NeuroImage 55 (2011) 113–132

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Bayesian symmetrical EEG/fMRI fusion with spatially adaptive priors

Martin Luessi [a,*], S. Derin Babacan [b], Rafael Molina [c], James R. Booth [d], Aggelos K. Katsaggelos [a]

[a] Department of Electrical Engineering and Computer Science, Northwestern University, 2145 Sheridan Road, Evanston IL 60208, USA
[b] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 N Mathews Ave, Urbana IL 61801, USA
[c] Departamento de Ciencias de la Computación e I.A., Universidad de Granada, 18071 Granada, Spain
[d] Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston IL 60208, USA

ARTICLE INFO

Article history:
Received 13 March 2010
Revised 10 November 2010
Accepted 11 November 2010
Available online 2 December 2010

Keywords:
Multimodal fusion
M/EEG source localization
Spatial adaptivity
Total variation
Variational Bayes

ABSTRACT

In this paper, we propose a novel symmetrical EEG/fMRI fusion method which combines EEG and fMRI by means of a common generative model. We use a total variation (TV) prior to model the spatial distribution of the cortical current responses and hemodynamic response functions, and utilize spatially adaptive temporal priors to model their temporal shapes. The spatial adaptivity of the prior model allows for adaptation to the local characteristics of the estimated responses and leads to high estimation performance for the cortical current distribution and the hemodynamic response functions. We utilize a Bayesian formulation with a variational Bayesian framework and obtain a fully automatic fusion algorithm. Simulations with synthetic data and experiments with real data from a multimodal study on face perception demonstrate the performance of the proposed method.

Introduction

Electroencephalography (EEG) is one of the most widely used functional brain mapping methods. A main advantage of EEG is that it provides a direct measure of electrical activity in the brain via voltage sensors on the scalp and thus can achieve a high temporal resolution. However, locating the sources of activity in the brain from the EEG measurements is a difficult problem as there is an indefinite number of source configurations which give rise to the same measurements. The same problem is also encountered in magnetoencephalography (MEG), where the electrical activity in the brain is measured using magnetic field sensors. Due to the problem that the same measurements can be generated by an indefinite number of source configurations, EEG and MEG source localization are referred to as ill-posed inverse problems (Hämäläinen et al., 1993).

In the last two decades a large number of EEG and MEG source localization methods have been proposed in the literature. Due to the similarity of the inverse problems most methods are applicable to either modality and can be divided into two groups. The first group assumes that there is a small number (typically 1–5) of sources, each modeled by an equivalent current dipole (ECD) (Scherg and Von Cramon, 1986). The locations of the dipoles are found by performing a nonlinear optimization which minimizes the discrepancy to the data with respect to the dipole locations. While ECD methods are popular in practice, they have some major limitations: First, the number of dipoles has to be specified by the user and second, the optimization algorithm can get trapped in a local minimum and thus might not be able to find the optimal dipole locations. In fact, ECD methods are known to be unreliable when more than one dipole is used (Yao and Dewald, 2005). The second, and more recently proposed group of methods is referred to as distributed methods (Hämäläinen et al., 1993). Methods in this group assume a large number, typically several thousands, of dipoles with fixed locations which are distributed over the cortical surface. Source localization then amounts to finding the current amplitudes for all dipoles simultaneously, which is still an ill-posed problem since the number of dipoles is much larger than the number of sensors. However, the use of dipoles with fixed locations means that the forward problem is linear and source localization can be regarded as solving an underdetermined linear system of equations, which is similar to problems encountered in signal and image processing.

In order to find a unique solution, it is necessary to make assumptions about the solution. Such assumptions can be formulated as deterministic regularization terms, such as in the minimum norm method (Hämäläinen and Ilmoniemi, 1994), which finds the source configuration with minimal energy or in the low resolution electromagnetic tomography (LORETA) method (Pascual-Marqui et al., 1994), where a regularization term based on a spatial Laplacian is used to enforce a smooth solution.

The source localization problem can also be formulated as a Bayesian inference problem (Baillet and Garnero, 1997), which allows for an elegant way to include *a priori* information about the solution in the form of priors, such as spatial and temporal smoothness priors (Baillet and Garnero, 1997). The priors can be either fixed or can be automatically selected from a set of candidate priors, by means of

[*] Corresponding author. Fax: +1 847 491 4455.
E-mail address: m-luessi@northwestern.edu (M. Luessi).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.11.037

Bayesian model selection. Examples of methods using fixed priors are ℓ_2-norm methods (Baillet et al., 2001), ℓ_1-norm methods (Uutela et al., 1999; Huang et al., 2006), as well as, the Bayesian formulation of the LORETA method (Pascual-Marqui et al., 1994). As stated in Wipf and Nagarajan (2009) there is a number of methods which attempt to perform Bayesian model selection. Examples of methods which automatically select priors using Bayesian model selection are methods which use a Gaussian prior with a linear combination of covariance components (Phillips et al., 2005; Mattout et al., 2006; Friston et al., 2006, 2008). These methods employ an empirical Bayesian scheme to estimate the hyperparameters controlling the contribution of each component. This formulation is very flexible and allows for the combination of priors such as spatial Laplacian, minimum norm, and depth constraints. Methods which use automatic relevance determination (ARD) (MacKay, 1992; Tipping, 2001; Ramírez, 2005; Wipf, 2006; Wipf et al., 2010) are based on similar ideas, i.e., the estimation of covariance components, but are more effective when the number of components is large. Typically, a separate hyperparameter is used for every diagonal element of the covariance matrix, which leads to a sparse solution, i.e., a solution with a small number of active dipoles, similar to ℓ_1-norm regularization. Many existing M/EEG source localization methods can be formulated in a unified Bayesian framework; we refer to Wipf and Nagarajan (2009) where the framework is introduced for a more thorough review of Bayesian M/EEG source localization methods.

The Bayesian treatment of M/EEG source localization offers advantages other than the automatic determination of relevant priors. The Bayesian formulation offers a formal way to include information from other functional neuroimaging modalities, such as functional magnetic resonance imaging (fMRI), into the source localization problem.

In recent years, fMRI has become a prominent neuroimaging method as it offers a very high spatial resolution. On the other hand, the temporal resolution is limited by technical and physical constraints, which limit the repetition time (TR) to be in the order of seconds, as well as, by the indirect mechanism fMRI uses to measure neuronal activity, i.e., the so-called blood oxygen level dependent (BOLD) contrast (Ogawa et al., 1990; Frahm et al., 1992), which depends on slow hemodynamic processes. However, the complementary advantages of EEG and fMRI and the fact that they can be acquired simultaneously (Laufs et al., 2008) make the modalities attractive candidates to be combined, or “fused”, with the goal of obtaining functional neuroimaging data with high spatial and temporal resolution.

A number of methods have been proposed for combining M/EEG and fMRI for source localization. They are all based on the assumption that a subset of the neuronal activity is detectable by both modalities (Pflieger and Greenblatt, 2001), thus fMRI data can be used to inform the source localization method about the location of the sources. In terms of ECD methods, it is possible to constrain the location of the dipoles to be within fMRI active areas (George et al., 1995) or to use them as starting points for the optimization algorithm (dipole seeding) (Hillyard et al., 1997). More recently, an ECD method using a Bayesian formulation with an fMRI location prior and Markov Chain Monte Carlo sampling has been proposed (Jun et al., 2008). In the distributed formulation, fMRI active areas can be assigned different weights when using a weighted minimum norm method (Liu et al., 1998), or principal component analysis (PCA) and independent component analysis (ICA) can be used to obtain basis signals which can explain both the EEG and fMRI observations (Brookings et al., 2009). Another method is based on an adaptive Wiener filter where it is assumed that the energy of the electrical activity at every location on the cortex is proportional to the magnitude of the BOLD response at the same location (Liu and He, 2008). It can also be assumed that the cortical activity is sparse, i.e., there are a small but unknown number of active dipoles, which are often located in fMRI active areas. This assumption can be formulated in a Bayesian framework using an ARD prior with different hyperparameters for fMRI active areas (Sato et al., 2004). Another approach is to employ a Bayesian EEG source localization method which can automatically select priors from a set of candidate priors (Phillips et al., 2005; Mattout et al., 2006). When using such a method for EEG/fMRI fusion, location priors can be derived from fMRI activation maps (Mattout et al., 2006). An advantage of this formulation is the possibility to include every fMRI active cluster as a separate location prior (Henson et al., 2010). Doing so enables the method to automatically adjust the relative prior weights by means of model evidence maximization, which is very powerful since it allows the method to emphasize valid fMRI priors (Henson et al., 2010).

All these methods are considered asymmetric since the fMRI data set is analyzed separately and location priors for source localization are derived from the obtained fMRI activation maps. Since some neuronal activity may only be visible in one modality, the introduction of a fixed fMRI based prior can cause an estimation bias which strongly depends on the way the fMRI prior is introduced (Mattout et al., 2006).

Symmetrical EEG/fMRI fusion methods, which analyze the EEG and fMRI jointly and do not use an explicit fMRI prior are believed to be more robust against possible discrepancies between EEG and fMRI. Recently, a method which combines EEG and fMRI symmetrically by means of a common generative model has been proposed (Daunizeau et al., 2007). The method links the modalities by means of a time invariant spatial profile and uses temporal smoothness priors for the cortical currents and the hemodynamic response functions, as well as, a spatial smoothness prior based on a spatial Laplacian, which is also used in the LORETA method (Pascual-Marqui et al., 1994). By using a fully Bayesian formulation and variational Bayesian (VB) inference (Jordan et al., 1999; Attias, 2000) the method can estimate all parameters from the data and does not depend on any user defined parameters. Recently, a method with a similar generative model structure has been proposed (Ou et al., 2010). A key difference is that the generative model is not fully symmetric since the hemodynamic response function for each voxel is treated as an input to the algorithm. Together with a gradient descent based optimization method, this leads to advantages in terms of computational efficiency. Another difference lies in the prior model, the method uses a spatially adaptive Laplacian spatial smoothness prior and does not use temporal smoothness priors.

In this paper, we propose a symmetrical EEG/fMRI fusion method which uses a common generative model and spatially adaptive priors. We extend the method by Daunizeau et al. (2007) in several directions and achieve a higher source localization performance. Specifically, we assume that the spatial profile can contain sharp boundaries between active and inactive regions. We model this by means of a total variation (TV) prior (Rudin et al., 1992) for the spatial profile of cortical activity. In contrast to LORETA-type, i.e., spatial Laplacian, priors (Pascual-Marqui et al., 1994), which are commonly employed in existing methods, the TV prior is spatially adaptive, that is, the degree of spatial smoothness imposed by the prior varies depending on the location. Our generative model can therefore explain abrupt changes in cortical activity, which typically occur at the boundaries of brain regions involved in event related processing, while simultaneously enforcing smoothness in the solution (we refer to Strong and Chan (2003) for a thorough analysis of the properties of the TV prior). A fundamental difference between the spatially adaptive Laplacian prior used in Ou et al. (2010) and the TV prior is that the former can only adapt the degree of spatial smoothness on a per-region basis while the TV prior can do so on a per-vertex basis. The spatially adaptive Laplacian prior therefore depends on an *a priori* segmentation of the cortex and changes in the degree of spatial smoothness can only occur at region boundaries. The TV prior on the other hand does not depend on such a segmentation and can explain changes in the degree of smoothness at arbitrary locations on the cortex. The TV prior was used in Adde et al. (2005) as a deterministic regularization term for the spatial current distribution at a single time instant. The use of the TV prior in this paper in the context of Bayesian inference is fundamentally different and also requires a different discretization. The proposed method also utilizes spatially adaptive temporal priors, allowing for adaptation of the amount of temporal smoothness according to the estimated activity in different brain regions. We use a fully Bayesian formulation and estimate all parameters from the data. Due to the form of the TV prior, it is not possible to directly apply standard variational Bayesian methods to estimate the posterior distribution. Therefore, in order to draw inference, we resort to a majorization method recently proposed in Babacan et al. (2008). The method employs a Gaussian approximation to the TV prior, which renders variational distribution approximation possible, but retains the spatial adaptivity of the TV prior.

We demonstrate the effectiveness of the proposed method using both simulation experiments with synthetic EEG and fMRI data and real data from a multimodal study on face perception. We also include comparisons with existing source localization algorithms and show that the proposed method provides higher performance than existing methods in terms of estimation of the spatio-temporal cortical current distribution. Due to the novel prior model, the proposed method also estimates the hemodynamic response functions more accurately than previous symmetrical fusion methods.

Organization of this paper

This paper consists of 5 sections. In the first section we model the EEG/fMRI fusion problem using the Bayesian paradigm and introduce new realistic prior distributions for the spatio-temporal cortical current distribution and the hemodynamic response functions. The Bayesian inference scheme is introduced in the second section. In the third section we report on experiments with simulated data and in the fourth section we apply the proposed method to real data from a multimodal study on face perception. The paper is discussed and conclusions are drawn in the last section. Appendices with a description of the anatomical parceling, a definition of the signal to noise ratio, an explanation of the quality metrics used, and a detailed derivation of the calculated posterior distributions using the variational framework complete the paper.

Notation

We use the following notation throughout this paper: $\mathbf{A}_{ij}$ and $\mathbf{A}_{i,j}$ denote the element at the i-th row and j-th column of matrix $\mathbf{A}$, while the i-th element of a vector $\mathbf{a}$ is denoted as a_i. $\mathbf{A}_{i\cdot}$ denotes a row vector containing the elements of the i-th row of $\mathbf{A}$, while $\mathbf{A}_{\cdot i}$ is a column vector containing the elements of the i-th column of $\mathbf{A}$. The operator $\mathrm{diag}(\mathbf{A})$ extracts the main diagonal of $\mathbf{A}$ as a column vector, whereas $\mathrm{Diag}(\mathbf{a})$ is a diagonal matrix with $\mathbf{a}$ as its diagonal. The operator $\mathrm{vec}(\mathbf{A})$ vectorizes $\mathbf{A}$ by stacking its columns, $\mathrm{tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$, and ⊗ denotes the Kronecker product.

Hierarchical Bayesian modeling

In this section we define the hierarchical generative model which forms the basis of the proposed method. In the first part we model the process which gives rise to the observed EEG and fMRI data when the current distribution on the cortex and the hemodynamic response function at every location are known. This constitutes the observation model which corresponds to the lowest level of the hierarchical model. In the second part we describe the spatio-temporal decomposition, which divides the cortex into a number of temporally coherent regions and establishes a connection between EEG and fMRI by means of an unknown time invariant spatial profile. We proceed by describing the spatio-temporal prior model, where we introduce the TV spatial prior, as well as, temporal priors which model varying degrees of temporal smoothness across the surface of the cortex. Following a fully Bayesian formulation, prior distributions for all hyperparameters of the model are defined next. At the end of this section, we combine the introduced probability density functions (pdf) to obtain a joint pdf over the observed data and all parameters of the model, which will enable us to obtain the Bayesian inference procedure defined in the next section.

Observation model

In the following we assume that the data is only related to a single event type. For EEG this means that the raw data is averaged over trials for the same event type in order to obtain event related potentials (ERPs) and for fMRI the event onset times for a single event type are used.

Using the distributed source framework (Hämäläinen et al., 1993) the EEG data is modeled as

$$\mathbf{M} = \mathbf{L}\mathbf{S} + \eta_1, \tag{1}$$

where $\mathbf{M}$ is an $m \times t_1$ matrix containing the EEG recordings with duration t_1 obtained from m electrodes placed on the scalp, $\mathbf{S}$ is an unknown $n \times t_1$ matrix representing the responses of n normal-oriented current dipoles distributed on the cortical surface, i.e., a spatio-temporal cortical current distribution, $\mathbf{L}$ is a known $m \times n$ forward operator, also known as lead-field matrix, which can be calculated from the head geometry and tissue conductivities, and η_1 is an $m \times t_1$ matrix representing noise.

We model the noise η_1 for EEG as zero-mean, independent and identically distributed (i.i.d.) Gaussian, resulting in

$$\mathrm{p}(\mathbf{M}|\mathbf{S},\alpha_1) = \prod_{i=1}^{t_1} \mathcal{N}\left(\mathbf{M}_{\cdot i}|\mathbf{L}\mathbf{S}_{\cdot i}, \alpha_1^{-1}\mathbf{I}_m\right), \tag{2}$$

where α_1 is the hyperparameter corresponding to the EEG noise precision.

In order to model the fMRI observations it is assumed that there is a linear relationship between the stimulus and the BOLD response, which leads to the following observation model (Marrelec et al., 2002)

$$\mathbf{Y} = \mathbf{B}\mathbf{H} + \eta_2, \tag{3}$$

where $\mathbf{Y}$ is the $t_2 \times n$ matrix containing the fMRI measurements at n voxels on the cortical surface (we assume here that the locations of the voxels coincide with the locations of the EEG current dipoles), $\mathbf{H}$ is an unknown $k \times n$ matrix representing the hemodynamic response function (HRF) of length k for each voxel, and η_2 is the $t_2 \times n$ matrix with additive noise. The $t_2 \times k$ matrix $\mathbf{B}$ is different from the design matrix in classical fMRI analysis (Friston et al., 1995). The matrix used here implements a convolution and is given by

$$\mathbf{B} = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{t_2-k+1} & x_{t_2-k} & \cdots & x_{t_2-2k+1} \\ 0 & x_{t_2-k+1} & \cdots & x_{t_2-2k+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{t_2-k+1} \end{bmatrix}, \tag{4}$$

where the experimental time course $(x_i)_{1 \le i \le t_2-k+1}$ is a discrete time series in which the i-th element encodes an event onset during the i-th fMRI acquisition, i.e., the time series is all zero except at indices corresponding to event onsets where we use $x_i = 1$ to encode the onset. From Eq. (3) and the structure of $\mathbf{B}$ in Eq. (4) it can be seen that the acquired fMRI time series of the j-th voxel is modeled as a convolution of the HRF with the experimental time course $\mathbf{x}$ plus additive noise, i.e.,

$$\mathbf{Y}_{\cdot j} = \mathbf{x} * \mathbf{H}_{\cdot j} + (\eta_2)_{\cdot j}, \tag{5}$$

where * denotes the (discrete) convolution operator.

For the fMRI noise we also assume that the noise is zero-mean, i.i.d. Gaussian, resulting in

$$\mathrm{p}(\mathbf{Y}|\mathbf{H},\alpha_2) = \prod_{i=1}^{n} \mathcal{N}\left(\mathbf{Y}_{\cdot i}|\mathbf{BH}_{\cdot i},\alpha_2^{-1}\mathbf{I}_{t_2}\right), \tag{6}$$

where α_2 is the hyperparameter corresponding to the fMRI noise precision.

