NAIST-IS-DD1161010

Doctoral Dissertation

State-Space Methods for Reconstructing Neuronal Current Sources

Makoto Fukushima

March 13, 2014

Department of Information Science
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Makoto Fukushima

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Yuji Matsumoto	(Co-supervisor)
Professor Mitsuo Kawato	(Co-supervisor)
Doctor Masa-aki Sato	(ATR-NIA, Co-supervisor)
Doctor Okito Yamashita	(ATR-NIA, Co-supervisor)

State-Space Methods for Reconstructing Neuronal Current Sources*

Makoto Fukushima

Abstract

Elucidating mechanisms of how functionally specialized brain regions dynamically interact has recently received attention in the neuroimaging community. Such dynamic integration of functional brain regions can be investigated by Magnetoencephalography (MEG) and Electroencephalography (EEG). To discover functional brain networks from MEG/EEG sensor measurements, it is indispensable to properly reconstruct neuronal current sources from these data and identify directed interactions (i.e., effective connectivity) between the current sources.

State-space approaches for MEG/EEG source reconstruction potentially provide ways to solve the above estimation problems. The state-space framework can incorporate *a priori* knowledge on neuronal current dynamics into the dynamic model of current sources. Imposing realistic priors on the source dynamics allows reconstructing current sources from MEG/EEG data more accurately. The richness of the prior assumptions also contributes to identification of functional brain networks. This can be achieved by first introducing model parameters of the source interactions based on prior knowledge, and then estimating these parameters from the measurements.

In this thesis, to realize accurate source reconstruction and discovery of functional brain networks, two novel extensions on state-space methods are applied. First, a limitation of previous state-space methods in reconstructing

*Doctoral Dissertation, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1161010, March 13, 2014.

spatially focal current sources has been resolved. By replacing spatially homogeneous dynamic source model in existing methods to spatially inhomogeneous one, focal current sources are successfully reconstructed under the state-space framework for the first time. Second, inference of functional brain networks has became available by incorporating long-range directed interactions into the dynamic source model, under prior knowledge on structural brain connectivity. The new state-space method extends previous dynamic models in which spatially local (or self) source interactions are only assumed and from which the functional networks cannot be identified.

Keywords:

MEG/EEG, Source reconstruction, State-space model, Neuronal dynamics, Prior knowledge, Effective connectivity

Acknowledgements

The work in my Ph.D. study was carried out at the Neural Information Analysis Laboratories, Advanced Telecommunications Research Institute International (ATR-NIA). First of all, I would like to thank my adviser Dr. Masa-aki Sato, Head of ATR-NIA, for his guidance in this study. By leaning a great number of (but only a small part of) his notes on mathematical backgrounds, I have earned skills required for developing the "dynamic" source reconstruction methods. I am also very grateful to Dr. Okito Yamashita, Head of CBI in ATR-NIA, for his helpful advices on this study. The thesis would not be possible without his encouragement and enthusiasm on our research topics.

Prof. Mitsuo Kawato, Director of ATR-BICR, provides me insightful suggestions throughout my doctoral course. I would like to extend my gratitude to his valuable (and critical) comments on my research, which certainly improved quality of the work in this thesis.

I would like to thank my thesis committee Prof. Kazushi Ikeda (NAIST), Prof. Yuji Matsumoto (NAIST), Prof. Mitsuo Kawato (ATR-CNS/NAIST), Dr. Masa-aki Sato (ATR-NIA), and Dr. Okito Yamashita (ATR-NIA) for their feedbacks on my research. It should be mentioned that Prof. Ikeda gave me helpful comments on my research, as well as an opportunity to talk my study at the regular seminar of the Mathematical Informatics Laboratory in NAIST. Prof. Matsumoto provided fruitful comments on the work of this thesis at the mid-term thesis presentation and the thesis defense.

The short visit at Max Planck Institute for Human Cognitive and Brain Sciences (MPI-CBS) for three months allows incorporating diffusion MRI data into the dynamic source model. I would like to thank Dr. Thomas Knösche, Group leader of CNCF in MPI-CBS, for his support during my stay and for research discussions at 15 times of Skype sessions after I came back to Japan.

I am grateful to Prof. Shin Ishii (Kyoto Univ./ATR-CMC) and Dr. Junichiro Yoshimoto (OIST/NAIST) for discussing technical details about variational Bayesian inference of state-space models. I would also like to thank Prof. Yukiyasu Kamitani (ATR-CNS/NAIST) and Prof. Tomohiro Shibata (Kyutech/NAIST) for their constructive comments on my research at the ATR student seminar.

During pursuing my doctoral course study, I was helped by many other people in ATR and NAIST. I would like to thank especially the following people for their helpful support and comments at ATR-NIA: Dr. Taku Yoshioka, Nobuo Hiroe, Dr. Yusuke Takeda, Dr. Takeaki Shimokawa, Dr. Atsunori Kanemura, Takao Sako, Ryosuke Hayashi, Noboru Nushi, and Noriko Masui; and for those at NAIST: Dr. Takashi Takenouchi, Dr. Kazuho Watanabe, Dr. Takatomi Kubo, Aya Tanimoto, and Satomi Adachi. I should also thank my academic colleagues, in particular Dr. Tomoyasu Horikawa, Kei Majima, Shigeyuki Ikeda, Satoshi Kozawa, Takashi Kosaka, and Ryota Aisu for useful discussions.

This research was supported by contracts with the National Institute of Information and Communications Technology entitled, 'Multimodal integration for brain imaging measurements' and 'Development of network dynamics modeling methods for human brain data simulation systems.' The study in this thesis was supported also by the Japan Society for the Promotion of Science through Grant-in-Aid for JSPS Fellows (23-3907).

Finally I would like to thank my family for their support in all my life.

Contents

List of Figures

List of Tables

Notational Conventions

Symbol	Definition
$P(x)$	Probability distribution of x
$P(x \mid y)$	Conditional probability distribution of x given y
$\mathcal{N}(x \mid \bar{x}, \sigma)$	Univariate Gaussian distribution over x with mean $\bar{x}$ and variance σ
$\mathcal{N}(\mathbf{x} \mid \bar{\mathbf{x}}, \mathbf{\Sigma})$	Multivariate Gaussian distribution over $\mathbf{x}$ with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{\Sigma}$
$\mathcal{G}(x \mid \bar{x}, \gamma)$	Gamma distribution over x with mean $\bar{x}$ and the shape parameter γ (the scale parameter corresponds to $\bar{x}\,\gamma^{-1}$)
$\Gamma(x)$	Gamma function of x
$\psi(x)$	Digamma function of x
$\langle x \rangle_{P(X)}$	Expectation value of x with respect to $P(X)$
$(\mathbf{x})_C$	A vector consisting of a subset of $\mathbf{x}$ with its indices included in a set C in ascending order
$(\mathbf{X})_{C,C}$	A matrix consisting of a subset of $\mathbf{X}$ with its row and column indices both included in C
$(\mathbf{X})_{k,l}$	The k,l-th entry of a matrix $\mathbf{X}$
$\mathrm{diag}(\mathbf{x})$	A diagonal matrix with its diagonal entries $\mathbf{x}$
$\mathrm{tr}(\mathbf{X})$	Trace of a matrix $\mathbf{X}$
$1\!:\!N$	A set of constants $1, \ldots, N$
$x_{1:N}$	A set of x with subscripts from 1 to N (i.e., $x_1, \ldots, x_N$)

Chapter 1

Introduction

1.1. MEG and EEG: Functional brain imaging with high temporal resolution

There are two fundamental functional principles of the brain: functional specialization and functional integration [23,24]. Identifying functionally specialized brain regions (e.g., for sensory processing, motor control, and cognitive processing) has been a long-term focus of neuroimaging studies. However, for a true understanding of the mechanisms underlying brain function, elucidating the scheme of dynamic integration between these functionally specialized brain regions is indispensable. This topic has received growing interest in recent years [47].

Magnetoencephalography (MEG) and electroencephalography (EEG) provide ways to investigate such dynamic integration of brain functions [65,74], because of their high temporal resolution and large reflection of neuronal electrical activity [37,60]. The richness of the temporal information in MEG/EEG allows capturing temporal propagation, or event-related dynamics, of neuronal activity occurring over millisecond time scales, which cannot be easily detected by slow blood oxygenation level dependent (BOLD) signals of functional magnetic resonance imaging (fMRI).

1.2. Reconstruction of neuronal current sources from MEG/EEG data

In contrast to the excellent temporal resolution, the spatial resolutions of MEG and EEG are limited; the spatial distribution of neuronal current sources cannot be uniquely determined from the measurements, unless *a priori* knowledge or assumptions are imposed as constraints on current sources [3]. Numerous source reconstruction methods have been proposed with various constraints: dipole methods assuming a small number of focal sources [56, 72], distributed source methods with purely spatial constraints [25, 38, 67, 71, 83], and distributed source methods with both spatial and temporal constraints [2,7,13,62,73,80]. Throughout this thesis, we employ distributed source modeling for formulation of the source reconstruction problem (details are shown in Chapter 2).

1.3. State-space methods for source reconstruction

In source reconstruction, introducing prior constraints on the spatiotemporal dynamics of source activities is of particular interest; this type of constraint complements other commonly used constraints (typically spatial) and introduces additional knowledge into the source reconstruction process, for example, on dynamic properties of neuronal populations, structural connections between brain areas, and transmission delays of neuronal activities. This knowledge potentially facilitates the extraction of information on directed interactions (i.e., effective connectivity) between sources, while reconstructing spatial source distributions from MEG/EEG data. The spatiotemporal dynamics reflects the generative nature of neuronal current sources, and is readily incorporated into a state-space representation. To formulate such dynamics, previous state-space methods have adopted linear autoregressive models with spatially local interactions [30, 51] and self-interactions [11, 87]. These methods extend an approach that imposes a simple prior assumption (such as a temporal smoothness prior in [73]) on the source dynamics (the effectiveness of imposing simple temporal smoothness is critically evaluated by [9]).

1.4. Extensions of state-space methods

In principle, dynamic integration of functional brain regions can be clarified by applying state-space methods in the context of MEG/EEG source reconstruction. However, limitations of previous state-space methods in source reconstruction accuracy and source dynamics modeling make the estimation of functional brain networks impossible. In this thesis, to overcome these problems, the reconstruction accuracy of state-space methods is first improved by extending the dynamic source model, with special emphasis on reconstructing spatially focal source activity. The dynamic model of current sources is then further extended to incorporate long-range source interactions, in order to allow identifying whole-brain functional brain networks.

1.4.1 Reconstruction of focal current sources (Chapter 3)

Previous state-space methods have not paid attention to reconstructing spatially focal current sources. Spatially focal solutions are important because they are consistent with the notion of functional specialization, which states that cortical activities specialized to specific stimulus attributes or experimental contexts are restricted to small regions [22]. In particular, these solutions provide good matches to brain activities evoked by external triggers such as visual, auditory, and somatosensory stimuli. To reconstruct such spatially focal current sources based on the state-space framework, we introduce a new model of the source dynamics in Chapter 3, by taking its spatially inhomogeneous temporal evolution into account, departing from a simplifying assumption that all current sources follow the same temporal evolution over the entire brain [11] or over a given cortical area [30, 87].

1.4.2 Reconstruction of sources with identifying long-range interactions (Chapter 4)

Representing spatially local interactions or self-interactions by simple autoregressive modeling [11,30,51,87] is insufficient to discover the functional brain networks. To identify all source interactions on these functional networks,

we therefore newly incorporate long-range interactions into the autoregressive modeling of source dynamics. A similar approach has been proposed by [61] using a full multivariate autoregressive (MAR) model. However, in this model, the spatiotemporal dynamics was formulated in a low-dimensional latent space rather than in the high-dimensional source space. In Chapter 4, we parameterize long-range interactions directly in a source space full MAR model, with help of prior knowledge of structural brain connectivity. Estimating parameters on this model allows identifying the functional brain networks in detail under a state-space framework.

1.5. Thesis organization

In Chapter 2, basic formulations of source reconstruction and state-space models, as well as estimation algorithms of current sources and model parameters, are described. Then, the extensions of state-space methods overviewed above are detailed in Chapters 3 and 4. Finally, contributions and future directions of the work in this thesis are presented in Chapter 5.

Chapter 2

Background

2.1. Model formulation

2.1.1 MEG/EEG source reconstruction

The distributed source approach of MEG/EEG source reconstruction is formulated as a linear inverse problem. In this problem formulation, the following linear relationship is provided between MEG/EEG sensor measurements and cortical source distributions [37]:

$$\mathbf{B}_t = \mathbf{G}\mathbf{J}_t + \varepsilon_t, \tag{2.1}$$

where $\mathbf{B}_t$ and $\mathbf{J}_t$ denote the measurement data[1] and current source activities, respectively, $\mathbf{G}$ is the lead field matrix (constant), ε_t is the observation noise, and t is an index of time samples (where $t \leq T$). $\mathbf{B}_t$ and $\mathbf{J}_t$ are M- and N-dimensional vectors that vertically concatenate the single channel data $B_{m,t}$ with $m \in \{1:M\}$ and the single source activity $J_{n,t}$ with $n \in \{1:N\}$, respectively. $\mathbf{G}$ is obtained by solving the forward problem [57], accounting for the structure of the human head.

To achieve source reconstruction with spatial resolution of several millimeters, intensities of thousands of current sources are necessary to be estimated from only hundreds of sensor data ($N > M$). Thus reconstructing current

[1] In the following chapters, we only assume MEG.

sources (i.e., estimating $\mathbf{J}_t$ with given $\mathbf{B}_t$ and $\mathbf{G}$) in such a high spatial resolution is an ill-posed inverse problem, where no unique solution is provided unless any constraint is imposed on the current sources.

The key to high accuracy in reconstructing current sources is how well prior assumptions of sources behind the constraint match real source characteristics (both in spatial and temporal). By reflecting a variety of real properties of current sources into prior assumptions, a great number of source reconstruction methods have been proposed for over the last two decades [3].

Norm Regularization

In a classical approach, imposing a constraint on current sources is implemented by norm regularization. Under this framework, reconstructed sources are obtained by minimizing the cost function

$$E = \sum_{t=1}^{T} \Big\{ (\mathbf{B}_t - \mathbf{G}\mathbf{J}_t)^{\mathrm{T}} (\mathbf{B}_t - \mathbf{G}\mathbf{J}_t) + \lambda g(\mathbf{J}_t) \Big\}, \tag{2.2}$$

where $g(\mathbf{J}_t)$ denotes a norm regularization term, which works as a constraint. The constant λ controls the strength of regularization. The simplest and most popular method uses $g(\mathbf{J}_t)$ as the $l2$-norm of $\mathbf{J}_t$ [38], where the total energy of current sources is regulated. Another example employs Laplacian of $\mathbf{J}_t$ as norm regularization, imposing a spatial smoothness prior on current sources [67]. The $l1$-norm of $\mathbf{j}_t$ serves as a constraint as well [55,81]. By using $l1$-norm regularization, one can reconstruct sources with a prior assumption of spatial sparse neuronal activities.

In addition, norm regularization methods can incorporate a temporal constraint into the regularization term. [73] regularizes differences between $\mathbf{J}_{t-1}$ and $\mathbf{J}_t$ to account for temporal smoothness of current sources. A more sophisticated method combines spatial sparseness and temporal smoothness priors [62]. A spatiotemporal $l1l2$-norm regularizer is adopted in this method, where $l1$-norm regularization is applied to a spatial domain while $l2$-norm regularization is imposed on a temporal domain (coefficients of temporal basis functions).

Bayesian Approach

Source reconstruction can be formulated within a Bayesian framework. The goal of Bayesian source reconstruction is to compute a posterior distribution of current sources using the well-known Bayes theorem:

$$\prod_{t=1}^{T} P(\mathbf{J}_t \mid \mathbf{B}_t) = \prod_{t=1}^{T} \frac{P(\mathbf{B}_t \mid \mathbf{J}_t) P(\mathbf{J}_t)}{P(\mathbf{B}_t)}. \tag{2.3}$$

If one assume that the likelihood function $P(\mathbf{B}_t \mid \mathbf{J}_t)$ and the prior distribution $P(\mathbf{J}_t)$ are linear Gaussian with a spherical noise covariance matrix, then the mean of the posterior distribution $P(\mathbf{J}_t \mid \mathbf{B}_t)$ becomes equivalent to the inverse solution with the $l2$-norm regularization [38]. The Bayesian approach imposes a constraint on current sources via the prior distribution. This allows incorporating prior assumptions on current sources in a more intuitive manner than employing a norm regularization term.

An example of the prior distribution on current sources is the automatic relevance determination (ARD) prior [71]. Similar to the $l1$-norm regularizer in [55, 81], the ARD prior imposes spatial sparseness on the inverse solution, which provides spatially focal source estimation. Variants of ARD-based source reconstruction have been recently proposed in, e.g., [25, 83]. Another example of source prior is temporal smoothness. This prior is incorporated into [2, 7, 13, 80] by method-specific forms.

Importantly, the Bayesian approach can combine knowledge of spatial brain activity patterns from functional magnetic resonance imaging (fMRI) data [8, 12, 42, 63, 71]. Integration of fMRI with MEG in source reconstruction improves the reliability of estimated spatial patterns of source activity, owing to relatively higher spatial resolution of fMRI (in the order of millimeters) than MEG.

In this thesis, source reconstruction is performed from the Bayesian approach. The observation noise $\boldsymbol{\varepsilon}_t$ in Eq. (2.1) is assumed to follow a Gaussian distribution $\mathcal{N}(\boldsymbol{\varepsilon}_t \mid \mathbf{0}, \beta^{-1}\mathbf{S})$, where β is a scaling parameter and $\mathbf{S}$ is the noise covariance matrix scaled by β ($\mathbf{S}$ is normalized to satisfy $\mathrm{tr}(\mathbf{S}) = M$ and is typically determined from pre-stimulus rest period measurements). Then the likelihood function for the assumed Gaussian observation noise can be writ-

ten by

$$P(\mathbf{B}_t \mid \mathbf{J}_t, \beta) = \mathcal{N}\left(\mathbf{B}_t \mid \mathbf{G}\mathbf{J}_t, \beta^{-1}\mathbf{S}\right). \tag{2.4}$$

2.1.2 State-space modeling

State-space modeling for MEG/EEG source reconstruction provides a framework of imposing prior assumptions on the spatiotemporal source dynamics. The state-space model consists of the observation process of sources from sensors (Eq. (2.1)), together with an equation that specifies the dynamics of neuronal current sources. Choices on which dynamic properties of sources are modeled in this equation determine which dynamic phenomena can be extracted from model parameters to be estimated from the data.

A general form of source dynamics can be represented by

$$\mathbf{J}_t = f(\mathbf{J}_{1:t-1}) + \boldsymbol{\omega}_t, \tag{2.5}$$

where the first term $f(\mathbf{J}_{1:t-1})$ is $\mathbf{J}_t$ predicted from $\mathbf{J}_{1:t-1}$ via a dynamic source model f. The second term $\boldsymbol{\omega}_t$ denotes the system noise; the input term to this dynamic model. By assuming a dynamic source model as a prior constraint on current sources, one can exploit past data samples in reconstructing the present state of current sources via the model prediction. Moreover, by parameterizing the dynamics f and estimating these model parameters from measurement data, latent factors specifying the brain dynamics (e.g., dynamic interactions between current sources) can be extracted.

The simplest dynamic model, proposed in [73], is

$$\mathbf{J}_t = \mathbf{J}_{t-1} + \boldsymbol{\omega}_t, \tag{2.6}$$

whose prior assumption on sources corresponds to the temporal smoothness prior. Similar linear source dynamics incorporating temporal smoothness are proposed in [2, 11, 13]. The dynamic model in Eq. (2.6) is extended by [51] to account for temporal smoothness between neighboring sources, where the model can be written as

$$\mathbf{J}_t = \mathbf{F}\mathbf{J}_{t-1} + \boldsymbol{\omega}_t, \tag{2.7}$$

however the matrix $\mathbf{F}$ is constant and its entries for other than neighboring source interactions are set to zero.

The ideal source modeling for extracting knowledge of directed source interactions on functional brain networks is to assume

$$\mathbf{J}_t = \mathbf{A}\,\mathbf{J}_{t-1} + \boldsymbol{\omega}_t, \tag{2.8}$$

where the MAR matrix $\mathbf{A}$ is a model parameter to be estimated from the data. Although previous methods [30, 61, 87] estimate the autoregressive parameter along with current sources, $\mathbf{A}$ is still parametrized in a region-wise manner; that is, a single autoregressive parameter (say a) is assumed for every entry of $\mathbf{A}$ in a corresponding brain region [30, 87] or $\mathbf{A}$ is assumed in a low-dimensional latent space [61]. Indeed, region-wise dynamic modeling has a limitation on reconstructing spatially focal sources. Focal sources typically mistakenly blurred according to shared source dynamics in inverse solutions. By parameterizing $\mathbf{A}$ directly in the high-dimensional source space, inverse solutions, where focal sources follow their own source dynamics, could be reconstructed.

As a first step to the model parameter estimation in the source space, we initially restrict the MAR matrix $\mathbf{A}$ in Eq. (2.8) as a diagonal matrix in Chapter 3, but decompose these diagonal entries in an element-wise manner:

$$\mathbf{J}_t = \mathrm{diag}(\mathbf{a})\,\mathbf{J}_{t-1} + \boldsymbol{\omega}_t. \tag{2.9}$$

This allows estimating spatially inhomogeneous source temporal evolution, where the inhomogeneity is crucial for reconstructing spatially focal sources (see Section 3.2.1 also).

Then, we incorporate non-diagonal entries into $\mathbf{A}$ in Chapter 4, also in an element-wise manner to identify directed source interactions. An intuitive form of the dynamic source model can be written as

$$\mathbf{J}_t = \mathbf{A}\,\mathbf{J}_{\mathrm{past}} + \boldsymbol{\omega}_t. \tag{2.10}$$

Here, entries of $\mathbf{A}$ are estimated for only those corresponding to structurally connected pairs of current sources. Since we further extend the model by assuming connection-specific time lags in the MAR model, $\mathbf{J}_{t-1}$ in Eq. (2.8) is

replaced with $\mathbf{J}_{\text{past}}$ in Eq. (2.10). Indeed, using the single vector $\mathbf{J}_{\text{past}}$ in Eq. (2.10) is not a proper form of expression for connection-specific time lags. The exact form of dynamic source model is shown in Eqs. (4.1) and (4.3) in Section 4.2.1.

In this thesis, the system noise $\boldsymbol{\omega}_t$ is assumed to follow a Gaussian distribution $\mathcal{N}(\boldsymbol{\omega}_t \,|\, \mathbf{0}, (\beta\, \mathrm{diag}(\mathbf{q}))^{-1})$. Adopting the dynamic source model in Eq. (2.8) and the above system noise is equivalent to imposing the following prior distribution on the current sources:

$$P(\mathbf{J}_t \,|\, \mathbf{J}_{t-1}, \beta, \mathbf{A}, \mathbf{q}) = \mathcal{N}\left(\mathbf{J}_t \,|\, \mathbf{A}\,\mathbf{J}_{t-1}, (\beta\, \mathrm{diag}(\mathbf{q}))^{-1}\right). \tag{2.11}$$

2.2. Estimation algorithm

2.2.1 Variational Bayesian inference

Since the number of source dimension in the state-space model is rather high (in the order of 1000), the well-known Kalman filter and smoother cannot be easily applied for the inference of source and parameter posterior distributions, due to computational limitations. Therefore, we derived a computationally efficient estimation algorithm based on the variational Bayesian framework [1,70].

The variational Bayesian inference allows computing approximate posterior distributions of the sources and parameters. When the likelihood function and the source prior are given by Eqs. (2.4) and (2.11), respectively, the approximate posterior distributions are obtained by taking the following steps. First, we assume that the set of variables of interest contained in X are independent

$$\begin{aligned} &Q\left(X = \{\mathbf{J}_{1:T}, \beta, \mathbf{A}, \mathbf{q}\}\right) \\ &\qquad = \prod_{t=1}^{T} \{Q(\mathbf{J}_t)\}\, Q(\beta)\, Q(\mathbf{A})\, Q(\mathbf{q}). \end{aligned} \tag{2.12}$$

Then, we minimize the Kullback-Leibler (KL) divergence of the true posterior distributions $P(X|\, \mathbf{B}_{1:T})$ relative to the approximate distributions. Note that by assuming independence among $Q(\mathbf{J}_1), \ldots, Q(\mathbf{J}_T)$ in Eq. (2.12), we

can reduce the computation time in deriving $Q(\mathbf{J}_{1:T})$ when the dimension of the dynamic source model is of order 1000 (cf., the variational Kalman smoother [5, 11], where the independency within $Q(\mathbf{J}_{1:T})$ does not assumed). Minimizing the KL divergence is identical to maximizing the free energy F, defined as

$$F = \langle \log P(\mathbf{B}_{1:T}, X) - \log Q(X) \rangle_{Q(X)}. \tag{2.13}$$

The free energy is monotonically increased by sequentially and iteratively updating the approximate distribution of x (where x is one of $\mathbf{J}_1, \ldots, \mathbf{J}_T$, β, $\mathbf{A}$, or $\mathbf{q}$) using the following update rule:

$$\log Q(x) = \langle \log P(\mathbf{B}_{1:T}, X) \rangle_{Q(X \backslash x)} + \text{const}, \tag{2.14}$$

where $Q(X \backslash x)$ is the product of all approximate distributions other than $Q(x)$, and const is independent of x. After convergence, the approximate posterior distributions maximally approach the true distributions in terms of the KL divergence, under the independence assumption of Eq. (2.12).

The actual forms of estimation algorithm are derived from model-specific parameterization in Chapters 3 and 4. A psudecode of the estimation algorithm is presented in each chapter.

Chapter 3

Reconstruction of Focal Current Sources

3.1. Introduction

In this chapter[1], we extend previous state-space methods to reconstruct focal current sources. The proposed method here incorporates temporal evolution of focal current activity into the probabilistic model of the hierarchical variational Bayesian method [71, 88], under a state-space representation. We focus on linear temporal dynamics, as in existing state-space methods, but introduce the dynamics model parameters in a spatially non-uniform manner; i.e., a current source at every cortical location is assumed to evolve under its individual temporal dynamics. The spatially focal sources are reconstructed by estimating both the dynamics parameters and the intensities of current sources from MEG data.

Since the assumed number of the current sources on the cortical surface

[1]

is large (of order 1000), our state-space model contains many degrees of freedom. As a result, the currents and the parameters are not readily determinable because of high sensitivity to unmodeled data components and high computational costs [52]. The sensitivity is reduced by introducing prior distributions of parameters that are common to all current sources. These prior distributions control the overall adaptability of each model parameter to the data. The computational cost is reduced by employing variational Bayesian inference (see Chapter 2), which yields an approximate posterior distribution of variables of interest. This approximation enables the complexity of the current source computation to grow linearly with the number of sources. By solving these two problems, we have developed the first state-space model-based source reconstruction method that can estimate spatially focal current sources.

3.2. Methods

3.2.1 Dynamic source model

We model the temporal evolution of a current source by a first-order autoregressive (AR(1)) process, which differs at each cortical location. In particular, we consider the following special class of linear dynamics:

$$\mathbf{J}_t = \mathrm{diag}(\mathbf{a})\mathbf{J}_{t-1} + \boldsymbol{\omega}_t \tag{3.1}$$

where the vector $\mathbf{a}$ is the AR(1) parameter and $\boldsymbol{\omega}_t$ is system noise. We assume that each current source on each cortical location evolves independently of the others; that is, $J_{n,t} = a_n J_{n,t-1} + \omega_{n,t}$ where $J_{n,t}$, a_n, and $\omega_{n,t}$ are the n-th elements of $\mathbf{J}_t$, $\mathbf{a}$, and $\boldsymbol{\omega}_t$, respectively. Although the linear model (Eq. (3.1)) is too simple to model complex brain dynamics, it is a significant improvement upon previous state-space methods, in which the temporal evolution is assumed uniform over the entire brain [2, 10, 13, 73] or within a cortical area [4, 30, 87]. The system noise $\boldsymbol{\omega}_t$ is assumed to follow a Gaussian distribution $\mathcal{N}\left(\boldsymbol{\omega}_t | \, \mathbf{0}, (\beta \, \mathrm{diag}(\mathbf{q}))^{-1}\right)$, whose variance also differs at each cortical location. Here $\mathbf{q}$ is the parameter vector and the variance for the n-th cortical site, denoted σ_n, is equal to $(\beta q_n)^{-1}$ (q_n is the n-th element of $\mathbf{q}$). The

scaling parameter β is introduced to the system noise for technical convenience [59,71,88] during application of the variational Bayesian inference.

