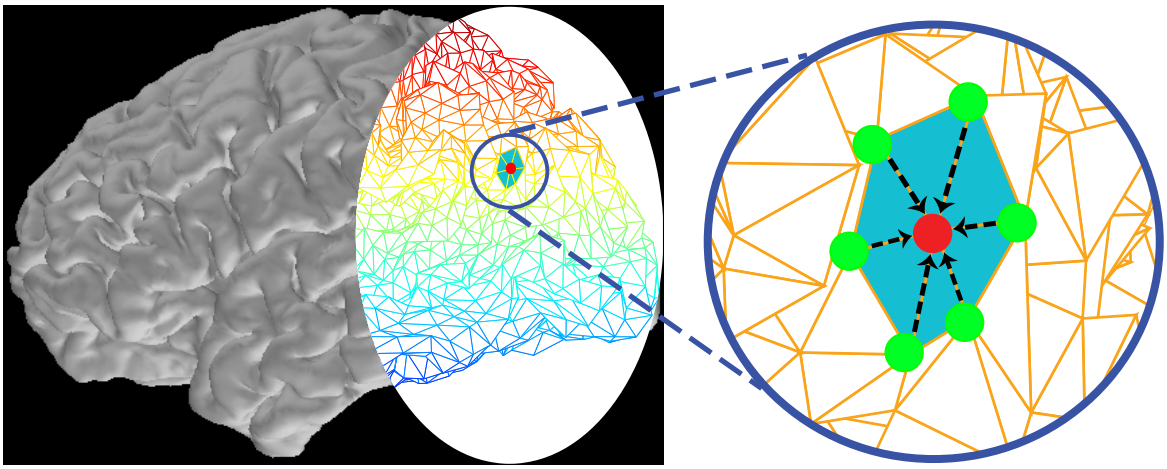

Figure 1-1: **Illustration of the spatiotemporal dynamic source model**. The left panel shows the reconstructed cortical surface where the caudal portion of cortex is depicted as a triangulated mesh of dipole sources. The zoomed-in panel in the right shows a dipole source (central red dot) and its nearest-neighbor 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.

$$(\mathbf{F})_{i,j} = \begin{cases} f_{i,j} & \text{if } j = i \text{ or } j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases}, \tag{1.5}$$

and $(\mathbf{F})_{i,j}$ is the ith row of the jth column of $\mathbf{F}$. The initial dipole source vector $\boldsymbol{\beta}_0$ is assumed to be Gaussian with zero mean and covariance matrix $\mathbf{C}_0$, and independent of $\{\boldsymbol{\omega}_t\}_{t=1}^T$.

1.3.3 Prior model for the parameters

The model defined in Sections 1.3.1 and 1.3.2 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 $\{\mathbf{X}, \mathbf{F}, \mathbf{Q}, \mathbf{C}_0\}$. 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 $\mathbf{X}$ can be computed using the boundary-element model based on high-resolution magnetic resonance images (MRI) [54]. The covariance $\mathbf{C}_0$ can be updated as described in Section 1.3.4. The state feedback matrix $\mathbf{F}$ is constructed as indicated in Section 1.3.2 to incorporate nearest-neighbor interaction in order to model basic local intracortical dynamic connections. The selection of the values of ϕ (Eq. 1.4) is discussed in Section 1.3.5.

We note that the process of fixing the lead field $\mathbf{X}$ and state feedback $\mathbf{F}$ 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 $\mathbf{F}$, which are fixed based on neurophysiological modeling considerations, determine how much the state $\boldsymbol{\beta}_t$ can be explained by its past $\boldsymbol{\beta}_{t-1}$, and the unexplained portion is left to the state noise term $\boldsymbol{\omega}_t$. In the same way, the regressors in $\mathbf{X}$, obtained from geometrical and biophysical knowledge, separate the measured MEG $\mathbf{y}_t$ between the signals coming from the brain $\boldsymbol{\beta}_t$, and the instrument and environmental noise $\boldsymbol{\varepsilon}_t$.

The remaining unknown parameters, which must be estimated from data, are the elements of the state noise covariance matrix $\mathbf{Q}$. We assume this matrix is diagonal, and parametrize it as follows in order more clearly intepret and normalize parameter estimates as well as to simplify the initialization of our algorithms:

$$\mathbf{Q}(\boldsymbol{\nu}) = [\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1} \text{diag}(\boldsymbol{\nu}), \tag{1.6}$$

where $\widehat{\boldsymbol{\Sigma}} = \mathbf{X}'\mathbf{X}/n$ is the sample covariance of the rows of $\mathbf{X}$, $\lambda > 0$ is equal to the inverse of the signal-to-noise ratio (SNR^2) of the data[1], and $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_p]'$ ($\nu_i > 0$) is the parameter vector we need to estimate. Therefore, the number of variance parameters we need to estimate is in the order of thousands ($p \approx 5000$)

A common probability model for the parameter vector $\boldsymbol{\nu}$ is the conjugate prior

[1]Specifically, if the matrix $\mathbf{F}$ is set equal to the identity matrix, and $\nu_i = 1$ $(i = 1, \ldots, p)$, the steady state covariance of $\boldsymbol{\beta}_\mathbf{t}$ becomes $\mathbf{C} = \lim_{t\to\infty} \text{Cov}(\boldsymbol{\beta}_t) = [\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}\mathbf{I}$[55]. If we define the power signal-to-noise ratio as $\text{SNR}^2 = \text{E}||\mathbf{X}\boldsymbol{\beta}_t||^2/\text{E}||\boldsymbol{\varepsilon}_t||^2$, then $\text{SNR}^2 = \text{tr}(\mathbf{X}'\mathbf{X})[\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}/n = 1/\lambda$.

distribution. In this case the conjugate prior follows an inverse gamma density:

$$\Pr(\boldsymbol{\nu}) = \prod_{i=1}^{p} \frac{b^{b-1}}{\Gamma(b-1)} \left(\frac{1}{\nu_i}\right)^{b} \exp\left(\frac{-b}{\nu_i}\right) \tag{1.7}$$

where the hyperparameter b can be set to a value slightly higher than 3 to make the mode of this prior close to 1, and thus consistent with the order of magnitude in the model units, and at the same time maximize its variance yielding a non-informative prior (see Section 1.7.2).

1.3.4 Empirical Bayes inference with the 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

$$\{\boldsymbol{\beta}_t\}_{t=1}^{T} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_t\} \tag{1.8}$$

and parameters

$$\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_p]' \tag{1.9}$$

in the model. Specifically, we have to find the Maximum a Posteriori (MAP) estimate of the parameters,

$$\hat{\boldsymbol{\nu}}_{\text{MAP}} = \arg\max_{\boldsymbol{\nu}} \Pr(\boldsymbol{\nu}|\{\mathbf{y}_t\}_{t=1}^{T}), \tag{1.10}$$

where $\Pr(\boldsymbol{\nu}|\{\mathbf{y}_t\}_{t=1}^{T})$ is the posterior density of the parameters $\boldsymbol{\nu}$ conditioned on the full set of measurements

$$\{\mathbf{y}_t\}_{t=1}^{T} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}, \tag{1.11}$$

and the Empirical Bayes estimate of the source amplitudes, i.e., the conditional mean

of the state vector given the full set of measurements $\{\mathbf{y}_t\}_{t=1}^T$ and the estimate of the parameters in the model $\hat{\boldsymbol{\nu}}_{\text{MAP}}$,

$$\boldsymbol{\beta}_{t|T} = \text{E}(\boldsymbol{\beta}_t|\{\mathbf{y}_t\}_{t=1}^T, \hat{\boldsymbol{\nu}}_{\text{MAP}}), \tag{1.12}$$

where the notation in the subscript of $\boldsymbol{\beta}_{t|T}$ indicates that we are conditioning on the measurements form time 1 until time T.

Once we obtain the MAP estimate of the parameters $\hat{\boldsymbol{\nu}}_{\text{MAP}}$, computing the expectation in the Empirical Bayes estimate of the source amplitudes (Eq. 1.12) can be readily obtained using the well-known Kalman Filter [56] and Fixed Interval Smoother algorithms [57]. However, the direct optimization required to compute $\hat{\boldsymbol{\nu}}_{\text{MAP}}$ (Eq. 1.10) 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. 1.10 and 1.12) based on the work of [58] and [59]. The Expectation-Maximization algorithm [60] 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, $\{\mathbf{y}_t\}_{t=1}^T$, $\{\boldsymbol{\beta}_t\}_{t=0}^T$ 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 in the E-step (Section 1.3.4) can be computed with the Kalman Filter [56], Fixed Interval Smoother [57], and lag-covariance [61] recursions, and the maximization in the M-step (Section 1.3.4) can be obtained in closed form. In each iteration we can evaluate the posterior density of the parameters $\boldsymbol{\nu}$ using the *innovations* form [62] (Section 1.3.4).

E-step

The dMAP-EM algorithm is initialized with the parameters $\mathbf{Q}(\boldsymbol{\nu}^{(0)})$ and $\mathbf{C}_0^{(0)}$. In the ith iterate of the E-step, the algorithm computes the conditional expectation of the complete data log-likelihood $\Pr(\{\mathbf{y}_t\}_{t=1}^T, \{\boldsymbol{\beta}_t\}_{t=0}^T|\boldsymbol{\nu})$, given the observed data $\{\mathbf{y}_t\}_{t=1}^T$ and the previous estimates of the parameters $\boldsymbol{\nu}^{(i-1)}$, with an added term for

the log-prior density:

$$U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)}) = \mathrm{E}\left[\log \Pr\left(\{\mathbf{y}_t\}_{t=1}^{T}, \{\boldsymbol{\beta}_t\}_{t=0}^{T}|\boldsymbol{\nu}\right)|\{\mathbf{y}_t\}_{t=1}^{T}, \boldsymbol{\nu}^{(i-1)}\right] + \log \Pr(\boldsymbol{\nu}). \tag{1.13}$$

We emphasize that the expectation in Equation (1.13) is computed with respect to $\Pr(\{\boldsymbol{\beta}_t\}_{t=0}^{T}|\{\mathbf{y}_t\}_{t=1}^{T}, \boldsymbol{\nu}^{(i-1)})$, 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

$$\boldsymbol{\beta}_{t|\tau}^{(i)} = \mathrm{E}(\boldsymbol{\beta}_t|\{\mathbf{y}_j\}_{j=1}^{\tau}, \boldsymbol{\nu}^{(i-1)}), \tag{1.14}$$

the conditional covariance

$$\mathbf{V}_{t|\tau}^{(i)} = \mathrm{Cov}(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{j=1}^{\tau}, \boldsymbol{\nu}^{(i-1)}), \tag{1.15}$$

and the conditional lag-covariance

$$\mathbf{V}_{t,t-1|\tau}^{(i)} = \mathrm{Cov}(\boldsymbol{\beta}_t, \boldsymbol{\beta}_{t-1}|\{\mathbf{y}_j\}_{j=1}^{\tau}, \boldsymbol{\nu}^{(i-1)}), \tag{1.16}$$

where the subscript in Equations 1.14, 1.15, and 1.16 indicate that we are conditioning on the measurements form time 1 until time τ, and the superscript i refers to the iteration number. For example, as we used in Equation (1.12), by setting τ equal the number of samples in time T we indicate that are conditioning on the full set of measurements.

As shown in 1.7.3, the computation of the conditional expectations in the function $U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)})$ (Eq. 1.13) can be obtained in closed form yielding

$$
\begin{aligned}
U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)}) = &-\frac{1}{2}\left\{c_1 + \log|\mathbf{C}_0| + \mathrm{tr}\left[\mathbf{C}_0^{-1}\left(\mathbf{V}_{0|T}^{(i)} + \boldsymbol{\beta}_{0|T}^{(i)}\boldsymbol{\beta}_{0|T}^{(i)}\right)\right]\right\} \\
&- \frac{1}{2}\left\{c_2 T + T\log|\mathbf{Q}(\boldsymbol{\nu})| + (1-\phi^2)^{-1}\mathrm{tr}\left[\mathbf{Q}(\boldsymbol{\nu})^{-1}\mathbf{A}^{(i)}\right]\right\} \\
&- \frac{1}{2}\left\{c_3 T + \mathrm{tr}\left[\mathbf{B}^{(i)}\right]\right\} \\
&- \frac{1}{2}\sum_{j=1}^{p}\left\{c_4 + 2b\log\nu_j + 2\frac{b}{\nu_j}\right\},
\end{aligned} \tag{1.17}
$$

where $\{c_i\}_{i=1}^4$ do not depend on $\boldsymbol{\nu}$, and

$$
\mathbf{A}^{(i)} = \mathbf{A}_1^{(i)} - \phi\mathbf{A}_2^{(i)}\mathbf{F}' - \phi\mathbf{F}\mathbf{A}_2^{(i)\prime} + \phi^2\mathbf{F}\mathbf{A}_3^{(i)}\mathbf{F}' \tag{1.18}
$$

$$
\mathbf{B}^{(i)} = \sum_{t=1}^{T}\left[\left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|T}^{(i)}\right)\left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|T}^{(i)}\right)' + \mathbf{X}\mathbf{V}_{t|T}^{(i)}\mathbf{X}'\right], \tag{1.19}
$$

where

$$
\begin{aligned}
\mathbf{A}_1^{(i)} &= \sum_{t=1}^{T}\left(\mathbf{V}_{t|T}^{(i)} + \boldsymbol{\beta}_{t|T}^{(i)}\boldsymbol{\beta}_{t|T}^{(i)\prime}\right) \\
\mathbf{A}_2^{(i)} &= \sum_{t=1}^{T}\left(\mathbf{V}_{t,t-1|T}^{(i)} + \boldsymbol{\beta}_{t|T}^{(i)}\boldsymbol{\beta}_{t-1|T}^{(i)\prime}\right) \\
\mathbf{A}_3^{(i)} &= \sum_{t=1}^{T}\left(\mathbf{V}_{t-1|T}^{(i)} + \boldsymbol{\beta}_{t-1|T}^{(i)}\boldsymbol{\beta}_{t-1|T}^{(i)}\right).
\end{aligned} \tag{1.20}
$$

The conditional expectations and covariances in Equation (1.20) can be computed with the Kalman Filter [56], Fixed Interval Smoother [57], and lag-covariance [61] recursions.

The Kalman Filter, Fixed Interval Smoother, and lag-covariance recursions We can use the forward and backward recursions [62] to compute the desired expectations and covariances. In the ith iteration we initialize the recursion with

$\boldsymbol{\beta}_{0|0}^{(i)} = 0$ and $\mathbf{V}_{0|0}^{(i)} = \mathbf{C}_0^{(i-1)}$. For the forward pass, $t = 1, 2, \ldots, T$, we compute the predictions:

$$\begin{aligned}\boldsymbol{\beta}_{t|t-1}^{(i)} &= \phi \mathbf{F} \boldsymbol{\beta}_{t-1|t-1}^{(i)} \\ \mathbf{V}_{t|t-1}^{(i)} &= \phi^2 \mathbf{F} \mathbf{V}_{t-1|t-1}^{(i)} \mathbf{F}' + (1-\phi^2)\mathbf{Q}(\boldsymbol{\nu}^{(i-1)}),\end{aligned} \tag{1.21}$$

and filtered estimates:

$$\begin{aligned}\mathbf{G}_t &= \mathbf{V}_{t|t-1}^{(i)} \mathbf{X}' \left(\mathbf{X} \mathbf{V}_{t|t-1}^{(i)} \mathbf{X}' + \mathbf{I} \right)^{-1} \\ \boldsymbol{\beta}_{t|t}^{(i)} &= \boldsymbol{\beta}_{t|t-1}^{(i)} + \mathbf{G}_t \left(\mathbf{y}_t - \mathbf{X} \boldsymbol{\beta}_{t|t-1}^{(i)} \right) \\ \mathbf{V}_{t|t}^{(i)} &= \left(\mathbf{I} - \mathbf{G}_t \mathbf{X} \right) \mathbf{V}_{t|t-1}^{(i)},\end{aligned} \tag{1.22}$$

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

$$\begin{aligned}\mathbf{J}_t &= \phi \mathbf{V}_{t|t}^{(i)} \mathbf{F}' {\mathbf{V}_{t+1|t}^{(i)}}^{-1} \\ \boldsymbol{\beta}_{t|T}^{(i)} &= \boldsymbol{\beta}_{t|t}^{(i)} + \mathbf{J}_t \left(\boldsymbol{\beta}_{t+1|T}^{(i)} - \boldsymbol{\beta}_{t+1|t}^{(i)} \right) \\ \mathbf{V}_{t|T}^{(i)} &= \mathbf{V}_{t|t}^{(i)} + \mathbf{J}_t \left(\mathbf{V}_{t+1|T}^{(i)} - \mathbf{V}_{t+1|t}^{(i)} \right) \mathbf{J}_t'.\end{aligned} \tag{1.23}$$

The lag-covariances can be obtained using the covariance smoothing algorithm [61]:

$$\mathbf{V}_{t,t-1|T}^{(i)} = \mathbf{P}_{t|T}^{(i)} \mathbf{J}_{t-1}'. \tag{1.24}$$

The Kalman filter and Fixed Interval Smoother recursions are summarized in Algorithms 1 and 2, respectively. We should note that the conditional mean and covariance that we ultimately use for source localization (Eq. 1.12) correspond to the Fixed Interval Smoother estimate (FIS, Eq. 1.23) obtained on the final iteration of the MAP-EM algorithm. The FIS provides an estimate of the state based on the full

Algorithm 1 The Kalman Filter

Inputs: $\{\mathbf{y}_t\}_{t=1}^T, \phi, \mathbf{F}, \mathbf{X}, \mathbf{C}_0, \mathbf{Q}(\boldsymbol{\nu})$

$\boldsymbol{\beta}_{0|0} \leftarrow \mathbf{0}$ ▷ Initialization

$\mathbf{V}_{0|0} \leftarrow \mathbf{C}_0$

for $t = 1, \dots T$ **do** ▷ Kalman's Recursion

$\quad \boldsymbol{\beta}_{t|t-1} \leftarrow \phi \mathbf{F} \boldsymbol{\beta}_{t-1|t-1}$

$\quad \mathbf{V}_{t|t-1} \leftarrow \phi^2 \mathbf{F} \mathbf{V}_{t-1|t-1} \mathbf{F}' + (1 - \phi^2) \mathbf{Q}(\boldsymbol{\nu})$

$\quad \mathbf{G}_t \leftarrow \mathbf{V}_{t|t-1} \mathbf{X}' \left(\mathbf{X} \mathbf{V}_{t|t-1} \mathbf{X}' + \mathbf{I} \right)^{-1}$

$\quad \boldsymbol{\beta}_{t|t} \leftarrow \boldsymbol{\beta}_{t|t-1} + \mathbf{G}_t \left(\mathbf{y}_t - \mathbf{X} \boldsymbol{\beta}_{t|t-1} \right)$

$\quad \mathbf{V}_{t|t} \leftarrow \left(\mathbf{I} - \mathbf{G}_t \mathbf{X} \right) \mathbf{V}_{t|t-1}$

end for

Outputs: $\{\boldsymbol{\beta}_{t|t-1}, \boldsymbol{\beta}_{t|t}, \mathbf{V}_{t|t-1}, \mathbf{V}_{t|t}\}_{t=1}^T$

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 [63].

In 1.7.4, 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) [11]. 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.

Algorithm 2 The Fixed Interval Smoother

Inputs: $\{\boldsymbol{\beta}_{t|t-1}, \boldsymbol{\beta}_{t|t}, \mathbf{V}_{t|t-1}, \mathbf{V}_{t|t}\}_{t=1}^T, \phi, \mathbf{F}$

for $t = T-1, \dots, 0$ **do** ▷ Rauch, Tung, and Striebel's Recursion

$\quad \mathbf{J}_t \leftarrow \phi \mathbf{V}_{t|t} \mathbf{F}' \mathbf{V}_{t+1|t}^{-1}$

$\quad \boldsymbol{\beta}_{t|T} \leftarrow \boldsymbol{\beta}_{t|t} + \mathbf{J}_t \left(\boldsymbol{\beta}_{t+1|T} - \boldsymbol{\beta}_{t+1|t} \right)$

$\quad \mathbf{V}_{t|T} \leftarrow \mathbf{V}_{t|t} + \mathbf{J}_t \left(\mathbf{V}_{t+1|T} - \mathbf{V}_{t+1|t} \right) \mathbf{J}_t'$

end for

Outputs: $\{\boldsymbol{\beta}_{t|T}, \mathbf{V}_{t|T}, \mathbf{J}_t\}_{t=1}^T$

M-step

The M-step in the ith iterate is achieved by maximizing with respect to the parameters $\boldsymbol{\nu}$ the function $U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)})$ computed in the E-step (Eq. 1.17), which equals the sum of two terms: 1) the conditional expectation of the complete data log-likelihood given the full set of measurements and the previous estimates of the parameters, and 2) the log-prior density,

$$\boldsymbol{\nu}^{(i)} = \arg\max_{\boldsymbol{\nu}} U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)}). \tag{1.25}$$

It is easy to see that the maxima is achieved at

$$\nu_j^{(i)} = \frac{a_{j,j}^{(i)} \frac{\lambda \mathrm{tr}(\widehat{\boldsymbol{\Sigma}})}{(1-\phi^2)n} + 2b}{T + 2b}, \tag{1.26}$$

where $a_{j,j}^{(i)}$ is the ith row of the ith column of $\mathbf{A}^{(i)}$ (Eq. 1.18), and b is the hyperparameter defined in Equation (1.7).

Evaluation of the log-posterior density

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

$$\log \Pr(\boldsymbol{\nu}|\{\mathbf{y}_t\}_{t=1}^T) = \underbrace{\log \Pr(\{\mathbf{y}\}_{t=1}^T|\boldsymbol{\nu})}_{\text{Log-likelihood}} + \underbrace{\log \Pr(\boldsymbol{\nu})}_{\text{Log-prior}} - \underbrace{\log \Pr(\{\mathbf{y}_t\}_{t=1}^T)}_{\text{Log-evidence}} \tag{1.27}$$

evaluated at $\boldsymbol{\nu} = \boldsymbol{\nu}^{(i)}$ reaches an asymptote at some iteration i_o.

We can evaluate the logarithm of the likelihood, i.e., the first term in Equation (1.27), using the *innovations* form [62],

$$\log \Pr(\{\mathbf{y}_t\}_{t=1}^T|\boldsymbol{\nu}^{(i)}) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{X}\mathbf{V}_{t|t-1}^{(i)}\mathbf{X}' + \mathbf{I}| \tag{1.28}$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|t-1}^{(i)}\right)'\left(\mathbf{X}\mathbf{V}_{t|t-1}^{(i)}\mathbf{X}' + \mathbf{I}\right)^{-1}\left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|t-1}^{(i)}\right), \tag{1.29}$$

and the logarithm of the prior is obtained is obtained from Equation (1.7):

$$\log \Pr(\boldsymbol{\nu}^{(i)}) = \sum_{j=1}^{p}\left\{c_4 - b\log\nu_j^{(i)} - \frac{b}{\nu_j^{(i)}}\right\}. \tag{1.30}$$

Since the evidence in the data, $\Pr(\{\mathbf{y}_t\}_{t=1}^T)$, is a constant not depending on $\boldsymbol{\nu}$, we do not need to compute it in any iteration.

Summary of the dMAP-EM algorithm

The algorithm is initialized with parameters $\boldsymbol{\nu}^{(0)}$ and $\mathbf{C}_0^{(0)}$. In the ith iteration, we set the state noise covariance $\mathbf{Q}(\boldsymbol{\nu}^{(i-1)}) = [\lambda\text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}\text{diag}(\boldsymbol{\nu}^{(i-1)})$ and initial state covariance $\mathbf{V}_{0|0}^{(i)} = \mathbf{C}_0^{(i-1)}$. We then perform an E-step (Section 1.3.4) by running the Kalman Filter, Fixed Interval Smoother, and lag-covariance recursion (Section 1.3.4), and perform an M-step (Section 1.3.4) to update the parameters $\boldsymbol{\nu}^{(i)}$. At each iteration we can update $\mathbf{C}_0$ with the heuristic $\mathbf{C}_0^{(i)} = \mathbf{V}_{0|T}^{(i)}$. The algorithm iterates for $i = 1, 2, \ldots, i_o$, performing an E-step followed by an M-step until the log-posterior density evaluated at $\boldsymbol{\nu}^{(i_o)}$ converges (Section 1.3.4). The Maximum a Posteriori (MAP) estimate of the parameters is then $\hat{\boldsymbol{\nu}}_{\text{MAP}} = \boldsymbol{\nu}^{(i_o)}$, and the Empirical Bayes estimate of the sources amplitudes is $\boldsymbol{\beta}_{t|T} = \boldsymbol{\beta}_{t|T}^{(i_o)}$. The dMAP-EM algorithm is summarized in Algorithm 3.

1.3.5 Initialization of algorithms

The state feedback matrix $\mathbf{F}$ was constructed as indicated in Section 1.3.2 to incorporate nearest-neighbor interaction in order to model basic local intracortical connec-

Algorithm 3 The dMAP-EM Algorithm

Inputs: $\{\mathbf{y}_t\}_{t=1}^T, \phi, \mathbf{F}, \mathbf{X}, \lambda \mathrm{tr}(\widehat{\mathbf{\Sigma}})/n$
Initialization:
$i \leftarrow 0$
$\boldsymbol{\nu}^{(i)} \leftarrow [1, 1, \ldots, 1]$
$\mathbf{C}_0^{(i)} \leftarrow [\lambda \mathrm{tr}(\widehat{\mathbf{\Sigma}})/n]^{-1}\mathbf{I}$
repeat
$i \leftarrow i + 1$
$\mathbf{Q}(\boldsymbol{\nu}^{(i-1)}) \leftarrow [\lambda \mathrm{tr}(\widehat{\mathbf{\Sigma}})/n]^{-1}\mathrm{diag}(\boldsymbol{\nu}^{(i-1)})$
E-step: ▷ Algorithms 1 and 2
$\{\boldsymbol{\beta}_{t|t-1}^{(i)}, \boldsymbol{\beta}_{t|t}^{(i)}, \mathbf{V}_{t|t-1}^{(i)}, \mathbf{V}_{t|t}^{(i)}\}_{t=1}^T \leftarrow \mathrm{KalmanFilter}(\{\mathbf{y}_t\}_{t=1}^T, \phi, \mathbf{F}, \mathbf{X}, \mathbf{C}_0^{(i-1)}, \mathbf{Q}(\boldsymbol{\nu}^{(i-1)}))$
$\{\boldsymbol{\beta}_{t|T}^{(i)}, \mathbf{V}_{t|T}^{(i)}, \mathbf{J}_t\}_{t=1}^T \leftarrow \mathrm{FixedIntervalSmoother}(\{\boldsymbol{\beta}_{t|t-1}^{(i)}, \boldsymbol{\beta}_{t|t}^{(i)}, \mathbf{V}_{t|t-1}^{(i)}, \mathbf{V}_{t|t}^{(i)}\}_{t=1}^T, \phi, \mathbf{F})$
for $t = 1, \ldots, T$ **do** ▷ Lag Covariance Recursion
$\mathbf{V}_{t,t-1|T}^{(i)} = \mathbf{P}_{t|T}^{(i)}\mathbf{J}_{t-1}'$
end for
$\mathbf{A}_1^{(i)} \leftarrow \sum_{t=1}^T \left(\mathbf{V}_{t|T}^{(i)} + \boldsymbol{\beta}_{t|T}^{(i)}\boldsymbol{\beta}_{t|T}^{(i)\prime}\right)$
$\mathbf{A}_2^{(i)} \leftarrow \sum_{t=1}^T \left(\mathbf{V}_{t,t-1|T}^{(i)} + \boldsymbol{\beta}_{t|T}^{(i)}\boldsymbol{\beta}_{t-1|T}^{(i)\prime}\right)$
$\mathbf{A}_3^{(i)} \leftarrow \sum_{t=1}^T \left(\mathbf{V}_{t-1|T}^{(i)} + \boldsymbol{\beta}_{t-1|T}^{(i)}\boldsymbol{\beta}_{t-1|T}^{(i)}\right)$
$\mathbf{A}^{(i)} \leftarrow \mathbf{A}_1^{(i)} - \phi\mathbf{A}_2^{(i)}\mathbf{F}' - \phi\mathbf{F}\mathbf{A}_2^{(i)\prime} + \phi^2\mathbf{F}\mathbf{A}_3^{(i)}\mathbf{F}'$ ▷ Equations (1.18) and (1.20)
M-step:
$\nu_j^{(i)} \leftarrow \left(a_{j,j}^{(i)} \frac{\lambda \mathrm{tr}(\widehat{\mathbf{\Sigma}})}{(1-\phi^2)n} + 2b\right)(T + 2b)^{-1}$
$\mathbf{C}_0^{(i)} \leftarrow \mathbf{V}_{0|T}^{(i)}$
Evaluation of log-posterior:
$\log \Pr(\boldsymbol{\nu}^{(i)}|\{\mathbf{y}_t\}_{t=1}^T) \leftarrow -\frac{1}{2}\sum_{t=1}^T \log|\mathbf{X}\mathbf{V}_{t|t-1}^{(i)}\mathbf{X}' + \mathbf{I}|$
$-\frac{1}{2}\sum_{t=1}^T \left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|t-1}^{(i)}\right)' \left(\mathbf{X}\mathbf{V}_{t|t-1}^{(i)}\mathbf{X}' + \mathbf{I}\right)^{-1} \left(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|t-1}^{(i)}\right)$
$\log \Pr(\boldsymbol{\nu}^{(i)}) \leftarrow -\sum_{j=1}^p \left[b \log \nu_j^{(i)} + \frac{b}{\nu_j^{(i)}}\right]$
until $\log \Pr(\boldsymbol{\nu}^{(i)}|\{\mathbf{y}_t\}_{t=1}^T) + \log \Pr(\boldsymbol{\nu}^{(i)})$ converges
$i_o \leftarrow i$
$\hat{\boldsymbol{\nu}}_{\mathrm{MAP}} \leftarrow \boldsymbol{\nu}^{(i_o)}$
$\boldsymbol{\beta}_{t|T} \leftarrow \boldsymbol{\beta}_{t|T}^{(i_o)}$
$\mathbf{V}_{t|T} \leftarrow \mathbf{V}_{t|T}^{(i_o)}$
Outputs: $\{\boldsymbol{\beta}_{t|T}, \mathbf{V}_{t|T}\}_{t=1}^T, \hat{\boldsymbol{\nu}}_{\mathrm{MAP}}$

tions. To evaluate the weights $f_{i,j}$ (Eqs. 1.3 and 1.5), the distance between dipoles was obtained from the triangular tessellation of the cortical surface. The value of ϕ in Equation (1.5) was set to 0.95: this constrains the modulus of the largest eigenvalue of $\phi\mathbf{F}$ to be strictly less that 1, and guarantees stability in the source model (Eq. 1.4).

