A mesostate-space model for EEG and MEG

Jean Daunizeau* and Karl J. Friston

The Wellcome Deparment of Imaging Neuroscience, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3BG, UK

We present a multi-scale generative model for EEG, that entails a minimum number of assumptions about evoked brain responses, namely: (1) bioelectric activity is generated by a set of distributed sources, (2) the dynamics of these sources can be modelled as random fluctuations about a small number of *mesostates*, (3) mesostates evolve in a temporal structured way and are functionally connected (i.e. influence each other), and (4) the number of mesostates engaged by a cognitive task is small (e.g. between one and a few).

A Variational Bayesian learning scheme is described that furnishes the posterior density on the models parameters and its evidence. Since the number of meso-sources specifies the model, the model evidence can be used to compare models and find the optimum number of meso-sources.

In addition to estimating the dynamics at each cortical dipole, the mesostate-space model and its inversion provide a description of brain activity at the level of the *mesostates* (i.e. in terms of the dynamics of meso-sources that are distributed over dipoles). The inclusion of a mesostate level allows one to compute posterior probability maps of each dipole being active (i.e. belonging to an active mesostate). Critically, this model accommodates constraints on the number of meso-sources, while retaining the flexibility of distributed source models in explaining data. In short, it bridges the gap between standard distributed and equivalent current dipole models. Furthermore, because it is explicitly spatiotemporal, the model can embed any stochastic dynamical causal model (e.g. a neural mass model) as a Markov process prior on the mesostate dynamics.

The approach is evaluated and compared to standard inverse EEG techniques, using synthetic data and real data. The results demonstrate the added-value of the mesostate-space model and its variational inversion.

Introduction

Electroencephalography (EEG) and Magnetoencephalography (MEG) provide direct physical observations of neuronal activity in the millisecond range. These measurements are thought to be generated by synchronized post-synaptic activity of neuronal macrocolumns (Baillet et al., 2001). However, the underlying current dipole sources cannot be estimated uniquely from the scalp data, without invoking priors or constraints on the inverse solution (Nunez, 1981).

As a consequence, EEG/MEG inverse techniques differ in the nature of their priors. For instance, assuming that each active area comprises a subset of temporally coherent neuronal macrocolumns has enabled the family of equivalent current dipole (ECD) methods (Scherg and von Cramon, 1986). Each ECD models the activity of a whole cortical area, which is described by a single temporal dynamic and its location. In contradistinction, distributed linear (DL) methods, rest on estimating the amplitude of a predefined highly dense ensemble of dipoles, typically distributed over the cortical sheet (Dale and Sereno, 1993). Usually, additional spatial and/or temporal constraints are used to finesse the under-determination of DL inverse solutions (Grova et al., 2006). More recently, models of the coupling between active brain areas (ECDs) have led to qualitatively different (i.e. biologically informed) inverse approaches (David et al., 2006; Kiebel et al., submitted for publication). These dynamic causal models are generally inverted to make inferences about the coupling parameters or effective connectivity but can also be used for conventional spatiotemporal source reconstruction.

The inversion of these models is usually formulated in Bayesian terms, where the constraints enter as priors and the objective of model inversion is to estimate the conditional or posterior probability of the model parameters. In addition, there has been a recent focus on model comparison and averaging, using the evidence or marginal (i.e. integrated) likelihood of different models. Model comparison allows one to explore model space (e.g. number of sources) in a principled way (Daunizeau et al., 2005; Penny et al., 2004).

In this work, we propose a generic spatiotemporal model that accommodates spatial and temporal priors on the deployment of bioelectric activity causing EEG–MEG signals. It is formulated to cover the sorts of constraints found in both DL and ECD models; it allows for distributed sources and precise anatomical prior information but, at the same time, can generate data using a small number of anatomically compact meso-sources. Critically, we use

* Corresponding author. Fax: +44 207 813 1445.
E-mail address: j.daunizeau@l.ion.ucl.ac.uk (J. Daunizeau).
Available online on ScienceDirect (www.sciencedirect.com).

a Bayesian inversion scheme that optimizes both the number of meso-sources and their spatiotemporal characteristics.

Brain activity is organized at many different scales: collectively, the microscopic behaviour of individual neurons is related to observed macroscopic dynamics through the average behaviour of local neuronal populations or ensembles (i.e. distributed sources comprising many macrocolumns or cortical dipoles). Indeed, most mean-field and neural mass models of neuronal activity are formulated at this level (David et al., 2006; Kiebel et al., submitted for publication). In this work, we model brain activity in terms of a set of locally distributed and temporally coherent *meso*-sources (c.f. Daunizeau et al., 2007; Daunizeau et al., 2006 for related work). Each meso-source can be regarded as a mean-field approximation to its underlying neuronal population dynamics. Critically, we model the activity of each source as random fluctuations around the average of sources that belong to each set (each meso-source). This average, over an ensemble of sources, represents a *mesoscopic* scale of activity, which we assume conforms to the same sorts of dynamics governing the ensemble of neurons in each source. These assumptions are implemented with a mixture of Gaussians (MoG) prior on the activity of sources distributed over the cortical mesh. This furnishes a flexible and structured spatial prior and induces a small number of hidden mesostates that describe the average dynamics of sources in the mixture. Priors on the temporal evolution of the averages or mesostates can be derived from neural mass models describing the interactions between excitatory and inhibitory subpopulations at the mesoscopic scale. This entails a (stochastic) dynamic causal model (DCM) that can be framed as a first-order Markov process prior on the mesostates. In short, the mesostate dynamics generate canonical state-space dynamics, about which the sources in each mixture fluctuate, hence the name mesostate-space model (MSM).

We describe an efficient Variational Bayesian (VB) inversion scheme, which allows one to invert the ensuing generative model, i.e. to derive the marginal posterior probability density functions (pdf) associated with the dipole and meso-source dynamics, the cortical parcelling of sources and parameters of the DCM mesostate. These densities can then be used for inference and construction of posterior probability maps (PPMs) of source activation in dipole space. This VB scheme optimizes a free-energy bound on the model evidence or marginal likelihood, which (after optimization) can be used for model comparison and selection.

This paper is organized as follows. The mesostate-space model is described in the Mesostate-space model of EEG–MEG section. In the Variational inversion section, we introduced the Variational Bayesian inversion framework. In the Comparative evaluations section, we describe comparative evaluations of the current and conventional models, using both numerical simulations and real data. Finally, our results and methodology are discussed in the Discussion section. Appendix A contains all the requisite technical details for implementation (i.e. the update equations for the inversion scheme).

Notations

In the following, $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}_i$, X_{ij} and $\mathrm{tr}(\mathbf{X})$ indicate the transpose of $\mathbf{X}$, the ith vector column of $\mathbf{X}$, the scalar element of the ith column and jth row of $\mathbf{X}$ and the trace of $\mathbf{X}$, respectively. $(x_i)_{1 \leq i \leq n}$ denotes the $n \times 1$ vector whose entries are x_i. $\mathbf{I}_n$, 1_n and 0_n stand for the $n \times n$

identity matrix, the $n \times 1$ vector of ones, and the $n \times 1$ null vector, respectively. For any $n \times 1$ vector $\mathbf{x}$, $\mathrm{Diag}(\mathbf{x})$ denotes the $n \times n$ diagonal matrix whose diagonal is $\mathbf{x}$. Conversely, $\mathrm{diag}(\mathbf{X})$ denotes the $n \times 1$ vector containing the diagonal entries of the $n \times n$ matrix $\mathbf{X}$. $\otimes$ denotes the Kronecker product and "$\propto$" means proportional. For two variables x and y, $x|y$ stands for "x given y", $p(x)$ for the probability of x, and $\langle x \rangle$ for its expectation. $\mathcal{N}(\mathbf{m}, \mathbf{V})$ is the Gaussian probability density function (pdf) with mean $\mathbf{m}$ and covariance matrix $\mathbf{V}$, $\mathcal{G}(a, b)$ is the Gamma pdf with a degrees of freedom (d.o.f.) and shape parameter b, $Multinom(n, \mathbf{c})$ is the multinomial pdf with sample sizes $\mathbf{c}$ out of n trials, $Dir(\mathbf{c})$ is the Dirichlet pdf with sample sizes $\mathbf{c}$ and $\psi(\cdot)$ is the digamma function.

Mesostate-space model of EEG–MEG

In this section, we describe the generative model through which we introduce priors on the spatial deployment and temporal evolution of brain activity. The model is specified in a hierarchical fashion, in terms of the conditional dependencies among the model parameters that generate spatiotemporal responses.

The EEG–MEG data likelihood

Solving the EEG–MEG inverse problem within a so-called distributed framework amounts to finding a unique solution to the linear system (Dale and Sereno, 1993):

$$\mathbf{y}_t = \mathbf{G}\mathbf{J}_t + \mathbf{e}_t, t = 1, \dots, T \tag{1}$$

where $\mathbf{y}_t$ stands for the $p \times 1$ vector of scalp or sensor data ($p \sim 10^2$: number of sensors, T: number of time samples), $\mathbf{e}_t$ is additive measurement noise, $\mathbf{J}_t$ is the $n \times 1$ vector describing the unknown dipole time courses ($n \sim 10^4$: number of dipoles distributed on the cortical surface) and $\mathbf{G}$ is the $p \times n$ gain matrix (the so-called forward operator). $\mathbf{G}$ is obtained by solving the quasi-static electromagnetic forward problem (de Munck, 1988) for a given set of dipoles with fixed position and orientation (here perpendicular to the cortical surface). Each column $\mathbf{G}_j$ of $\mathbf{G}$ indicates the putative contribution of dipole j to the data in sensor space (its forward field).

The likelihood results from distributional assumptions about the measurement noise. We consider $\mathbf{e}_t$ as a realization of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and (unknown) precision σ. In this case:

$$p(\mathbf{y}_t|\mathbf{J}_t, \sigma) = \mathcal{N}(\mathbf{G}\mathbf{J}_t, \sigma^{-1}\mathbf{I}_p) \tag{2}$$

Each of these dipoles has a known position in source space. In the following, we will refer to $\mathbf{X}$ as the $3 \times n$ matrix containing the spatial coordinates of the dipoles.

Mesostates and meso-sources

We assume that cortical activity can be described by a set of K mesostates, which describe the ensemble activity of all dipoles belonging to a particular mesostate. These ensembles correspond to (spatially extended) meso-sources of constituent dipoles that populate a particular cortical region.

As with the ECD model, each meso-source or dipole ensemble has a mean temporal expression and spatial position. However, here we consider each meso-source as an ensemble, i.e. a *cluster*, of dipoles. This means we also model the variability inherent in the temporal dynamics and position of the dipoles in each meso-

source. These can be regarded as *local perturbations of the mesostates*.

Let:

- $\boldsymbol{\xi}$ be a binary $n \times K$ matrix such that: $\xi_{ik}=1$ when the ith dipole belongs to the kth cluster (by convention, the Kth cluster contains all inactive dipoles). $\boldsymbol{\xi}$ will be referred to as the labelling process.
- $\boldsymbol{\mu}$ be the $K \times T$ matrix containing the time course of each of the meso-sources (the Kth time course is supposed to be zero).
- $\boldsymbol{\eta}$ be the $3 \times K$ matrix containing the mean position of each of the meso-sources.