From a Bayesian point of view, specifying the dynamic source model (Eq. (3.1)) corresponds to setting the following prior distribution on the current sources:

$$P(\mathbf{J}_t|\,\mathbf{J}_{t-1},\beta,\mathbf{a},\mathbf{q}) = \mathcal{N}\left(\mathbf{J}_t|\,\mathrm{diag}(\mathbf{a})\mathbf{J}_{t-1},(\beta\,\mathrm{diag}(\mathbf{q}))^{-1}\right). \tag{3.2}$$

3.2.2 Prior distributions on model parameters

To approach the source reconstruction problem from Bayesian perspectives, we introduce prior distributions of the model parameters β, $\mathbf{a}$, and $\mathbf{q}$. For these parameters, the following conjugate priors [6] are used to simplify the derivation of their (approximate) posterior distributions. The prior distribution of the scaling parameter β is set as

$$P(\beta) \propto \beta^{-1}, \tag{3.3}$$

which corresponds to a non-informative prior [6]. For prior distribution of the AR(1) parameter $\mathbf{a}$, we select a zero mean Gaussian distribution:

$$P(\mathbf{a}) = \prod_{n=1}^{N} \mathcal{N}(a_n|\,0,\eta_0^{-1}) \tag{3.4}$$

where a_n is the n-th element of $\mathbf{a}$, and η_0 is a hyperparameter that controls the reliability of this prior (the higher the η_0, the more reliable the prior). For the parameter $\mathbf{q}$, the prior distribution is assumed as

$$P(\mathbf{q}) = \prod_{n=1}^{N} \mathcal{G}(q_n|\,q_0,\gamma_0) \tag{3.5}$$

where both q_0 and γ_0 are hyperparameters (the mean and shape parameter of Gamma distribution, respectively). High γ_0 means high reliability of this prior. The hyperparameters η_0, q_0, and γ_0 are shared by all current sources. Their values are determined as described in Section 3.2.5. It should be noted that these prior distributions are dynamic extensions of the ARD prior [58] that gives non-dynamic focal estimation [71,88]. Indeed, when all entries of $\mathbf{a}$

are set to 0 (i.e., $\eta_0 \to \infty$), Eqs. (3.2) and (3.5) reduce to the hierarchical prior distributions of ARD.

The non-informative priors for $\mathbf{a}$ and $\mathbf{q}$, derived by setting both η_0 and γ_0 to 0, were used in the simulations of Sections 3.3.1, 3.3.2, and 3.3.3, whereas the informative priors for these parameters (non-zero η_0 and γ_0) were used in the more realistic simulations of Section 3.3.4 and the real data analysis of Section 3.3.5. Choosing informative priors in the realistic data setting reduces the effects of modeling errors on parameter estimation. Since the dimension of $\mathbf{a}$ and $\mathbf{q}$ is large, these parameters are highly sensitive to modeling errors, such as discrepancies between the assumed model and the data generation process. Thus, to render our estimation robust to these errors, we reduce the sensitivity by constraining the parameter space through the informative priors.

3.2.3 Joint probability distribution

The likelihood function and all of the prior distributions comprise the joint probability distribution:

$$P(\mathbf{B}_{1:T}, \mathbf{J}_{1:T}, \beta, \mathbf{a}, \mathbf{q}) = \prod_{t=1}^{T} \{P(\mathbf{B}_t|\,\mathbf{J}_t, \beta)\}\, P(\mathbf{J}_1|\,\bar{\mathbf{J}}_0, \beta, \mathbf{a}, \mathbf{q}) \times \prod_{t=2}^{T} \{P(\mathbf{J}_t|\,\mathbf{J}_{t-1}, \beta, \mathbf{a}, \mathbf{q})\}\, P(\beta)P(\mathbf{a})P(\mathbf{q}) \tag{3.6}$$

whose graphical representation is shown in Fig. 3.1. The initial current $\bar{\mathbf{J}}_0$ is set as a zero vector rather than a vector of random variables, as appropriate for analysis of data whose starting point immediately follows a rest period. This setting was used for all analyses in Section 3.3. A minor alteration to the estimation algorithm would allow us to estimate $\bar{\mathbf{J}}_0$ also.

3.2.4 Estimation algorithm

Estimates of the current sources $\mathbf{J}_{1:T}$ and model parameters β, $\mathbf{a}$, and $\mathbf{q}$ are obtained by the variational Bayesian inference, calculating their approximate

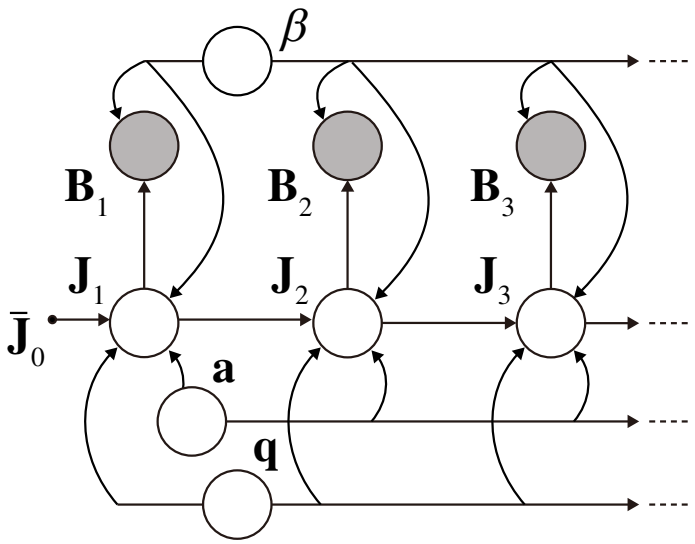

Figure 3.1: Graphical representation of the joint probability distribution. Each circle node represents a random variable (gray: observed, white: unobserved) and each arrow indicates a dependency between variables. $\bar{\mathbf{J}}_0$ (black dot) is constant.

posterior distributions $Q(\mathbf{J}_1), \ldots, Q(\mathbf{J}_T)$, $Q(\beta)$, $Q(\mathbf{a})$, and $Q(\mathbf{q})$. The estimation process of the proposed method is summarized as follows:

1. Set the initial values required for updating $Q(\mathbf{J}_1)$, and also set the hyperparameters.
2. **J**-step: For $t = 1:T$, update $Q(\mathbf{J}_t)$.
3. β-step: Update $Q(\beta)$, and compute the free energy.
4. **a**-step: Update $Q(\mathbf{a})$.
5. **q**-step: Update $Q(\mathbf{q})$.
6. If the change of the free energy relative to the value of the free energy at the previous step is smaller than 10^{-9}, then exit; otherwise, go to 2).

Each update rule is derived from Eq. (2.14) by substituting the corresponding variable of interest for x. Detailed derivation of the update rules is presented in Appendix A.1.

3.2.5 Initial values and hyperparameters

Since the proposed method adopts an iterative estimation algorithm, it requires initial values for the first update of $Q(\mathbf{J}_1)$ in the **J**-step. In our algorithm the mean and variance of $Q(a_n)$ and the mean of $Q(q_n)$ must be initialized (see Eqs. (A.2)–(A.4) in Appendix A.1). To avoid the need to initialize the means of $Q(\mathbf{J}_{2:T})$, the mean and variance of $Q(a_n)$ are initialized to 0 for all n. The initial mean of $Q(q_n)$ is set as the estimated inverse variance parameter of the hierarchical variational Bayesian method (VB) described in [88], which is equivalent to the mean of $Q(q_n)$ when no temporal evolution is assumed (i.e., when all a_n are fixed at 0).

In addition, three hyperparameters must be specified. The two hyperparameters η_0 and γ_0, both of which control reliability of the priors, are selected from an appropriate range of values explored in a grid-search manner (see Section 3.3.4). As noted above, η_0 and γ_0 are set to 0 during the estimations of **a** and **q** with the non-informative priors (Sections 3.3.1, 3.3.2, and 3.3.3). The remaining hyperparameter q_0 is determined such that the product $(\beta q_0)^{-1}$, corresponding to the prior for the system noise variance σ_n for any n, becomes the average variance of all current sources over the cortex. When specifying this variance, we use the currents estimated by the minimum norm estimation (MN) [38] and we set β^{-1} as the variance of the observation noise estimated by MN, with the optimal regularization parameter of MN derived from variational Bayesian inference.

3.2.6 Required computation time

All analyses were performed in MATLAB 7.5.0. Approximately 15 minutes were required for convergence of the iterative process using a standard multicore PC (2 cores, 2.66 GHz CPU, and 8 GB RAM) for real data analysis in Section 3.3.5 ($M = 400$, $N = 2000$, $T = 300$, and $\eta_0 = \gamma_0 = 10^2$). In this case, 89 iterations were required to maximize the free energy. When the free energy calculation was bypassed, the computation time was reduced to around 3 minutes only. Note that the number of iterations required to converge depends on the values of the hyperparameters. For smaller η_0 and γ_0, more

iterations are needed. For the first simulation in Section 3.3.1, which assumes non-informative priors, 926 were required for free energy convergence.

3.2.7 Previous methods for comparison

We present here two previously proposed methods against which our method is compared in Sections 3.3.1 and 3.3.5. This comparison reveals the effects of temporal constraints on the estimated currents, and also the effects of spatially inhomogeneous and data-adaptive dynamics parameters.

To meet the first purpose, our model is compared with VB [71, 88]. The model in VB can be viewed as a specialized form of our model with all entries of **a** set to 0. Under this simplification, the prior distribution of the currents is not dependent on past and future currents, thereby removing temporal constraints.

For the second purpose, our model is compared with that of [10], which is similar to a random walk model (denoted RW hereafter). Our model reduces to the RW model when all entries of **a** are set to 1 (constant) and the vector **q** is replaced by a scalar q. In this chapter, all entries of **a** are set to 0.9 (i.e., less than 1), to avoid gradual random-walk deviation of the amplitudes from the baseline (slight changing the value of **a** does not substantially alter the results).

3.2.8 Local spatial smoothing

For all methods, a local spatial smoothness constraint is imposed on the current sources by applying a spatial smoothing filter. The relationship between the smoothed currents $\mathbf{J}_t$ and the unsmoothed currents $\mathbf{Z}_t$ is assumed to be

$$\mathbf{J}_t = \mathbf{W}\mathbf{Z}_t \tag{3.7}$$

where **W** represents a Gaussian smoothing filter with 8 mm full width at half maximum as used in [71, 88]. By substituting Eq. (3.7) into Eq. (2.1), the observation process is rewritten as

$$\mathbf{B}_t = \hat{\mathbf{G}}\mathbf{Z}_t + \boldsymbol{\epsilon}_t \tag{3.8}$$

where $\hat{\mathbf{G}} = \mathbf{GW}$ is a smoothed lead field matrix. To obtain the solutions $\mathbf{J}_{1:T}$, we first solve the source reconstruction problem of $\mathbf{Z}_{1:T}$ with the smoothed lead field matrix $\hat{\mathbf{G}}$, then convert $\mathbf{Z}_{1:T}$ to $\mathbf{J}_{1:T}$ by computing (3.7).

3.3. Results

3.3.1 Simulation study 1: Method comparison

First, we conducted simulation experiments to compare the performance of the three source reconstruction methods (the proposed method, VB, and RW). Time courses and cortical positions of the simulated active current sources are summarized in the leftmost column of Fig. 3.2. Three active sources were assumed, following time courses of non-smooth with high amplitude, smooth with high amplitude, and smooth with low amplitude (hereafter denoted Sources 1, 2, and 3, respectively). Chosen positions of the three sources are indicated by the three arrows in Fig. 3.2. Each current source was blurred with a local spatial smoothing filter (see Section 3.2.8). Although the time course of Source 1 appears physiologically unrealistic, it usefully illustrates how non-smooth temporal profiles affect our estimation results.

The simulated MEG, for which sampling frequency was assumed to be 1 kHz, was generated by multiplying the simulated currents with the lead field matrix and then adding white Gaussian noise. The lead field matrix was based on the anatomical structure of the cortex extracted from a T1 image of magnetic resonance imaging (MRI). The covariance matrix of Gaussian noise was determined from real MEG data recorded without a subject and its scale was adjusted to a signal-to-noise ratio (SNR)[2] -5 dB, similar to that of single-trial measurements of stimulus-evoked responses. In concordance with the real data analysis in Section 3.3.5, the structure of this covariance was estimated from rest period measurements, which contained only Gaussian noise in this simulation.

The proposed method and VB successfully reconstructed all sources as

[2]The SNR is defined as $10 \log_{10} \frac{\sum_{m=1}^{M} \sum_{t=1}^{T} \hat{B}_{m,t}^2}{MT\beta^{-1}}$, where $\hat{B}_{m,t}$ is the noise-free sensor measurements on the m-th sensor at time point t.

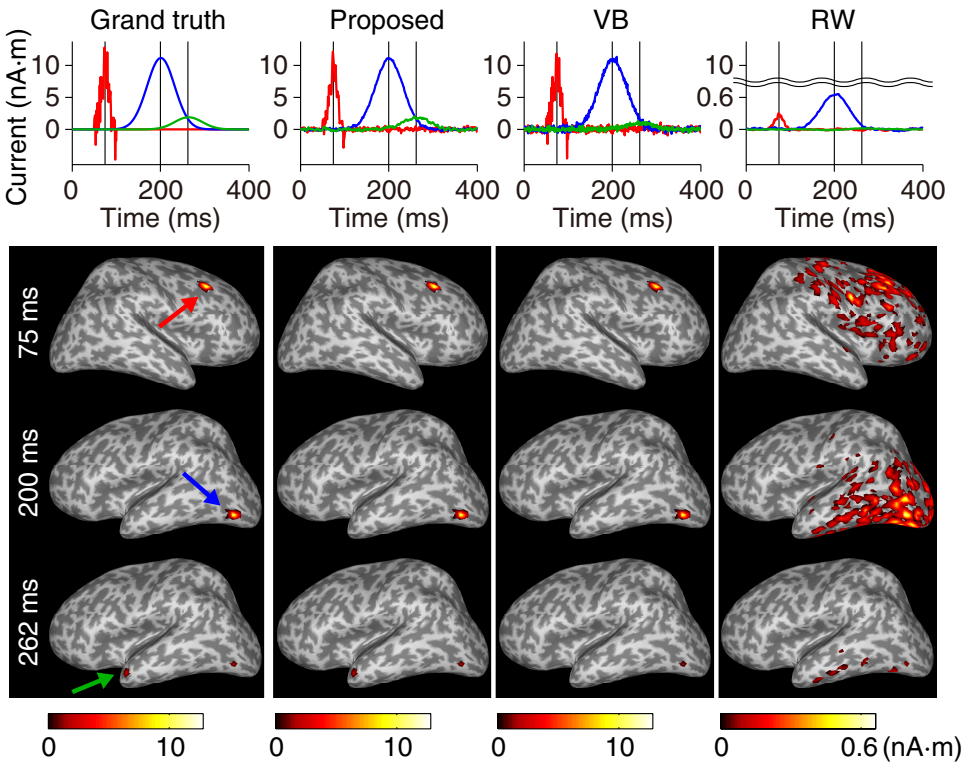

Figure 3.2: Spatiotemporal profiles of the true and estimated currents in simulation study 1. The left-most column illustrates the grand truth. The other columns (left to right) show the estimates from the proposed method, VB, and RW, respectively. Upper: time courses of the currents at the three active positions (Sources 1, 2 and 3 plotted in red, blue and green, respectively). Lower: spatial maps of the currents at peak time points 75 ms, 200 ms, and 262 ms. The estimated currents are mapped onto an inflated cortical surface, and the currents exceeding one tenth of the maximum amplitude are shown.

spatially focal, whereas RW reconstructed them as spatially diffuse (see Fig. 3.2). The waveform of Source 1 was best estimated by VB. In contrast, it was slightly over-smoothed by the proposed method. Both methods reconstructed Source 2 similarly. For Source 3, the proposed method generated better amplitude estimates than VB. The amplitudes of all activities estimated by RW were suppressed, being below one tenth of the assumed amplitudes.

To statistically evaluate the performance of the three methods, we repeated the above simulation 100 times for different positions of the three sources. We also considered two SNR cases, 5 dB and -5 dB. The cross correlation and the

reconstruction gain were used as performance measures. The cross correlation is defined as the correlation between the time courses of the estimated and true currents. The reconstruction gain is the ratio of the estimated amplitude of the currents to the true one. Both measures were computed at the cortical locations of the true sources. The results are summarized as boxplots in Fig. 3.3. For Source 1, the cross correlation was highest in VB but the proposed method yielded the higher reconstruction gain. For Source 2, the estimates from the proposed method and VB were very similar. For Source 3, both cross correlation and reconstruction gain were closest to 1 for the proposed method (detailed investigation reveals that the lead-field norm [71] of the true source position, rather than the distance between sources, influences the estimation accuracy). The cross correlation of RW was comparable to those of the proposed method and VB, but RW yielded low reconstruction gains in all three cases.

To further understand the results of the proposed method, we investigated

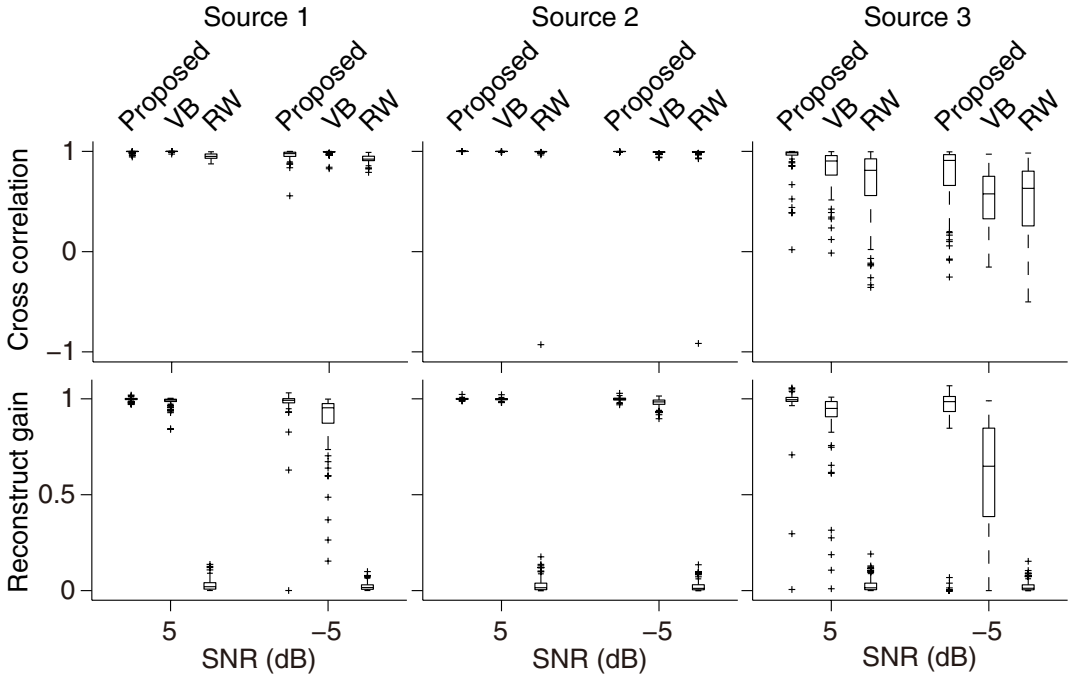

Figure 3.3: Boxplot representation of cross correlation (upper) and reconstruction gain (lower) of the currents estimated by the proposed method, VB, and RW at SNRs ± 5 dB.

the relationship between the estimated currents and the estimated dynamics parameters (AR(1) parameter and system noise variance). These estimates, resulting from the simulation data shown in Fig. 3.2, are presented in Fig. 3.4. The currents were non-zero at positions where either the AR(1) parameter or the system noise variance was estimated as non-zero. When the SNR was high (5 dB), the AR(1) parameter was estimated as non-zero only at the three positions of the grand truth. When the SNR was low (-5 dB), negative AR(1) parameter and large system noise variance were estimated at several positions of false positive currents. In both SNR cases, temporally smooth currents (Sources 2 and 3) assessed the AR(1) parameter as close to 1 and produced low system noise variance. For the non-smooth current (Source 1), the AR(1) parameter was lower than for the previous two cases, while the

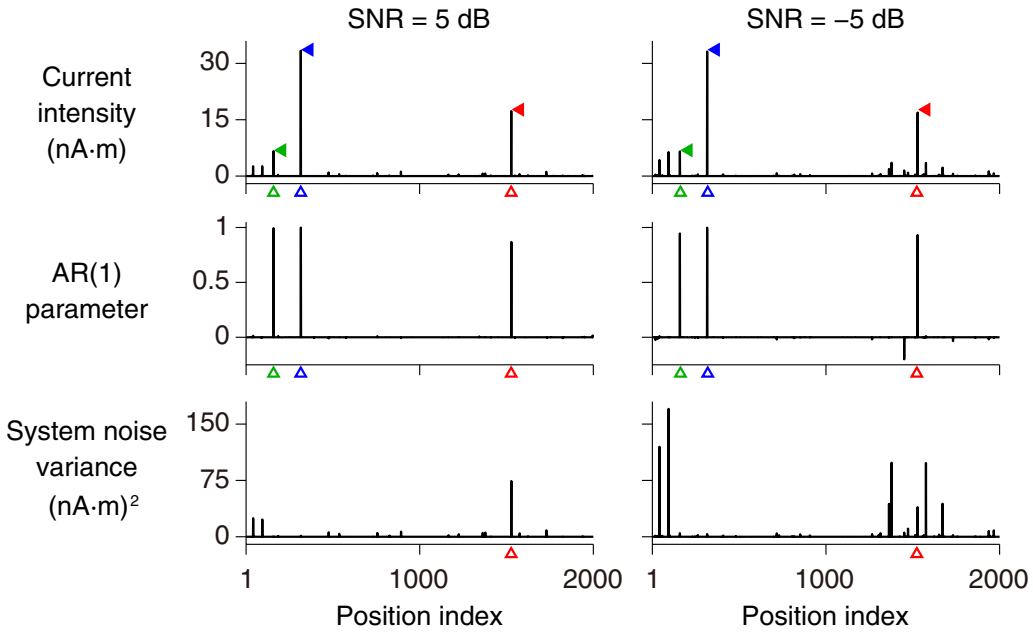

Figure 3.4: Estimated currents and dynamics parameters from the proposed method. Upper: current intensity (shown as standard deviation). Center: AR(1) parameter. Lower: system noise variance. Positions of Sources 1, 2, and, 3 are indicated by red, blue, and, green unfilled triangles, respectively. The amplitude of the grand truth currents is indicated by filled triangles. In this figure, the currents $\mathbf{Z}_{1:T}$ before applying the local spatial filter are presented rather than the smoothed currents $\mathbf{J}_{1:T}$ to clearly show the correspondence of position indices between the currents and the parameters.

system noise variance was much higher. Although the estimated AR(1) parameter appears unexpectedly high, this result is reasonable because the low frequency component in Source 1 generates some auto-correlation.

3.3.2 Simulation study 2: Estimation for correlated sources

To examine performance of the proposed method for correlated sources, we changed the source configuration from simulation study 1 to a situation in which two nearby sources (10–15 mm apart) have an identical waveform (that of Source 2 in simulation study 1). This simulation was repeated 100 times for different positions of the paired sources.

The proposed method successfully reconstructed the correlated sources for most of the simulations when the SNR was 5 dB. The reconstruction gain averaged over the two true sources was almost 1 for 77 times. In these cases, the false positive sources were not observed around the peak time point (one tenth of the maximum amplitude was used as a threshold). Even at the low SNR (-5 dB), the reconstruction gain was almost 1 for over half (56 times) of the simulations, while the false positives were appeared in 15 out of 56 times.

3.3.3 Simulation study 3: Validation of parameter estimation

We conducted a further set of simulation experiments to check whether the dynamics parameters can be estimated correctly by variational Bayesian inference. To assess the accuracy of the parameter estimation, we generated time courses of the grand truth currents directly from our temporal model shown in Eq. (3.1). Three active sources were assumed, with AR(1) parameter values a_n of 1) 0.7, 2) 0.8, and 3) 0.9, and system noise variance σ_n of 70 $(\mathrm{nA} \cdot \mathrm{m})^2$ for all cases. For the other non-active sources, AR(1) parameters and system noise variances were both set to 0. Apart from the waveform of the simulated currents and the SNRs, the simulation setting was that of the repeated experiments of simulation study 1.

As shown in Fig. 3.5, the dynamics parameters were estimated correctly even under the variational Bayesian approximation when the SNR was ± 5 dB. At a SNR of -15 dB, far below that of single-trial measurements, the AR(1)

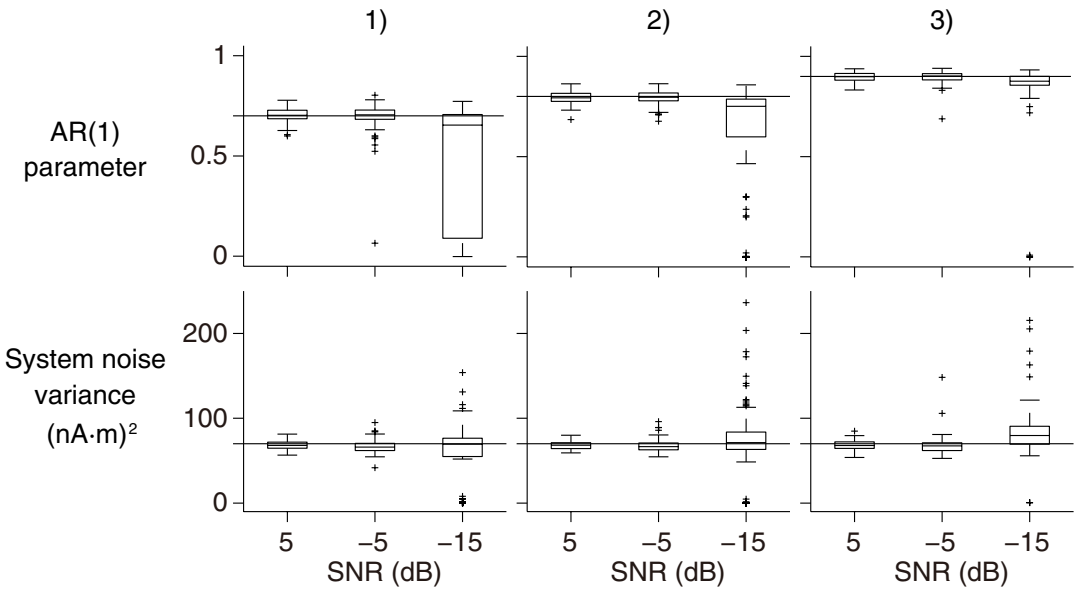

Figure 3.5: Boxplot representation of the AR(1) parameter (upper) and system noise variance (lower) estimated by the proposed method at SNRs 5 dB, −5 dB, and −15 dB. The true parameter values of the three active sources are represented by horizontal lines slicing the boxes.

parameter was underestimated and the system noise variance was overestimated several times. In addition, the estimated parameters were close to 0 in some cases, as a result of mislocalization of the active sources. Nevertheless, the median of the parameters remained close to the true values.