We should note that although it would be interesting to estimate the parameters $f_{i,j}$ and ϕ, this would add ~ 30000 degrees of freedom to the estimation task—each dipole has about 6 nearest neightbors and there are about 5000 dipoles—making this option unfeasible.

The measurement noise covariance was set equal to the identity matrix since the model has been spatially whitened, i.e., the original raw data model $\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}\boldsymbol{\beta}_t + \tilde{\boldsymbol{\varepsilon}}_t$ has been premutiplied by the inverse of a matrix square root of the sample covariance of $\tilde{\boldsymbol{\varepsilon}}_t$, which was in turn estimated from from empty room recordings. To initialize the algorithms we set the source covariance at time zero ($\mathbf{C}_0^{(0)}$) and the input noise covariance as a multiple of the identity:

$$\mathbf{C}_0^{(0)} = \mathbf{Q}(\boldsymbol{\nu}^{(0)}) = [\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}\mathbf{I} \tag{1.31}$$

thus approximating the power signal-to-noise ratio (SNR) of the measurements.

1.4 Data analysis

1.4.1 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, $\boldsymbol{\beta}_t^{(\text{MNE})}$ in Equation (1.50)) [11]; 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 [27] and [21]; 3) The FIS without the estimation of the parameters $\boldsymbol{\nu}$ ($\boldsymbol{\beta}_{t|T}^{(0)}$ in Equation (1.23), i.e., the smoothed estimate in the first iteration of our algorithm); and 4) And our dMAP-EM algorithm ($\boldsymbol{\beta}_{t|T}^{(i_o)}$ in Equation (1.23), 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. 1.7) for the source variances. Therefore, the sMAP-EM is similar to our method, but with the state feedback matrix $\mathbf{F}$ set to zero, i.e., $\mathbf{F} = \mathbf{0}$. It provides a baseline to compare how the matrix $\mathbf{F}$, as specified to model local intracortical

connections, can be used as covariates or regressors to explain the source activity $\boldsymbol{\beta}_t$ in terms of its past $\boldsymbol{\beta}_{t-1}$.

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 (Figure 1-2 top panel);
- *Small Patch:* We chose a small active region over primary auditory cortex (Figure 1-4 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 [54] using a single-compartment boundary-element model (BEM) based on high-resolution MRIs processed with *Freesurfer* [52, 53] from a human subject. The measurement equation (Eq. 1.1) 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.

1.4.2 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 Section 1.4.1. 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 $\boldsymbol{\nu}$. 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 $\mathbf{V}_{t|T}^{(1)}$ (Eq. 1.23), while the CIs of the dMAP-EM source estimates equal ± 2 times the diagonal elements of $\mathbf{V}_{t|T}^{(i_o)}$, 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 $\mathbf{F} = 0$ in our formulation.

Figure 1-2 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 $\boldsymbol{\beta}_{j,t}$ 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 patch 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.

Figure 1-3 compares the temporal tracking performance for representative dipole sources located inside the active area for the large patch simulation. The MNE time series greatly under-estimates the amplitude of the true simulated time series and present wide CIs. The sMAP-EM either underestimates (sub-panels B and C) or overestimates (sub-panel D) the source amplitude and presents larger CIs that 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 (SI1.mov) a video with the results of this simulation study that visualizes dynamically

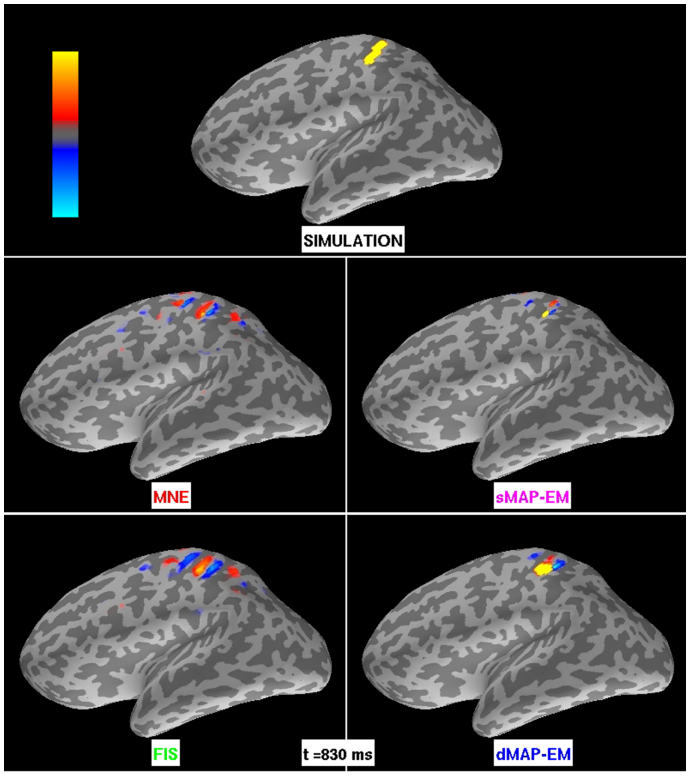

Figure 1-2: **Large cortical patch simulation results for MNE, sMAP-EM, FIS, and dMAP-EM methods**. The intensity maps represent the amplitude of simulated and estimated sources. The colorbar's maximum (bright yellow) and minimum (bright blue) corresponds to ± 3 nAm for all methods. The top panel shows 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.

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 Figure 1-4. 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.

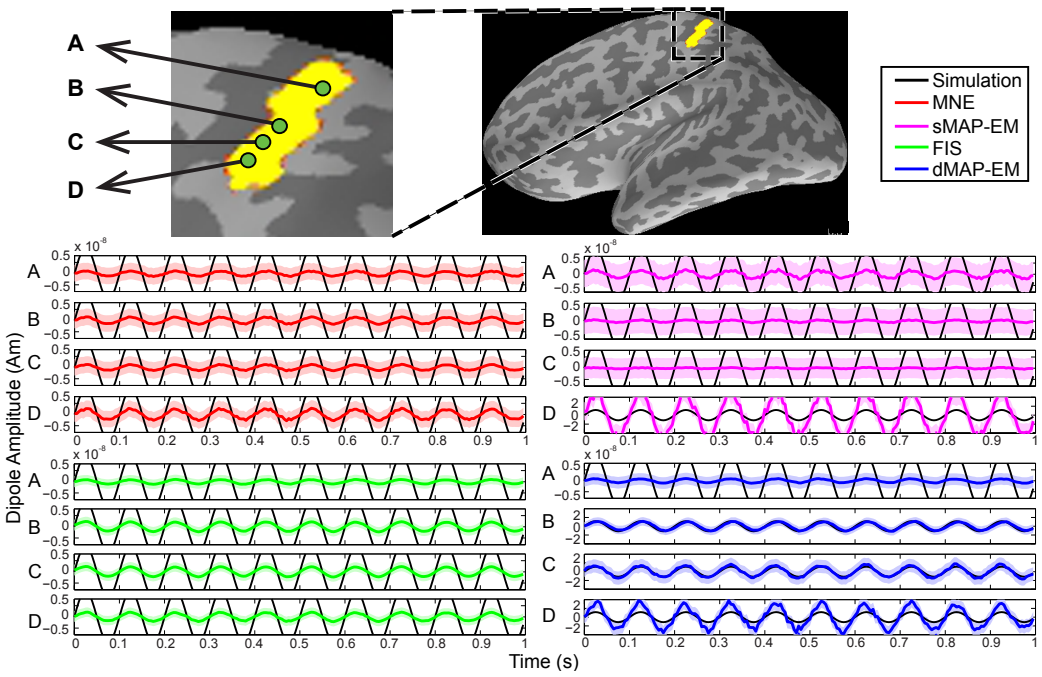

Figure 1-3: **Time course estimates for the large patch simulation**. The upper panel shows a zoomed-in view of the simulated cortical area, where green dots represent the selected dipoles labeled A, B, C, and D. The black lines represent simulated sources, while the colored lines represent estimated 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 amplitrude 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.

Figure 1-5 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

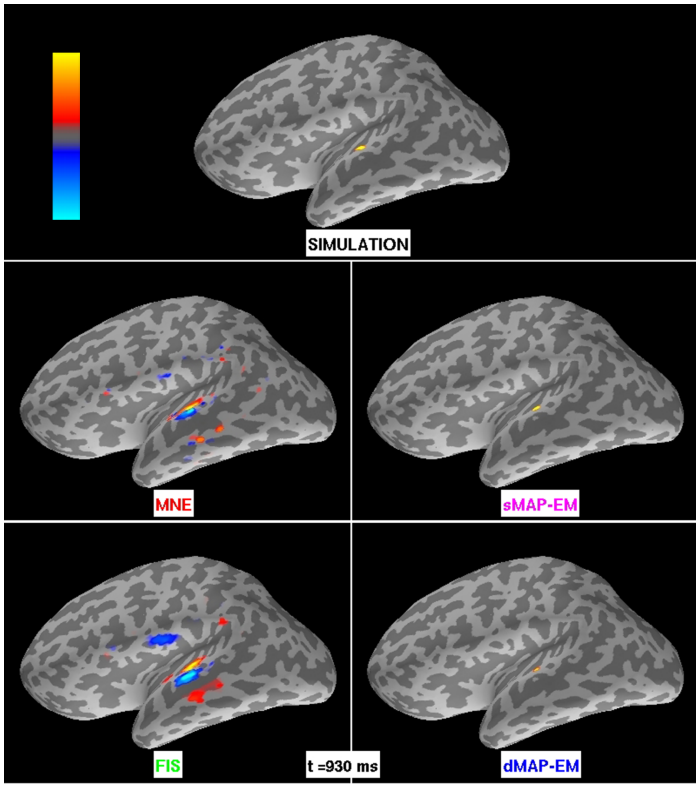

Figure 1-4: **Small cortical patch simulation results for MNE, sMAP-EM, FIS, and dMAP-EM methods.** The intensity maps represent the amplitude of simulated and estimated sources. The colorbar's maximum (bright yellow) and minimum (bright blue) corresponds to ± 3.5 nAm for MNE and FIS, and ± 70 nAm for 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.

(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 (SI2.mov) a video with the results of this simulation study.

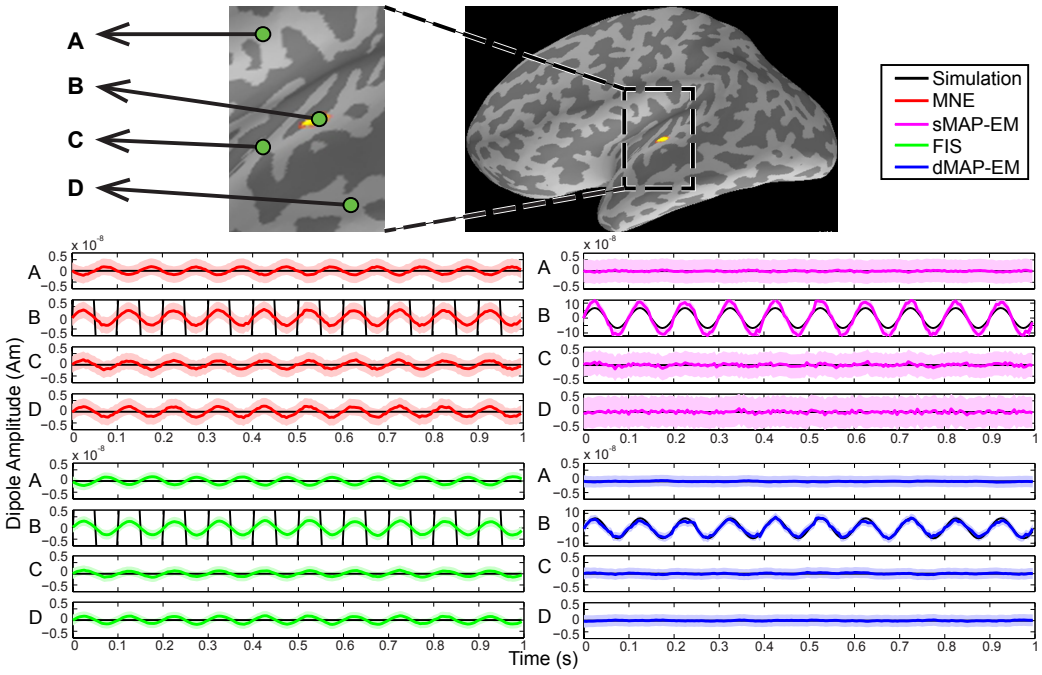

Figure 1-5: **Time course of estimation results for the small patch simulation.** The upper shows a zoomed-in view of the simulated cortical area where green dots represent selected dipoles labeled A, B, C, and D. The black lines represent simulated sources, while the colored lines represent estimated 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, D, 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.

1.4.3 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 [64], or [65] for an application in the MEG/EEG inverse problem) and root mean square errors (RMSE)

of sources estimates in the simulation studies described in Sections 1.4.1 and 1.4.2.

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,

$$\mathrm{pr}_D = \Pr(\text{reject } \mathrm{H}_0 | \mathrm{H}_A \text{ is true}), \tag{1.32}$$

and the false alarm probability,

$$\mathrm{pr}_{FA} = \Pr(\text{reject } \mathrm{H}_0 | \mathrm{H}_0 \text{ is true}), \tag{1.33}$$

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 jth dipole source at time t equals zero ($\beta_{j,t}^{(\mathrm{SIM})} = 0$), while a source signal is considered active (H_A is true) when $\beta_{j,t}^{(\mathrm{SIM})} \neq 0$. For a given estimation method, we define $\hat{\beta}_{j,t}$ as the jth 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 $|\hat{\beta}_{j,t}| > c$. Consequently, for a specific threshold c, an estimate of the detection probability $\widehat{\mathrm{pr}}_D$ is given by the fraction of events where we correctly detected an active source, i.e., when the dipole source estimate was considered active ($|\hat{\beta}_{j,t}| > c$) given that the underlying true source was active ($\beta_{j,t}^{(\mathrm{SIM})} \neq 0$). Similarly, the estimate of false alarm probability $\widehat{\mathrm{pr}}_{FA}$ is given by the proportion of events where we considered the source estimate to be active ($|\hat{\beta}_{j,t}| > c$) but the underlying true source was inactive ($\beta_{j,t}^{(\mathrm{SIM})} = 0$) and made a "false alarm" (See 1.7.5 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 study are shown in the sub-

panels A and B of Figure 1-6, 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 ($\widehat{\text{pr}}_D \approx 0.9$) with as few as $\sim 2\%$ false alarms ($\widehat{\text{pr}}_{FA} \approx 0.02$), 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 Figure 1-6), 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 $\hat{\beta}_{j,t}$ in relation to the true simulated sources $\beta_{j,t}^{(\text{SIM})}$ 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 ($j = 1, 2, \ldots, p$) over the length in time ($T = 200$) of the simulation,

$$\text{RMSE}_j = \sqrt{\frac{\sum_{t=1}^{T} \left(\hat{\beta}_{j,t} - \beta_{j,t}^{(\text{SIM})}\right)^2}{T}}, \tag{1.34}$$

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. 1-6, sub-panels C and D). However, because the number of dipole source located outside the active patch was relatively large, we computed box-plot summaries of the RMSE of sources from this region, thereby

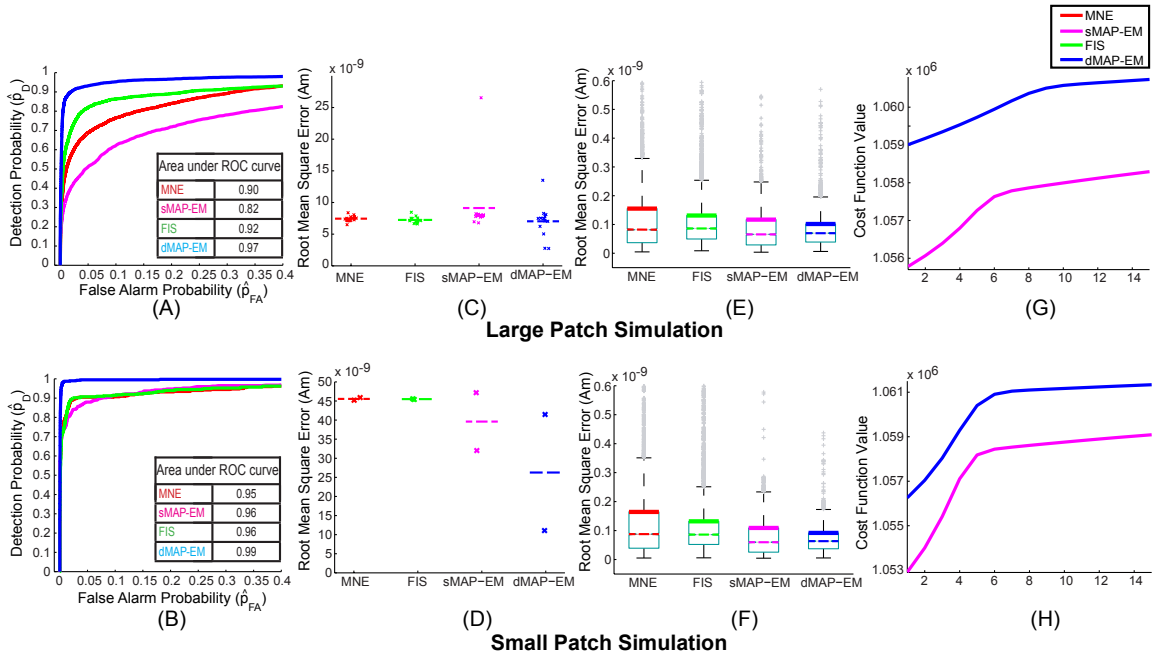

Figure 1-6: **ROC curves, RMSE, and convergence of algorithms.** ROC curves and area under the ROC curves from the large patch simulation (sub-panel A) and the small patch simulation (sub-panel B) show that the dMAP-EM method outperforms MNE, FIS, and sMAP-EM methods as it detects more active dipole sources while making significantly less false alarms. Sub-panels C and D show scatter plots of the RMSE of dipoles 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.

displaying 0.01, 0.25, 0.5, 0.75, and 0.99 approximate quantiles (Fig. 1-6, 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 grey crosses outside whiskers are considered

outliers.

Sub-panels C and D of Figure 1-6 show the scatter plots and the average (dash 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 of the average RMSE in relation to the other methods. In the large patch simulation the dMAP-EM method yielded an average RMSE of 7 nAm, which represents an reduction of $\sim$ 5.4%, $\sim$ 2.7%, and $\sim$ 23% in the average of errors with respect to MNE, FIS, and sMAP-EM methods, respectively. Similarly, the dMAP-EM method resulted with an average RMSE of 26 nAm in the small patch simulation showing a reduction of $\sim$ 42% with respect to MNE and FIS, and of $\sim$ 33% in relation 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 Figure 1-6 show the RMSE box-plots of dipole sources located outside the active region in large and small patch simulation, respectively. In both simulated scenarios, our dMAP-EM method yielded the smallest quantiles among the analyzed methods. The reduction in RMSE relative to MNE, FIS, and sMAP-EM are shown in Table 1.1. 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.

	Method \ **Quantile**		**0.5**	**0.75**	**0.99**
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%

Table 1.1: Percentage reduction of the RMSE quantiles for dipole sources outside the active region under dMAP-EM, compared to MNE, FIS, and sMAP-EM

1.4.4 Convergence and computational requirements

The convergence of the dMAP-EM algorithm was assessed by tracking the logarithm of the posterior density (Eq. 1.27),

$$\text{cost}(\boldsymbol{\nu}^{(i)}) = \log \Pr(\{\mathbf{y}\}_{t=1}^{T} | \boldsymbol{\nu}^{(i)}) + \log \Pr(\boldsymbol{\nu}^{(i)}) \tag{1.35}$$

omitting the log-evidence term, which does change during EM iteration. Sub-panels G and H of Figure 1-6 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 $O(Tp^3)$ [66], where we assumed that the number of sensors n is fixed and much smaller than the number of dipole sources p ($n << p$), 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 $p = 5124$, the number of sensor was $n = 204$, 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 hours per EM iteration.

1.4.5 Analysis of experimental data from human subjects

We also 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 minutes of data at a sampling frequency of 601 Hz with a bandwidth of 0.1 to 200 Hz, and later downsampled to 200.3 Hz. The data were visually inspected to select 1 second of strong *mu*-rhythm activity, as evidenced by

its characteristic "μ" shape, for subsequent analyses. In Figure 1-7, we show the results of the three methods where the intensity maps represent the amplitude of source estimates. This figure also includes the topography of the magnetic field component normal to the sensor surface at the same time instant [11, 67]. Similar to Figures 1-2 and 1-4 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 Figure 1-8. The 95% CIs (light colored areas) are obtained as described in Section 1.4.2. 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 *mu*-rhythm signals [68]. 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 *mu*-shape that characterizes these data, and presented smaller CIs than those of the MNE and sMAP-EM methods. In the Supplementary Information 3 (SI3.mov) we present a video of the results of this analysis.

1.5 Discussion

Intracranial electrophysiological recordings have shown that on a local scale, brain activity is spatially and temporally correlated [36, 37, 38, 39]. Similarly, non-invasive fMRI and PET studies have shown that brain activation is temporally coherent in

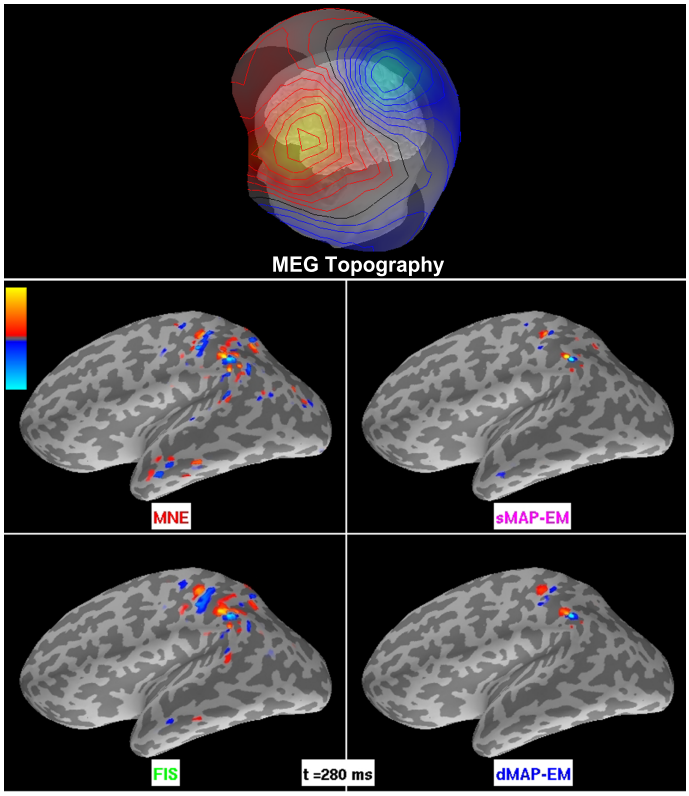

Figure 1-7: Analysis of human MEG *mu*-rhythm data. The top panel shows the topography of the magnetic field component normal to the sensor surface, with isocontour lines indicating a field change of 100 fT. In the remaining panels, the intensity maps represent the amplitude of source estimates. The 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.

a spatially distributed network [48, 49, 50, 51]. 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 [40, 41, 42, 43, 44, 45, 46, 47]. 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

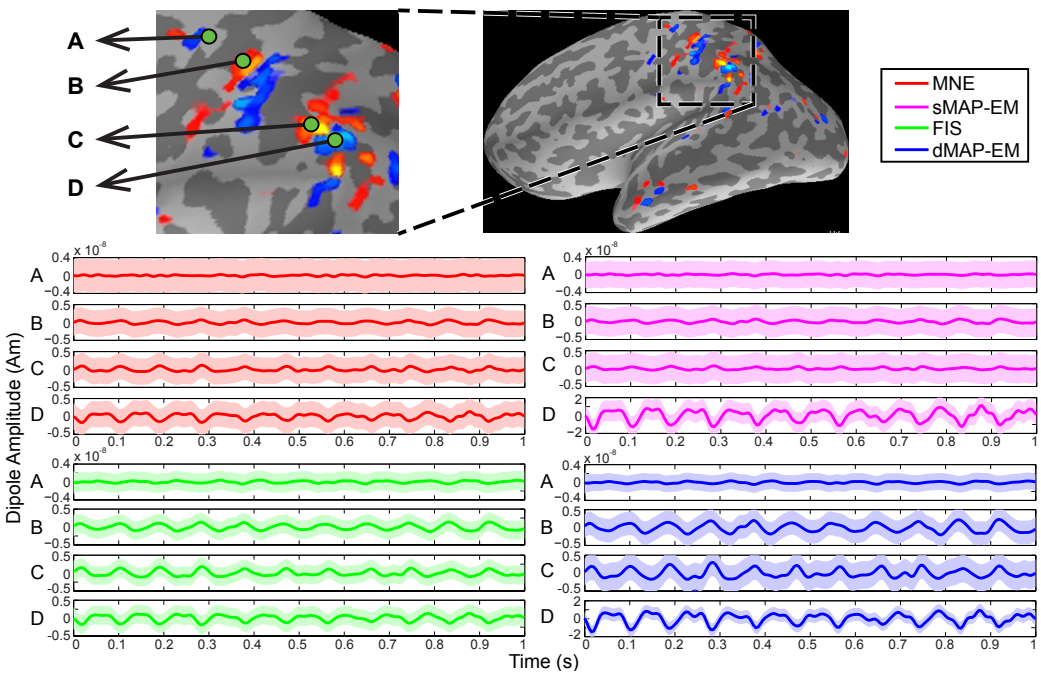

Figure 1-8: Time course of estimation results for human MEG *mu*-rhythm data. The upper panel shows a zoomed-in view of the cortical activity map obtained with the MNE method, where the green dots represent dipoles labeled A, B, C, and D. The colored lines are estimated sources with 95% CIs represented 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 resulted in 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 *mu*-shape that characterizes these data (dipoles B, C, and D).

model acts as a soft *a priori* constraint on the evolution of spatiotemporal cortical trajectories, in a way that allows 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 [23, 24, 25, 26, 27, 28, 29, 30, 31, 32] 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 [33] or an approximate version of the Kalman filter [34]. 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 1.7.4). 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 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 [69] and [35], 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 spatiotemporal 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 *mu*-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 *mu*-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 [21] 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 [70, 71]. 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.

1.6 Conclusions and future work

In this work: 1) we developed a distributed stochastic dynamic model based on a nearest-neighbors autoregression on the cortical surface to represent spatiotemporal cortical dynamics; 2) we derived the dMAP-EM algorithm for optimal dynamic esti-

mation 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) we developed expressions to relate our dynamic estimation method to standard static algorithms; and 4) we 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.

1.7 Supplemental materials

1.7.1 Robustness of source model against variations in the feedback matrix F

In this appendix we analyze how modifications in our spatial model ($\mathbf{F}$) 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 $\mathbf{F}$ matrix. To simplify our notation, we combine the parameter ϕ into the definitions of $\mathbf{F}$ and $\mathbf{Q}$. To show the robustness our our model in this area, we first take a modified state model (Eq. 1.4) where the feedback matrix $\tilde{\mathbf{F}} = \mathbf{F} + \mathbf{\Delta}_F$ has been perturbed by $\mathbf{\Delta}_F$. Then, we consider the prior source covariance of the original model $\mathbf{C}$ and that

of the modified model $\tilde{\mathbf{C}}$, when they have reached equilibrium (steady-state), i.e., $\lim_{t\to\infty} \mathrm{Cov}(\boldsymbol{\beta}_t)$. At last, we derive an upper bound for the matrix 2-norm of the difference between the equilibrium covariances $\boldsymbol{\Delta}_C = \tilde{\mathbf{C}} - \mathbf{C}$, as a function of the perturbation of the spatial model $\boldsymbol{\Delta}_F$.