Spatio-temporal decomposition model

In this section we introduce the spatio-temporal decomposition model, which allows us to link EEG and fMRI by means of a common time invariant spatial profile. We adopt the model proposed in Daunizeau et al. (2007) as it provides an elegant way to combine EEG and fMRI. The model utilizes a hierarchical description of the cortical current distribution and the hemodynamic response functions. In order to obtain the hierarchical description, it is assumed that the cortical activity can be described by a set of regions where the responses within a region have similar temporal characteristics, i.e., the responses within a region are temporally coherent. In order to introduce the spatio-temporal decomposition, let us first define a fixed segmentation of the cortex into q regions, or parcels, which we encode using a fixed $n\times q$ segmentation matrix $\mathbf{C}$ defined as

$$\mathbf{C}_{ij} = \begin{cases} 1 & \text{if } i\text{–th vertex is in the } j\text{–th parcel}, \\ 0 & \text{otherwise}. \end{cases} \tag{7}$$

In this work the matrix $\mathbf{C}$ is obtained by a segmentation procedure which uses a region growing algorithm; the procedure is described in Appendix A. However, we note that the segmentation procedure itself is not an integral part of the proposed EEG/fMRI fusion method. By assuming that the electrical responses within each region have the same shape with different scales, the coherency assumption for EEG is formalized by

$$\mathbf{S} = \mathrm{Diag}\left(\mathbf{w}^{\mathrm{EEG}}\right)\mathbf{CX} + \rho_1, \tag{8}$$

where $\mathbf{w}^{\mathrm{EEG}}$ is a $n\times 1$ vector representing the unknown spatial profile of the cortical currents, $\mathbf{X}$ is a $q\times t_1$ matrix with the unknown temporal shape of the currents for each region, and ρ_1 is a $n\times t_1$ matrix representing residual activity which cannot be explained by the model. From Eqs. (7) and (8) it can be seen that if the i-th dipole lies within the j-th parcel, the current waveform of the dipole is modeled as the waveform of the j-th parcel $\mathbf{X}_{j\cdot}$ scaled by the scaling variable for the i-dipole w_i^{EEG}, i.e.,

$$\mathbf{S}_{i\cdot} = w_i^{\mathrm{EEG}}\mathbf{X}_{j\cdot} + (\rho_1)_{i\cdot}. \tag{9}$$

We assume that all the residuals in ρ_1 are zero-mean, i.i.d. Gaussian distributed and obtain the following hierarchical prior for the cortical currents

$$\mathrm{p}\left(\mathbf{S}|\mathbf{X},\mathbf{w}^{\mathrm{EEG}},\epsilon_1\right) = \prod_{i=1}^{t_1}\mathcal{N}\left(\mathbf{S}_{\cdot i}|\mathrm{Diag}\left(\mathbf{w}^{\mathrm{EEG}}\right)\mathbf{CX}_{\cdot i},\epsilon_1^{-1}\mathbf{I}_n\right). \tag{10}$$

Utilizing the same coherency assumption for the HRFs leads to

$$\mathbf{H}^T = \mathrm{Diag}\left(\mathbf{w}^{\mathrm{fMRI}}\right)\mathbf{CZ} + \rho_2, \tag{11}$$

where $\mathbf{Z}$ is a $q\times k$ matrix containing the unknown HRFs of the parcels, $\mathbf{w}^{\mathrm{fMRI}}$ is a $n\times 1$ vector describing the spatial profile, and ρ_2 is a $n\times k$ matrix representing the modeling residual. Note that we use $\mathbf{H}^T$ instead of $\mathbf{H}$ in Eq. (11) since $\mathbf{H}^T$ and $\mathbf{S}$ have the same spatio-temporal structure, i.e., the rows correspond to waveforms at different locations on the cortex. Therefore, by using $\mathbf{H}^T$ in Eq. (11) the equation has the same form as Eq. (8).

As for EEG, we assume that ρ_2 is zero-mean, i.i.d., Gaussian and obtain the following hierarchical prior for the HRFs

$$\mathrm{p}\left(\mathbf{H}^T|\mathbf{Z},\mathbf{w}^{\mathrm{fMRI}},\epsilon_2\right) = \prod_{i=1}^{k}\mathcal{N}\left(\left(\mathbf{H}^T\right)_{\cdot i}|\mathrm{Diag}\left(\mathbf{w}^{\mathrm{fMRI}}\right)\mathbf{CZ}_{\cdot i},\epsilon_2^{-1}\mathbf{I}_n\right). \tag{12}$$

In order to establish a connection between the imaging modalities, a common spatial profile is assumed, i.e.,

$$\mathbf{w}^{\mathrm{EEG}} = \mathbf{w}^{\mathrm{fMRI}} = \mathbf{w}. \tag{13}$$

Note how the temporal characteristics of EEG and fMRI are modeled by $\mathbf{X}$ and $\mathbf{Z}$, respectively, while the time invariant spatial profile $\mathbf{w}$ is responsible for the scale. Therefore, the hierarchical generative model represents a spatio-temporal decomposition and no assumptions are made about the relationship between the temporal shapes of the HRFs and cortical currents. The spatio-temporal decomposition is illustrated in Fig. 1 where the cortical currents and HRFs are shown for two parcels.

Spatial prior model

It is widely known that event related processing in the brain occurs in a number of specialized brain regions. Based on this, we assume that the spatial profile $\mathbf{w}$ contains sharp boundaries between active and inactive regions. In this work, this *a priori* knowledge is incorporated by utilizing a total variation (TV) prior, given by

$$\mathrm{p}(\mathbf{w}|\gamma) = \frac{1}{Z(\gamma)}\exp(-\gamma\mathrm{TV}(\mathbf{w})), \tag{14}$$

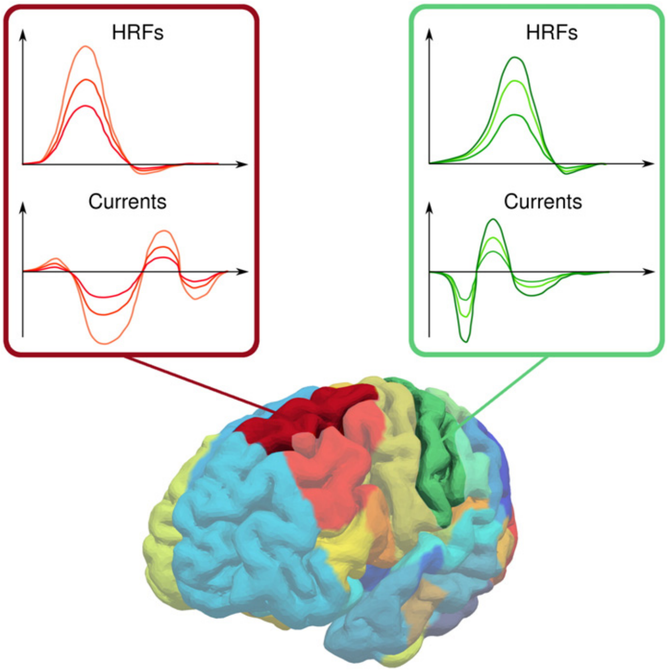


Fig. 1. Illustration of the spatio-temporal decomposition model. The cortical currents and the HRFs within a parcel are assumed to be temporally coherent, i.e., the temporal shape is the same but with different scales, which are modeled by the time invariant spatial profile **w**. The spatial profile links the EEG and fMRI modalities since w_i controls the scale of the current response as well as the scale of the HRF at the i-th vertex. This is illustrated here for two parcels and three waveforms per parcel; the waveforms belonging to the same vertex are drawn with the same color.

where Z(γ) is the partition function and TV(·) is a discrete version of the total variation integral, which is given by

$$\mathrm{TV}_{\mathrm{integral}}(f) = \int_{\Omega} ||\nabla f(\mathbf{x})|| d\mathbf{x}, \tag{15}$$

where Ω denotes the domain over which $f(\cdot)$ is defined and $||\nabla f(\cdot)||$ denotes the magnitude of the gradient of $f(\cdot)$. The hyperparameter γ is similar to the precision (inverse variance) parameter of a Gaussian prior, i.e., it controls the strength of the prior. As will be shown later, following a fully Bayesian approach γ will be treated as unknown and estimated from the data. Total variation priors have been used with great success in a number of inverse problems, such as image denoising and restoration (Rudin et al., 1992; Babacan et al., 2008). A property of the TV prior is that it promotes piecewise smooth solutions, which matches well with our assumption that the spatial profile contains sharp boundaries between smooth regions. An intuitive explanation for the promotion of piecewise smooth solutions can be obtained by thinking of TV regularization as ℓ_1-norm regularization of the magnitude of the gradient. While regular ℓ_1-norm regularization leads to a sparse solution, i.e., a solution where few entries are non-zero, TV regularization leads to a solution where only few locations have non-zero gradient magnitudes, which corresponds to a piecewise smooth solution.

There are two main difficulties in utilizing a TV-prior on the spatial profile **w**. First, the spatial profile **w** is defined on the folded surface of the cortex, such that the calculation of the gradient is not straightforward as in image processing applications where the image is defined on a rectangular 2-D lattice. The second difficulty is that the partition function Z(γ) in Eq. (14) is intractable. Both these difficulties are addressed below.

We address the first problem by defining the gradient of the spatial profile on a differentiable 2-manifold representing the cortical surface embedded in $\mathbb{R}^3$. In practice, the geometry of the manifold is approximated by a triangular mesh denoted by $M = (V,E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of n vertices, and E denotes the set of edges each connecting a pair of vertices. Let us denote $\nabla_M w_i$ the gradient of **w** at vertex v_i. This gradient $\nabla_M w_i$ is the result of discretizing the gradient on a 2-manifold, i.e., the gradient is in the tangent space of M at v_i, which is a Euclidean space in $\mathbb{R}^2$ orthogonal to the surface normal vector at v_i. As the surface normal vector at a vertex we utilize the angle-weighted average of the surface normal vectors of the adjacent triangles (Thürmer and Wuthrich, 1998). In order to calculate the gradient, we project the neighboring vertices $v_j \forall j \in \mathcal{N}_i$ onto the tangent space at v_i, where $\mathcal{N}_i$ denotes the ordered set of neighborhood vertex indices defined as $\mathcal{N}_i = (j | (v_i\ v_j) \in E)$. By doing so we obtain for every neighbor a vector $\mathbf{e}_{ij}$ in $\mathbb{R}^2$ which points from vertex v_i to the projected location of v_j, as depicted in Fig. 2. To calculate the gradient, note that the gradient can be used to obtain a first order approximation, i.e.,

$$w_i + \mathbf{e}_{ij}^T \nabla_M w_i = w_j + r \quad \forall j \in \mathcal{N}_i, \tag{16}$$

where r denotes the residual error. By using all neighbors and rewriting Eq. (16) in matrix form we obtain

$$\mathbf{r} = \underbrace{\begin{bmatrix} \mathbf{e}_{i\mathcal{N}_i(1)}^T \\ \mathbf{e}_{i\mathcal{N}_i(2)}^T \\ \vdots \\ \mathbf{e}_{i\mathcal{N}_i(|\mathcal{N}_i|)}^T \end{bmatrix}}_{\mathbf{E}_i} \nabla_M w_i - \underbrace{\begin{bmatrix} w_{\mathcal{N}_i(1)} - w_i \\ w_{\mathcal{N}_i(2)} - w_i \\ \vdots \\ w_{\mathcal{N}_i(|\mathcal{N}_i|)} - w_i \end{bmatrix}}_{\mathbf{d}_i}, \tag{17}$$

which enables us to estimate the gradient by minimizing the residual $\|\mathbf{r}\|_2$, resulting in

$$\nabla_M w_i = \left(\mathbf{E}_i^T \mathbf{E}_i\right)^{-1} \mathbf{E}_i^T \mathbf{d}_i = \mathbf{G}_i \mathbf{d}_i. \tag{18}$$

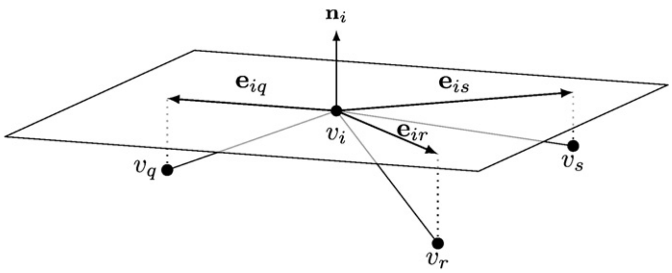


Fig. 2. Illustration of the tangent plane at vertex v_i, which is assumed to have three neighbors $\mathcal{N}_i = (q, r, s)$. The tangent plane is a Euclidean space in $\mathbb{R}^2$ oriented orthogonal to the vertex normal $\mathbf{n}_i$. By projecting the neighboring vertices $\{v_q, v_r, v_s\}$ onto the tangent plane the vectors $\mathbf{e}_{iq}$, $\mathbf{e}_{ir}$, and $\mathbf{e}_{is}$ in $\mathbb{R}^2$ are obtained. The vectors are utilized for calculating the gradient operator matrix at vertex v_i.

Note that since the $2 \times |\mathcal{N}_i|$ gradient matrix $\mathbf{G}_i$ for vertex v_i solely depends on the geometry of the mesh, the gradient matrices for all vertices of the mesh have to be computed only once.

We also note that

$$\mathbf{d}_i = \begin{bmatrix} w_{\mathcal{N}_i(1)} - w_i \\ w_{\mathcal{N}_i(2)} - w_i \\ \vdots \\ w_{\mathcal{N}_i(|\mathcal{N}_i|)} - w_i \end{bmatrix} = \Delta_i \mathbf{w} \tag{19}$$

where Δ_i is a $|\mathcal{N}_i| \times n$ matrix whose j-th row consists of zeros except at the columns i and $\mathcal{N}_i(j)$ where it has the values -1 and 1, respectively.

Finally, the discrete version of the total variation integral in Eq. (14) can be expressed as

$$\mathrm{TV}(\mathbf{w}) = \sum_{i=1}^{n} ||\nabla_M w_i||_2 = \sum_{i=1}^{n} \sqrt{\mathbf{w}^T \Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \mathbf{w}}. \tag{20}$$

A second difficulty arising from the use of a TV prior is that the partition function Z(γ) in Eq. (14) has to be calculated as

$$\mathrm{Z}(\gamma) = \int \exp(-\gamma \mathrm{TV}(\mathbf{w})) d\mathbf{w}, \tag{21}$$

which is intractable since the integral cannot be calculated analytically. Note that we cannot resort to numerical methods, such as Monte Carlo integration, to calculate the partition function as it would require drawing samples from p(**w**|γ) and there is no known method for this task. To address this difficulty, we use the following method to approximate the partition function. We can express the gradient at the i-th vertex as $\mathbf{g} = [g_1 g_2]^T = \mathbf{G}_i \Delta_i \mathbf{w}$ and thus $\mathbf{g}^T \mathbf{g} = g_1^2 + g_2^2$. Using this we can calculate the partition function for a single vertex as follows

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\gamma \sqrt{g_1^2 + g_2^2}\right) dg_1 dg_2 = \frac{2\pi}{\gamma^2}. \tag{22}$$

By combining the partition functions of all n vertices of the mesh we use this to approximate p(**w**|γ) in Eq. (14) as

$$\mathrm{p}(\mathbf{w}|\gamma) = c\gamma^{\varphi n} \exp(-\gamma \mathrm{TV}(\mathbf{w})), \tag{23}$$

where c is a constant and φ is a parameter with a value of $\varphi = 2.0$ if the gradient at every vertex is assumed to be independent from the gradients at all other vertices. Due to the dependency between the gradient values, we empirically found that using $\varphi = 1.0$ improves the performance of the algorithm and we therefore used this value throughout the rest of this paper.

Temporal prior model

We also make the assumption that the HRFs and the cortical currents are smooth in the temporal dimension. This assumption can be expressed by a Gaussian prior which penalizes the second order temporal derivative; a prior of this form was also used in Marrelec et al. (2002) and Daunizeau et al. (2007). In contrast to previous work, we assume that the degree of temporal smoothness varies across the surface of the cortex. We model this by utilizing a separate Gaussian prior for every parcel, i.e., for the temporal shapes of the cortical currents we use

$$p(\mathbf{X}|\beta_1) \propto \prod_{i=1}^{q} \exp\left(-\frac{(\beta_1)_i}{2}(\mathbf{X}_{i\cdot})\mathbf{T}_1^T\mathbf{T}_1(\mathbf{X}_{i\cdot})^T\right), \tag{24}$$

where $\mathbf{T}_1$ is a $t_1 \times t_1$ matrix given by

$$(\mathbf{T}_1)_{ij} = \begin{cases} -2 & \text{if} \quad i = j, \\ 1 & \text{if} \quad j = i \pm 1, \\ 0 & \text{otherwise}, \end{cases} \tag{25}$$

and β_1 is a $q \times 1$ vector with per-parcel precision hyperparameters, each controlling the smoothness and scale of the cortical current waveform of a parcel. The use of separate hyperparameters allows for spatially adaptive temporal smoothness of the cortical currents, i.e., the model can reduce the degree of temporal smoothness in active regions while enforcing a higher degree of smoothness in inactive regions.

For the temporal shape of the hemodynamic response functions we use

$$p(\mathbf{Z}|\beta_2) \propto \prod_{i=1}^{q} \exp\left(-\frac{(\beta_2)_i}{2}(\mathbf{Z}_{i\cdot})\mathbf{T}_2^T\mathbf{T}_2(\mathbf{Z}_{i\cdot})^T\right), \tag{26}$$

where $\mathbf{T}_2$ is a $k \times k$ matrix that is defined analogously to $\mathbf{T}_1$ and β_2 is a $q \times 1$ vector with per-parcel precision hyperparameters. As with the cortical currents, the use of separate hyperparameters allows for spatially adaptive temporal smoothness of the HRF.

Hyperparameter prior model

Following the Bayesian approach we proceed by defining priors for all hyperparameters of the model. In order to obtain priors for the EEG and fMRI noise precisions, we obtain pre-stimulus data segments $\mathbf{M}^0$ for EEG and $\mathbf{Y}^0$ for fMRI containing only noise with sizes $m \times t_1^0$ and $t_2^0 \times n$, respectively. From the Gaussian noise assumption it follows that $p\left(\alpha_1|\mathbf{M}^0\right)$ and $p\left(\alpha_2|\mathbf{Y}^0\right)$ are gamma distributed (Daunizeau et al., 2007), which motivates the use of the following prior distribution for the EEG noise precision hyperparameter

$$\begin{aligned} p(\alpha_1) &= p\left(\alpha_1|\mathbf{M}^0\right) = \Gamma\left(\alpha_1|a^0_{\alpha_1}, b^0_{\alpha_1}\right), \\ a^0_{\alpha_1} &= \frac{mt_1^0}{2}, \quad b^0_{\alpha_1} = \frac{\mathrm{tr}\left(\mathbf{M}^{0T}\mathrm{M}^0\right)}{2}. \end{aligned} \tag{27}$$

The gamma distribution is defined as

$$\Gamma(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx), \tag{28}$$

where $a > 0$ and $b > 0$ are the shape and inverse scale parameters, respectively. Similarly, we use the following prior distribution for the fMRI noise precision hyperparameter

$$\begin{aligned} p(\alpha_2) &= p\left(\alpha_2|\mathbf{Y}^0\right) = \Gamma\left(\alpha_2|a^0_{\alpha_2}, b^0_{\alpha_2}\right), \\ a^0_{\alpha_2} &= \frac{nt_2^0}{2}, \quad b^0_{\alpha_2} = \frac{\mathrm{tr}\left(\mathbf{Y}^{0T}\mathrm{Y}^0\right)}{2}. \end{aligned} \tag{29}$$

Note that the prior distributions become more sharply peaked as the lengths of the pre-stimulus segments increase. Longer pre-stimulus segments cause the fusion algorithm to rely more on the initial noise estimates, i.e., the noise estimated by the algorithm becomes almost entirely decided by the initial estimates. On the other hand, as the length of the pre-stimulus segments goes towards zero, the prior distributions become flat and the noise precision is estimated solely by the fusion algorithm.

For the precision parameter vectors β_1 and β_2, which control the per-parcel temporal smoothness and scale of the cortical currents and hemodynamic response functions, respectively, we use a hyperparameter prior model which allows us to control the degree of spatial adaptivity. In order to do so, we use gamma priors as follows

$$p(\beta_1|\delta_1) = \prod_{i=1}^{q} \Gamma\left((\beta_1)_i|a^0_{\beta_1}, \delta_1\right), \tag{30}$$

$$p(\beta_2|\delta_2) = \prod_{i=1}^{q} \Gamma\left((\beta_2)_i|a^0_{\beta_2}, \delta_2\right), \tag{31}$$

where $a^0_{\beta_1}$ and $a^0_{\beta_2}$ are fixed shape parameters and the unknown inverse scale parameters are denoted by δ_1 and δ_2. The use of fixed shape parameters allows us to control the degree of spatial adaptivity. As will become clear after the derivation of the approximate posterior distribution in the next section, by using a value close to zero for $a^0_{\beta_1}$ the posterior distributions of $(\beta_1)_i, \ldots, (\beta_1)_q$ can be drastically different. Hence, the model is fully spatially adaptive. On the other hand, when $a^0_{\beta_1}$ is very large, all posterior distributions will be almost identical and the prior model is not spatially adaptive, which is similar to the temporal prior model in Daunizeau et al. (2007). We empirically find that the proposed method performs best when the degree of spatial adaptivity for the EEG side is limited by using $a^0_{\beta_1} = 100$ while using a higher degree of spatial adaptivity for the fMRI side with $a^0_{\beta_2} = 10^{-3}$. These values are used throughout the rest of this paper. We note here that the proposed method is not very sensitive to the exact values of the shape parameters, i.e., a value in the range $10, \ldots, 200$ works well for $a^0_{\beta_1}$ while any value close to 0 works well for $a^0_{\beta_2}$.