3.3.4 Simulation study 4: Effects of hyperparameters

Next, we conducted more realistic simulation experiments to assess the importance of appropriate hyperparameter values. Two active current sources in the left and right temporal lobes were assumed, mimicking auditory evoked responses (see the blue and red arrows of the leftmost column in Fig. 3.6). MEG measurements were generated as for simulation study 1, but replacing white Gaussian observation noise with more realistic background noise. The background noise was derived from 300 ms pre-stimulus periods of trial-averaged data from the real auditory experiment described in Section 3.3.5 (SNR was around 7 dB). Since such real background brain activities are nei-

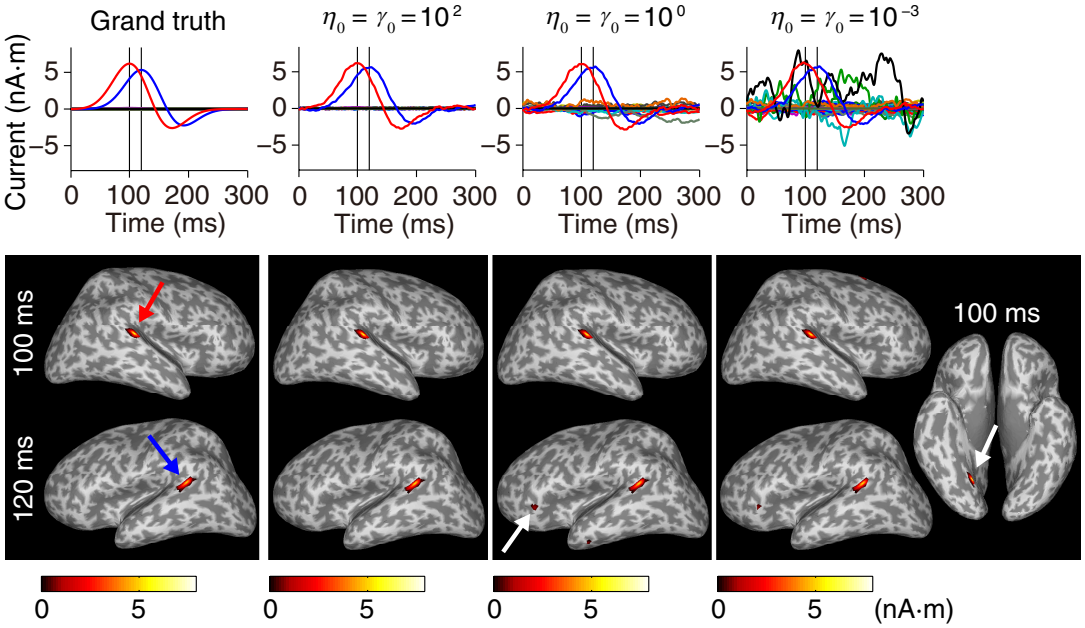

Figure 3.6: Spatiotemporal profiles of the true and estimated currents in simulation study 4. The left-most column illustrates the grand truth. The other columns (left to right) show the estimates from the proposed method for both η_0 and γ_0 equaling 10^2, 10^0, and 10^{-3}, respectively. Upper: time courses of the currents over the whole cortex. Lower: spatial maps of the currents at peak time points 100 ms and 120 ms (white arrows indicate false positives). The mapping threshold is that used in Fig. 3.2.

ther Gaussian nor white, these components could not be properly modeled in our method. Therefore, the hyperparameters η_0 and γ_0 were carefully chosen to minimize the effect of modeling errors on the simulation results.

Fig. 3.6 illustrates the estimates from the simulated measurements. With the hyperparameters set to $\eta_0 = \gamma_0 = 10^2$, the two current activities were correctly extracted and the others were suppressed. When η_0 and γ_0 were reduced, i.e., the prior distributions of the parameters were closer to non-informative ones, parameter estimation became more sensitive to disturbance. In this case (both η_0 and γ_0 set to 10^0 or 10^{-3}), the estimated current activities appeared not only at the positions of the grand truth but also at other positions.

To determine the appropriate range of the hyperparameter values, we repeated the estimation by varying η_0 and γ_0 from 10^{-3} to 10^6 (10 degrees). Fig.

3.7(a) shows an interpolated contour map of the sum of squared error (SSE) between the true and estimated currents. Fig. 3.7(b) presents the results of receiver operating characteristic (ROC) analysis [77]. In this map, the hyperparameter settings for which the area under the ROC curve (AUC[3]), averaged over 90–130 ms, equals 1, are delineated by dark and light gray regions. The dark gray region is regarded as the "correct" source reconstruction case, in which the current activities are obtained only at the two true positions, using one tenth of the maximum amplitude as a threshold. The best case of the estimates in Fig. 3.6 is indicated by a small circle, and the other cases, yielding false positives, are indicated by a small triangle and a small square. The SSE and AUC maps also show that large hyperparameter values result in incorrect source reconstruction. As shown in Fig. 3.7(b), correct estimates were obtained for a relatively wide range of γ_0 when η_0 was 10^2 or 10^3. On the other hand, setting γ_0 to 10^2 enabled a large range of η_0 to be successfully used. Thus, appropriate ranges for the hyperparameters, at least in the tested simulation setting, are $\eta_0 = 10^2$–10^3 and $\gamma_0 \sim 10^2$.

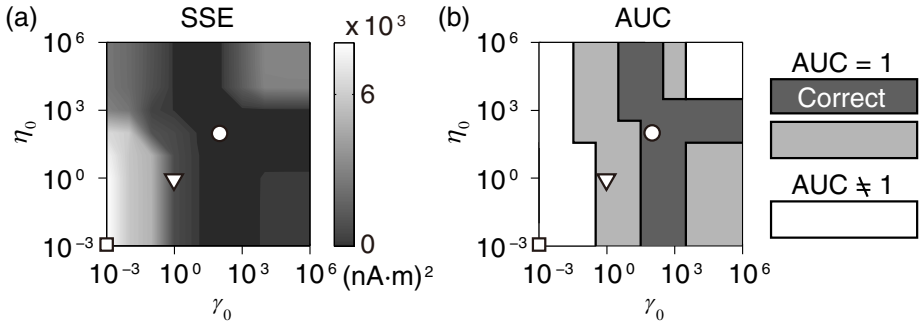

Figure 3.7: (a) Interpolated contour map of the SSE between the true and estimated currents. (b) Evaluation of source reconstruction results based on AUC. Sets of hyperparameters resulting in AUC equal to 1 are colored gray (dark and light) while those resulting in AUC not equal to 1 are white. The former region is further divided into two sub-regions, indicated by dark and light gray, based on quantitative differences (see text). The small circle, triangle, and square in each map correspond to the hyperparameter sets shown in Fig. 3.6.

[3]A potential bias caused by discrepancy between the active and inactive sources is not corrected in this chapter.

3.3.5 Real data analysis: Auditory evoked response

Finally, we examined whether the above-specified hyperparameter values provide a physiologically plausible result in a real data analysis. We applied the proposed method to the auditory evoked fields from one right-handed healthy subject (a 27-year-old male). A signed informed consent approved by the ATR Human Subject Review Committee was obtained prior to experiment. Measurement data were acquired from a 400 channel MEG system (MEGvision PQ1400RM, Yokogawa Electric Co.) with a sampling frequency of 1 kHz. The cortical surface model was constructed from a segmented gray matter image of T1-weighted MRI collected by a 3 T scanner (MAGNETOM Trio, A Tim System 3T, Siemens).

An 800 Hz tone burst of duration 500 ms was presented to the left ear of the subject via an eartip attach to 20 feet (6.096 m) of silicone tubing (ER-30, Etymotic Research, Inc.). To keep the subject awake, we asked him to press a button 2–3 s after each stimulus. The trial-averaged data in the time window 0 ms to 300 ms with respect to the stimulus onset were used to estimate the currents (time delays of sounds in a silicone tube are not taken into account). The noise covariance matrix was determined in advance from non-averaged rest period data in a 100 ms pre-stimulus window.

The following preprocessing operations were applied: baseline correction, high-pass filtering (cutoff 0.5 Hz), low-pass filtering (cutoff 200 Hz), trial segmentation, and trial rejection. Artifactual trials, determined from the amplitudes of MEG and electrooculogram (EOG) and from subject behavior (button press), were rejected. As a result, 112 out of 128 trials were accepted for further analysis.

Preprocessed trial-averaged MEG data, showing typical time courses and scalp maps of the auditory evoked fields, are displayed in Fig. 3.8(a). Fig. 3.8(b) presents the estimated currents for $\eta_0 = \gamma_0 = 10^2$. Focal current activities were estimated at the contra-lateral and ipsilateral primary auditory cortices (cA1 and iA1, respectively), consistent with previous findings from a cat electrophysiological study [48], a human intracranial study [32], and a human MEG study [39]. The estimated time courses show that the latency and amplitude of cA1 activities are shorter and larger respectively, than those

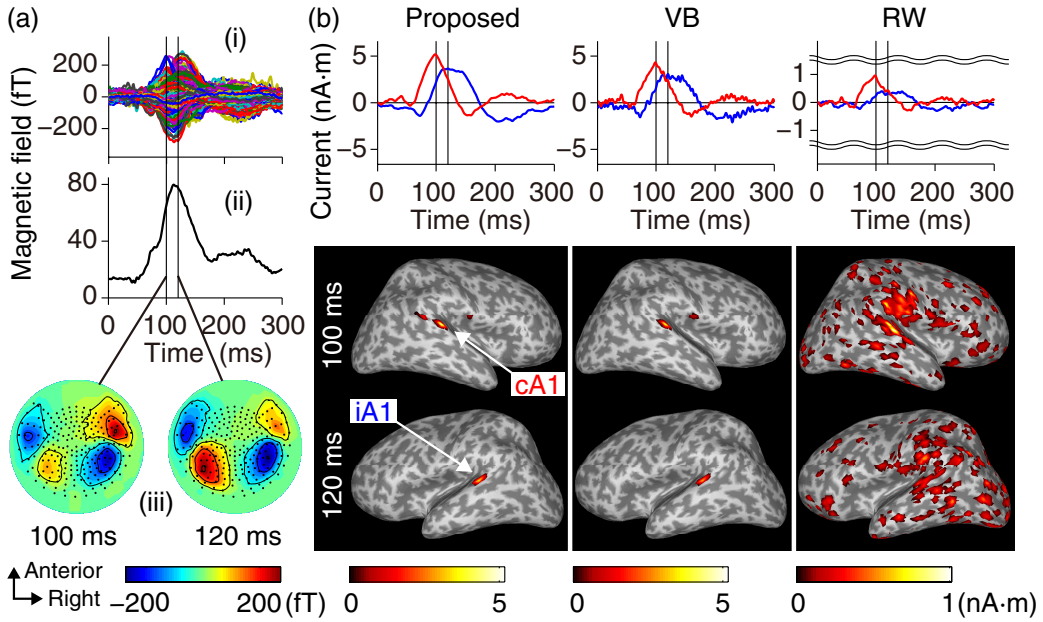

Figure 3.8: (a) Preprocessed trial-averaged MEG data from the auditory evoked fields: (i) superposition of the sensor signals; (ii) square root of power averaged over all the sensors; (iii) 2D scalp maps at two time points, 100 ms and 120 ms. (b) Spatiotemporal profiles of the estimated currents from the MEG data shown in (a). Upper: time courses of the currents at cA1 (red) and iA1 (blue) (plotted vertices are common to the three methods). Lower: spatial maps of the currents at peak time points 100 ms and 120 ms. The mapping threshold is the same as for Figs. 3.2 and 3.6. Small difference in the pre-specified threshold did not affect the basic properties of the mapped results.

from iA1. These results are also consistent with previous physiological findings [29, 66]. When η_0 was increased to 10^3, the number of active positions slightly decreased while the current time courses at cA1 and iA1 did not effectively change.

Similar results were obtained from VB using the same hyperparameter values, but some differences were observed: the amplitude and temporal smoothness of the currents were lower than those of the proposed method, and the active brain region was more limited. RW generated spatially diffuse currents with the lowest amplitudes as shown in Section 3.3.1.

3.4. Discussion

We have developed a state-space model-based source reconstruction method that achieves current source reconstruction to high spatial resolution (i.e., reconstruction of focal current source patterns). To our knowledge, the method is the first to reconstruct spatially focal current sources utilizing temporal constraints from the state-space model. The key to our method is to assume different temporal evolution for each current source and to estimate all parameters in the state equation from data. The joint estimation of both the current sources and the model parameters is based on the Bayesian framework. We overcome difficulties presented by the high dimensionality of the problem by 1) introducing prior distributions for the parameters, which reduces the sensitivity of the parameter estimation to data, especially when discrepancies exist between the model assumptions and the data generation process, and 2) employing variational Bayesian inference, which reduces the computational cost of the estimation algorithm.

The proposed method is a dynamic extension of the hierarchical variational Bayesian method (VB) [71, 88]. Its advantage over VB is high reconstruction gain, especially for low-amplitude currents (see Fig. 3.3). This indicates that, by introducing temporal constraints, considerable improvement is achieved over a simple smoothing operation such as low-pass filtering. The enhanced reconstruction gain of our method demonstrates that the amplitude of current source underestimated by VB can be compensated by the effective use of the current sources at adjacent time points, owing to the estimated temporal dynamics. Indeed, VB, along with FOCUSS [34] and empirical Bayesian methods [25, 53] is a special form of covariance component estimation methods [82]. Unlike the proposed method, none of these established methods incorporate temporal constraints. Introducing state-space modeling into FOCUSS and empirical Bayesian methods might enhance their performance in reconstructing weak sources, as for VB.

A key factor in reconstructing focal activities is estimating the temporal dynamics of each source individually, and this is not apparent from the form of our probabilistic model unlike the explicit constraint term of the $l1$ or $l0$ norm-regularization methods [55, 81, 85]. To understand how our method can

reconstruct focal activities, we must consider the prior distribution for any single current source, given by $\mathcal{N}(J_{n,t}|\, a_n J_{n,t-1}, \sigma_n)$, and the update rules of a_n and q_n (see Eqs. (A.11) and (A.14) in Appendix A.1, respectively). The prior distribution ensures that when the AR(1) parameter a_n and the system noise variance σ_n are close to 0, the currents are softly constrained to 0. For such near-zero currents, the update rules dictate that a_n and σ_n (which is proportional to q_n^{-1}) are close to 0 also (σ_n is not exactly zero when $\gamma_0 \neq 0$). These two mechanisms provide positive feedback effects on promoting the most of $J_{n,1:T}$, a_n, and σ_n estimated to be near 0, resulting in spatially focal activities (see Fig. 3.4). We have confirmed that this property also holds under initialization with non-focal estimates such as MN. In contrast to the proposed method, homogeneous random walk (RW) dynamics imposes no zero-value constraints on any of the current sources, since all sources fluctuate around their past values in a spatially uniform manner.

There are another type of focal source reconstruction methods in the subspace approach, such as MUSIC [56] and FINES [20, 86]. As reported in [20], these methods have difficulty in reconstructing perfectly correlated and closely located sources; however, simulation study 2 shows that these sources can be reconstructed by the proposed method. This is a possible advantage of our method over MUSIC and FINES, in addition to a desirable property that the number of active sources does not need to be estimated beforehand.

Variational Bayesian inference [1, 70] has enabled the computationally demanding inference of the high-dimensional state-space model with little detriment to accuracy. A main drawback of our estimation algorithm appears to be the decrease in parameter estimation accuracy due to the independence assumption on the approximate posterior distribution of the current sources $Q(\mathbf{J}_{1:T}) = Q(\mathbf{J}_1) \cdots Q(\mathbf{J}_T)$. This presents the major difference between our algorithm and the variational Kalman smoother (VKS) algorithm [5, 11]. However, as shown in simulation study 3, parameter estimation of our method was reasonably accurate for the modest SNR case (± 5 dB) (see Fig. 3.5). In fact, we also observed little difference in accuracy between our algorithm and the VKS algorithm in a small-size simulation. The estimation accuracy of the dynamics would come from the fact that temporal dependency between

current $\mathbf{J}_t$ is propagated through the means of the posterior distributions at neighboring time samples $\mathbf{J}_{t-1}$ and $\mathbf{J}_{t+1}$ in our algorithm (see Eqs. (A.2)–(A.4) in Appendix A.1).

To reduce the sensitivity of the parameter estimation to unmodeled components, in particular for real data analysis, appropriate values of the hyperparameters are required. An example of such unmodeled components for our method is background brain activity, which cannot be modeled by white Gaussian observation noise. With no informative prior distributions, our source reconstruction could be affected by these factors. We searched the appropriate ranges of the hyperparameters in simulation study 4, in which the background activities of the real experiment were used, and confirmed that those hyperparameter settings derived physiologically reasonable results from real MEG data of the auditory evoked fields (see Fig. 3.8(b)). In fact, by applying the same hyperparameters to somatosensory evoked fields, we have also reconstructed plausible current sources that were consistent with literature results (data not shown).

Recently, both spatially focal and temporally smooth current source reconstruction has been achieved in [62] and [7], via group-wise $l1$ regularization and temporal basis functions. Although such methods might yield similar results to ours, the state-space modeling approach can model spatiotemporal properties of the current sources, and is thus more flexible in constraining inverse solutions.

The temporal dynamics assumed in this study might appear to be oversimplified; however, our model substantially advances those of previous state-space methods. The model could be further extended by using a matrix form of the AR(1) parameter to account for source-space effective connectivity. This extension would be much more demanding because the number of parameters increases with square order of the state dimension. The large degrees of freedom could be reduced by incorporating structural connections as prior knowledge. This possibility is pursued in the next chapter.

Chapter 4

Reconstruction of Sources with Identifying Long-Range Interactions

4.1. Introduction

In this chapter[1], we extend previous state-space methods to allow long-range interactions to be directly estimated in the source space. To achieve this goal, the full MAR model is implemented in the high-dimensional source space. The structure of the MAR model is informed by whole-brain structural networks inferred from diffusion MRI (dMRI). More specifically, the MAR coefficients (entries of the MAR matrix) associated with pairs of structurally connected sources according to dMRI, are estimated from the data, while the others are fixed at zero. The time lags of the MAR model are determined from the mean fiber lengths between pairs of source locations.

The use of structural long-range connectivity to constrain the source dynamics is supported by previous findings; the whole-brain structural connectivity patterns determined by dMRI are closely related to the resting-state functional connectivity networks imaged by fMRI [19, 45]. The relationship

[1] A part of the work in Chapter 4 has been submitted to NeuroImage on November 29, 2013 (current status on March 13, 2014: under revision). A preliminary study on this chapter was presented in a conference proceeding [28].

between structure and dynamics of the brain has been assumed in forward modeling of neuronal dynamics [18, 31, 46] and in estimating the effective connectivity from fMRI data [76,84]. The *a priori* knowledge of structural connectivity also enables a feasible estimation by reducing the prohibitively large number of model parameters (of order 1000×1000 in our scenario).

To further improve the reliability of source reconstruction, we apply an fMRI prior on the spatial patterns of source activity. While the fMRI prior is used as a spatial constraint frequently in *non-dynamic* (or not temporally constrained) reconstruction methods [8, 12, 42, 63, 71], it has yet to be applied in the above-mentioned *dynamic* (or state-space) methods. The fMRI prior in the proposed method is implemented similarly to our previous non-dynamic method [71,88]. In forming this prior, the variance of the system noise (an input term driving the spatiotemporal dynamics of the MAR model) is weighted by the fMRI t-values. If all MAR coefficients are fixed at zero, this prior becomes identical to the fMRI prior proposed in [71,88].

The present study unifies the MAR model, prior knowledge on the model parameters, and the measurement process of the current sources into a Bayesian framework. To improve stability of the estimated source dynamics, this framework also includes a sparse prior on the MAR coefficients. All hidden parameters in the unified probabilistic model (such as source amplitudes and the MAR coefficients) are jointly estimated by a variational Bayesian algorithm [1,70] (see Chapter 2). The update rules are similar to those proposed in Chapter 3, enabling inference of a high-dimensional dynamic model within a reasonable computation time.

Our method estimates the effective connectivity in the source space without requiring the selection of regions of interest (ROIs). To this end, the source dynamics are formulated using the full MAR model, and the source amplitudes and interactions are estimated simultaneously over the whole brain. These extensions allow exploratory analysis of the integration of brain functions, which complements the confirmatory approach of dynamic causal modeling (DCM; [16,26]). In contrast to our method, DCM initially assigns a small number of ROIs as network nodes, and then examines the validity of the network solutions by post hoc comparison of the model evidence.

The proposed method is quantitatively and qualitatively evaluated on simulation and experimental data, respectively. The results are compared with those obtained from the non-dynamic equivalent. First, we examine the identification accuracy of the MAR model, using data generated from the adopted dynamic source model. We then investigate the estimation performance under more realistic conditions by mimicking stimulus-evoked responses by a network of neural mass models [14, 15, 49]. Finally, we examine the physiological plausibility of the estimates by application to a publicly available experimental dataset on face perception [43].

This chapter is organized as follows. Section 4.2 explains the model formulation and the adopted estimation algorithm. Model construction from the data and schemes for evaluating the estimation performance are described in Section 4.3. Sections 4.4 and 4.5 present the settings and results of the evaluation studies. Section 4.6 highlights the significance of the present study, and discusses the advantages and limitations of the proposed method.

4.2. Theory

4.2.1 Dynamic source model

The spatiotemporal dynamics of the current sources is directly formulated in the source space by a full MAR model, in which the MAR coefficients represent the source interactions across the whole brain. The MAR coefficients to be estimated, and the time lags between pairs of sources, are based on structural brain connectivity inferred from dMRI.

We begin by modeling the dynamics of a single source $J_{n,t}$, which linearly interacts with its structurally connected sources as follows:

$$J_{n,t} = \sum_{k \in C_n} \left(a_{n,k} J_{k,\, t-\delta_{n,k}} \right) + \omega_{n,t}, \tag{4.1}$$

where C_n is an index set of source locations that are structurally connected to the n-th source (including self-connection), $a_{n,k}$ and $\delta_{n,k}$ are the MAR coefficient and time lag, respectively, for interconnection between the n,k-th source pair, and $\omega_{n,t}$ is the system noise of the n-th source. The MAR coefficient $a_{n,k}$

quantifies the effective connectivity from the k-th source to the n-th source. The time lag is determined by the length of the structural connection $l_{n,k}$, the axonal conduction velocity v, and the local delay constant τ, as follows:

$$\delta_{n,k} = \frac{l_{n,k}}{v} + \tau. \tag{4.2}$$

Ways of determining the structural connections and the delay parameters are explained in Section 4.3.

The dynamic source model (Eq. (4.1)) of all current sources over the whole brain can be written in a single equation:

$$\mathbf{J}_t = \sum_{l=1}^{L} \left(\mathbf{A}_l \, \mathbf{J}_{t-\Delta_l} \right) + \boldsymbol{\omega}_t, \tag{4.3}$$

where $\Delta_{1:L}$ contains all time lags in Eq. (4.1) for all n in ascending order. Since a single MAR coefficient is assigned to a pair of structurally connected sources with a specific time lag, only one of $\mathbf{A}_l$ for $l \in \{1:L\}$ is non-zero in each given matrix entry. For simplicity, the MAR matrix $\mathbf{A}$ is constructed by allocating all the non-zero entries of $\mathbf{A}_{1:L}$ to an N-dimensional zero matrix. The matrix $\mathbf{A}$ can also be represented by

$$\mathbf{A} = \sum_{l=1}^{L} \mathbf{A}_l, \tag{4.4}$$

where any original values of the MAR coefficients are not superimposed. The system noise for all current sources is denoted by $\boldsymbol{\omega}_t$ and is assumed to follow a Gaussian distribution $\mathcal{N}(\boldsymbol{\omega}_t \,|\, \mathbf{0}, (\beta \, \mathrm{diag}(\mathbf{q}))^{-1})$. Here the parameter $\mathbf{q}$ in the variance term is multiplied by β for technical convenience during the application of the variational Bayesian algorithm [59,71,88]. The above model of source dynamics and system noise is equivalent to imposing the following prior distribution on the current sources:

$$\begin{aligned} &P(\mathbf{J}_t \,|\, \mathbf{J}_{t-\Delta_1}, \ldots, \mathbf{J}_{t-\Delta_L}, \beta, \mathbf{A}, \mathbf{q}) \\ &\qquad = \mathcal{N}\left(\mathbf{J}_t \;\Big|\; \sum_{l=1}^{L} \left(\mathbf{A}_l \mathbf{J}_{t-\Delta_l} \right), (\beta \, \mathrm{diag}(\mathbf{q}))^{-1} \right). \end{aligned} \tag{4.5}$$

The system noise $\boldsymbol{\omega}_t$ in Eq. (4.3) can be regarded as an input to the dynamic source model. Thus, regions with higher system noise yield higher source amplitude, although activity depends not only on the input term but also on the source activity predicted by the dynamic model. Therefore, the variance of the system noise can be used to specify the spatial prior from fMRI data. In our method, the fMRI prior is constructed by assigning higher-magnitude input terms (i.e., higher system noise variance) to regions of higher t-value. This is achieved by basing the prior distribution of $\mathbf{q}$ on the magnitude of fMRI t-values (as described in the next subsection).

4.2.2 Prior distributions on model parameters

The unknown parameters estimated from the data in the observation and dynamic source models (Eqs. (2.4) and (4.5), respectively) are β, $\mathbf{A}$, and $\mathbf{q}$. For these parameters, the following prior distributions are assumed. The prior distribution of the scaling parameter β is a non-informative Jeffreys prior:

$$P(\beta) \propto \beta^{-1}. \tag{4.6}$$

For the MAR matrix $\mathbf{A}$, the following Gaussian prior distribution is imposed on the MAR coefficients in an element-wise manner:

$$\begin{aligned} P(\mathbf{A} \mid \boldsymbol{\eta}_{1:N}) = \prod_{n=1}^{N} P(\mathbf{a}_n \mid \boldsymbol{\eta}_n) &= \prod_{n=1}^{N} \mathcal{N}\left(\mathbf{a}_n \mid \mathbf{0}, \mathrm{diag}(\boldsymbol{\eta}_n)^{-1}\right) \\ &= \prod_{n=1}^{N} \prod_{k=1}^{K_n} \mathcal{N}\left(a_{n,k} \mid 0, \eta_{n,k}^{-1}\right), \end{aligned} \tag{4.7}$$

where $\mathbf{a}_n$ is a column vector containing all non-zero entries of the n-th row of $\mathbf{A}$ in ascending order, and K_n is the number of sources structurally connected to the n-th source. In addition, we impose the Gamma prior distribution on the inverse variance $\eta_{n,k}$ in Eq. (4.7), and thereby apply the ARD sparse prior [58] to the MAR coefficients:

$$\begin{aligned} P(\boldsymbol{\eta}_{1:N}) = \prod_{n=1}^{N} P(\boldsymbol{\eta}_n) &= \prod_{n=1}^{N} \prod_{k=1}^{K_n} P(\eta_{n,k}) \\ &= \prod_{n=1}^{N} \prod_{k=1}^{K_n} \mathcal{G}(\eta_{n,k} \mid \eta_0, g_0), \end{aligned} \tag{4.8}$$

where η_0 is the mean and g_0 is the shape parameter of the Gamma prior distribution (common to all entries of $\boldsymbol{\eta}_{1:N}$). The ARD sparse prior, comprising Eqs. (4.7) and (4.8), effectively prunes the coefficients associated with very small source amplitudes, ensuring that solutions are robust toward noise and modeling errors. Finally, we set the prior distribution of the parameter $\mathbf{q}$ as

$$P(\mathbf{q}) = \prod_{n=1}^{N} P(q_n) = \prod_{n=1}^{N} \mathcal{G}(q_n \,|\, \bar{\nu}_n^{-1}, \gamma_0), \tag{4.9}$$

where $\bar{\nu}_n$ is the prior mean of the system noise variance (scaled by β), and γ_0 quantifies the reliability of this prior. To incorporate the fMRI prior into Eq. (4.9), $\bar{\nu}_n$ is weighted by the fMRI t-value w_n, normalized from zero to one:

$$\bar{\nu}_n = \nu_0 + (m_0 - 1)\, \nu_0 \cdot w_n^2. \tag{4.10}$$

In Eq. (4.10), ν_0 denotes the variance of the source amplitude averaged over the whole brain during a pre-stimulus period, obtained by Bayesian minimum norm estimation [88], and $m_0\,(>1)$ specifies the relative difference of variance intensity, ranked from the lowest (or statistically thresholded) to the highest t-value region. In this weighting scheme, regions of higher t-value receive larger task-related input activity; up to m_0 times larger than the changes associated with averaged resting activity.

4.2.3 Joint probability distribution

The likelihood function and all prior distributions (Eqs. (2.4) and (4.5)–(4.9)) constitute the joint probability distribution:

$$\begin{aligned} &P(\mathbf{B}_{1:T}, \mathbf{J}_{1:T}, \beta, \mathbf{A}, \boldsymbol{\eta}_{1:N}, \mathbf{q}) \\ &\quad = \prod_{t=1}^{T} \left\{ P(\mathbf{B}_t \,|\, \mathbf{J}_t, \beta)\, P(\mathbf{J}_t \,|\, \mathbf{J}_{t-\Delta_1}, \ldots, \mathbf{J}_{t-\Delta_L}, \beta, \mathbf{A}, \mathbf{q}) \right\} \\ &\qquad\qquad \times P(\beta)\, P(\mathbf{A} \,|\, \boldsymbol{\eta}_{1:N})\, P(\boldsymbol{\eta}_{1:N})\, P(\mathbf{q}), \end{aligned} \tag{4.11}$$

The following analyses (other than that of the simulated MAR time series) assume that the analyzed data are trial-averaged evoked responses and that the time samples $-\Delta_L \leq t \leq 0$ ($\Delta_L \approx 70$ ms) constitute part of the rest

period. Therefore, the current sources $\mathbf{J}_t$ (where $t \leq 0$), are set to zero. By slightly modifying the dynamic source model, $\mathbf{J}_t$ $(t \leq 0)$ could be set as model parameters estimated from the data.