Then, we can postulate that (Penny and Friston, 1999):

1. the probability of observing J_{ti} is a mixture of Gaussians (MoG) associated with each meso-source such that:

$$p(J_{ti}|\boldsymbol{\xi}, \boldsymbol{\mu}_t, \alpha_1, \ldots, \alpha_K) = \prod_{k=1}^{K} \left[\mathcal{N}(\mu_{tk}, \alpha_k^{-1}) \right]^{\xi_{ik}}, \quad t = 1, \ldots, T,$$

(3)

where α_k is the precision of the kth Gaussian (i.e. mesostate) of the mixture. A high temporal coherence within the kth meso-source is parameterized with a high value of α_k. Note that $\mu_{tK}=0$ (non-activated cluster).

2. the model of random fluctuations also applies to the position of the dipoles, i.e. the probability of observing $\mathbf{X}_i$ is a mixture of 3D-Gaussians associated with meso-source; such that:

$$p(\mathbf{X}_i|\boldsymbol{\xi}, \boldsymbol{\eta}, \gamma_1, \ldots, \gamma_K) = \prod_{k=1}^{K} \left[\mathcal{N}(\boldsymbol{\eta}_k, \gamma_k^{-1}\mathbf{I}_3) \right]^{\xi_{ik}},$$

(4)

where γ_k is the precision of the kth Gaussian of the mixture (in the anatomical space). A low γ_k value generates a spatially extended cluster.

Mesostate dynamics

Commonly, prior expectations about the dynamics of electrical meso-sources come from considering them as *nodes in a network*. Usually, the interactions among meso-sources are modelled with some form of dynamic causal model (Friston et al., 2003; Horwitz et al., 1999). Moreover, we may want to consider random fluctuations in the (hidden) meso-dynamics (of coupled meso-sources) (Freeman et al., 2001). In this situation, we have to infer the mesostates, given our observations and our prior beliefs about their dynamical behaviour.

The resulting stochastic dynamical causal model is used as a (hierarchical) prior on the temporal evolution of mesostates. In other words, they are constrained to lie in the vicinity of a manifold defined by the models evolution equation. These temporal priors constrain the EEG–MEG inverse problem substantially and help regularize conditional estimates.

Let us assume that the mesostates obey the following evolution equation:

$$\dot{\boldsymbol{\mu}}(t) = \mathbf{h}(\boldsymbol{\mu}(t)) + \mathbf{f}^*(t),$$

(5)

where $\mathbf{h}$ is the evolution function parameterizing the dynamic causal model, and $\mathbf{f}^*(t)$ is some random (Wiener) process. This can

be rewritten as a first-order multivariate autoregressive model of the form:[1]

$$\boldsymbol{\mu}_t = \mathbf{A}\boldsymbol{\mu}_{t-1} + \mathbf{f}_{t-1}, \qquad t = 1, \ldots, T;$$

(6)

where

$$\mathbf{A} = \exp(\Delta t \mathbf{H}) \approx I + \Delta t \mathbf{H}$$

(7)

is a multivariate autoregression (MAR) matrix and Δt is the time between EEG samples. Here, $\mathbf{H} = \partial \mathbf{h}/\partial \mu$ is the systems Jacobian and $\mathbf{f}_t = \Delta t \mathbf{f}^*(t)$ is accumulated system noise, which we will assume is i.i.d. Gaussian. Strictly speaking, this equality only holds when the evolution function is linear in the mesostates. Generally, the autoregressive form relying on a local linearization of the non-linear evolution function is sufficient. Given the fast sampling of EEG, this is a reasonable assumption. Under this assumption we can use Eq. (6) to build a first-order Markov process prior for $\boldsymbol{\mu}_t$:

$$p(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}, \beta) = \mathcal{N}(\mathbf{A}\boldsymbol{\mu}_{t-1}, \beta^{-1}\mathbf{I}_K),$$

(8)

where β is the (unknown) precision of $\mathbf{f}_t$.

For simplicity, in this introductory paper we will use $\mathbf{A}=\mathbf{I}_K$, which gives a simple AR(1) smoothness prior on the evolution of mesostates. This comes from assuming the following form for the evolution function: $\mathbf{h}(\boldsymbol{\mu}(t))=(1/\tau)\boldsymbol{\mu}(t)$ where $\tau > \Delta t$.[2]

Spatial priors and hyperpriors

To complete the MoG formulation in Eqs. (3) and (4), we have to specify the prior on the labelling process $\boldsymbol{\xi}$: we assume that each dipole belongs to the kth meso-source with a priori probability λ_k, which is the same for all dipoles. In other words, the columns $\boldsymbol{\xi}_i$ of the labelling process $\boldsymbol{\xi}$ are a priori i.i.d. multinomial:

$$p(\boldsymbol{\xi}|\boldsymbol{\lambda}) \propto \prod_{i=1}^{n} \prod_{k=1}^{K} \lambda_k^{\xi_{ik}}.$$

(9)

Note that the standard MoG formulation, in terms of a weighted sum of Gaussians, is equivalent to this; the weights being the class frequencies $(\lambda_k)_{k=1,\ldots,K}$.

We further suppose λ is a priori Dirichlet-distributed (for conjugacy purposes):

$$p(\boldsymbol{\lambda}) \propto \prod_{k=1}^{K} \lambda_k^{b_k-1},$$

(10)

where $\mathbf{b}_k$ are prior hyperparameters that we suppose are known. Note that, *a priori*: $\langle \lambda_k \rangle = b_k/b_{00}$, where $b_{00} = \sum_{k=1}^{K} b_k$. Hence, for non-informative (equiprobable) prior on $\boldsymbol{\lambda}$, the b_k can be fixed as: $b_k = n/K$. This parameterization postulates $b_{00} = n$ "prior (virtual) observations" for defining prior class frequencies and their associated prior uncertainty. Note that the Dirichlet distribution

---

[1] Note that the time index $t=0, \ldots, T$ in the MAR model (Eq. (6)) is different from the continuous independent variable $t \in [0, +\infty]$ of the stochastic differential equation (Eq. (5)).

[2] Which is a weakly compromising assumption, given the EEG/MEG data sampling.

enforces a normalization constraint, which is sensible since the λ_k are frequencies:

$$1 = \sum_{k=1}^{K} \lambda_k. \tag{11}$$

A direct consequence of the assumption of Gaussian measurement noise is that the measurement noise precision σ given data behaves as a Gamma variate. The definition of its prior pdf thus pertains to the derivation of the conditional pdf of σ given (previously observed) "pure-noise" data $\mathbf{y}_0$ (e.g. obtained by anti-averaging[3]):

$$p(\sigma) \equiv p(\sigma|\mathbf{y}_0) = \mathcal{G}\left(\frac{pt_0}{2}, \frac{1}{2}\,\mathrm{tr}\left[\mathbf{y}_0^T \mathbf{y}_0\right]\right), \tag{12}$$

where t_0 is the number of time samples in the null data $\mathbf{y}_0$.

Equivalently, we may have some prior information regarding the expected spatial extent of the active regions. This may be included in the generative model using a Gamma prior pdf for the position precisions of the MoG:

$$p(\gamma) = \prod_{k=1}^{K} \mathcal{G}(a_{\gamma,0}^k, b_{\gamma,0}^k). \tag{13}$$

For the remaining precision hyperparameters, we use non-informative Jeffreys priors (uniform over log-scale):

$$\begin{cases} p(\beta) \propto (\beta)^{-1} \\ p(\alpha) \propto \prod_{k=1}^{K} (\alpha_k)^{-1} \end{cases}, \tag{14}$$

where we made use of the condensed notations: $\boldsymbol{\alpha} = (\alpha_k)_{1 \le k \le K}$, and $\gamma = (\gamma_k)_{1 \le k \le K}$. Finally, we will use flat priors on the position of the meso-sources:

$$p(\boldsymbol{\eta}) \propto 1. \tag{15}$$

In other contexts, one may want using information from fMRI or previous EEG studies (to place spatial priors on the location). This could be implemented here as a Gaussian prior (see Discussion).

The choice of priors above may seem rather ad hoc; however, the priors for generative models of this sort are usually self-evident and are based on positivity and conjugacy constraints, in a principled way. Conjugate priors are necessary for variational learning because they prescribe the form of the conditional density and its sufficient statistics that need updating. In the present context, the use of Dirichlet prior for the class frequencies $\boldsymbol{\lambda}$ is standard because; (i) a Dirichlet density enforces to be positive and sum to one and (ii) a Dirichlet prior is conjugate to the multinomial density. Similarly, the use of a Gamma hyperprior for the precision parameters (Eqs. (12) and (13)) is a standard choice, which can be motivated by the fact that: (i) a Gamma density enforces positivity constraints (required for precisions) and (ii) they are conjugate to the Gaussian likelihood implicit in our assumptions about measurement noise. A direct consequence of the assumption of Gaussian measurement noise is that σ, given data, behaves as a Gamma variate. In other words, the forms of the prior and posterior densities are identical. This is due to the use of conjugate priors.

---

[3] Anti-averaging refers to the estimation of correlations in the noise by eliminating signal from empirical data. Signal is eliminated by taking the average of differences between successive realizations of each trial.
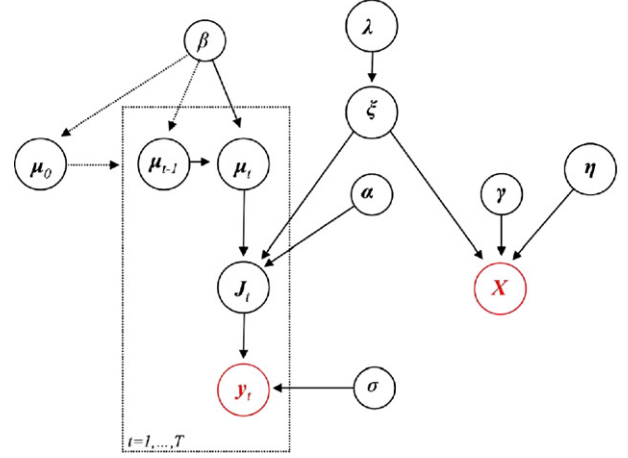

Fig. 1. Graph representing the hierarchical dependencies among the generative model parameters. This graphical model $\mathcal{M}$ depicts the mesostate-space model. Each sequence $\{\mathbf{y}_1, ..., \mathbf{y}_T\}$ is represented as the panel over T pairs of hidden variables.

Posterior or conditional density

The graphical model associated with the generative model described above is depicted in Fig. 1.

This graphical model encompasses all the assumptions associated with the hierarchical generative model $\mathcal{M}$. The Markov properties of the graphical model enables us to write the joint posterior density of the model parameters as the following product:

$$\begin{aligned} p(\mathbf{J}, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta}, \sigma, \beta, \alpha, \gamma | \mathbf{y}, \mathbf{X}) &\propto p(\boldsymbol{\xi}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}) p(\sigma) p(\beta) p(\alpha) p(\gamma) \\ &\times \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{J}_t, \sigma) p(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}, \beta). \\ &\times \prod_{i=1}^{n} p(\mathbf{X}_i|\boldsymbol{\xi}, \boldsymbol{\eta}, \gamma) \prod_{t=1}^{T} p(J_{ti}|\boldsymbol{\xi}, \boldsymbol{\mu}_t, \alpha) \end{aligned} \tag{16}$$

Variational inversion

There are two goals of Bayesian inversion: the first is to estimate the posterior distribution over the model parameters. The second is to assess the quality of the model $\mathcal{M}$ in terms of its evidence $p(\mathbf{y}|\mathcal{M})$ (where we leave the dependence on $\mathbf{X}$ implicit). The evidence or marginal likelihood allows one to perform model comparison and averaging.