We make no assumptions about the remaining hyperparameters and consequently use noninformative Jeffreys priors given by

$$p(\theta_i) = \Gamma(\theta_i|0,0) \propto (\theta_i)^{-1} \quad \forall \quad \theta_i \in \theta, \tag{32}$$

where $\theta = \{\delta_1, \delta_2, \epsilon_1, \epsilon_2, \gamma\}$, to define

$$p(\theta) = \prod_{\theta_i \in \theta} p(\theta_i). \tag{33}$$

We note here that an important reason for selecting gamma distributions as priors for the hyperparameters is that the gamma distribution is the conjugate prior for the precision of a Gaussian distribution, as well as, for the inverse scale parameter of the gamma distribution, which simplifies the Bayesian inference since the posterior distributions of the hyperparameters will also be gamma distributions. As will be shown in the next section, in order to draw inference we employ a quadratic approximation to the energy of the TV prior in the form of a Gaussian distribution and consequently the conjugate prior for γ is a gamma distribution.

Global modeling

By combining all distributions introduced above, we obtain the joint probability density function as follows

$$\begin{aligned} p(\Theta, \mathbf{M}, \mathbf{Y}) = {} & p(\mathbf{M}|\mathbf{S}, \alpha_1)p(\mathbf{S}|\mathbf{X}, \mathbf{w}, \epsilon_1)p(\mathbf{X}|\beta_1) \\ & \times p(\mathbf{Y}|\mathbf{H}, \alpha_2)p(\mathbf{H}|\mathbf{Z}, \mathbf{w}, \epsilon_2)p(\mathbf{Z}|\beta_2)p(\mathbf{w}|\gamma) \\ & \times p(\alpha_1)p(\alpha_2)p(\beta_1|\delta_1)p(\beta_2|\delta_2)p(\theta), \end{aligned} \tag{34}$$

where $\Theta = \{\mathbf{S}, \mathbf{H}, \mathbf{w}, \mathbf{X}, \mathbf{Z}, \alpha_1, \alpha_2, \beta_1, \beta_2,\} \cup \theta$ is the set of all unknowns. The dependencies between the variables in the joint pdf are illustrated as a directed acyclic graphical model in Fig. 3.

The joint pdf allows us to derive a fusion algorithm using Bayesian inference, which is described in the next section.

Bayesian inference

Inference is based on the posterior distribution

$$\mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y}) = \frac{\mathrm{p}(\Theta, \mathbf{M}, \mathbf{Y})}{\mathrm{p}(\mathbf{M}, \mathbf{Y})}. \tag{35}$$

However, the posterior $\mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y})$ is intractable since

$$\mathrm{p}(\mathbf{M}, \mathbf{Y}) = \int \mathrm{p}(\mathbf{M}, \mathbf{Y}, \Theta) d\Theta \tag{36}$$

cannot be calculated analytically. Therefore, we utilize an approximation to the posterior. In this work, we employ the Variational Bayesian (VB) method using the mean field approximation (Jordan et al., 1999; Attias, 2000), i.e., we approximate the true posterior by a distribution which factorizes over the nodes of the graphical model

$$\begin{aligned} \mathrm{q}(\Theta) = {} & \mathrm{q}(\mathbf{S})\mathrm{q}(\mathbf{H})\mathrm{q}(\mathbf{X})\mathrm{q}(\mathbf{Z})\mathrm{q}(\mathbf{w})\mathrm{q}(\alpha_1)\mathrm{q}(\alpha_2) \\ & \times \left(\prod_{i=1}^{q} \mathrm{q}((\beta_1)_i) \right) \left(\prod_{i=1}^{q} \mathrm{q}((\beta_2)_i) \right) \\ & \times \mathrm{q}(\delta_1)\mathrm{q}(\delta_2)\mathrm{q}(\epsilon_1)\mathrm{q}(\epsilon_2)\mathrm{q}(\gamma). \end{aligned} \tag{37}$$

As stated in Jaakkola and Jordan (1998), mean field theory (Parisi, 1998) provides an intuitive explanation of the mean field approximation. That is, in a dense graph each node is influenced by many other nodes such that the influence from each other node is weak and the total influence is approximately additive. Hence, each node can be characterized by its mean value, which is unknown and related to the mean values of all other nodes. The task then becomes finding the relation between the mean values and designing an algorithm which can find a consistent assignment of mean values. This is exactly what we will do in the following. First, we will find a distribution for each node in the graphical model shown in Fig. 3. The distributions describe the relation to all other nodes in the model and allow us to obtain an inference algorithm in which we iteratively update the distribution of each node leading to a consistent assignment of distributions.

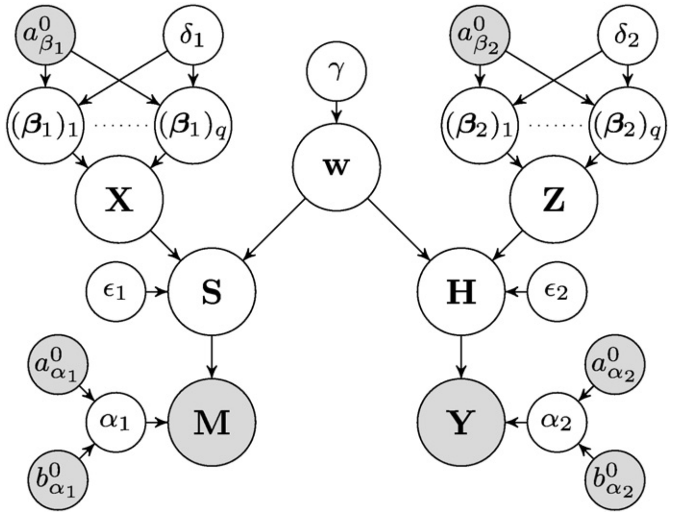


Fig. 3. Directed acyclic graphical model describing the joint pdf (gray: known, white: unknown).

The posterior approximation $\mathrm{q}(\Theta)$ is found by performing a variational minimization of the Kullback–Leibler (KL) divergence, which is given by

$$\begin{aligned} & C_{KL}(\mathrm{q}(\Theta) \| \mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y})) \\ & \quad = \int \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y})} \right) d\Theta \\ & \quad = \int \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{M}, \mathbf{Y})} \right) d\Theta + \mathrm{const} \\ & \quad = \mathcal{K}(\mathrm{q}(\Theta)) + \mathrm{const}, \end{aligned} \tag{38}$$

and is non-negative and equal to zero only if $\mathrm{q}(\Theta) = \mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y})$. In variational Bayesian analysis, the optimal $\mathrm{q}(\Theta)$ is found by

$$\begin{aligned} \mathrm{q}(\Theta) = {} & \underset{\mathrm{q}(\Theta)}{\arg\min}\, C_{KL}(\mathrm{q}(\Theta) \| \mathrm{p}(\Theta|\mathbf{M}, \mathbf{Y})) \\ = {} & \underset{\mathrm{q}(\Theta)}{\arg\min}\, \mathcal{K}(\mathrm{q}(\Theta)). \end{aligned} \tag{39}$$

Using a standard result from variational Bayesian analysis (Bishop, 2006), for each variable the distribution which minimizes Eq. (38) is given by

$$\mathrm{q}(\Theta_i) \propto \exp\left(\mathbb{E}_{\Theta \backslash \Theta_i}[\ln \mathrm{p}(\Theta, \mathbf{M}, \mathbf{Y})] \right), \tag{40}$$

where $\mathbb{E}_{\Theta \backslash \Theta_i}[\cdot]$ denotes the expectation with respect to all variables except the variable of interest.

Unfortunately, the form of the TV prior prevents us from calculating the expectation in Eq. (40) and thus from finding an analytical form of $\mathrm{q}(\Theta)$. Therefore, we resort to a majorization method which approximates $\mathcal{K}(\mathrm{q}(\Theta))$ by upper-bounding functionals which render the calculation of the expectation tractable (Babacan et al., 2008). First, let us consider the geometric–arithmetic mean inequality (Hardy et al., 1988) which states that for two positive numbers $a \geq 0$ and $b > 0$

$$\sqrt{ab} \leq \frac{a+b}{2} \Rightarrow \sqrt{a} \leq \frac{a+b}{2\sqrt{b}}. \tag{41}$$

We proceed by defining for $\mathbf{w}$, γ, and an $n \times 1$ vector $\mathbf{u} \in (\mathbb{R}^+)^n$, the following functional:

$$\mathrm{F}(\mathbf{w}, \mathbf{u}, \gamma) = c\gamma^{\varphi n} \exp\left(-\frac{\gamma}{2} \sum_{i=1}^{n} \frac{\mathbf{w}^T \Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \mathbf{w} + u_i}{\sqrt{u_i}} \right). \tag{42}$$

Using inequality Eq. (41) in Eq. (23) with $a = \mathbf{w}^T \Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \mathbf{w}$ and $b = u_i$, and comparing with Eq. (42), we obtain

$$\mathrm{p}(\mathbf{w}|\gamma) \geq \mathrm{F}(\mathbf{w}, \mathbf{u}, \gamma). \tag{43}$$

The auxiliary variable $\mathbf{u}$ is related to the spatial smoothness in $\mathbf{w}$ and needs to be updated by the inference algorithm, as will be shown later. Using Eq. (43) in Eq. (34), we obtain a lower bound of the joint probability density function, i.e.,

$$\begin{aligned} \mathrm{p}(\Theta, \mathbf{M}, \mathbf{Y}) \geq {} & \mathrm{p}(\mathbf{M}|\mathbf{S}, \alpha_1)\mathrm{p}(\mathbf{S}|\mathbf{X}, \mathbf{w}, \epsilon_1)\mathrm{p}(\mathbf{X}|\beta_1) \\ & \times \mathrm{p}(\mathbf{Y}|\mathbf{H}, \alpha_2)\mathrm{p}(\mathbf{H}|\mathbf{Z}, \mathbf{w}, \epsilon_2)\mathrm{p}(\mathbf{Z}|\beta_2) \\ & \times \mathrm{p}(\alpha_1)\mathrm{p}(\alpha_2)\mathrm{p}(\beta_1|\delta_1)\mathrm{p}(\beta_2|\delta_2) \\ & \times \mathrm{p}(\theta)\mathrm{F}(\mathbf{w}, \mathbf{u}, \gamma) \\ = {} & \mathrm{F}(\Theta, \mathbf{u}, \mathbf{M}, \mathbf{Y}), \end{aligned} \tag{44}$$

which allows us to derive an inference procedure, as will be shown below. It should be noted that the proposed method therefore does not employ the TV prior directly; doing so would not lead to a tractable inference. Instead, the proposed method uses the lower bound $\mathrm{F}(\mathbf{w}, \mathbf{u}, \gamma)$ to the TV prior, which retains many of its desirable characteristics, i.e., the ability to model sharp boundaries, and allows for a tractable inference.

To derive the inference procedure, let us now define

$$\tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u}) = \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{F}(\Theta,\mathbf{u},\mathbf{M},\mathbf{Y})}\right)d\Theta, \tag{45}$$

which is the KL divergence between $\mathrm{q}(\Theta)$ and $\mathrm{F}(\Theta,\mathbf{u},\mathbf{M},\mathbf{Y})$. By using Eqs. (38) and (44), we obtain

$$\mathcal{K}(\mathrm{q}(\Theta)) \le \min_{\mathbf{u}} \tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u}). \tag{46}$$

Therefore we can obtain a sequence of distributions $\{\mathrm{q}(\Theta)\}$ which monotonically decreases $\tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u})$ for a fixed $\mathbf{u}$. From Eq. (46) it can be seen that this leads to a monotonically decreasing upper bound to $C_{KL}(\mathrm{q}(\Theta)\|\mathrm{p}(\Theta|\mathbf{M},\mathbf{Y}))$ and therefore leads to an approximation of the true posterior distribution. Moreover, we can minimize $\tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u})$ with respect to $\mathbf{u}$ for each distribution $\mathrm{q}(\Theta)$, which tightens the upper bound to the KL divergence and thus leads to a more accurate distribution approximation. The two interleaved minimization steps naturally lead to the iterative distribution estimation algorithm. During each iteration the algorithm first minimizes the functional $\tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u})$ with respect to $\mathrm{q}(\Theta)$; the distribution approximation which minimizes this functional has the same form as in standard VB analysis (see Eq. (40)) and the distribution approximation of the node $\Theta_i \in \Theta$ is given by

$$\mathrm{q}(\Theta_i) \propto \exp\left(\mathbb{E}_{\Theta\backslash\Theta_i}[\ln \mathrm{F}(\Theta,\mathbf{u},\mathbf{M},\mathbf{Y})]\right). \tag{47}$$

Using Eq. (47) we obtain a distribution for every node of the graphical model. The distributions of the nodes **S**, **H**, **X**, **Z**, and **w** are found to be Gaussian while the hyperparameter distributions are found to be gamma distributions (since conjugate priors were used). The form of the distributions obtained by applying Eq. (47) is given in Table 1 and the corresponding derivations are shown in Appendix D. In order to update the distributions and therefore to minimize $\tilde{\mathcal{K}}(\mathrm{q}(\Theta),\mathbf{u})$ in the first step of the algorithm, the algorithm updates the parameters of the distributions in Table 1 using the most recently updated parameters, i.e., either from the previous or from the current iteration. The distributions are updated in the following order: $\mathrm{q}(\mathbf{S})$, $\mathrm{q}(\alpha_1)$, $\mathrm{q}(\mathbf{X})$, $\mathrm{q}(\epsilon_1)$, $\mathrm{q}((\beta_1)_1)$, …, $\mathrm{q}((\beta_1)_q)$, $\mathrm{q}(\delta_1)$, $\mathrm{q}(\mathbf{H})$, $\mathrm{q}(\alpha_2)$, $\mathrm{q}(\mathbf{Z})$, $\mathrm{q}(\epsilon_2)$, $\mathrm{q}((\beta_2)_1)$, …, $\mathrm{q}((\beta_2)_q)$, $\mathrm{q}(\delta_2)$, and $\mathrm{q}(\mathbf{w})$.

Table 1
Distributions for the nodes of the graphical model obtained using Eq. (47). Derivations are shown in Appendix D, where matrices **Q**, $\mathbf{P}_1$, $\mathbf{P}_2$, and $W(\mathbf{u})$ and the cov(·) operator are defined. The matrix $\mathbf{R}_{(k,q)}$ is a $kq \times kq$ permutation matrix with the property $\mathbf{R}_{(k,q)}\mathrm{vec}(\mathbf{Z}^T) = \mathrm{vec}(\mathbf{Z})$ (the matrix $\mathbf{R}_{(t_1,q)}$ is defined analogously).

Functional form	Parameters
$\mathrm{q}(\mathbf{S}) = \mathcal{N}(\mathrm{vec}(\mathbf{S})\,|\,\mathrm{vec}(\langle\mathbf{S}\rangle), \mathbf{I}_{t_1}\otimes\sum_{\mathbf{S}})$	$\langle\mathbf{S}\rangle = \sum_{\mathbf{S}}(\langle\alpha_1\rangle\mathbf{L}^T\mathbf{M} + \langle\epsilon_1\rangle\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{X}\rangle)$ $\sum_{\mathbf{S}} = \left(\langle\alpha_1\rangle\mathbf{L}^T\mathbf{L} + \langle\epsilon_1\rangle\mathbf{I}_n\right)^{-1}$
$\mathrm{q}(\mathbf{H}) = \mathcal{N}(\mathrm{vec}(\mathbf{H})\,|\,\mathrm{vec}(\langle\mathbf{H}\rangle), \mathbf{I}_n\otimes\sum_{\mathbf{H}})$	$\langle\mathbf{H}\rangle = \sum_{\mathbf{H}}\left(\langle\alpha_2\rangle\mathbf{B}^T\mathbf{Y} + \langle\epsilon_2\rangle\langle\mathbf{Z}\rangle^T\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right)$ $\sum_{\mathbf{H}} = \left(\langle\alpha_2\rangle\mathbf{B}^T\mathbf{B} + \langle\epsilon_2\rangle\mathbf{I}_k\right)^{-1}$
$\mathrm{q}(\mathbf{X}) = \mathcal{N}(\mathrm{vec}(\mathbf{X})\,|\,\mathrm{vec}(\langle\mathbf{X}\rangle), \sum_{\mathbf{X}})$	$\mathrm{vec}(\langle\mathbf{X}\rangle) = \langle\epsilon_1\rangle\sum_{\mathbf{X}}\left(\mathbf{I}_{t_1}\otimes\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right)\mathrm{vec}(\langle\mathbf{S}\rangle)$ $\sum_{\mathbf{X}} = \left(\langle\epsilon_1\rangle(\mathbf{I}_{t_1}\otimes\mathbf{Q}) + \mathbf{R}^T_{(t_1,q)}\left(\mathrm{Diag}(\langle\beta_1\rangle)\otimes\mathbf{T}_1^T\mathbf{T}_1\right)\mathbf{R}_{(t_1,q)}\right)^{-1}$
$\mathrm{q}(\mathbf{Z}) = \mathcal{N}(\mathrm{vec}(\mathbf{Z})\,|\,\mathrm{vec}(\langle\mathbf{Z}\rangle), \sum_{\mathbf{Z}})$	$\mathrm{vec}(\langle\mathbf{Z}\rangle) = \langle\epsilon_2\rangle\sum_{\mathbf{Z}}\left(\mathbf{I}_k\otimes\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right)\mathrm{vec}\left(\langle\mathbf{H}\rangle^T\right)$ $\sum_{\mathbf{Z}} = \left(\langle\epsilon_2\rangle(\mathbf{I}_k\otimes\mathbf{Q}) + \mathbf{R}^T_{(k,q)}\left(\mathrm{Diag}(\langle\beta_2\rangle)\otimes\mathbf{T}_2^T\mathbf{T}_2\right)\mathbf{R}_{(k,q)}\right)^{-1}$
$\mathrm{q}(\mathbf{w}) = \mathcal{N}(\mathbf{w}\,|\,\langle\mathbf{w}\rangle, \sum_{\mathbf{w}})$	$\langle\mathbf{w}\rangle = \sum_{\mathbf{w}}\mathrm{diag}\left(\langle\epsilon_1\rangle\langle\mathbf{S}\rangle\langle\mathbf{X}\rangle^T\mathbf{C}^T + \langle\epsilon_2\rangle\langle\mathbf{H}\rangle^T\langle\mathbf{Z}\rangle^T\mathbf{C}^T\right)$ $\sum_{\mathbf{w}} = (\langle\epsilon_1\rangle\mathbf{P}_1 + \langle\epsilon_2\rangle\mathbf{P}_2 + \langle\gamma\rangle W(\mathbf{u}))^{-1}$
$\mathrm{q}(\alpha_1) = \Gamma(\alpha_1\|a_{\alpha_1}, b_{\alpha_1})$	$a_{\alpha_1} = \frac{mt_1}{2} + a^0_{\alpha_1}$ $b_{\alpha_1} = \frac{1}{2}\,\mathrm{tr}\left((\mathbf{M}-\mathbf{L}\langle\mathbf{S}\rangle)^T(\mathbf{M}-\mathbf{L}\langle\mathbf{S}\rangle)\right) + \frac{t_1}{2}\mathrm{tr}\left(\sum_{\mathbf{S}}\mathbf{L}^T\mathbf{L}\right) + b^0_{\alpha_1}$
$\mathrm{q}(\alpha_2) = \Gamma(\alpha_2\|a_{\alpha_2}, b_{\alpha_2})$	$a_{\alpha_2} = \frac{nt_2}{2} + a^0_{\alpha_2}$ $b_{\alpha_2} = \frac{1}{2}\,\mathrm{tr}\left((\mathbf{Y}-\mathbf{B}\langle\mathbf{H}\rangle)^T(\mathbf{Y}-\mathbf{B}\langle\mathbf{H}\rangle)\right) + \frac{n}{2}\mathrm{tr}\left(\sum_{\mathbf{H}}\mathbf{B}^T\mathbf{B}\right) + b^0_{\alpha_2}$
$\mathrm{q}(\epsilon_1) = \Gamma(\epsilon_1\|a_{\epsilon_1}, b_{\epsilon_1})$	$a_{\epsilon_1} = \frac{t_1 n}{2}$ $b_{\epsilon_1} = \frac{1}{2}\left[\mathrm{tr}\left(\langle\mathbf{S}\rangle^T\langle\mathbf{S}\rangle - 2\langle\mathbf{S}\rangle^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{X}\rangle + \langle\mathbf{X}\rangle^T\mathbf{Q}\langle\mathbf{X}\rangle\right) + t_1\mathrm{tr}(\sum_{\mathbf{S}}) + \mathrm{tr}(\sum_{\mathbf{X}}(\mathbf{I}_{t_1}\otimes\mathbf{Q}))\right]$
$\mathrm{q}(\epsilon_2) = \Gamma(\epsilon_2\|a_{\epsilon_2}, b_{\epsilon_2})$	$a_{\epsilon_2} = \frac{kn}{2}$ $b_{\epsilon_2} = \frac{1}{2}\left[\mathrm{tr}\left(\langle\mathbf{H}\rangle\langle\mathbf{H}\rangle^T - 2\langle\mathbf{H}\rangle\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{Z}\rangle + \langle\mathbf{Z}\rangle^T\mathbf{Q}\langle\mathbf{Z}\rangle\right) + n\mathrm{tr}(\sum_{\mathbf{H}}) + \mathrm{tr}(\sum_{\mathbf{Z}}(\mathbf{I}_k\otimes\mathbf{Q}))\right]$
$\mathrm{q}((\beta_1)_i) = \Gamma((\beta_1)_i\,\|\,(\mathbf{a}_{\beta_1})_i, (\mathbf{b}_{\beta_1})_i)$	$(\mathbf{a}_{\beta_1})_i = \frac{t_1}{2} + a^0_{\beta_1}$ $(\mathbf{b}_{\beta_1})_i = \frac{1}{2}\,\langle\mathbf{X}_{i\cdot}\rangle\mathbf{T}_1^T\mathbf{T}_1\langle\mathbf{X}_{i\cdot}\rangle^T + \frac{1}{2}\mathrm{tr}\left(\mathbf{T}_1^T\mathbf{T}_1\mathrm{cov}\left((\mathbf{X}_{i\cdot})^T\right)\right) + \langle\delta_1\rangle$
$\mathrm{q}((\beta_2)_i) = \Gamma((\beta_2)_i\,\|\,(\mathbf{a}_{\beta_2})_i, (\mathbf{b}_{\beta_2})_i)$	$(\mathbf{a}_{\beta_2})_i = \frac{k}{2} + a^0_{\beta_2}$ $(\mathbf{b}_{\beta_2})_i = \frac{1}{2}\,\langle\mathbf{Z}_{i\cdot}\rangle\mathbf{T}_2^T\mathbf{T}_2\langle\mathbf{Z}_{i\cdot}\rangle^T + \frac{1}{2}\mathrm{tr}\left(\mathbf{T}_2^T\mathbf{T}_2\mathrm{cov}\left((\mathbf{Z}_{i\cdot})^T\right)\right) + \langle\delta_2\rangle$
$\mathrm{q}(\delta_1) = \Gamma(\delta_1\|a_{\delta_1}, b_{\delta_1})$	$a_{\delta_1} = a_{\beta_1}{}^0 q$ $b_{\delta_1} = \sum_{i=1}^{q}\langle(\beta_1)_i\rangle$
$\mathrm{q}(\delta_2) = \Gamma(\delta_2\|a_{\delta_2}, b_{\delta_2})$	$a_{\delta_2} = a_{\beta_2}{}^0 q$ $b_{\delta_2} = \sum_{i=1}^{q}\langle(\beta_2)_i\rangle$
$\mathrm{q}(\gamma) = \Gamma(\gamma\|a_\gamma, b_\gamma)$	$a_\gamma = \varphi n$ $b_\gamma = \sum_{i=1}^{n}\sqrt{u_i}$