4.2.4 Estimation algorithm

All of the unknown variables in Eq. (4.11) are simultaneously estimated by calculating their approximate posterior distributions using the variational Bayesian algorithm. This estimation scheme enables an exploratory analysis of the effective connectivity across the whole brain. The pseudocode of the estimation algorithm is given below:

1. Initialize the values required for updating $Q(\mathbf{A})$, and the hyperparameters η_0, g_0, m_0, and γ_0 in Eqs. (4.8)–(4.10).

2. **A**-step: Update $Q(\mathbf{A})$.

3. $\boldsymbol{\eta}$-step: Update $Q(\boldsymbol{\eta}_{1:N})$.

4. **q**-step: Update $Q(\mathbf{q})$.

5. **J**-step: Update $Q(\mathbf{J}_1), \ldots, Q(\mathbf{J}_T)$ in this order.

6. β-step: Update $Q(\beta)$, and compute the free energy.

7. If the maximization of the free energy converges, then exit; otherwise, return to the **A**-step.

Each update rule for computing the approximate posterior distribution is derived from Eq. (2.14) by substituting a variable of interest for x (the specific forms of the update equations are presented in Appendix B.1).

In practice, the algorithm was terminated after 100 iterations, when the relative change in free energy was typically of order 10^{-5} or 10^{-6}. The computation time ranged from approximately 12 to 48 hours (running on a multicore PC: 8 cores, 3.60 GHz CPU, and 96 GB RAM), depending on the number of structural connections in the individual brain.

4.2.5 Initial values and hyperparameters

At the first update of the approximate posterior distribution $Q(\mathbf{A})$ in the **A**-step, initialization of the remaining approximate posterior distributions is necessary. In particular, the means of all these distributions and the noise covariance matrices of $Q(\mathbf{J}_1), \ldots, Q(\mathbf{J}_T)$ must be specified beforehand. The initial mean for all the elements of $Q(\boldsymbol{\eta}_{1:N})$ is set to their prior mean η_0. The rest of the required initial values is determined from the non-dynamic solutions obtained by the hierarchical variational Bayesian method [71, 88], where the probabilistic model in [88] is identical to the model in the proposed dynamic method if all entries of **A** are fixed to zero.

In addition, the hyperparameters η_0, g_0, m_0, and γ_0 that control the prior distributions of $\boldsymbol{\eta}_{1:N}$ and $\mathbf{q}$ must be pre-specified. The default values used in the simulations are listed below.

- The hyperparameter η_0, also used as the initial values for $\boldsymbol{\eta}_{1:N}$, suppresses instabilities in the estimated dynamics. If η_0 is very small, the MAR coefficients can amplify, which consequently lead to divergence of the source amplitude. On the other hand, if η_0 is extremely large, all estimated MAR coefficients are close to zero. Thus, we set the default η_0 to a moderate value (such as 100).

- The hyperparameter m_0 in the fMRI spatial prior regulates the maximum strength of the system noise variance relative to the baseline. We set the default m_0 to 100, a value recommended by [88] in the non-dynamic counterpart to the proposed method.

- In Section 4.4, the hyperparameters g_0 and γ_0 were set by one of two strategies. When the data were generated from the MAR model identical to the adopted dynamic source model, both g_0 and γ_0 were set to zero, yielding non-informative priors. When discrepancies were found between our model formulation and the data generation process, g_0 and γ_0 were set to 100 to realize informative priors. Such priors enhance the robustness of the estimates to modeling errors (see Chapter 3).

The optimal hyperparameter values in the event of modeling errors were

searched by replacing the values specified above (all 100) in Section 4.4, and these values were applied in Section 4.5.

4.3. Methods

This section briefly introduces the analyzed dataset, and presents the data processing procedures from which the constituents of the model are derived, namely, the structural connections, the time lags, and the lead field matrix. We then explain how the estimation performance is evaluated in simulations. Finally we describe the procedure to estimate the source amplitudes and interactions, used in the non-dynamic counterpart to the proposed method.

4.3.1 Dataset

Simulation and real data analyses were conducted on a publicly available experimental dataset[2] (the data acquisition and task conditions are detailed in [43]). The dataset contains the structural images (T1- and diffusion-weighted MRI) and functional data (MEG and fMRI data collected during face perception) required for our method. Within this dataset, the entire set of MRI, dMRI, MEG, and fMRI data were archived for 11 out of 16 subjects (Sub01–Sub06, Sub09, Sub12–15); these data were used in the real data analysis. The incomplete data of the remaining six subjects were not used; dMRI data are absent for Sub07, Sub08, Sub10, Sub11, and Sub16. Simulation data were generated from the brain data of a single subject (Sub12). In the following, the ten original subject codes (in ascending order) are renamed Subject I-Subject XI.

[2]The dataset can be downloaded from `ftp://ftp.mrc-cbu.cam.ac.uk/personal/rik.henson/wakemandg_hensonrn/`. We would like to thank Daniel Wakeman and Dr. Richard Henson for providing the data.

4.3.2 Diffusion MRI analysis

Data processing

The structural connections and time lags for the dynamic source model were inferred from fiber tracking with structural and diffusion MRI data (T1 weighted and diffusion weighted images). All MRI data (including fMRI) were acquired on a 3T Trio (Siemens, Erlangen, Germany). The spatial resolutions of the T1-weighted and the diffusion weighted images were 1 mm and 2 mm isotropic, respectively. During dMRI acquisition, the number of gradient directions was 64 and the b-value was 1000 $\mathrm{s/mm^2}$. To pre-process the dMRI data, subject motion during the image acquisition was corrected by FSL[3]; the consequent bias on the gradient directions was compensated afterwards by rotating the b-vector. The fractional anisotropy image was then calculated from the corrected images and was used for registering the diffusion-space to the T1-space by the non-linear registration tool (FNIRT) in FSL.

The seed and target ROIs used for fiber tracking were obtained by parcellating the cortical surfaces extracted by FreeSurfer[4]. In the cortical parcellation, 2000 vertices were first specified by applying the Matlab function `reducepatch` to the high resolution white/gray matter boundary surfaces provided by FreeSurfer. All of the highly-resolved surface vertices were then clustered into 2000 parcels in a nearest-neighbor manner. The vertices in subcortical regions were discarded, and the remaining parcels (numbering around 1840) constituted the final ROIs used for fiber tracking. These surface ROIs were converted to volume ROIs, and then transformed to the diffusion-space.

Fibers were tracked using MRtrix[5]. The local model of fiber orientations was the fiber orientation distribution (FOD), reconstructed at each voxel by constrained spherical deconvolution [79] with six-dimensional spherical harmonics for the response function. Based on the reconstructed FOD, fibers were probabilistically tracked with a step length of 0.2 mm within a mask of white matter volume. The fiber tracks were generated 10^5 times from each ROI and unidirectionally stepped until they entered another ROI or until the

[3] `http://www.fmrib.ox.ac.uk/fsl`

[4] `http://surfer.nmr.mgh.harvard.edu/`

[5] `http://www.nitrc.org/projects/mrtrix/`

total fiber length reached 300 mm. The fibers were also terminated at voxels of FOD amplitude less than 0.1. The minimum fiber length and radius of curvature were set to 10 mm and 1 mm, respectively.

Structural connections

The structural connectivity was quantified by the fiber counts. The strength of the connectivity was computed as the number of fibers within each ROI pair f_t divided by the total number of fibers generated from its seed ROI f_s with voxel size normalization; that is, $(f_t/v_t)\,/\,(f_s/v_s)$, where v_t and v_s are the number of voxels in the target and the seed ROI, respectively. Since structural connectivity conveys no directional information, the connectivity strengths were symmetrized by assigning the higher strength to both directions.

To determine the structurally connected ROI pairs, the structural connectivity strengths were binarized. The binarized connections were used for specifying pairs of structurally connected *sources* (and hence for determining the MAR coefficients to be estimated from the data), since we placed a single current source at each ROI when computing the lead field matrix (as explained later in this section). The binarization threshold was set as low as possible (in our case, 10^{-4}). While a lower threshold prolongs the computation time, it reduces the risk of missing true structural connections. Indeed, a lower threshold is important for estimating the source interactions in our method. The false negative structural connections set the MAR coefficients to zero, even when the corresponding pair of sources is effectively connected. On the other hand, MAR coefficients with false positive structural connections are not necessarily increased; the sparse prior on the MAR matrix can make these coefficients go to zero. After the thresholding, between 5–9% of all ROI pairs were labeled as structurally connected (the precise percentage depended on the individual subject).

Time lags

The time lags in the MAR model were computed from the inter-ROI fiber length, the conduction velocity, and the local delay constant (see Eq. (4.2)).

- The fiber length was obtained by first summing the tracking steps of each fiber track. The lengths of fibers included in each pair of ROIs were then averaged. The mean fiber lengths were finally symmetrized by weighted averaging of both directional fiber lengths by their (pre-symmetrized) connectivity strengths.

- The conduction velocity was set to 6 m/s, as assumed for myelinated axons in a previous simulation study [31].

- The local delay constant was defined as the time between an input signal and peak population-level neuronal activity evoked by this input. It was set to 27 ms, estimated from the peak to peak time difference between the source activities reconstructed in the primary and secondary visual areas (about 30 ms; [68]). Here we accounted for the time delay introduced by propagation of the source activity along the structural connection. This delay was estimated at about 3 ms, based on the distance between the primary and secondary visual areas and the above velocity setting.

The time lags in the self-connections (the diagonal entries in the MAR matrix) were set to the local delay constant. The impact of conduction velocity and local delay constant on the estimation performance was investigated in Section 4.4.

4.3.3 Lead field matrix

To calculate the lead field for each MEG sensor location, current dipoles of unit strength were successively placed at all cortical vertex positions specified in the ROI generation, and the MEG forward solution was computed. The sensor positions were matched to the positions of 102 magnetometers in a VectorView system (Elekta Neuromag, Helsinki, Finland), and were linearly registered to the T1-space using the fiducials and head points provided in the real dataset. The dipole was oriented perpendicular to the white/gray matter boundary surface. The forward solution was computed by a boundary

element method (BEM) using a single-shell head model, based on the inner skull surface obtained from FreeSurfer.

The lead field matrix was locally smoothed by a Gaussian smoothing filter with 8 mm full width at half maximum (FWHM), equivalent to assuming a spatial smoothness prior on the current sources [88]. The smoothing relationship is given by $\mathbf{J}_t = \mathbf{W}\mathbf{Z}_t$, where $\mathbf{J}_t$ and $\mathbf{Z}_t$ are the smoothed and the unsmoothed current sources, respectively, and $\mathbf{W}$ is the spatial smoothing matrix. Substituting this equation into Eq. (2.1), the actual observation model is obtained as $\mathbf{B}_t = \hat{\mathbf{G}}\mathbf{Z}_t + \varepsilon_t$, where $\hat{\mathbf{G}} = \mathbf{G}\mathbf{W}$ is a smoothed lead field matrix. Using this smoothness prior, we estimated the MAR matrix $\mathbf{A}$ for $\mathbf{Z}_{1:T}$ rather than $\mathbf{J}_{1:T}$. To clearly demonstrate the correspondence between the source dynamics and the MAR coefficients, all results presented in the following sections are based on the unsmoothed sources $\mathbf{Z}_{1:T}$, unlike Chapter 3.

4.3.4 Performance evaluation

The estimation performance in simulations was quantified by two threshold-free evaluation scores; the area under the ROC curve (AUC) and the normalized root mean square (RMS) error.

The AUC measures the detection accuracy [35]. The AUC scores for the current sources and the MAR matrix were calculated from the RMS of the estimated source amplitudes and the estimated absolute MAR coefficients, respectively. Prior to calculating the AUC, both quantities were normalized by their maximum values over all sources and connections. By comparing the normalized estimates to the ground truth, we computed the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates for thresholds ranging from zero to one. We then calculated the sensitivity and specificity for each threshold, defined as

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ and specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{4.12}$$

A plot of sensitivity versus (1−specificity) yields the ROC curve. The AUC was determined by the trapezoidal rule (the potential bias on AUC intro-

duced by the discrepancy between the numbers of actives and inactives was corrected as described in Appendix B.2).

The normalized root mean square error (nRMSE) quantifies the differences between the estimates and the ground truth. The nRMSE scores of the current sources (or the MAR matrix if the ground truth existed) were computed as the RMS error between the simulated and estimated source amplitudes over all sources and time instances (or MAR coefficients over all connections), divided by the RMS of the simulated ones.

When reconstructing the simulated MAR time series, we evaluated the estimation accuracy by AUC, nRMSE, and two additional evaluation measures that provide an intuitive understanding (see Simulation 1 in Section 4.4). One additional measure is the reconstruction gain, defined as the ratio of the time-averaged estimated source amplitude to the true source amplitude. The other is the non-normalized RMSE of the MAR coefficients.

We also applied thresholds on the estimated sources and MAR coefficients when mapping the results onto the cortical surfaces and when distinguishing between active and inactive sources and connections. In particular, these thresholds were used for qualitative evaluation of the estimation performance when the method is applied to experimental data. Other than when calculating AUC, a source was considered active if its magnitude once exceeded one-tenth of the maximum over all sources. A connection was regarded as effectively connected if its absolute MAR coefficients exceeded 0.1. Although these thresholds were determined in an *ad-hoc* manner, small differences in the thresholds did not alter the relative characteristics of the estimates when comparing the proposed dynamic method with its non-dynamic counterpart (the non-dynamic method is described below).

4.3.5 Method comparison

To demonstrate advantages of dynamic modeling and simultaneous estimation, the estimates obtained by our method were compared with those of its non-dynamic equivalent, in which every pair of sources are assumed non-interacting during source reconstruction. In the non-dynamic method, the current sources were initially reconstructed with all MAR matrix entries set

to zero. This is equivalent to reconstructing the current sources by the hierarchical variational Bayesian method described in [88]. The MAR coefficients were then separately estimated from the reconstructed current sources. These coefficients were computed only for structurally connected pairs of sources; the coefficients of unconnected pairs were set to zero. The MAR coefficients were estimated from the algorithm used in the proposed dynamic method, with modification of omitting the **J**- and β-steps.

In Appendix B.4, we compared with the non-dynamic method whose estimates were obtained from a conventional approach. Here the MAR coefficients were not computed from sources over the whole brain but only from active sources. This procedure of estimating dynamic parameters only from pre-specified regions is similar to the estimation steps of DCM [16, 26]. As in comparison with the non-dynamic method in Sections 4.4 and 4.5, the MAR coefficients were separately estimated from the reconstructed sources. These coefficients were also computed for structurally connected pairs but only from active sources, i.e., the coefficients of unconnected pairs and pairs including any inactive sources were set to zero. These were estimated from the algorithm used in the dynamic method, with the following modifications: the **J**- and β-steps were omitted and a step for extracting the active sources was added. By the method comparison in Appendix B.4, we investigated advantages of the proposed dynamic method, including those coming from whole-brain estimation of the effective connectivity.

GoF analysis

When comparing the estimates obtained from real data, we used a goodness of fit (GoF) measure:

$$\mathrm{GoF} = 100\left\{1 - \frac{\sum_{m=1}^{M}\sum_{t=1}^{T}(B_{m,t} - (\mathbf{G}\bar{\mathbf{J}}_t)_{m,t})^2}{\sum_{m=1}^{M}\sum_{t=1}^{T}B_{m,t}^2}\right\}. \tag{4.13}$$

to quantify the accuracy of reconstructed sources $\bar{\mathbf{J}}_{1:T}$ and estimated dynamics. GoF increases when the error between observed and reconstructed measurements decreases. Although a high GoF score does not always grantee

high quality of source reconstruction in an underdetermined setting ($M < N$), a high value of GoF is a necessary condition of a good inverse solution.

In addition to simply calculating GoF, from sensor measurements averaged across all trials (see Fig. 4.1A), the GoF score was computed from the data averaged across a half of all trials (see Fig. 4.1B). Here, for evaluating generalizability of the methods, model parameters (i.e., random variables other than mean and covariance of current sources) were pre-estimated from the data averaged over the other half of all trials. The parameters were fixed to the pre-estimated values when reconstructing sources for the GoF calculation.

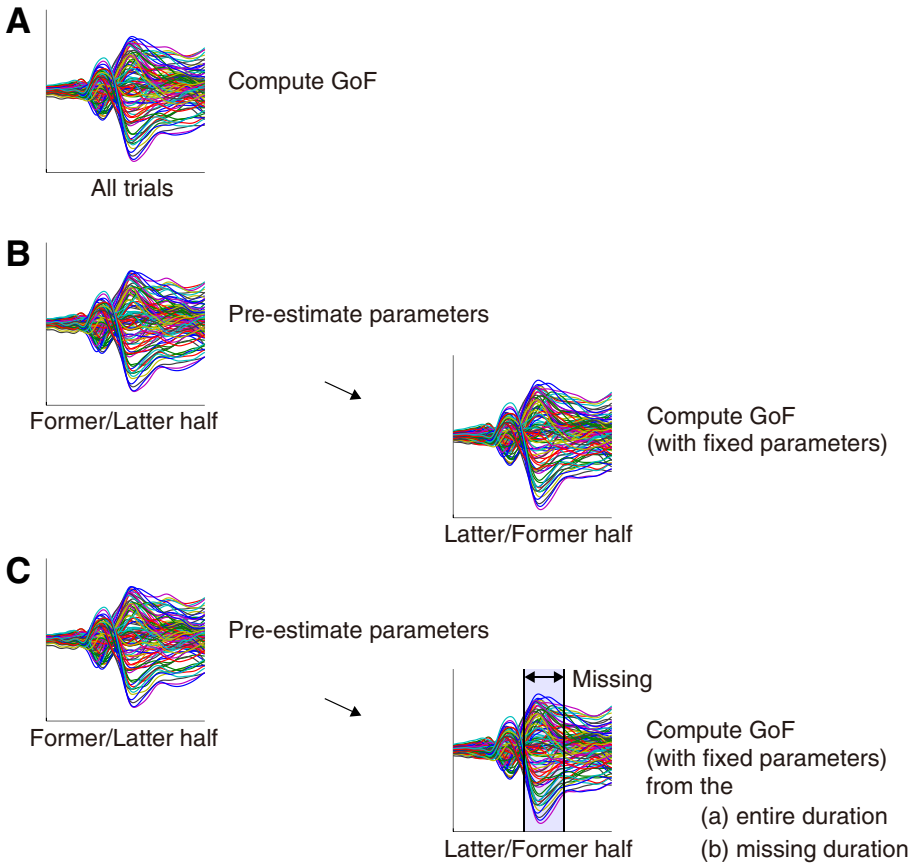

Figure 4.1: A schematic view of GoF calculation (detailed description for the three ways of GoF calculation (A–C) are presented in text). When computing GoF for (B) and (C), all trials were splitted into the former and latter half. A GoF score for each subject was obtained by averaging a set of two GoF scores computed from the former and latter half.

Furthermore, we assessed prediction accuracy of the dynamic model to evaluate estimated source dynamics. The procedure of computing GoF was almost the same when the generalizability was evaluated, but here the data contained a missing duration (135–210 ms after the stimulus onset; see Fig. 4.1C). The model parameters were estimated from intact data, whereas the GoF measure was calculated using sources reconstructed from data containing a missing duration. The sources at a missing duration were computed by forward prediction of the estimated dynamic model (no system noise was assumed in forward prediction). Along with computing GoF from the entire duration of data, we also calculated the GoF score only from the missing duration.

4.4. Simulations

4.4.1 Simulation 1: MAR time series

We first examined the estimation performance on a simulated MAR time series. The data were generated from the dynamic source model assumed in the proposed method. The purpose of this simulation was to evaluate the identification accuracy of the MAR model with varying numbers of active sources and effective connections.

Settings

The number of active sources was set to 5, 10, 20, or 30 out of 1841 cortical sources. The time courses of the active sources were generated from the dynamic source model. The diagonal MAR matrix entries of active sources were set to 0.4, while structurally-connected non-diagonal entries were assigned random values uniformly distributed between -0.4 and 0.4. All other entries of the MAR matrix were set to zero. The system noise variance of active and inactive sources was set to 20 nAm^2 and zero, respectively. Gaussian simulated observation noise was assumed; the normalized noise covariance matrix was computed from empty room data obtained from the real dataset. The scaling parameter was adjusted to yield an SNR of 5 dB in the simulated sen-

sor measurements, a typical value in trial-averaged measurements of simple stimulus-evoked responses.

For each number of active sources specified above, we conducted 20 Monte Carlo simulation trials. In each simulation, the active source positions were sequentially and randomly selected from source positions that were structurally connected to one of the selected source positions. The current sources and the MAR matrix were estimated from the simulated measurements (duration 400 ms and sampling frequency 1 kHz) with known noise covariance structure and time lags. Throughout this simulation study, non-informative priors were imposed on the model parameters (i.e., $g_0 = \gamma_0 = 0$) and no fMRI information was provided.

Results

The estimation accuracies of the non-dynamic and dynamic methods were first compared by computing the AUC and nRMSE of their estimated current sources and MAR matrices (see Fig. 4.2). When calculating the AUC scores, the simulated source positions and non-zero MAR coefficients were assumed as the true sources and connections. Both methods yielded almost identical AUCs. However, the nRMSE scores were improved in the dynamic method. This demonstrates the advantage of explicitly accounting for the source dynamics, and simultaneously reconstructing the current sources and MAR coefficients.

To investigate the estimates of the dynamic method in detail, the reconstruction gain of the current sources and the RMSE of the MAR matrix were computed. The results are listed in Table 4.1. The reconstruction gains were obtained by averaging the gains across the active source positions (upper line) and the positions at which the true positive sources were successfully estimated (lower line). The RMSE scores were calculated from a subset of the MAR matrix, restricting the row and column indices to the locations of the originally active sources (upper line) and the true positive sources (lower line). This table reveals near-perfect source reconstruction and model identification for 5 active sources. As the number of active sources increased, the estimation performance deteriorated. In addition, sources at less observable

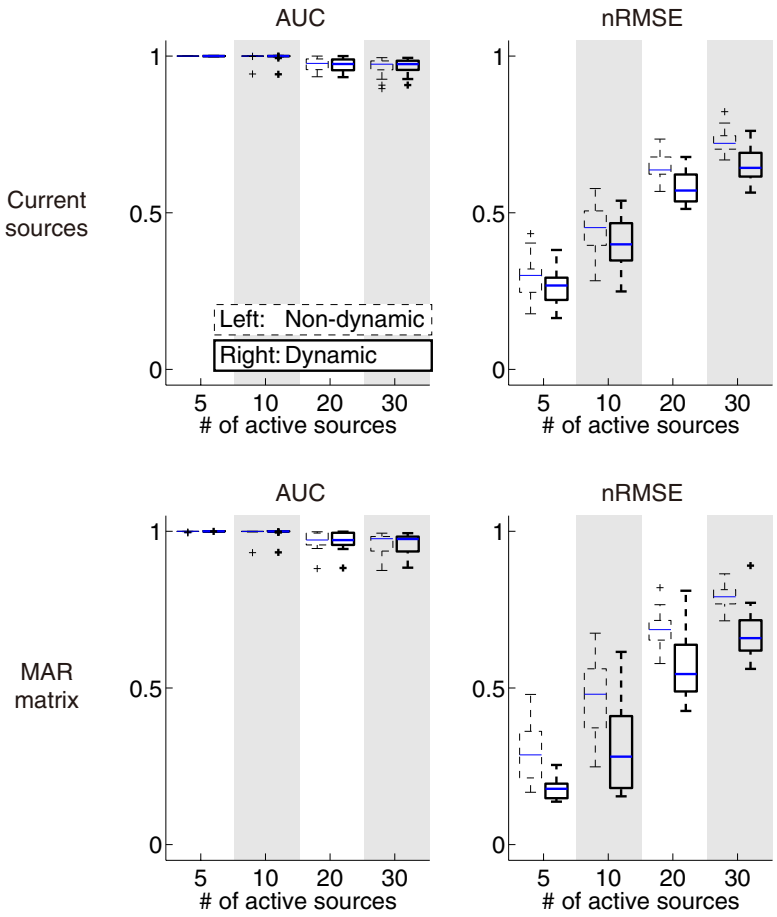

Figure 4.2: Evaluation of the estimation accuracy in Simulation 1. AUC and nRMSE of the current sources (upper) and the MAR matrix (lower) are presented as boxplots. Each panel displays the distributions of AUC or nRMSE estimated by the non-dynamic and dynamic methods. The blue bar of each box indicates the median and the edges show the upper and the lower quartiles. The length of the whisker is the default length set in the Matlab function `boxplot`. Method comparison with a conventional approach of the non-dynamic method is presented in Fig. B.1.

cortical positions (with small lead field norms) failed to be identified (data not shown). Similarly, connections with small products of the corresponding lead field norms tended to be missed. Moreover, the estimation accuracy decreased for smaller-magnitude current sources and MAR coefficients among the simulation data. Despite these anomalies, Table 4.1 shows that even for source numbers of 20 and 30, more than 75% (true positive rate) of the orig-

Table 4.1:
Reconstruction gains of the current sources and RMSEs of the MAR matrix, estimated from the originally active and true positive sources in Simulation 1.

Number of active sources	5	10	20	30
Reconstruction gain				
of the active sources	0.95	0.88	0.73	0.66
of the true positive sources	0.95	0.91	0.85	0.84
RMSE of the MAR matrix				
between the active sources	0.05	0.09	0.15	0.18
between the true positive sources	0.05	0.07	0.10	0.12
True positive rate	1.00	0.97	0.85	0.78

The displayed values were computed from a subset of the estimates obtained by the dynamic method. Each entry displays the mean of 20 simulations.

inally active sources were reconstructed with 1) reconstruction gains greater than 0.80 and 2) RMSE of the MAR matrix less than 0.15.

4.4.2 Simulaiton 2: Stimulus-evoked responses

Next, to examine the estimation accuracy in a more realistic simulation setting, we applied the methods to synthesized stimulus-evoked responses. The event-related dynamics inherent in the simulation data was generated from a functional network of the neural mass models with external inputs.

Settings

A schematic of the simulation setting is shown in Fig. 4.3. The time courses of stimulus-evoked responses, with sampling frequency 1 kHz, were generated from a network of non-linear neural mass models (the constituents of this network model are detailed in Appendix B.3). When computing the exact time courses of the active sources, the waveforms generated from the network were multiplied by a constant to ensure that the resulting sensor measurements and

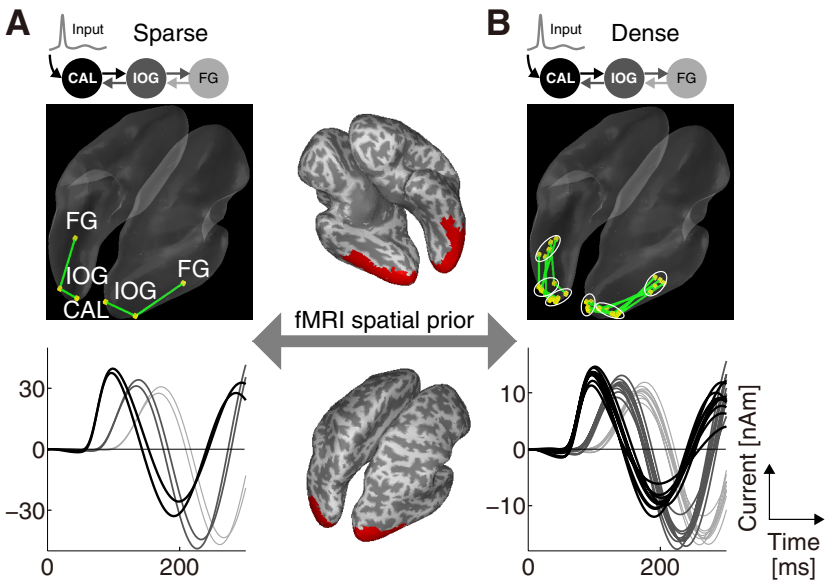

Figure 4.3: Data generation settings in Simulation 2. (A) sparse setting (B) dense setting. The time courses of the active sources in CAL, IOG, and FG are plotted in black, dark gray, and light gray, respectively. The red regions on the cortical surfaces are the fMRI positive regions assumed in this simulation study.

the real measurements in the dataset were of comparable order of magnitude. We adopted a simplified network model of a face perception task comprising six regions of interest; the bilateral lower visual cortices around the calcarine sulcus (CAL) and the higher visual cortices located in the core regions of face perception [40], the inferior occipital gyrus (IOG) and the fusiform gyrus (FG). The remaining core region, the superior temporal sulcus (STS), was excluded from the network model because direct structural connections between STS and IOG/FG were not found in previous dMRI studies [36, 69].