We assume that both $\mathbf{F}$ and $\tilde{\mathbf{F}}$ 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 ($\mathbf{C}$) and that of the modified model ($\tilde{\mathbf{C}}$) correspond to the respective (unique) solutions of the discrete Lyapunov equations [72]:

$$\begin{aligned} \mathbf{C} &= \mathbf{F}\mathbf{C}\mathbf{F}' + \mathbf{Q} \\ \tilde{\mathbf{C}} &= \tilde{\mathbf{F}}\tilde{\mathbf{C}}\tilde{\mathbf{F}}' + \mathbf{Q}. \end{aligned} \tag{1.36}$$

We now express the difference in the equilibrium covariances $\boldsymbol{\Delta}_C$ by subtracting the equalities above to obtain:

$$\boldsymbol{\Delta}_C = \mathbf{F}\boldsymbol{\Delta}_C\mathbf{F}' + \mathbf{F}\tilde{\mathbf{C}}\boldsymbol{\Delta}_F' + (\mathbf{F}\tilde{\mathbf{C}}\boldsymbol{\Delta}_F')' + \boldsymbol{\Delta}_F\tilde{\mathbf{C}}\boldsymbol{\Delta}_F'. \tag{1.37}$$

Now we take the matrix induced 2-norm in Equation (1.37), and apply repeatedly the triangle inequality and the sub-multiplicative property of induced norms:

$$||\boldsymbol{\Delta}_C||_2 \leq ||\mathbf{F}||_2^2||\boldsymbol{\Delta}_C||_2 + 2||\mathbf{F}||_2||\tilde{\mathbf{C}}||_2||\boldsymbol{\Delta}_F||_2 + ||\tilde{\mathbf{C}}||_2||\boldsymbol{\Delta}_F||_2^2. \tag{1.38}$$

We rearrange terms in Equation (1.38) to obtain:

$$||\boldsymbol{\Delta}_C||_2 \leq \frac{||\tilde{\mathbf{C}}||_2(2||\mathbf{F}||_2 + \boldsymbol{\Delta}_F||_2)}{1 - ||\mathbf{F}||_2^2} \cdot ||\boldsymbol{\Delta}_F||_2 \tag{1.39}$$

The fact that both feedback matrices are stable is equivalent to $||\mathbf{F}||_2$, $||\tilde{\mathbf{F}}||_2 < 1$. Using the inverse triangle inequality and setting $\mathbf{F} = \tilde{\mathbf{F}} - \boldsymbol{\Delta}_F$, we obtain $|\,||\tilde{\mathbf{F}}||_2 - ||\boldsymbol{\Delta}_F||_2\,| \leq ||\mathbf{F}||_2 < 1$. This inequality implies that $||\boldsymbol{\Delta}_F||_2 < 2$. We plug the previous result and the fact that $||\mathbf{F}||_2 < 1$ in the inequality above (Eq. 1.39), and

find that the matrix 2-norm of the difference between the equilibrium covariances $\mathbf{\Delta}_C$ is bounded above by a constant multiplied by the 2-norm of the perturbation in the feedback matrix $\mathbf{\Delta}_F$:

$$||\mathbf{\Delta}_C||_2 < c \cdot ||\mathbf{\Delta}_F||_2, \tag{1.40}$$

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

From Equation (1.40) we can conclude that as long as the misspecification of the feedback matrix $\mathbf{\Delta}_F$ is small, the smoothness modeled *a priori* in the dipole sources is not dramatically altered.

1.7.2 A non-informative prior for the state noise covariance $\mathbf{Q}(\boldsymbol{\nu})$

In this section we show that setting the parameter b in our prior (Eq. 1.7) to a value slightly larger than 3, makes our prior on ν_j non-informative (flat). This can be achieved by giving a large variance to the prior while fixing its mode to a value consistent with the order of magnitude in the model. In our case, the order of magniture of ν_j ($j \in [1, \ldots, p]$) is 1 since this makes our model consistent with SNR considerations as well as model units. We should note that for $b > 3$, the mode of the prior is 1. This results from the way we parametrice the inverse gamma prior. The variance of our prior is given by:

$$\mathrm{Var}(\nu_i) = \frac{b^2}{(b-2)^2(b-3)}. \tag{1.41}$$

From this equation (Eq. 1.41) we can see that setting $b = 3+\delta$, where $0 < \delta << b$, makes the variance very large and thus give a non-informative (flat) prior.

1.7.3 Derivation of $U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)})$ in the E-step

From Equation (1.13) we have that

$$U(\boldsymbol{\nu}|\boldsymbol{\nu}^{(i-1)}) = \mathrm{E}\left[\log \Pr\left(\{\mathbf{y}_t\}_{t=1}^T, \{\boldsymbol{\beta}_t\}_{t=0}^T|\boldsymbol{\nu}\right)|\{\mathbf{y}_t\}_{t=1}^T, \boldsymbol{\nu}^{(i-1)}\right] + \log \Pr(\boldsymbol{\nu}). \tag{1.42}$$

where the complete data log-likelihood is derived from the measurement (Eq. 1.1) and source (Eq. 1.4) models:

$$\log \Pr(\{\mathbf{y}_t\}_{t=1}^T, \{\boldsymbol{\beta}_t\}_{t=0}^T|\boldsymbol{\nu}) = \sum_{t=1}^T \log \Pr(\mathbf{y}_t|\boldsymbol{\beta}_t, \boldsymbol{\nu}) + \sum_{t=1}^T \log \Pr(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\nu}) + \log \Pr(\boldsymbol{\beta}_0|\boldsymbol{\nu}), \tag{1.43}$$

with,

$$\begin{aligned}
\log \Pr(\boldsymbol{\beta}_0|\boldsymbol{\nu}) &= -\frac{1}{2}\{c_1 + \log|\mathbf{C}_0| + \boldsymbol{\beta}_0'\mathbf{C}_0^{-1}\boldsymbol{\beta}_0\} \\
\log \Pr(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\nu}) &= -\frac{1}{2}\{c_2 + \log|\mathbf{Q}(\boldsymbol{\nu})| + (1-\phi^2)^{-1}(\boldsymbol{\beta}_t - \phi\mathbf{F}\boldsymbol{\beta}_{t-1})'\mathbf{Q}(\boldsymbol{\nu})^{-1}(\boldsymbol{\beta}_t - \phi\mathbf{F}\boldsymbol{\beta}_{t-1})\} \\
\log \Pr(\mathbf{y}_t|\boldsymbol{\beta}_t, \boldsymbol{\nu}) &= -\frac{1}{2}\{c_3 + (\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_t)'(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_t)\},
\end{aligned} \tag{1.44}$$

where c_1, c_2, and c_3 are constants not depending on $\boldsymbol{\nu}$.

We apply Equation (1.44) to the complete data log-likelihood (Eq. 1.43) and compute its expectation with respect to $\Pr(\{\boldsymbol{\beta}_t\}_{t=0}^T|\{\mathbf{y}_t\}_{t=1}^T, \boldsymbol{\nu}^{(i-1)})$, where we have conditioned on the full set of measurements and the parameter estimate of the previous iteration [58]. We then add the logarithm of the prior density of $\boldsymbol{\nu}$ and obtain Equation (1.17).

1.7.4 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) [11]. We emphasize that in this analysis the parameters $\boldsymbol{\nu}$ 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 [11, 21], 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 [62, 73], 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 $\{\mathbf{y}_k\}_{k=1}^{t-1}$ and future $\{\mathbf{y}_k\}_{k=t+1}^{T}$ measurements is optimally accounted by the estimate; and 2) to present the formulas for the KF and FIS estimates in a way that parallel those of the well-known MNE equation as an attempt to further introduce these dynamic estimation techniques in the broad 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 $\Pr(\cdot) = \Pr(\cdot|\boldsymbol{\nu})$. We begin by recalling that the MNE assumes the probability density of the source amplitude vector $\Pr(\boldsymbol{\beta}_t)$ to be Gaussian with mean $\boldsymbol{\mu}_t = \boldsymbol{\mu}^{(\text{MNE})} = \mathbf{0}$ and covariance, or regularization matrix, $\mathbf{V}_t = \mathbf{C}^{(\text{MNE})}$. Therefore, the MNE maximizes the posterior density [11],

$$\Pr(\boldsymbol{\beta}_t|\mathbf{y}_t) \propto \underbrace{\Pr(\mathbf{y}_t|\boldsymbol{\beta}_t)}_{\text{Likelihood}} \underbrace{\Pr(\boldsymbol{\beta}_t)}_{\text{“Prior”}}, \tag{1.45}$$

where the likelihood $\Pr(\mathbf{y}_t|\boldsymbol{\beta}_t)$ is Gaussian with mean $\mathbf{X}\boldsymbol{\beta}_t$ and covariance $\mathbf{I}$ (Eq. 1.1). We must emphasize that the MNE "prior" density $\Pr(\boldsymbol{\beta}_t)$ does not contain any information about the measurements.

A similar interpretation can be given to the Kalman Filter estimate of $\boldsymbol{\beta}_t$ given data up to time t, as shown in [62, 73]. In this case, the "prior" corresponds to the conditional density of $\boldsymbol{\beta}_t$ given data up to time $t-1$, $\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1})$, which is Gaussian with mean $\boldsymbol{\mu}_t = \boldsymbol{\beta}_{t|t-1}$ and covariance $\mathbf{V}_t = \mathbf{V}_{t|t-1}$ given by the prediction step of Kalman Filter recursions (Eq. 1.21). Then the Kalman Filter computes the Maximum a Posteriori estimate of $\boldsymbol{\beta}_t$ given data up to time t by maximizing the posterior density:

$$\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t}) \propto \underbrace{\Pr(\mathbf{y}_t|\boldsymbol{\beta}_t)}_{\text{Likelihood}} \underbrace{\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1})}_{\text{"Prior"}}. \tag{1.46}$$

We highlight that the Kalman Filter "prior" density $\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1})$ contains information only from past measurements $\{\mathbf{y}_k\}_{k=1}^{t-1}$.

The Fixed Interval Smoother estimate can be interpreted similarly, where the "prior" density corresponds to the conditional density of $\boldsymbol{\beta}_t$ given previous data up to time $t-1$ and future data after time $t+1$, $\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1}, \{\mathbf{y}_k\}_{k=t+1}^{T})$. 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 $\boldsymbol{\mu}_t = \boldsymbol{\beta}_{t|T\setminus t}$ and covariance $\mathbf{V}_t = \mathbf{V}_{t|T\setminus t}$ can be obtained by standard methods with computationally costly algebraic manipulation. In this case, the notation of the subscript $t|T\setminus t$ reflects that we are conditioning on all data except the immediate data point $\mathbf{y}_t$. Therefore, the Fixed Interval Smoother estimate maximizes the posterior density of the state given all measurements:

$$\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{T}) \propto \underbrace{\Pr(\mathbf{y}_t|\boldsymbol{\beta}_t)}_{\text{Likelihood}} \underbrace{\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1}, \{\mathbf{y}_k\}_{k=t+1}^{T})}_{\text{"Prior"}}. \tag{1.47}$$

We emphasize that the Fixed Interval Smoother "prior" density $\Pr(\boldsymbol{\beta}_t|\{\mathbf{y}_k\}_{k=1}^{t-1}, \{\mathbf{y}_k\}_{k=t+1}^{T})$ includes information from both past $\{\mathbf{y}_k\}_{k=1}^{t-1}$ and future measurements $\{\mathbf{y}_k\}_{k=t+1}^{T}$.

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 Equations 1.45, 1.46, and 1.47, 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

$$\arg\min_{\boldsymbol{\beta}_t} \underbrace{||\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_t||^2}_{\text{Data fit}} + \underbrace{||\boldsymbol{\beta}_t - \boldsymbol{\mu}_t||^2_{\mathbf{V}_t^{-1}}}_{\text{Penalty}}, \tag{1.48}$$

where $\boldsymbol{\mu}_t$ and $\mathbf{V}_t$ are the "prior" mean and covariance as defined above for each method. We should note at this point that Equation (1.48) 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 Equation (1.48) is given by

$$\hat{\boldsymbol{\beta}}_t(\boldsymbol{\mu}_t, \mathbf{V}_t) = \boldsymbol{\mu}_t + \mathbf{V}_t\mathbf{X}'(\mathbf{X}\mathbf{V}_t\mathbf{X}' + \mathbf{I})^{-1}(\mathbf{y}_t - \mathbf{X}\boldsymbol{\mu}_t). \tag{1.49}$$

Now we can simply replace the corresponding "prior" mean and covariance for each method to obtain the MNE ($\boldsymbol{\beta}_t^{(\text{MNE})}$), the Kalman Filter estimate ($\boldsymbol{\beta}_{t|t}$) and the Fixed Interval Smoother estimate ($\boldsymbol{\beta}_{t|T}$):

$$\begin{aligned}
\boldsymbol{\beta}_t^{(\text{MNE})} &= 0 && + \mathbf{C}^{(\text{MNE})}\mathbf{X}'(\mathbf{X}\mathbf{C}^{(\text{MNE})}\mathbf{X}' + \mathbf{I})^{-1}(\mathbf{y}_t - 0) \\
\boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} && + \mathbf{V}_{t|t-1}\mathbf{X}'(\mathbf{X}\mathbf{V}_{t|t-1}\mathbf{X}' + \mathbf{I})^{-1} \ (\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|t-1}) \\
\boldsymbol{\beta}_{t|T} &= \underbrace{\boldsymbol{\beta}_{t|T\setminus t}}_{\text{Prior mean}} && + \underbrace{\mathbf{V}_{t|T\setminus t}\mathbf{X}'(\mathbf{X}\mathbf{V}_{t|T\setminus t}\mathbf{X}' + \mathbf{I})^{-1}}_{\text{Gain}} \ \underbrace{(\mathbf{y}_t - \mathbf{X}\boldsymbol{\beta}_{t|T\setminus t})}_{\text{Residual}}.
\end{aligned} \tag{1.50}$$

Equation (1.50) 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. 1.22), 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 $\boldsymbol{\beta}_t$ using the respective "prior" mean, and then updates this prediction using the measurement at the same time point $\mathbf{y}_t$ by computing a 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 $\{\mathbf{y}_k\}_{k=1}^{t-1}$. Finally, the Fixed Interval Smoother goes further to build a prediction based on previous data $\{\mathbf{y}_k\}_{k=1}^{t-1}$ as well as future data $\{\mathbf{y}_k\}_{k=t+1}^{T}$. 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 $\mathbf{y}_t$ by adding a term that is proportional to the residuals.

1.7.5 Computation of ROC curves

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

$$\begin{aligned}\hat{\text{pr}}_D(c) =& \frac{\sum_{j=1}^{p}\sum_{t=1}^{T}\text{indic}(|\hat{\beta}_{j,t}| > c)\cdot\text{indic}(\beta_{j,t}^{(\text{SIM})} \neq 0)}{\sum_{j=1}^{p}\sum_{t=1}^{T}\text{indic}(\beta_{j,t}^{(\text{SIM})} \neq 0)} \\ \hat{\text{pr}}_{FA}(c) =& \frac{\sum_{j=1}^{p}\sum_{t=1}^{T}\text{indic}(|\hat{\beta}_{j,t}| > c)\cdot\text{indic}(\beta_{j,t}^{(\text{SIM})} = 0)}{\sum_{j=1}^{p}\sum_{t=1}^{T}\text{indic}(\beta_{j,t}^{(\text{SIM})} = 0)},\end{aligned} \tag{1.51}$$

where indic$(\cdot)$ in Equation (1.51) is the indicator function. We should note that $p \approx 5000$ represents the number of dipole sources and $T = 200$ is the number of time samples.

Chapter 2

An Analysis of How Spatiotemporal Dynamic Models of Brain Activity Could Dramatically Improve MEG/EEG Inverse Solutions

2.1 Abstract

MEG and EEG are noninvasive functional neuroimaging techniques that provide recordings of brain activity with high temporal resolution, and thus provide a unique window to study fast time-scale neural dynamics in humans. However, the accuracy of brain activity estimates resulting from these data is limited since 1) the number of sensors is much smaller than the number of sources, and 2) the low sensitivity of the recording device to deep or radially oriented sources. These factors limit the number of sources that can be recovered and bias source estimates to superficial cortical areas, resulting in the need to include a priori information about the source activity. The question of how to specify this information and how it might lead to improved solutions remains a critical open problem. In this paper we show that the incorporation of knowledge about the brain's underlying connectivity and spatiotemporal

dynamics could dramatically improve inverse solutions. To do this, we develop the concept of the *dynamic lead field mapping*, which expresses how information about source activity at a given time is mapped not only to the immediate measurement, but to a time series of measurements. With this mapping we show that the number of source parameters that can be recovered could increase by up to a factor of ~ 20, and that such improvement is primarily represented by deep cortical areas. Our result implies that future developments in MEG/EEG analysis that model spatialtemporal dynamics have the potential to dramatically increase source resolution.

2.2 Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) are functional neuroimaging tools that provide noninvasive recordings of the magnetic and electric fields at the scalp generated by neuronal currents. Because MEG and EEG have a high sampling rate in the order of kHz, they hold particular promise as tools to noninvasively study fast time-scale neural dynamics of human brain function in health and disease [1]. This makes MEG/EEG unique amongst other functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), or diffuse optical tomography (DOT), which instead provide indirect measures of brain activity related to slower neurovascular changes. However, to properly interpret MEG and EEG recordings and fully realize their potential, one needs to estimate the source current activity underlying the measured electric potentials and magnetic fields at the scalp surface, i.e., to solve the ill-posed MEG/EEG inverse problem. Although significant progress has been made on this problem in the past few decades [11, 15, 74, 21, 23, 27, 31, 32, 75, 34, 35, 76] at present MEG/EEG inverse solutions are viewed as rough estimates characterized by poor spatial resolution (in the order of a few cm), and insensitive the majority of deep cortical and subcortical regions.

Two main factors make it difficult to obtain accurate sources estimates. These can be analyzed if we consider the *lead field matrix* $\mathbf{X}$, which maps the cortical activity

of a few thousand dipole sources, $\boldsymbol{\beta}_t$, to the recordings in a couple hundred scalp sensors, $\mathbf{y}_t$, at an individual time instant t [1, 77]:

$$\mathbf{y}_t = \mathbf{X}\boldsymbol{\beta}_t + \textit{“noise”}.$$

The first factor is due to the dimensionality in MEG/EEG. Since the number of sources to be estimated is an order of magnitude larger than the number of sensors, different source configurations can produce identical scalp recordings, making solutions to this inverse problem non-unique. This issue is exacerbated by the fact that adjacent sources can produce very similar recordings. As a result, the number of sources that we can expect to recover from measurements in an individual time point t, which is determined by the rank of $\mathbf{X}$, is much smaller than the number of sources. The second factor is due to the biophysics of MEG/EEG. This is because the amplitude of the scalp fields and potentials rapidly decays with the square of the distance from the sensor to the source [1]. In addition, radially-oriented sources can be magnetically silent [78]. Therefore, the sensitivity of the recording device—the signal power generated by an individual active source across all sensors given by the norm of the columns of $\mathbf{X}$—is very low for a large percentage of cortical areas. These two factors limit the number of source parameters that can be recovered and effectively restrict estimates to cortical areas whose activity is most easily detected by the sensors, ultimately imposing the need to include a priori information about the source currents. The question of how to specify this a priori information and how it might lead to improved solutions is one of the most important problems in bioelectromagnetism and neuroimaging.

Prior information about the current dipole distribution has been used in the MEG/EEG inverse problem literature in order to obtain unique estimates. This information has taken the form of a probabilistic model or optimization penalty that implicitly or explicitly assumes cortical activity is either independent across time [11, 15, 74, 21], temporally or spatio-temporally smooth [23, 27, 31, 32, 75], or follow a linear dynamic process [34, 35, 76]. While these priors alleviate issues related

to the non-uniqueness of source estimates, they do not necessarily improve on the limitations that stem from the rank deficiency and restricted sensitivity of the lead field matrix. However, one way to ameliorate these issues could be achieved by incorporating prior knowledge about the spatiotemporal physiological relations present in brain activity [36, 37, 38, 39]. This is because in such a spatiotemporal system, information about source activation at a given time instant t is contained not only in the immediate measurement $\mathbf{y}_t$ but also in multiple observations over a time interval $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$. As a result of this dynamic flow of information, we can imagine a mapping that relates brain activity at a given moment in time to the measurements in the complete analysis interval—a *dynamic* lead field mapping. The rank and sensitivity of the dynamic lead field mapping could be substantially better than those of the static lead field matrix $\mathbf{X}$, not only because of its increased dimensionality but also because of how information can flow across structural connections from brain areas that are harder to detect at the scalp to those that are easier to detect.

Inspired by electrophysiology and neuroanatomy studies [36, 37, 38, 39], in this paper we show how the incorporation of information about the source spatial connectivity and temporal dynamics could dramatically increase both the number of sources that can be recovered, as well as the sensitivity for detecting such sources. To do this, we develop the concept of the *dynamic lead field mapping*, a dynamic extension of the lead field matrix that allows us to analyze the rank and sensitivity of the mapping between source activity at a given time instant to the measurements both forward and backwards in time over the complete analysis interval. This dynamic lead field mapping, though developed for the specific problem of MEG/EEG source imaging, is firmly grounded in dynamic systems theory. With this mapping we show that the number of sources that can be effectively recovered increases by up to a factor of ~ 20 by modeling the most basic local cortical dynamic connections. Furthermore, we show that the inclusion of such local cortical connections increases the sensitivity for detecting sources distributed across the brain, and that the increase in sensitivity is more pronounced in sources located within sulci and other deeper areas. At the core of our technical development is a projection operation that allows us to analyze

the rank and sensitivity resulting from different prior source models. In particular, we analyze a space-time separable model, as well as a static model that assumes temporal independence. We find that improvements in the number and sensitivity of sources that can be recovered occur only when dynamic spatiotemporal connections are modeled. Our results imply that future developments in MEG/EEG analysis that explicitly model dynamic connections between brain areas have the potential to dramatically increase spatiotemporal resolution by taking full advantage of the brain's inherent spatiotemporal structure, connectivity, and dynamics.

2.3 Methods

2.3.1 The spatiotemporal dynamic source model and the MEG/EEG measurement model

Converging lines of evidence suggest that brain activity is a spatiotemporal dynamic process with structure varying at different scales. At the smaller spatial scale, intracranial recordings in different species have revealed that cortical activity exhibits strong correlations that persist up to a distance of 10 mm during rest and task periods [36, 37, 38]. These local cortico-cortical dynamic interactions can be supported neuroanatomically by the fact that axonal collateral projections from pyramidal cells spread laterally approximately 6 mm along the cortical surface [39]. In addition, long-distance correlations can exist by means of white matter tracts that connect distant brain regions, like those which are thought to support large-scale brain network activity such as the resting-state networks studied via fMRI [79, 80, 81].

In order to provide a conservative analysis where spatiotemporal dynamics consistent with neurophysiological evidence are considered, we choose to focus on a parsimonious model that incorporates local spatiotemporal interactions. One way to model local spatiotemporal connections of this type is to use a first order linear dynamic process. In this model, neuronal currents at a given point in time t and spatial location i, $\beta_{i,t}$, are a function of past neuronal currents at the same location, $\beta_{i,t-1}$,

as well as past currents, $\beta_{j,t-1}$, at locations j within a local neighborhood $\mathcal{N}(i)$:

$$\beta_{i,t} = \phi[f_{i,i}\beta_{i,t-1} + \sum_{j\in\mathcal{N}(i)} f_{i,j}\beta_{j,t-1}] + (1-\phi^2)^{1/2}\omega_{i,t}. \tag{2.1}$$

In Equation (2.1), the parameter ϕ ($0 \leq \phi < 1$) represents the strength of the history dependance in the dynamics. The weighting factors $f_{i,j}$ represent the interaction between sources at locations i and j. At each location i, the weights $f_{i,j}$ corresponding to its neighbors $j \in \mathcal{N}(i)$ are assumed to be positive and inversely proportional to the distance between sources. In addition, they are normalized such that the contribution of the neighbors to the dynamics of the ith source equals its self contribution, while the total contribution is equal to one: $\sum_{j\in\mathcal{N}(i)} f_{i,j} = f_{i,i}$ and $\sum_{j\in\mathcal{N}(i)} f_{i,j} + f_{i,i} = 1$. Furthermore, the input process $\omega_{i,t}$ is assumed to be Gaussian with zero-mean and independent across time and space. The spatiotemporal model in Equation (2.1), which has been previously used in dynamic source localization analysis [34, 76], can be readily expressed in vector form as:

$$\boldsymbol{\beta}_t = \phi\mathbf{F}\boldsymbol{\beta}_{t-1} + (1-\phi^2)^{1/2}\boldsymbol{\omega}_t, \tag{2.2}$$

where $\boldsymbol{\beta}_t$ is a vector of dimension p ($\sim 10^3$), and the input process $\boldsymbol{\omega}_t$ is Gaussian with zero mean and independent across time with spatial covariance matrix $\mathbf{Q}$, which we assume diagonal and positive definite. We should note that since the matrix $\phi\mathbf{F}$ is stable and $(1-\phi^2)\mathbf{Q}$ is positive definite, we can assume that the process $\boldsymbol{\beta}_t$ has reached a steady-state where its spatial covariance matrix $\mathbf{C} = \mathrm{Cov}(\boldsymbol{\beta}_t)$ is time invariant and invertible[1]. Without loss of generality, we can parametrize the covariance matrix $\mathbf{Q} = [\lambda \mathrm{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}\mathrm{diag}(\nu_1, \nu_2, \ldots, \nu_p)$, where $\widehat{\boldsymbol{\Sigma}} = \mathbf{X}'\mathbf{X}/n$ is the sample covariance of the rows of $\mathbf{X}$, $\lambda > 0$, and $\nu_i > 0$. Doing so would allow us to interpret λ as

[1]The matrix $\phi\mathbf{F}$ is stable, i.e., the magnitude of its largest eigenvalue is equal to $\phi < 1$, since $\mathbf{F}$ is a stochastic matrix (it has nonnegative entries with the sum of its rows equal to 1) and because the largest eigenvalue of a stochastic matrix is 1. The matrix $(1-\phi^2)\mathbf{Q}$ is positive definite because it is diagonal with positive elements in its diagonal. Since $\phi\mathbf{F}$ is stable, the process covariance reaches a steady-state, $\mathbf{C} = \lim_{t\to\infty}\mathrm{Cov}(\boldsymbol{\beta}_t)$, which is given by the solution of the Lyapunov equation $\mathbf{C} = \phi^2\mathbf{F}\mathbf{C}\mathbf{F}' + (1-\phi^2)\mathbf{Q}$. Furthermore, since $(1-\phi^2)\mathbf{Q}$ is positive definite and invertible, so it is the source covariance $\mathbf{C}$ [55].

the inverse of the power signal-to-noise ratio (SNR^2) [2].

In an MEG/EEG experiment, we obtain a recording of the magnetic field and electric potential from hundreds of sensors located on or above the scalp at times $t \in [1, 2, \ldots, T]$. At time t, the resulting data vector $\mathbf{y}_t$, which is of dimension n ($\sim 10^2$), is related to the source vector $\boldsymbol{\beta}_t$ by the observation equation [1, 77]:

$$\mathbf{y}_t = \underbrace{\mathbf{X}\boldsymbol{\beta}_t}_{\text{Signal}} + \underbrace{\boldsymbol{\varepsilon}_t}_{\text{Noise}}, \tag{2.3}$$

where $\mathbf{X}$ is the $n \times p$ lead field matrix computed using a quasistatic approximation of the Maxwell's equations, and $\boldsymbol{\varepsilon}_t$ is the Gaussian white noise vector with zero mean and spatial covariance equal to the identity matrix $\mathbf{I}$. We should emphasize that in our notation we assume the observation model (Eq. 2.3) has been spatially whitened, i.e., that the original raw data model $\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}\boldsymbol{\beta}_t + \tilde{\boldsymbol{\varepsilon}}_t$ has been premutiplied by the inverse of a matrix square root of the covariance of $\tilde{\boldsymbol{\varepsilon}}_t$. In Equation (2.3), $\mathbf{X}\boldsymbol{\beta}_t$ represents the *signal* portion of the model, and the *noise* term $\boldsymbol{\varepsilon}_t$ is independent from $\boldsymbol{\beta}_t$ for all time points.

2.3.2 The dynamic lead field mapping

In the observation model (Eq. 2.3), the signals generated by brain activity are represented by the product of the lead field matrix and the source vector: $\mathbf{X}\boldsymbol{\beta}_t$. Since the resulting product and the source vector are independent of the noise term $\boldsymbol{\varepsilon}_t$, the lead field matrix $\mathbf{X}$ contains all the information related to the mapping of the brain source vector $\boldsymbol{\beta}_t$ at a particular point in time t to the measurement $\mathbf{y}_t$ at that same point in time t (Figure 2-1A). From this static point of view, it is clear that the maximum number of independent variables that could be determined from a single measurement in time is limited by the rank of the $\mathbf{X}$ matrix, which unfortunately is less than or equal to the number of sensors, and much smaller than the number of sources [82]. However, if we consider the brain's source activity as a spatiotemporal

[2] Roughly speaking, if the matrix $\mathbf{F} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and $\nu_i = 1$ $(i = 1, \ldots, p)$, the steady state covariance of $\boldsymbol{\beta}_\mathbf{t}$ becomes $\mathbf{C} = [\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}\mathbf{I}$. If we define the power signal-to-noise ratio as $\text{SNR}^2 = \text{E}||\mathbf{X}\boldsymbol{\beta}_t||^2/\text{E}||\boldsymbol{\varepsilon}_t||^2$, then $\text{SNR}^2 = \text{tr}(\mathbf{X}'\mathbf{X})[\lambda \text{tr}(\widehat{\boldsymbol{\Sigma}})/n]^{-1}/n = 1/\lambda$.