After updating $\mathrm{q}(\Theta)$ in the first step of an iteration of the algorithm, the algorithm minimizes the functional $\tilde{\mathcal{K}}(\mathrm{q}(\Theta), \mathbf{u})$ with respect to $\mathbf{u}$ in the second step of an iteration, which is equivalent to

$$\mathbf{u} = \arg\min_{\mathbf{u}} \sum_{i=1}^{n} \frac{\mathbb{E}\left[\mathbf{w}^T \Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \mathbf{w}\right] + u_i}{\sqrt{u_i}}. \quad (48)$$

Since Eq. (48) is a linear combination of n functions where the i-th function is convex with respect to u_i, the minimizer is found by calculating the derivative with respect to u_i and equating to zero, which results in the following update

$$\begin{aligned} u_i &= \mathbb{E}\left[\mathbf{w}^T \Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \mathbf{w}\right] \\ &= \mathrm{tr}\left[\Delta_i^T \mathbf{G}_i^T \mathbf{G}_i \Delta_i \left(\textstyle\sum_{\mathbf{w}} + \langle \mathbf{w} \rangle \langle \mathbf{w} \rangle^T\right)\right], \end{aligned} \quad (49)$$

for $i = 1, \ldots, n$. It is clear from Eq. (49) that the auxiliary vector $\mathbf{u}$ is related to the gradient of the estimated spatial profile $\mathbf{w}$. Moreover, as can be seen from $\mathrm{q}(\mathbf{w})$ (shown in Table 1), the vector $\mathbf{u}$ introduces spatially adaptive smoothing through the matrix $W(\mathbf{u})$ into the estimation process (see Appendix D). This matrix controls the amount of smoothing at each vertex depending on the local variation of the spatial profile.

Computational complexity

To conclude this section we discuss the per-iteration computational complexity of the proposed method. Note that this does not take into account the computational cost of obtaining the parcellation of the cortex and the cost of computing the gradient projection matrices, as these operations only have to be performed once for a given cortical mesh. Excluding these operations from the discussion is also justified by the fact that the time required to perform them is typically shorter than the time required for one iteration of the proposed method. The per-iteration computational complexity of the proposed method is governed by the complexity of the matrix inversions needed to compute the covariance matrices in Table 1. For many applications it is possible to avoid the explicit inversion of matrices by employing efficient linear system solvers, such as the conjugate gradient method. Unfortunately, this is not possible in fully Bayesian methods, such as the one proposed in this work, since the covariance matrices are required for the computation of hyperparameters. By assuming that the inversion of an $N \times N$ matrix has complexity $O(N^3)$ and by taking into account the sizes of the covariance matrices in Table 1, the per-iteration complexity of the proposed method is found to be $O(n^3 + q^3(t_1^3 + k^3))$. From this one can see how the number of parcels q, which is in the range $[1,n]$, affects the computational complexity. Ideally one would like to use a large number of parcels, such that parcels are small and the probability of having multiple sources in the same parcel is low. However, doing so can lead to prohibitively high computational demands and one has to chose $q \ll n$ in order to satisfy the constraints imposed by the computational resources available.

Simulations

In this section we evaluate the proposed method using simulations with synthetic EEG and fMRI data. The use of synthetic data enables us to compare the proposed method and existing methods by means of objective quality metrics.

At the end of this section we evaluate the results and compare the proposed method to several existing methods. Two EEG/fMRI fusion methods are used for the comparison. The first method is the symmetrical BASTERF method (Daunizeau et al., 2007), which is similar to the proposed method but uses a different prior model. The second method is the fMRI weighted minimum norm method (fWMN) (Liu et al., 1998), which can be considered one of the simplest methods for asymmetrical EEG/fMRI fusion. As an additional reference we include several EEG-only source localization methods in the comparison. The MSP method (Friston et al., 2008) is a recently proposed method that uses multiple sparse priors (256 per hemisphere are used here) with an empirical Bayesian modeling and can be considered a state of the art EEG source localization method. We also include two classic EEG source localization methods, namely the LORETA method (Pascual-Marqui et al., 1994), and the minimum norm method (MNE) with Tikhonov noise regularization (Dale and Sereno, 1993).

EEG forward model

The lead field matrix $\mathbf{L}$ used for the simulations was calculated as follows. First, the template cortical mesh included in SPM8 (http://www.fil.ion.ucl.ac.uk/spm) with a total of 8196 vertices was downsampled to $n = 1000$ vertices. While the coarser mesh provides a less accurate geometrical description of the cortex, it significantly reduces the computational requirements. The lead field matrix was then computed using the BEM method from FieldTrip (http://fieldtrip.fcdonders.nl) with standard sensor locations for a 64 channel montage and canonical scalp, outer skull, and inner skull meshes, which are included in SPM8.

Simulated EEG and fMRI data

In order to simulate a range of source configurations and various degrees of agreement between EEG and fMRI a total of 5 different simulation scenarios are used in our evaluation. In the first simulation scenario we use a complex source configuration with more widespread sources, such sources are for example known to occur in children (Friedrich and Friederici, 2004; Sanders et al., 2006). We denote the scenario CPX and use a total of 4 sources, among which 2 are more widespread. All sources are hemodynamically, as well as, electrically active. Due to the complexity of source configuration, it can be expected that EEG/fMRI fusion methods have a significant advantage over EEG-only methods for this scenario. The remaining simulation scenarios use simpler source configurations with only 2 sources and are used to depict situations where some sources can be detectable by either only one modality or both (a similar experiment was presented in Daunizeau et al. (2007)). In practice such situations can for example occur when a source is active for a short time and can be detected by EEG but does not generate a BOLD response strong enough to be detectable by fMRI. On the other hand, it is possible that a source is far from the surface of the scalp, and thus generates a weak EEG signal while having a strong BOLD response. The scenarios are denoted as MM for the scenario with 2 multimodal, i.e., electrically and hemodynamically active, sources, ME for the scenario with one multimodal source and another source that only exhibits electrical activity, MH with one multimodal source and another source that is only hemodynamically active, and EH where one source is electrically active and the other is hemodynamically active. The EH scenario is included for completeness and it should be noted that it fundamentally violates the assumption which motivates fusion of EEG and fMRI, that is, the assumption that a subset of the neuronal activity is detectable by either modality. An overview of the simulation scenarios is given in Table 2. For each scenario, two sources each with a spatial extent of either 8 or 16 vertices are placed at random, non-overlapping locations on the cortical surface. Note that we assume no knowledge about the parcellation used by our algorithm when placing the sources on the cortex. It is therefore possible that the sources overlap parcel boundaries or that multiple sources are within the same parcel.

Table 2
Simulation scenarios used in the empirical evaluation. A multimodal source is denoted as "M" while sources which are only electrically or hemodynamically active are denoted as "E" and "H", respectively. The numbers indicate the spatial extent in vertices of the source, e.g., M(16) denotes a multimodal source with a spatial extent of 16 vertices. The source waveforms of the various sources are depicted in Fig. 4.

Scenario	Source 1	Source 2	Source 3	Source 4
CPX	M(8)	M(8)	M(16)	M(16)
MM	M(8)	M(8)		
ME	M(8)	E(8)		
MH	M(8)	H(8)		
EH	E(8)	H(8)		

Table 3
Summary of simulation parameters.

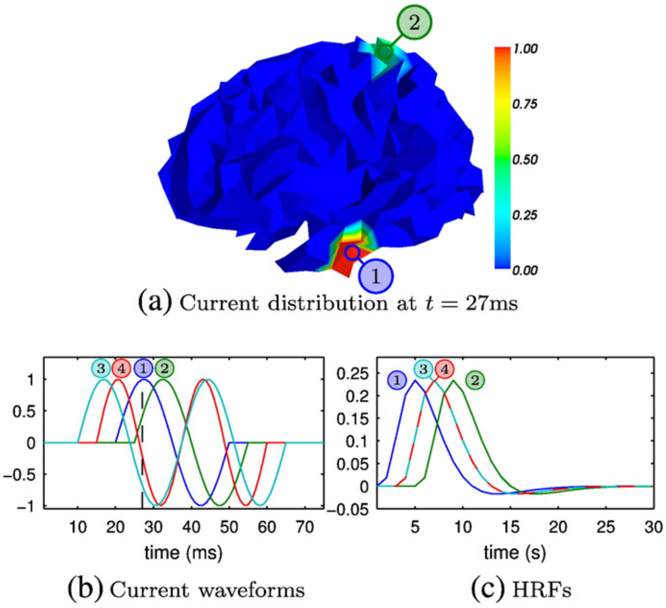
Common			
Size cortical mesh			$n = 1000$
Number of parcels			$q = 32$
EEG		fMRI	
Number of sensors	$m = 64$	Length HRF	$k = 30$
Time points	$t_1 = 75$	Time points	$t_2 = 1000$
Sampling rate	1 kHz	Sampling rate	1 Hz
SNR	15 dB, 20 dB, 25 dB	SNR	5 dB

To simulate source waveforms, we use sinusoids with different starting points and frequencies as the current waveforms of electrically active sources and a shifted canonical HRF from SPM8 with a positive peak at 5 s and a smaller negative peak at 12 s for hemodynamically active sources. The source waveforms of the sources, as well as, an example of the source distribution on the cortex for the MM scenario are illustrated in Fig. 4. The rest of the simulation parameters are as follows. For EEG we use $m = 64$ sensors, $t_1 = 75$ (we assume a sampling rate of 1 kHz), and signal to noise ratios (SNRs) of 15 dB, 20 dB and 25 dB (refer to Appendix B for a definition of the SNR). For fMRI we use $t_2 = 1000$, $k = 30$ with 30 random occurrences of the event of interest, and an SNR of 5 dB. We use $q = 32$ anatomical parcels which are obtained using the procedure described in Appendix A. Note that we use the same parceling for the proposed method and for the BASTERF method. A summary of all parameters used for the simulations is shown in Table 3.

We perform 25 simulations per scenario and SNR configuration for each algorithm. For all algorithms the same random source configurations and noise manifestations are used in order to provide a fair comparison.

Initialization

In order to start the iterative inference procedure we initialize the parameters of the proposed method as follows. For the EEG noise precision we assume that the noise only data window $\mathbf{M}^0$ is one third of the length of $\mathbf{M}$, i.e., 25 columns, and use $a_{\alpha_1} = a^0_{\alpha_1} = 25m/2$, $b_{\alpha_1} = b^0_{\alpha_1} = a_{\alpha_1}\sigma^2_{\mathrm{EEG}}$, where σ^2_{EEG} is the EEG noise variance. Similarly, we use for the fMRI noise precision hyperparameters $a_{\alpha_2} = a^0_{\alpha_2} = 250n/2$, $b_{\alpha_2} = b_{\alpha_2}{}^0 = a_{\alpha_2}\sigma^2_{\mathrm{fMRI}}$. The expectations of the remaining hyperparameters and the vector $\mathbf{u}$ are initialized with small values of 10^{-3}. The variables $\langle \mathbf{Z} \rangle$, $\langle \mathbf{X} \rangle$, and $\langle \mathbf{w} \rangle$ and their covariance matrices are initialized with all zero values, while minimum norm estimates are used for $\langle \mathbf{S} \rangle$ and $\langle \mathbf{H} \rangle$ together with all zero covariance matrices. After the initialization the algorithm is started and the variables are updated in the order given in the previous section. While we do not provide a detailed analysis of the convergence properties of the proposed method, we note here that we find that the method is insensitive to parameter initialization, which agrees with earlier work where the same inference scheme is used (Babacan et al., 2008). For example, the proposed method typically converges to the same solution when it is initialized using the method stated above as when it is initialized with the solution found by the BASTERF method.



(a) Current distribution at $t = 27$ms
(b) Current waveforms
(c) HRFs

Fig. 4. Source configurations used for simulations. The upper panel illustrates an example current distribution of a simulation with the MM scenario (two multimodal sources); the lower panels show the current waveforms and HRFs used for the simulations. The numbers refer to the source numbers in Table 2.

Results

Estimated cortical current waveforms and their spatial distribution on the cortex in one simulation for scenario MM where both sources are electrically and hemodynamically active are shown in Fig. 5. The currents estimated by the proposed method are closer to the ground truth than those estimated by existing methods, i.e., the spatial distribution of the currents contains sharper transitions between active and inactive regions and the temporal waveforms have an appropriate degree of temporal smoothness. While currents estimated by the BASTERF method are both spatially and temporally smooth, the method fails to recover the sharp transitions at the boundaries of the sources and therefore provides a lower localization performance than the proposed method. This behavior can be explained by the fact that the BASTERF method uses LORETA-type spatial prior which is not spatially adaptive. Due to the lack of spatial smoothness priors the current distribution obtained by the fWMN method is more widespread than the distribution obtained by the proposed and the BASTERF methods. Considering the simplicity of the fWMN method, the results obtained by the fWMN method are surprisingly good. It should be noted however that in our evaluation the fWMN method has an unfair advantage over the symmetrical fusion methods (proposed and BASTERF) since the true locations of the hemodynamically active sources are used to obtain the weights for the fWMN method. Among the EEG-only methods, the MSP method clearly outperforms the other methods (LORETA and MNE) but due to the lack of fMRI information does not recover the spatio-temporal source distribution as well as the evaluated EEG/fMRI fusion methods. The advantage of spatially adaptive priors can also be seen when comparing the HRFs estimated by the proposed method and the BASTERF method, as shown in Fig. 6. As with the cortical currents, spatial adaptivity enables the proposed method to obtain estimates which are closer to the ground truth with sharper transitions between active and inactive regions and a more accurate degree of temporal smoothness.

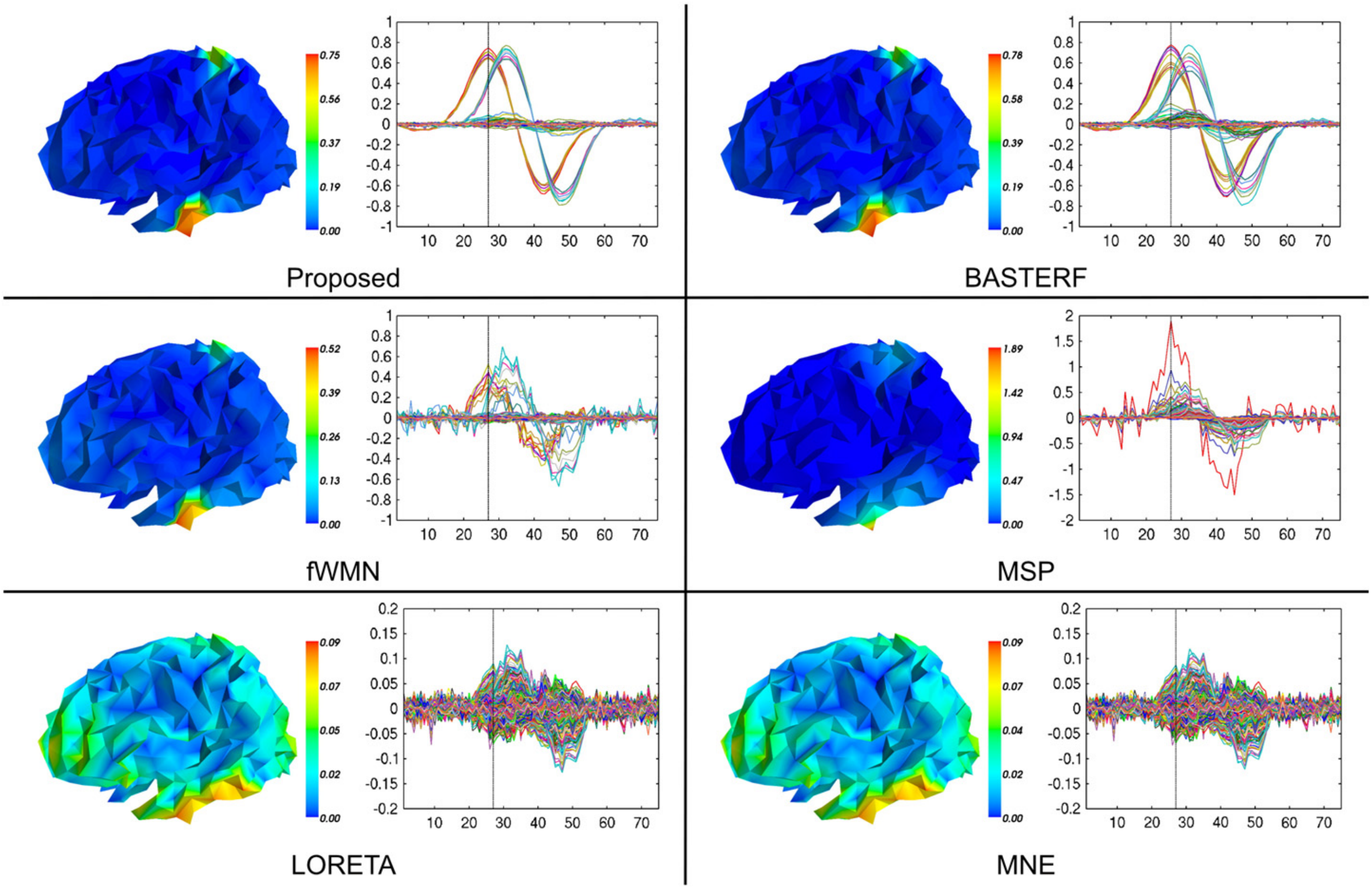


Fig. 5. Butterfly plots of the estimated currents ($\hat{\mathbf{S}}$) and their projection onto the cortical mesh at t = 27 ms for one simulation of the scenario MM (SNR EEG = 20 dB). The ground truth for this simulation is depicted in Fig. 4. Note that the color scales are adjusted for each method to show the full range of the source distribution and that the y-axis of the butterfly plots for the MSP, LORETA, and MNE methods has been adjusted to allow for a clear depiction of the estimated current waveforms.