For most generative models, the posterior dependencies among the parameters render $p(\mathbf{y}|\mathcal{M})$ difficult to evaluate. The main purpose of variational Bayes (VB) is to finesse this problem by approximating the joint posterior of the unknown parameters with a simpler (variational) density $q(\ldots)$. This approximation usually posits independence assumptions beyond those implied by the generative model. In this work, we use the *mean-field approximation* (Ghahramani and Beal, 2000), in which the approximate density factorizes into the product of the variational marginal posteriors on the nodes of the graphical model $\mathcal{M}$:

$$\begin{aligned} & q(\mathbf{J}, \boldsymbol{\xi}, \boldsymbol{\lambda}\boldsymbol{\mu}, \boldsymbol{\eta}, \sigma, \beta, \alpha, \gamma) \\ & \approx q(\boldsymbol{\mu}_{1:T}) q(\boldsymbol{\eta}) q(\boldsymbol{\xi}) q(\boldsymbol{\lambda}) q(\sigma) q(\beta) q(\alpha) q(\gamma) \prod_{t=1}^{T} q(\mathbf{J}_t). \end{aligned} \tag{17}$$

where $q(\cdot)$ denotes the variational approximation of any marginal posterior pdf $p(\cdot|\mathbf{y}, \mathcal{M})$. The mean field approximation enables us to decompose the model evidence $p(\mathbf{y}|\mathcal{M})$ the following way:

$$\ln p(\mathbf{y}|\mathcal{M}) = \underbrace{\langle \ln p(\ldots, y|\mathcal{M}) \rangle_{\prod q(\cdot)} + \sum S(q(\cdot))}_{\mathcal{F}(q)} \\ + \mathcal{D}_{KL}(q(\ldots), p(\ldots|\mathbf{y}, \mathcal{M})),$$ (18)

where $\mathcal{S}(\cdot)$ is the Shannon entropy, $\mathcal{D}_{KL}(\cdot)$ is the Kullback–Leibler divergence and "…" (resp. "·") denotes the full set of nodes (resp. each node) of the graphical model.

The quantity $\mathcal{F}(q)$ is called the negative free energy. Since the model evidence is fixed, for a given generative model $\mathcal{M}$, maximizing the negative free energy is equivalent to minimizing the Kullback–Leibler divergence between the mean-field approximation and the true posterior. In other words, the higher the negative free energy, the more $q(\cdot)$ approaches the true marginal posterior pdf.

The functional form of the variational marginal density $q(\cdot)$ comes from optimizing the free energy:

$$\frac{\partial \mathcal{F}(q)}{\partial q(\cdot)} = 0 \Rightarrow \ln q(\cdot) = \langle \ln p(\ldots, \mathbf{y}|\mathcal{M}) \rangle \prod_{q(\cdot)} + cst,$$ (19)

where the expectation is under the marginals of the Markov blanket of each node.[4] The VB scheme is simply an iterative optimization of the negative free energy through updates of the sufficient statistics of the variational marginal of each node. After convergence, the negative free energy can be used as an approximation (a lower bound) to the models log-evidence. In other words, maximizing the negative free energy allows one to approximate the marginal posterior of the model parameters and, coincidently, the model evidence.

Under the mean-field approximation, the variational density of location and scale (precision) parameters are Gaussian and Gamma densities, respectively. Moreover, the class frequencies $\boldsymbol{\lambda}$ are Dirichlet-distributed, and each labelling process $\boldsymbol{\xi}_i$, $i=1, \ldots, n$ follows a multinomial pdf. The derivation of each variational marginal $q(\cdot)$, and the update equations that form our VB scheme, are provided in Appendix A.

Critically, the number of meso-sources K is an implicit part of the model $\mathcal{M}$. In our VB scheme, meso-sources can be 'switched off' by eliminating meso-sources that are not supported by the data. This is possible because we estimate the posterior probability that each dipole belongs to each meso-source ($p(\xi_{ik}=1|\mathbf{y}_{1:T})$). Therefore, during the VB iterative process, we may find a set of *empty* meso-sources k_0, i.e. such that:

$$k_0 = \left\{ k : \sum_{i=1}^{n} p(\xi_{ik} = 1|\mathbf{y}_{1:T}) = 0 \right\}.$$ (20)

As a consequence, the VB scheme can remove unnecessary components, i.e. the final number of meso-sources may be less than the starting number. This is referred to as automatic relevance determination (ARD) (Choudrey and Roberts, 2003). This ARD process allows us to estimate the number of meso-sources (i.e. the

effective dimension of the hidden mesostates) that explain the data optimally, as part of the inversion itself. However, this ARD-like behaviour depends on starting with the optimum number of meso-sources or more and on finding the global free energy extremum. Therefore, we will still resort to model comparison to show how the negative free energy $\mathcal{F}(q) = \mathcal{F}_K(q)$ depends on the number of meso-sources (K). In this context, the best model satisfies:

$$K^* = \underset{K}{\arg\max} \, \mathcal{F}_K(q).$$ (21)

In the next section, we will ignore the inactive or background ensemble (cluster) so that when we talk about K active meso-sources we imply there are $K+1$ dipole ensembles in total.

Comparative evaluations

We evaluated the variational schemes ability to recover the spatial support of the meso-sources (i.e. cortical parcelling) and their temporal dynamics in relation to other standard EEG–MEG inverse approaches using Monte-Carlo simulations. We also illustrate inversion of the mesostate-space model using real data. The following section describes the simulations and the real data used for the evaluations.

Simulations

Comparisons

Inversion of the mesostate-space model (MSM) provides conditional estimates of dipole dynamics, as well as the mesoscale descriptions of brain activity, i.e. the meso-sources (described by the multinomial labelling process), and their dynamics (the mesostates). Furthermore, we can create PPMs of dipole activation; in terms of the probability that each dipole does not belong to the Kth inactive ensemble.

In contrast, existing approaches to the EEG inverse problem furnish only the conditional expectation of dipole dynamics. Therefore, we compared our MSM to three classical methods (minimum norm estimator—MNE (Hämäläinen and Ilmoniemi, 1994), weighted MNE—MWNE (Kohler et al., 1996), LORETA (Pascual-Marqui et al., 1994) and the standard restricted maximum-likelihood (ReML)). In fact, ReML was used to embed all these classical methods by specifying appropriate prior covariance components in dipole space; For the ReML solution, the prior covariance Σ of the dipole dynamics had two components: $\Sigma = \sum h_i \mathbf{Q}_i$, where $\mathbf{Q}_1$ was the identity matrix (i.i.d. assumption) and $\mathbf{Q}_2$ was the inverse of a discrete Laplacian (smoothness constraint). The hyperparameters $(h_i)_{1 \leq i \leq 2}$ were estimated using restricted maximum-likelihood (Mattout et al., 2006).

The conditional estimates of dipole activities were evaluated using three adequacy scores:

- the SSE score, which is the normalized (time-integrated) sum of squared errors on the estimated dipole activities:

$$SSE = \log \left[\frac{\sum_{t=1}^{T} \| \mathbf{J}_t - \hat{\mathbf{J}}_t \|^2}{\sum_{t=1}^{T} \| \mathbf{J}_t \|^2} \right];$$ (22)

where $\mathbf{J}$ and $\hat{\mathbf{J}}$ are the simulated and estimated bioelectric dipole activities respectively.

---

[4] The Markov blanket refers to the neighborhood in the moralized graph of the generative model (c.f., Fig. 1). The Markov blanket of each node comprises its parents, its children and the parents of its children. The mean-field approach uses the Markov blanket of each node to build the sufficient statistics of its marginal (variational) posterior pdf.

- the (normalized) area under the ROC curve (ROC):

$$\text{ROC} = \log\left[\frac{P}{1-P}\right], \tag{23}$$

where P is the area under the ROC curve (Egan, 1975). For each Monte-Carlo simulation, the ROC curve was built by varying the threshold on the power of (sum of squared) dipole-wise dynamics, and comparing the resulting activation maps with the known spatial support of the simulated sources.

- the correlation coefficient (CC) between the true and estimated dipole activities. The CC was calculated using the entire time series of all distributed dipoles.

These accuracy scores were calculated for all models (MSM, MNE, WMNE, LORETA and ReML).

In addition, to these comparative analyses, we evaluated the MSMs ability to recover the correct number of the mesostates; i.e. the number of active meso-sources. We used model comparison (as described in the Variational inversion section) with different numbers of meso-sources: $1 < K < 6$ (plus one inactive ensemble of dipoles).

The definition of the prior regarding the mesostate spatial precision parameters may be context-dependent. In both the simulations and the visual ERP application, we used:

$$\begin{cases} a_{\gamma,0}^k = \dfrac{n}{2(K-1)}, & k = 1, \dots, K, \\ b_{\gamma,0}^k = 20 \times a_{\gamma,0}^k \end{cases}$$

which assumes that the mesosource has a spatial extension of about 20 cm² (see, e.g. Ebersole, 1997).

Synthetic data

A cortical surface was extracted from a structural MRI of a normal subject and down-sampled to about 500 vertices ($n = 458$). The sampling frequency was chosen to be 1 kHz (with $T = 200$ time samples), there were $p = 128$ sensors, the gain matrix was calculated using a three-sphere analytical model (Mosher et al., 1999), and measurement noise was drawn from i.i.d. Gaussian pdf, such that the signal-to-noise ratio was SNR = 10. We ran four series of simulations, with one ($S1$), two ($S2$), three ($S3$) and four ($S4$) activated meso-sources, each series consisting of 50 randomly chosen source configurations. In total, we then ran 200 different Monte-Carlo trials.

Results

Fig. 2 shows an example of a simulation with three active meso-sources with overlapping time courses ($S3$ series). These data were modelled using the MSM with different numbers of meso-sources. Fig. 3 shows the corresponding free energies $\mathcal{F}(K)$ and the final number of active meso-sources (K^*; after ARD component elimination) as a function of K. As expected the inferred K^* corresponds to the K with the highest negative free energy.

In this example, the VB scheme removed all unnecessary components (except for the $K = 4$ case), i.e. the true number of activated regions ($K^* = 3$) was found by model comparison schemes.[5]

---

[5] The negative free energy of the models with $K = 3$ (after ARD) were significantly higher than the other models.
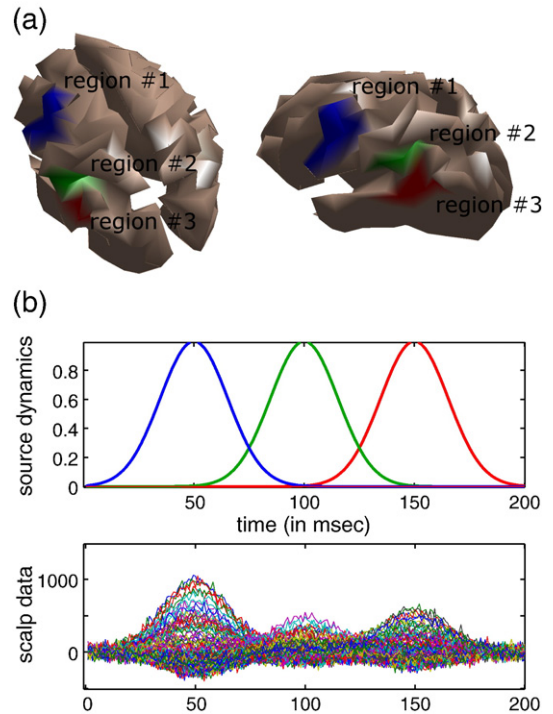

Fig. 2. Example of source configuration for the simulations with three meso-sources (S3). (a) The three labelled active meso-sources; (b)-top: time courses of the three active meso-sources; (b)-bottom: corresponding (noisy) scalp EEG measurements.