We examined the estimation performance under two simulation scenarios (see Figs. 4.3A and 4.3B) with different numbers of active sources in each ROI; one (the sparse setting) and five (the dense setting). Each active source in the CAL of both hemispheres received an input signal. The CAL activity was sequentially transmitted to the active sources in IOG and FG in a one-to-one manner via the bidirectional effective connectivity between CAL–IOG and IOG–FG (the input signal and connection parameters are also detailed in Appendix B.3). Note that all forward and backward effective connections

were consistent with their structural connections. For simplicity, the lateral effective connectivity within each ROI was ignored in this simulation.

An identical fMRI prior was assumed in both sparse and dense settings. To obtain this spatial prior, we first binarized the t-value map of the real fMRI data in the dataset, and then masked the binarized areas by extracting regions close to CAL, IOG, and FG (the t-value map was computed as described in Section 4.5). The fMRI positive regions, spread across 69 cortical source positions, include all active sources placed in both simulation settings. Since the t-values were binarized, the fMRI positive and negative regions were assigned $w_n = 1$ and $w_n = 0$, respectively, in Eq. (4.10). The advantages of applying the fMRI prior were assessed by comparing the results with those obtained without the fMRI prior (assigning $w_n = 1$ to the whole cortex).

The observation noise was real background MEG data in the dataset, comprising trial-averaged sensor measurements taken during a 400 ms rest period. During the trial averaging, potential event-related components in the resting data were suppressed by randomly forward-shifting the time course of each trial, then translating the fragmented data projected from the time window to the end of the shifted time series. The magnitudes of the resulting trial-averaged measurements were comparable to those of the actual resting data. The first 100 ms duration of the averaged background MEG data were used to fit the normalized noise covariance matrix; the remaining 300 ms were added to the noise-free simulated measurements.

Various modeling errors typically encountered in real data applications, such as non-linear source dynamics and non-Gaussian observation noise, were introduced to this simulation study. To prevent the entry of possible estimation errors caused by such model discrepancies, we applied informative priors (including the fMRI prior) on the model parameters; that is, we set g_0 and γ_0 to non-zero values. Optimal values of g_0 and γ_0, as well as η_0 and m_0, were searched by varying the values of the hyperparameter sets (η_0, g_0) in $P(\boldsymbol{\eta}_{1:N})$ and (m_0, γ_0) in $P(\mathbf{q})$ from their default settings $(100, 100)$.

Besides varying the hyperparameters, we also examined the impact of changing the delay parameters in the MAR model, namely, the conduction velocity v and the local delay constant τ, from their default values ($v = 6$ m/s

and $\tau = 27$ ms).

Results

The AUC and nRMSE scores with and without the fMRI prior are displayed in Table 4.2. Clearly, applying the fMRI prior improved the estimation performance. In both simulation scenarios, the best score was consistently achieved by the dynamic method with the fMRI prior. In the dense simulation setting, however, the dynamic method with the fMRI prior still yielded an nRMSE score around unity, reflecting the difficulty of reconstructing a large number of closely-spaced active sources. Nevertheless, the mean activity across sources within an ROI was reasonably reconstructed in this scenario. The nRMSE score computed from the time courses averaged over the five active sources within each ROI (denoted by ROI-nRMSE in Table 4.2) was improved to 0.39. The waveforms of current sources were also appropriately estimated;

Table 4.2:
AUC and nRMSE of the current sources and AUC of the MAR matrix with (and without) applying the fMRI prior in Simulation 2.

	Non-dynamic		Dynamic	
	Sparse	Dense	Sparse	Dense
Current sources				
AUC	**1.00** (0.77)	0.99 (0.51)	**1.00** (0.99)	**1.00** (0.94)
nRMSE	0.18 (0.66)	0.97 (1.19)	**0.15** (0.64)	**0.86** (1.13)
ROI-nRMSE		0.45 (0.80)		**0.39** (0.77)
MAR matrix				
AUC	**1.00** (0.69)	0.89 (0.41)	**1.00** (0.87)	**0.91** (0.76)

The scores with no fMRI prior applied are listed in parentheses. The best scores in each simulation scenario across the methods, applied with and without the prior, are indicated in bold font. ROI-nRMSE displays the nRMSE of the ROI-averaged time courses. Method comparison with a conventional approach of the non-dynamic method is presented in Table B.1.

correlation coefficient between the true and estimated activities (not shown in the table) was computed as 0.65, using the entire time series of all sources.

The current sources and the MAR coefficients estimated from the methods with the fMRI prior are shown in Fig. 4.4. In the sparse setting, both the non-dynamic and the dynamic methods successfully reconstructed the active sources and identified their effective connectivity (see Fig. 4.4A), although the non-dynamic method detected a single false positive source. On the other hand, the dense cortical activity and effective connectivity was imperfectly estimated by both methods (see Fig. 4.4B). False positive sources were estimated within the regions covered by the fMRI prior; and false positive connections emerged due to the limitations of the linear autoregressive modeling. In fact, such errors in the estimated connections occurred even when the MAR coefficients were fitted to the true current sources (data not shown). Despite these shortcomings, the dynamic method showed higher sensitivity of current source reconstruction and identification of effective connectivity than its non-dynamic counterpart (see Fig. 4.4B). In the dynamic method, 23 out of 30 active sources were reconstructed and at least one single effective connection was discovered, i.e., the absolute MAR coefficient was larger than 0.1, at each functional ROI pair. In contrast, the non-dynamic method missed half of the active sources (15 out of 30), and detected no effective connection for several functional ROI pairs.

The evaluation scores obtained under various hyperparameter settings are presented in Fig. 4.5A. The hyperparameters η_0 and g_0 exerted little effect on the AUC and nRMSE of the current sources. However, increasing η_0 and g_0 in the dense setting enlarged the AUC of the MAR matrix. Nevertheless, we recommend a moderate value of η_0 (say 100) because $\eta_0 = 1000$ tended to suppress the number of estimated effective connections (data not shown). When η_0 was increased from 100 to 1000, the sensitivity of the effective connectivity decreased about from 0.3 to 0.1. In determining g_0, we observed that the perfect source reconstruction in the sparse setting (no false positives or negatives) was achieved only for $g_0 = 100$ or 1000 (and $\eta_0 = 100$). Thus, our recommended hyperparameter setting in Eq. (4.8) is $(\eta_0, g_0) = (100, 100)$ or $(100, 1000)$. The right hand side of Fig. 4.5A shows that high values of the

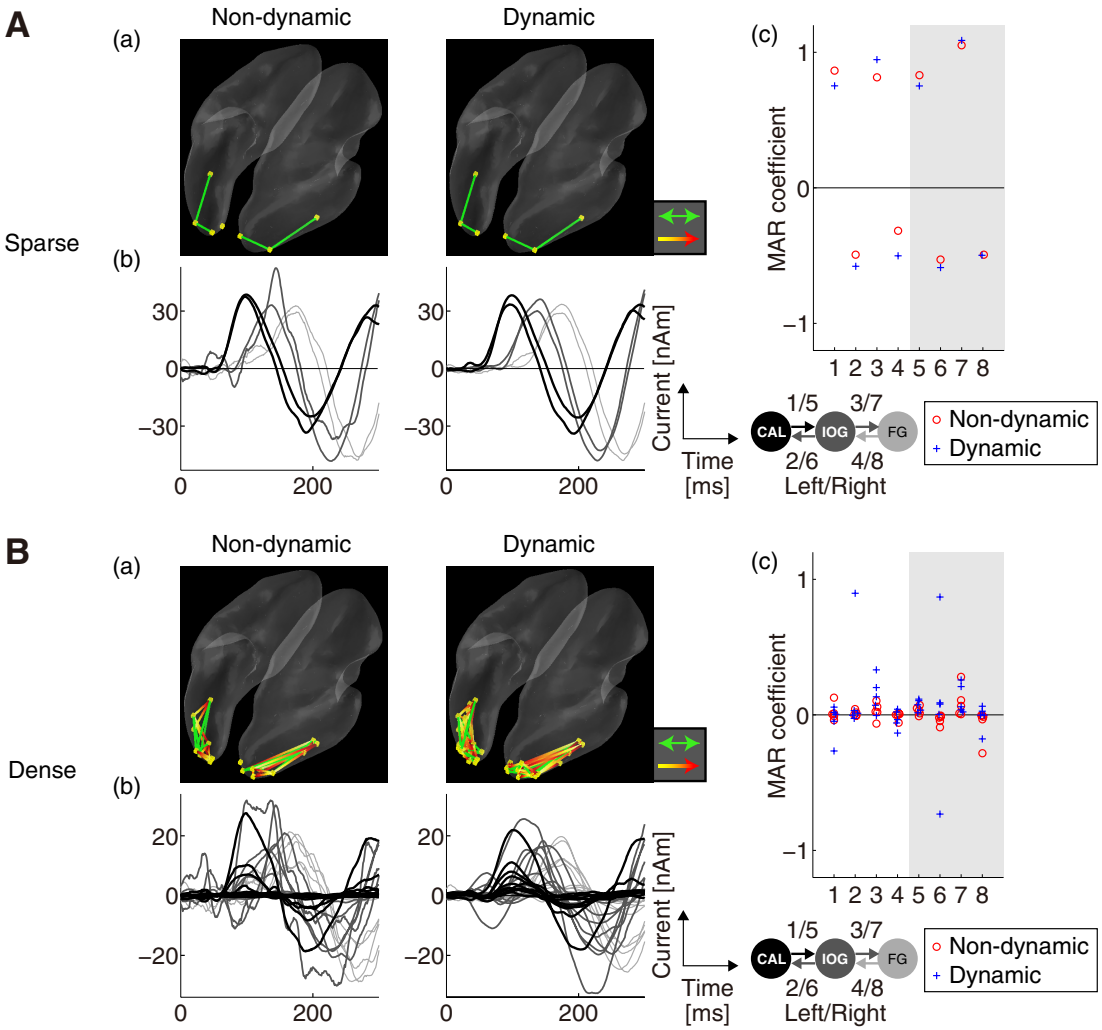

Figure 4.4: Results obtained from the non-dynamic and dynamic methods with the fMRI prior in Simulation 2. (A) sparse setting and (B) dense setting. (a) Estimated effective connectivity among the positive sources indicated by yellow dots. Green lines, and lines graduated from yellow to red, denote bi- and uni-directional effective connectivity, respectively. Note that line color is not related to the magnitude of the MAR coefficients. (b) Estimated time courses of the current sources whose cortical positions matched the originally active sources. The source time courses (gray scale) are explained in the caption of Fig. 4.3. (c) Estimated MAR coefficients at source pairs of true effective connectivity. The horizontal axis indicates the locations of the inter-ROI connections (1–4: connections in the left hemisphere; 5–8: connections in the right hemisphere). Method comparison with a conventional approach of the non-dynamic method is presented in Fig. B.2.

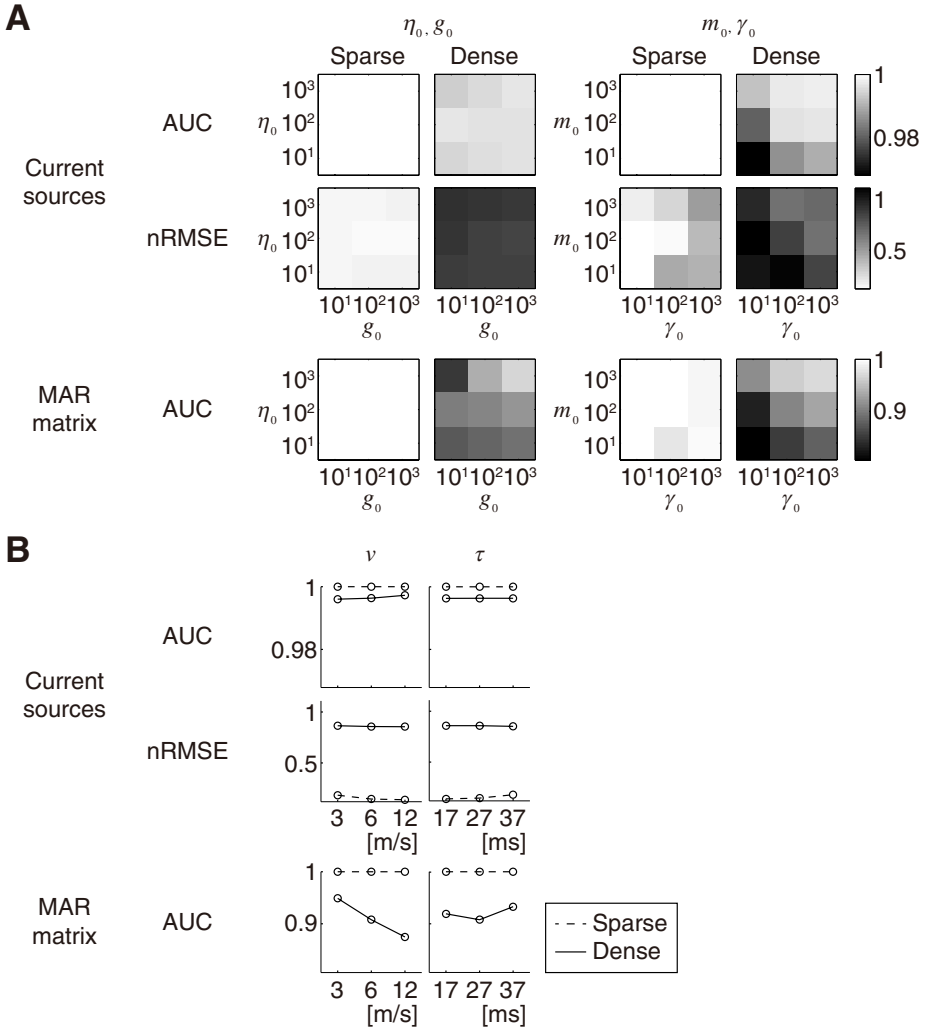

Figure 4.5: Dependence of the hyperparameters and the delay parameters on the estimation accuracy in Simulation 2. (A) Variation in AUC and nRMSE scores as the hyperparameters η_0, g_0, m_0, and γ_0 are varied from 10 to 1000. Scores with the default setting (all hyperparameters set to 100) are shown in the center of each panel. The gray scale is common to each evaluation measure. (B) The AUC and nRMSE scores obtained by varying the conduction velocity ($v = 3, 6, 12\,\mathrm{m/s}$) and local delay constant ($\tau = 17, 27, 37\,\mathrm{ms}$). The ranges of the vertical axes correspond to the gray scales in (A).

hyperparameters m_0 and γ_0 yielded higher estimation accuracy in the dense setting and the reverse was true in the sparse setting. Since the central value of 100 achieved relatively high scores in both settings, our recommended hyperparameter setting in Eqs. (4.9) and (4.10) is $(m_0, \gamma_0) = (100, 100)$.

Fig. 4.5B shows the effect of changing the conduction velocity v and the local delay constant τ on the evaluation measures. Although the AUC of the MAR matrix was noticeably altered in the dense setting, the AUC and nRMSE scores were relatively insensitive to perturbations of v and τ. This indicates that misspecification of the delay parameters does not severely degrade the estimation accuracy.

4.5. Application to real data

Finally, the estimated source distributions and interactions were evaluated on experimental data. To this end, the methods were tested on a publicly available multimodal dataset [43]. The task data were acquired while pictures of human faces were presented to subjects. The physiological plausibility of the estimated current sources and effective connectivity were qualitatively evaluated from previous literature reports.

4.5.1 Settings

In evaluating the estimation performance, we focused on the temporal propagation of the stimulus-evoked responses. For this purpose, the MEG measurements were trial averaged and the t-values for the fMRI prior were computed with a task versus baseline contrast. Here, we used magnetometer measurements recorded in the dataset; the reconstructed sources from these measurements are presented as MEG results in [43]. All hyperparamters for the informative priors were set to 100. As demonstrated in Simulation 2, these hyperparamter settings optimized the reconstruction accuracy in both the sparse and dense settings.

The continuous MEG data in the dataset had been pre-processed by temporal signal-space separation [78] to remove external noise. These data were further pre-processed by baseline correction, high-pass filtering (cutoff 0.25 Hz), low-pass filtering (cutoff 40 Hz), down sampling (from 1.1 kHz to 1 kHz), trial segmentation (0–300 ms after the stimulus onset), trial rejection, and trial averaging (concatenating the entire cohort of face stimulus conditions; see Fig.

4.6A). When segmenting the data, the last 100 ms prior to stimulus onset were also extracted to estimate the noise covariance matrix from the trial-averaged data. The trial rejection step eliminated trials with amplitude of MEG exceeding 3000 fT and that of EOG exceeding 100 μV. 548 trials (averaged across subjects) remained after the trial rejection step.

The fMRI data (3 mm isotropic) were pre-processed by running a batch Matlab script accompanying the dataset. This script executed a standard data pre-processing pipeline of SPM8[6]. The fMRI prior was computed from a t-value map ($p < 0.001$, uncorrected) in which the face stimulus conditions are contrasted against the baseline (the group-level t-value map is illustrated in Fig. 4.6B). Note that the t-value maps used to construct the fMRI prior were not group-averaged, but were separately computed from each subject.

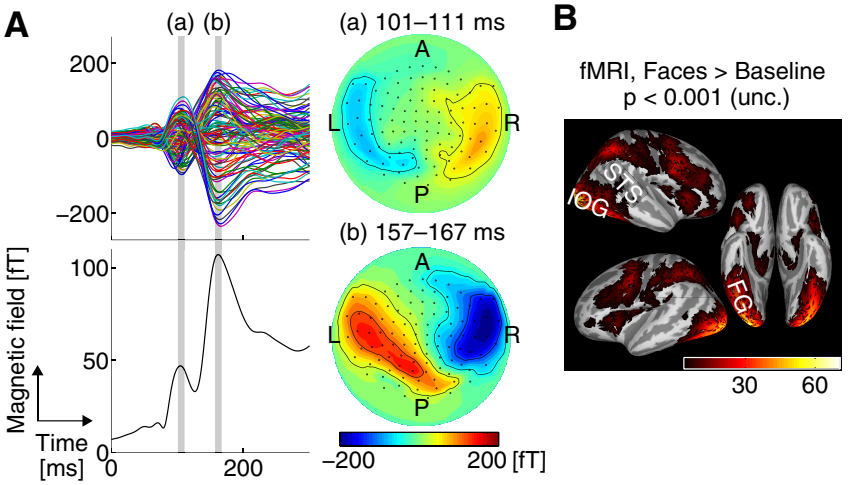

Figure 4.6: Real MEG and fMRI data acquired during a face perception task. (A) Evoked responses to visual face stimuli (averaged across trials and subjects). The top-left plot shows the time courses of sensor measurements, and the bottom-left plot shows the square root of the measured power (averaged over sensors). The right figures display 2D topography maps at the two times of peak activity, indicated by the gray lines (a) and (b) in the left plots. (B) Group-averaged fMRI t-value maps with a contrast of face stimulus conditions against the baseline.

[6] http://www.fil.ion.ucl.ac.uk/spm/

4.5.2 Results

Fig. 4.7 presents the GoF scores, obtained from the three ways of calculation described in Fig. 4.1. For all the ways of computing GoF, larger mean scores were obtained from the dynamic method. Fig. 4.7A shows that the dynamic method significantly improved the normal GoF measure. However, the score differences between the non-dynamic and dynamic methods were small for the other cases of GoF calculation (see Figs. 4.7B and 4.7C). Significance advantages of the dynamic method in generalizability and model prediction accuracy were not found. Despite that, the mean GoF score over 30 in Fig. 4.7C(b) illustrates that the forward model prediction helped recovering the

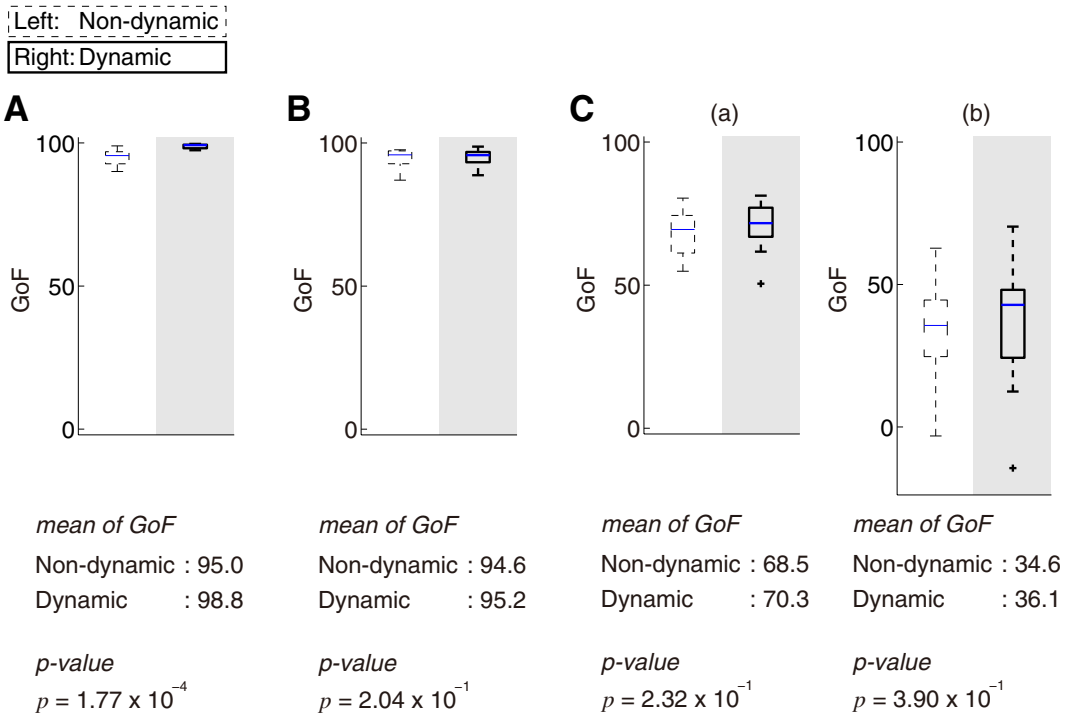

Figure 4.7: Results of GoF analysis on real experimental data. Boxplots of GoF over eleven subjects, mean scores of GoF, and p-value of one sided paired t-test ($\mathrm{GoF}_{\mathrm{Dynamic}} > \mathrm{GoF}_{\mathrm{Non-dynamic}}$) are presented. (A) GoF computed from data averaged across all trials. (B) GoF computed from data averaged across a half of all trials with fixed parameters pre-estimated from the other half. (C) GoF computed as (B) with data containing a missing duration; (a) GoF of the entire duration; (b) GoF of the missing duration. Method comparison with a conventional approach of the non-dynamic method is presented in Fig. B.3.

unobserved data. This suggests that the estimated dynamic model reflected true spatiotemporal dynamics of cortical activity during face processing in some extent.

Fig. 4.8 displays the reconstructed cortical sources and their estimated effective connectivity. The spatially clustered active sources estimated from the dynamic method were more consistent among subjects than the activity estimated from the non-dynamic method. The active sources were mainly estimated in the occipitotemporal cortex containing the face-selective regions IOG and FG [40]. In contrast, no sources were estimated in another face-

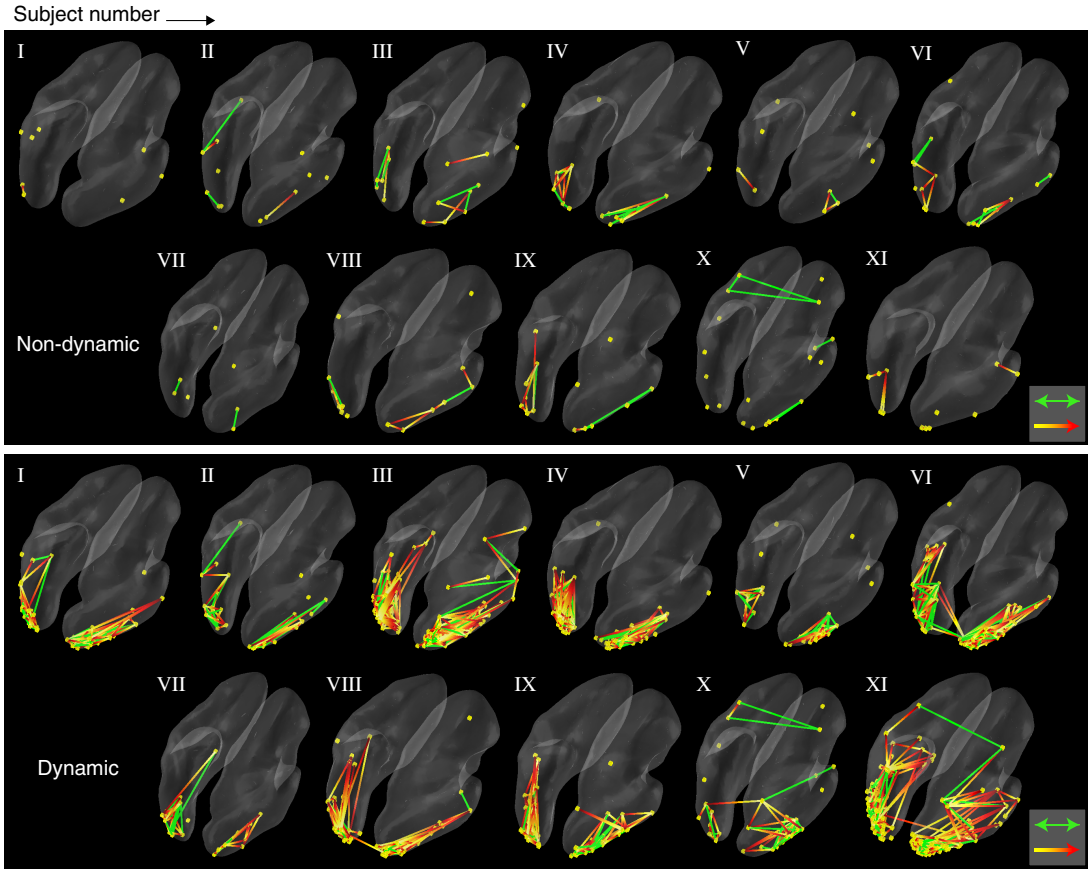

Figure 4.8: Reconstructed current sources (shown by yellow dots) and their effective connectivity during a face perception task undertaken by eleven subjects. Color legend of the effective connectivity is described in the caption of Fig. 4.4. Method comparison with a conventional approach of the non-dynamic method is presented in Fig. B.4.

selective region STS; source activity in this region is not easily reconstructed from MEG data because of its insensitivity to radial source components [44]. The effective connectivity was estimated across the occipitotemporal cortex linking the lower and higher visual cortices. The reconstructed connections structurally and functionally overlapped with the ventral visual pathways [33] responsible for processing visual objects and faces [41].