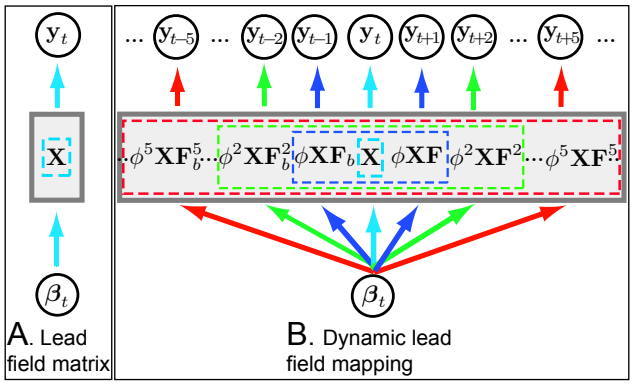

Figure 2-1: **The dynamic lead field mapping** A: The (static) lead field matrix $\mathbf{X}$ determines how information from the source vector $\boldsymbol{\beta}_t$ propagates to the measurement $\mathbf{y}_t$ at the same point in time t. B: When spatiotemporal dynamics are modeled, the dynamic lead field mapping $\mathbf{D}_t$ determines how information from the source vector $\boldsymbol{\beta}_t$ at a given time t is mapped to the complete time series of measurements $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$.

dynamic process, we can dramatically improve on these limitations.

In a spatiotemporal system, the information about the source vector $\boldsymbol{\beta}_t$ at a particular time t is contained not only in the immediate observation $\mathbf{y}_t$ but also in the previous $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}]$ and future $[\mathbf{y}_{t+1}, \mathbf{y}_{t+2}, \ldots, \mathbf{y}_T]$ observations. Because of this, information from the source vector $\boldsymbol{\beta}_t$ is effectively mapped to the complete time series of measurements through a function (Figure 2-1B), which we call the *dynamic lead field mapping*, whose rank and sensitivity would be greater than those of the static lead field matrix $\mathbf{X}$.

To obtain the dynamic lead field mapping, we must derive an observation model for the complete set of measurements $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ where: i) the *signal* portion of the model is a function of only the source vector $\boldsymbol{\beta}_t$ at a particular time t, and ii) the *noise* term is independent of the source vector $\boldsymbol{\beta}_t$. As we will show below, the signal function for this observation model is indeed linear. Because of this, the complete observation model can be expressed as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \underbrace{\mathbf{D}_t \boldsymbol{\beta}_t}_{\text{Signal}} + \underbrace{\mathbf{n}_t}_{\text{Noise}}, \tag{2.4}$$

where the matrix $\mathbf{D}_t$ of dimension $nT \times p$ is the *dynamic lead field mapping*, and the noise term $\mathbf{n}_t$ of dimension nT is independent of the source vector $\boldsymbol{\beta}_t$. In analogy to the lead field matrix $\mathbf{X}$ in the *static* observation model (Eq. 2.3), the *dynamic* lead field mapping $\mathbf{D}_t$ can be used to determine the number of independent variables we can recover as well as the sensitivity for detecting such sources when we consider the complete time series of observations in a spatiotemporal dynamic framework.

In order to derive $\mathbf{D}_t$ in Equation (2.4), we consider three separate cases depending on whether the observation vector $\mathbf{y}_{t\pm k}$ $(k = 0, 1, \ldots)$ corresponds to a future, present, or past observation with respect to the source vector $\boldsymbol{\beta}_t$ a the present time t. Because of this partition we will consider the dynamic lead field mapping $\mathbf{D}_t$ as a block matrix composed of submatrices $\mathbf{P}_{t\pm k,t}$:

$$\mathbf{D}_t = \begin{bmatrix} \mathbf{P}_{1,t} \\ \mathbf{P}_{2,t} \\ \vdots \\ \mathbf{P}_{T,t} \end{bmatrix},$$

where we call the blocks $\mathbf{P}_{t\pm k,t}$ the projection matrices.

In the case where the measurement vector corresponds to the present observation $(k = 0)$, we use the static observation model (Eq. 2.3) to obtain the relation between $\boldsymbol{\beta}_t$ and $\mathbf{y}_t$ in the block number t of Equation (2.4):

$$\mathbf{y}_t = \underbrace{\mathbf{X}\boldsymbol{\beta}_t}_{\text{Signal}} + \underbrace{\boldsymbol{\varepsilon}_t}_{\text{Noise}}.$$

From the equation above we can identify that $\mathbf{P}_{t,t} = \mathbf{X}$.

To obtain the projection matrices corresponding to future observations, we start

by using the recursion defining the source vector spatiotemporal dynamics (Eq. 2.2) and iterate it $k \in [1, 2, \ldots]$ times to obtain:

$$\begin{aligned}
\boldsymbol{\beta}_{t+1} &= \phi \mathbf{F} \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \boldsymbol{\omega}_{t+1} \\
\boldsymbol{\beta}_{t+2} &= \phi^2 \mathbf{F}^2 \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \phi \mathbf{F} \boldsymbol{\omega}_{t+1} + (1-\phi^2)^{1/2} \boldsymbol{\omega}_{t+1} \\
&\vdots \\
\boldsymbol{\beta}_{t+k} &= \phi^k \mathbf{F}^k \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{F}^{k-j} \boldsymbol{\omega}_{t+j}.
\end{aligned} \tag{2.5}$$

Using this k-step iteration into the future along with the static measurement model (Eq. 2.3), we obtain the relation between $\boldsymbol{\beta}_t$ and the future measurement $\mathbf{y}_{t+k}$:

$$\mathbf{y}_{t+k} = \underbrace{\phi^k \mathbf{X} \mathbf{F}^k \boldsymbol{\beta}_t}_{\text{Signal}} + \underbrace{(1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X} \mathbf{F}^{k-j} \boldsymbol{\omega}_{t+j} + \boldsymbol{\varepsilon}_{t+k}}_{\text{Noise}}. \tag{2.6}$$

We should note that Equation (2.6) represents the $(t+k)$th block of Equation (2.4), and therefore we identify that $\mathbf{P}_{t+k,t} = \phi^k \mathbf{X} \mathbf{F}^k$.

To derive the past projection matrices we make use of an equivalent representation of the source dynamics (Eq. 2.2) known as the *backwards Markovian* model [55]:

$$\boldsymbol{\beta}_t = \phi \mathbf{F}_b \boldsymbol{\beta}_{t+1} + (1-\phi^2)^{1/2} \boldsymbol{\omega}_t^b. \tag{2.7}$$

In this representation, where the state vector $\boldsymbol{\beta}_t$ evolves backwards in time, the time-reversed transition matrix is given by $\mathbf{F}_b = \mathbf{C}\mathbf{F}'\mathbf{C}^{-1}$. The backwards input process $\boldsymbol{\omega}_t^b$, which is independent across time and Gaussian, has zero mean and covariance matrix $\mathbf{Q}_b = (1-\phi^2)^{-1}(\mathbf{C} - \phi^2 \mathbf{F}_b \mathbf{C} \mathbf{F}_b')$. With this equivalent backwards dynamic representation (Eq. 2.7), we proceed to obtain the relation between $\boldsymbol{\beta}_t$ and the previous measurement $\mathbf{y}_{t-k}$ just as we did when we considered future measurement. In this case the $k \in [1, 2, \ldots]$ step backwards iteration is given by:

$$
\begin{aligned}
\boldsymbol{\beta}_{t-1} =& \phi \mathbf{F}_b \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \boldsymbol{\omega}_{t-1}^b \\
\boldsymbol{\beta}_{t-2} =& \phi^2 \mathbf{F}_b^2 \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \phi \mathbf{F}_b \boldsymbol{\omega}_{t-1}^b + (1-\phi^2)^{1/2} \boldsymbol{\omega}_{t-2}^b \\
& \vdots \\
\boldsymbol{\beta}_{t-k} =& \phi^k \mathbf{F}_b^k \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{F}_b^{k-j} \boldsymbol{\omega}_{t-j}^b .
\end{aligned} \tag{2.8}
$$

We use the static observation model (Eq. 2.3) along with the k-step past iteration to obtain the $(t-k)$th block of Equation (2.4):

$$
\mathbf{y}_{t-k} = \underbrace{\phi^k \mathbf{X} \mathbf{F}_b^k \boldsymbol{\beta}_t}_{\text{Signal}} + \underbrace{(1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X} \mathbf{F}_b^{k-j} \boldsymbol{\omega}_{t-j}^b + \boldsymbol{\varepsilon}_{t-k}}_{\text{Noise}}, \tag{2.9}
$$

where we see that $\mathbf{P}_{t-k,t} = \phi^k \mathbf{X} \mathbf{F}_b^k$.

At this point we are ready to explicitly write the complete observation model (Eq. 2.4) as well as the dynamic lead field mapping $\mathbf{D}_t$. For a particular time t, we use the equations mapping the source vector $\boldsymbol{\beta}_t$ to the present (Eq. 2.3), future (Eq. 2.6), and past (Eq. 2.9) to obtain the complete observation model:

$$
\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{t-1} \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \underbrace{\begin{bmatrix} \phi^{t-1} \mathbf{X} \mathbf{F}_b^{t-1} \\ \vdots \\ \phi \mathbf{X} \mathbf{F}_b \\ \mathbf{X} \\ \phi \mathbf{X} \mathbf{F} \\ \vdots \\ \phi^{T-t} \mathbf{X} \mathbf{F}^{T-t} \end{bmatrix}}_{\mathbf{D}_t} \boldsymbol{\beta}_t + \underbrace{\begin{bmatrix} \mathbf{e}_{1,t} \\ \vdots \\ \mathbf{e}_{t-1,t} \\ \mathbf{e}_{t,t} \\ \mathbf{e}_{t+1,t} \\ \vdots \\ \mathbf{e}_{T,t} \end{bmatrix}}_{\mathbf{n}_t}, \tag{2.10}
$$

where error terms are given by,

$$\mathbf{e}_{t\pm k,t} = \begin{cases} (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X}\mathbf{F}_b^{k-j} \boldsymbol{\omega}_{t-j}^b + \boldsymbol{\varepsilon}_{t-k}, & -k > 0 \\ \boldsymbol{\varepsilon}_t, & k = 0 \\ (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X}\mathbf{F}^{k-j} \boldsymbol{\omega}_{t+j} + \boldsymbol{\varepsilon}_{t+k}, & +k > 0 \end{cases} .$$

From the complete observation model (Eq. 2.10) we can see that the dynamic lead field mapping is:

$$\mathbf{D}_t = \left[\phi^{t-1} \mathbf{F}_b^{t-1'} \mathbf{X}', \ldots, \phi \mathbf{F}_b' \mathbf{X}', \mathbf{X}', \phi \mathbf{F}' \mathbf{X}', \ldots, \phi^{T-t} \mathbf{F}^{T-t'} \mathbf{X}' \right]' . \tag{2.11}$$

In addition, we can observe that Equation (2.10) expresses the complete set of measurements $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ as the sum of the *signal* function $\mathbf{D}_t \boldsymbol{\beta}_t$ and *noise* term $\mathbf{n}_t = [\mathbf{e}'_{1,t}, \mathbf{e}'_{2,t}, \ldots, \mathbf{e}'_{T,t}]'$. We should note that the noise term $\mathbf{n}_t$ is indeed independent of the source vector $\boldsymbol{\beta}_t$, since the vectors $\mathbf{e}_{t\pm k,t}$ correspond to the projection errors that result from projecting the measurements $\mathbf{y}_{t\pm k}$ onto the source vector $\boldsymbol{\beta}_t$. Such projections are obtained via the matrices $\mathbf{P}_{t\pm k,t}$ (For details see Section 2.6.3 in Supplemental materials). Consequently, the *dynamic* lead field mapping $\mathbf{D}_t$ contains all the information describing how the brain source vector $\boldsymbol{\beta}_t$, at a particular point in time t, propagates to the complete set of measurements. Therefore, we can use $\mathbf{D}_t$ to determine the number of independent variables we can recover, as well as the sensitivity for detecting such sources, from the complete time series of observations in a spatiotemporal dynamic framework.

2.3.3 Extension to general Gaussian source models

The dynamic lead field ($\mathbf{D}_t$) allows us to analyze the mapping from a source vector $\boldsymbol{\beta}_t$ to the complete time series of measurements $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ when considering that the prior source model follows the spatiotemporal dynamics in Equation (2.1). While our dynamic source model is inspired by neurophysiological evidence and neuroanatomy [36, 37, 38, 39], it is natural to consider other mappings resulting from more general Gaussian source models. Doing so would allow us to place our dynamic

lead field analysis within a more general framework, and to compare the properties of different source models in MEG/EEG imaging.

In the models we now consider, the joint distribution of the sources has a zero mean and is Gaussian, just like in our spatiotemporal dynamic source model (Eqs. 2.1 and 2.2), but the spatiotemporal covariance structure differs from that in our model. Specifically, under a model $\mathfrak{m}$, the joint distribution of the source vectors $[\boldsymbol{\beta}_1^{(\mathfrak{m})}, \boldsymbol{\beta}_2^{(\mathfrak{m})}, \ldots, \boldsymbol{\beta}_T^{(\mathfrak{m})}]$ is Gaussian with zero mean ($\mathrm{E}[\boldsymbol{\beta}_t^{(\mathfrak{m})}] = \mathbf{0}$) and arbitrary cross-covariances $\mathrm{E}[\boldsymbol{\beta}_k^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})'}]$, where E denotes the expectation operator. The MEG/EEG measurement are obtained via the static observation model (Eq. 2.3):

$$\mathbf{y}_t^{(\mathfrak{m})} = \mathbf{X}\boldsymbol{\beta}_t^{(\mathfrak{m})} + \boldsymbol{\varepsilon}_t. \tag{2.12}$$

Just as we did in the case of the dynamic lead field mapping, when considering a model $\mathfrak{m}$ we would like to express the complete time series of measurements $[\mathbf{y}_1^{(\mathfrak{m})}, \mathbf{y}_2^{(\mathfrak{m})}, \ldots, \mathbf{y}_T^{(\mathfrak{m})}]$ as the sum of a function that only depends on the source vector $\boldsymbol{\beta}_t^{(\mathfrak{m})}$ at the present time t and a noise term that is independent of $\boldsymbol{\beta}_t^{(\mathfrak{m})}$. As we will show below, this can be done via the linear relation,

$$\begin{bmatrix} \mathbf{y}_1^{(\mathfrak{m})} \\ \mathbf{y}_2^{(\mathfrak{m})} \\ \vdots \\ \mathbf{y}_T^{(\mathfrak{m})} \end{bmatrix} = \underbrace{\mathbf{D}_t^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})}}_{\text{Signal}} + \underbrace{\mathbf{n}_t^{(\mathfrak{m})}}_{\text{Noise}}, \tag{2.13}$$

where $\mathbf{D}_t^{(\mathfrak{m})}$ is the equivalent of the dynamic lead field mapping for the model $\mathfrak{m}$, and the noise term $\mathbf{n}_t^{(\mathfrak{m})}$ is independent of the source vector $\boldsymbol{\beta}_t^{(\mathfrak{m})}$.

To obtain Equation (2.13) we again consider the matrix $\mathbf{D}_t^{(\mathfrak{m})}$ as composed of sub-matrix blocks $\mathbf{P}_{t\pm k,t}^{(\mathfrak{m})}$:

$$\mathbf{D}_t^{(m)} = \begin{bmatrix} \mathbf{P}_{1,t}^{(\mathfrak{m})} \\ \mathbf{P}_{2,t}^{(\mathfrak{m})} \\ \vdots \\ \mathbf{P}_{T,t}^{(\mathfrak{m})} \end{bmatrix}.$$

To obtain the $t \pm k$th block of Equation (2.13) we make use of the fact that, under a model $\mathfrak{m}$, the measurement vector $\mathbf{y}_{t\pm k}^{(\mathfrak{m})}$ at time $t \pm k$ can be expressed as the sum of its projection onto the source vector $\boldsymbol{\beta}_t^{(\mathfrak{m})}$ at time t and the projection error:

$$\mathbf{y}_{t\pm k}^{(\mathfrak{m})} = \mathbf{P}_{t\pm k,t}^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})} + \mathbf{e}_{t\pm k,t}^{(\mathfrak{m})},$$

where the projection error $\mathbf{e}_{t\pm k,t}^{(\mathfrak{m})}$ is independent of the source vector at time t, and the projection matrix $\mathbf{P}_{t\pm k,t}^{(\mathfrak{m})}$ is the solution of the orthogonality equation[3]:

$$\mathrm{E}\left[(\mathbf{y}_{t\pm k}^{(\mathfrak{m})} - \mathbf{P}_{t\pm k,t}^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})}) \boldsymbol{\beta}_t^{(\mathfrak{m})'}\right] = \mathbf{0}. \tag{2.14}$$

By making use of the static observation model (Eq. 2.12), it is easy to see that the projection matrices that solve the orthogonality equation (Eq. 2.14) under a given source model $\mathfrak{m}$ are given by:

$$\mathbf{P}_{t\pm k,t}^{(\mathfrak{m})} = \mathbf{X} \left(\mathrm{E}[\boldsymbol{\beta}_{t\pm k}^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})'}]\right) \left(\mathrm{E}[\boldsymbol{\beta}_t^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})'}]\right)^{-1}. \tag{2.15}$$

Therefore, for a given Gaussian model $\mathfrak{m}$ with a specific spatiotemporal source covariance structure, we can compute the corresponding projection matrices $\mathbf{P}_{t\pm k,t}^{(\mathfrak{m})}$ and, just as we did in the development of the dynamic lead field mapping (Eqs. 2.4 and 2.10), obtain an equivalent observation model for the complete series of measurements by stacking up the projection matrices.

To give concrete examples of the $\mathbf{D}_t^{(\mathfrak{m})}$ matrix for general Gaussian source models, we will analyze two models commonly used in the source localization literature: i) the first, which we denote as the *IND* model, assumes the source vectors are independent across time [11], i.e., $\mathrm{E}[\boldsymbol{\beta}_{t\pm k}^{(ind)} \boldsymbol{\beta}_t^{(ind)'}] = \mathbf{0}$ for $k \neq 0$; and ii) a space-time separable (*STS*) model in which the joint source covariance factors via the Kronecker product $\boldsymbol{\Gamma} \otimes \mathbf{C}$ into a purely spatial covariance matrix $\mathbf{C}$ and a purely temporal covariance

[3] The projection error $\mathbf{e}_{t\pm k,t}^{(\mathfrak{m})} = \mathbf{y}_{t\pm k}^{(\mathfrak{m})} - \mathbf{P}_{t\pm k,t}^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})}$ is uncorrelated with $\boldsymbol{\beta}_t^{(\mathfrak{m})}$, namely $\mathrm{E}[(\mathbf{y}_{t\pm k}^{(\mathfrak{m})} - \mathbf{P}_{t\pm k,t}^{(\mathfrak{m})} \boldsymbol{\beta}_t^{(\mathfrak{m})}) \boldsymbol{\beta}_t^{(\mathfrak{m})'}] = \mathbf{0}$ [55]. Since $\mathbf{y}_{t\pm k}^{(\mathfrak{m})}$ and $\boldsymbol{\beta}_t^{(\mathfrak{m})}$ are jointly Gaussian, the projection error $\mathbf{e}_{t\pm k,t}^{(\mathfrak{m})}$ and $\boldsymbol{\beta}_t^{(\mathfrak{m})}$ are also jointly Gaussian. Therefore, the projection error and the source vector are uncorrelated and jointly Gaussian, and thus are independent.

matrix $\mathbf{\Gamma}$ [27]. We should note that in the *STS* model, the covariance between source vectors at times $t \pm k$ and t is given by $\mathrm{E}[\boldsymbol{\beta}_{t\pm k}^{(sts)} \boldsymbol{\beta}_t^{(sts)'}] = \gamma_{t\pm k,t}\mathbf{C}$, where $\gamma_{k,t}$ is the element of the temporal covariance matrix $\mathbf{\Gamma}$ at position $t \pm k, t$. With the covariance structure defined for these two models, and using Equation (2.15), we can see that the projection matrices for the *IND* and *STS* models are

$$\mathbf{P}_{t\pm k,t}^{(ind)} = \begin{cases} \mathbf{X} & k = 0 \\ \mathbf{0} & k \neq 0 \end{cases} \quad \text{and} \quad \mathbf{P}_{t\pm k,t}^{(sts)} = \frac{\gamma_{t\pm k,t}}{\gamma_{t,t}}\mathbf{X},$$

respectively. Therefore, the $\mathbf{D}_t^{(\mathfrak{m})}$ matrix for the *IND* and *STS* models, which constitute the equivalent of the $\mathbf{D}_t$ matrix for these Gaussian models, are given by:

$$\mathbf{D}_t^{(ind)} = \mathbf{1}_t \otimes \mathbf{X} \quad \text{and} \quad \mathbf{D}_t^{(sts)} = \frac{1}{\gamma_{t,t}}\boldsymbol{\gamma}_t \otimes \mathbf{X} \tag{2.16}$$

respectively, where $\mathbf{1}_t$ is the unit vector with the tth entry set equal to one, and $\boldsymbol{\gamma}_t$ is the tth column of $\mathbf{\Gamma}$

2.4 Results

2.4.1 Rank and singular value spectrum of the dynamic lead field mapping $\mathbf{D}_t$

We used the dynamic lead field mapping construct in conjunction with an MRI-based MEG forward model from a human subject (see Section 2.6.2 in Supplemental materials) to estimate the number of independent sources we can recover in a model that includes local spatiotemporal cortical dynamics (Eq. 2.1). In order to avoid the computational challenges associated with finding the rank and singular values of the very large $\mathbf{D}_t$ matrix (Eq. 2.11), we chose to analyze the truncated versions of $\mathbf{D}_t$ that correspond to a model that includes $k \in [1, 2, 5, 10, 20]$ measurements into the past and future of a given time t, $[\mathbf{y}_{t-k}, \ldots, \mathbf{y}_t, \ldots, \mathbf{y}_{t+k}]$. The resulting mappings, which we denote by $\mathbf{D}_t(k)$, are given by:

$$\mathbf{D}_t(k) = \left[\phi^k \mathbf{F}_b^{k'} \mathbf{X}', \ldots, \phi \mathbf{F}_b' \mathbf{X}', \mathbf{X}', \phi \mathbf{F}' \mathbf{X}', \ldots \phi^k \mathbf{F}^{k'} \mathbf{X}'\right]'. \tag{2.17}$$

This arrangement was chosen to compare the relative contributions of increasing numbers of observations to the number of independent sources that can be effectively recovered. Furthermore, since the matrices $\mathbf{D}_t(k)$ correspond to a truncated version of the much larger matrix $\mathbf{D}_t$, our computations serve as a lower bound on the number of sources that could be recovered from the time series of measurements in the complete interval $[1, 2, \ldots, T]$.

Figure 2-2A and Table 2.1 show the evolution of the singular value spectrum and the rank of the $\mathbf{D}_t(k)$ matrix, respectively, as the number of observed data points $2k + 1$ increases. With each successive increase in k, the singular value spectrum and the rank of the $\mathbf{D}_t(k)$ increases monotonically, reaching a value of $\text{rank}(\mathbf{D}_t(20)) = 4551$. This indicates the number of sources that could be recovered increases by up to a factor of 20 by modeling only local cortical dynamic connections.

Table 2.1: **Rank of $\mathbf{D}_t(k)$.**

k	0	1	2	5	10	20
$\text{rank}(\mathbf{D}_t(k))$	204	612	1020	2244	3889	4551

Similarly, we analyzed the the number of sources that could be recovered in the *IND* and *STS* models by evaluating the rank and singular values spectrum of the $\mathbf{D}_t^{(ind)}$ and $\mathbf{D}_t^{(sts)}$ matrices, respectively (Eq. 2.16). Just as we did in the case of our spatiotemporal dynamic model, we evaluated the truncated versions of these matrices given by:

$$\begin{aligned} \mathbf{D}_t(k)^{(ind)} &= \left[\mathbf{0}', \ldots, \mathbf{0}', \mathbf{X}', \mathbf{0}', \ldots, \mathbf{0}'\right]' \\ \mathbf{D}_t(k)^{(sts)} &= \frac{1}{\gamma_{t,t}} \left[\gamma_{t-k,t} \mathbf{X}', \ldots, \gamma_{t-1,t} \mathbf{X}', \mathbf{X}', \gamma_{t+1,t} \mathbf{X}', \ldots \gamma_{t+k,t} \mathbf{X}'\right] \end{aligned} \tag{2.18}$$

For both the *IND* and *STS* models, the rank of these matrices did not increase with

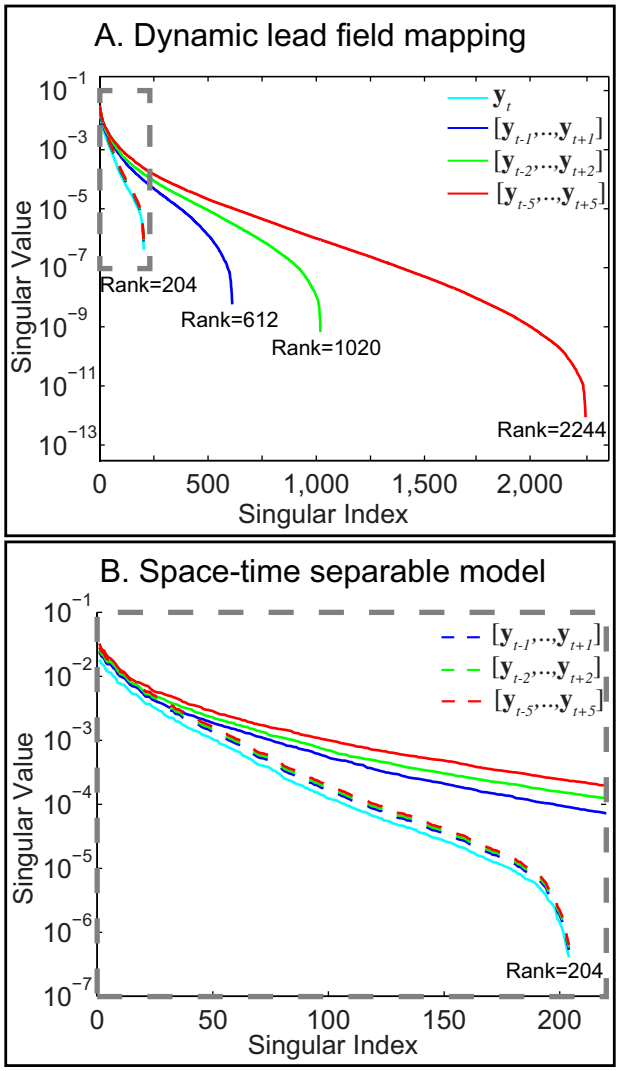

Figure 2-2: **Spectrum of the dynamic lead field mapping $\mathbf{D}_t(k)$.** A: The evolution of the singular value spectra of $\mathbf{D}_t(k)$ (solid lines) with increasing number of temporal measurements k. The spectrum of this matrix increases thus indicating an increase in the number of sources that could be recovered in the spatiotemporal dynamic model with local connections. B: Zoom-in to subpanel (A) shows the singular value spectra of $\mathbf{D}_t(k)^{(sts)}$ for the space-time separable (STS) source model (dashed line) indicating that the number of sources that could be recovered does not increase in the STS source model.

the inclusion of more temporal measurements: $\text{rank}(\mathbf{D}_t(k)^{(ind)}) = \text{rank}(\mathbf{D}_t(k)^{(sts)}) = 204$ for all values of k. In the case of the STS model, we saw that the 204 singular values of $\mathbf{D}_t(k)^{(sts)}$ were slightly increased (Fig. 2-2B). However, such increase was uniform across singular values, as it can be shown that the vector with the ordered singular

values of $\mathbf{D}_t(k)^{(sts)}$ is equal to a scaled version of the vector with the singular values of $\mathbf{X}$: $\text{svd}(\mathbf{D}_t(k)^{(sts)}) \propto \text{svd}(\mathbf{X})$. Importantly, the fact that the rank of $\mathbf{D}_t(k)^{(ind)}$ and $\mathbf{D}_t(k)^{(sts)}$ does not increase can be seen from the Equation (2.18). In both cases, these matrices are obtained by stacking the lead field matrix $\mathbf{X}$ above and below with either a matrix of zeros or scaled versions of $\mathbf{X}$.