Objective quality metric scores from all simulations are shown in Fig. 7. To evaluate the reconstruction of the current distribution we use the mean squared error (MSE), denoted MSE EEG, as well as, the area under the ROC curve (AUC EEG). For fMRI we evaluate the reconstruction of the HRFs using the MSE, which we denote MSE fMRI. Refer to Appendix C for the definition of the quality metrics used.

We observe that the proposed method clearly outperforms the other evaluated methods for medium and high EEG SNRs (20 dB and 25 dB), except for the EH scenario where the MSP method performs better. Note, however, that such a result is not unexpected since the EH scenario, which uses one source that is only electrically active and another source that is only hemodynamically active, fundamentally violates the assumption which motivates EEG/fMRI fusion, i.e., that a subset of activity is detectable by both modalities. A method which does not use fMRI information has an advantage in this case since it does not have a bias towards fMRI active locations. From the results for scenario EH it can also be seen that the proposed method is more robust against disagreements between EEG and fMRI than the other EEG/fMRI fusion methods (BASTERF and fWMN). Also note that whenever there is a strong agreement between EEG and fMRI (scenarios CPX and MM), the fusion methods (proposed, BASTERF and fWMN) clearly outperform the EEG-only methods (MSP, LORETA and MNE). It is also interesting to note that the performance for all fusion algorithms is worse when there are current sources which are hemodynamically inactive (scenario ME) than when there are spurious hemodynamic sources (scenario MH), which is in agreement with previously reported results (Liu et al., 1998; Ahlfors and Simpson, 2004; Daunizeau et al., 2005; Daunizeau et al., 2007). As expected, the performance of all evaluated methods degrades when lowering the EEG SNR to 15 dB. It should be noted that the performance of some methods degrades more than that of others, e.g., the advantage

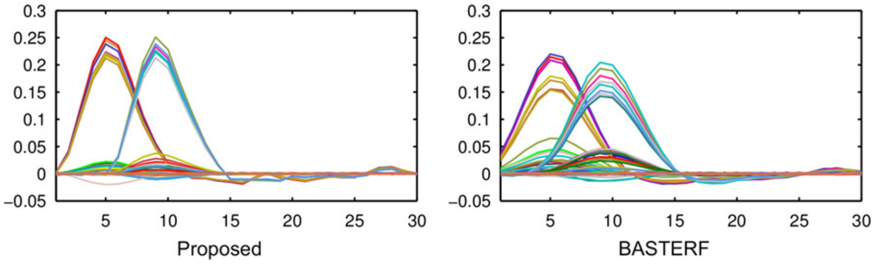


Fig. 6. Estimated HRFs ($\hat{\mathbf{H}}$) by the proposed method and the BASTERF method for one simulation of the scenario MM (SNR EEG = 20 dB). The ground truth for this simulation is depicted in Fig. 4 (hemodynamic sources 1 and 2 are active).

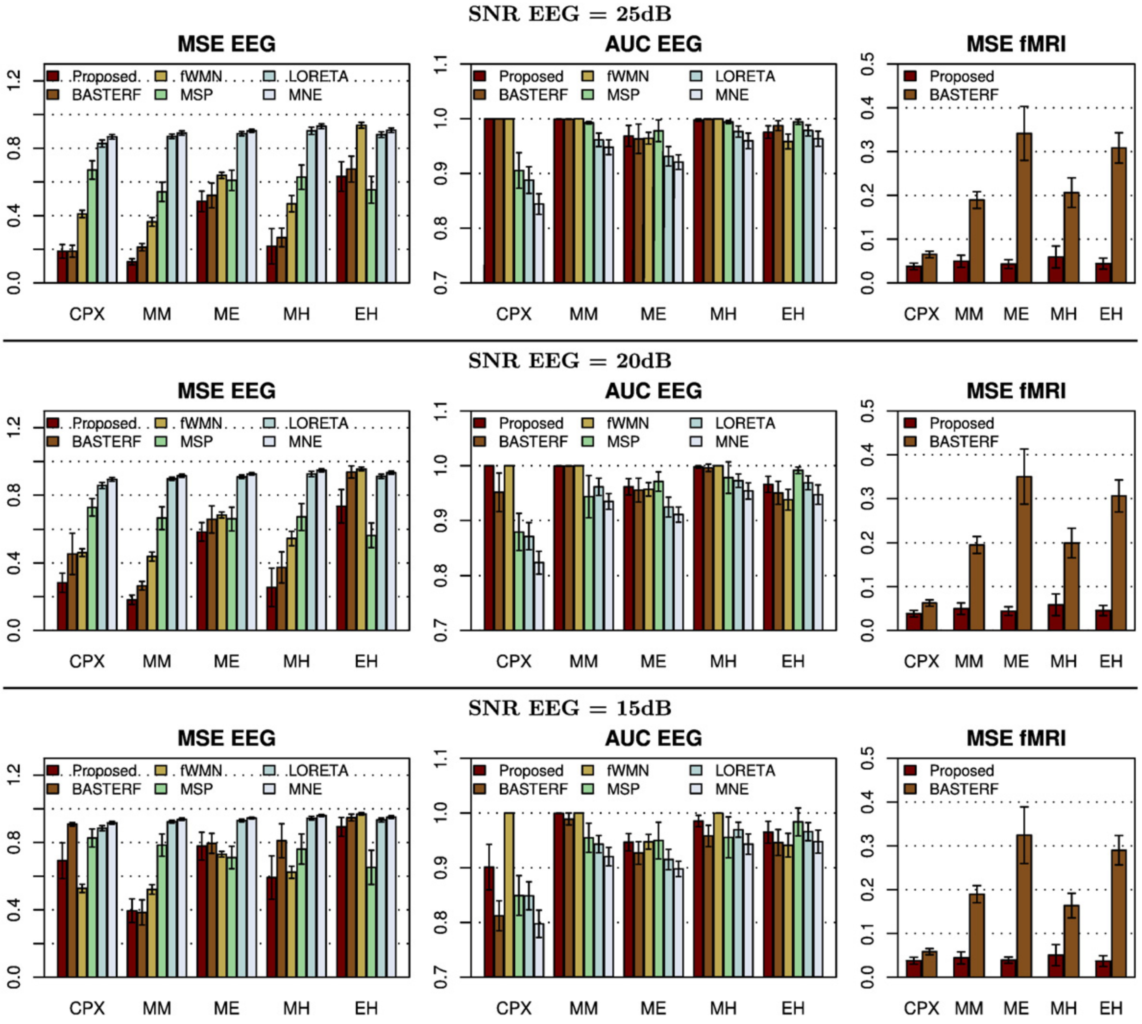


Fig. 7. Objective quality metric scores for different simulation scenarios. The mean squared error scores for the estimated currents and hemodynamic response functions are denoted as MSE EEG and MSE fMRI, respectively. The area under the ROC curve for EEG is denoted as AUC EEG. For mean squared error scores lower values are better while a value of 1.0 indicates the best performance in terms of AUC EEG. The error bars indicate the 95% confidence intervals.

of the proposed method over the BASTERF method typically becomes clearer when lowering SNR. A surprising result is that the fWMN method performs better than the other fusion methods for the CPX scenario at a low SNR. However, the same is not true for the other simulation scenarios. Potentially, this is again due to the fact that the fWMN has an unfair advantage over the other methods since the true source locations are used to obtain the weights used in the method. From Fig. 7 it can also be seen that the proposed method clearly outperforms the BASTERF method in terms of MSE of the hemodynamic response function, which can mainly be attributed to the use of spatially adaptive temporal smoothness priors in the proposed method. Another observation is that the reconstruction of the HRFs is largely unaffected by the EEG SNR and the agreement between EEG and fMRI and mainly depends on the number of hemodynamically active sources (CPX: 4 sources, MM,MH: 2 sources, ME,EH: 1 source). This result is not unexpected since unlike the estimation of **S**, the estimation of **H** does not amount to a localization problem, i.e., it is not possible to use a source configuration with different source locations and obtain the same observation (assuming no noise). Hence, it can be concluded that for realistic fMRI SNRs the estimation of the HRFs does not benefit from the EEG information.

The advantage of the proposed method comes from the improved prior model, consisting of a spatially adaptive TV prior for the spatial profile and spatially adaptive temporal priors for the estimated currents and HRFs. An interesting question is how is the estimation performance affected by each prior? We try to answer this question by repeating the simulations of the CPX scenario with two modified versions of the proposed method, where one prior is replaced with the prior used in the BASTERF method. More specifically, the first method (denoted by ALG1) adopts the spatial Laplacian prior from BASTERF to model **w** and employs spatially adaptive temporal priors to model **X** and **Z**, while the second method (denoted by ALG2) uses a TV prior together with the temporal priors from BASTERF, which are not spatially adaptive. As can be seen from the results in Fig. 8, both additional priors contribute to the

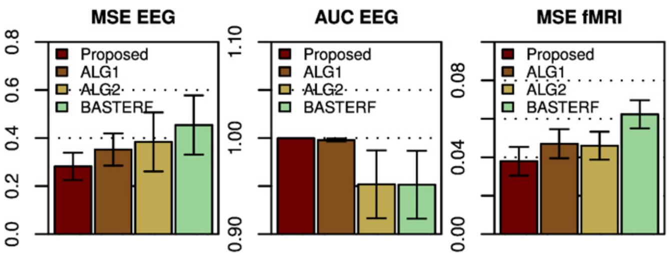


Fig. 8. Results for the CPX scenario (SNR EEG = 20 dB) for the proposed method, the BASTERF method, and intermediate methods, denoted by ALG1 and ALG2. The method ALG1 uses a Laplacian spatial prior (as in BASTERF) together with spatially adaptive temporal priors (as in the proposed method) and ALG2 uses a TV prior (as in the proposed method) together with temporal priors that are not spatially adaptive (as in BASTERF). It can be seen that the improved spatial prior as well as the improved temporal priors contribute to the higher performance of the proposed method. The error bars indicate the 95% confidence intervals.

improved performance in terms of MSE EEG and MSE HRF. An interesting observation is that for the area under the ROC curve (AUC EEG), methods that use spatially adaptive temporal priors (proposed and ALG1) have higher scores than methods that use temporal priors without spatial adaptivity (ALG2 and BASTERF). While we only show results for the CPX scenario, these results are typical and correspond well with our observations that both parts (spatial and temporal) of the improved prior model contribute to the higher performance of the proposed method.

To conclude this evaluation, we also mention run times and convergence properties of the evaluated algorithms. Naturally, while using a more complex symmetrical model, as with the proposed and the BASTERF methods, allows for higher performance, doing so comes at the cost of higher computational complexity. For the simulations used in this evaluation, all methods except the proposed method and the BASTERF method require less than 1 s to perform one simulation. The symmetrical fusion methods (proposed and BASTERF) are significantly more complex and both require about 10 s for one iteration (on a standard 2.6 GHz PC). Note that the time required for one iteration is about the same since the computationally most expensive operations are matrix inversions and both methods perform matrix inversions of the same order during each iteration, i.e., the proposed method and the BASTERF method have the same per-iteration time complexity. The time required for one simulation is in the order of 1 h, as both methods typically require several hundred iterations to reach convergence.

Application to real data

In this section, we demonstrate the performance of the proposed method in a real data set. The EEG and fMRI data was acquired for a multimodal study on face perception; details of the experimental paradigm can be found in Henson et al. (2003) and the data is available at http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces/. The experiment involved the subjects making symmetry judgments for pictures of familiar faces, unfamiliar faces, and scrambled faces. In the following, familiar and unfamiliar faces are combined to create the face condition (F) whereas scrambled faces form the scrambled face condition (S). The data set available contains the data for one subject (male, 33 years old, neurologically healthy).

EEG data

The EEG data was collected using a 128-channel BioSemi ActiveTwo system with two additional electrodes, one on each earlobe, and a sampling rate of 2048 Hz. Faces and scrambled faces were presented in random order for 600 ms, every 3600 ms. Data was collected in two (identical) sessions; 86 faces and 86 scrambled faces were presented in each session. The EEG data was downsampled to 200 Hz, referenced to the average across all channels, and epoched from −100 ms to 600 ms. Trials for which the voltage exceeded 120 μV at any channel were rejected, leaving a total of 136 trials for faces and 134 trials for scrambled faces. The remaining trials were baseline corrected from −100 ms to 0 ms and averaged to create one ERP for the face condition and one ERP for the scrambled face condition.

EEG forward model

The EEG forward operator **G** was calculated using a BEM method implemented in FieldTrip (http://fieldtrip.fcdonders.nl). Subject specific meshes were used for the calculation; the cortex mesh was obtained from a high resolution T1-weighted structural MRI (1 mm^3 resolution) of the subject using BrainVisa 3.2 (http://brainvisa.info). The high resolution cortex mesh obtained by BrainVisa was downsampled to 5998 vertices. The remaining meshes needed for the BEM calculation, namely the scalp, outer skull, and inner skull meshes, were obtained as follows. A nonlinear inverse normalization transform using the T1-weighted structural MRI of the subject was calculated using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). The transform was used to warp template scalp, outer skull, inner skull, and cortex meshes from a standard space into a subject specific space (the template meshes are included in SPM8). The meshes were then used together with electrode locations, which were obtained using a Polhemus Isotrak digitizer, as inputs to the BEM method.

fMRI data

The fMRI data was collected in 2 sessions; 64 faces and 86 scrambled faces were presented in each session. The experimental paradigm was slightly different from that used for EEG, i.e., the stimuli were presented for 600 ms but the time between trials was randomly distributed between 3 s and 18 s to allow for an estimation of the HRF. The data was acquired using a gradient-echo EPI sequence on a 3 T Siemens TIM Trio scanner with 32 slices, voxel size 3 × 3 × 3 mm (skip 0.75 mm), and a TR of 2 s. For each session 390 volumes were obtained. The fMRI data was preprocessed using SPM8, which involved the following steps: Slice timing correction to account for descending slice order, realignment for motion correction using 4-th degree b-spline interpolation, co-registration with the T1-weighted structural MRI of the subject, and spatial smoothing using a symmetric Gaussian kernel with a full width at half maximum (FWHM) of 8 mm. In order to be able to use fMRI data as input to the fusion algorithm, the volumetric data has to be interpolated onto the cortical surface, i.e., the cortical mesh of 5998 vertices which was also used for the EEG BEM model. We use the method proposed in Grova et al. (2006) to perform the interpolation. The method uses a binary gray matter mask to construct a 3D geodesic Voronoi diagram with one Voronoi cell for each vertex of the mesh. The interpolated value at a given vertex is then obtained by averaging the voxels belonging to the Voronoi cell which is associated with the vertex. Compared to simplistic interpolation methods, such as integrating over a sphere around each vertex, this interpolation method has the advantage that each gray matter voxel is associated with exactly one vertex. Therefore no signal mixing occurs between neighboring vertices and no signal is lost due to gray matter voxels being too far away from the closest vertex. Here, the gray matter mask was obtained from the T1-weighted structural MRI using BrainVisa 3.2. After interpolation of the fMRI data for each session onto the cortical mesh, low frequency drifts were removed by fitting and subtracting a third order polynomial to the fMRI waveform of each vertex. The interpolated data from the two sessions were then concatenated and upsampled by a factor of 2 to obtain a pseudo TR of 1 s resulting in an fMRI data matrix **Y** of size 1560 × 5998.

Noise estimates

The proposed method uses two noise-only data segments $\mathbf{M}^0$ and $\mathbf{Y}^0$, for EEG and fMRI, respectively, to obtain noise precision hyperparameters using Eqs. (27) and (29). The pre-stimulus time window from −100 ms to −5 ms was used to obtain an EEG noise matrix $\mathbf{M}^0$ of size 128 × 20. For fMRI, ideally the data segment $\mathbf{Y}^0$ is obtained from a sufficiently long time window during which no event onsets occurred, i.e., it can be assumed that the data segment only contains noise (consisting of measurement noise from the MRI scanner and noise from other sources such as spontaneous brain activity). Unfortunately, the fMRI data provided in the dataset does not contain data from a long period during which no event onsets occurred. In order to obtain an initial noise estimate, first note that the SNR for fMRI is very low and only a small number of brain regions exhibit significant task induced hemodynamic activity. Therefore, calculated across the whole brain and over a long time window, the power of the event related signal is negligible compared to the noise power. Hence, we simply used data from the first 30 s of the experiment, i.e., the first 30 rows in **Y**, as $\mathbf{Y}^0$. Due to the above arguments the noise parameter $b^0_{\alpha_2}$ is quite accurate but may be slightly larger than the “true” $b^0_{\alpha_2}$ due to event onsets during the

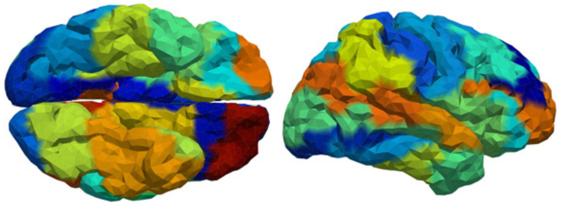


Fig. 9. Ventral (left) and right lateral (right) views of the cortical mesh showing the parcellation of 5998 vertices into 48 regions.

first 30 s of the experiment. It can be expected that this inaccuracy does not affect the result since the noise precision is mostly estimated by the fusion algorithm itself.

Application of the fusion algorithm

The preprocessed EEG and fMRI data were used as inputs to the proposed EEG/fMRI fusion method, as well as, to the BASTERF method (Daunizeau et al., 2007), which was included for comparison purposes. The fusion methods were applied for each condition (face and scrambled face) separately. Prior to applying the algorithms, the cortical mesh was parcellated into 48 regions using the procedure described in Appendix A; the parcellation is illustrated in Fig. 9. The size EEG data matrix **M** was 128 × 61 corresponding to a time window from 0 ms to 300 ms after event onset. The length of the HRF for fMRI was chosen to 20 s, resulting in a design matrix **B** of size 1560 × 20. The design matrix was obtained using Eq. (4) with an experimental time course which was zero everywhere except at locations corresponding to the onset times of the condition of interest, where the value of the time course was set equal to 1.

Results

Previous EEG studies (Henson et al., 2003) have shown that the difference between the face (F) and scrambled face (S) conditions is apparent in the negative component of the right occipito-temporal channels at 170 ms after event onset, which is known as N170. This effect is clearly visible in the estimated current waveforms of the dipoles in the right fusiform region, as illustrated in Fig. 10. Notice that the difference between the F and the S condition is larger for the proposed method than for the BASTERF method. The difference between the methods can be attributed to the spatial adaptivity of the proposed method which allows for more focal sources with adaptive temporal smoothness.