Fig. 4.9 presents representative examples of the event-related dynamics of the stimulus-evoked responses. Examples differently and similarly estimated by the two methods are shown in Figs. 4.9A and 4.9B, respectively. In the dynamic method, source activity was reconstructed in IOG and FG for both subjects. The effective connectivity between IOG–FG was dominant for the direction from IOG to FG, as inferred from larger absolute MAR coefficients in this direction. The MAR coefficients, estimated by the dynamic method on these extracted connections from IOG to FG, were closely related to the

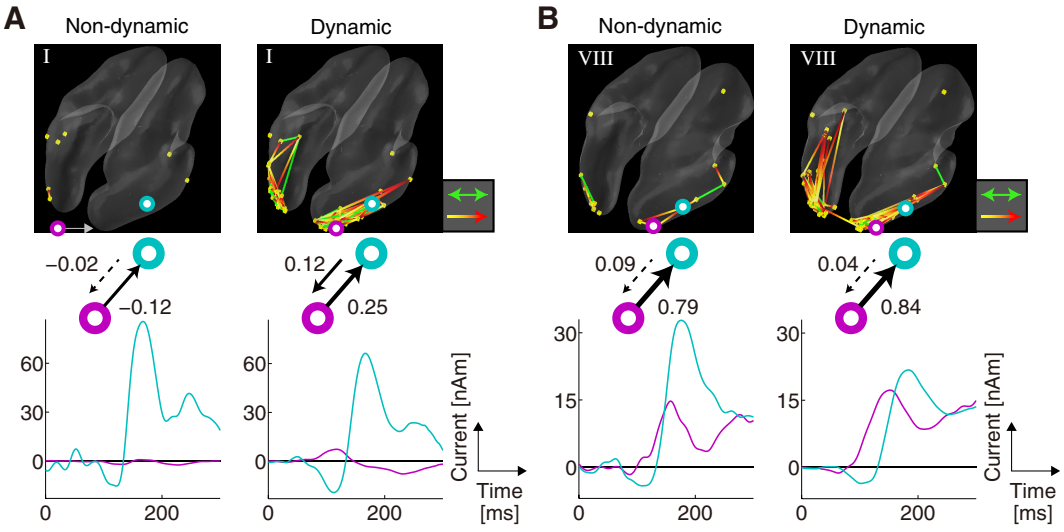

Figure 4.9: Representative examples of the event-related dynamics of stimulus-evoked responses (IOG-FG in Subject I (A) and Subject VIII (B)). Cortical positions of the plotted current sources are indicated by small circles on the cortical surfaces. The numbers beside the arrows connecting the two circles denote the corresponding estimated MAR coefficients. The time courses of each source (magenta and cyan) are plotted in the respective colors of each circle. Method comparison with a conventional approach of the non-dynamic method is presented in Fig. B.5.

temporally lagged correlation between the source time courses of IOG and FG.

4.6. Discussion

In this chapter, we present a new state-space method of MEG source reconstruction that simultaneously estimates the source amplitudes and interactions across the whole brain. Directed interactions between sources are represented by the full MAR model in the source space. The unknown network structure of this MAR model is informed by prior knowledge of the existence and lengths of structural connections, inferred from dMRI data. To improve the reliability and robustness of the estimates, we incorporate the following priors into the model; a spatial prior derived from fMRI activity patterns, and an ARD sparse prior on the MAR coefficients. Source reconstruction using the MAR model with the above prior assumptions is formulated by a state-space representation within a Bayesian framework. The current source activity and model parameters are jointly estimated by a variational Bayesian algorithm. The estimation performance of the proposed dynamic method was quantitatively and qualitatively evaluated on simulation and experimental data, respectively, comparing with its non-dynamic counterpart.

As demonstrated in Simulation 1, the source amplitudes and the MAR coefficients were reasonably estimated by the proposed method, even when the number of active sources was as high as 20 or 30. For over 75% of these originally active sources, the reconstruction gains exceeded 0.80 and the RMSEs of the MAR coefficients were below 0.15. However, the estimates yielded a non-negligible number of missing sources and connections with higher number of active sources. Close inspection of the estimates revealed that these misspecifications occurred more frequently when the source locations were less observable from the MEG sensors. In general, a less observable source (i.e., a source with a smaller lead field norm) is difficult to correctly estimate, especially when other active sources with larger lead field norms produce a similar spatial pattern of the sensor data. Since such source placements increase with increasing number of active sources in the Monte Carlo simulations, this

phenomenon would explain the decay in the source reconstruction accuracy, reported also in [64] using standard methods, with higher number of active sources. This phenomenon might also explain the declining performance of the connectivity estimation. Overall, however, our results suggest that the source amplitudes and the model parameters were adequately estimated by our proposed method.

In Simulation 2, we examined the estimation accuracy of the methods using the synthesized evoked responses. This simulation demonstrated that: 1) our method can correctly estimate the active sources and their effective connectivity in spatially sparse source configurations, and 2) although all of the active sources are not reconstructed in spatially dense source settings, our method identifies at least one single source interaction for every pair of ROIs with true effective connectivity. Moreover, while the nRMSE score of the sources was rather high in the dense setting (0.86), a reasonable score (0.39) was obtained by computing the nRMSE from the region-wise source time courses. These results indicate that, even when the sources are densely distributed over the cortex, ROI-wise activity and effective connectivity can be appropriately estimated. We emphasize that these estimation performances were obtained under realistic simulation settings, which introduce significant discrepancies between our model formulation and the data generation process. In particular, the assumed dynamic model was based on linear autoregressive modeling, whereas the data were generated from a network of nonlinear neural mass models. The limitation of our source dynamics modeling manifests as false positive effective connectivity in the dense setting. This indicates a deficiency in the assumed model, since it also appeared when the MAR model was fitted to the true source time courses. It should be mentioned that the above discussion of Simulation 2 was based on estimates informed by the fMRI prior. Our analysis with and without the fMRI prior confirmed that the prior certainly improves the estimation accuracy in both sparse and dense source configurations.

Applying our method to real data collected during a face perception task, we obtained physiologically plausible estimates that were appropriately consistent among subjects. The estimated source positions were mainly located

along the occipitotemporal cortex, which contains the ventral part of the face-selective regions IOG and FG [40]. The estimated effective connectivity largely overlapped with the ventral visual pathways mediating transmission between the lower and higher visual areas during object recognition [33]. The averaged GoF score, larger than 30, computed from the missing duration data indicates that the dynamic model captured spatiotemporal dynamics of cortical activity during face perception in some extent. Since the face-selective regions have been extensively studied in the literature [40, 41, 50], a face perception task is useful for properly evaluating the source distributions estimated from the methods. In contrast, how these face-selective regions dynamically interact is not known in detail, although several studies have investigated structural connectivity [36, 69], functional connectivity [17], and effective connectivity [21] among these areas. Therefore, although our estimated effective connectivity roughly corresponds to the ventral visual pathways, rigorous evaluation of the existence and directions of the source interactions is beyond the scope of this thesis. For improved future evaluation, the consistency of the estimates in the present study could be examined using other neuroimaging modalities with higher spatial resolution than MEG, such as the electrocorticogram (ECoG).

Major differences between the dynamic and non-dynamic methods exist in whether sources are reconstructed with or without temporal constraints, and the source amplitudes and interactions are estimated simultaneously or not. These differences resulted in better scores of nRMSE and normal GoF in the dynamic method. It should be noted that the higher normal GoF score, presented in Fig. 4.7, does not necessarily mean higher performance of the dynamic method in source reconstruction, since the number of parameters in the dynamic method is quite larger than that of the non-dynamic method. The dynamic method did not significantly improve the AUC scores in simulations and the GoF measures in examining generalizability and model prediction accuracy. Indeed, improvements on these scores became obvious when the MAR matrix of the non-dynamic method was computed only from pairs of active sources (see Appendix B.4). This indicates that methodological advancements of our method allowing estimation of effective connectivity over

the whole brain also have a crucial role in obtaining good results, which overcomes limitations of conventional two-stage approaches.

Prior knowledge on the existence of structural connections in the dynamic method may reciprocally compensate for the estimated magnitudes of the source amplitudes and interactions. This is because the estimation steps of the current sources and their associated MAR coefficients are mutually dependent under the variational Bayesian algorithm. In contrast, the non-dynamic method does not update the reconstructed current sources after estimating the MAR coefficients. Therefore, the estimated connectivity is more sensitive to source reconstruction errors, because if two reconstructed sources are once incorrectly assigned to pairs of structurally non-connected locations, the non-dynamic method no longer estimates their effective connectivity in successive iterations. The modeling and algorithmic differences between the two methods are highlighted by the higher physiological plausibility and inter-subject consistency in the dynamic method, when both methods were applied to the face perception dataset.

Using the proposed method, we can estimate the effective connectivity without requiring the selection of a limited number of ROIs as network nodes. This was achieved by extending the source-dimensional autoregressive models adopted in previous dynamic methods of MEG/EEG source reconstruction [30, 51, 87], in contrast to the approach of [61] where they assumed a full MAR model in a low-dimensional latent space. The non-diagonal MAR coefficients newly introduced in the proposed method are derived from the existence and lengths of structural connections. These coefficients allow long-range source interactions to be estimated within whole-brain structural networks. Our method adopts an exploratory approach to estimating the effective connectivity. This approach complements the confirmatory approach of DCM [16, 26], in which the effective connectivity between brain regions is inferred using a procedure of Bayesian model comparison.

When incorporating the structural connectivity information of dMRI into the MAR model, we require a threshold value of fiber counts to binarize the structural connectivity. We must also determine the conduction velocity and the local delay constant that specify the time lags between two sources. We

selected a small threshold value of fiber counts (10^{-4}) to minimize false negative effective connectivity. Our model assumes that effective connectivity exists only between structurally connected pairs of sources. Therefore, a higher threshold would increase false negative effective connectivity by increasing the number of false negative structural connections. By choosing a small threshold, we reduced the possibility of false negative structural connections, while increasing the number of MAR coefficients to be estimated. Although more MAR coefficients should increase the false positive effective connectivity, this effect is partially negated by the sparse prior applied to the MAR coefficients. The conduction velocity and local delay constant were set to 6 m/s and 27 ms, respectively. In Simulation 2, we investigated the impact of these parameters on the estimation accuracy, and found that the estimation performance was relatively insensitive to changes in both parameters. This indicates that appropriate estimates could be obtained if the time lags in the MAR model were only rough approximations.

The hyperparameters m_0 and γ_0, controlling the magnitude and the reliability of the fMRI prior, respectively, must be carefully determined. Such care is required because non-negligible differences might exist between the spatial patterns of the electrocortical activity and those observed in fMRI. Increasing the dependence on the fMRI prior may not improve the results. Indeed, imposing a large weighting on the spatially dense fMRI prior (by setting γ_0 high) reduced the estimation accuracy in the sparse source setting of Simulation 2, where a moderate value of γ_0 provided higher estimation accuracy. Therefore, in applying the methods to experimental data, these hyperparameters were selected to yield reasonable evaluation scores in both the sparse and the dense simulation settings.

Representing the source dynamics by linear autoregressive modeling would introduce errors to the estimated (actually non-linear) source interactions. When the interactions were governed by realistic non-linear network dynamics, the limitation of our dynamic model emerged as false positive estimates of the effective connectivity. This observation confirms the necessity of a non-linear extension of the linear MAR source model to embrace a wide range of dynamic phenomena. In future work, we will develop a means of incorpo-

rating non-linear neurobiological dynamics into the source model, while preserving the tractability of estimation in the high-dimensional source space.

Chapter 5

Contributions and Future Directions

In this thesis, previous state-space methods for MEG/EEG source reconstruction are extended to reconstruct focal current sources (Chapter 3) and to identify long-range directed source interactions over the whole brain (Chapter 4). By parameterizing the MAR coefficients of the dynamic source model in spatially fine resolution, method extensions in both chapters were achieved.

5.1. Contributions

Reconstruction of focal sources was achieved by extending previous dynamic models of state-space methods. The temporal evolution of source activity, used to be common across or within brain regions, has been properly parameterized in a spatially inhomogeneous manner in Chapter 3. Through this extension, spatially focal sources are obtained for the first time under the state-space framework[1]. Acquiring the ability of reconstructing focal sources is an important improvement for state-space methods, since the focal sources are consistent with the property that cortical activities specialized to specific brain functions are restricted to small brain regions [22].

[1]Reconstructing focal sources was succeeded also in a state-space method [51] at around the same time when the work in Chapter 3 [27] was published.

Incorporating long-range directed source interactions into the dynamic model of current source allows estimating whole-brain functional networks. The new state-space method in Chapter 4 first provides a mean of elucidating mechanisms in dynamic integration of functional segregated current sources, by discovering functional brain networks in the source space. Identifying long-range directed interactions between whole-brain sources is a fundamental goal of method developments in neuroimaging fields. Successes in this line of researches allow conducting an exploratory approach of effective connectivity estimation. This approach complements a confirmatory approach of DCM; the most widely used method for elucidating brain effective connectivity [16,26]. In DCM, only a limited number of ROIs can be selected as network nodes for discovering functional brain networks. Inclusion of whole brain areas as network nodes in DCM is intractable due to combinatorial explosion of the ROI selections [54]. The state-space method extended in Chapter 4 allows conducting the effective connectivity analysis over the whole brain, which cannot be dealt with DCM.

5.2. Future directions

For both chapters, the dynamic model of neuronal current sources is extended into new source reconstruction methods. By developing new modeling and parameterization of the source dynamics, each chapter-specific purpose of method extension was realized. The principle idea on the model extensions is to model more realistic source dynamics. Assuming the spatially inhomogeneous temporal evolution of sources in Chapter 3 and the long-range directed source interactions on whole-brain structural networks in Chapter 4 are the examples of introducing realistic source dynamics.

Other representative examples, not incorporated into the dynamic model in this thesis, are non-linearity and non-stationarity of the source dynamics. These components are important building blocks to represent real neurobiological dynamic behaviors. Although estimating non-linear and non-stationary dynamics in a high-dimensional source space is a challenging direction of future research, inclusion of these properties into the dynamic

source model would allow estimating the source amplitudes and interactions more closely reflecting real characteristics of neuronal current source dynamics. Development on state-space methods combining with the non-linearity and non-stationarity is a possible direction of future research.

Appendix A

Appendix of Chapter 3

A.1. Algorithm details

We present here the detailed derivation of the update rules for the currents (**J**-step) and for the parameters (β-step, **a**-step, and **q**-step).

In the **J**-step, the approximate posterior distributions of the currents $Q(\mathbf{J}_t)$ for $t = 1:T$ are updated. The distribution $Q(\mathbf{J}_t)$ is obtained by substituting $\mathbf{J}_t$ into Eq. (2.14):

$$\log Q(\mathbf{J}_t) = \langle \log P(\mathbf{B}_{1:T}, X) \rangle_{Q(X \setminus \mathbf{J}_t)} + \text{const.} \tag{A.1}$$

By calculating the integral in the right-hand side of Eq. (A.1), we obtain the linear and quadratic terms with respect to $\mathbf{J}_t$. Therefore, $Q(\mathbf{J}_t)$ is a Gaussian distribution:

$$Q(\mathbf{J}_t) = \mathcal{N}(\mathbf{J}_t | \bar{\mathbf{J}}_t, \bar{\beta}^{-1} \bar{\mathbf{V}}_t) \tag{A.2}$$

where $\bar{\beta}$ represents the mean of $Q(\beta)$. The (scaled) covariance matrix $\bar{\mathbf{V}}_t$ and the mean $\bar{\mathbf{J}}_t$ are derived as

$$\begin{cases} \bar{\mathbf{V}}_t = \left(\mathbf{G}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{G} + \hat{\mathbf{V}}_t^{-1} \right)^{-1} \\ \bar{\mathbf{J}}_t = \bar{\mathbf{V}}_t \left(\mathbf{G}^{\mathrm{T}} \mathbf{S}^{-1} \mathbf{B}_t + \hat{\mathbf{V}}_t^{-1} \hat{\mathbf{J}}_t \right) \end{cases} \tag{A.3}$$

where we have introduced auxiliary parameters $\hat{\mathbf{V}}_t$ and $\hat{\mathbf{J}}_t$. These are defined

as

$$\hat{\mathbf{V}}_t = \begin{cases} \left(\left(\mathbf{I} + \langle \mathrm{diag}(\mathbf{a})\mathrm{diag}(\mathbf{a})\rangle\right)\mathrm{diag}(\bar{\mathbf{q}})\right)^{-1}, & \text{for } t \neq T \\ \left(\mathrm{diag}(\bar{\mathbf{q}})\right)^{-1}, & \text{for } t = T \end{cases}$$
$$\hat{\mathbf{J}}_t = \begin{cases} \left(\mathbf{I} + \langle \mathrm{diag}(\mathbf{a})\mathrm{diag}(\mathbf{a})\rangle\right)^{-1} \\ \qquad \times \mathrm{diag}(\bar{\mathbf{a}})(\bar{\mathbf{J}}_{t-1} + \bar{\mathbf{J}}_{t+1}), & \text{for } t \neq T \\ \mathrm{diag}(\bar{\mathbf{a}})\bar{\mathbf{J}}_{t-1}, & \text{for } t = T \end{cases} \tag{A.4}$$

with $\bar{\mathbf{a}}$, $\bar{\mathbf{q}}$, $\bar{\mathbf{J}}_{t-1}$, and $\bar{\mathbf{J}}_{t+1}$ being the means of their approximate posteriors, which are updated at the previous iteration, and $\mathbf{I}$ being the identity matrix. Eq. (A.3) can be rewritten in a computationally efficient form using the Woodbury's matrix inversion lemma [6]:

$$\begin{cases} \bar{\mathbf{V}}_t = \left(\mathbf{I} - \mathbf{K}\mathbf{G}\right)\hat{\mathbf{V}}_t \\ \bar{\mathbf{J}}_t = \hat{\mathbf{J}}_t + \mathbf{K}\left(\mathbf{B}_t - \mathbf{G}\hat{\mathbf{J}}_t\right) \end{cases} \tag{A.5}$$

where $\mathbf{K} = \hat{\mathbf{V}}_t\mathbf{G}^{\mathrm{T}}(\mathbf{G}\hat{\mathbf{V}}_t\mathbf{G}^{\mathrm{T}} + \mathbf{S})^{-1}$ plays a role similar to the Kalman gain. By adopting this form, the matrix inversions of large size (in our application 2000×2000) in Eq. (A.3) are avoided. It should be noted that since $\hat{\mathbf{V}}_t$ is diagonal owing to the independence approximation on $Q(\mathbf{J}_{1:T})$, the computational complexity required for obtaining $\bar{\mathbf{V}}_t$ and $\bar{\mathbf{J}}_t$ increases linearly with number of current sources.

In the β-step, the approximate posterior distribution of the scaling parameter $Q(\beta)$ is updated. The distribution $Q(\beta)$ is obtained by substituting β into Eq. (2.14):

$$\log Q(\beta) = \langle \log P(\mathbf{B}_{1:T}, X)\rangle_{Q(X\setminus\beta)} + \text{const.} \tag{A.6}$$

By calculating the integral in the right-hand side of Eq. (A.6), we obtain the linear and logarithmic terms with respect to β. Therefore, $Q(\beta)$ is a gamma distribution:

$$Q(\beta) = \mathcal{G}(\beta|\,\bar{\beta}, \gamma_\beta) \tag{A.7}$$

where the shape parameter γ_β and the mean $\bar{\beta}$ are derived as

$$\begin{cases} \gamma_\beta = \frac{1}{2}MT \\ \bar{\beta} = \gamma_\beta \Big\{ \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\Big(\mathbf{S}^{-1}(\mathbf{B}_t\mathbf{B}_t^{\mathrm{T}} - \mathbf{G}\bar{\mathbf{J}}_t\mathbf{B}_t^{\mathrm{T}} - \mathbf{B}_t\bar{\mathbf{J}}_t^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} + \mathbf{G}\bar{\mathbf{J}}_t\bar{\mathbf{J}}_t^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}) \Big) \\ \qquad + \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\Big(\mathrm{diag}(\bar{\mathbf{q}})(\bar{\mathbf{J}}_t\bar{\mathbf{J}}_t^{\mathrm{T}} - 2\,\mathrm{diag}(\bar{\mathbf{a}})\bar{\mathbf{J}}_{t-1}\bar{\mathbf{J}}_t^{\mathrm{T}} \\ \qquad\qquad + \langle \mathrm{diag}(\mathbf{a})\mathrm{diag}(\mathbf{a}) \rangle \, \bar{\mathbf{J}}_{t-1}\bar{\mathbf{J}}_{t-1}^{\mathrm{T}}) \Big) \Big\}^{-1}. \end{cases} \tag{A.8}$$

In the **a**-step, the approximate posterior distribution of the AR(1) parameter $Q(\mathbf{a})$ is updated. The distribution $Q(\mathbf{a})$ is obtained by substituting **a** into Eq. (2.14):

$$\log Q(\mathbf{a}) = \langle \log P(\mathbf{B}_{1:T}, X) \rangle_{Q(X\setminus\mathbf{a})} + \mathrm{const.} \tag{A.9}$$

By calculating the integral in the right-hand side of Eq. (A.9), we obtain the sum of the linear and quadratic terms with respect to a_n. Therefore, $Q(\mathbf{a})$ is a product of Gaussian distributions:

$$Q(\mathbf{a}) = \prod_{n=1}^{N} \mathcal{N}(a_n | \, \bar{a}_n, \eta_n^{-1}) \tag{A.10}$$

where the variance η_n^{-1} and the mean $\bar{a}_n$ are derived as

$$\begin{cases} \eta_n^{-1} = \Big\{ \eta_0 + \sum_{t=1}^{T} \bar{\beta} \Big(\mathrm{diag}(\bar{\mathbf{q}}) \Big\langle \mathbf{J}_{t-1}\mathbf{J}_{t-1}^{\mathrm{T}} \Big\rangle \Big)_{n,n} \Big\}^{-1} \\ \bar{a}_n = \eta_n^{-1} \sum_{t=1}^{T} \bar{\beta} \big(\mathrm{diag}(\bar{\mathbf{q}})\bar{\mathbf{J}}_{t-1}\bar{\mathbf{J}}_t^{\mathrm{T}} \big)_{n,n}. \end{cases} \tag{A.11}$$

In the **q**-step, the approximate posterior distribution of the vector parameterizing the system noise $Q(\mathbf{q})$ is updated. The distribution $Q(\mathbf{q})$ is obtained by substituting **q** into Eq. (2.14):

$$\log Q(\mathbf{q}) = \langle \log P(\mathbf{B}_{1:T}, X) \rangle_{Q(X\setminus\mathbf{q})} + \mathrm{const.} \tag{A.12}$$

By calculating the integral in the right-hand side of Eq. (A.12), we obtain the sum of the linear and logarithmic terms with respect to q_n. Therefore, $Q(\mathbf{q})$ is a product of gamma distributions:

$$Q(\mathbf{q}) = \prod_{n=1}^{N} \mathcal{G}(q_n | \, \bar{q}_n, \gamma_n) \tag{A.13}$$

where the shape parameter γ_n and the mean $\bar{q}_n$ are derived as

$$\begin{cases} \gamma_n = \gamma_0 + \frac{1}{2}\,T \\ \bar{q}_n = \gamma_n \Big\{ \gamma_0\, q_0^{-1} + \frac{1}{2} \sum_{t=1}^{T} \bar{\beta} \Big(\Big\langle \mathbf{J}_t \mathbf{J}_t^{\mathrm{T}} \Big\rangle - 2\mathrm{diag}(\bar{\mathbf{a}}) \bar{\mathbf{J}}_{t-1} \bar{\mathbf{J}}_t^{\mathrm{T}} \\ \qquad\qquad + \langle \mathrm{diag}(\mathbf{a}) \mathrm{diag}(\mathbf{a}) \rangle \Big\langle \mathbf{J}_{t-1} \mathbf{J}_{t-1}^{\mathrm{T}} \Big\rangle \Big)_{n,n} \Big\}^{-1}. \end{cases} \tag{A.14}$$

Remaining expectations in the update equations are evaluated as follows ($\boldsymbol{\eta}$ is a vector whose entries are η_n in Eq. (A.11)):

$$\langle \mathrm{diag}(\mathbf{a}) \mathrm{diag}(\mathbf{a}) \rangle = \mathrm{diag}(\bar{\mathbf{a}}) \mathrm{diag}(\bar{\mathbf{a}}) + (\mathrm{diag}(\boldsymbol{\eta}))^{-1} \tag{A.15}$$

$$\sum_{t=1}^{T} \Big\langle \mathbf{J}_{t-1} \mathbf{J}_{t-1}^{\mathrm{T}} \Big\rangle = \bar{\mathbf{J}}_0 \bar{\mathbf{J}}_0^{\mathrm{T}} + \sum_{t=2}^{T} \Big(\bar{\mathbf{J}}_{t-1} \bar{\mathbf{J}}_{t-1}^{\mathrm{T}} + \bar{\beta}^{-1} \bar{\mathbf{V}}_t \Big) \tag{A.16}$$

$$\sum_{t=1}^{T} \Big\langle \mathbf{J}_t \mathbf{J}_t^{\mathrm{T}} \Big\rangle = \sum_{t=1}^{T} \Big(\bar{\mathbf{J}}_t \bar{\mathbf{J}}_t^{\mathrm{T}} + \bar{\beta}^{-1} \bar{\mathbf{V}}_t \Big). \tag{A.17}$$

Free energy

The free energy F (Eq. (2.13)) in Chapter 3 is computed by being decomposed into L, H_β, H_a, and H_q:

$$\begin{aligned} F &= \left\langle \log \frac{P(\mathbf{B}_{1:T}, \mathbf{J}_{1:T}, \beta, \mathbf{a}, \mathbf{q})}{Q(\mathbf{J}_{1:T}, \beta, \mathbf{a}, \mathbf{q})} \right\rangle \\ &= \left\langle \sum_{t=1}^{T} \left(\log \frac{P(\mathbf{B}_t \,|\, \mathbf{J}_t, \beta) P(\mathbf{J}_t \,|\, \mathbf{J}_{t-1}, \beta, \mathbf{a}, \mathbf{q})}{Q(\mathbf{J}_t)} \right) \right\rangle \\ &\qquad + \left\langle \log \frac{P(\beta)}{Q(\beta)} \right\rangle + \left\langle \log \frac{P(\mathbf{a})}{Q(\mathbf{a})} \right\rangle + \left\langle \log \frac{P(\mathbf{q})}{Q(\mathbf{q})} \right\rangle \\ &= L + H_\beta + H_a + H_q. \end{aligned} \tag{A.18}$$

Each term composing the free energy is expanded as follows:

$$
\begin{aligned}
L &= \sum_{t=1}^{T} \Bigg\{ \Big\langle \log \mathcal{N}\Big(\mathbf{B}_t \,|\, \mathbf{G}\mathbf{J}_t, \beta^{-1}\mathbf{S}\Big) \Big\rangle \\
&\quad + \Big\langle \log \mathcal{N}\Big(\mathbf{J}_t | \operatorname{diag}(\mathbf{a})\mathbf{J}_{t-1}, (\beta \operatorname{diag}(\mathbf{q}))^{-1}\Big) \Big\rangle \\
&\quad - \Big\langle \log \mathcal{N}(\mathbf{J}_t \,|\, \bar{\mathbf{J}}_t, \bar{\beta}^{-1}\bar{\mathbf{V}}_t) \Big\rangle \Bigg\} \\
&= \sum_{t=1}^{T} \Bigg\{ -\frac{M}{2}\log 2\pi + \frac{M}{2}\langle \log \beta \rangle - \frac{1}{2}\log |\mathbf{S}| \\
&\qquad - \frac{\bar{\beta}}{2} \operatorname{tr}\Big(\mathbf{S}^{-1} \big(\mathbf{B}_t\mathbf{B}_t^{\mathrm{T}} - \mathbf{G}\bar{\mathbf{J}}_t\mathbf{B}_t^{\mathrm{T}} - \mathbf{B}_t\bar{\mathbf{J}}_t^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} + \mathbf{G}\Big\langle \mathbf{J}_t\mathbf{J}_t^{\mathrm{T}} \Big\rangle \mathbf{G}^{\mathrm{T}} \big) \Big) \\
&\quad - \frac{N}{2}\log 2\pi + \frac{N}{2}\langle \log \beta \rangle + \frac{1}{2}\sum_{n=1}^{N} \langle \log q_n \rangle \\
&\qquad - \frac{\bar{\beta}}{2} \operatorname{tr}\Bigg(\operatorname{diag}(\bar{\mathbf{q}}) \Bigg(\Big\langle \mathbf{J}_t\mathbf{J}_t^{\mathrm{T}} \Big\rangle - 2\operatorname{diag}(\bar{\mathbf{a}})\bar{\mathbf{J}}_{t-1}\bar{\mathbf{J}}_t^{\mathrm{T}} \\
&\qquad\quad + \langle \operatorname{diag}(\mathbf{a})\operatorname{diag}(\mathbf{a}) \rangle \Big\langle \mathbf{J}_{t-1}\mathbf{J}_{t-1}^{\mathrm{T}} \Big\rangle \Bigg) \Bigg) \\
&\quad + \frac{N}{2}(1 + \log 2\pi) - \frac{N}{2}\log \bar{\beta} + \frac{1}{2}\log |\bar{\mathbf{V}}_t| \Bigg\},
\end{aligned} \tag{A.19}
$$

$$
\begin{aligned}
H_\beta &= \langle \log \beta^{-1} \rangle - \langle \log \mathcal{G}(\beta \,|\, \bar{\beta}, \gamma_\beta) \rangle \\
&= \log \Gamma(\gamma_\beta) - \gamma_\beta\, \psi(\gamma_\beta) + \gamma_\beta,
\end{aligned} \tag{A.20}
$$

$$
\begin{aligned}
H_a &= \sum_{n=1}^{N} \Bigg\{ \Big\langle \log \mathcal{N}(a_n | \, 0, \eta_0^{-1}) \Big\rangle - \Big\langle \log \mathcal{N}(a_n | \, \bar{a}_n, \eta_n^{-1}) \Big\rangle \Bigg\} \\
&= -\frac{N}{2}\log 2\pi + \frac{N}{2}\log \eta_0 - \frac{\eta_0}{2} \operatorname{tr}\Big(\langle \operatorname{diag}(\mathbf{a})\operatorname{diag}(\mathbf{a}) \rangle \Big) \\
&\quad + \frac{N}{2}(1 + \log 2\pi) - \frac{1}{2}\sum_{n=1}^{N} \Big(\log \eta_n \Big),
\end{aligned} \tag{A.21}
$$

$$
\begin{aligned}
H_q &= \sum_{n=1}^{N} \Bigg\{ \Big\langle \log \mathcal{G}(q_n \,|\, q_0, \gamma_0) \Big\rangle - \Big\langle \log \mathcal{G}(q_n \,|\, \bar{q}_n, \gamma_n) \Big\rangle \Bigg\} \\
&= \sum_{n=1}^{N} \Bigg\{ \gamma_0 \Big(\log(\bar{q}_n q_0^{-1}) - \bar{q}_n q_0^{-1} + 1 \Big) + H(\gamma_n, \gamma_0) \Bigg\},
\end{aligned} \tag{A.22}
$$

where the term H in H_q is defined as

$$H(x,y) = \Big(\log\Gamma(x) - x\psi(x) + x\Big) - \Big(\log\Gamma(y) - y\psi(y) + y\Big) + y\Big(\psi(x) - \log x\Big), \tag{A.23}$$

and expectation of the logarithmic terms in Eq. (A.19) are evaluated as

$$\langle \log\beta \rangle = \log\bar{\beta} + \psi(\gamma_\beta) - \log\gamma_\beta, \tag{A.24}$$

$$\langle \log q_n \rangle = \log\bar{q}_n + \psi(\gamma_n) - \log\gamma_n. \tag{A.25}$$

Appendix B

Appendix of Chapter 4

B.1. Algorithm details

In this appendix, the update rules for the current sources and the model parameters are derived in detail. The update equations are presented in order of the updating steps in the estimation algorithm (**A**-step, $\boldsymbol{\eta}$-step, **q**-step, **J**-step, and β-step). For notational simplicity, $\mathbf{J}_t$ and $\bar{\mathbf{J}}_t$ for $t \leq 0$ are included in the update equations, although they are set to zero in the model, and the subscripts in expectation terms are omitted.