2.4.2 Sensitivity analysis of dynamic lead field mapping $\mathbf{D}_t$

The biophysics of the MEG/EEG forward problem dictate that some regions of the brain are more difficult to observe and measure than others. This is because the measured signal decays with the inverse of the square distance from the sensor to the source, and in the case of MEG, sources oriented near the radial direction are magnetically silent. Figure 2-3 illustrates this phenomenon where we show arrows representing dipole sources located in the right cingulate cortex (deep dipole, white), the trough of a gyrus (mildly deep dipole, red), and the side of a gyrus (superficial dipole, yellow). In this hypothetical example, the deep white dipole located at a distance d_1 from its closest sensor produces very low amplitude signals which renders it practically impossible to detect; the mildly deep red dipole at a distance d_2 from its closest sensor generates relatively weak signals, and this makes it difficult to measure; in contrast to the previous two cases, the yellow dipole located on the superficial side of the gyrus at a distance d_3 from its closest sensor produces strong signals which makes it easy to recover. A natural way to quantify this ease or difficulty for detecting the electric and magnetic fields at the scalp surface when only the instantaneous $\mathbf{y}_t$ is observed, is to compute the signal power measured across sensors generated by a single active dipole source of unit amplitude. Specifically, if we fix the source vector to represent a unit amplitude active dipole at the ith cortical location, the total signal power it produces at the sensors is given by:

$$\varsigma_{i,t}(0) = ||\mathbf{X}\mathbf{1}_i||_2, \tag{2.19}$$

Performing such computation for all cortical locations i results in a static profile

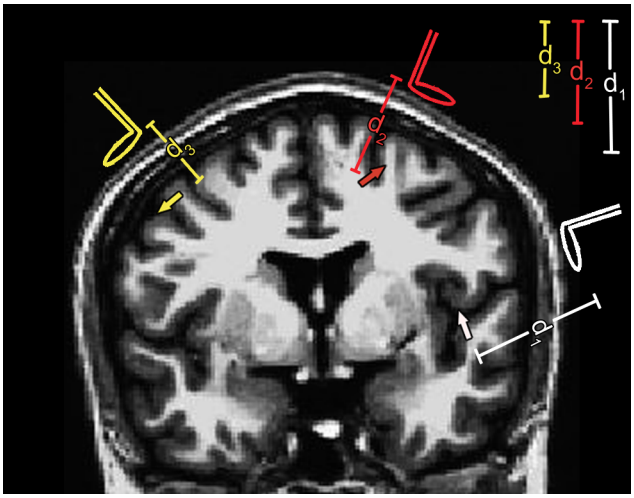

Figure 2-3: **Illustration of sensitivity of the static lead field matrix to dipoles at different depths.** The white dipole located very deep in the insular cortex at a distance d_1 from its closest sensor produces very weak signals which makes it practically undetectable. The red dipole near the trough of a gyrus at a distance d_2 from its closest sensor produces relatively weak signals and is difficult to detect. The superficial yellow dipole in the side of a sulcus at a distance d_3 from its closest sensor produces strong signals. For each dipole, the sensitivity can be calculated using Equation (2.20).

of the *absolute* sensitivity gain of the lead field matrix $\mathbf{X}$. Figure 2-4 shows this static sensitivity profile in the top panel ($k = 0$) computed using the lead field matrix described in the previous section. We saw that the sensitivity was highest in some portions of gyri and the more superficial cortical areas (red tones), but it was low in the troughs of sulci, insula, and inferior-frontal regions.

While the sensitivity of the lead field matrix gives us a static image of the signal quality obtained in the instantaneous measurement $\mathbf{y}_t$, it does not inform us of the quality of the measured signals in time when we account for the spatiotemporal dynamics of the underlying cortical activity. One way to asses such signal quality is to extend the sensitivity analysis done for the lead field matrix (Eq. 2.19) to the case of the dynamic lead field $\mathbf{D}_t(k)$. Specifically, if we assume a unit amplitude dipole is active at time t and cortical location i in our spatio-temporal dynamic model (Eq. 2.2), the total measured power in the time interval $[t-k, \ldots, t+k]$ across all sensor is given by:

$$\varsigma_{i,t}(k) = ||\mathbf{D}_t(k)\mathbf{1}_i||_2, \tag{2.20}$$

Figure 2-4A (left panels) shows the absolute sensitivity of the dynamic lead field mapping as we increase the number of measurement in time $(2k+1)$. We saw that the sensitivity increased in most cortical regions including sulci when more temporal information is included, but this increase was not as pronounced in deeper regions such as some portions of insula and the inferior-frontal cortex. We performed an equivalent sensitivity analysis assuming that the cortical activation was generated by the (STS) model. To compute this sensitivity, we used Equation (2.20) but in this case replaced $\mathbf{D}_t(k)$ with the truncated matrix $\mathbf{D}_t(k)^{(sts)}$ (Eq. 2.18). Figure 2-4B (center panels) shows the sensitivity obtained in the space-time separable model. We saw an increase in sensitivity in most cortical regions, but this increase was not as broadly spread as it was in the case of the dynamic lead field mapping (Fig. 2-4A). Figure 2-4C (right panels) shows the difference in absolute sensitivity between our spatiotemporal dynamic model and the STS model for each value of k: $\varsigma_{i,t}(k) - \varsigma_{i,t}(k)^{(sts)}$. We should note that the color scale in this case is different from that in panels A and B: the red and blue tones indicate small positive and negative differences of $\sim 0.2 \times 10^{-3}$, respectively; large positive difference of $\sim 2 \times 10^{-3}$ are shown in yellow, while large negative differences are shown in light blue. For the majority of cortical areas, and all values of k, the sensitivity was higher in the spatiotemporal dynamic model as indicated by the red and yellow tones. The small areas where the sensitivity of the STS model was slightly higher (blue areas where the difference is $\sim -0.2 \times 10^{-3}$) were mainly represented by superficial portions of sulci which are indeed already detectable from a static point of view, i.e., if we only consider the immediate measurement for analysis. Interestingly, for values of $k \geq 5$ the sensitivity in the spatiotemporal dynamic model is higher in deeper areas such as sulci and cingulate cortex, which are precisely the areas that are more difficult to detect from a static point of view.

To quantify the improvement in the dynamic lead field sensitivity $\varsigma_{i,t}(k)$ as we include more temporal information in relation to the static case, i.e., the static lead

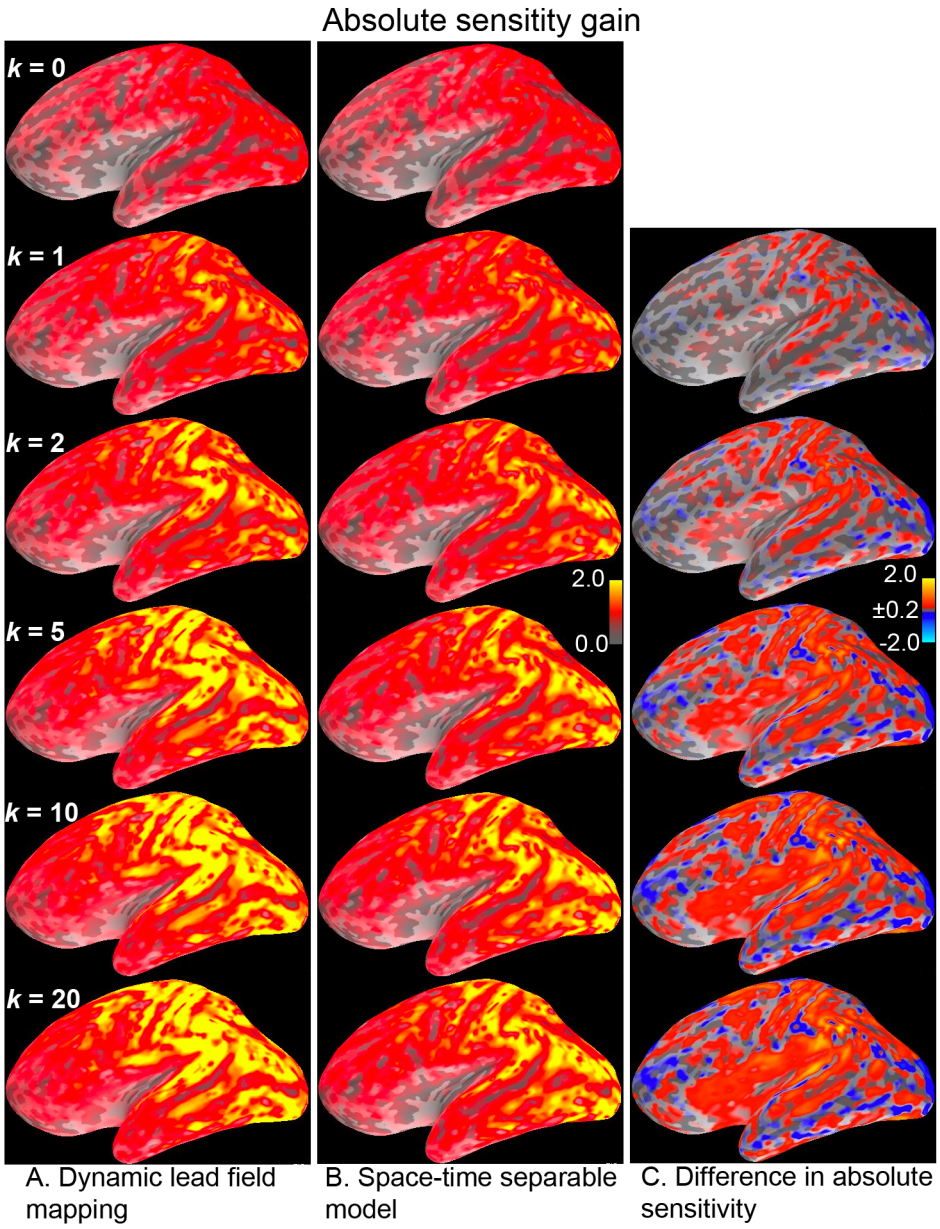

Figure 2-4: **Absolute sensitivity of the dynamic lead field mapping.** A: Absolute sensitivity of $\mathbf{D}_t(k)$ as a function of the number of incorporated measurements $2k+1$. When $k=0$, the sensitivity of the static lead field matrix $\mathbf{X} = \mathbf{D}_t(0)$ is shown. For $k \geq 1$, the sensitivity increases in most cortical regions including sulci. B: Absolute sensitivity for the space-time separable model $\mathbf{D}_t(k)^{(sts)}$ as a function of k. The sensitivity increases in this case but is not as widespread nor as high as it was in the case of the dynamic lead field as indicated by the differences in sensitivities between these two models (C, right panels). The color-scale is multiplied by 10^{-3}.

field matrix sensitivity $\varsigma_{i,t}(0)$, we computed the *relative* sensitivity gain: $\varsigma_{i,t}(k)/\varsigma_{i,t}(0)$. Figure 2-5A (left panels) shows the relative sensitivity gain of the dynamic lead field mapping. We found that the relative sensitivity increases in the majority of cortical regions with additional temporal information. Interestingly, the increments in relative sensitivity were higher in regions of low absolute sensitivity, such as sulci, insula, and inferior frontal cortex (yellow tones). We computed the relative sensitivity in the case of the space-time separable (STS) model by using instead truncated versions of the $\mathbf{D}_t^{(sts)}$ matrix (Eq. 2.18, bottom). The resulting relative increments in the STS model were spatially uniform across cortex (Figure 2-5B). Compared to the relative sensitivity of the dynamic lead field, the increments in the STS model were not as pronounced in regions that are more difficult to detect from a static point of view, such as sulci, insula, and inferior-frontal regions.

We saw in the previous section that, in a model that includes local spatiotemporal cortical dynamics, adding temporal measurements (k) increases the rank of the $\mathbf{D}_t(k)$ matrices. We therefore wanted to characterize the dynamic lead field sensitivity that corresponded to the rank increments, which thus represent the sensitivity gains in the newly accessible dimensions. Specifically, if we consider two dynamic lead field matrices $\mathbf{D}_t(k_1)$ and $\mathbf{D}_t(k_2)$ with $k_1 < k_2$, the sensitivity in the dimension accessible at k_2 but not at k_1 can be obtained by: first projecting the rows of $\mathbf{D}_t(k_2)$ onto the null space of $\mathbf{D}_t(k_1)$; and then compute the sensitivity of the matrix constructed with the projected rows. Figure 2-6A shows the sensitivity in the newly accessible areas for the pairs of values corresponding to consecutive number in $k \in [1, 2, 5, 10, 20]$. As k increased, the distribution of sensitivity for the newly accessible dimensions was highest in some portions of sulci, insula, and inferior-frontal regions, which are poorly detected areas in the static case (Fig. 2-6A, left panels). We performed an equivalent analysis for the space-time separable (STS) model. In this case there were no gains in sensitivity in the previously unaccessible dimensions (Fig. 2-6B, right panels). This results from the fact that the rank of the $\mathbf{D}_t(k)^{(sts)}$ matrix does not increase irrespective to the number of analyzed measurements (k).

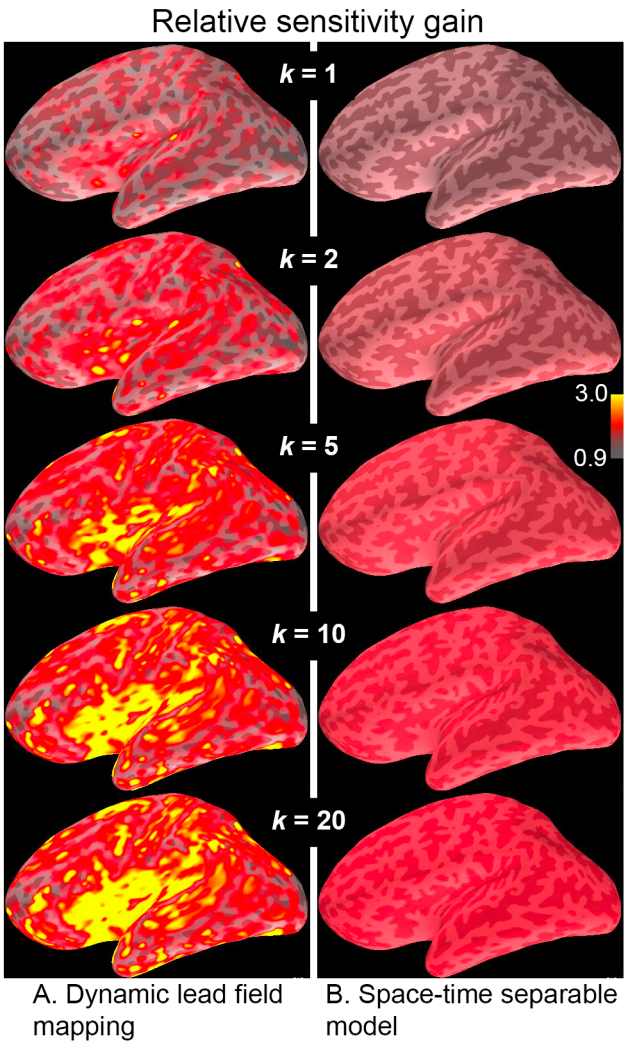

Figure 2-5: **Relative sensitivity of the dynamic lead field mapping.** A: The relative sensitivity of the dynamic lead field mapping increases with k in most cortical regions. This relative increase is more pronounced in areas with low static sensitivity such as sulci, insula, and inferior frontal cortex. B: The increments in relative sensitivity for the space-time separable model are spatially uniform and not as pronounced in regions that are difficult to detect such as sulci, insula, and inferior frontal cortex.

2.5 Conclusions and discussion

In this work we have shown that the number of independent sources that can be recovered from MEG recordings can increase up to a factor of ~ 20 by modeling basic first-order and linear source dynamics and local cortical connections. To do this, we developed and analyzed the concept of the *dynamic lead field mapping*. This

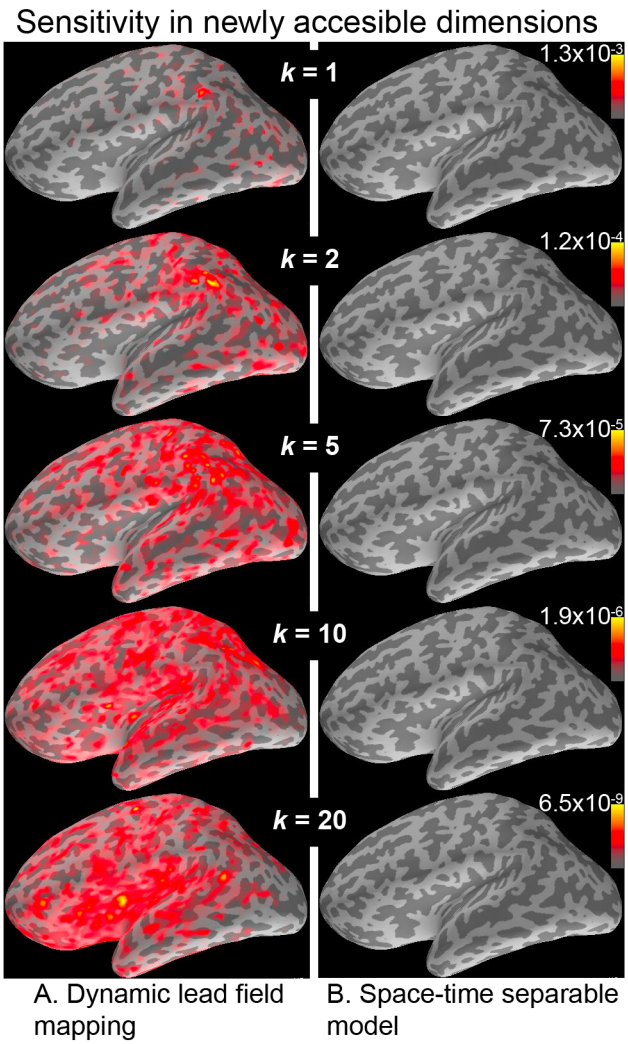

Figure 2-6: **Sensivity of dynamic lead field in newly accessible dimensions.** A: The sensitivity of the dynamic lead field mapping in the newly accessible dimension increases as a function of k and is highest in some portions of sulci, insula, and inferior frontal regions. B: In the case of the space-time separable model, there are no sensitivity gains in previously unaccessible dimensions since the rank of the matrix $\mathbf{D}_t(k)^{(sts)}$ does not increase.

dynamic mapping extends the static observation model that originates from the lead field biophysics (Eq. 2.3) to to account for spatiotemporal dynamics [36, 37, 38, 39, 79, 80, 81]. The dynamic lead field mapping expresses the relation between the cortical source vector at any given time t and the MEG/EEG measurements in the complete experiment interval $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$, creating an observation model for the complete time series of measurements that maintains the crucial signal-plus-noise structure

present in its static counterpart. Since the rank of the lead field matrix determines the number of independent variables we can expect to recover from an individual measurement in time, the rank of the dynamic lead field mapping determines the number of variables we can recover in the dynamic case from the complete time series of observations. Therefore, in a typical MRI-based source model with ~ 5124 dipole sources and $\sim$ 204 sensors, while the static (IND) or space-time separable (STS) models can recover at most $\sim 4\%$ ($204/5124 \times 100$) of the independent variables, the approximate $4551/204 \approx 20$ fold rank improvement of the dynamic lead field implies that we could recover up to $\sim 89\%$ ($4551/5124 \times 100$) of the independent variables when we consider the brain's source activity as a spatiotemporal dynamic process.

While the static lead field matrix provides information about the source vector from a single measurement at a given point in time, the dynamic lead field provides additional information by accounting for past and future measurements that reflect the trajectory followed by the dynamic cortical state. Along this trajectory, states hidden within a "blind spot" of the static lead field at a given time can evolve into a "visible" portion of the static lead field at some other time due to the brain's high level of connectivity. Since the dynamic lead field mapping explicitly describes how the measurement at one time point contains information about the cortical state at another time point, this mapping effectively shows how information can flow from blind spots in the static lead field to more visible locations. In principle, depending on the cortical connectivity and dynamics, it is then possible to access and potentially estimate a much larger number of independent variables. Our work shows that modeling the simplest spatiotemporal relationships could dramatically improve source localization. Moreover, it suggests that more comprehensive models of connectivity and spatiotemporal dynamics could provide even greater improvements.

We performed a sensitivity analysis of the dynamic lead field mapping to determine the degree to which different cortical areas produce detectable recordings as we increased the number of measurements in time. This analysis showed that, in our spatiotemporal dynamic paradigm, brain areas that are difficult to detect from a static point of view, such as those with low sensitivity in troughs of sulci, insula, and inferior-

frontal regions—"blind spots" as described above—become more accessible with the addition of temporal measurements. Thus, activity within these "blind spots" become accessible through their connection to more visible regions. This could explain the improvements in MEG/EEG source localization accuracy obtained via algorithms that explicitly model source dynamics with local cortical connections [34, 76].

We have generalized the development of the dynamic lead field mapping to include arbitrary Gaussian source models using a projections approach. With this generalization, we performed a detailed analysis of two Gaussian source models commonly used in the MEG/EEG inverse problem literature, namely the temporally independent model (*IND*) and the space-time separable model (*STS*). We showed that these source models do not produce rank improvements in their respective observation matrices as the number of temporal measurements increased, nor do they obtain sensitivity gains in cortical "blind spots" of the static lead field. This finding suggests that source models that either ignore dynamics or restrict the spatiotemporal source structure do not improve with additional temporal measurements. This is consistent with our previous discussion, since such source models do not account for dynamic connections between blind and visible portions of the static lead field.

We envision a possible way in which our dynamic lead field construct could be extended to more general source models, to potentially include non-Gaussian models or more realistic spatiotemporal neural source dynamics. In principle, this could be achieved by taking techniques from control theory designed to analyze deterministic systems, and adapting them to characterize properties of stochastic dynamical systems. In control theory, a basic question is to determine if different initial states of a deterministic dynamical system necessarily produce different measurements. When this is the case, the system is called *observable*, since the system's initial states can be distinguished based on measurements alone [55]. In our linear dynamic state-space model (Eqs. 2.2 and 2.3), for example, assuming that the state input $\boldsymbol{\omega}_t$ is a priori known and deterministic, and that the measurements are noiseless ($\boldsymbol{\varepsilon}_t = 0$), effectively converts our model into a deterministic system. In this deterministic case, an unknown initial state $\boldsymbol{\beta}_0$ of the system can be distinguished

based on noiseless measurements when the rank of the so-called observability matrix $\mathcal{O} = [\mathbf{X}', \phi\mathbf{F}'\mathbf{X}', \ldots, \phi^{p-1}\mathbf{F}'^{p-1}\mathbf{X}']'$ is equal to the dimension of the state (p). We can see here that the observability matrix $\mathcal{O}$ is related to the mapping given by the dynamic lead field from the state $\boldsymbol{\beta}_t$ to only the present and future observations $[\mathbf{y}_t, \mathbf{y}_{t+1}, \ldots, \mathbf{y}_{t+p-1}]$ (this can be seen in Equation (2.11)). We therefore hypothesize that the dynamic lead field concept developed here could be extended by incorporating insights from the analysis of observability in deterministic nonlinear systems [83], and by applying the time-reversed representation of nonlinear stochastic dynamic systems [84].

A long-standing conjecture in functional neuroimaging is that the integration of different imaging modalities, such as MEG, EEG, fMRI, and DOT, could improve spatiotemporal resolution, due to the complementary nature of the physics and physiology underlying these modalities. Extensions of our analysis to multimodal neuroimaging could provide a formal paradigm to characterize and maximize the spatiotemporal resolution that can be obtained from multimodal data. For example, an immediate multimodal effort could be aimed at including long-distance connections—both between distant cortical locations, and between cortex and low sensitivity subcortical areas—by using anatomical connectivity derived from diffusion MRI [79]. We predict that the inclusion of these long-distance connections in a stochastic dynamics framework could have a dramatic impact on the number and location of source parameters that could be recovered from MEG/EEG time series data. While the focus of our work is in the field of neuroimaging, our dynamic lead field approach could have application in other areas where an underlying spatiotemporal stochastic dynamical system is observed through noisy serial measurements. Therefore, we hypothesize that our approach could find applications in areas such as geophysics, network theory, and epidemiology, where physical spatiotemporal dynamic relationships may exist, but are not yet exploited in inverse solutions.

Our analysis and results have been focused on the number, spatial distribution, and sensitivity of sources that can be recovered under a spatiotemporal dynamic framework. In practical applications, model misspecification, as well as the signal-to-

noise of the measurements, will likely limit performance gains to some extent below the levels we have reported. However, the large improvements in rank and sensitivity that we have observed, even under a very simple spatiotemporal model, suggest that substantial performance improvements would be achievable in practical source localization applications. In future work, we will analyze source localization performance under this spatiotemporal framework and characterize the influence of different spatiotemporal model choices.

Due to the disparity between the number of unknown cortical sources and the number of sensors, and the low sensitivity of MEG/EEG to deeper sources, accurate source estimates are thought to be unattainable. In this paper, we have demonstrated how source models that delineate the brain's dynamic and spatial connectivity structure could dramatically improve MEG/EEG inverse solutions. Our work therefore suggests a new way forward in this field, where information from neurophysiology and neuroanatomy studies about brain connectivity and dynamics could be used to guide the development of models of brain activity in functional neuroimaging. Given the dramatic improvements we have observed under a very simple model, our hope is that our results can inspire the development of more detailed models of spatiotemporal brain activity, and the use of these models as a fundamental component of MEG/EEG source localization analyses.

Acknowledgment

We would like to thank Dr. Simona Temereanca for providing us with the mu-rhythm MEG data and structural MRI used in this project.

2.6 Supplemental materials

2.6.1 Ethics statement

Human studies were approved by the Human Research Committee of Massachusetts General Hospital, Boston, MA.

2.6.2 Data description and preprocessing

In our analysis, the lead field matrix $\mathbf{X}$ was computed with the *MNE* software using an MRI-based boundary element model from a human subject [1, 77], with $p = 5124$ dipole sources oriented normal to the cortical surface. This arrangement yielded an average distance between nearest neighbors of ~ 6.2 mm, resulting in a model that is consistent with intracranial electrophysiology and neuroanatomy [36, 37, 38, 39]. Due to the difference in orders of magnitude and physical units in the elements of the lead field matrix $\mathbf{X}$ corresponding to planar gradiometers and magnetometers, we restricted our analysis to the 204 planar gradiometers of the Neuromag Vectorview system at Massachusetts General Hospital. The $\mathbf{F}$ matrix was specified using the nearest-neighbor dynamic formulation described by Equations (2.1) and (2.2), with the stability parameter $\phi = 0.95$. The state input covariance matrix $\mathbf{Q}$ was estimated from *mu*-rhythm data from one subject using the *dMAP-EM* algorithm presented in [76]. The *mu*-rhythm originates from motor and somatosensory cortices, and consists of synchronous oscillations with 10 and 20-Hz components. Data were collected from one subject using a 306-channel Neuromag Vectorview MEG system at the Massachusetts General Hospital. The signals were acquired at 601 Hz with a bandwidth of 0.1 to 200 Hz, and downsampled to 204.8 Hz for subsequent analysis. In the *STS* model, the elements for the temporal covariance matrix $\mathbf{\Gamma}$ were set as suggested in [27]: $\gamma_{t\pm k,t} = \sum_{j=1}^{T} \exp\{-1/2[(t \pm k - j)^2 + (j - t)^2]\Delta^{-2}\psi^{-2}\}$, where $\Delta = 4 \times 10^{-3}$ s and $\psi = 204.8$ Hz.

2.6.3 The dynamic lead field mapping projects $[\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_T']'$ onto $\boldsymbol{\beta}_t$

In this section we show that the observation model for the complete set of measurements $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ under the dynamic lead field mapping possesses the desired *signal* plus *noise* structure in which the source vector at time t is independent of the *noise* term (Eq. 2.10). In order to do this, we will show that the matrices $\mathbf{P}_{t\pm k,t}$ that make up the blocks of the dynamic lead field mapping are indeed matrices projecting

the corresponding measurement $\mathbf{y}_{t\pm k}$ onto the source vector $\boldsymbol{\beta}_t$. Doing so would imply that the dynamic lead field mapping achieves the desired independence condition.

We should first point to a few facts about the linear projection $\mathbf{P}_{t\pm k,t}\boldsymbol{\beta}_t$ of the measurement vector $\mathbf{y}_{t\pm k}$ onto the source vector $\boldsymbol{\beta}_t$. By definition, the projection error $\mathbf{e}_{t\pm k,t} = \mathbf{y}_{t\pm k} - \mathbf{P}_{t\pm k,t}\boldsymbol{\beta}_t$ is uncorrelated with $\boldsymbol{\beta}_t$[55]:

$$\mathrm{E}\left[(\mathbf{y}_{t\pm k} - \mathbf{P}_{t\pm k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'\right] = \mathbf{0}. \tag{2.21}$$

In addition, since $\mathbf{y}_{t\pm k}$ and $\boldsymbol{\beta}_t$ are jointly Gaussian, the projection error $\mathbf{e}_{t\pm k,t}$ and $\boldsymbol{\beta}_t$ are also jointly Gaussian. Therefore, the projection error and the source vector are uncorrelated and jointly Gaussian, and thus are independent. As a result, we can use the definition of the projection error to express any measurement vector $\mathbf{y}_{t\pm k}$ as the sum of the projection $\mathbf{P}_{t\pm k,t}\boldsymbol{\beta}_t$ and projection error $\mathbf{e}_{t\pm k,t}$:

$$\mathbf{y}_{t\pm k} = \mathbf{P}_{t\pm k,t}\boldsymbol{\beta}_t + \mathbf{e}_{t\pm k,t}.$$

If we do this for all values of k we obtain an observation model for the complete time series of measurements where the source vector $\boldsymbol{\beta}_t$ is indeed independent of the noise term $[\mathbf{e}_{1,t}', \mathbf{e}_{2,t}', \ldots, \mathbf{e}_{T,t}']'$:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{t-1} \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{P}_{1,t} \\ \vdots \\ \mathbf{P}_{t-1,t} \\ \mathbf{P}_{t,t} \\ \mathbf{P}_{t+1,t} \\ \vdots \\ \mathbf{P}_{T,t} \end{bmatrix}}_{\text{Signal}} \boldsymbol{\beta}_t + \underbrace{\begin{bmatrix} \mathbf{e}_{1,t} \\ \vdots \\ \mathbf{e}_{t-1,t} \\ \mathbf{e}_{t,t} \\ \mathbf{e}_{t+1,t} \\ \vdots \\ \mathbf{e}_{T,t} \end{bmatrix}}_{\text{Noise}}.$$

Therefore, if we show that the choices we made for the $\mathbf{P}_{t\pm k,t}$ matrices for the dynamic lead field mapping in Section 2.3.2, namely,

$$\mathbf{P}_{t,t} = \mathbf{X},$$
$$\mathbf{P}_{t+k,t} = \phi^k \mathbf{X}\mathbf{F}^k,$$
$$\mathbf{P}_{t-k,t} = \phi^k \mathbf{X}\mathbf{F}_b^k,$$

satisfy the orthogonality condition (Eq. 2.21), it would follow that the *signal* and *noise* portions in Equation (2.10) are independent.