Fig. 4 shows the posterior probability map of being active for the best model. The probability of a source or mesh-dipole being active is one minus the probability it belongs to the $(K+1)$th inactive ensemble. These probabilities are shown in our PPMs; with red denoting an active source and blue an inactive source. Fig. 5 shows the associated mesostates dynamics. In addition, we show the labelling process (defined by assigning each dipole to its most probable meso-source), and the spatial position of the meso-sources. This model ($K^* = 3$) estimated correctly the mean position and spatial extent of the meso-sources (see the multinomial labelling process (a) and the temporal dynamics of the regions (b) in Fig. 5 and posterior probability map (PPM) of being active (Fig. 4). Moreover, the mesostate dynamics were correctly estimated, and with high precision (note the low posterior confidence intervals $\mathbf{m}_{1:T}$ of $\boldsymbol{\mu}_{1:T}$).

Fig. 6 shows the histogram of the inferred number of active meso-sources K^*, using the model selection, as a function of the true number of active regions. In most simulation series (S1, S2, S3), the peak of the distribution corresponds to the true number of active regions. The higher error rate for the (S4) series is most likely due to the depth issue of the EEG inverse problem. Since the simulated regions are randomly deployed on the cortical surface, there is more chance that at least one of them falls in a deep structure. These sources are less detectable on the scalp compared to other sources. Therefore, as the number of true meso-sources increases, the inferred K^* may be too small. This demonstrates the slight undercomplexity bias of the mesostate-space model VB inversion (see Discussion).

Fig. 7 shows the empirical cdf of the accuracy metric SSE, ROC and CC for all 200 simulations, for the five inverse methods (LORETA, MNE, WMNE, ReML and MSM). Except for the SSE score (Fig. 7(a)), better performance is indicated by curves that are
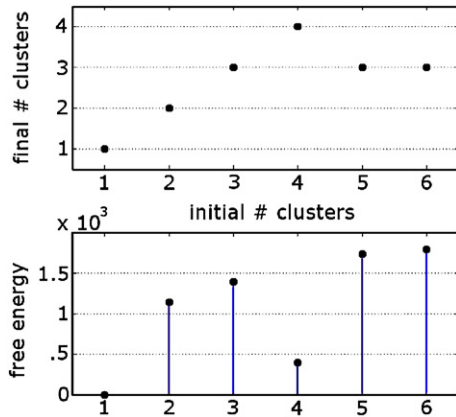
Fig. 3. The figures show the negative variational free energy (bottom) and the number of active meso-sources after ARD (top) as a function of the initial number of meso-sources. The best model is the sixth one (bottom), which, after ARD removal of spurious dimensions, is associated to only $K=3$ mesosources (top). Note that the three best models (models 3, 5 and 6) were all ARD-reduced to the $K=3$ model.

shifted to the right. Table 1 details the medians of these scores for these inverse methods. The MSM is the most accurate (lowest SSE, and highest ROC and CC). Notably, the LORETA method gave the worst results, irrespective of the true number of active sources; both in terms of spatial deployment and dynamics. The ReML scheme, which is a generalization of the standard methods LORETA, MNE and WMNE, performs better that the other methods, but fails to attain the accuracy of MSM. This means that the ability to estimate dipole dynamics improves significantly when augmenting a "naive" spatiotemporal structure with meso-scopic constraints.

Application to real data

To illustrate MSM in a realistic context, we used EEG data measuring an event-related response to the visual presentation of scrambled faces, which actually constituted the control condition in a face-recognition study.

The EEG data were acquired on a 128-channel Active Two System, sampled at 2048 Hz, plus electrodes on the left earlobe, right earlobe, and two measuring HEOG and VEOG. The data were epoched and re-referenced to the average. Trials with artefacts were rejected; artefacts were defined as time-points exceeding an absolute threshold of 120 μV (mainly in the VEOG), or expressing a significant linear drift (correlation >.8) over the epoch. A total of 29 of the 172 trials were rejected due to artefacts.

An anatomical MRI of the subject (acquired at 1.5 T) was also available, with a resolution of 1 mm × 1 mm × 1 mm. The cortical surface was segmented using the SPM5 (http://www.fil.ion.ucl.ac.uk/spm), and co-registered to the EEG frame of reference. The cortical surface was then down-sampled to 4000 faces using the shape-preserving reducepatch.m function in MATLAB. The gain matrix was calculated using a three-sphere analytical model using the SPM5 software.

Fig. 8 shows the resulting ERP. It exhibits three peaks of activity, the first at 110 ms, the second 170 ms, and the third 230 ms following stimulus onset.

The MSM was inverted using VB as described above and model comparison was performed over $K=1$ to $K=8$. Fig. 9 shows

the free energies of the ensuing models. The model with $K^*=3$ meso-sources had the highest negative free energy, and was selected as the most plausible solution.

Fig. 10 shows the PPM of being active, as well as the centroids of the active meso-sources for the $K^*=3$ model. Interestingly, this map showed very little variability, when varying K. Figs. 11 and 12 show respectively, the estimated mesostates (meso-source dynamics) and the most probable meso-source responsibilities (from the labelling processes). Two regions were found in the left and right occipital lobes. Their time courses show a damped oscillation which explains the three peaks observed in the EEG scalp data. Note that the left-occipital meso-source elaborates its response sooner than the right-occipital area. The third region actually lies on the right frontal lobe. It is activated largely during the second peak in the scalp EEG.

Discussion

We have described a generic scheme within which to introduce prior beliefs on the spatiotemporal dynamics of cortical activity causing measured EEG and MEG signals. This mesostate-space
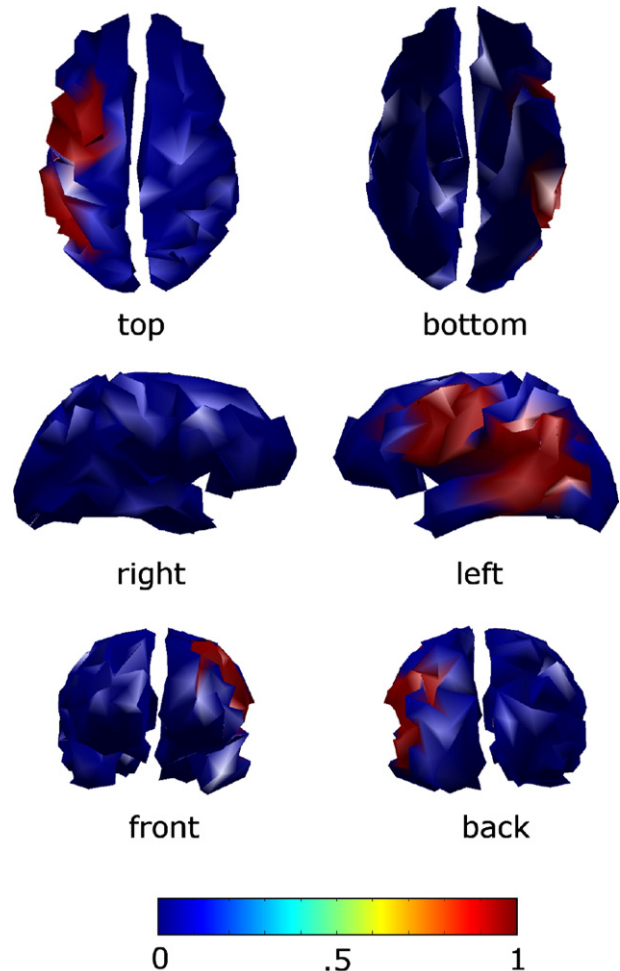

Fig. 4. PPM associated with the VB inversion of the mesostate-space model: best model ($K^*=3$). The probability of a source or mesh-dipole being active is one minus the probability it belongs to the ($K+1$)th inactive ensemble. These probabilities are shown in our PPMs; with red denoting an active source and blue an inactive source.
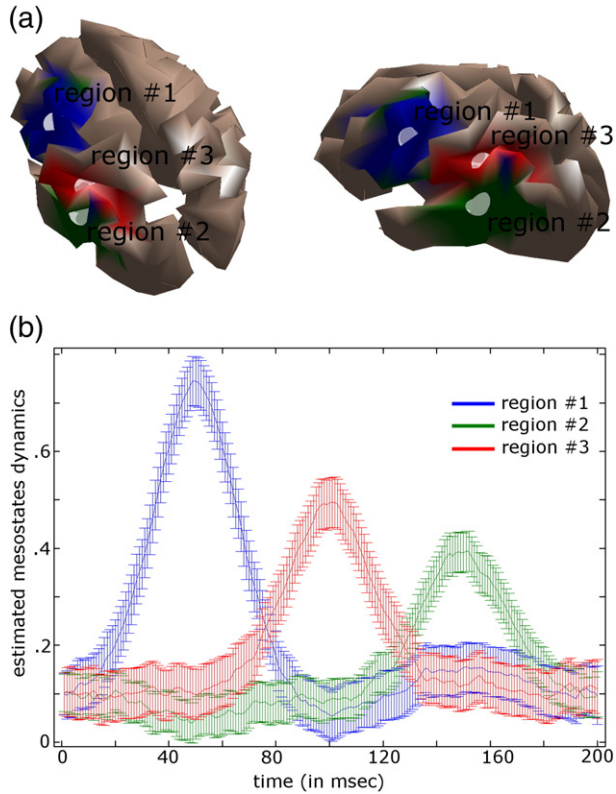
Fig. 5. VB inversion of the mesostate-space model: best model ($K^* = 3$). (a) Multinomial labelling process (cream: inactive ensemble, blue/red/green: active meso-sources); (b) estimated mesostates dynamics. The spheres show the 99% confidence interval on the estimated mean position of each cluster.

model (MSM) is formulated as a stochastic dynamical system that governs dynamics at the mesoscale. This system is cast as first-order Markov process prior, which can be inverted using standard (variational) Bayesian techniques. A mesoscale representation means that we can distinguish between the dynamics of meso-sources and individual cortical dipoles. This means that dipoles can be assigned probabilistically to a small number of meso-sources and that number can be optimized using model selection.

Inference on the model

When specifying a generative model for data it is necessary to define the form of the model in terms of specific quantities and how these quantities depend upon each other. These dependencies can make it difficult to deconstruct the model and assess which components are useful and which are not. For example, in our model are the temporal priors a useful addition to the model or do they induce unnecessary complexity? The evaluation of different model components can, however, proceed in a principled way using the marginal-likelihood or model-evidence. This involves switching off various model components by choosing different hyper-priors and then evaluating the models with and without a particular component in terms of their relative marginal-likelihood, as described above. For example, we could switch off the temporal priors by imposing very precise hyper-priors that make the elements of the matrix **A** zero. This means that the temporal

dynamics are modelled entirely by the random innovation and, a priori, would have no temporal structure. This an advantage of having a full generative model with hierarchically organized (hyper) priors. We will pursue this in subsequent cross-validation papers which will compare our mesoscopic model to alternative models based upon classical weighted minimum-norm solutions with multiple covariance components (Daunizeau et al., in preparation).