The **A**-step updates the approximate posterior distribution of the MAR matrix $Q(\mathbf{A})$. Substituting **A** for x in Eq. (2.14) and calculating its expectation term, we obtain a sum of linear and quadratic terms of $\mathbf{a}_n$, where $\mathbf{a}_n^{\mathrm{T}}$ denotes the non-zero entries of the n-th row of **A**. Therefore, $Q(\mathbf{A})$ is a product of Gaussian distributions:

$$Q(\mathbf{A}) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{a}_n \mid \bar{\mathbf{a}}_n, \bar{\mathbf{U}}_n). \tag{B.1}$$

The covariance $\bar{\mathbf{U}}_n$ and the mean $\bar{\mathbf{a}}_n$ are derived as

$$\begin{cases} \bar{\mathbf{U}}_n = \Big\{ \mathrm{diag}(\bar{\boldsymbol{\eta}}_n) \\ \qquad\qquad + \bar{\beta}\, \bar{q}_n \sum_{t=1}^{T} \Big(\sum_{l=1}^{L} \mathbf{E}_{n,\Delta_l} .* \left\langle \mathbf{J}_{t-\Delta_l} \mathbf{J}_{t-\Delta_l}^{\mathrm{T}} \right\rangle \Big)_{C_n, C_n} \Big\}^{-1} \\ \bar{\mathbf{a}}_n = \bar{\mathbf{U}}_n\, \bar{\beta}\, \bar{q}_n \sum_{t=1}^{T} \left(\left(\sum_{l=1}^{L} \mathbf{e}_{n,\Delta_l} .* \bar{\mathbf{J}}_{t-\Delta_l} \right)_{C_n} \bar{J}_{n,t} \right), \end{cases} \tag{B.2}$$

where $\bar{\boldsymbol{\eta}}_n$, $\bar{\beta}$, $\bar{q}_n$, $\bar{\mathbf{J}}_{t-\Delta_l}$, and $\bar{J}_{n,t}$ are the means of the approximate posterior distributions, which are updated in the previous iteration, and C_n is a set of indices corresponding to the cortical locations structurally connected to the n-th source position. A dot product $.*$ denotes an element-wise multiplication of a pair of vectors or matrices. The N-dimensional column vector $\mathbf{e}_{n,\Delta_l}$ is an indicator of binary values. If the locations corresponding to the index entries of $\mathbf{e}_{n,\Delta_l}$ are structurally connected to the n-th source position with time lag Δ_l, the relevant indices in $\mathbf{e}_{n,\Delta_l}$ are set to one. Another indicator, the N-dimensional matrix $\mathbf{E}_{n,\Delta_l}$ is computed as $\mathbf{e}_{n,\Delta_l}\,\mathbf{e}_{n,\Delta_l}^{\mathrm{T}}$. Here, the terms subscripted C_n, C_n and C_n in Eq. (B.2) can be considered as the covariance and mean, respectively, of a vector of current sources that directly affects $J_{n,t}$.

The $\boldsymbol{\eta}$-step updates the approximate posterior distribution of the inverse variances of the MAR coefficients $Q(\boldsymbol{\eta}_{1:N})$. Substituting $\boldsymbol{\eta}_{1:N}$ for x in Eq. (2.14) and calculating its expectation term, we obtain a sum of linear and logarithmic terms of $\eta_{n,k}$. Therefore, $Q(\boldsymbol{\eta}_{1:N})$ is a product of gamma distributions:

$$Q(\boldsymbol{\eta}_{1:N}) = \prod_{n=1}^{N}\prod_{k=1}^{K_n} \mathcal{G}(\eta_{n,k} \mid \bar{\eta}_{n,k}, g_{n,k}) \tag{B.3}$$

whose shape parameter $g_{n,k}$ and mean $\bar{\eta}_{n,k}$ are given by

$$\begin{cases} g_{n,k} = g_0 + \frac{1}{2} \\ \bar{\eta}_{n,k} = g_{n,k}\left\{ g_0\,\eta_0^{-1} + \frac{1}{2}\left(\left\langle \mathbf{a}_n\,\mathbf{a}_n^{\mathrm{T}} \right\rangle\right)_{k,k} \right\}^{-1}. \end{cases} \tag{B.4}$$

The $\mathbf{q}$-step updates the approximate posterior distribution of the inverse variance of the scaled system noise $Q(\mathbf{q})$. Substituting $\mathbf{q}$ for x in Eq. (2.14) and calculating its expectation term, we obtain a sum of linear and logarithmic terms of q_n. Therefore, $Q(\mathbf{q})$ is a product of gamma distributions:

$$Q(\mathbf{q}) = \prod_{n=1}^{N} \mathcal{G}(q_n \mid \bar{q}_n, \gamma_n) \tag{B.5}$$

whose shape parameter γ_n and mean $\bar{q}_n$ are given by

$$\begin{cases} \gamma_n = \gamma_0 + \frac{1}{2}\,T \\ \bar{q}_n = \gamma_n \Bigg\{ \gamma_0\,\bar{v}_n + \frac{1}{2}\,\bar{\beta} \sum_{t=1}^{T} \Bigg(\left\langle \mathbf{J}_t\,\mathbf{J}_t^{\mathrm{T}} \right\rangle \\ \qquad - \left(\sum_{l=1}^{L} \left(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l} \right) \right) \bar{\mathbf{J}}_t^{\mathrm{T}} - \bar{\mathbf{J}}_t \left(\sum_{l=1}^{L} \left(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l} \right) \right)^{\mathrm{T}} \\ \qquad + \left\langle \left(\sum_{l=1}^{L} \left(\mathbf{A}_l\,\mathbf{J}_{t-\Delta_l} \right) \right) \left(\sum_{l=1}^{L} \left(\mathbf{A}_l\,\mathbf{J}_{t-\Delta_l} \right) \right)^{\mathrm{T}} \right\rangle \Bigg)_{n,n} \Bigg\}^{-1}. \end{cases} \tag{B.6}$$

$\bar{\mathbf{A}}_l$ and $\bar{\mathbf{J}}_{t-\Delta_l}$ are the means of the approximate posterior distributions, where $\bar{\mathbf{A}}_l$ is obtained from $\bar{\mathbf{a}}_n$ for $n \in \{1\!:\!N\}$ (see Eq. (4.3) for the definition of $\mathbf{A}_l$).

The **J**-step updates the approximate posterior distributions of the current sources $Q(\mathbf{J}_t)$ for $t \in \{1\!:\!T\}$. Substituting $\mathbf{J}_t$ for x in Eq. (2.14) and calculating its expectation term, we obtain linear and quadratic terms of $\mathbf{J}_t$. Therefore, $Q(\mathbf{J}_t)$ is a Gaussian distribution:

$$Q(\mathbf{J}_t) = \mathcal{N}(\mathbf{J}_t \,|\, \bar{\mathbf{J}}_t, \bar{\beta}^{-1}\bar{\mathbf{V}}_t) \tag{B.7}$$

where $\bar{\beta}$ is the mean of $Q(\beta)$. The (scaled) covariance matrix $\bar{\mathbf{V}}_t$ and the mean $\bar{\mathbf{J}}_t$ are derived as

$$\begin{cases} \bar{\mathbf{V}}_t = \left(\mathbf{G}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{G} + \hat{\mathbf{V}}_t^{-1} \right)^{-1} \\ \bar{\mathbf{J}}_t = \bar{\mathbf{V}}_t \left(\mathbf{G}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{B}_t + \hat{\mathbf{V}}_t^{-1}\hat{\mathbf{J}}_t \right) \end{cases} \tag{B.8}$$

where we have introduced auxiliary parameters $\hat{\mathbf{V}}_t$ and $\hat{\mathbf{J}}_t$. These are defined as

$$\hat{\mathbf{V}}_t = \begin{cases} \left(\mathrm{diag}(\bar{\mathbf{q}}) + \sum_{l'=1}^{L'} \left\langle \mathbf{A}_{l'}^{\mathrm{T}}\,\mathrm{diag}(\bar{\mathbf{q}})\,\mathbf{A}_{l'} \right\rangle \right)^{-1}, & \text{for } 1 \le t \le T - \Delta_1 \\ \left(\mathrm{diag}(\bar{\mathbf{q}}) \right)^{-1}, & \text{for } T - \Delta_1 < t \le T \end{cases}$$

$$\hat{\mathbf{J}}_t = \begin{cases} \hat{\mathbf{V}}_t \Bigg(\mathrm{diag}(\bar{\mathbf{q}}) \sum_{l=1}^{L} \left(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l} \right) + \sum_{l'=1}^{L'} \left(\bar{\mathbf{A}}_{l'}^{\mathrm{T}}\,\mathrm{diag}(\bar{\mathbf{q}})\,\bar{\mathbf{J}}_{t+\Delta_{l'}} \right) \\ \qquad - \sum_{l'=1}^{L'} \left\langle \mathbf{A}_{l'}^{\mathrm{T}}\,\mathrm{diag}(\bar{\mathbf{q}}) \left(\sum_{l=1(l\neq l')}^{L} \mathbf{A}_l\,\bar{\mathbf{J}}_{t-\Delta_l+\Delta_{l'}} \right) \right\rangle \Bigg), \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } 1 \le t \le T - \Delta_1 \\ \sum_{l=1}^{L} \left(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l} \right), \qquad\qquad \text{for } T - \Delta_1 < t \le T. \end{cases} \tag{B.9}$$

The variables denoted with an upper bar are obtained from the means of the approximate posterior distributions. L is the total number of time lags. L' equals a number of integers within the range of $t + \Delta_{L'} \leq T < t + \Delta_{L'+1}$ for $T - \Delta_L < t \leq T$, and $L' = L$ for $1 < t \leq T - \Delta_L$. $\bar{\mathbf{V}}_t$ and $\bar{\mathbf{J}}_t$ in Eq. (B.8) is calculated from the following computationally efficient form:

$$\begin{cases} \bar{\mathbf{V}}_t = \big(\mathbf{I} - \mathbf{K}\mathbf{G}\big)\hat{\mathbf{V}}_t \\ \bar{\mathbf{J}}_t = \hat{\mathbf{J}}_t + \mathbf{K}\big(\mathbf{B}_t - \mathbf{G}\hat{\mathbf{J}}_t\big) \end{cases} \tag{B.10}$$

where $\mathbf{K} = \hat{\mathbf{V}}_t\mathbf{G}^{\mathrm{T}}(\mathbf{G}\hat{\mathbf{V}}_t\mathbf{G}^{\mathrm{T}} + \mathbf{S})^{-1}$ is a gain matrix for the sensor measurements and $\mathbf{I}$ is the identity matrix.

The β-step updates the approximate posterior distribution of the scaling parameter $Q(\beta)$. Substituting β for x in Eq. (2.14) and calculating its expectation term, we obtain linear and logarithmic terms of β. Therefore, $Q(\beta)$ is a gamma distribution:

$$Q(\beta) = \mathcal{G}(\beta \,|\, \bar{\beta}, \gamma_\beta) \tag{B.11}$$

whose shape parameter γ_β and mean $\bar{\beta}$ are given by

$$\begin{cases} \gamma_\beta = \frac{1}{2}MT \\ \bar{\beta} = \gamma_\beta \Bigg\{ \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left(\mathbf{S}^{-1}\Big(\mathbf{B}_t\mathbf{B}_t^{\mathrm{T}} - \left(\mathbf{G}\bar{\mathbf{J}}_t\right)\mathbf{B}_t^{\mathrm{T}} - \mathbf{B}_t\left(\mathbf{G}\bar{\mathbf{J}}_t\right)^{\mathrm{T}} \right. \right. \\ \qquad \left. \left. + \left(\mathbf{G}\bar{\mathbf{J}}_t\right)\left(\mathbf{G}\bar{\mathbf{J}}_t\right)^{\mathrm{T}} \Big) \right) + \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left(\mathrm{diag}(\bar{\mathbf{q}}) \left(\bar{\mathbf{J}}_t\bar{\mathbf{J}}_t^{\mathrm{T}} \right. \right. \\ \qquad \left. \left. - \Big(\sum_{l=1}^{L}\Big(\bar{\mathbf{A}}_l\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)\bar{\mathbf{J}}_t^{\mathrm{T}} - \bar{\mathbf{J}}_t\Big(\sum_{l=1}^{L}\Big(\bar{\mathbf{A}}_l\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)^{\mathrm{T}} \right. \right. \\ \qquad \left. \left. + \left\langle \Big(\sum_{l=1}^{L}\Big(\mathbf{A}_l\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)\Big(\sum_{l=1}^{L}\Big(\mathbf{A}_l\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)^{\mathrm{T}} \right\rangle \right) \right) \Bigg\}^{-1}. \end{cases} \tag{B.12}$$

The expectation terms in the update equations are evaluated as follows. These terms, which are used in the **A**-step, $\boldsymbol{\eta}$-step, and **q**-step, are derived as

$$\left\langle \mathbf{J}_{t-\Delta_l}\mathbf{J}_{t-\Delta_l}^{\mathrm{T}} \right\rangle = \bar{\mathbf{J}}_{t-\Delta_l}\bar{\mathbf{J}}_{t-\Delta_l}^{\mathrm{T}} + \bar{\mathbf{V}}_{t-\Delta_l}, \tag{B.13}$$

$$\left\langle \mathbf{a}_n\mathbf{a}_n^{\mathrm{T}} \right\rangle = \bar{\mathbf{a}}_n\bar{\mathbf{a}}_n^{\mathrm{T}} + \bar{\mathbf{U}}_n, \tag{B.14}$$

$$\left\langle \mathbf{J}_t\mathbf{J}_t^{\mathrm{T}} \right\rangle = \bar{\mathbf{J}}_t\bar{\mathbf{J}}_t^{\mathrm{T}} + \bar{\mathbf{V}}_t. \tag{B.15}$$

After some rearrangement, the final term of the **q**-step in Eq. (B.6) is evaluated with the above expectation terms as:

$$\left(\left\langle \left(\sum_{l=1}^{L} \left(\mathbf{A}_l \, \mathbf{J}_{t-\Delta_l} \right) \right) \left(\sum_{l=1}^{L} \left(\mathbf{A}_l \, \mathbf{J}_{t-\Delta_l} \right) \right)^{\mathrm{T}} \right\rangle \right)_{n,n}$$
$$= \operatorname{tr} \left(\left\langle \mathbf{a}_n \, \mathbf{a}_n^{\mathrm{T}} \right\rangle \left(\sum_{l=1}^{L} \mathbf{E}_{n,\Delta_l} .* \left\langle \mathbf{J}_{t-\Delta_l} \mathbf{J}_{t-\Delta_l}^{\mathrm{T}} \right\rangle \right)_{C_n, C_n} \right). \tag{B.16}$$

The trace of the final term of the β-step in Eq. (B.12) is derived from Eq. (B.16) by replacing the expectation of the quadratic term of the current sources with its mean. The remaining expectation terms reside in Eq. (B.9) for the **J**-step. These are obtained by computing the following expectation term for all pairs of l' and l:

$$\left\langle \mathbf{A}_{l'}^{\mathrm{T}} \operatorname{diag}(\bar{\mathbf{q}}) \, \mathbf{A}_l \right\rangle = \sum_{n=1}^{N} \bar{q}_n \left\langle \tilde{\mathbf{a}}_{l'n} \, \tilde{\mathbf{a}}_{l\,n}^{\mathrm{T}} \right\rangle, \tag{B.17}$$

where $\tilde{\mathbf{a}}_{l\,n}^{\mathrm{T}}$ denotes the n-th row of $\mathbf{A}_l$. The expectation in the right-hand side of Eq. (B.17) is obtained by first expanding Eq. (B.14) into an N-dimensional matrix (based on the original source index of $\mathbf{a}_n$). Then, the matrix entries not corresponding to row and column source indices that structurally connect the n-th source index with time lags $\Delta_{l'}$ and Δ_l, respectively, are padded with zeros.

Free energy

The free energy F in Chapter 4 can be decomposed as

$$\begin{aligned} F &= \left\langle \log \frac{P(\mathbf{B}_{1:T}, \mathbf{J}_{1:T}, \beta, \mathbf{A}, \boldsymbol{\eta}_{1:N}, \mathbf{q})}{Q(\mathbf{J}_{1:T}, \beta, \mathbf{A}, \boldsymbol{\eta}_{1:N}, \mathbf{q})} \right\rangle \\ &= \left\langle \sum_{t=1}^{T} \left(\log \frac{P(\mathbf{B}_t \mid \mathbf{J}_t, \beta) P(\mathbf{J}_t \mid \mathbf{J}_{t-\Delta_1}, \ldots, \mathbf{J}_{t-\Delta_L}, \beta, \mathbf{A}, \mathbf{q})}{Q(\mathbf{J}_t)} \right) \right\rangle \\ &\quad + \left\langle \log \frac{P(\beta)}{Q(\beta)} \right\rangle + \left\langle \log \frac{P(\mathbf{A} \mid \boldsymbol{\eta}_{1:N})}{Q(\mathbf{A})} \right\rangle + \left\langle \log \frac{P(\boldsymbol{\eta}_{1:N})}{Q(\boldsymbol{\eta}_{1:N})} \right\rangle + \left\langle \log \frac{P(\mathbf{q})}{Q(\mathbf{q})} \right\rangle \\ &= L + H_\beta + H_A + H_\eta + H_q. \end{aligned} \tag{B.18}$$

Each constituent of the free energy is obtained by computing

$$
\begin{aligned}
L = \sum_{t=1}^{T} \Bigg\{ & \Big\langle \log \mathcal{N}\Big(\mathbf{B}_t \,|\, \mathbf{G}\mathbf{J}_t, \beta^{-1}\mathbf{S}\Big) \Big\rangle \\
& + \Big\langle \log \mathcal{N}\Big(\mathbf{J}_t \,\Big|\, \sum_{l=1}^{L}\Big(\mathbf{A}_l \mathbf{J}_{t-\Delta_l}\Big), (\beta\,\mathrm{diag}(\mathbf{q}))^{-1}\Big) \Big\rangle \\
& - \Big\langle \log \mathcal{N}(\mathbf{J}_t \,|\, \bar{\mathbf{J}}_t, \bar{\beta}^{-1}\bar{\mathbf{V}}_t) \Big\rangle \Bigg\} \\
= \sum_{t=1}^{T} \Bigg\{ & -\frac{M}{2}\log 2\pi + \frac{M}{2}\langle \log\beta\rangle - \frac{1}{2}\log|\mathbf{S}| \\
& \quad - \frac{\bar{\beta}}{2}\,\mathrm{tr}\Big(\mathbf{S}^{-1}\big(\mathbf{B}_t\mathbf{B}_t^{\mathrm{T}} - (\mathbf{G}\bar{\mathbf{J}}_t)\mathbf{B}_t^{\mathrm{T}} - \mathbf{B}_t(\mathbf{G}\bar{\mathbf{J}}_t)^{\mathrm{T}} + \mathbf{G}\left\langle \mathbf{J}_t\mathbf{J}_t^{\mathrm{T}}\right\rangle \mathbf{G}^{\mathrm{T}}\big)\Big) \\
& \quad - \frac{N}{2}\log 2\pi + \frac{N}{2}\langle\log\beta\rangle + \frac{1}{2}\sum_{n=1}^{N}\langle \log q_n\rangle \\
& \quad\quad - \frac{\bar{\beta}}{2}\,\mathrm{tr}\Bigg(\mathrm{diag}(\bar{\mathbf{q}})\Bigg(\Big\langle \mathbf{J}_t\mathbf{J}_t^{\mathrm{T}}\Big\rangle \\
& \quad\quad\quad - \Big(\sum_{l=1}^{L}\Big(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)\bar{\mathbf{J}}_t^{\mathrm{T}} - \bar{\mathbf{J}}_t\Big(\sum_{l=1}^{L}\Big(\bar{\mathbf{A}}_l\,\bar{\mathbf{J}}_{t-\Delta_l}\Big)\Big)^{\mathrm{T}} \\
& \quad\quad\quad + \Big\langle\Big(\sum_{l=1}^{L}\Big(\mathbf{A}_l\,\mathbf{J}_{t-\Delta_l}\Big)\Big)\Big(\sum_{l=1}^{L}\Big(\mathbf{A}_l\,\mathbf{J}_{t-\Delta_l}\Big)\Big)^{\mathrm{T}}\Big\rangle\Bigg)\Bigg) \\
& \quad + \frac{N}{2}(1+\log 2\pi) - \frac{N}{2}\log\bar{\beta} + \frac{1}{2}\log|\bar{\mathbf{V}}_t|\Bigg\},
\end{aligned}
\tag{B.19}
$$

$$
\begin{aligned}
H_\beta &= \langle \log\beta^{-1}\rangle - \langle \log\,\mathcal{G}(\beta\,|\,\bar{\beta},\gamma_\beta)\rangle \\
&= \log\Gamma(\gamma_\beta) - \gamma_\beta\,\psi(\gamma_\beta) + \gamma_\beta,
\end{aligned}
\tag{B.20}
$$

$$
\begin{aligned}
H_A &= \sum_{n=1}^{N}\Big\{\Big\langle \log\mathcal{N}\Big(\mathbf{a}_n\,|\,\mathbf{0},\mathrm{diag}(\boldsymbol{\eta}_n)^{-1}\Big)\Big\rangle - \Big\langle\log\mathcal{N}(\mathbf{a}_n\,|\,\bar{\mathbf{a}}_n,\bar{\mathbf{U}}_n)\Big\rangle\Big\} \\
&= \sum_{n=1}^{N}\Bigg\{-\frac{K_n}{2}\log 2\pi + \frac{1}{2}\sum_{k=1}^{K_n}\langle\log\eta_{n,k}\rangle - \frac{1}{2}\mathrm{tr}\Big(\mathrm{diag}(\bar{\boldsymbol{\eta}}_n)\Big\langle \mathbf{a}_n\,\mathbf{a}_n^{\mathrm{T}}\Big\rangle\Big) \\
&\qquad + \frac{K_n}{2}(1+\log 2\pi) + \frac{1}{2}\log|\bar{\mathbf{U}}_n|\Bigg\},
\end{aligned}
\tag{B.21}
$$

$$H_\eta = \sum_{n=1}^{N}\sum_{k=1}^{K_n}\left\{\left\langle \log\, \mathcal{G}(\eta_{n,k} \mid \eta_0, g_0)\right\rangle - \left\langle \log\, \mathcal{G}(\eta_{n,k} \mid \bar{\eta}_{n,k}, g_{n,k})\right\rangle\right\}$$
$$= \sum_{n=1}^{N}\sum_{k=1}^{K_n}\left\{g_0\Big(\log\left(\bar{\eta}_{n,k}\,\eta_0^{-1}\right) - \bar{\eta}_{n,k}\,\eta_0^{-1} + 1\Big) + H(g_{n,k}, g_0)\right\}, \tag{B.22}$$

$$H_q = \sum_{n=1}^{N}\left\{\left\langle \log\, \mathcal{G}(q_n \mid \bar{v}_n^{-1}, \gamma_0)\right\rangle - \left\langle \log\, \mathcal{G}(q_n \mid \bar{q}_n, \gamma_n)\right\rangle\right\}$$
$$= \sum_{n=1}^{N}\left\{\gamma_0\Big(\log\left(\bar{q}_n\bar{v}_n\right) - \bar{q}_n\bar{v}_n + 1\Big) + H(\gamma_n, \gamma_0)\right\}. \tag{B.23}$$

The term H in H_η and H_q is defined by Eq. (A.23) and the logarithmic expections in L and H_A are evaluated as follows:

$$\langle \log \beta\rangle = \log\bar{\beta} + \psi(\gamma_\beta) - \log\gamma_\beta, \tag{B.24}$$
$$\langle \log q_n\rangle = \log\bar{q}_n + \psi(\gamma_n) - \log\gamma_n, \tag{B.25}$$
$$\langle \log \eta_{n,k}\rangle = \log\bar{\eta}_{n,k} + \psi(g_{n,k}) - \log g_{n,k}. \tag{B.26}$$

B.2. Bias correction of AUC

To interpret AUC as an unbiased detection accuracy score, one should provide the same number of active and inactive sources (or connections) in the ROC analysis [35]. In our simulations, the number of actives was actually far less than the number of inactives, as usually assumed in distributed source methods. Thus, to compute the AUC, we randomly and sequentially extracted the inactives until the number of actives and inactives matched. We computed two types of AUC by extracting the inactives from a close neighborhood of actives ($\mathrm{AUC}_{\mathrm{close}}$) and from the far local maxima of false positives ($\mathrm{AUC}_{\mathrm{far}}$). The final score of AUC was defined as the mean of $\mathrm{AUC}_{\mathrm{close}}$ and $\mathrm{AUC}_{\mathrm{far}}$, both computed from the average scores of 50 different drawings from a set of inactives. $\mathrm{AUC}_{\mathrm{close}}$ and $\mathrm{AUC}_{\mathrm{far}}$ for sources were computed similarly to [35]. By an equivalent procedure, $\mathrm{AUC}_{\mathrm{close}}$ and $\mathrm{AUC}_{\mathrm{far}}$ for connections were obtained (with some modifications, as described below).

$\mathrm{AUC}_{\mathrm{close}}$ was used to assess the reconstruction accuracy of the spatial extent of sources or connections. The spatial neighborhood in computing $\mathrm{AUC}_{\mathrm{close}}$ was specified by a 5 cm-radius sphere centered on the actives. The area of this sphere is comparable to the spatial neighborhood adopted in [35]. When calculating this score for connections, the spatial neighborhood of a connection was defined as connections of which seed and target sources are included in the spatial neighborhoods of the seed and target sources of the centered connection, respectively.