The hemodynamic response functions estimated by both methods look mostly similar as shown in Fig. 11. The similarity between the methods indicates that for this particular example the improved prior model has little influence on the estimates. An explanation for this is that there is a large amount of fMRI data available (86 event onsets for each condition) for the estimation of the HRFs. Hence, the Bayesian methods reduce the weight of the priors and the particular type of prior used has less influence on the estimate. The distributions of the current magnitudes for the F and S conditions at 170 ms are shown in Fig. 12. The results for both, the proposed and the BASTERF methods, are generally consistent with previously reported EEG source localization results for the same data (Trujillo-Barreto et al., 2008; Friston et al., 2008). There is bilateral activity in the fusiform region with emphasis on the right side, as well as, activity in the right superior temporal sulcus and the right middle frontal gyrus. Compared to previously reported results, the current sources, especially the ones in the bilateral fusiform regions, are more clearly separated from inactive regions. This is clear from the sharp boundaries between active and inactive regions shown in Fig. 12. This effect can be explained by the fact that the evaluated EEG/fMRI fusion methods use fMRI information, which allows for more accurate source localization and estimation of the spatial extent of the sources. While the current distributions estimated by the proposed method and the BASTERF method are quite similar, notice that the proposed method obtains sharper boundaries and therefore a better localization of the brain activity. Both methods also find some activity in the medial superior frontal region, which is inconsistent with previous EEG source localization results (Trujillo-Barreto et al., 2008; Friston et al., 2008). Notice that for the BASTERF method, the dipole with the largest magnitude at 170 ms is located in the medial superior frontal region and not in the right fusiform region. More recent MEG results (Henson et al., 2007) show some activity in the medial superior frontal region for some subjects, which suggests that it is possible that previous EEG source localization studies did not report this activity since the employed source localization methods simply failed to detect the activity in the medial superior frontal region. On the other hand, activity in the medial superior frontal region for fMRI and positivity in the frontocentral electrodes for EEG at 550 ms has been reported to be related to the familiarity of faces (Henson et al., 2003). While not shown here, both fusion methods find some hemodynamic medial superior frontal activity. This activity is much weaker than the activity in the fusiform region but may in fact be related to electrical activity that occurs at 550 ms, i.e., outside the EEG time window used in our analysis. The currents in the medial superior frontal regions found by the fusion algorithms may therefore be spurious estimates caused by hemodynamic activity which is related to electrical activity outside the time window of interest. This behavior illustrates a possible shortcoming of EEG/fMRI fusion methods: As the estimated hemodynamic response function is much longer than the EEG time window of interest, information about cortical activity occurring after 300 ms is included into the fusion process, which causes invalid fMRI location priors in the time invariant spatial profile **w**. While both the proposed method and

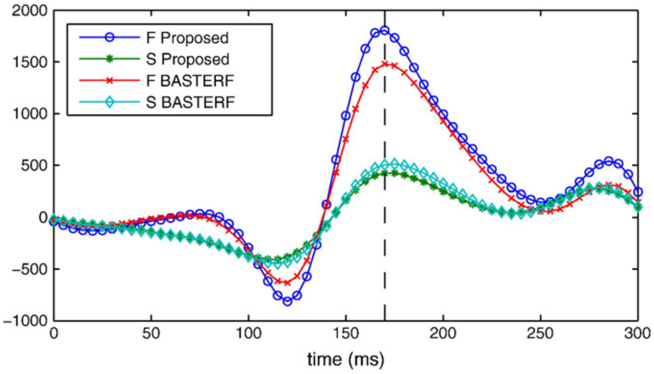


Fig. 10. Estimated current waveforms for a dipole in the right fusiform region. The dipole was selected as the dipole with the maximum current magnitude over all time instants for the face condition and the proposed method. The difference between the face (F) and the scrambled face (S) condition at $t = 170$ ms is clearly visible. Note that the difference is larger for the proposed method than for the BASTERF method.

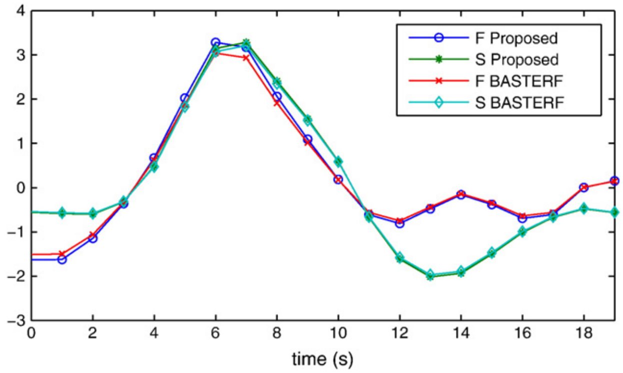


Fig. 11. Estimated hemodynamic response functions for a vertex in the right fusiform region corresponding to the location of the dipole used in Fig. 10.

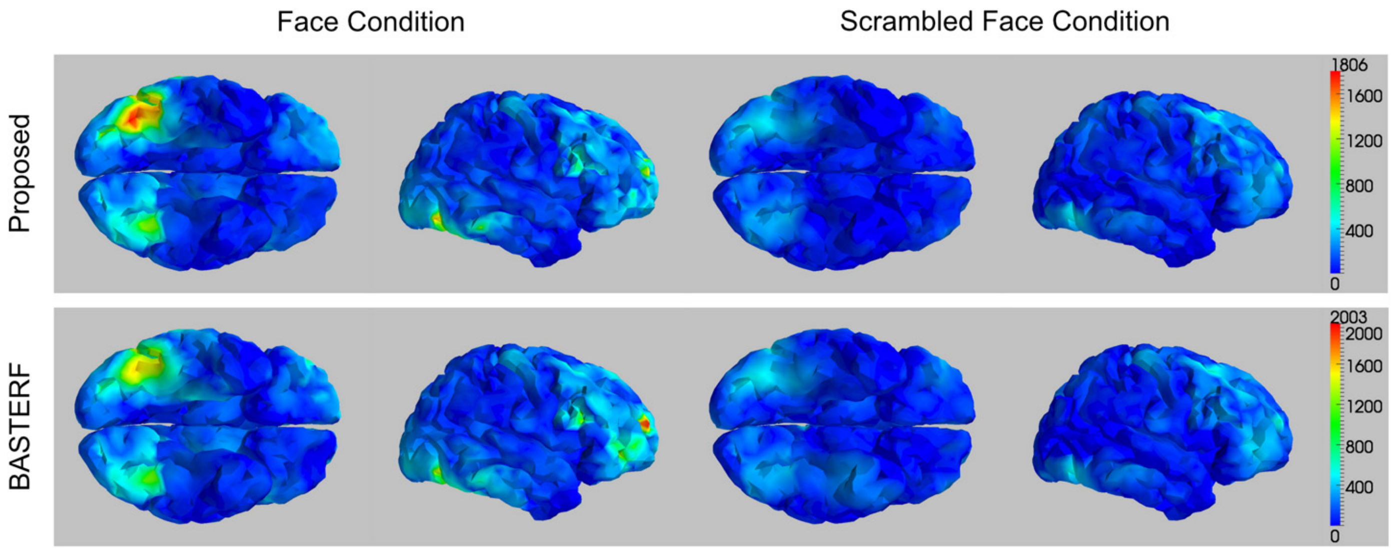


Fig. 12. Distributions of the current magnitudes at $t=170$ ms for the multimodal face data. Results obtained by the proposed method are shown in the top panel while the bottom panel shows the results obtained by the BASTERF method. The color maps are scaled to the range of the current magnitudes for the face condition for each algorithm.

the BASTERF method have some robustness against spurious hemodynamic sources, the current estimates are still biased towards regions with hemodynamic activity and the currents in the medial superior frontal region at 170 ms may in fact be spurious current estimates caused by invalid fMRI location priors.

Conclusions

In this paper we proposed a novel symmetrical EEG/fMRI fusion method. The method utilizes a hierarchical generative model with symmetrical structure which explains both EEG and fMRI observations. In contrast to previous symmetrical fusion methods, the proposed method uses spatially adaptive signal priors, leading to an improved performance. Specifically, the use of a total variation (TV) prior allows sharp boundaries between active and inactive brain regions. Unlike LORETA-type (Pascual-Marqui et al., 1994) spatial priors, the TV prior is spatially adaptive, such that it not only imposes spatial smoothness but also allows for abrupt changes in brain activity at the boundaries of active regions. We also assume that although each response is temporally smooth, the degree of smoothness varies from one spatial location to another, which is incorporated by utilizing a spatially adaptive temporal smoothness prior. We use a fully Bayesian formulation with a variational Bayesian inference method. The method utilizes a spatially adaptive bound to the TV prior which makes the calculation of the variational posterior distribution approximation possible.

We used simulations with synthetic EEG and fMRI data and objective quality metrics to evaluate the proposed method and to compare it to existing methods. In terms of estimation of the spatiotemporal cortical current distribution, our results show that the proposed method outperforms existing methods for simulation scenarios with high agreement between EEG and fMRI, i.e., scenarios where the sources of cortical activity are detectable by either modality. In situations where there is a strong disagreement between EEG and fMRI, the performance of the proposed method was slightly lower than that of the EEG-only MSP method but higher than the performance of other fusion methods, suggesting that the proposed method is more robust against disagreement between EEG and fMRI. In terms of estimation of the hemodynamic response function, the proposed method consistently outperformed the BASTERF method (Daunizeau et al., 2007), which can be attributed to the improved prior model.

We also demonstrated the performance of the proposed method using a multimodal EEG/fMRI dataset from an experiment with face evoked responses (Henson et al., 2003). For comparison purposes, we also applied the BASTERF method to the same data. The results of both methods generally agree with previously reported results for the same data (Trujillo-Barreto et al., 2008; Friston et al., 2008), i.e., 170 ms after event onset the cortical current distribution exhibits clusters of activity in the bilateral fusiform region, as well as, activity in the right superior temporal sulcus and in the right middle frontal gyrus. Compared to previously reported results and to the current distribution obtained by the BASTERF method, the proposed method delineates the clusters in the bilateral fusiform more clearly. The proposed method also obtains a larger difference in terms of current amplitudes between the conditions than the BASTERF method. This can be attributed to the use of the spatially adaptive prior model in the proposed method, which allows for sharp transitions in the cortical current density and for adaptation of the degree of temporal smoothness.

Acknowledgments

The authors would like to thank Rik Henson for making the multimodal EEG/fMRI dataset available and permitting the use of the data for our work. The authors would also like to thank two anonymous reviewers for their comments which helped improving our work considerably. Furthermore, the authors would like to acknowledge support from the National Institute of Child Health and Human Development (HD042049) to James R. Booth. This work was also supported in part by the "Comisión Nacional de Ciencia y Tecnología" under Contract TIC2007-65533 and the Spanish research program Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

Appendix A. Anatomical parceling

In this work we assume a fixed cortical parceling, which is encoded by the matrix **C**. Since there has been no published method to obtain a functional parceling jointly based on EEG and fMRI data, we resort to parceling based on anatomical information. We empirically find that the proposed method, as well as the BASTERF method (Daunizeau et al., 2007), performs better when all parcels are approximately equal in size. Therefore, we use a simple parcellation procedure which tries to segment the cortical mesh in a number of compact parcels with equal size. The parcellation procedure is similar to that in Daunizeau et al. (2007), i.e., the cortical mesh is first down-sampled to obtain a number of seed vertices and then a region growing algorithm is used to obtain the final parcellation. More specifically, in order to obtain a parcellation with q parcels of a cortical mesh $M=(V,E)$ with n vertices, we first down-sample the mesh of each hemisphere to a mesh with $q/2$ vertices using the Matlab function “reducepatch”. Note that we require that q is an even number. The down-sampled meshes are then combined to obtain a mesh $M_\mathrm{D}=(V_\mathrm{D},E_\mathrm{D})$ with a total of q vertices. The vertices in V_D are used as initial labels for the region growing algorithm. In order to start the algorithm we define a label assignment map L_init of length n as

$$L_\mathrm{init}(i) = \begin{cases} j & \text{if } v_i = v_j, v_i \in V, v_j \in V_\mathrm{D}, \\ 0 & \text{otherwise}, \end{cases} \tag{A.1}$$

where $v_i \in V$ and $v_j \in V_\mathrm{D}$ denote the i-th and j-th vertices of the meshes, M and M_D, respectively. The map L_init and the mesh M are then used as inputs to the region growing algorithm in Fig. A.13. The algorithm keeps a map F which indicates if a parcel cannot be grown any further. During each iteration, the algorithm first selects the smallest parcel which can still be grown. In a second step, the neighboring vertex with the largest number of edges connecting the vertex to the selected parcel is added to the parcel. Finally, the algorithm terminates when all vertices have been assigned to a parcel. The $n \times q$ parcellation matrix $\mathbf{C}$ used in the proposed method is then obtained from L as follows

$$\mathbf{C}_{ij} = \begin{cases} 1 & \text{if } L(i) = j, \\ 0 & \text{otherwise}. \end{cases} \tag{A.2}$$

Appendix B. Definition of the SNR

Throughout this paper we use the following definition for the EEG signal to noise ratio

$$\mathrm{SNR}_\mathrm{EEG} = 10\log_{10}\frac{||\mathrm{vec}(\mathbf{LS})||_\infty^2}{\sigma_1^2}, \tag{B.1}$$

where $||\cdot||_\infty$ denotes the infinity norm; i.e., the largest absolute value of the vector, and σ_1^2 is the noise variance. This definition corresponds to the peak signal to noise ratio and has the advantage that it is not affected by the length of silent periods before and after the evoked responses (a similar definition is used in Lapalme et al. (2006)). Similarly, we use the following definition for the fMRI signal to noise ratio

$$\mathrm{SNR}_\mathrm{fMRI} = 10\log_{10}\frac{||\mathrm{vec}(\mathbf{BH})||_\infty^2}{\sigma_2^2}, \tag{B.2}$$

```
INPUTS: $M = (V, E)$, $L_\mathrm{init}$
OUTPUTS: $L$
Initialization: $L \leftarrow L_\mathrm{init}$, $F \leftarrow [0, 0, \ldots, 0]$ (size: $q \times 1$)
while any vertices unassigned do
    Select parcel $i$ with
    $|\{v_j : i = L(j),\ F(i) = 0,\ v_j \in V\}|$ minimal
    Select vertex $v_s$ such that $p$ is maximal with
    $p = |\{(v_r\ v_s) : (v_r\ v_s) \in E,\ L(r) = i,\ L(s) = 0\}|$
    if $p > 0$ then
        Add vertex $v_s$ to parcel $i$: $L(s) \leftarrow i$
    else
        Parcel $i$ is complete: $F(i) \leftarrow 1$
    end if
end while
```

Fig. A.13. Region growing algorithm used to obtain a parcellation of the cortical mesh.

where σ_2^2 denotes the noise variance. One advantage of this definition of the SNR is that the signal power, and thus the SNR, is not affected by the number of voxels for which we assume no hemodynamic response.

Appendix C. Quality metrics

The following objective quality metrics are used in the evaluation. The mean squared error (MSE) score for EEG measures the deviation of the estimated currents $\hat{\mathbf{S}}$ from the true currents $\mathbf{S}$ and is defined as

$$\text{MSE EEG} = \frac{\|\hat{\mathbf{S}}-\mathbf{S}\|_F^2}{\|\mathbf{S}\|_F^2}, \tag{C.1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. In addition to the MSE, we use the area under the ROC curve, denoted as EEG AUC, to evaluate the EEG source localization performance. In order to calculate the AUC we calculate the power map $\mathbf{P}_S$ (Daunizeau et al., 2007) of size $n \times 1$ from the estimated currents $\hat{\mathbf{S}}$ as follows

$$(\mathbf{P}_S)_i = \hat{\mathbf{S}}_{i\cdot}\hat{\mathbf{S}}_{i\cdot}^T, \tag{C.2}$$

i.e., $(\mathbf{P}_S)_i$ contains the power of the estimated source waveform of the i-th dipole. The AUC is then calculated from $(\mathbf{P}_S)_i$ and a binary mask encoding the true locations of vertices belonging to electrically active sources. Unlike the MSE, the AUC does not measure the quality of the estimation based on the spatio-temporal shape of the estimated currents but measures the ability of a method to correctly classify dipoles as either active or inactive based on the energy of the estimated source waveforms. The AUC lies in the range $[0,1]$ where 1 corresponds to perfect classification performance. To evaluate the quality of the estimation of the HRFs we use the MSE, which is analogously defined to the EEG side, i.e.,

$$\text{MSE fMRI} = \frac{\|\hat{\mathbf{H}}-\mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2}, \tag{C.3}$$

with $\hat{\mathbf{H}}$ and $\mathbf{H}$ being the estimated and the true HRFs, respectively.

Appendix D. Derivation of the approximate posterior distribution

In this appendix we show the derivations to obtain the approximate posterior distribution shown in Table 1.

To obtain $\mathrm{q}(\mathbf{S})$, we use Eq. (47) and write

$$\ln \mathrm{q}(\mathbf{S}) = \mathbb{E}_{\Theta\backslash\mathbf{S}}[\ln \mathrm{p}(\mathbf{M}|\mathbf{S},\alpha_1) + \ln \mathrm{p}(\mathbf{S}|\mathbf{X},\mathbf{w},\epsilon_1)] + c, \tag{D.1}$$

where all terms that do not depend on $\mathbf{S}$ have been absorbed into the additive normalization constant c.[1] To perform the calculations it is more convenient to rewrite both $\mathrm{p}(\mathbf{M}|\mathbf{S},\alpha_1)$ and $\mathrm{p}(\mathbf{S}|\mathbf{X},\mathbf{w},\epsilon_1)$ in vector form. They are given by

$$\mathrm{vec}(\mathbf{M})_{(m\cdot t_1)\times 1} = \left(\mathbf{I}_{t_1} \otimes \mathbf{L}_{m\times n}\right)\mathrm{vec}(\mathbf{S})_{(n\cdot t_1)\times 1} + \mathrm{vec}(\eta_1)_{(m\cdot t_1)\times 1}, \tag{D.2}$$

$$\mathrm{vec}(\eta_1)_{(m\cdot t_1)\times 1} \sim \mathcal{N}\left(0, \alpha_1^{-1}\mathbf{I}_{m\cdot t_1}\right), \tag{D.3}$$

and

$$\mathrm{vec}(\mathbf{S})_{(n\cdot t_1)\times 1} = \left(\mathbf{I}_{t_1} \otimes \mathrm{Diag}(\mathbf{w})_{n\times n}\mathbf{C}_{n\times q}\right)\mathrm{vec}(\mathbf{X})_{(q\cdot t_1)\times 1} + \mathrm{vec}(\rho)_{(n\cdot t_1)\times 1}, \tag{D.4}$$

[1] Note that in this appendix c is used for simplicity to denote any terms which are not of interest for a particular derivation. Therefore, the value of c can be different for every equation shown.

$$\mathrm{vec}(\rho_1)_{(n\cdot t_1)\times 1} \sim \mathcal{N}\left(0, \epsilon_1^{-1}\mathbf{I}_{n\cdot t_1}\right), \tag{D.5}$$

respectively. Note that we include the sizes of the matrices and vectors in the subscripts as a reference. Using these equations we can write Eq. (D.1) as

$$\begin{aligned}\ln q(\mathbf{S}) = \mathbb{E}_{\Theta\backslash\mathbf{S}}\Big[&-\frac{\alpha_1}{2}\left(\mathrm{vec}(\mathbf{M})-\left(\mathbf{I}_{t_1}\otimes\mathbf{L}\right)\mathrm{vec}(\mathbf{S})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{M})-\left(\mathbf{I}_{t_1}\otimes\mathbf{L}\right)\mathrm{vec}(\mathbf{S})\right)\\ &-\frac{\epsilon_1}{2}\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)\Big] + c.\end{aligned} \tag{D.6}$$

Due to the conjugacy of the priors (Gaussian for the mean and gamma for the precision) we know that q(**S**) will be Gaussian as well and we can find vec(⟨**S**⟩) by taking the derivative with respect to vec(**S**), equating to zero, and calculating the expectation; by doing so we obtain

$$\begin{aligned}\mathrm{vec}(\langle\mathbf{S}\rangle) = &\left(\langle\alpha_1\rangle\left(\mathbf{I}_{t_1}\otimes\mathbf{L}^T\mathbf{L}\right) + \langle\epsilon_1\rangle\mathbf{I}_{n\cdot t_1}\right)^{-1}\\ &\times\left(\langle\alpha_1\rangle\left(\mathbf{I}_{t_1}\otimes\mathbf{L}^T\right)\mathrm{vec}(\mathbf{M}) + \langle\epsilon_1\rangle\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\right)\mathrm{vec}(\langle\mathbf{X}\rangle)\right),\end{aligned} \tag{D.7}$$

where we can see by inspection that the first part corresponds to the covariance matrix. The covariance matrix can also be obtained by calculating the second derivative of Eq. (D.6) with respect to vec(**S**), equating to zero, and calculating the expectation. Using the properties of the Kronecker product and vec(·) operators, Eq. (D.7) can also be written as

$$\begin{aligned}\langle\mathbf{S}\rangle = &\underbrace{\left(\langle\alpha_1\rangle\mathbf{L}^T\mathbf{L} + \langle\epsilon_1\rangle\mathbf{I}_n\right)^{-1}}_{=\sum_\mathbf{S}}\\ &\times\left(\langle\alpha_1\rangle\mathbf{L}^T\mathbf{M} + \langle\epsilon_1\rangle\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{X}\rangle\right),\end{aligned} \tag{D.8}$$

which is the form given in Table 1.