In the case where the projection matrix corresponds to the present observation ($k = 0$), we substitute Equation (2.3) in Equation (2.21), and note that setting $\mathbf{P}_{t,t} = \mathbf{X}$ achieves the desired orthogonality condition since the source vector $\boldsymbol{\beta}_t$ is independent of the measurement noise $\boldsymbol{\varepsilon}_t$:

$$\begin{aligned}
\mathrm{E}\left[(\mathbf{y}_t - \mathbf{P}_{t,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'\right] &= \mathrm{E}\left[(\mathbf{X}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t - \mathbf{P}_{t,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'\right] \\
&= \mathrm{E}\left[\boldsymbol{\varepsilon}_t\boldsymbol{\beta}_t'\right] \\
&= \mathbf{0}.
\end{aligned}$$

To analyze the case where the measurement corresponds to a future observation, we should first make note of two facts in relation to the source vector process. The first is that, by definition, the input process to the source dynamics is independent across time, i.e., $\boldsymbol{\omega}_t$ is independent of $\boldsymbol{\omega}_{t\pm j}$ for $j \in [1, 2, \ldots]$. The second is that given the recursive definition of the source process (Eq. 2.2), the source vector at a particular time t is a function of only the present input $\boldsymbol{\omega}_t$ and past inputs $\boldsymbol{\omega}_{t-j}$, $j \in [1, 2, \ldots]$. Because of these two facts, the source vector a time t, $\boldsymbol{\beta}_t$, is independent of the future inputs $\boldsymbol{\omega}_{t+j}$, $j \in [1, 2, \ldots]$.

With this result at hand, we proceed to deduce the future projection matrices $\mathbf{P}_{t+k,t}$, $k \in [1, 2, \ldots]$ by using the forward iteration in Equation (2.5):
$\boldsymbol{\beta}_{t+k} = \phi^k \mathbf{F}^k \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{F}^{k-j} \boldsymbol{\omega}_{t+j}$. We substitute Equations (2.3) in Equation (2.21), then substitute Equation (2.5) in the obtained result, and note

that by setting $\mathbf{P}_{t+k,t} = \phi^k \mathbf{X}\mathbf{F}^k$ we achieve the orthogonality condition since $\boldsymbol{\beta}_t$ is independent of both $\boldsymbol{\varepsilon}_{t+k}$, and $\boldsymbol{\omega}_{t+j}$, $j \in [1, 2, \ldots]$:

$$\begin{aligned}
\mathrm{E}[(\mathbf{y}_{t+k} - \mathbf{P}_{t+k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] &= \mathrm{E}[(\mathbf{X}\boldsymbol{\beta}_{t+k} + \boldsymbol{\varepsilon}_{t+k} - \mathbf{P}_{t+k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] \\
&= \mathrm{E}[(\phi^k \mathbf{X}\mathbf{F}^k\boldsymbol{\beta}_t + (1-\phi^2)^{1/2}\sum_{j=1}^{k}\phi^{k-j}\mathbf{X}\mathbf{F}^{k-j}\boldsymbol{\omega}_{t+j} + \boldsymbol{\varepsilon}_{t+k} - \mathbf{P}_{t+k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] \\
&= \mathrm{E}[((1-\phi^2)^{1/2}\sum_{j=1}^{k}\phi^{k-j}\mathbf{X}\mathbf{F}^{k-j}\boldsymbol{\omega}_{t+j} + \boldsymbol{\varepsilon}_{t+k})\boldsymbol{\beta}_t'] \\
&= \mathbf{0}.
\end{aligned}$$

To analyze the past observations $\mathbf{y}_{t-k}$, $k \in [1, 2, \ldots]$, we proceed similarly to the case of the future projections but use instead the *backwards Markovian* representation (Eq. 2.7) and its k-step past iteration (Eq. 2.8)[4]: $\boldsymbol{\beta}_{t-k} = \phi^k \mathbf{F}_b^k \boldsymbol{\beta}_t + (1-\phi^2)^{1/2}\sum_{j=1}^{k}\phi^{k-j}\mathbf{F}_b^{k-j}\boldsymbol{\omega}_{t-j}^b$. We should note that in the *backwards Markovian* representation (Eq. 2.7), the source vector $\boldsymbol{\beta}_t$ is a function of only the present $\boldsymbol{\omega}_t^b$ and future inputs $\boldsymbol{\omega}_{t+j}^b$, $j \in [1, 2, \ldots]$. In addition, given that the backwards input process $\boldsymbol{\omega}_t^b$ is independent across time, in this equivalent representation the source vector $\boldsymbol{\beta}_t$ is independent of the past inputs $\boldsymbol{\omega}_{t-j}^b$, $j \in [1, 2, \ldots]$. Using this result, we can readily obtain the past projection matrices. We use Equations (2.3), (2.21), and (2.8) to deduce that by setting $\mathbf{P}_{t-k,t} = \phi^k \mathbf{X}\mathbf{F}_b^k$ we obtain the desired orthogonality condition since $\boldsymbol{\beta}_t$ is independent of both $\boldsymbol{\varepsilon}_{t-k}$, and $\boldsymbol{\omega}_{t-j}^b$, $j \in [1, 2, \ldots]$:

[4]We should point out that directly reversing the time direction of the equation defining the source spatiotemporal dynamics (Eq. 2.2), i.e., $\boldsymbol{\beta}_{t-1} = \phi^{-1}\mathbf{F}^{-1}\boldsymbol{\beta}_t - \phi^{-1}(1-\phi^2)^{1/2}\mathbf{F}^{-1}\boldsymbol{\omega}_t$, instead of using the *backwards Markovian* model does not yield the desired orthogonality condition. This is because the k-step backwards iteration resulting from this direct time reversal, namely, $\boldsymbol{\beta}_{t-k} = \phi^{-k}\mathbf{F}^{-k}\boldsymbol{\beta}_t - (1-\phi^2)^{1/2}\sum_{j=0}^{k-1}\phi^{j-k}\mathbf{F}^{j-k}\boldsymbol{\omega}_{t-j}$, contains the terms $\boldsymbol{\omega}_t$ and $\boldsymbol{\omega}_{t-j}$ which are not independent of the source vector $\boldsymbol{\beta}_t$.

$$
\begin{aligned}
\mathrm{E}[(\mathbf{y}_{t-k} - \mathbf{P}_{t-k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] &= \mathrm{E}[(\mathbf{X}\boldsymbol{\beta}_{t-k} + \boldsymbol{\varepsilon}_{t-k} - \mathbf{P}_{t-k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] \\
&= \mathrm{E}[(\phi^k \mathbf{X}\mathbf{F}_b^k \boldsymbol{\beta}_t + (1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X}\mathbf{F}_b^{k-j} \boldsymbol{\omega}_{t-j}^b + \boldsymbol{\varepsilon}_{t-k} - \mathbf{P}_{t-k,t}\boldsymbol{\beta}_t)\boldsymbol{\beta}_t'] \\
&= \mathrm{E}[((1-\phi^2)^{1/2} \sum_{j=1}^{k} \phi^{k-j} \mathbf{X}\mathbf{F}_b^{k-j} \boldsymbol{\omega}_{t-j}^b + \boldsymbol{\varepsilon}_{t-k})\boldsymbol{\beta}_t'] \\
&= \mathbf{0}.
\end{aligned}
$$

Chapter 3

Spatial Sparsity, Long Range Connectivity, and Dynamics in a Multi-Scale Solution to the MEG/EEG Inverse Problem

3.1 Abstract

MEG/EEG are important tools that non-invasively record brain function with millisecond temporal resolution, allowing us to study fast scale dynamics of brain activation. In order to fully grasp the power of these techniques, one needs to estimate the cortical activity that generates these recordings, i.e., to solve the MEG/EEG inverse problem. However, due to the biophysics of how the MEG/EEG signals are generated as well as the fact that the number of sensors is much smaller that the number of sources, this inverse problem is highly ill-posed and has no unique solution. Because of this, source estimation methods need to include additional information about the nature of the source activity. In this work we present a source localization method that models source dynamics, local and long distance brain structural connections derived from diffusion MRI, as well as spatial sparsity—we refer to this method as the

Multi-Scale Kalman Filter based Group Subspace Pursuit algorithm. We show via simulations and experimental data from a human subject the improvements in source localization accuracy obtained using our method. Our fast algorithm is competitive in term of computational speed, while greatly improves source estimation accuracy. Our results show the great importance of proper modeling of activation dynamics and brain structural connectivity, as well as algorithmic development, in the MEG/EEG inverse problem. This result therefore suggests a way forward in MEG/EEG source imaging in which detailed models of the brain's structure and function can yield significant improvements in source localization accuracy.

3.2 Introduction

The localization of current sources inside the brain from electromagnetic fields measured at the scalp surface alone is an ill-posed problem since it solution is generally not unique. This is due to the fact that different source configurations can produce identical measurements at the scalp [1]. In EEG/MEG brain imaging, the estimation of current sources can be even more challenging since the sensitivity of the recording device decays rapidly with the square of the distance from the sensor to the sources, as well as due to the fact that the number of sensors is typically much smaller that the number of discrete sources distributed across the brain tissue. However, incorporating brain structural connectivity may provide a means to dramatically improve source localization. This is because source activity in areas that are more difficult to estimate due to its orientation and distance to the sensors can be better inferred if they are connected to sources in easier to estimate areas. Therefore, if brain connectivity could be modeled properly, incorporating both short and long distance connections, this approach could provide substantial improvements in source localization performance.

3.2.1 MEG/EEG measurement model

In a MEG/EEG experiment we record data on or above the scalp at a few hundred sensors. This data is regularly sampled over a time interval. Typically, the experi-

ment is repeated between $m = 50$ and $m = 100$ trials or epochs. In some cases, such as in the analysis of resting state data or in the study of epilepsy, the trials cannot be naturally repeated and we can only obtain a single trial. We define $\mathbf{y}_t^{(i)} \in \mathbb{R}^n$ as the measurement vector at time $t \in [1, \ldots, T]$ and trial $i \in [1, \ldots, m]$ in all measured locations n (~ 300). The MEG/EEG measurements are generated by the postsynaptic currents in synchronously active neuronal populations across the brain's cortices. This process can be accurately modeled by a few thousand current dipoles uniformly distributed over the cortical mantle and perpendicular to it [1]. We define $\beta_{j,t}$ as the current dipole at time t and cortical location $j \in [1, \ldots, p]$, where $p \sim 5000$, and the source vector as $\boldsymbol{\beta}_t = [\beta_{1,t}, \ldots, \beta_{p,t}]'$. At any point in time t and trial i, the measurement and source vector are related via the linear model:

$$\mathbf{y}_t^{(i)} = \mathbf{X}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t^{(i)}, \tag{3.1}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the gain matrix that maps the source activity to the measured fields and potentials. This matrix is obtained via a quasi-static approximation to Maxwell's equations using a boundary element model derived from high-resolution MRI [1, 54]. Without loss of generality, we can assume that our measurement model (Eq. 3.1) has been spatially whitened, i.e., that the original raw data model $\tilde{\mathbf{y}}_t^{(i)} = \tilde{\mathbf{X}}\boldsymbol{\beta}_t + \tilde{\boldsymbol{\varepsilon}}_t^{(i)}$ has been pre-multiplied by the inverse of a matrix square root of the covariance of $\tilde{\boldsymbol{\varepsilon}}_t$. Therefore, we can assume that the noise vector $\boldsymbol{\varepsilon}_t^{(i)}$ has a zero mean and covariance equal to the identity matrix. We further assume that $\boldsymbol{\varepsilon}_t^{(i)}$ is independent across time as well as it is independent of the modeled signal $\mathbf{X}\boldsymbol{\beta}_t$.

At each point in time t, the data is average over the m trials or epochs to produce an evoked response: $\bar{\mathbf{y}}_t = \frac{1}{m}\sum_{i=1}^{m} \mathbf{y}_t^{(i)}$. The evoked responses can thus be model via:

$$\bar{\mathbf{y}}_t = \mathbf{X}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \tag{3.2}$$

where $\boldsymbol{\varepsilon}_t$ is the noise in the evoked response at time t, and has zero mean and covariance equal to $\frac{1}{m}\mathbf{I}$.

3.2.2 Long distance connections can reduce entropy of source estimates

One way to incorporate structural connectivity of this kind in the MEG/EEG source localization problem is to use a linear dynamical system:

$$\beta_{i,t} = \underbrace{f_{i,i}\beta_{i,t-1}}_{\text{Past activity}} + \underbrace{\sum_{j\in\mathcal{N}_{Short}(i)} f_{i,j}\beta_{j,t-1}}_{\text{Short Distance Connections}} + \underbrace{\sum_{j\in\mathcal{N}_{Long}(i)} f_{i,j}\beta_{j,t-1}}_{\text{Long Distance Connections}} + \underbrace{\omega_{i,t}}_{\text{Unaccounted factors}}, \tag{3.3}$$

where $\beta_{i,t}$ represents the activity of the ith dipole source at time t, the coefficients $f_{i,j}$ indicate the strength of the spatiotemporal interactions between the ith and jth dipoles, $\mathcal{N}_{Short}(i)$ and $\mathcal{N}_{Long}(i)$ are the set of close and distantly connected dipoles to the ith source, and $\omega_{i,t}$ is a white noise process driving the dynamics of the ith source. We can readily write Equation 3.3 in vector form as $\boldsymbol{\beta}_t = \mathbf{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t$, where $\mathbf{F}$ is the feedback connectivity matrix in our model. In this work, we model the short distance connections as those corresponding to nearest-neighbor sources. We then model the long-distance connection as those corresponding to axonal bundles derived from diffusion MRI (Figure 3-1, left panels).

The potential improvement of our proposed approach that includes long-distance connection can be readily quantified in simulation studies. To do this we evaluated the source localization accuracy using a dynamic model that includes: (1) nearest-neighbor short distance connections (Figure 3-1, bottom right panel), and (2) the long distance neuroanatomical connections that are thought to link the Default Mode Network (DMN) [81, 85] (Figure 3-1, top right panel). We computed the MEG gain matrix $\mathbf{X}$ using the Freesurfer and MNE software packages [52, 53, 54] using high-resolution MRI from a human subject. The nearest-neighbor connectivity was obtained from the MNE software, while the long distance connections were derived from Freesurfer's automatic cortical parcellation [86] by linking the cortical areas in the DMN that have been show to be anatomically connected [81]. In each scenario,

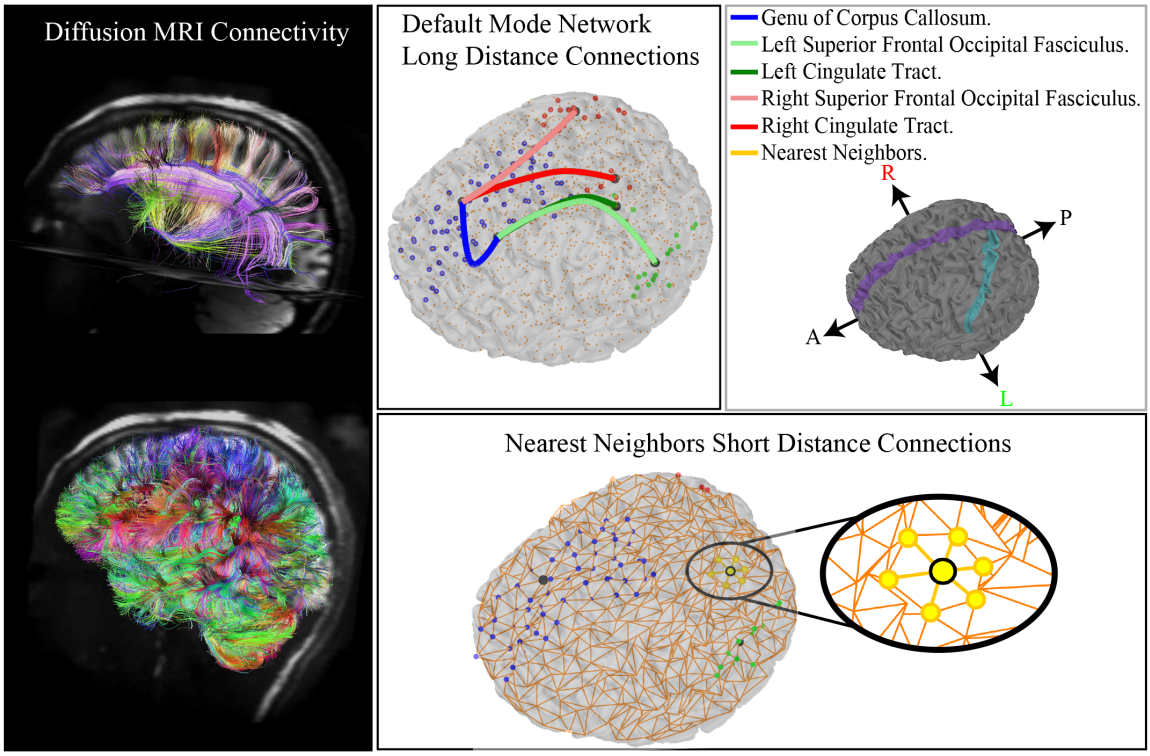

Figure 3-1: **Long-distance and nearest-neighbor connectivity in the MEG/EEG inverse problem**. Anatomically derived local and long distance structural connections

we simulated the source activity using the autoregressive model in Equation 3.3 with connectivity parameters ($f_{i,j}$, $i \neq j$) set equal to 0.45 divided by the number of neighbors, and $f_{i,i} = 0.45$. In this way, the connections attempt to explain approximately one half of the amplitude of the sources and the source history explains the other half. At the same time, since the largest eigenvalue of $\mathbf{F}$ is less than one we obtain stable dynamics. We then generated the MEG time series by mapping the simulated source activity via the gain matrix $\mathbf{X}$ and added noise to achieve a signal-to-noise ratio of approximately 5. At last, under each model parametrization, we estimated the source activity using the Kalman Filter [56].

In Figure 3-2 we show the reduction in the uncertainty of the estimated dipole sources due to the inclusion of both types of connections by computing the entropy of the source estimates. The entropy is a widely used measure that combines the uncertainty or noise in a group of estimates. In our case, high entropy implies that

the source estimates have high noise and large confidence intervals (Figure 3-2, top left panel), while a low entropy values indicate more accurate source estimates (Figure 3-2, bottom left panel). We specifically computed the entropy of the posterior distribution of the source vector $\boldsymbol{\beta}_t$ given the MEG measurements $\mathbf{y}_t$ in the time interval $0 \leq t \leq t_e$ while we varied the feedback connectivity matrix (Figure 3-2, right panel):

$$S(\mathbf{F}) = \mathrm{E}\left[\ln \Pr(\boldsymbol{\beta}_{t_e} | \mathbf{y}_t,\, 0 \leq t \leq t_e)\right] \tag{3.4}$$

where the expectation operation E is computed with respect the posterior distribution, and t_e is the time where the posterior distribution reaches an equilibrium. We first computed the entropy for the model that only includes nearest-neighbor connections $S(\mathbf{F}_{NN})$, and then evaluated how the entropy decreased while sequentially adding long distance DMN connection. We found that the entropy decreases when the connections that link cortical areas within the left hemisphere in the DMN (green circles) are added to the nearest-neighbor connections already present in the model. In addition, we found that the entropy was further reduced when we also included the inter-hemispheric connection (blue circle, genu of corpus callosum). Lastly, we found that the entropy reached its lowest value when we included the remaining DMN long distance connections to the model (red circles, right hemisphere). Interestingly, randomly adding edges to the nearest-neighbor connectivity model could reduce the entropy (dashed line, diamond marks). However, in this randomly connected model, thousand of added edges were necessary to reduce the entropy to the level observed in the DMN, which consists only of 5 connections. Furthermore, we found that the entropy increases when we modeled only the spatial correlations emerging from the connectivity model but eliminated the temporal dynamics, yielding a prior model with spatially correlations equivalent to those of the spatially connected model (dashed line).

This preliminary result indicates that the inclusion of well-known local and long distance neuroanatomic connections in a dynamical systems framework significantly improves source localization accuracy in the EEG/MEG brain imaging. As suggested

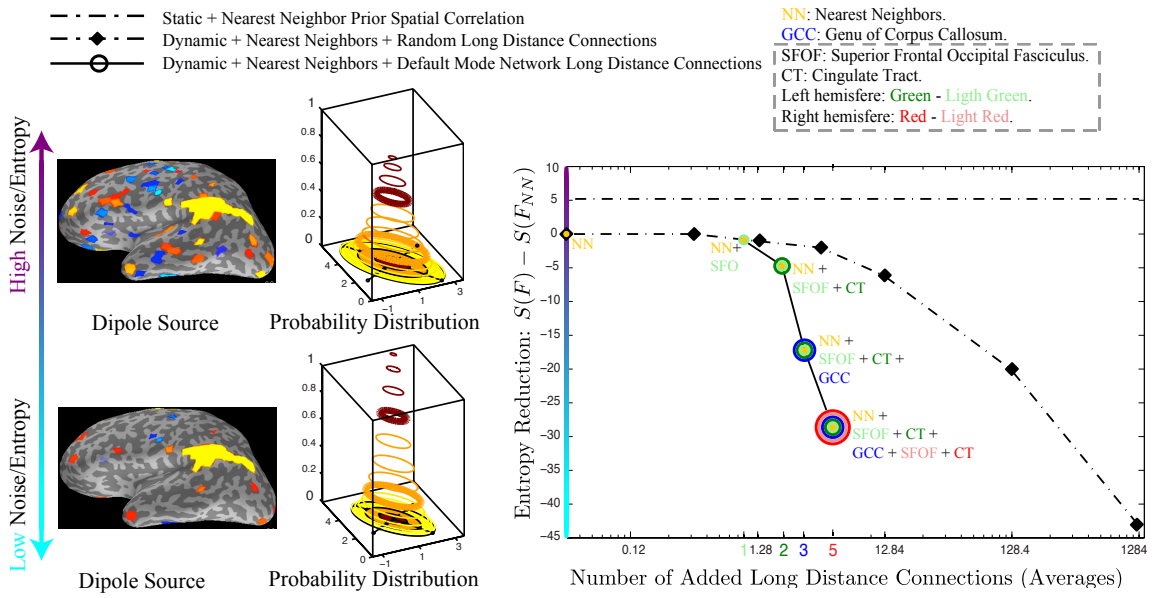

Figure 3-2: **Reduction in source estimation entropy by including dMRI long-distance connections.** The left panels show how high entropy estimates result in noisier estimates with larger confidence intervals. The right panel shows the decrease in the entropy of the estimate by the inclusion of short and long distance anatomical connection that support the Default Mode Network

by the theoretical analysis presented in Chapter 2, this improvement cannot be explained alone by the addition of connections (random connection model) or by the inclusion of spatial correlations (temporally independent model), but by the interaction of underlying spatial connections and dynamics. Therefore, in the following sections we will present a dynamic source localization algorithm, that not only includes these types of connections, but also exploits the spatially sparse nature of cortical activation and is orders of magnitude faster than a standard dynamic source localization algorithm such as dMAP-EM (See [76]).

3.3 Dimension reduction at different spatial scales using anatomically defined cortical parcellations and principal components analysis

A major challenge faced when developing accurate source estimation algorithms for MEG/EEG data (Eq. 3.2) is that the number of sensors is about one order of magnitude smaller than the number of sources to be estimated ($n << p$). In the last few years—and thanks to recent developments in the field of high-dimensional statistics and compressed sensing [87, 88]—several source localization algorithms that explicitly models spatial sparsity have been proposed [21, 89, 75, 90]. These algorithms exploit the fact that, in most MEG/EEG experiments, the cortical areas that are active and effectively relate to the given experimental paradigm are indeed spatially sparse. These sparse active areas are usually formed by small set of focal active dipoles, or by small group of active cortical areas of different sizes. Based on this insight, these sparse source localization algorithms are design to search for sparse cortical activity of a relative size a much smaller that the number of dipoles ($a << p$).

The guarantees for these sparse modeling algorithms in term of selection of the correct variables, as well as to the reconstruction of the noise-free measured signal, are essentially tied to the correlations between the regressors, i.e, the columns of the measurement matrix $\mathbf{X}$ [87, 88]. One can imagine that it would difficult, or perhaps impossible, to correctly select a few truly active sources if their lead field vectors (columns of $\mathbf{X}$) where highly correlated with the lead field vectors of sources that are not active. These scenario can often occur in MEG/EEG when a group of sources located near each other are simultaneously active, or when only a portion of these nearby sources are active. Because of this, sparse reconstruction algorithms that work directly with the lead field matrix $\mathbf{X}$ do not perform well in those scenarios, and are thus better suited to experiments where the activation is thought to be generated by a few focal areas of activity.

To overcome this challenge, a recent approach in the MEG/EEG literature has

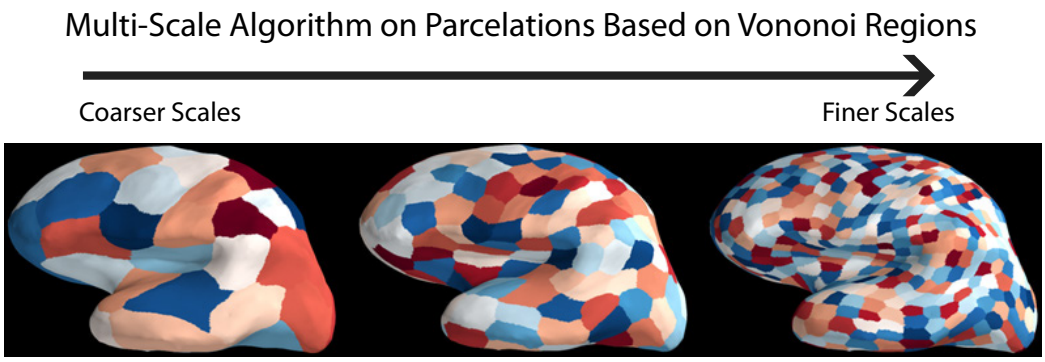

Figure 3-3: **Multi-Scale analysis using Voronoi regions.** The Multi-Scale algorithm traverses a hierarchy of spatial scales. In each scale, it produces an estimate that is then refined in the next scale. These scales are defined as the Voronoi regions that sequentially parcellate cortex. From left to right, the parcellations contain 84, 324, and 1284 regions.

been to both model and estimate sparse cortical activity at several spatial scales [91]. The spatial scales range from a coarser scale comprised of less than a hundred cortical areas to a finer spatial scale made of about a thousand regions. These scales are constructed as Voronoi regions that sequentially subdivide or parcellate cortex. The proposed algorithm traverses the hierarchy of scales, and in each scale produces a refined estimate of what was obtained in the precious coarser scale (Figure 3-3). Importantly, these algorithm operate on a source space of reduced dimension obtained in a manner similar to principal components regression. Specifically, at a given scale s, the cortex is parcellated into $|s|$ regions or areas (in this case the operator $|\cdot|$ indicates the number of regions and not the absolute value) yielding the observation model:

$$\bar{\mathbf{y}}_t = \sum_{i=1}^{|s|} \mathbf{X}_i^{(s)} \boldsymbol{\beta}_{i,t}^{(s)} + \boldsymbol{\varepsilon}_t, \tag{3.5}$$

where $\mathbf{X}_i^{(s)}$ and $\boldsymbol{\beta}_{i,t}^{(s)}$ are the columns of the lead field matrix $\mathbf{X}$ and the components of the source vector $\boldsymbol{\beta}_t$ corresponding to the ith cortical region at scale s, respectively.

To obtain a model of reduced dimension, it is common to use the low rank approximation of $\mathbf{X}_i^{(s)}$ given by its first few principal components [92, 91]. Concretely,

if $\mathbf{U}_i^{(s)}\text{diag}(\mathbf{d}_i^{(s)})\mathbf{V}_i^{(s)'}$ is the singular value decomposition of $\mathbf{X}_i^{(s)}$, the measured signal coming from the ith cortical area can be accurately approximated by:

$$\mathbf{X}_i^{(s)}\boldsymbol{\beta}_{i,t}^{(s)} \approx \mathbf{Z}_i^{(s)}\boldsymbol{\theta}_{i,t}^{(s)}, \tag{3.6}$$

where

$$\mathbf{Z}_i^{(s)} = [\mathbf{u}_{1,i}^{(s)}, \mathbf{u}_{2,i}^{(s)}, \ldots, \mathbf{u}_{p_i^{(s)},i}^{(s)}]\text{diag}(d_{1,i}^{(s)}, d_{2,i}^{(s)}, \ldots, d_{p_i^{(s)},i}^{(s)}) \tag{3.7}$$

is the new regression (gain) matrix composed of the first $p_i^{(s)}$ principal component of $\mathbf{X}_i^{(s)}$, and $\boldsymbol{\theta}_{i,t}^{(s)} \in \mathbb{R}^{p_i^{(s)}}$ are the parameters in the ith cortical area that we need to estimate. To make this approximation accurate, it is necessary to select enough principal components. To this end, it is recommended to select $p_i^{(s)}$ such that the representation accuracy exceeds 90%:

$${R_i^{(s)}}^2 = \frac{\sum_{j=1}^{p_i^{(s)}} {d_{j,i}^{(s)}}^2}{\sum_{\text{all } j} {d_{j,i}^{(s)}}^2} > 0.9. \tag{3.8}$$

As a result of the dimension reduction, as well as the representation over cortical areas, the measurement model becomes:

$$\bar{\mathbf{y}}_t = \sum_{i=1}^{|s|} \mathbf{Z}_i^{(s)}\boldsymbol{\theta}_{i,t}^{(s)} + \boldsymbol{\varepsilon}_t. \tag{3.9}$$

We should note that this representation not only reduces the dimensionality of the problem at any given scale, but also automatically ameliorates the issues related to the high correlations between the lead field vectors. This is because, in the new representation, the columns of $\mathbf{X}_i^{(s)}$, which represent nearby sources in a given cortical area, are now represented by a set of orthogonal vectors in $\mathbf{Z}_i^{(s)}$.