Inference on the model (or the parameters of any given model) requires a specification of our uncertainty about competing models or the values of their parameters. At the level of parameters, this is achieved simply by reference to the conditional distribution of the parameters: the conditional uncertainty is a variational free parameter, which is optimized during inversion. The same principle can be applied to model-space and implicitly those model parameters that specify different models. This involves computing the conditional probability over a set of models using their marginal likelihoods. Under at priors on the models, the conditional probability mass on each model is proportional to the marginal likelihood, where this mass sums to one. This can be particularly useful when each model is indexed by a categorical parameter. An example of this would be the number of meso-sources K. If we consider that the ith model corresponds to $K = i$, then we can reformulate the profile of model-evidences in Fig. 3 as the conditional probability of K over the values of K for which the model evidence was evaluated. In other words, Fig. 3 encodes explicitly our uncertainty about the number of meso-sources. It can be seen that the $K = 3$ model is the most likely, given our data and that this is $\exp(0.2 \times 10^3)$ times more likely than the $K = 7$ model! This sort of analysis can be pursued for any model parameter that defines a point mass in model space (i.e. when the parameter is a discrete variable). This highlights the close connection between conditional inference on model and parameter space, through optimization of the same free-energy functional.

Having said that, it is important to understand the limitation of the model-evidence or marginal-likelihood and its variational free-energy approximation. These quantities are based upon the probability of the observed data given a specific model or family of models. They only have meaning in relation to the model or
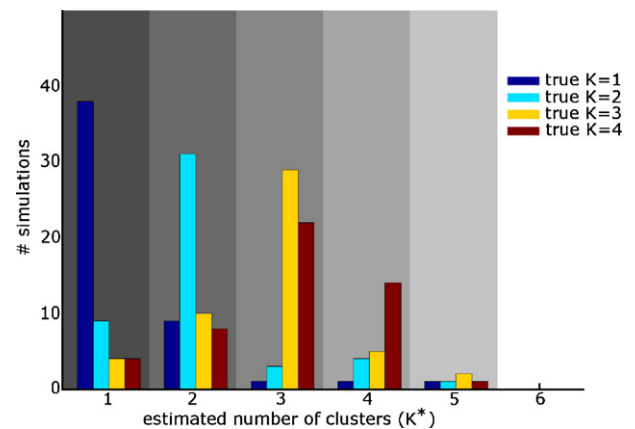

Fig. 6. Performance of model selection. The figure shows a histogram of final (inferred) number of active meso-sources K^* for all simulations (x-axis: number of final active regions; y-axis: number of simulations). The four colours correspond to different numbers of true active meso-sources (series S1: blue, S2: turquoise, S3: yellow, S4: red).
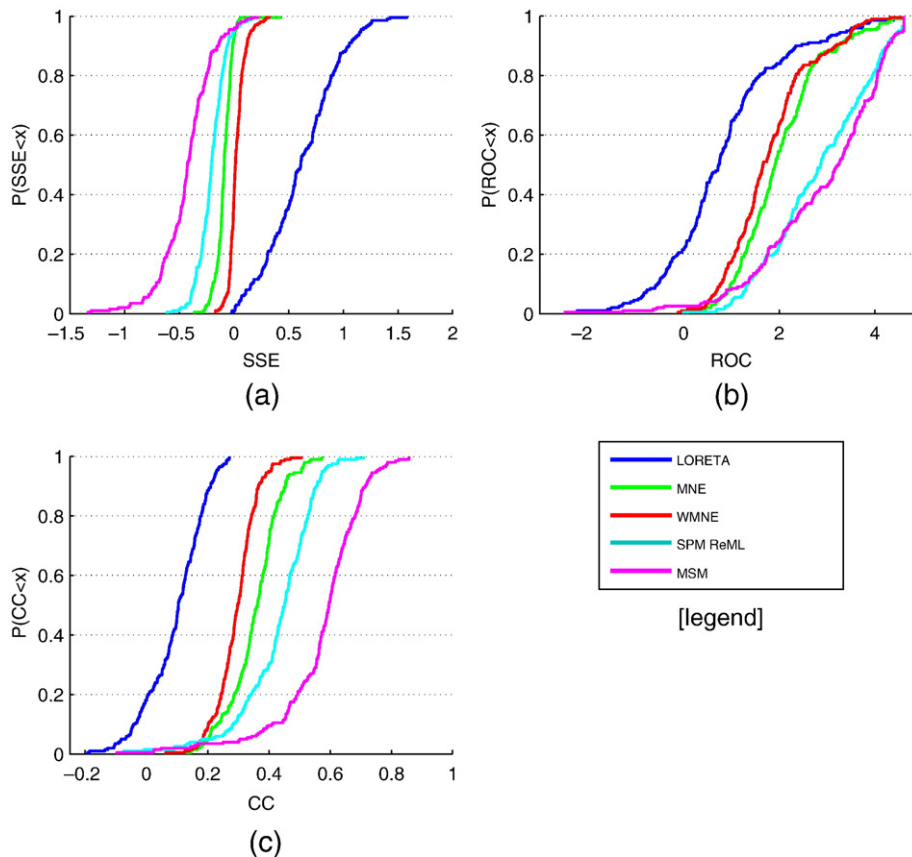
Fig. 7. Comparison of the five inverse methods in terms of their ability to recover the dipole dynamics. The figure shows empirical cumulative distribution functions (cdf) of the adequacy scores for all Monte-Carlo simulations (x-axis: adequacy score; y-axis: cdf). The five colours correspond to the four inverse methods (MSM: pink; ReML: blue turquoise; MNE: green; WMNE: red; LORETA: blue sea). (a) SSE score (the x-axis corresponds to the log-transformed sum of squared errors SSE, see the Comparisons section); (b) ROC score (the x-axis corresponds to $\ln(P/(1-P))$, where P is the area under the ROC curve); (c) CC score (the x-axis corresponds to the correlation coefficient between true and estimated dipole-wise dynamics). Except for the SSE score (a), better performance is indicated by curves that are shifted to the right.

models considered and can only be interpreted in a relative sense, in terms of model comparison. Clearly, all the models that one might consider will be suboptimal, compared to the unknown best model. However, the marginal likelihood can be very useful when trying to explore aspects of model space. Another important issue here is that the model-evidence depends upon the data. The best model for one particular dataset will be different from the best model for another dataset. For example, if we had access to intra-cranial recordings or local field potentials, the best models might be much more complicated and realistic, as assessed by the model-evidence. In short, the model-evidence is really a statement about the relative goodness of models and how those models are informed by the particular data that are available.

Table 1
Medians of the adequacy scores (SSE, ROC and CC) based on the dipole dynamics **J** for all 200 simulations

	LORETA	MNE	WMNE	ReML	MSM
SSE	0.61	−0.09	0.01	−0.21	−0.43
ROC	0.74	1.92	1.66	2.87	3.23
CC	0.10	0.36	0.30	0.45	0.60

Relation to existing approaches

Our generative model and its variational inversion belongs to the family of parametric empirical Bayes (PEB) approaches (Friston et al., 2002). PEB requires a hierarchical observation model where the parameters and hyperparameters at any particular level can be treated as priors on the level below. In the context of the EEG inverse problem, this modelling strategy has already been employed successfully in (Mattout et al., 2006; Phillips et al., 2002) for constraining the spatial deployment of distributed sources. These models parameterize the prior covariance of dipole dynamics with a linear mixture of a priori covariance components. The ensuing ReML (a variant of expectation maximization; EM) hyperparameters estimates are then used to construct an optimal approximation to the posterior covariance structure.

Despite the MoG form of the MSM, the variational marginal pdf of the dipole dynamics is Gaussian. As a consequence, the VB posterior covariance estimate of the MSM also relies on a linear mixing of covariance components (see Eq. (39)). In other words, inversion of the MoG prior can be reduced to an empirical Bayesian estimation of a single parametric block-diagonal covariance matrix.

When inverting the MSM, the scheme estimates the posterior probability $p(\xi_{i,K}=1|\mathbf{y})$, that the ith dipole belongs to the cluster
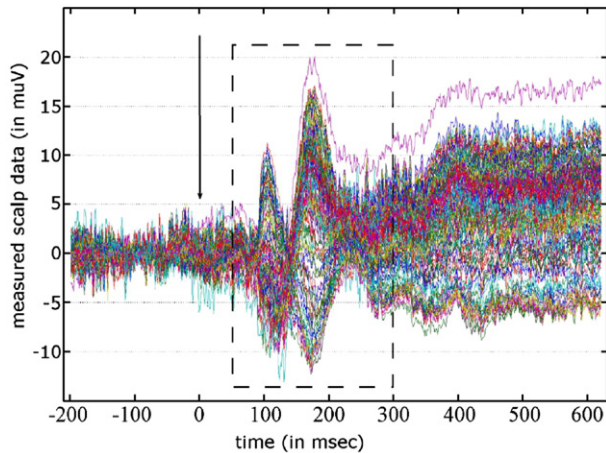
Fig. 8. EEG event-related potential, following "scrambled faces" visual stimulation. These scalp measurements exhibit three peaks, at 110, 170 and 230 ms after stimulus onset (depicted by the arrow). The scalp data actually analysed are depicted using the dashed rectangle.

having zero mean (the background or inactive ensemble). As the precision α_K increases, the (K)th Gaussian of the mixture asymptotes to a Dirac distribution. At the zero variance limit $(\alpha_K \rightarrow \infty)$, the background dipoles do not contribute to the measured signal. Therefore, the quantity $p(\xi_{i,K}=1|\mathbf{y})$ can be considered as the posterior probability of the ith dipole being inactive. Similarly, $1-p(\xi_{i,K}=1|\mathbf{y})$ is the posterior probability of it being active. This allows us to assess the significance of any dipole being active, i.e. form a posterior probability map (PPM) on dipole space (Friston and Penny, 2003).

On evaluation results

So far, we have focused on the added-value of the MSM as a tool for EEG/MEG source reconstruction. Hence, we used minimal assumptions regarding the expected temporal behaviour of the mesostates, i.e. we only used a smoothness prior. However, it is possible to incorporate more constraints by specifying a more realistic dynamic causal model on the interactions between meso-

Fig. 10. Posterior probability map (PPM) of being active (blue=1, red=0) for the $K^*=3$ model. The probability of a source or mesh-dipole being active is one minus the probability that it belongs to the $(K+1)$th inactive ensemble. This map identifies three active meso-sources in the left and right occipital lobes, and in the right frontal lobe.

sources. These models are currently being evaluated in a non-stochastic setting using DCM for ERPs.

The MSM has been evaluated and compared to standard EEG inverse approaches (MNE, WMNE and LORETA) using simulated data. These simulations demonstrate the added-value of the approach w.r.t. accuracy. This is probably due to the nature of the generative model: the better match between the prior assumptions and the underlying structure of brain activity, the better the inversion. In other words, the gain in precision afforded by the prior information, introduced through the meso-sources, more than compensates for the increase in complexity.