To obtain the distribution q(**H**) we use the same procedure, i.e., we first write

$$\ln q(\mathbf{H}) = \mathbb{E}_{\Theta\backslash\mathbf{H}}[\ln p(\mathbf{Y}|\mathbf{H},\alpha_2) + \ln p(\mathbf{H}|\mathbf{Z},\mathbf{w},\epsilon_2)] + c \tag{D.9}$$

and use vector notation to obtain

$$\begin{aligned}\ln q(\mathbf{H}) = \mathbb{E}_{\Theta\backslash\mathbf{H}}\Big[&-\frac{\alpha_2}{2}(\mathrm{vec}(\mathbf{Y})-(\mathbf{I}_n\otimes\mathbf{B})\mathrm{vec}(\mathbf{H}))^T\\ &\times(\mathrm{vec}(\mathbf{Y})-(\mathbf{I}_n\otimes\mathbf{B})\mathrm{vec}(\mathbf{H}))\\ &-\frac{\epsilon_2}{2}\left(\mathrm{vec}(\mathbf{H})-\left(\mathbf{C}^T\mathrm{Diag}(\mathbf{w})\otimes\mathbf{I}_k\right)\mathrm{vec}\left(\mathbf{Z}^T\right)\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{H})-\left(\mathbf{C}^T\mathrm{Diag}(\mathbf{w})\otimes\mathbf{I}_k\right)\mathrm{vec}\left(\mathbf{Z}^T\right)\right)\Big] + c.\end{aligned} \tag{D.10}$$

Since q(**H**) is Gaussian, we can obtain the mean by calculating the derivative with respect to vec(**H**) and equating to zero. By doing so and by using the properties of the Kronecker product and vec(·) operators we get

$$\begin{aligned}\langle\mathbf{H}\rangle = &\underbrace{\left(\langle\alpha_2\rangle\mathbf{B}^T\mathbf{B} + \langle\epsilon_2\rangle\mathbf{I}_k\right)^{-1}}_{=\sum_\mathbf{H}}\\ &\times\left(\langle\alpha_2\rangle\mathbf{B}^T\langle\mathbf{Y}\rangle + \langle\epsilon_2\rangle\langle\mathbf{Z}\rangle^T\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right).\end{aligned} \tag{D.11}$$

The distribution q(**X**) is obtained similarly, i.e., we collect all terms that depend on **X** and write

$$\ln q(\mathbf{X}) = \mathbb{E}_{\Theta\backslash\mathbf{X}}[\ln p(\mathbf{S}|\mathbf{X},\mathbf{w},\epsilon_1) + \ln p(\mathbf{X}|\beta_1)] + c. \tag{D.12}$$

Next we rewrite p(**X**|β₁) in vector form as

$$\mathrm{vec}\left(\mathbf{X}^T\right)_{(t_1\cdot q)\times 1} = 0 + \mathrm{vec}(\nu_1)_{(t_1\cdot q)\times 1}, \tag{D.13}$$

$$\mathrm{vec}(\nu_1)_{(t_1\cdot q)\,\tilde{n}\,1} \sim \mathcal{N}\left(0, \left(\mathrm{Diag}(\beta_1)_{q\times q}\otimes\left(\mathbf{T}_1^T\mathbf{T}_1\right)_{t_1\times t_1}\right)^{-1}\right). \tag{D.14}$$

Using this we can write Eq. (D.12) as

$$\begin{aligned}\ln q(\mathbf{X}) = \mathbb{E}_{\Theta\backslash\mathbf{X}}\Big[&-\frac{\epsilon_1}{2}\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)\\ &-\frac{1}{2}\mathrm{vec}\left(\mathbf{X}^T\right)^T\left(\mathrm{Diag}(\beta_1)\otimes\mathbf{T}_1^T\mathbf{T}_1\right)\mathrm{vec}\left(\mathbf{X}^T\right)\Big] + c.\end{aligned} \tag{D.15}$$

Since the prior used for **X** is conjugate, we know that q(**X**) is Gaussian. In order to be able to calculate the derivative with respect to vec(**X**), we define the $t_1\cdot q\times t_1\cdot q$ permutation matrix $\mathbf{R}_{(t_1,q)}$ with the property

$$\mathbf{R}_{(t_1,q)}\mathrm{vec}\left(\mathbf{X}^T\right) = \mathrm{vec}(\mathbf{X}), \tag{D.16}$$

which allows us to rewrite Eq. (D.15) as

$$\begin{aligned}\ln q(\mathbf{X}) = \mathbb{E}_{\Theta\backslash\mathbf{X}}\Big[&-\frac{\epsilon_1}{2}\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)\\ &-\frac{1}{2}\mathrm{vec}(\mathbf{X})^T\mathbf{R}_{(t_1,q)}^T\left(\mathrm{Diag}(\beta_1)\otimes\mathbf{T}_1^T\mathbf{T}_1\right)\mathbf{R}_{(t_1,q)}\mathrm{vec}(\mathbf{X})\Big] + c.\end{aligned} \tag{D.17}$$

By taking the derivative with respect to vec(**X**), equating to zero, and calculating the expectation we obtain

$$\begin{aligned}\mathrm{vec}(\langle\mathbf{X}\rangle) = &\underbrace{\left(\langle\epsilon_1\rangle\left(\mathbf{I}_{t_1}\otimes\mathbf{Q}\right) + \mathbf{R}_{(t_1,q)}^T\left(\mathrm{Diag}(\langle\beta_1\rangle)\otimes\mathbf{T}_1^T\mathbf{T}_1\right)\mathbf{R}_{(t_1,q)}\right)^{-1}}_{=\sum_\mathbf{X}}\\ &\times\langle\epsilon_1\rangle\left(\mathbf{I}_{t_1}\otimes\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right)\mathrm{vec}(\langle\mathbf{S}\rangle),\end{aligned} \tag{D.18}$$

where

$$\begin{aligned}\mathbf{Q} &= \mathbb{E}\left[\mathbf{C}^T\mathrm{Diag}(\mathbf{w})^T\mathrm{Diag}(\mathbf{w})\mathbf{C}\right],\\ &= \mathbf{C}^T\left(\mathrm{Diag}(\langle\mathbf{w}\rangle)^T\mathrm{Diag}(\langle\mathbf{w}\rangle) + \mathrm{Diag}(\mathrm{diag}(\textstyle\sum_\mathbf{w}))\right)\mathbf{C}.\end{aligned} \tag{D.19}$$

To derive q(**Z**) we write

$$\ln q(\mathbf{Z}) = \mathbb{E}_{\Theta\backslash\mathbf{Z}}\left[\ln p\left(\mathbf{H}^T|\mathbf{Z},\mathbf{w},\epsilon_2\right) + \ln p(\mathbf{Z}|\beta_2)\right] + c. \tag{D.20}$$

By comparing the distributions in Eq. (D.20) with those in Eq. (D.12) we see that the distributions have the same form and consequently q(**Z**) has the same form as q(**X**). Therefore, by applying the same steps that we used for the EEG side we obtain

$$\begin{aligned}\mathrm{vec}(\langle\mathbf{Z}\rangle) = &\underbrace{\left(\langle\varepsilon_2\rangle(\mathbf{I}_k\otimes\mathbf{Q}) + \mathbf{R}_{(k,q)}^T\left(\mathrm{Diag}(\langle\beta_2\rangle)\otimes\mathbf{T}_2^T\mathbf{T}_2\right)\mathbf{R}_{(k,q)}\right)^{-1}}_{=\sum_\mathbf{Z}}\\ &\times\langle\varepsilon_2\rangle\left(\mathbf{I}_k\otimes\mathbf{C}^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\right)\mathrm{vec}\left(\langle\mathbf{H}\rangle^T\right).\end{aligned} \tag{D.21}$$

To obtain the distribution $\mathrm{q}(\mathbf{w})$ for the spatial profile, we collect all the terms depending on $\mathbf{w}$, which results in

$$\ln \mathrm{q}(\mathbf{w}) = \mathbb{E}_{\Theta\backslash\mathbf{w}}\left[\ln \mathrm{p}(\mathbf{S}|\mathbf{X},\mathbf{w},\epsilon_1) + \ln \mathrm{p}\left(\mathbf{H}^T|\mathbf{Z},\mathbf{w},\epsilon_2\right) + \ln \mathrm{M}(\mathbf{w},\mathbf{u},\gamma)\right] + c. \tag{D.22}$$

This can be rewritten as

$$\begin{aligned}\ln \mathrm{q}(\mathbf{w}) = \mathbb{E}_{\Theta\backslash\mathbf{w}}\Big[&-\frac{\epsilon_1}{2}\mathrm{tr}\left((\mathbf{S}-\mathrm{Diag}(\mathbf{w})\mathbf{CX})^T(\mathbf{S}-\mathrm{Diag}(\mathbf{w})\mathbf{CX})\right)\\ &-\frac{\epsilon_2}{2}\mathrm{tr}\left(\left(\mathbf{H}^T-\mathrm{Diag}(\mathbf{w})\mathbf{CZ}\right)^T\left(\mathbf{H}^T-\mathrm{Diag}(\mathbf{w})\mathbf{CZ}\right)\right)\\ &-\frac{\gamma}{2}\sum_{i=1}^{n}\frac{\mathbf{w}^T\Delta_i^T\mathbf{G}_i^T\mathbf{G}_i\Delta_i\mathbf{w}+u_i}{\sqrt{u_i}}\Big] + c.\end{aligned} \tag{D.23}$$

Note that there are several terms which do not depend on $\mathbf{w}$. By absorbing all of them into the additive normalization constant and rewriting the remaining terms using $\mathbf{w}$ instead of $\mathrm{Diag}(\mathbf{w})$ we obtain

$$\begin{aligned}\ln \mathrm{q}(\mathbf{w}) = \mathbb{E}_{\Theta\backslash\mathbf{w}}\Big[&\epsilon_1\mathbf{w}^T\mathrm{diag}\left(\mathbf{SX}^T\mathbf{C}^T\right)-\frac{\epsilon_1}{2}\mathbf{w}^T\mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{CXX}^T\mathbf{C}^T\right)\right)\mathbf{w}\\ &+\epsilon_2\mathbf{w}^T\mathrm{diag}\left(\mathbf{H}^T\mathbf{Z}^T\mathbf{C}^T\right)-\frac{\epsilon_2}{2}\mathbf{w}^T\mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{CZZ}^T\mathbf{C}^T\right)\right)\mathbf{w}\\ &-\frac{\gamma}{2}\mathbf{w}^T\left(\sum_{i=1}^{n}\frac{\Delta_i^T\mathbf{G}_i^T\mathbf{G}_i\Delta_i}{\sqrt{u_i}}\right)\mathbf{w}\Big] + c,\end{aligned} \tag{D.24}$$

which has the form of a multivariate Gaussian distribution. We find the mean of the distribution by setting the derivative with respect to $\mathbf{w}$ to zero, resulting in

$$\begin{aligned}\langle\mathbf{w}\rangle = &\underbrace{\left(\langle\epsilon_1\rangle\mathbf{P}_1+\langle\epsilon_2\rangle\mathbf{P}_2+\langle\gamma\rangle W(\mathbf{u})\right)^{-1}}_{\sum_{\mathbf{w}}}\\ &\times\mathrm{diag}\left(\langle\epsilon_1\rangle\langle\mathbf{S}\rangle\langle\mathbf{X}\rangle^T\mathbf{C}^T+\langle\epsilon_2\rangle\langle\mathbf{H}\rangle^T\langle\mathbf{Z}\rangle^T\mathbf{C}^T\right),\end{aligned} \tag{D.25}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are given by

$$\mathbf{P}_1 = \mathbb{E}\left[\mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{CXX}^T\mathbf{C}^T\right)\right)\right] = \mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{C}\left[\langle\mathbf{X}\rangle\langle\mathbf{X}\rangle^T+\sum_{i=1}^{t_1}\sum\nolimits_{\mathbf{X}}^{[i]}\right]\mathbf{C}^T\right)\right), \tag{D.26}$$

$$\mathbf{P}_2 = \mathbb{E}\left[\mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{CZZ}^T\mathbf{C}^T\right)\right)\right] = \mathrm{Diag}\left(\mathrm{diag}\left(\mathbf{C}\left[\langle\mathbf{Z}\rangle\langle\mathbf{Z}\rangle^T+\sum_{i=1}^{k}\sum\nolimits_{\mathbf{Z}}^{[i]}\right]\mathbf{C}^T\right)\right), \tag{D.27}$$

where $\sum_{\mathbf{X}}^{[i]}$ and $\sum_{\mathbf{Z}}^{[i]}$ denote the i-th block of size $q\times q$ on the main diagonal of the corresponding covariance matrix. The $n\times n$ matrix $W(\mathbf{u})$ is defined as

$$W(\mathbf{u}) = \sum_{i=1}^{n}\frac{\Delta_i^T\mathbf{G}_i^T\mathbf{G}_i\Delta_i}{\sqrt{u_i}}. \tag{D.28}$$

Distributions for hyperparameters

Next, we show the derivations of the approximate posterior distributions for the hyperparameters. To obtain the distribution for the EEG noise precision we write

$$\ln \mathrm{q}(\alpha_1) = \mathbb{E}_{\Theta\backslash\alpha_1}\left[\ln \mathrm{p}(\mathbf{M}|\mathbf{S},\alpha_1) + \ln \mathrm{p}\left(\alpha_1|a_{\alpha_1}^0,b_{\alpha_1}^0\right)\right] + c. \tag{D.29}$$

By using vector notation, calculating the logarithms, absorbing constant parts into the constant c, and rearranging, we obtain

$$\begin{aligned}\ln \mathrm{q}(\alpha_1) = \mathbb{E}_{\Theta\backslash\alpha_1}\Big[&\left(\frac{mt_1}{2}+a_{\alpha_1}^0-1\right)\ln(\alpha_1)-\frac{\alpha_1}{2}\left(\mathrm{vec}(\mathbf{M})-\left(\mathbf{I}_{t_1}\otimes\mathbf{L}\right)\mathrm{vec}(\mathbf{S})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{M})-\left(\mathbf{I}_{t_1}\otimes\mathbf{L}\right)\mathrm{vec}(\mathbf{S})\right)-b_{\alpha_1}^0\alpha_1\Big] + c.\end{aligned} \tag{D.30}$$

By comparing this with the functional form of a gamma distribution, i.e.,

$$\mathrm{p}(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}\mathrm{e}^{-bx}, \tag{D.31}$$

where $\Gamma(\cdot)$ denotes the gamma function, we see that $\mathrm{q}(\alpha_1)$ is gamma distributed with parameters

$$a_{\alpha_1} = \frac{mt_1}{2}+a_{\alpha_1}^0, \tag{D.32}$$

$$b_{\alpha_1} = \frac{1}{2}\mathrm{tr}\left((\mathbf{M}-\mathbf{L}\langle\mathbf{S}\rangle)^T(\mathbf{M}-\mathbf{L}\langle\mathbf{S}\rangle)\right)+\frac{t_1}{2}\mathrm{tr}\left(\Sigma_{\mathbf{S}}\mathbf{L}^T\mathbf{L}\right)+b_{\alpha_1}^0, \tag{D.33}$$

where we have used the properties of the $\mathrm{vec}(\cdot)$ and Kronecker product operators to write b_{α_1} in a compact form using the trace operator. The term $t_1\mathrm{tr}\left(\sum_{\mathbf{S}}\mathbf{L}^T\mathbf{L}\right)$ comes from the term that is quadratic with respect to $\mathbf{S}$ in Eq.(D.30), i.e.,

$$\begin{aligned}\mathbb{E}\left[\mathrm{vec}(\mathbf{S})^T\left(\mathbf{I}_{t_1}\otimes\mathbf{L}^T\mathbf{L}\right)\mathrm{vec}(\mathbf{S})\right] &= \mathrm{vec}(\langle\mathbf{S}\rangle)^T\left(\mathbf{I}_{t_1}\otimes\mathbf{L}^T\mathbf{L}\right)\mathrm{vec}(\langle\mathbf{S}\rangle)\\ +\mathrm{tr}\left(\left(\mathbf{I}_{t_1}\otimes\sum\nolimits_{\mathbf{S}}\right)\left(\mathbf{I}_{t_1}\otimes\mathbf{L}^T\mathbf{L}\right)\right) &= \mathrm{tr}\left(\langle\mathbf{S}\rangle^T\mathbf{L}^T\mathbf{L}\langle\mathbf{S}\rangle\right)+t_1\mathrm{tr}\left(\sum\nolimits_{\mathbf{S}}\mathbf{L}^T\mathbf{L}\right).\end{aligned} \tag{D.34}$$

To obtain the distribution for the noise precision of the fMRI side we collect all the terms depending on α_2 and obtain

$$\ln \mathrm{q}(\alpha_2) = \mathbb{E}_{\Theta\backslash\alpha_2}\left[\ln \mathrm{p}(\mathbf{Y}|\mathbf{H},\alpha_2) + \ln \mathrm{p}\left(\alpha_2|a_{\alpha_2}^0,b_{\alpha_2}^0\right)\right] + c. \tag{D.35}$$

Clearly, since the distributions in Eq. (D.29) have exactly the same form as the distributions in Eq. (D.35), $\mathrm{q}(\alpha_2)$ is gamma distributed with parameters that have the same form as the parameters of $\mathrm{q}(\alpha_1)$; they are given by

$$a_{\alpha_2} = \frac{nt_2}{2}+a_{\alpha_2}^0, \tag{D.36}$$

$$b_{\alpha_2} = \frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{B}\langle\mathbf{H}\rangle)^T(\mathbf{Y}-\mathbf{B}\langle\mathbf{H}\rangle)\right)+\frac{n}{2}\mathrm{tr}\left(\sum\nolimits_{\mathbf{H}}\mathbf{B}^T\mathbf{B}\right)+b_{\alpha_2}^0. \tag{D.37}$$

The distribution of the hyperparameter ϵ_1, which controls the strength of the hierarchical prior obtained from the spatio-temporal decomposition model on the EEG side, is obtained by

$$\ln \mathrm{q}(\epsilon_1) = \mathbb{E}_{\Theta\backslash\epsilon_1}[\ln \mathrm{p}(\mathbf{S}|\mathbf{X},\mathbf{w},\epsilon_1) + \mathrm{p}(\epsilon_1)] + c, \tag{D.38}$$

which we can write as

$$\begin{aligned}\ln \mathrm{q}(\epsilon_1) = \mathbb{E}_{\Theta\backslash\epsilon_1}\Big[&\left(\frac{t_1n}{2}-1\right)\ln(\epsilon_1)-\frac{\epsilon_1}{2}\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)^T\\ &\times\left(\mathrm{vec}(\mathbf{S})-\left(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right)\Big] + c.\end{aligned} \tag{D.39}$$