$\mathrm{AUC}_{\mathrm{far}}$ was computed to detect false positives far from the actives. To compute this score for sources, a coarse cortical parcel was initially selected at random. Within this parcel subtracted by the area occupied by the spatial neighborhood of the actives, the false positive source with maximum magnitude was then extracted. The coarse cortical parcel was obtained by a nearest-neighbor manner parcellation of the whole cortex into 86 parcels, the number of parcels specified in [35]. When calculating this score for connections, we extracted the false positive connection with maximum absolute MAR coefficient from a randomly selected pair of parcels, whose connections were not included in the spatial neighborhood of the actives. When selecting parcels or pairs of parcels, we excluded in advance those candidates lacking a source or structural connection beyond the spatial neighborhood.

B.3. Network of neural mass models

This appendix details the functional network of neural mass models used for generating the stimulus-evoked responses in Simulation 2. The equations of this network model were integrated by a standard Runge-Kutta method with an integration time step of 0.1 ms (the generated waveforms were afterwards downsampled to 1 kHz). The parameter settings used in each component of the network model, namely, the neural mass model, the extrinsic coupling, and the input signal, are provided below.

The neural mass model was based on a model of two subpopulations [14]. The convolution kernels for the excitatory and the inhibitory synapses, represented as $(H_e t/\tau_e)\exp(-t/\tau_e)$ and $(H_i t/\tau_i)\exp(-t/\tau_i)$ for $t \geq 0$, respec-

tively, were assumed in both subpopulations. The parameters τ_e and τ_i for the two subpopulations were set to the values adopted by [14]. The parameter specifying the relative proportion of the subpopulation with larger τ_e and τ_i was set to 0.7. Under this parameter setting, the activity peaks occurred at 100 ms and 170 ms following the input, mimicking the stimulus-evoked responses during face perception. For each subpopulation, the parameters H_e and H_i were derived by dividing the values of $H_e\tau_e$ and $H_i\tau_i$ in [15] by the above-specified values of τ_e and τ_i, respectively. The parameters governing the intrinsic couplings and the sigmoid functions of the neural mass model were set according to [15], while the relative strength among the intrinsic couplings was specified by the setting popularly used in the literature [14,16,49,75].

The extrinsic coupling parameters in the network model were set to reproduce typical waveforms of stimulus-evoked responses. Following [15], who extensively investigated the relationships between extrinsic coupling parameters and response waveforms, we set these parameters to 50 and 10 for forward and backward connections, respectively. The time lags in the extrinsic couplings were determined solely from the fiber transmission delay. The local delay constant τ in Eq. (4.2) was omitted because it had been embodied in the convolution kernels of the neural mass model.

The input signal to the network model was generated by a gamma density function and a set of discrete cosine functions, as described in [16]. The shape and scale parameters of the gamma density function were set to 64 and 1024, respectively, yielding activity peaks at 100 ms and 170 ms. The parameters in the discrete cosine functions were set to those adopted in [16].

B.4. The non-dynamic method in a conventional approach

In this appendix[1], we compare the results of Chapter 4 with the non-dynamic method whose estimates of the MAR matrix were computed after the ex-

[1]The manuscript submitted to NeuroImage on November 29th, 2013, presents a part of the results in this appendix, instead of those in Chapter 4.

traction of active sources (i.e., a conventional approach of the non-dynamic method; see Section 4.3.5). Advantages of the dynamic method were investigated including those owing to the whole-brain parameter estimation, which has not been adopted by standard methods of effective connectivity estimation (e.g., DCM [16, 26]).

Fig. B.1 shows AUC and nRMSE of the MAR matrix in Simulation 1. In addition to nRMSE, clear differences of AUC were found between the non-dynamic and dynamic methods. The AUC score was also largely improved by the dynamic method in the dense setting of Simulation 2 (see Table B.1). The estimated effective connectivity among active sources presented in Fig. B.2

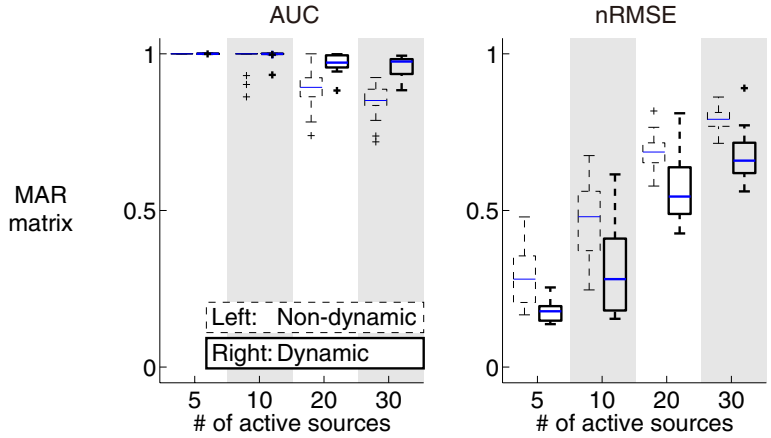

Figure B.1: Evaluation of the estimation accuracy in Simulation 1, comparing with a conventional approach of the non-dynamic method. See the caption for the lower part of Fig. 4.2.

Table B.1:
AUC of the MAR matrix with (and without) applying the fMRI prior in Simulation 2, comparing with a conventional approach of the non-dynamic method (see the caption for the lower part of Table 4.2).

	Non-dynamic		Dynamic	
	Sparse	Dense	Sparse	Dense
MAR matrix				
AUC	**1.00** (0.75)	0.55 (0.50)	**1.00** (0.87)	**0.91** (0.76)

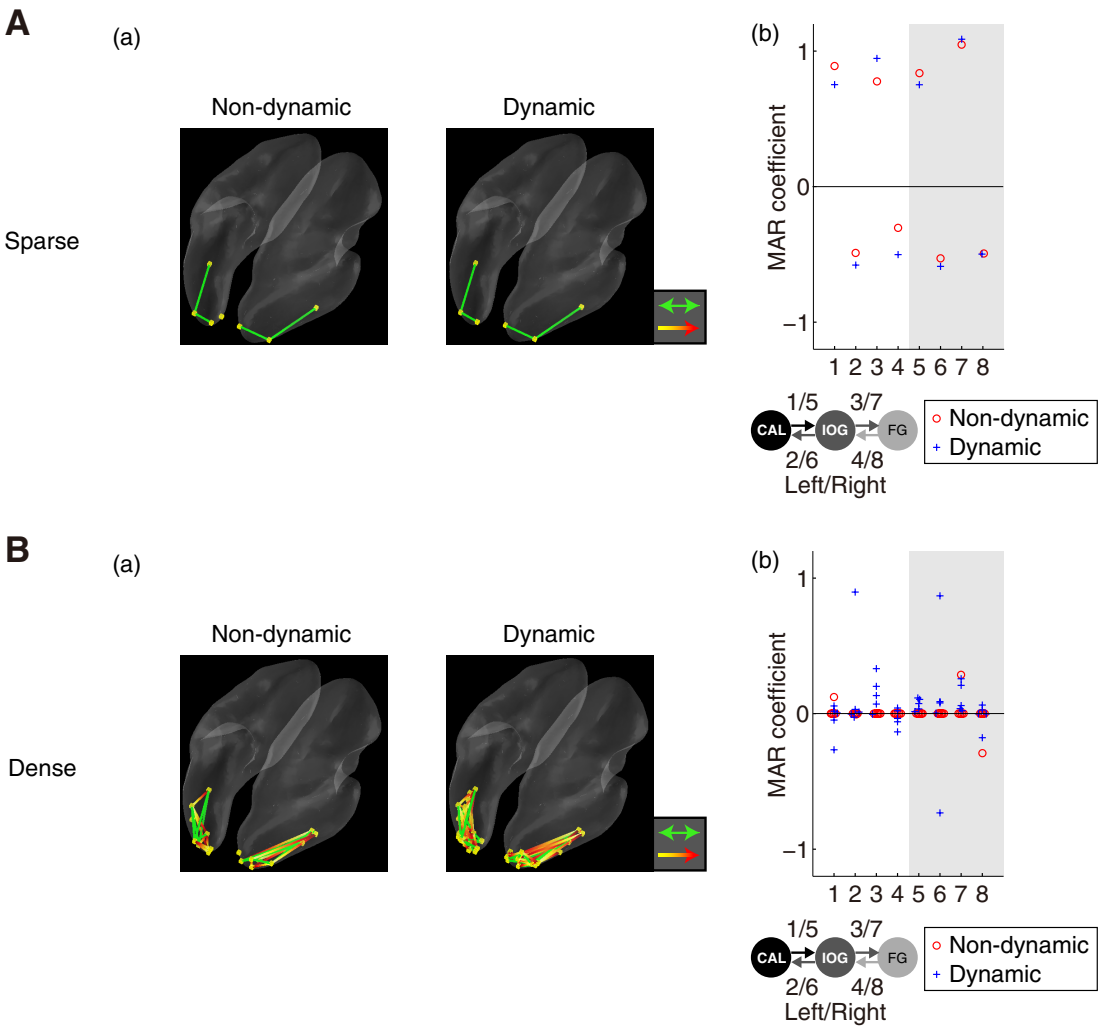

Figure B.2: Results obtained from the non-dynamic and dynamic methods with the fMRI prior in Simulation 2, comparing with a conventional approach of the non-dynamic method. See the caption for (a) and (c) of Figs. 4.4A, and 4.4B.

was similar to that of Fig. 4.4, while the MAR coefficients of the non-dynamic method were more likely to be regarded as zero in Fig. B.2, because of the extraction of active sources when computing the MAR matrix.

Fig. B.3 presents that advantages in the prediction accuracy of the dynamic method became noticeable when comparing with a conventional approach of the non-dynamic method. Similar to the results of Simulation 2, the estimated effective connectivity among active sources (see Fig. B.4) was almost

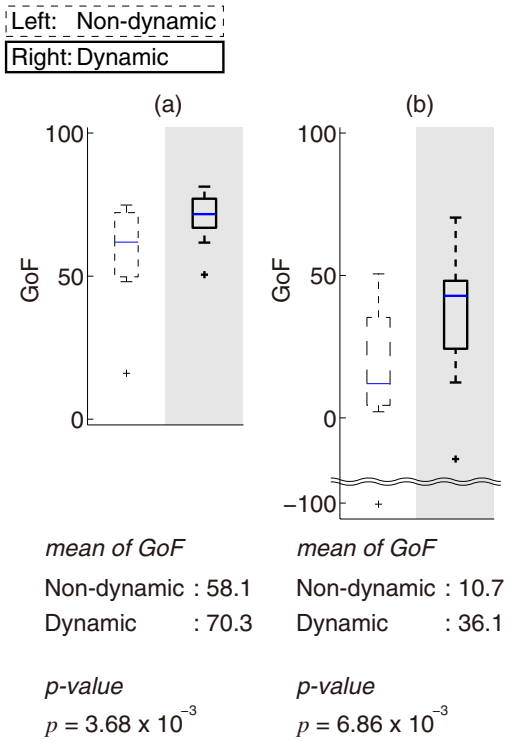

Figure B.3: Results of GoF analysis on real experimental data that contain a missing duration, comparing with a conventional approach of the non-dynamic method. See the caption for Fig. 4.7C.

the same as that presented in Fig. 4.8. For Subject I, the MAR coefficients of the non-dynamic method between IOG–FG were regarded as zero because IOG sources were absent (see Fig. B.5A). For Subject VIII, there were no essential differences in these coefficients of the non-dynamic method between Figs. 4.9B and B.5B.

Overall, differences in evaluation scores between the non-dynamic and dynamic methods were enlarged by adopting the conventional approach of the non-dynamic method; where the MAR matrix was computed only from pairs of active sources. This characteristic suggests that the procedure of estimating effective connectivity over the whole brain itself is also an important component for obtaining accurate solutions, as well as dynamic modeling and simultaneous estimation.

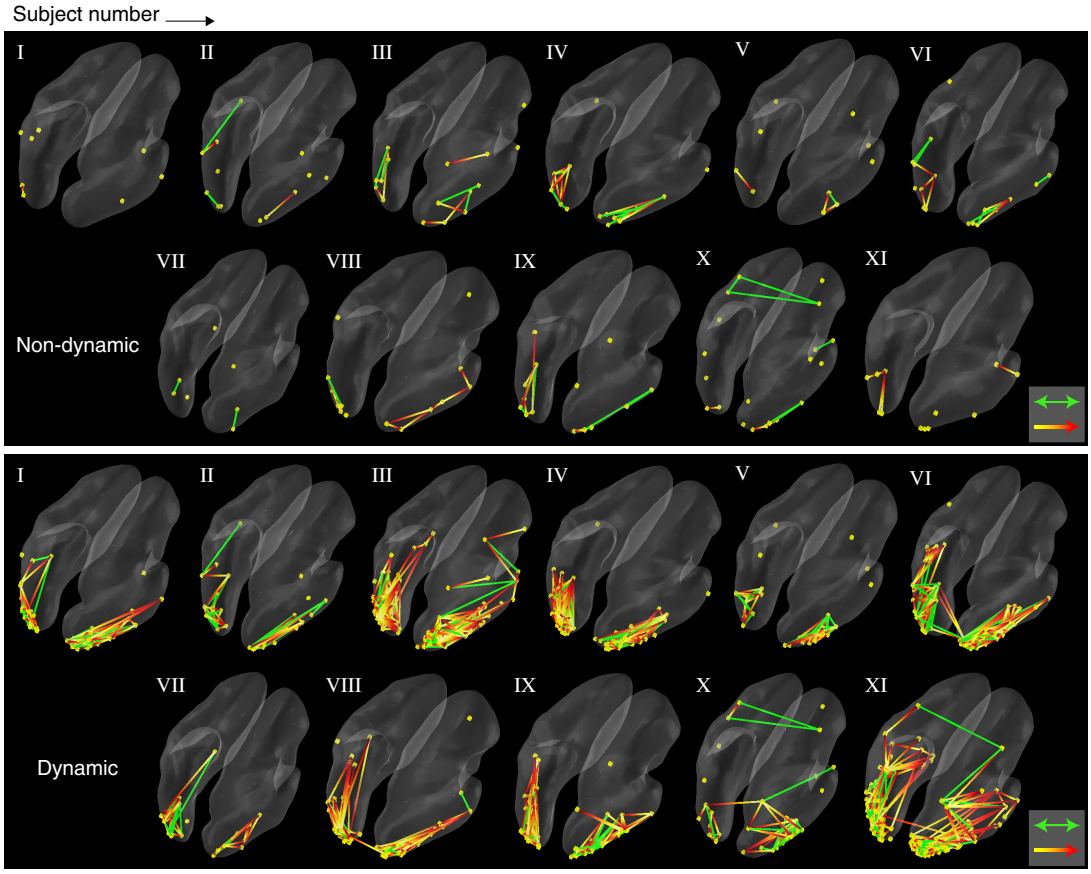

Figure B.4: Reconstructed current sources and their effective connectivity, comparing with a conventional approach of the non-dynamic method. See the caption for Fig. 4.8.

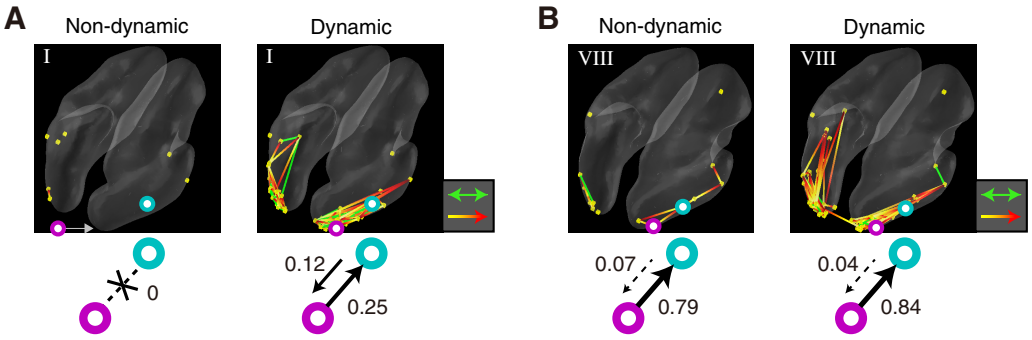

Figure B.5: Representative examples of the event-related dynamics of stimulus-evoked responses, comparing with a conventional approach of the non-dynamic method. See the caption for the upper part of Fig. 4.9.

Bibliography

[1] H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes," in *Proc. 15th Conf. Uncertainty in Artificial Intelligence*, 1999, pp. 21–30.

[2] 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.

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

[4] M. J. Barton, P. A. Robinson, S. Kumar, A. Galka, H. F. Durrant-Whyte, J. Guivant, and T. Ozaki, "Evaluating the performance of Kalman-filter-based EEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 122–136, 2009.

[5] M. J. Beal, "Variational algorithm for approximate Bayesian inference," Ph.D. dissertation, Univ. College London, London, UK, 2003.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[7] 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, 2009.

[8] A. Dale, A. Liu, B. Fischl, and R. Buckner, "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity," *Neuron*, vol. 26, no. 1, pp. 55–67, 2000.

[9] M. Dannhauer, E. Lämmel, C. H. Wolters, and T. R. Knösche, "Spatio-temporal regularization in linear distributed source reconstruction from EEG/MEG: A critical evaluation." *Brain Topogr.*, vol. 26, no. 2, pp. 229–246, 2013.

[10] F. Darvas, U. Schmitt, A. K. Louis, M. Fuchs, G. Knoll, and H. Buchner, "Spatio-temporal current density reconstruction (stCDR) from EEG/MEG-data," *Brain Topogr.*, vol. 13, no. 3, pp. 195–207, 2001.

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

[12] J. Daunizeau, C. Grova, G. Marrelec, J. Mattout, S. Jbabdi, M. Pélégrini-Issac, J.-M. Lina, and H. Benali, "Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework." *NeuroImage*, vol. 36, no. 1, pp. 69–87, 2007.

[13] 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.

[14] O. David and K. Friston, "A neural mass model for MEG/EEG: coupling and neuronal dynamics." *NeuroImage*, vol. 20, no. 3, pp. 1743–1755, 2003.

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

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

[17] J. Davies-Thompson and T. J. Andrews, "Intra- and interhemispheric connectivity between face-selective regions in the human brain." *J. Neurophysiol.*, vol. 108, no. 11, pp. 3087–3095, 2012.

[18] G. Deco, V. Jirsa, A. R. McIntosh, O. Sporns, and R. Kötter, "Key role of coupling, delay, and noise in resting brain fluctuations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 29, pp. 10 302–10 307, 2009.

[19] G. Deco, V. K. Jirsa, and A. R. McIntosh, "Emerging concepts for the dynamical organization of resting-state activity in the brain." *Nat. Rev. Neurosci.*, vol. 12, no. 1, pp. 43–56, 2011.

[20] L. Ding and B. He, "Spatio-temporal EEG source localization using a three-dimensional subspace FINE approach in a realistic geometry inhomogeneous head model." *IEEE Trans. Biomed. Eng.*, vol. 53, no. 9, pp. 1732–1739, 2006.

[21] S. L. Fairhall and A. Ishai, "Effective connectivity within the distributed cortical network for face perception." *Cereb. Cortex*, vol. 17, no. 10, pp. 2400–2406, 2007.

[22] R. S. J. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. T. Ashburner, and W. D. Penny, *Human Brain Function*, 2nd ed. New York: Academic Press, 2004.

[23] K. Friston, "Beyond phrenology: what can neuroimaging tell us about distributed circuitry?" *Annu. Rev. Neurosci.*, vol. 25, pp. 221–250, 2002.

[24] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Mapp.*, vol. 2, no. 1-2, pp. 56–78, 1994.

[25] K. J. Friston, L. M. Harrison, J. Daunizeau, S. J. 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, 2008.

[26] K. J. Friston, L. M. Harrison, and W. D. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[27] M. Fukushima, O. Yamashita, A. Kanemura, S. Ishii, M. Kawato, and M. Sato, "A state-space modeling approach for localization of focal cur-

rent sources from MEG." *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1561–1571, 2012.

[28] M. Fukushima, O. Yamashita, T. R. Knösche, and M. Sato, "MEG source reconstruction constrained by diffusion MRI based whole brain dynamical model," in *Proc. IEEE 10th Int. Symp. Biomed. Imaging*, 2013, pp. 990–993.

[29] D. Gabriel, E. Veuillet, R. Ragot, D. Schwartz, A. Ducorps, A. Norena, J. D. Durrant, A. Bonmartin, F. Cotton, and L. Collet, "Effect of stimulus frequency and stimulation site on the N1m response of the human auditory cortex," *Hear. Res.*, vol. 197, no. 1-2, pp. 55–64, 2004.

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

[31] A. Ghosh, Y. Rho, A. McIntosh, R. Kötter, and V. Jirsa, "Noise during rest enables the exploration of the brain's dynamic repertoire." *PLoS Comput. Biol.*, vol. 4, no. 10, p. e1000196, 2008.

[32] B. Godey, D. Schwartz, J. B. de Graaf, P. Chauvel, and C. Liégeois-Chauvel, "Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients," *Clin. Neurophysiol.*, vol. 112, no. 10, pp. 1850–1859, 2001.

[33] M. A. Goodale and A. Milner, "Separate visual pathways for perception and action," *Trends Neurosci.*, vol. 15, no. 1, pp. 20–25, 1992.

[34] 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, 1995.

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

[36] M. Gschwind, G. Pourtois, S. Schwartz, D. Van De Ville, and P. Vuilleumier, "White-matter connectivity between face-responsive regions in the human brain." *Cereb. Cortex*, vol. 22, no. 7, pp. 1564–1576, 2012.

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

[38] 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.

[39] R. Hari, K. Aittoniemi, M. L. Järvinen, T. Katila, and T. Varpula, "Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators," *Exp. Brain Res.*, vol. 40, no. 2, pp. 237–240, 1980.

[40] J. Haxby, E. Hoffman, and M. Gobbini, "The distributed human neural system for face perception." *Trends Cogn. Sci.*, vol. 4, no. 6, pp. 223–233, 2000.

[41] J. Haxby, B. Horwitz, L. Ungerleider, J. Maisog, P. Pietrini, and C. Grady, "The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations," *J. Neurosci.*, vol. 14, no. 11, pp. 6336–6353, 1994.

[42] R. N. Henson, G. Flandin, K. J. Friston, and J. Mattout, "A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction." *Hum. Brain Mapp.*, vol. 31, no. 10, pp. 1512–1531, 2010.

[43] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston, "A parametric empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration," *Front. Hum. Neurosci.*, vol. 5, no. 76, pp. 1–16, 2011.

[44] A. Hillebrand and G. R. Barnes, "A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex," *NeuroImage*, vol. 16, no. 3, pp. 638–650, 2002.

[45] 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," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 6, pp. 2035–2040, 2009.

[46] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns, "Network structure of cerebral cortex shapes functional connectivity on multiple time scales." *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 24, pp. 10 240–10 245, 2007.

[47] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, "Dynamic functional connectivity: promise, issues, and interpretations." *NeuroImage*, vol. 80, pp. 360–378, 2013.

[48] T. J. Imig and H. O. Adrián, "Binaural columns in the primary field (A1) of cat auditory cortex," *Brain Res.*, vol. 138, no. 2, pp. 241–257, 1977.

[49] B. Jansen and V. Rit, "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns," *Biol. Cybern.*, vol. 366, pp. 357–366, 1995.

[50] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: A module in human extrastriate cortex specialized for face perception." *J. Neurosci.*, vol. 17, no. 11, pp. 4302–4311, 1997.

[51] 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.

[52] C. Lamus, C. J. Long, M. S. Hämäläinen, E. N. Brown, and P. L. Purdon, "Parameter estimation and dynamic source localization for the magnetoencephalography (MEG) inverse problem," in *4th IEEE Int. Symp. Biomedical Imaging*, 2007, pp. 1092–1095.

[53] X. Lei, P. Xu, C. Luo, J. Zhao, D. Zhou, and D. Yao, "fMRI functional networks for EEG source imaging." *Hum. Brain Mapp.*, vol. 32, no. 7, pp. 1141–1160, 2011.

[54] G. Lohmann, K. Erfurth, K. Müller, and R. Turner, "Critical comments on dynamic causal modelling." *NeuroImage*, vol. 59, no. 3, pp. 2322–2329, 2012.

[55] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 6, pp. 608–615, 1995.

[56] 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.

[57] 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, 1999.

[58] R. M. Neal, *Bayesian Learning for Neural Networks*. New York: Springer, 1996.

[59] 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, 2007.

[60] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. New York: Oxford University Press, 2006.

[61] I. Olier, N. J. Trujillo-Barreto, and W. El-Deredy, "A switching multi-scale dynamical network model of EEG/MEG." *NeuroImage*, vol. 81, pp. 262–287, 2013.

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

[63] W. Ou, A. Nummenmaa, J. Ahveninen, J. W. Belliveau, M. S. Hämäläinen, and P. Golland, "Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation." *NeuroImage*, vol. 52, no. 1, pp. 97–108, 2010.

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

[65] S. Palva and J. M. Palva, "Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs." *Trends Cogn. Sci.*, vol. 16, no. 4, pp. 219–230, 2012.

[66] C. Pantev, B. Ross, P. Berg, T. Elbert, and B. Rockstroh, "Study of the human auditory cortices using a whole-head magnetometer: left vs. right hemisphere and ipsilateral vs. contralateral stimulation," *Audiol. Neurootol.*, vol. 3, no. 2-3, pp. 183–190, 1998.

[67] 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, 1994.

[68] K. Portin, S. Vanni, V. Virsu, and R. Hari, "Stronger occipital cortical activation to lower than upper visual field stimuli. Neuromagnetic recordings." *Exp. Brain Res.*, vol. 124, no. 3, pp. 287–294, 1999.

[69] J. A. Pyles, T. D. Verstynen, W. Schneider, and M. J. Tarr, "Explicating the face perception network with white matter connectivity." *PLoS ONE*, vol. 8, no. 4, p. e61611, 2013.

[70] M. Sato, "Online model selection based on the variational Bayes," *Neural Comput.*, vol. 13, no. 7, pp. 1649–1681, 2001.

[71] M. 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, 2004.

[72] M. Scherg and D. Von Cramon, "Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model," *Electroencephalogr. Clin. Neurophysiol.*, vol. 62, no. 1, pp. 32–44, 1985.

[73] U. Schmitt, A. K. Louis, F. Darvas, H. Buchner, and M. Fuchs, "Numerical aspects of spatio-temporal current density reconstruction from EEG-/MEG-data," *IEEE Trans. Med. Imaging*, vol. 20, pp. 314–324, 2001.

[74] J.-M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG." *Hum. Brain Mapp.*, vol. 30, no. 6, pp. 1857–1865, 2009.

[75] A. Spiegler, S. J. Kiebel, F. M. Atay, and T. R. Knösche, "Bifurcation analysis of neural mass models: Impact of extrinsic inputs and dendritic time constants." *NeuroImage*, vol. 52, no. 3, pp. 1041–1058, 2010.

[76] K. E. Stephan, M. Tittgemeyer, T. R. Knösche, R. J. Moran, and K. J. Friston, "Tractography-based priors for dynamic causal models." *NeuroImage*, vol. 47, no. 4, pp. 1628–1638, 2009.

[77] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.

[78] S. Taulu and J. Simola, "Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements." *Phys. Med. Biol.*, vol. 51, no. 7, pp. 1759–1768, 2006.

[79] J. Tournier, F. Calamante, and A. Connelly, "Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution." *NeuroImage*, vol. 35, no. 4, pp. 1459–72, 2007.

[80] 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, 2008.

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

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

[83] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG." *NeuroImage*, vol. 49, no. 1, pp. 641–655, 2010.

[84] M. W. Woolrich and K. E. Stephan, "Biophysical network models and the human connectome." *NeuroImage*, vol. 80, pp. 330–338, 2013.

[85] P. Xu, Y. Tian, X. Lei, and D. Yao, "Neuroelectric source imaging using 3SCO: a space coding algorithm based on particle swarm optimization and l_0 norm constraint." *NeuroImage*, vol. 51, no. 1, pp. 183–205, 2010.

[86] X.-L. Xu, B. Xu, and B. He, "An alternative subspace approach to EEG dipole source localization," *Phys. Med. Biol.*, vol. 49, no. 2, pp. 327–343, 2004.

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

[88] T. Yoshioka, K. Toyama, M. Kawato, O. Yamashita, S. Nishina, N. Yamagishi, and M. Sato, "Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals." *NeuroImage*, vol. 42, no. 4, pp. 1397–1413, 2008.