Moreover, the evaluation part of this work suggests that the method is able to correctly identify the number, position, spatial extent and dynamics of the underlying active meso-sources. In addition, the variational free energy enabled us to select the optimal model given the measurements. This model selection approach, accounting for any uncertainty about the parameterization of the model, is likely to be important when comparing more complex models (see next section).
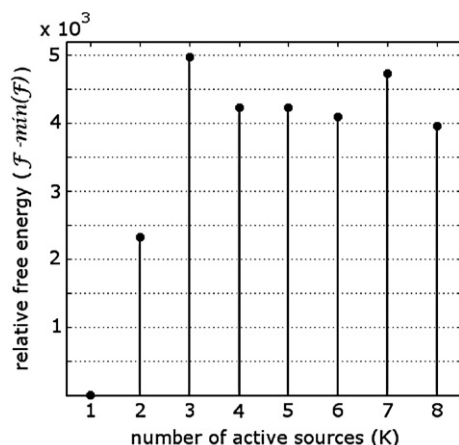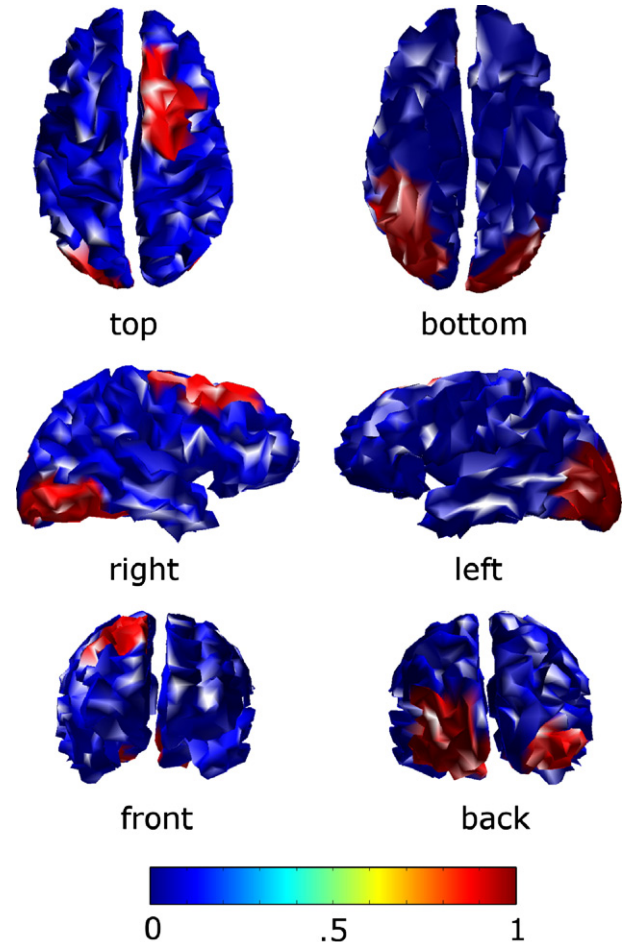

Fig. 9. Variational negative free energy of the eight models with different numbers of active meso-sources. The ARD scheme did not eliminate any component (final and initial K values are identical). Nevertheless, the $K^*=3$ model exhibits a significantly higher negative free energy.
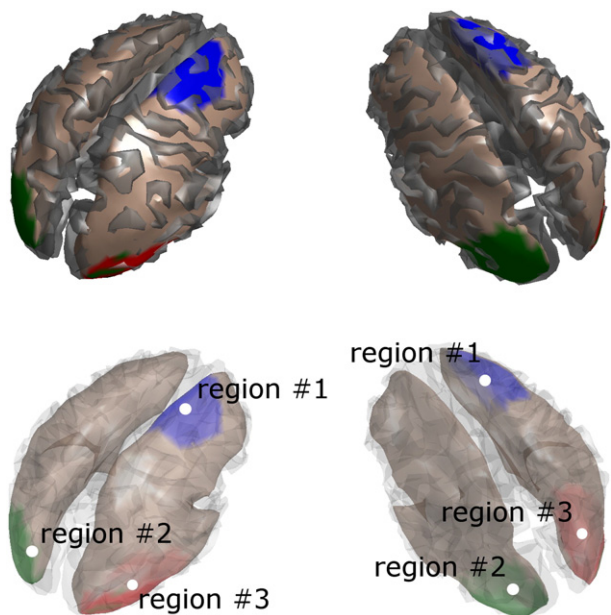
Fig. 11. Spatial distribution of the most plausible multinomial labelling process (ensemble responsibilities) for the $K^* = 3$ model. This map is built by assigning each dipole to the meso-source it is most likely to belong to. The labelling process is displayed on the inflated cortical mesh, with the high-resolution cortex shown in transparency (cream: inactive ensemble, blue/green/red: active meso-sources). The white spheres correspond to the estimated mean position and 99% confidence intervals for each meso-source.

Lastly, although the model comparison capabilities of the methods prove able to correctly detecting the number of meso-sources for most of encountered situations, it definitely exhibited a slight undercomplexity bias (see Fig. 3). An intuitive understanding for this bias may be derived the following way: Recall that the free energy approximation to the log model evidence (or so-called "marginal likelihood") is actually a lower bound (cf. Eq. (18)):

$$\mathcal{L}(K) \equiv \ln p(\mathbf{y}|\mathcal{M}, K) = \mathcal{F}(q, K) + \mathcal{D}_{KL}(q, p),$$

where $\mathcal{L}(K)$ (respectively $\mathcal{F}(q, K)$) is the true marginal likelihood (resp. variational free energy) for a model with K components, and $\mathcal{D}_{KL}(q, p)$ is the KL divergence between the true posterior and its variational (mean-field) approximation.

Let us assume that every component (meso-source) in our model contributes a constant KL divergence penalty of $\mathcal{D}$. The difference in log marginal likelihoods, $\mathcal{L}(K + 1) - \mathcal{L}(K)$, is the quantity we wish to estimate, but if we base this on the lower bounds the difference becomes:

$$\begin{aligned} \mathcal{L}(K + 1) - \mathcal{L}(K) &= (\mathcal{F}(K + 1) + (K + 1)\mathcal{D}) - (\mathcal{F}(K) + K\mathcal{D}) \\ &= \mathcal{F}(K + 1) - \mathcal{F}(K) + \mathcal{D} \\ &\neq \mathcal{F}(K + 1) - \mathcal{F}(K) \end{aligned}$$

where the last line is the result we would have basing the difference on lower bounds. Therefore, if each component contributes independently to the KL divergence term, and since the KL divergence is strictly positive, the use of free energy for model comparison might introduce a systematic bias towards simpler models.

Perspectives

In the context of multimodal information fusion, the MSM model can be easily extended by introducing priors onto the position of each active region. This way of introducing fMRI priors allows us to bypass standard weighted minimum norm techniques, which require the delicate parameterization of the prior variances of the sources.[6]

As mentioned above (see the Mesostate dynamics section), the natural extension of the MSM approach is the incorporation of neural mass models to prescribe interactions among mesostates.

Neural mass models describe the dynamics of an ensemble of atoms (macro-columns) as a point mass, i.e. the mean or expected dynamics of a population (David et al., 2006). In our formulation we allow for random fluctuations about this mass to provide a stochastic dynamic causal model for neuronal interactions (see the Mesostates and meso-sources section). The use of stochastic DCM priors, as part of the MSM would allow one to infer not only the meso-source dynamics but also the coupling parameters shaping interactions among meso-sources. In fact, under the mean-field approximation (see Variational inversion section), exactly the same variational inversion can be used if the evolution function can be written as a *multilinear* function of joint hidden states and parameters (see the Mesostate dynamics section; Beal, 2003).

Notably, if the neural mass model can not be formulated using a multilinear evolution function, we may then resort to, e.g. Laplace approximations to the formal Variational Bayesian learning scheme (Friston et al., in preparation).

Two interesting applications of this scheme are worth mentioning:

1. stochastic DCM (Friston et al., 2003). This family of models would encompass four basic characteristics of neuronal activity and organization: (i) neurons are dynamic units, (ii) driven by stochastic forces, (iii) organized into populations with similar biophysical properties and response characteristics and (iv) multiple populations interact to form functional networks (Harrison et al., 2005);
2. EEG/fMRI fusions (Riera et al., 2006). Noticeably, recent advances in understanding physiological mechanisms at different spatiotemporal scales have provided a framework within which to develop sophisticated biophysical models that permit an integration of different imaging modalities, each sharing a common aetiology (Riera et al., 2005). More precisely, evolution/observation equations encoding the relationship between bioelectric and hemodynamic mesostates can be motivated using both physiological and physical facts (Friston et al., 2000; Kilner et al., 2005).

Variational Bayesian inference schemes allow one to compare different models. This comparison can be used to select any dynamical model introduced through the state-space formulation in a principled way. For example, different wirings of the mesoscopic network (e.g. specific assumptions regarding the hierarchy within

---

[6] With a fixed zero-mean prior, the prior variance of the source is the prior *power* of the sources. In weighted MNE-like methods, fMRI maps of activation are used as a shrinkage prior of the spatial distribution of power (see Daunizeau et al., 2005).
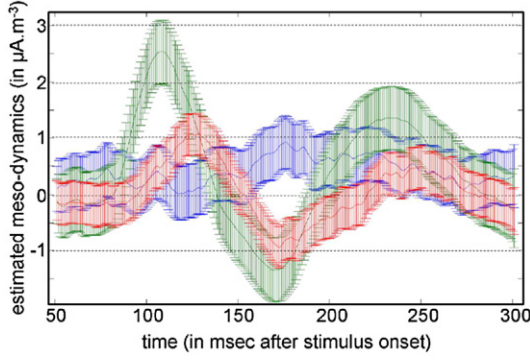
Fig. 12. Temporal dynamics of the mesostates for the $K^*=3$ model (active meso-source time courses). The error bars depict the 99% confidence intervals. The left and right occipital active areas exhibit a damped oscillation that may be responsible for the three peaks seen in the scalp EEG (see Fig. 8). The right frontal meso-source is active during the second peak (N170).

the network) can be compared, or different candidate processes for neurovascular coupling could be tested.

Appendix A

This appendix focuses on the variational inversion used to derive the marginal posterior pdf of each node of the MSM. All the densities belong to the exponential family: they are parameterized using a finite number of so-called sufficient statistics.[7] These sufficient statistics are the only quantities that we need to update during the iterations of the VB algorithm.

A.1. Sufficient statistics

For notational convenience, define the following matrices as:

$$\tilde{\mathbf{Y}} = \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t^T$$

$$\dot{\mathbf{Y}} = \sum_{t=1}^{T} \langle \mathbf{J}_t \rangle \mathbf{y}_t^T$$

$$\tilde{\mathbf{J}} = \sum_{t=1}^{T} \langle \mathbf{J}_t \mathbf{J}_t^T \rangle$$

$$\tilde{\mathbf{M}} = \sum_{t=1}^{T} \langle \boldsymbol{\mu}_t \boldsymbol{\mu}_t^T \rangle$$

$$\dot{\mathbf{M}} = \sum_{t=1}^{T} \langle \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}_t^T \rangle$$

$$\check{\mathbf{M}} = \sum_{t=1}^{T} \langle \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}_{t-1}^T \rangle = \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^T + \sum_{t=1}^{T-1} \langle \boldsymbol{\mu}_t \boldsymbol{\mu}_t^T \rangle$$

$$\mathbf{S} = \sum_{t=1}^{T} \langle \mathbf{J}_t \boldsymbol{\mu}_t^T \rangle = \sum_{t=1}^{T} \langle \mathbf{J}_t \rangle \langle \boldsymbol{\mu}_t \rangle^T$$

$$\mathbf{Z} = \langle \boldsymbol{\xi} \rangle$$

$$\tilde{\mathbf{X}} = \mathbf{X}^T \mathbf{X}$$

$$\tilde{\mathbf{N}} = \mathrm{Diag}(\mathrm{diag}(\langle \boldsymbol{\eta}^T \boldsymbol{\eta} \rangle))$$

$$\mathbf{T} = \mathbf{X}^T \langle \boldsymbol{\eta} \rangle \tag{24}$$

----
[7] These statistics are quantities that define a density given its form. They are usually linked to the first two moments of the distribution.