In contrast to the parcellations based on Voronio regions used in [91], we propose to used the hierarchy of gyral based anatomically defined parcellations 3-4 developed by the *Connectome Mapper* [93]. This parcellation is obtained by sequentially subdividing *freesurfer*'s parcellation [94]. One important difference between the Voronoi

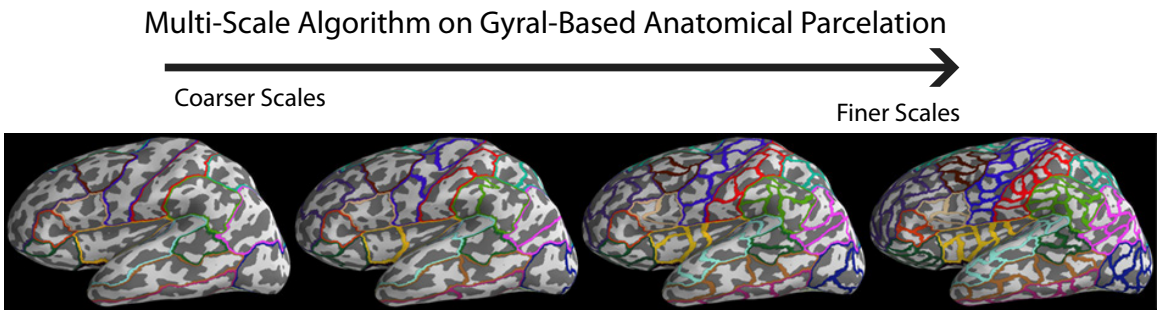

Figure 3-4: **Multi-Scale analysis using gyral-based parcellation**. The Multi-Scale traverses a hierarchy of spatial scales, and it sequentially refines the estimates found in a scale based on what it estimated in the previous coarser scale. These scales are defined via the sequential subdivision of a gyral-based anatomical parcellation. From left to right, the parcellations contain 72, 118, 219, and 452 regions.

based parcellations and the ones we use in the dynamic Multi-Scale analysis, is that source estimates derived from the latter could benefit from the naturally introduced anatomical constraints. Furthermore, since these gyral-based anatomical parcellations can be mapped between different subjects, they could provide a way to combine source localization results between different subjects.

3.4 Dynamic and sparse source estimation with the Kalman Filter based Group Subspace Pursuit algorithm (KF-GSP)

Converging lines of evidence suggest that dynamics are a central feature of brain function [40, 41, 42, 43, 44, 45, 46, 47]. Because of this, algorithms for source localization that explicitly model dynamics have been shown to dramatically improve source estimates since they attempt to mimic those spatiotemporal interactions [34, 35, 76](See Chapters 1 and 2 for detailed analysis and discussion). While some of these algorithms can yield spatially sparse source estimates [76], they do not explicitly model spatial sparsity nor do they take advantage of the fast algorithms and optimization techniques that are generally used in the high-dimensional and sparse modeling do-

mains. On the other hand, the fast techniques used for sparse source localization, such as [91], do not account for the source dynamics that are thought to be central to accurately model brain function.

Because of this, in this section we present a novel algorithm for dynamic and sparse source localization that we call the Kalman Filter based Group Subspace Pursuit algorithm (KF-GSP). This greedy algorithm form the basis for a Multi-Scale source localization analysis similar to the one presented in [91]. Specifically, the KF-GSP algorithm is used at each spatial scale s to estimate the brain's source activity. The algorithm is first applied in the coarsest spatial scale $s = 1$ (Figure 3-4, left-most panel) and takes the whole cortical mantle as candidate region to produce the source estimates. These estimates in turn define the candidate regions that the KF-GSP algorithm will consider in the next scale. This process is repeated until the finest spatial scale in reached (Figure 3-4, right-most panel). The details of how the algorithm traverses through the spatial scales are described in Section 3.5

At any given scale s, using the candidate areas $\mathcal{C}^s$ defined in the previous scale, the KF-GSP algorithm takes the reduced model corresponding to the candidate areas:

$$\bar{\mathbf{y}}_t = \sum_{i \in \mathcal{C}^s} \mathbf{Z}_i^{(s)} \boldsymbol{\theta}_{i,t}^{(s)} + \boldsymbol{\varepsilon}_t. \tag{3.10}$$

To simplify our notation, we will get rid or the reference to the scale, the candidate regions, and the summation in Equation (3.10) and compactly write it as: $\bar{\mathbf{y}}_t = \mathbf{Z}\boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t$. The algorithm is similar to the Subspace Pursuit (SP) and Ridge-Regression SP algorithms [95, 91] in that it is made of 4 main steps: 1) identification, 2) merging and estimation, 3) active set pruning, and 4) residual update. Our KF-GSP algorithm, however, performs each of these steps in a different way.

The algorithm is given a desired number of active variables a, a predefined correlation history and smoothness parameter ϕ, and an estimate of the inverse of the power signal-to-noise of the data $\lambda = 1/\text{SNR}^2$. It is then initialized ($i = 0$) with an empty active set $\hat{\mathcal{A}}^i = \{\}$, source parameter estimates set equal to zero, the residual equal to the data, and the residual sum of squares equal to the sum of square of the

data ($RSS^i = \sum_{t=1}^{T} ||\bar{\mathbf{y}}_t||_2^2$). Specifically, for each $t \in [1, \dots T]$:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_t &= \mathbf{0} \\ \hat{\boldsymbol{\varepsilon}}_t &= \bar{\mathbf{y}}_t. \end{aligned} \tag{3.11}$$

Then the algorithm iterates ($i = 0, 1, \dots$) through the 4 steps (See Alg. alg:kfgsp for a summary). First, is the identification step. The idea of this step is to use a computationally inexpensive technique to determine which variables can explain the current residual through time. To do this, we compute the correlation between each regressor and the residual at every point in time, and then square it:

$$r_{j,t}^2 = (\mathbf{z}_j' \hat{\boldsymbol{\varepsilon}}_t)^2. \tag{3.12}$$

We should point out that in the case that the regressors have unit norm, this quantity can be interpreted as a scaled coefficient of determination for regressing each variable independently on the residuals through time. We then average the $r_{j,t}^2$ values through time to obtain the averaged residual square correlation:

$$\bar{r}_j^2 = \frac{1}{T} \sum_{t=1}^{T} r_{j,t}^2. \tag{3.13}$$

The algorithm then selects the set of size a with the variables j that have the largest averaged residual square correlation, and merges these to the current active set $\hat{\mathcal{A}}^i$:

$$\hat{\mathcal{M}} = \hat{\mathcal{A}}^i \cup \{\text{set of size } a \text{ with the variables } j \text{ with the largest } \bar{r}_j^2\}. \tag{3.14}$$

Therefore, in the identification step we select the a variables that better explain the current residual—as measured by their correlation to the regressors—giving more importance to the residuals through time ($\hat{\boldsymbol{\varepsilon}}_t$) and regressors ($\mathbf{z}_j$) whose magnitude

Algorithm 4 The Kalman Filter in Information Form

Inputs: $\{\bar{\mathbf{y}}_t\}_{t=1}^T, \mathbf{Z}, \mathcal{S}, \phi, \lambda$

$\hat{\boldsymbol{\Sigma}}_{\mathcal{S}} \leftarrow \mathbf{Z}_{\mathcal{S}}'\mathbf{Z}_{\mathcal{S}}/n$ ▷ Initialization

$\boldsymbol{\theta}_{0|0} \leftarrow \mathbf{0}$

$\mathbf{V}_{0|0} \leftarrow \left(\frac{\lambda \text{tr}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}})}{n}\right)^{-1} \mathbf{I}$

for $t = 1, \dots T$ **do** ▷ Kalman's Recursion in Information Form

$\quad \boldsymbol{\theta}_{t|t-1} \leftarrow \phi \boldsymbol{\theta}_{t-1|t-1}$

$\quad \mathbf{V}_{t|t-1} \leftarrow \phi^2 \mathbf{V}_{t-1|t-1} + \frac{(1-\phi^2)}{\lambda \text{tr}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}})/n} \mathbf{I}$

$\quad \mathbf{V}_{t|t} \leftarrow \left(\mathbf{V}_{t|t-1}^{-1} + \hat{\boldsymbol{\Sigma}}_{\mathcal{S}}\right)^{-1}$

$\quad \boldsymbol{\theta}_{t|t} \leftarrow \boldsymbol{\theta}_{t|t-1} + \frac{1}{n}\mathbf{V}_{t|t}\mathbf{Z}_{\mathcal{S}}' \left(\bar{\mathbf{y}}_t - \mathbf{Z}_{\mathcal{S}}\boldsymbol{\theta}_{t|t-1}\right)$

end for

Outputs: $\{\boldsymbol{\theta}_{t|t}\}_{t=1}^T$

is larger. We should emphasize that even though the set of variables j can be large, the computations involved in this step are inexpensive (pT inner product of vectors of size $n \sim 300$).

Then the algorithm uses the Kalman Filter on the current merged set $\hat{\mathcal{M}}$ to obtain a new estimate of the parameters $\hat{\boldsymbol{\theta}}_t$. Specifically, for $t \in [1, \dots T]$ (Alg. 4):

$$\{\hat{\boldsymbol{\theta}}_{\hat{\mathcal{M}},t}\}_{t=1}^T = \text{KalmanFilterInformationForm}(\{\bar{\mathbf{y}}_t\}_{t=1}^T, \mathbf{Z}, \hat{\mathcal{M}}, \phi, \lambda). \tag{3.15}$$

Since the size of the current merged set $\hat{\mathcal{M}}$ is equal to $2a$, the algorithm then performs a pruning step to pick the best a variables in the current merged set. To do this, from the current estimate it selects the variables with the a largest squared sum through time. That is,

$$\begin{aligned} i &= i + 1 \\ \hat{\mathcal{A}}^i &= \{j \in \hat{\mathcal{M}} : \bar{\theta}_j^2 \text{ is one of the } a \text{ largest variables}\}, \end{aligned} \tag{3.16}$$

where

$$\bar{\theta}_j^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\theta}_{j,t}^2. \tag{3.17}$$

At this point, the algorithm has obtained an updated active set of the desired size a. However, to check for convergence as well as to obtain a residual that reflects this pruned active set, we need to perform an additional Kalman Filtering step (Alg. 4). We should note that, in comparison to the previous Kalman Filter step, this time the Kalman recursion is executed over a set of a variables:

$$\{\hat{\boldsymbol{\theta}}_{\hat{\mathcal{A}}^i,t}\}_{t=1}^{T} = \text{KalmanFilterInformationForm}(\{\bar{\mathbf{y}}_t\}_{t=1}^{T}, \mathbf{Z}, \hat{\mathcal{A}}^i, \phi, \lambda). \tag{3.18}$$

With this updated estimate, the algorithm then computes the residual ($t \in [1, \ldots, T]$),

$$\hat{\boldsymbol{\varepsilon}}_t = \bar{\mathbf{y}}_t - \mathbf{Z}_{\hat{\mathcal{A}}^i}\hat{\boldsymbol{\theta}}_{\hat{\mathcal{A}}^i,t}, \tag{3.19}$$

yielding the updated residual sum of squares

$$RSS^i = \sum_{t=1}^{T} ||\hat{\boldsymbol{\varepsilon}}_t||_2^2. \tag{3.20}$$

The last step of the iteration consist of checking if the residual sum of squares did not decrease in relation to the previous iteration. If this is the case, the algorithm stops and returns the Kalman Filter estimate and active set of size a corresponding to the previous iteration. If not, the algorithm goes to the beginning to perform an additional iteration (Eq. 3.12). The KF-GSP procedure is summarized in Algorithm 5.

We should note that this algorithm is fast in practice since the steps involving the more computationally costly Kalman Filter are performed on active set of less than 100 variables. At the same time, the identification step involving the complete set of variables is computationally inexpensive. This is because it consists of several independent inner product computations. Lastly, it is worth emphasizing that the Kalman Filter, over any set $\mathcal{S}$, explicitly solves the following optimization problem for every $t \in [1, \ldots, T]$:

Algorithm 5 The Kalman Filter Group Subspace Pursuit

Inputs: $\{\bar{\mathbf{y}}_t\}_{t=1}^T, \mathbf{Z}, \phi, \lambda, a$
Initialization:
$i \leftarrow 0$
$\mathcal{A}^{(i)} \leftarrow \{\}$
$\hat{\boldsymbol{\theta}}_t \leftarrow \mathbf{0}$
$\hat{\boldsymbol{\varepsilon}}_t \leftarrow \bar{\mathbf{y}}_t$
$RSS^i \leftarrow \sum_{t=1}^T ||\hat{\boldsymbol{\varepsilon}}_t||_2^2$
repeat
  Identification:
  for $j = 1, \ldots, p$ **do**
    $\bar{r}_j^2 \leftarrow \frac{1}{T}\sum_{t=1}^T (\mathbf{z}_j' \hat{\boldsymbol{\varepsilon}}_t)^2$
  end for
  $\mathcal{I} \leftarrow \{\text{set of size } a \text{ with the variables } j \text{ with the largest } \bar{r}_j^2\}$
  Merging and estimation:
  $\hat{\mathcal{M}} \leftarrow \hat{\mathcal{A}}^i \cup \mathcal{I}$
  $\{\hat{\boldsymbol{\theta}}_{\hat{\mathcal{M}},t}\}_{t=1}^T \leftarrow \text{KalmanFilterInformationForm}(\{\bar{\mathbf{y}}_t\}_{t=1}^T, \mathbf{Z}, \hat{\mathcal{M}}, \phi, \lambda)$ ▷ Alg. 4
  $i \leftarrow i + 1$
  for $j \in \mathcal{M}$ **do**
    $\bar{\theta}_j^2 \leftarrow \frac{1}{T}\sum_{t=1}^T \hat{\theta}_{j,t}^2$
  end for
  Active set pruning and re-estimation:
  $\hat{\mathcal{A}}^i \leftarrow \{j \in \hat{\mathcal{M}} : \bar{\theta}_j^2 \text{ is one of the } a \text{ largest variables}\}$
  $\{\hat{\boldsymbol{\theta}}_{\hat{\mathcal{A}}^i,t}\}_{t=1}^T \leftarrow \text{KalmanFilterInformationForm}(\{\bar{\mathbf{y}}_t\}_{t=1}^t, \mathbf{Z}, \hat{\mathcal{A}}^i, \phi, \lambda)$ ▷ Alg. 4
  Residual update:
  $\hat{\boldsymbol{\varepsilon}}_t \leftarrow \bar{\mathbf{y}}_t - \mathbf{Z}_{\hat{\mathcal{A}}^i}\hat{\boldsymbol{\theta}}_{\hat{\mathcal{A}}^i,t}$
  $RSS^i \leftarrow \sum_{t=1}^T ||\hat{\boldsymbol{\varepsilon}}_t||_2^2$
until $RSS^i > RSS^{i-1}$
$\hat{\boldsymbol{\theta}}_{\setminus\hat{\mathcal{A}}^{i-1},t} \leftarrow \mathbf{0}$
Outputs: $\hat{\boldsymbol{\theta}}_{\hat{\mathcal{A}}^{i-1},t}\hat{\boldsymbol{\theta}}_{\setminus\hat{\mathcal{A}}^{i-1},t}$

$$\min_{\{\boldsymbol{\theta}_{\mathcal{S},\tau}\}_{\tau=1}^t} \frac{1}{n}\sum_{\tau=1}^t ||\bar{\mathbf{y}}_\tau - \mathbf{Z}_{\mathcal{A}}\boldsymbol{\theta}_{\mathcal{S},\tau}||_2^2 + \frac{\lambda \text{tr}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}})}{(1-\phi^2)n}\sum_{\tau=2}^t ||\boldsymbol{\theta}_{\mathcal{S},\tau} - \phi\boldsymbol{\theta}_{\mathcal{S},\tau-1}||_2^2 + \frac{\lambda \text{tr}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}})}{n}||\boldsymbol{\theta}_{\mathcal{S},1}||_2^2, \quad (3.21)$$

where $\hat{\boldsymbol{\Sigma}}_{\mathcal{S}} = \mathbf{Z}_{\mathcal{S}}'\mathbf{Z}_{\mathcal{S}}/n$ is the sample correlation of the considered regressors. It is important to note that the Kalman Filter recursively solves this optimization problems in a fast manner.

3.5 Linking spatial scales using local and long-range connectivity derived from diffusion MRI

In order to perform the Multi-Scale analysis, once we have obtained an estimate of the spatially sparse source activity with the KS-GSP algorithm over a given scale s, as well as the corresponding set of active cortical regions, we need to define the candidate regions that will be considered in the following spatial scale $(s+1)$. The approach used in [91] is to set as candidate regions those that match the current set of active cortical regions as well as their nearest-neighboring regions. We illustrate this process in the bottom panel of Figure 3-5. If a hypothetical active region is active in a scale s (solid orange), the candidate regions that the next scale considers are the light orange areas as well as the orange active regions. We can imagine that in a MEG/EEG experiment, if a cortical region is active, and it is connected to a neighboring region, it is possible that the second region is also active. Therefore, this strategy can correct for those cases in which the estimated active regions misses some of their nearest neighbors.

However, it is also possible to add distant areas that are structurally connected via white matter bundles, such as those that can be estimated from diffusion MRI(Fig. 3-1, left). We illustrate how the process of adding dMRI connected regions would work in the top panel of Figure 3-5. In this case, if a hypothetical active region is active in a scale s (solid orange), the candidate regions that the next scale considers are the light orange areas as well as the orange active regions. In contrast to the nearest-neighbor strategy, we can see that the diffusion MRI connected regions can span both hemispheres. As we saw in Section 3.2.2, the inclusion of such long distance connections can greatly reduce the source estimation error (entropy). In addition, a reason why this might improve the quality of our source estimates comes from neuroscience literature using other brain imaging modalities. Specifically, recent imaging studies provide evidence linking the brain's structural connectivity with its functional activation. For instance, structural connectivity derived from diffusion MRI exhibits a strong correspondence with resting-state fMRI connectivity[79], and have been used

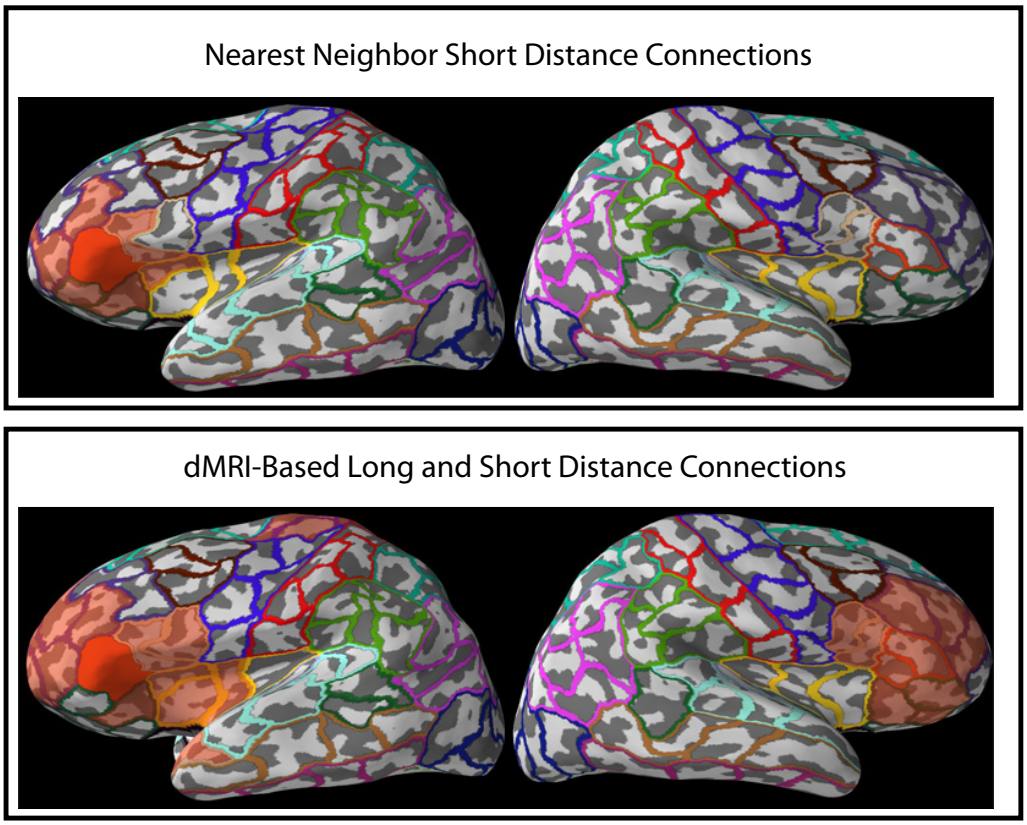

Figure 3-5: **Linking spatial scales with nearest-neighbor connections and connectivity derived from dMRI**. The top panels show the nearest neighbor regions of a cortical patch in the left hemisphere of the inferior frontal cortex (orange). The bottom panels show the distantly connected regions (derived from dMRI) of the same inferior frontal patch.

to predict resting-state functional connectivity[80] as well as fMRI activation in experimental task conditions[96]. This indicates that structural connectivity does carry information about functional activation and thus suggest that a Multi-Scale source localization approach could benefit from adding these distantly connected regions.

We propose a Multi-Scale approach that adds these two kinds of connected areas—the nearest-neighbors as well as distant regions that are structurally connected. We call the resulting algorithm the Multi-Scale KF-GSP (MS-KF-GSP). This algorithm first estimates activity in the coarsest scale ($s = 1$) via the KF-GSP algorithm by searching for the a_s regressor vectors that best explain the data. With the estimated source parameters and active regions, the algorithm then moves to the following finer

scale $s+1$. To do this, it determines which areas in the current parcellations correpond to the previously found active areas. Next, it add the regions that are connected to these active areas via diffusion MRI as well as the nearest-neighbor areas. This then forms the current set of candidate areas. The KF-GSP algorithm is then used to estimate the active source parameters in this candidate set. This process is repeated as the Multi-Scale analysis traverses through the spatial scales, until it reaches the finest scale (Fig. 3-6).

3.6 Model selection and hyper-parameter tunning via K-fold cross-validation

Source estimate obtained with the MS-KF-GSP algorithm depend on two sets of parameters: The number of active regions in each spatial scale a_s and the correlation history parameter ϕ that determines the temporal smoothness of source estimates. These parameters can be automatically selected from data in a principled way using K-fold cross-validation[97]. K-fold cross-validation is a general method for model selection and hyper-parameter tunning. It works by splitting the data into K parts of approximately equal size. Then, for a given combination of the hyper-parameter, the model or algorithm is fitted to the data with the jth partition of the data left aside. This process is repeated leaving out each of the K partitions. Then, each of the K fitted models is evaluated against the corresponding partition that was left out using a loss function, such as the deviance from a statistical model, and the results are averaged over the K partitions to produce a cross-validation (CV) score. This technique has been effectively used in several data analysis application, and most importantly, it has been shown that it is a generalization of methods for model selection such as Akaike's Information Criterion (AIC) [98].

In MEG/EEG analyses, K-fold cross-validation has not yet been used. This is probably because there is not an immediately natural way to partition the data into sets that are statistically independent. Specifically, if one leaves a portion of the time

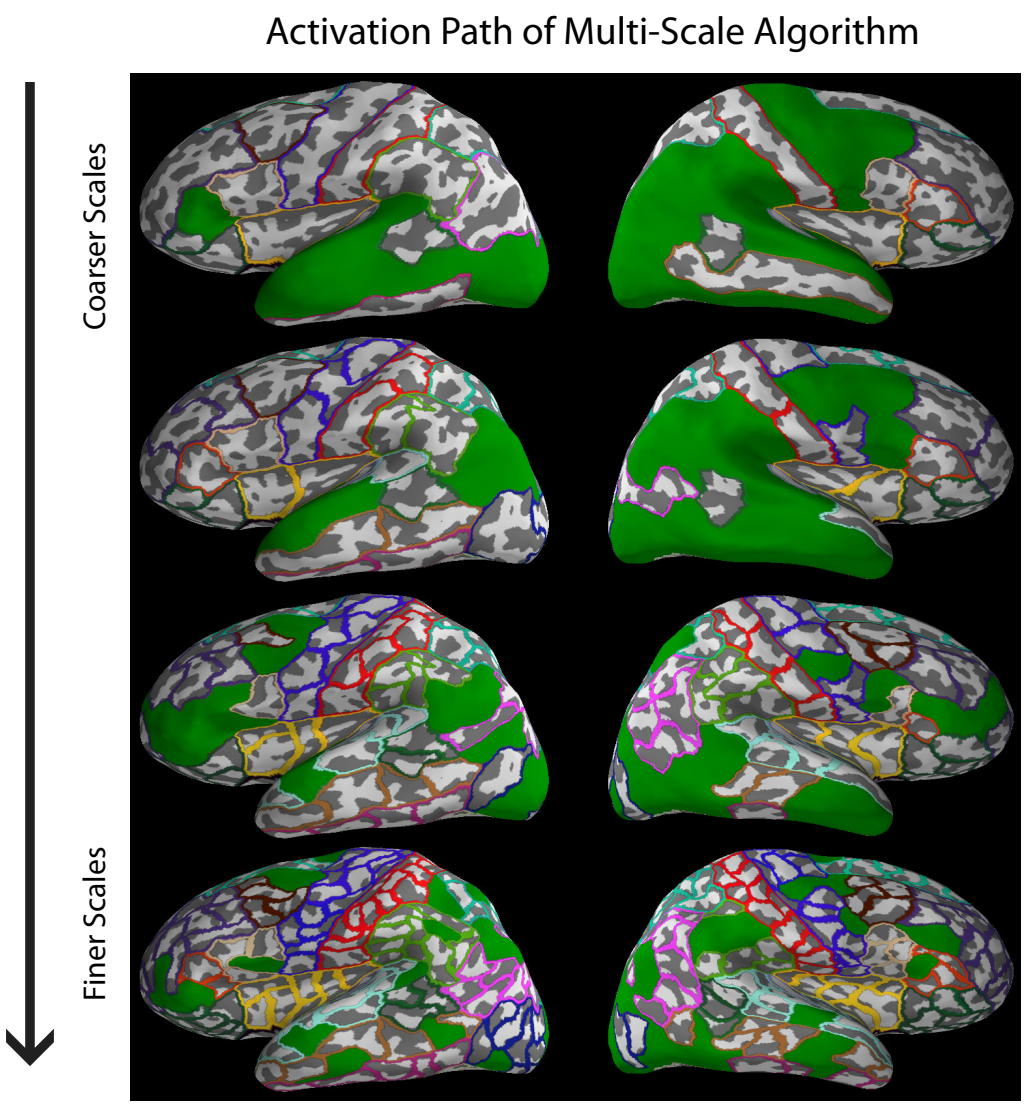

Figure 3-6: **Activation path of Multi-Scale algorithm**. The top panels show in green a hypothetical set of regions that the Multi-Scale algorithm finds active. As the algorithm traverses the spatial scales, from coarser (top) to finer (bottom) scales, it successively refines the the extent of the active areas.

samples in an evoked response $\bar{\mathbf{y}}_t$ or some of the MEG/EEG channels aside to perform the cross validation technique, the results might not be reliable. This is because the measured evoked responses across time and space are indeed highly dependent. One way to overcome this issue is to perform the partition of the data over the trials or epochs in an evoked response study. A reason why this approach could perform well is because, conditioning or fixing the source activity and for a given time instant t,

the measurement of a trial i is likely to be independent from a measurement in a different trial. The premise about the independence of the trials constitute the basis for averaging trials to obtain evoked responses.

In practice, the sizes of the active set a_s in each scale that we would like to evaluate range from 1 to 50. Because of this, it would be computationally prohibitive to perform K-fold cross-validation over 3 scales, since this would require the evaluation of the $\sim 50^3$ combination of parameters. Therefore, we propose to define a single parameter that encompasses the sparsity levels throughout all scales. To do this, we use the following simplification. If in any given scale, consisting of p_s candidate regressors (i.e., the number of columns in $\mathbf{Z}$ in the candidate set), the number of regressors necessary to explain the data is a_s, then the relative number of degrees of freedom used by the scale is $\text{df} = a_s/p_s$. It might then be desirable to fix *df* through the scales. This is because, as the algorithms goes from a coarser scale to a finer scale, the number of candidate regressors increase due to the fact that the number of modes or principal components necessary to approximate the the lead field matrices in the parcellation increases. Then, this might make it necessary to adjust a_s accordingly to allow the model to match the expressive power across the different scales. This heuristic, beside working well in practice, is effectively similar to the doubling heuristic proposed in [91].