Like for the previous hyperparameter distributions, we can see by inspection that $\mathrm{q}(\epsilon_1)$ is gamma distributed with a shape parameter $a_{\epsilon_1} = t_1n/2$. In order to obtain the parameter b_{ϵ_1} we have to calculate the expectation of the second term in Eq. (D.39). We break the calculation of the expectation into several parts. The calculation of $\mathbb{E}\left[\mathrm{vec}(\mathbf{S})^T\mathrm{vec}(\mathbf{S})\right]$ is similar to Eq. (D.34), i.e.,

$$\begin{aligned}\mathbb{E}\left[\mathrm{vec}(\mathbf{S})^T\mathrm{vec}(\mathbf{S})\right] &= \mathrm{vec}(\langle\mathbf{S}\rangle)^T\mathrm{vec}(\langle\mathbf{S}\rangle)+\mathrm{tr}\left(\left(\mathbf{I}_{t_1}\otimes\sum\nolimits_{\mathbf{S}}\right)\right)\\ &= \mathrm{tr}\left(\langle\mathbf{S}\rangle^T\langle\mathbf{S}\rangle\right)+t_1\mathrm{tr}\left(\sum\nolimits_{\mathbf{S}}\right).\end{aligned} \tag{D.40}$$

The expectation of the second quadratic term is calculated as follows

$$\mathbb{E}\left[\mathrm{vec}(\mathbf{X})^T\left(\mathbf{I}_{t_1}\otimes\mathbf{C}^T\mathrm{Diag}(\mathbf{w})^T\mathrm{Diag}(\mathbf{w})\mathbf{C}\right)\mathrm{vec}(\mathbf{X})\right] = \mathbb{E}\left[\mathrm{vec}(\mathbf{X})^T\left(\mathbf{I}_{t_1}\otimes\mathbf{Q}\right)\mathrm{vec}(\mathbf{X})\right]$$
$$= \mathrm{tr}\left(\langle\mathbf{X}\rangle^T\mathbf{Q}\langle\mathbf{X}\rangle\right) + \mathrm{tr}\left(\sum\nolimits_{\mathbf{X}}\left(\mathbf{I}_{t_1}\otimes\mathbf{Q}\right)\right). \tag{D.41}$$

By combining Eqs. (D.40) and (D.41) and by also including $\mathbb{E}\left[\mathrm{vec}(\mathbf{S})^T(\mathbf{I}_{t_1}\otimes\mathrm{Diag}(\mathbf{w})\mathbf{C})\mathrm{vec}(\mathbf{X})\right]$ we obtain

$$\begin{aligned} b_{\epsilon_1} = \frac{1}{2}\Big[&\mathrm{tr}\left(\langle\mathbf{S}\rangle^T\langle\mathbf{S}\rangle - 2\langle\mathbf{S}\rangle^T\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{X}\rangle + \langle\mathbf{X}\rangle^T\mathbf{Q}\langle\mathbf{X}\rangle\right) \\ &+ t_1\mathrm{tr}(\sum\nolimits_{\mathbf{S}}) + \mathrm{tr}\left(\sum\nolimits_{\mathbf{X}}\left(\mathbf{I}_{t_1}\otimes\mathbf{Q}\right)\right)\Big]. \end{aligned} \tag{D.42}$$

To obtain the distribution $q(\epsilon_2)$ we again make use of the symmetry of the model by realizing that the distributions in

$$\ln q(\epsilon_2) = \mathbb{E}_{\Theta\backslash\epsilon_2}\left[\ln p\left(\mathbf{H}^T|\mathbf{Z},\mathbf{w},\epsilon_2\right) + p(\epsilon_2)\right] + c \tag{D.43}$$

have the same form as the distributions in Eq. (D.38). Therefore, $q(\epsilon_2)$ is gamma distributed with parameters

$$a_{\epsilon_2} = \frac{kn}{2} \tag{D.44}$$

$$\begin{aligned} b_{\epsilon_2} = \frac{1}{2}\Big[&\mathrm{tr}\left(\langle\mathbf{H}\rangle\langle\mathbf{H}\rangle^T - 2\langle\mathbf{H}\rangle\mathrm{Diag}(\langle\mathbf{w}\rangle)\mathbf{C}\langle\mathbf{Z}\rangle + \langle\mathbf{Z}\rangle^T\mathbf{Q}\langle\mathbf{Z}\rangle\right) \\ &+ n\mathrm{tr}(\sum\nolimits_{\mathbf{H}}) + \mathrm{tr}(\sum\nolimits_{\mathbf{Z}}(\mathbf{I}_k\otimes\mathbf{Q}))\Big]. \end{aligned} \tag{D.45}$$

Next, we show the derivation of $q((\beta_1)_i)$, i.e., the distribution of the hyperparameter $(\beta_1)_i$ which controls the degree of temporal smoothness and scale of the current waveforms in the i-th parcel. As before, we only need to keep distributions depending on $(\beta_1)_i$ when applying Eq. (47), resulting in

$$\ln q\left((\beta_1)_i\right) = \mathbb{E}_{\Theta\backslash(\beta_1)_i}[\ln p(\mathbf{X}|\beta_1) + \ln p(\beta_1|\delta_1)] + c. \tag{D.46}$$

Note that we can assign all parts of $p(\mathbf{X}|\beta_1)$ and $p(\beta_1|\delta_1)$ which are independent of $(\beta_1)_i$ to the additive normalization constant, which allows us to write

$$\begin{aligned} \ln q\left((\beta_1)_i\right) = \mathbb{E}_{\Theta\backslash(\beta_1)_i}\Bigg[&\ln \mathit{det}\left(2\pi\left((\beta_1)_i\mathbf{T}_1^T\mathbf{T}_1\right)\right)^{\frac{1}{2}} - \frac{(\beta_1)_i}{2}\mathbf{X}_{i\cdot}\mathbf{T}_1^T\mathbf{T}_1\mathbf{X}_{i\cdot}^T \\ &-\delta_1(\beta_1)_i + \ln((\beta_1)_i)\left(a_{\beta_1}^0 - 1\right)\Bigg] + c, \end{aligned} \tag{D.47}$$

where $\mathit{det}(\cdot)$ denotes the determinant. By using the properties of the determinant and the logarithm, calculating the expectation, and rearranging we obtain

$$\begin{aligned} \ln q((\beta_1)_i) = &-\frac{(\beta_1)_i}{2}\left(\langle\mathbf{X}_{i\cdot}\rangle\mathbf{T}_1^T\mathbf{T}_1\langle\mathbf{X}_{i\cdot}\rangle^T + \mathrm{tr}\left(\mathbf{T}_1^T\mathbf{T}_1\mathrm{cov}\left((\mathbf{X}_{i\cdot})^T\right)\right)\right) \\ &-(\beta_1)_i\langle\delta_1\rangle + \ln((\beta_1)_i)\left(\frac{t_1}{2} + a_{\beta_1}^0 - 1\right) + c, \end{aligned} \tag{D.48}$$

where $\mathrm{cov}\left((\mathbf{X}_{i\cdot})^T\right)$ denotes the $t_1 \times t_1$ covariance matrix of the i-th row of $\mathbf{X}$; it can be extracted from $\sum_{\mathbf{X}}$ as follows

$$\mathrm{cov}\left((\mathbf{X}_{i\cdot})^T\right)_{r,c} = (\sum\nolimits_{\mathbf{X}})_{i\,+\,(r-1)q,\ i\,+\,(c-1)q}. \tag{D.49}$$

By comparing Eq. (D.48) with the functional form of a gamma distribution (Eq. (D.31)) we see that $q((\beta_1)_i)$ is gamma distributed with parameters

$$\left(\mathbf{a}_{\beta_1}\right)_i = \frac{t_1}{2} + a_{\beta_1}^0, \tag{D.50}$$

$$\left(\mathbf{b}_{\beta_1}\right)_i = \frac{1}{2}\left[\langle\mathbf{X}_{i\cdot}\rangle\mathbf{T}_1^T\mathbf{T}_1\langle\mathbf{X}_{i\cdot}\rangle^T + \mathrm{tr}\left(\mathbf{T}_1^T\mathbf{T}_1\mathrm{cov}\left((\mathbf{X}_{i\cdot})^T\right)\right)\right] + \langle\delta_1\rangle. \tag{D.51}$$

To obtain $q((\beta_2)_i)$, we write

$$\ln q\left((\beta_2)_i\right) = \mathbb{E}_{\Theta\backslash(\beta_2)_i}[\ln p(\mathbf{Z}|\beta_2) + \ln p(\beta_2|\delta_2)] + c \tag{D.52}$$

and again notice that due to the symmetry of the model the distributions have the exact same form as the distributions in Eq. (D.46). Thus, by following the same procedure that we used to obtain $q((\beta_1)_i)$ we find that $q((\beta_2)_i)$ is gamma distributed with parameters

$$\left(\mathbf{a}_{\beta_2}\right)_i = \frac{k}{2} + a_{\beta_2}^0, \tag{D.53}$$

$$\left(\mathbf{b}_{\beta_2}\right)_i = \frac{1}{2}\left[\langle\mathbf{Z}_{i\cdot}\rangle\mathbf{T}_2^T\mathbf{T}_2\langle\mathbf{Z}_{i\cdot}\rangle^T + \mathrm{tr}\left(\mathbf{T}_2^T\mathbf{T}_2\mathrm{cov}\left((\mathbf{Z}_{i\cdot})^T\right)\right)\right] + \langle\delta_2\rangle. \tag{D.54}$$

The distribution $q(\delta_1)$ is obtained by calculating

$$\ln q(\delta_1) = \mathbb{E}_{\Theta\backslash\delta_1}[\ln p(\beta_1|\delta_1) + \ln p(\delta_1)] + c, \tag{D.55}$$

which, by absorbing terms into c, can be written as

$$\ln q(\delta_1) = \mathbb{E}_{\Theta\backslash\delta_1}\left[-\delta_1\sum_{i=1}^{q}(\beta_1)_i + \ln(\delta_1)\left(qa_{\beta_1}^0 - 1\right)\right] + c. \tag{D.56}$$

From this it can be seen that $q(\delta_1)$ is gamma distributed with parameters

$$a_{\delta_1} = qa_{\beta_1}^0, \quad b_{\delta_1} = \sum_{i=1}^{q}\langle(\beta_1)_i\rangle. \tag{D.57}$$

Similarly, we find that $q(\delta_2)$ is gamma distributed with parameters

$$a_{\delta_2} = qa_{\beta_2}^0, \quad b_{\delta_2} = \sum_{i=1}^{q}\langle(\beta_2)_i\rangle, \tag{D.58}$$

by calculating

$$\ln q(\delta_2) = \mathbb{E}_{\Theta\backslash\delta_2}[\ln p(\beta_2|\delta_2) + \ln p(\delta_2)] + c. \tag{D.59}$$

Finally, we show the derivation of $q(\gamma)$, i.e., the distribution of the hyperparameter which controls the strength of the TV prior. By collecting all terms that depend on γ and absorbing independent parts into the additive constant we obtain

$$\ln q(\gamma) = \mathbb{E}_{\Theta\backslash\gamma}[\ln F(\mathbf{w},\mathbf{u},\gamma) + \ln p(\gamma)] + c. \tag{D.60}$$

By calculating the logarithm and absorbing parts independent of γ into c we obtain

$$\ln q(\gamma) = \mathbb{E}_{\Theta\backslash\gamma}\left[\ln(\gamma)(\varphi n - 1) - \frac{\gamma}{2}\sum_{i=1}^{n}\frac{\mathbf{w}^T\Delta_i^T\mathbf{G}_i^T\mathbf{G}_i\Delta_i\mathbf{w} + u_i}{\sqrt{u_i}}\right] + c. \tag{D.61}$$

From which we can see that $q(\gamma)$ is gamma distributed and that the shape parameter is given by $a_\gamma = \varphi n$. To calculate b_γ we use Eq. (49) to obtain

$$b_\gamma = \frac{1}{2}\mathbb{E}_{\Theta\backslash\gamma}\left[\sum_{i=1}^{n}\frac{\mathbf{w}^T\Delta_i^T\mathbf{G}_i^T\mathbf{G}_i\Delta_i\mathbf{w} + u_i}{\sqrt{u_i}}\right] = \sum_{i=1}^{n}\sqrt{u_i}. \tag{D.62}$$

References

Adde, G., Clerc, M., Keriven, R., 2005. Imaging methods for MEG/EEG inverse problem. International Journal of Bioelectromagnetism 7 (2), 111–114.

Ahlfors, S.P., Simpson, G.U., 2004. Geometrical interpretation of fMRI-guided MEG/EEG inverse estimates. Neuroimage 22 (1), 323–332 May.

Attias, H., 2000. A variational Bayesian framework for graphical models. Advances in Neural Information Processing Systems 12 (1–2), 209–215.

Babacan, S.D., Molina, R., Katsaggelos, A.K., 2008. Parameter estimation in TV image restoration using variational distribution approximation. IEEE Transactions on Image Processing 17 (3), 326–339.

Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Transactions on Biomedical Engineering 44 (5), 374–385 August.

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Processing Magazine 18 (6), 14–30.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Brookings, T., Ortigue, S., Grafton, S., Carlson, J., 2009. Using ICA and realistic bold models to obtain joint EEG/fMRI solutions to the problem of source localization. Neuroimage 44 (2), 411–420 September.

Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. Journal of Cognitive Neuroscience 5, 162–176.

Daunizeau, J., Grova, C., Mattout, J., Marrelec, G., Clonda, D., Goulard, B., Pelegrini-Issac, M., Lina, J.M., Benali, H., 2005. Assessing the relevance of fMRI-based prior in the EEG inverse problem: a Bayesian model comparison approach. IEEE Transactions on Signal Processing 53 (9), 3461–3472.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pelegrini-Issac, M., Lina, J.M., Benali, H., 2007. Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. Neuroimage 36 (1), 69–87 May.

Frahm, J., Bruhn, H., Merboldt, K.D., Math, D., 1992. Dynamic MR imaging of human brain oxygenation during rest and photic stimulation. Journal of Magnetic Resonance Imaging 2 (5), 501–505.

Friedrich, M., Friederici, A.D., 2004. N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. Journal of Cognitive Neuroscience 16 (8), 1465–1477.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. Neuroimage 2 (1), 45–53 March.

Friston, K., Henson, R., Phillips, C., Mattout, J., 2006. Bayesian estimation of evoked and induced responses. Human Brain Mapping 27 (9), 722–735.

Friston, K.J., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N.J., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39 (3), 1104–1120 February.

George, J., Aine, C., Mosher, J., Schmidt, D., Ranken, D., Schlitt, H., Wood, C., Lewine, J., Sanders, J., Belliveau, J., 1995. Mapping function in the human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. Journal of Clinical Neurophysiology 12 (5), 406.

Grova, C., Makni, S., Flandin, G., Ciuciu, P., Gotman, J., Poline, J., 2006. Anatomically informed interpolation of fMRI data on the cortical surface. Neuroimage 31 (4), 1475–1486 July.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Medical & Biological Engineering & Computing 32 (1), 35–42.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of Modern Physics 65 (2), 413–497 Apr.

Hardy, G.H., Littlewood, J.E., Pólya, G., 1988. Inequalities. Cambridge University Press.

Henson, R.N., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., Rugg, M.D., 2003. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cerebral Cortex 13 (7), 793–805 July.

Henson, R., Mattout, J., Singh, K., Barnes, G., Hillebrand, A., Friston, K., 2007. Population-level inferences for distributed MEG source localization under multiple constraints: application to face-evoked fields. Neuroimage 38 (3), 422–438.

Henson, R. N., Flandin, G., Friston, K. J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Human Brain Mapping 31 (10), 1512–1531.

Hillyard, S.A., Hinrichs, H., Tempelmann, C., Morgan, S.T., Hansen, J.C., Scheich, H., Heinze, H.J., 1997. Combining steady-state visual evoked potentials and fMRI to localize brain activity during selective attention. Human Brain Mapping 5 (4), 287–292.

Huang, M.X., Dale, A.M., Song, T., Halgren, E., Harrington, D.L., Podgorny, I., Canive, J.M., Lewis, S., Lee, R.R., 2006. Vector-based spatial–temporal minimum L1-norm solution for MEG. Neuroimage 31 (3), 1025–1037.

Jaakkola, T.S., Jordan, M.I., 1998. Improving the mean field approximation via the use of mixture distributions. Learning in Graphical Models 89, 163–173.

Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Machine Learning 37 (2), 183–233.

Jun, S.C., George, J.S., Kim, W., Paré-Blagoev, J., Plis, S., Ranken, D.M., Schmidt, D.M., 2008. Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC. Neuroimage 40 (4), 1581–1594 May.

Lapalme, E., Lina, J., Mattout, J., 2006. Data-driven parceling and entropic inference in MEG. Neuroimage 30 (1), 160–171 March.

Laufs, H., Daunizeau, J., Carmichael, D.W., Kleinschmidt, A., 2008. Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. Neuroimage 40 (2), 515–528 April.

Liu, Z., He, B., 2008. fMRI–EEG integrated cortical source imaging by use of time-variant spatial constraints. Neuroimage 39 (3), 1198–1214 February.

Liu, A.K., Belliveau, J.W., Dale, A.M., 1998. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. Proceedings of the National Academy of Sciences U.S.A. 95 (15), 8945–8950 July.

MacKay, D.J.C., 1992. Bayesian interpolation. Neural Computation 4 (3), 415–447.

Marrelec, G., Benali, H., Ciuciu, P., Poline, J.B., 2002. Bayesian estimation of the hemodynamic response function in functional MRI. Bayesian Inference and Maximum Entropy Methods in Science and Engineering 617 (1), 229–247.

Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. Meg source localization under multiple constraints: an extended Bayesian framework. Neuroimage 30 (3), 753–767 April.

Ogawa, S., Lee, T., Kay, A., Tank, D., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences 87 (24), 9868.

Ou, W., Nummenmaa, A., Ahveninen, J., Belliveau, J.W., Hämäläinen, M.S., Golland, P., 2010. Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation. Neuroimage 52 (1), 97–108.

Parisi, G., 1998. Statistical Field Theory. Westview Press.

Pascual-Marqui, R., Michela, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology 18 (1), 49–65 October.

Pflieger, M.E., Greenblatt, R.E., 2001. Nonlinear analysis of multimodal dynamic brain imaging data. Int. J. Bioelectromagnetism 3.

Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J., 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. Neuroimage 24 (4), 997–1011 February.

Ramírez, R. R., May 2005. Neuromagnetic source imaging of spontaneous and evoked human brain dynamics. Ph.D. thesis, New York Univ., New York.

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Physica D 259–268.

Sanders, L.D., Stevens, C., Coch, D., Neville, H.J., 2006. Selective auditory attention in 3- to 5-year-old children: an event-related potential study. Neuropsychologia 44 (11), 2126–2138.

Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. Neuroimage 23 (3), 806–826 November.

Scherg, M., Von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. Electroencephalography and Clinical Neurophysiology 65 (5), 344.

Strong, D., Chan, T., 2003. Edge-preserving and scale-dependent properties of total variation regularization. Inverse Problems 19, S165.

Thürmer, G., Wuthrich, C.A., 1998. Computing vertex normals from polygonal facets. Journal of Graphics Tools 3 (1), 43–46.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1, 211–244.

Trujillo-Barreto, N.J., Aubert-Vazquez, E., Penny, W.D., 2008. Bayesian M/EEG source reconstruction with spatio-temporal priors. Neuroimage 39 (1), 318–335 January.

Uutela, K., Hämäläinen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. Neuroimage 10 (2), 173–180.

Wipf, D. P., 2006. Bayesian methods for finding sparse representations. Ph.D. thesis, University of California, San Diego.

Wipf, D.P., Nagarajan, S.S., 2009. A unified Bayesian framework for MEG/EEG source imaging. Neuroimage 44 (3), 947–966.

Wipf, D.P., Owen, J.P., Attias, H.T., Sekihara, K., Nagarajan, S.S., 2010. Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. Neuroimage 49 (1), 641–655.

Yao, J., Dewald, J.P.A., 2005. Evaluation of different cortical source localization methods using simulated and experimental EEG data. Neuroimage 25 (2), 369–382 April.