These (conditional covariance) matrices are necessary for the deriving the variational marginal posterior pdf of the model parameters.

A.2. Precision parameters

A.2.1. Measurement noise variance
The Markov neighborhood of σ is $\mathbf{y}_t$, $\mathbf{J}_t$, $t=1, \ldots, T$:

$$\ln q(\sigma) = \langle \ln p(\sigma | \mathbf{y}, \mathbf{J}, \cdot) \rangle_{q(\mathbf{J})} + cst. \tag{25}$$

Therefore, σ is Gamma-distributed:

$$q(\sigma) = \mathcal{G}(a_\sigma, b_\sigma) \Rightarrow \langle \sigma \rangle = \hat{\sigma} = \frac{a_\sigma}{b_\sigma}, \tag{26}$$

with parameters (a_σ, b_σ) such that:

$$\begin{cases} a_\sigma = \dfrac{p(T + t_0)}{2} \\ b_\sigma = \dfrac{1}{2}\left(\mathrm{tr}[\tilde{\mathbf{Y}}] - 2\mathrm{tr}[\mathbf{G}\dot{\mathbf{Y}}] + \mathrm{tr}[\mathbf{G}^T \mathbf{G}\tilde{\mathbf{J}}] + \mathrm{tr}[\mathbf{y}_0^T \mathbf{y}_0]\right). \end{cases} \tag{27}$$

A.2.2. System noise variance
The Markov neighborhood of β is $\boldsymbol{\mu}_t$, $t=1, \ldots, T$:

$$\ln q(\beta) = \langle \ln p(\beta | \boldsymbol{\mu}, \cdot) \rangle_{q(\boldsymbol{\mu})} + cst. \tag{28}$$

Therefore, β is Gamma-distributed:

$$q(\beta) = \mathcal{G}(a_\beta, b_\beta) \Rightarrow \langle \beta \rangle = \hat{\beta} \frac{a_\beta}{b_\beta}, \tag{29}$$

with parameters (a_β, b_β) such that:

$$\begin{cases} a_\beta = \dfrac{KT}{2} \\ b_\beta = \dfrac{1}{2}\left(\mathrm{tr}[\tilde{\mathbf{M}}] - 2\mathrm{tr}[\mathbf{A}\dot{\mathbf{M}}] + \mathrm{tr}[\mathbf{A}^T \mathbf{A}\check{\mathbf{M}}]\right). \end{cases} \tag{30}$$

A.2.3. Mesostate temporal precision parameters
The Markov neighborhood of α_k is $\{\boldsymbol{\xi}, \boldsymbol{\mu}_t, \mathbf{J}_t, t=1, \ldots, T\}$. Hence:

$$\ln q(\alpha_k) = \langle \ln p(\alpha_k | \mathbf{J}, \boldsymbol{\mu}, \boldsymbol{\xi}, \cdot) \rangle_{q(\boldsymbol{\mu})q(\mathbf{J})\tilde{q}(\boldsymbol{\xi})} + cst. \tag{31}$$

Therefore, α_k is Gamma-distributed:

$$q(\alpha_k) = \mathcal{G}(a_\alpha^k, b_\alpha^k) \Rightarrow \langle \alpha_k \rangle = \frac{a_\alpha^k}{b_\alpha^k}, \tag{32}$$

with parameters (a_α^k, b_α^k) such that:

$$\begin{cases} a_\alpha^k = \dfrac{T}{2}\mathbf{Z}_k^T 1_n \\ b_\alpha^k = \dfrac{1}{2}\left(\mathbf{Z}_k^T\left(\mathrm{diag}(\tilde{\mathbf{J}}) + \tilde{M}_{kk} 1_n - 2S_k\right)\right) \end{cases}. \tag{33}$$

Note that for the inactive ensemble: $\tilde{M}_{KK}=0$ and $\mathbf{S}_K=0$. In the following, we will use the vector forms to denote the expectation (under the Gamma posterior pdf) of the mesostate precisions:

$$\alpha = \left(\frac{a_\alpha^k}{b_\alpha^k}\right)_{1 \le k \le K}.$$

A.2.4. Mesostate spatial precision parameters
The Markov neighborhood of γ_k is $\{\boldsymbol{\xi}, \boldsymbol{\eta}, \mathbf{X}\}$. Hence:

$$\ln q(\gamma_k) = \langle \ln p(\gamma_k | \mathbf{X}, \boldsymbol{\eta}, \boldsymbol{\xi}, \cdot) \rangle_{q(\boldsymbol{\eta})q(\boldsymbol{\xi})} + cst. \tag{34}$$

Therefore, γ_k is Gamma-distributed:

$$q(\gamma_k) = \mathcal{G}\left(a_\alpha^k, b_\alpha^k\right) \Rightarrow \langle \gamma_k \rangle = \frac{a_\gamma^k}{b_\gamma^k}, \tag{35}$$

with parameters (a_γ^k, b_γ^k) such that:

$$\begin{cases} a_\gamma^k = \dfrac{3}{2}\mathbf{Z}_k^T 1_n + a_{\gamma,0}^k \\ b_\gamma^k = \dfrac{1}{2}\mathbf{Z}_k^T \left(\mathrm{diag}(\tilde{\mathbf{X}}) + \tilde{N}_{kk} 1_n - 2\mathbf{T}_k\right) + b_{\gamma,0}^k. \end{cases} \tag{36}$$

Again, for convenience, we will use the vector forms to denote the expectation (under the Gamma posterior pdf) of the precisions:

$$\hat{\gamma} = \left(\frac{a_\gamma^k}{b_\gamma^k}\right)_{1 \le k \le K}.$$

A.3. Dipole dynamics

The Markov neighborhood of $\mathbf{J}_t$ is $\{\boldsymbol{\xi}, \mathbf{y}_t, \boldsymbol{\mu}_t, \alpha, \sigma\}$:

$$\ln q(\mathbf{J}_t) = \langle \ln p(\mathbf{J}_t | \boldsymbol{\xi}, \boldsymbol{\mu}_t, \alpha, \sigma) \rangle_{q(\boldsymbol{\xi})q(\boldsymbol{\mu}_t)q(\alpha)q(\sigma)} + cst. \tag{37}$$

Eq. (37) is a quadratic form in $\mathbf{J}_t$. Therefore, under the mean-field approximation, the variational marginal of $\mathbf{J}_t$ is Gaussian: $q(\mathbf{J}_t) = \mathcal{N}(\hat{\mathbf{J}}_t, \mathbf{W})$, with:

$$\hat{\mathbf{J}}_t = \mathbf{W}\left(\frac{a_\sigma}{b_\sigma}\mathbf{G}^T \mathbf{y}_t + \mathbf{Z}\mathrm{Diag}(\hat{\alpha})\begin{bmatrix} \mathbf{m}_t \\ 0 \end{bmatrix}\right), \tag{38}$$

where the posterior covariance matrix $\mathbf{W}$ is defined as:

$$\begin{aligned} \mathbf{W} &= (\hat{\sigma}\mathbf{G}^T\mathbf{G} + \mathrm{Diag}(\mathbf{Z}\hat{\alpha}))^{-1} \\ &= \mathrm{Diag}(\mathbf{Z}\hat{\alpha})^{-1}\left(\mathbf{I}_n - \mathbf{G}^T\left(\frac{1}{\hat{\sigma}}\mathbf{I}_p + \mathbf{G}\mathrm{Diag}(\mathbf{Z}\hat{\alpha})^{-1}\mathbf{G}^T\right)^{-1}\mathbf{G}\mathrm{Diag}(\mathbf{Z}\hat{\alpha})^{-1}\right) \end{aligned} \tag{39}$$

where the second line requires only a $p \times p$ matrix inversion.

The sufficient statistics $\dot{\tilde{\mathbf{J}}}$ and $\dot{\mathbf{Y}}$ are:

$$\begin{cases} \dot{\mathbf{Y}} = \hat{\mathbf{J}}\mathbf{y}^T \\ \tilde{\mathbf{J}} = T\mathbf{W} + \hat{\mathbf{J}}\hat{\mathbf{J}}^T. \end{cases} \tag{40}$$

A.4. The labelling parameters

A.4.1. The class frequencies

The Markov neighborhood of $\boldsymbol{\lambda}$ is $\boldsymbol{\xi}$:

$$\ln q(\boldsymbol{\lambda}) = \langle \ln p(\boldsymbol{\lambda}|\boldsymbol{\xi}, \cdot) \rangle_{q(\boldsymbol{\xi})}. \tag{41}$$

The class frequencies λ_k, $k = 1, \ldots, K$ are, by definition, frequencies, i.e. they satisfy the normalization constraint given by Eq. (11). Therefore, together with the Eq. (41), we have a Dirichlet posterior pdf for $\boldsymbol{\lambda}$:

$$q(\boldsymbol{\lambda}) = \mathrm{Dir}(\mathbf{c}), \tag{42}$$

where the $(K) \times 1$ vector $\mathbf{c}$ is such that:

$$\mathbf{c} = \mathbf{b} + \mathbf{Z}^T 1_n. \tag{43}$$

Therefore, under the normalization constraint, the expectation of $\boldsymbol{\lambda}$ is:

$$\langle \boldsymbol{\lambda} \rangle = \frac{1}{c_0}\mathbf{c}, \tag{44}$$

where c_0 is such that:

$$c_0 = \sum_{k=1}^{K} c_k. \tag{45}$$

A.4.2. Labelling process

The Markov neighborhood of $\boldsymbol{\xi}$ is $\{\mathbf{J}, \mathbf{X}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\eta}\}$. Hence:

$$\ln q(\boldsymbol{\xi}) = \langle \ln p(\boldsymbol{\xi}|\mathbf{J}, \mathbf{X}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\eta}, \cdot) \rangle_{q(\mathbf{J})q(\boldsymbol{\lambda})q(\boldsymbol{\alpha})q(\boldsymbol{\gamma})q(\boldsymbol{\mu})q(\boldsymbol{\eta})} \tag{46}$$

By expanding the right-hand term of Eq. (46), one notes that there is no cross-term between any of the columns $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$ of $\boldsymbol{\xi}$. Therefore, the joint posterior of the columns $\boldsymbol{\xi}_i$ is the product of the marginals. Since for each dipole, the probability of belonging to all classes should sum to one, the following constraint holds:

$$1 = \sum_{k=1}^{K} q(\xi_{ik} = 1), \quad i = 1, \ldots, n. \tag{47}$$

The functional form of the variational marginal on $\boldsymbol{\xi}_i$ is then given by:

$$\ln q(\boldsymbol{\xi}_i) = \sum_{k=1}^{K} \xi_{ik} \ln \pi_{ik} + cst, \tag{48}$$

with π_{ik} being defined as:

$$\begin{aligned} \pi_{ik} = \exp&\left(\psi(c_k) - \psi(c_0) + \frac{T}{2}\left(\psi(a_\alpha^k) - \ln b_\alpha^k\right) + \frac{3}{2}\left(\psi(a_\gamma^k) - \ln b_\gamma^k\right)\right) \\ \times \exp&\left(-\frac{1}{2}\frac{a_\gamma^k}{b_\gamma^k}\left(\tilde{X}_{ii} + \tilde{N}_{kk} - 2T_{ik}\right) - \frac{1}{2}\frac{a_\alpha^k}{b_\alpha^k}\left(\tilde{J}_{ii} + \tilde{M}_{kk} - 2S_{ik}\right)\right), \end{aligned} \tag{49}$$

where $\psi(\cdot)$ is the digamma function.