To determine the hyper-parameters in our algorithm we propose the following K-fold cross validation scheme. The MEG/EEG signals are measured in m trials or epochs. We should recall that the measurement at time t and trial i is given by Equation (3.1) ($\mathbf{y}_t^{(i)} = \mathbf{X}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t^{(i)}$, $i \in [1, \ldots, m]$). We randomly partition the m trials into K groups of approximately equal size. Then, we compute an evoked response leaving out each group of trials $\mathcal{G}_j$ ($j \in [1, \ldots, K]$) one at the time. We then use the MS-KF-GSP algorithm to estimate the source activity until we reach the estimates in the finer spatial scale $\hat{\boldsymbol{\theta}}_t^{\setminus \mathcal{G}_j}$. With the estimated activity, we produce the predicted evoked response on the sensor space as $\mathbf{Z}\hat{\boldsymbol{\theta}}_t^{\setminus \mathcal{G}_j}$. At last, we produce the cross-validation score:

$$CV(\mathrm{df}, \lambda) = \frac{1}{mnT} \sum_{j=1}^{K} \sum_{i \in \mathcal{G}_j} \sum_{t=1}^{T} ||\mathbf{y}_t^{(i)} - \mathbf{Z}\hat{\boldsymbol{\theta}}_t^{\setminus \mathcal{G}_j}||_2^2. \tag{3.22}$$

3.7 Data description and preprocessing

In order to evaluate the proposed algorithm, whole head MEG data was recorded from a human subject using the 306-channel Neuromag Vectorview MEG system at Massachusetts General Hospital. The data was sampled at 1 kHz and band pass filter between 1 and 40 Hz. The gain matrix $\mathbf{X}$ was computed from the patient's high-resolution MRI using a 3-layer boundary element model with the mne-python package[99], the MNE software [1, 54], and the *freesurfer* software[52, 53]. The algorithm for multi-scale source localization (MS-KF-GSP) was coded in the *Python* programming language.

The MEG experimental paradigm consisted of a semantic decision task in which 160 English nouns were visually presented to the subject. The subject were asked to decide if the presented words were concrete or abstract, and to respond accordingly using different button presses. Each word was presented for 1 second, and the inter-stimulus interval was 2 seconds. The total recording time of ~ 8 minutes. The epochs (trials) were visually inspected for artifact resulting in $m = 105$ trials.

Diffusion MRI data was collected from the same subject. This data was process with the Connectome Mapper[93] using a diffusion tensor model and deterministc streamlining algorithm to produce connectivity matrices corresponding to the 4 spatial scales analyzed by our algorithm. In our algorithms, we consider two cortical region in a given parcellation to be connected if they are connected by any streamline fiber.

3.8 Cross-validation results

We used the proposed k-fold cross validation technique for model selection purposes, as well as to evaluate contribution of the connectivity information derived from diffu-

sion MRI in improving source estimates. Figure 3-7 show the K-fold cross-validation results. In general, it is recommended that K is set equal to 5 or 10[97]. We chose $K = 5$ as this seemed to provide estimates of the prediction error (CV) scores with less variability. The results that include both the nearest-neighbor as well as the connectome information (dMRI) are presented as solid lines. The dashed lines indicate that only nearest neighbor connectivity is included. We evaluated the sparsity parameter a_1 on a grid of values ranging from 10 to 50. We evaluated the temporal smoothness parameter ϕ taking the values 0.5 and 0.75. The results form cross-validation suggest that the inclusion of long distance connectome (dMRI) information yields smaller values of CV scores as well as more parsimonious models. Specifically, the inclusion of dMRI connectivity allow the models to achieve a small prediction error (CV score) at a smaller sparsity level. In the case of the smoothness parameter ϕ, it is difficult to appreciate a difference between the evaluated smoothness parameters. The effect of this parameter is appears more difficult to disambiguate.

3.9 Localization of language processing network

We estimated cortical activity using the MEG data described in Section 3.7 to study the language networks involved in the semantic decision task. The source estimates were obtained using a time interval between $250-550$ ms after stimulus. This interval was selected because the language network areas are prominently active during this period of time. The sparsity parameter a_1 was selected via 5-fold cross-validation as described in Section 3.8. Specifically, for the analysis that contains the connectome dMRI information, the sparsity parameter found via cross-validation was $a_1 = 17$. In the case where the algorithm only uses nearest neighbor information, the parameter found was $a_1 = 25$. Since cross-validation results for the parameter ϕ were not conclusive, we chose to set $\phi = 0.75$. Doing so produced smoother and less noisy cross-validation curves. The source localization results are shown in Figure 3-8. The top panels show the results that include both dMRI connectome information as well as nearest-neighbors, while the lower panels show results that only use nearest-neighbor

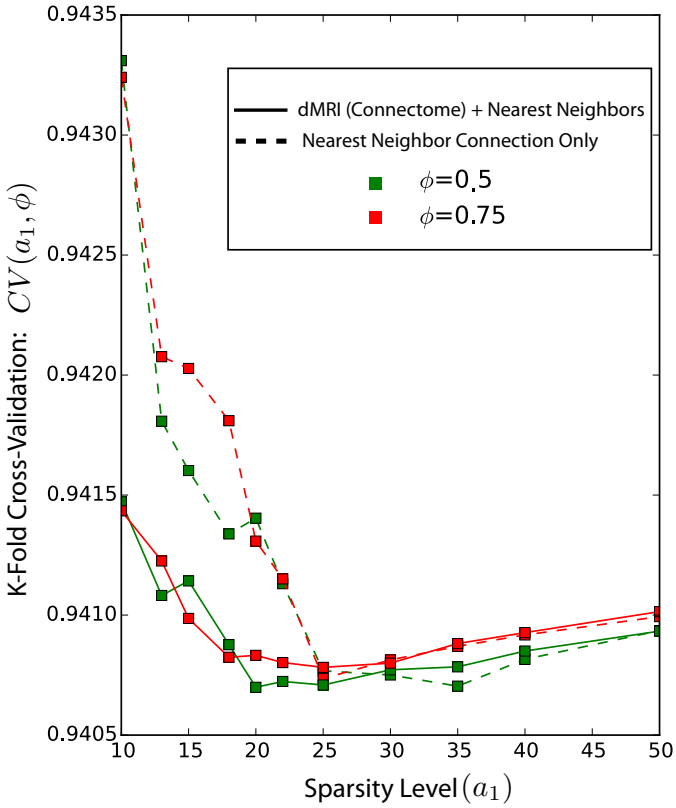

Figure 3-7: **k-fold cross validation for model hyper-parameters**. The dashed lines indicate the model used only neareast neighbor connection, while the solid line indicates the model uses both the connectivity derived from dMRI as well as nearest neighbors. The evaluated values of the temporal smoothness of correlation history parameter ϕ where $0.5, 0.75$. The x-axis shows the sparsity level a_1.

information. The absolute value of the estimated activation is shown in the red to yellow color scale. Estimate obtained with dMRI connectome information (Fig. 3-8, top) seem to cover more closely the areas that are thought to be involved in the semantic task, such us the opercular and triangular portions of the inferior frontal gyrus, superior temporal gyrus, and supramarginal gyrus in both hemispheres[100]. On the other hand, estimates obtained without the dMRI connectome seem to miss some of the relevant areas, including the supramarginal gyrus and the inferior frontal gyrus.

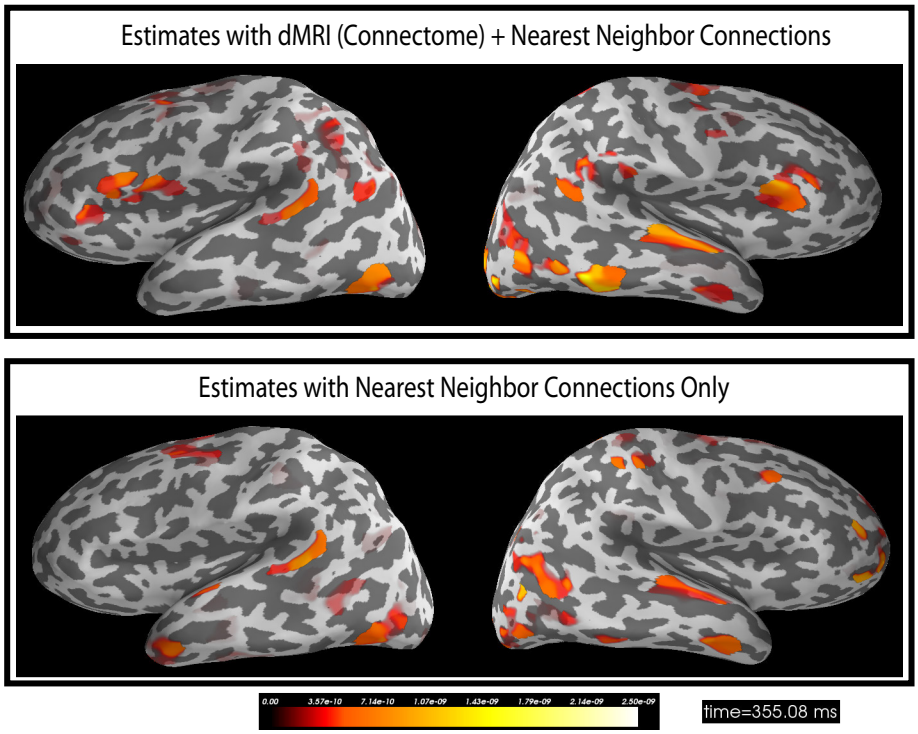

Figure 3-8: **Source localization results with the MS-KF-GSP algorithm in semantic processing task.**. The top panel show the source estimates obtained using the dMRI connectome information as well as nearest neighbors. The bottom panel shows the source estimation results when we only use the nearest neighbor information. The absolute value of the estimated source activity is shown in the red to yellow color scale. In both cases, the displayed sample correspond to 355 ms after the presentation of the word (visual stimulus). The estimates obtained using dMRI connectome information more closely detect the language areas that are believed to be involved in this task. In addition, the estimates that do not use dMRI information seem to produce a larger number of false positives.

3.10 Conclusions and discussion

In this work we demonstrated the importance of modeling and developing algorithms that include 1) spatial sparsity, 2) dynamics, 3) local nearest-neighbor connections, and 4) long distance connections derived from diffusion MRI in the MEG inverse problem. We showed with a simulation study as well as using MEG experimental data from a human subject the benefits of including these four modeling components in terms of the accuracy of source estimates. At the core of our result is the development of the fast, multi-scale, and dynamic greedy algorithm for source localization, namely, the

Multi-Scale Kalman Filter Based Group Subspace Pursuit (MS-KF-GSP) algorithm. Our algorithm traverses different spatial scales—from coarser to finer scales—and sequentially refines the spatial extent of the dynamic source estimates.

In our simulation study we investigated the importance of including the long distance connections that are thought to support the default mode network (DMN) in terms of the reduction of the entropy of source estimates. We showed that the reduction of the estimation error (entropy) when including the DMN connections could not be explained alone by the addition of connections (random connection model) or by the inclusion of spatial correlations (temporally independent model). Therefore, our result showed that the dramatic decrease in entropy result from the interaction of the spatial connection and dynamics in the source localization problem.

In order to make dynamic source localization feasible in practical applications, we developed the greedy and dynamic Kalman Filter based Group Subspace Pursuit algorithm (KF-GSP). Our algorithm exploits the fact that, in most MEG/EEG experimental scenarios, the spatial extent of brain activity is much smaller that the area representing the complete cortical mantle, i.e., that brain activity is spatially sparse. Our algorithm is a dynamic extension of the Ridge-Regression GSP [91]. The KF-GSP algorithm, as well as the Ridge-Regression GSP algorithm, greedily searches for cortical areas that can help explain the measured data. This is done by computing the correlation between the regressors corresponding to these cortical areas and the unexplained data residual. Because of these greedy computations have very low computational complexity, the algorithm is very fast and can obtain source estimates in seconds to minutes.

The main difference between our dynamic algorithm (KF-GSP) and the Ridge-Regression GSP lies on the way source estimates are obtained withing the estimated active sets. Whereas the the Ridge-Regression variant uses a static estimate that penalizes the L2-norm of the source parameters, our algorithm estimates source activity dynamically using the Kalman Filter. Because of this, our KF-GSP algorithm not only produce estimates that are temporally smooth, but also attempts to more closely mimic cortical activation dynamics that are thought to be a central feature

of brain function [40, 41, 42, 43, 44, 45, 46, 47]. Furthermore, our algorithm models both local nearest-neighbor as well as long distance connections derived from diffusion MRI. To do this, our Multi-Scale algorithm includes these two kinds of connections to determine the candidate activation areas as it traverses different spatial scales while refining the spatial extent of the estimated activity. Interestingly, our experimental source localization results in the semantic decision task highlight the importance of including these long distance connections: In the absence of these connection, the algorithm misses language areas that are thought to be active under these experimental conditions.

We develop a K-fold cross validation technique to estimate the model hyper-parameters as well as to perform model selection. The principal insight in developing this method was to recognize that the most appropriate way to split the data was across trials. By doing this, we could clearly determine the most appropriate size of the desired active set (sparsity level) in the MS-KF-GSP algorithm. In addition, the cross-validation technique allowed us to recognize that the inclusion of long distance connections produce more spatially parsimonious source estimates with equal or better predictive power.

Altogether, our results demonstrate for the first time the importance and clear improvements in terms of estimation accuracy that result from appropriately modeling dynamics, the brain's long and short distance structural connections, and spatial sparsity in the MEG/EEG inverse problem.

Acknowledgment

We would like to thank Dr. Steve Stufflebeam for providing us with the experimental MEG data, structural MRI, and dMRI data used in this project.

Bibliography

[1] M. S. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography – theory, intrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413–497, 1993.

[2] J. C. Mosher, R. M. Leahy, and P. S. Lewis, "EEG and MEG: Forward solutions for inverse methods," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 3, pp. 245–259, Mar. 1999.

[3] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 6, pp. 541–557, 1992.

[4] K. Uutela, M. S. Hämäläinen, and R. Salmelin, "Global optimization in the localization of neuromagnetic sources," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 6, pp. 716–723, 1998.

[5] C. Bertrand, M. Ohmi, R. Suzuki, and H. Kado, "A probabilistic solution to the MEG inverse problem via MCMC methods: the reversible jump and parallel tempering algorithms," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 533–542, 2001.

[6] E. Somersalo, A. Voutilainen, and J. P. Kaipio, "Non-stationary magnetoencephalography by Bayesian filtering of dipole models," *Inverse Probl*, vol. 19, no. 5, pp. 1047–1063, 2003.

[7] S. C. Jun, J. S. George, J. Paré-Blagoev, S. M. Plis, D. M. Ranken, D. M. Schmidt, and C. C. Wood, "Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data," *NeuroImage*, vol. 28, no. 1, pp. 84–98, Oct. 2005.

[8] S. J. Kiebel, J. Daunizeau, C. Phillips, and K. J. Friston, "Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG." *NeuroImage*, vol. 39, no. 2, pp. 728–741, Jan. 2008.

[9] C. Campi, A. Pascarella, A. Sorrentino, and M. Piana, "A Rao-Blackwellized particle filter for magnetoencephalography," *Inverse Probl*, vol. 24, no. 2, p. 025023, 2008.

[10] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Proc. Mag.*, vol. 18, no. 6, pp. 14–30, 2001.

[11] M. S. Hämäläinen and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain: Minimum norm estimates," *Med. Biol. Eng. Comput.*, vol. 32, no. 1, pp. 35–42, 1994.

[12] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–880, 1997.

[13] K. Uutela, M. S. Hämäläinen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *NeuroImage*, vol. 10, no. 2, pp. 173–180, Aug. 1999.

[14] T. Auranen, A. Nummenmaa, M. S. Hämäläinen, I. P. Jääskeläinen, J. Lampinen, A. Vehtari, and M. Sams, "Bayesian analysis of the neuromagnetic inverse problem with l(p)-norm priors." *NeuroImage*, vol. 26, no. 3, pp. 870–884, Jul. 2005.

[15] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, no. 1, pp. 49–65, Oct. 1994.

[16] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 231–251, Oct. 1995.

[17] C. Phillips, J. Mattout, M. D. Rugg, P. Maquet, and K. J. Friston, "An empirical Bayesian solution to the source reconstruction problem in EEG," *NeuroImage*, vol. 24, no. 4, pp. 997–1011, Feb. 2005.

[18] J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston, "MEG source localization under multiple constraints: An extended Bayesian framework," *NeuroImage*, vol. 30, no. 3, pp. 753–767, Apr. 2006.

[19] M.-a. Sato, T. Yoshioka, S. Kajihara, K. Toyama, N. Goda, K. Doya, and M. Kawato, "Hierarchical Bayesian estimation for MEG inverse problem," *NeuroImage*, vol. 23, no. 3, pp. 806–826, Nov. 2004.

[20] A. Nummenmaa, T. Auranen, M. S. Hämäläinen, I. P. Jääskeläinen, J. Lampinen, M. Sams, and A. Vehtari, "Hierarchical Bayesian estimates of distributed MEG sources: Theoretical aspects and comparison of variational and MCMC methods," *NeuroImage*, vol. 35, no. 2, pp. 669–685, Apr. 2007.

[21] D. P. Wipf and S. S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947–966, Feb. 2009.

[22] G. Buzsaki and A. Draguhn, "Neuronal oscillations in cortical networks," *Science*, vol. 304, no. 5679, pp. 1926–1929, 2004.

[23] S. Baillet and L. Garnero, "A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 5, pp. 374–385, 1997.

[24] F. Greensite, "The temporal prior in bioelectromagnetic source imaging problems," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1152–1159, 2003.

[25] J. Daunizeau, J. Mattout, D. Clonda, B. Goulard, H. Benali, and J. M. Lina, "Bayesian spatio-temporal approach for EEG source reconstruction: Conciliating ECD and distributed models," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 503–516, 2006.

[26] J. Daunizeau and K. J. Friston, "A mesostate-space model for EEG and MEG," *NeuroImage*, vol. 38, no. 1, pp. 67–81, Oct. 2007.

[27] K. J. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. J. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, Feb. 2008.

[28] N. J. Trujillo-Barreto, E. Aubert-Vázquez, and W. D. Penny, "Bayesian M/EEG source reconstruction with spatio-temporal priors." *NeuroImage*, vol. 39, no. 1, pp. 318–335, Jan. 2008.

[29] J. M. Zumer, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data." *NeuroImage*, vol. 41, no. 3, pp. 924–940, Jul. 2008.

[30] T. Limpiti, B. D. Van Veen, H. T. Attias, and S. S. Nagarajan, "A spatiotemporal framework for estimating trial-to-trial amplitude variation in event-related MEG/EEG." *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 633–645, Mar. 2009.

[31] W. Ou, M. S. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, Feb. 2009.

[32] A. Bolstad, B. D. Van Veen, and R. Nowak, "Space-time event sparse penalization for magneto-/electroencephalography," *NeuroImage*, vol. 46, no. 4, pp. 1066–1081, Jul. 2009.

[33] O. Yamashita, A. Galka, T. Ozaki, R. Biscay, and P. A. Valdes-Sosa, "Recursive penalized least squares solution for dynamical inverse problems of EEG generation," *Hum. Brain Mapp.*, vol. 21, no. 4, pp. 221–235, 2004.

[34] A. Galka, O. Yamashita, T. Ozaki, R. Biscay, and P. A. Valdes-Sosa, "A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering," *NeuroImage*, vol. 23, no. 2, pp. 435–453, 2004.

[35] C. J. Long, P. L. Purdon, S. Temereanca, N. U. Desai, M. S. Hämäläinen, and E. N. Brown, "State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing." *Ann. Appl. Stat.*, vol. 5, no. 2B, pp. 1207–1228, Jun. 2011.

[36] T. H. Bullock, M. C. McClune, J. Z. Achimowicz, V. J. Iragui-Madoz, R. B. Duckrow, and S. S. Spencer, "Temporal fluctuations in coherence of brain waves," *PNAS*, vol. 92, no. 25, pp. 11 568–11 572, Dec. 1995.

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

[38] D. A. Leopold, Y. Murayama, and N. K. Logothetis, "Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging," *Cereb. Cortex*, vol. 13, no. 4, pp. 422–433, Apr. 2003.

[39] P. L. Nunez, *Neocortical dynamics and human EEG rhythms*, 1st ed. New York: Oxford University Press, 1995.

[40] V. K. Jirsa, K. J. Jantzen, A. Fuchs, and J. A. S. Kelso, "Spatiotemporal forward solution of the EEG and MEG using network modeling," *IEEE Trans. Med. Ima.*, vol. 21, no. 5, pp. 493–504, 2002.

[41] J. J. Wright, C. J. Rennie, G. J. Lees, P. A. Robinson, P. D. Bourke, C. L. Chapman, E. Gordon, and D. L. Rowe, "Simulated electrocortical activity at microscopic, mesoscopic and global scales," *Int. J. Bifurcat. Chaos*, vol. 14, no. 2, pp. 853–872, 2004.

[42] P. A. Robinson, C. J. Rennie, D. L. Rowe, S. C. O'Connor, and E. Gordon, "Multiscale brain modelling," *Philos. Trans. R. Soc. B*, vol. 360, no. 1457, pp. 1043–1050, May 2005.

[43] O. David, L. M. Harrison, and K. J. Friston, "Modelling event-related responses in the brain," *NeuroImage*, vol. 25, no. 3, pp. 756–770, Apr. 2005.

[44] R. C. Sotero, N. J. Trujillo-Barreto, Y. Iturria-Medina, F. Carbonell, and J. C. Jimenez, "Realistically coupled neural mass models can generate EEG rhythms." *Neural Comput.*, vol. 19, no. 2, pp. 478–512, Feb. 2007.

[45] J. W. Kim and P. A. Robinson, "Compact dynamical model of brain activity," *Phys Rev E*, vol. 75, no. 3, p. 031907, 2007.

[46] E. M. Izhikevich and G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *PNAS*, vol. 105, no. 9, pp. 3593–3598, Mar. 2008.

[47] J. Gross, J. Kujala, M. S. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin, "Dynamic imaging of coherent sources: Studying neural interactions in the human brain," *PNAS*, vol. 98, no. 2, pp. 694–699, Jan. 2001.

[48] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *PNAS*, vol. 98, no. 2, pp. 676–682, 2001.

[49] D. A. Gusnard and M. E. Raichle, "Searching for a baseline: Functional imaging and the resting human brain," *Nat. Rev. Neurosci.*, vol. 2, no. 10, pp. 685–694, 2001.

[50] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anti-correlated functional networks," *PNAS*, vol. 102, no. 27, pp. 9673–9678, 2005.

[51] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nat. Rev. Neurosci.*, vol. 8, no. 9, pp. 700–711, Sep. 2007.

[52] A. M. Dale, B. R. Fischl, and M. I. Sereno, "Cortical surface-based analysis: I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, Feb. 1999.

[53] B. R. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system," *NeuroImage*, vol. 9, no. 2, pp. 195–207, 1999.

[54] M. S. Hämäläinen and J. Sarvas, "Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 2, pp. 165–171, Feb. 1989.

[55] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimationâĂŐ*, 1st ed. Upper Saddle River: Prentice Hall, 2000.

[56] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.*, vol. 82D, pp. 35–45, 1960.

[57] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.

[58] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Ser. Anal.*, vol. 3, no. 4, pp. 253–264, 1982.

[59] P. J. Green, "On use of the EM algorithm for penalized likelihood estimation," *J. R. Stat. Soc. Ser. B*, vol. 52, no. 3, pp. 443–452, 1990.

[60] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[61] P. de Jong and M. J. Mackinnon, "Covariances for smoothed estimates in state space models," *Biometrika*, vol. 75, no. 3, pp. 601–602, 1988.

[62] G. Kitagawa and W. Gersch, *Smoothness priors analysis of time seriesâĂŐ*. New York: Springer, 1996.

[63] B. D. O. Anderson and J. B. Moore, *Optimal FilteringâĂŐ*. Mineola, NY: Dover Publications, 2005.

[64] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processingâĂŐ*. Upper Saddle River, NJ: Prentice Hall, 2000.

[65] F. Darvas, D. Pantazis, E. Kucukaltun-Yildirim, and R. M. Leahy, "Mapping human brain function with MEG and EEG: methods and validation," *NeuroImage*, vol. 23 Suppl 1, pp. S289–99, 2004.

[66] J. M. Mendel, "Computational requirements for a discrete Kalman filter," *IEEE Trans. Autom. Control*, vol. 16, no. 6, pp. 748–758, Dec. 1971.

[67] J. Numminen, S. P. Ahlfors, R. J. Ilmoniemi, J. Montonen, and J. Nenonen, "Transformation of multichannel magnetocardiographic signals to standard grid form," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 72–78, 1995.

[68] R. Hari and R. Salmelin, "Human cortical oscillations: A neuromagnetic view through the skull," *Trends Neurosci.*, vol. 20, no. 1, pp. 44–49, 1997.

[69] N. U. Desai, "Souce localization for magnetoencephalography by spatio-temporal Kalman fitering and Fixed-Interval smoothing," Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2005.

[70] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston, "Dynamic causal modeling of evoked responses in EEG and MEG," *NeuroImage*, vol. 30, no. 4, pp. 1255–1272, May 2006.

[71] S. J. Kiebel, M. I. Garrido, R. Moran, C.-C. Chen, and K. J. Friston, "Dynamic causal modeling for EEG and MEG," *Hum. Brain Mapp.*, vol. 30, no. 6, pp. 1866–1876, 2009.

[72] P. J. Brockwell and R. A. Davis, *Time series*, 2nd ed., ser. Theory and methodsâĂŐ. New York: Springer, 2006.

[73] S. Roweis and Z. Ghahramani, "A unifying review of linear gaussian models," *Neural Comput.*, vol. 11, no. 2, pp. 305–345, 1999.

[74] A. Nummenmaa, T. Auranen, M. S. Hämäläinen, I. P. Jääskeläinen, M. Sams, A. Vehtari, and J. Lampinen, "Automatic relevance determination based hierarchical Bayesian MEG inversion in practice," *NeuroImage*, vol. 37, no. 3, pp. 876–889, Sep. 2007.

[75] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, and M. Kowalski, "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410–422, Apr. 2013.

[76] C. Lamus, M. S. Hämäläinen, S. Temereanca, E. N. Brown, and P. L. Purdon, "A spatiotemporal dynamic distributed solution to the MEG inverse problem." *NeuroImage*, vol. 63, no. 2, pp. 894–909, 2012.

[77] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear appoach," *J. Cogn. Neurosci.*, vol. 5, no. 2, pp. 162–176, 1993.

[78] S. P. Ahlfors, J. Han, J. W. Belliveau, and M. S. Hämäläinen, "Sensitivity of MEG and EEG to source orientation," *Brain Topogr.*, vol. 23, no. 3, pp. 227–232, Jul. 2010.

[79] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biol.*, vol. 6, no. 7, pp. 1479–1493, 2008.

[80] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *PNAS*, vol. 106, no. 6, pp. 2035–2040, Feb. 2009.

[81] M. P. van den Heuvel, R. C. W. Mandl, R. S. Kahn, and H. E. Hulshoff Pol, "Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain," *Hum. Brain Mapp.*, vol. 30, no. 10, pp. 3127–3141, Oct. 2009.

[82] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, 1st ed. New York: Springer, Aug. 2013.

[83] E. D. Sontag, *Mathematical control theory*, 2nd ed., ser. Deterministic finite dimensional systems. New York: Springer, Jul. 1998.

[84] B. D. O. Anderson, "Reverse-time diffusion equation models," *Stoch. Proc. Appl.*, vol. 12, no. 3, pp. 313–326, May 1982.

[85] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, "Resting-state functional connectivity reflects structural connectivity in the default mode network," *Cereb. Cortex*, vol. 19, no. 1, pp. 72–78, Apr. 2008.

[86] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, Oct. 2010.

[87] P. Bühlmann and S. A. van de Geer, *Statistics for high-dimensional data*, ser. Methods, theory and applications. New York: Springer, Jun. 2011.

[88] T. Hastie, R. Tibshirani, and M. J. Wainwright, *Statistical learning with sparsity*, 1st ed., ser. The Lasso and generalizations. Boca Raton: CRC Press, May 2015.

[89] D. Strohmeier, J. Haueisen, and A. Gramfort, "Improved MEG/EEG source localization with reweighted mixed-norms," in *International Workshop on Pattern Recognition in Neuroimaging*. Tubingen: IEEE, 2014, pp. 1–4.

[90] A. Gramfort, M. Kowalski, and M. S. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods." *Phys. Med. Biol.*, vol. 57, no. 7, pp. 1937–1961, Apr. 2012.

[91] B. Babadi, G. Obregon-Henao, C. Lamus, M. S. Hämäläinen, E. N. Brown, and P. L. Purdon, "A Subspace Pursuit-based Iterative Greedy Hierarchical solution to the neuromagnetic inverse problem." *NeuroImage*, vol. 87, pp. 427–443, Feb. 2014.

[92] T. Limpiti, B. D. Van Veen, and R. T. Wakai, "Cortical patch basis model for spatially extended neural activity." *IEEE Trans. Biomed. Eng.*, vol. 53, no. 9, pp. 1740–1754, Sep. 2006.

[93] A. Daducci, S. Gerhard, A. Griffa, A. Lemkaddem, L. Cammoun, X. Gigandet, R. Meuli, P. Hagmann, and J.-P. Thiran, "The Connectome Mapper: An open-source processing pipeline to map connectomes with MRI." *PLoS One*, vol. 7, no. 12, p. e48121, 2012.

[94] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006.

[95] W. Dai and O. Milenkovic, "subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, 2009.

[96] Z. M. Saygin, D. E. Osher, K. Koldewyn, G. Reynolds, J. D. E. Gabrieli, and R. R. Saxe, "Anatomical connectivity patterns predict face selectivity in the fusiform gyrus," vol. 15, no. 2, pp. 321–327, Feb. 2012.

[97] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd ed., ser. Data mining, inference, and prediction. New York: Springer, 2009.

[98] B. Efron, "The estimation of prediction error," *J. Am. Stat. Assoc.*, vol. 99, no. 467, pp. 619–632, Sep. 2004.

[99] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014.

[100] N. Tanaka, H. Liu, C. Reinsberger, J. R. Madsen, B. F. Bourgeois, B. A. Dworetzky, M. S. Hämäläinen, and S. M. Stufflebeam, "Language lateralization represented by spatiotemporal mapping of magnetoencephalography," *American Journal of Neuroradiology*, vol. 34, no. 3, pp. 558–563, Mar. 2013.