Together with the normalization constraint given by Eq. (47), Eq. (48) states that $\boldsymbol{\xi}_i$ follows a multinomial distribution such that:

$$q(\boldsymbol{\xi}_i) \propto \prod_{k=1}^{K} \pi_{ik}^{*\xi_{ik}} = \mathrm{Multinom}(1, \{\pi_{ik}^*\}_{1 \le k \le K}), \tag{50}$$

where the π_{ik}^* are derived from the normalization constraint:

$$\pi_{ik}^* = \frac{\pi_{ik}}{\sum_{k=1}^{K} \pi_{ik}} = q(\xi_{ik} = 1) = \langle \xi_{ik} \rangle = Z_{ik}. \tag{51}$$

A.4.3. The mean spatial position

The Markov neighborhood of $\boldsymbol{\eta}_k$ is $\{\mathbf{X}, \gamma_k, \boldsymbol{\xi}\}$:

$$\ln q(\boldsymbol{\eta}_k) = \langle \ln p(\boldsymbol{\eta}_k|\mathbf{X}, \boldsymbol{\xi}, \gamma, \cdot) \rangle_{q(\mathbf{X})q(\gamma)q(\boldsymbol{\xi})} + cst, \quad k = 1, \ldots, K. \tag{52}$$

Eq. (52) is a quadratic form in $\boldsymbol{\eta}_k$. Therefore, the variational posterior pdf of $\boldsymbol{\eta}_k$ is Gaussian: $q(\boldsymbol{\eta}_k) = \mathcal{N}(\hat{\boldsymbol{\eta}}_k, \mathbf{V}^{(k)})$ with:

$$\begin{cases} \hat{\boldsymbol{\eta}}_k = \dfrac{1}{\mathbf{Z}_k^T \mathbf{1}_n} \mathbf{X} \mathbf{Z}_k \\[2mm] \mathbf{V}^{(k)} = \dfrac{1}{\mathbf{Z}_k^T \mathbf{1}_n} \dfrac{1}{\hat{\gamma}_k} \mathbf{I}_3 \end{cases}. \tag{53}$$

This states that the mean position of each meso-source is simply the weighted (according to the posterior probabilities $\mathbf{Z}$) average position of the constituent dipoles. When introducing a Gaussian prior on the position of meso-sources (as derived from fMRI, for example), this estimate is modified in proportion to the prior variance.

Therefore, the sufficient statistics are:

$$\begin{cases} \mathbf{T} = \mathbf{X}^T \hat{\eta} \\[2mm] \tilde{\mathbf{N}} = \mathrm{Diag}\left(\left(\hat{\boldsymbol{\eta}}_k^T \hat{\boldsymbol{\eta}}_k + \dfrac{3}{\hat{\gamma}_k \mathbf{Z}_k^T \mathbf{1}_n} \right)_{k=1,\dots,K} \right) \end{cases} \tag{54}$$

A.5. The mesostate dynamics: a variational Kalman smoother

The remaining variational marginal is $q(\boldsymbol{\mu}_{1:T})$. This posterior can be obtained by vectorizing the matrix $\boldsymbol{\mu}$, accounting for the inter time-step covariance structure (through the known matrix $\mathbf{A}$), and averaging the parameters of this (Gaussian) $(K-1)T-D$ pdf under the variational marginals of the Markov neighborhood of $\boldsymbol{\mu}_{1:T}$:

$$\ln q(\boldsymbol{\mu}_{1:T}) = \langle \ln p(\boldsymbol{\mu}_{1:T}|\mathbf{J}_{1:T}, \boldsymbol{\xi}, \beta, \alpha) \rangle_{q(\mathbf{J}_{1:T})q(\boldsymbol{\xi})q(\beta)q(\alpha)} \tag{55}$$

However, this is not generally feasible because of the dimensionality of the problem. Instead, we use a variational Kalman–Rauch smoother, which is computationally much more expedient because it only evaluates the (instantaneous) marginal $q(\boldsymbol{\mu}_t)$, as opposed to the joint posterior of the whole sequence $q(\boldsymbol{\mu}_{1:T})$.

Operationally, the variational Kalman smoother (Beal and Ghahramani, 2002) entails a two-step procedure involving:

1. the *forward pass*, which iterates the following equations from $t=2$ to $t=T$:

$$\begin{cases} \mathbf{R}_{t|t-1} = (1/\hat{\beta})\mathbf{I}_K + \mathbf{A}\mathbf{R}_{t-1|t-1}\mathbf{A}^T \\[1mm] \mathbf{m}_{t|t-1} = \mathbf{A}\mathbf{m}_{t-1}^* \\[1mm] \mathbf{R}_{t|t} = (\mathrm{Diag}(\mathrm{Diag}(\tilde{\boldsymbol{\alpha}})\tilde{\mathbf{Z}}^T \mathbf{1}_n) + \mathbf{R}_{t|t-1}^{-1})^{-1} \\[1mm] \mathbf{m}_t^* = \mathbf{R}_{t|t}(\mathbf{R}_{t|t-1}^{-1}\mathbf{m}_{t|t-1} + \mathrm{Diag}(\tilde{\boldsymbol{\alpha}})\tilde{\mathbf{Z}}^T \bar{\mathbf{J}}_t), \end{cases} \tag{56}$$

where $\tilde{\mathbf{Z}}$ and $\tilde{\boldsymbol{\alpha}}$ are truncated to include only "activated" meso-sources:

$$\begin{cases} \tilde{\mathbf{Z}} = \mathbf{Z}_{1:K-1} \\ \tilde{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}_{1:K-1}, \end{cases}$$

where K is the number of dipole ensembles;

2. the *backward pass*, which iterates the following equations from $t=T-1$ to $t=1$:

$$\begin{cases} \mathbf{K}_t = \mathbf{R}_{t|t}\mathbf{A}^T \mathbf{R}_{t+1|t}^{-1} \\[1mm] \boldsymbol{\Upsilon}_{t,t} = \mathbf{R}_{t|t} + \mathbf{K}_t(\boldsymbol{\Upsilon}_{t+1,t+1} - \mathbf{R}_{t+1|t})\mathbf{K}_t^T \\[1mm] \mathbf{m}_t = \mathbf{m}_t^* + \mathbf{K}_t(\mathbf{m}_{t+1} - \mathbf{A}\mathbf{m}_t^*) \\[1mm] \boldsymbol{\Upsilon}_{t,t+1} = \mathbf{R}_{t+1|t}\mathbf{A}^T \boldsymbol{\Upsilon}_{t+1,t+1}, \end{cases} \tag{57}$$

the last expression being necessary for both the sufficient statistics of the parameters of the evolution function and the calculation of the variational free energy (see Section A.6).

Using the Markov properties of the mesostate time series, it is relatively straightforward to show that the variational marginal $q(\boldsymbol{\mu}_t)$ of $\boldsymbol{\mu}_t$ is Gaussian such that:

$$q(\boldsymbol{\mu}_t) = \mathcal{N}(\mathbf{m}_t, \boldsymbol{\Upsilon}_{t,t}).$$

Then, the sufficient statistics of the mesostates are assembled through the backward pass as follows:

$$\begin{cases} \tilde{\mathbf{M}} = \displaystyle\sum_{t=1}^{T} \mathbf{m}_t \mathbf{m}_t^T + \boldsymbol{\Upsilon}_{t,t} \\[3mm] \dot{\mathbf{M}} = \displaystyle\sum_{t=1}^{T} \mathbf{m}_{t-1}\mathbf{m}_t^T + \boldsymbol{\Upsilon}_{t-1,t} \\[3mm] \check{\mathbf{M}} = \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^T + \displaystyle\sum_{t=1}^{T} \mathbf{m}_t \mathbf{m}_t^T + \boldsymbol{\Upsilon}_{t,t} \\[3mm] \mathbf{S} = \displaystyle\sum_{t=1}^{T} \hat{\mathbf{J}}_t \mathbf{m}_t^T \end{cases} \tag{58}$$

A.6. Derivation of the variational free energy

Under the mean field approximation, the calculation of the free energy requires the evaluation of the sum of the Shannon entropy of the marginal variational posterior pdfs. Except for the mesostates pdf, these are relatively straightforward. However, due to the use of the Kalman–Rauch scheme in the derivation of the posterior $q(\boldsymbol{\mu}_t|\mathbf{y}_{1:T})$, the derivation of the actual joint entropy $\mathcal{S}(q(\boldsymbol{\mu}_{1:T}))$ requires special considerations.

First, let us not note that the joint $q(\boldsymbol{\mu}_{1:T})$ factorizes over instantaneous transition pdf (Chapman–Kolmogorov equation):

$$\begin{aligned} q(\boldsymbol{\mu}_{1:T}) &= q(\boldsymbol{\mu}_1|\boldsymbol{\mu}_0)\prod_{t=2}^{T} q(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}) \\[2mm] &= q(\boldsymbol{\mu}_1)\frac{\prod\limits_{t=2}^{T} q(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t-1})}{\prod\limits_{t=2}^{T} q(\boldsymbol{\mu}_{t-1})}. \end{aligned} \tag{59}$$

Therefore, its entropy may be decomposed into:

$$\begin{aligned} \mathcal{S}(q(\boldsymbol{\mu}_{1:T})) = &-\sum_{t=2}^{T} \int \log q(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t-1})\mathrm{d}q(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t-1}) \\ &+ \sum_{t=2}^{T-1} \int \log q(\boldsymbol{\mu}_t)\mathrm{d}q(\boldsymbol{\mu}_t). \end{aligned} \tag{60}$$

Noting that:

$$q(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t-1}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_t \\ \mathbf{m}_{t-1} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Upsilon}_{t,t} & \boldsymbol{\Upsilon}_{t,t-1}^T \\ \boldsymbol{\Upsilon}_{t,t-1} & \boldsymbol{\Upsilon}_{t,t} \end{bmatrix} \right), \tag{61}$$

this gives:

$$\begin{aligned} \mathcal{S}(q(\boldsymbol{\mu}_{1:T})) = &\frac{1}{2}\sum_{t=1}^{T-1} \left(\log \begin{vmatrix} \boldsymbol{\Upsilon}_{t+1,t+1} & \boldsymbol{\Upsilon}_{t+1,t}^T \\ \boldsymbol{\Upsilon}_{t+1,t} & \boldsymbol{\Upsilon}_{t,t} \end{vmatrix} - \log|\boldsymbol{\Upsilon}_{t,t}| \right) \\ &+ \frac{1}{2}\log|\boldsymbol{\Upsilon}_{1,1}| + \frac{1}{2}(K-1)T(\log 2\pi + 1). \end{aligned} \tag{62}$$

The entropy $\mathcal{S}(q(\boldsymbol{\mu}_{1:T}))$ is then evaluated during the backward pass of the Kalman–Rauch scheme, using the iterative calculation of the matrix determinants of the right-hand side of Eq. (62).

References

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 18, 14–30.

Beal, M.J., 2003. Variational algorithms for approximate Bayesian inference. PhD thesis, University of London, United Kingdom.

Beal, M.J., Ghahramani, Z., 2002. The variational Kalman smoother. Neural Comput. 14, 2647–2692.

Choudrey, R.A., Roberts, S.J., 2003. Variational mixture of Bayesian independent component analysers. Neural Comput. 15, 213–252.

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

Daunizeau, J., Grova, C., Mattout, J., Marrelec, G., Clonda, D., Goulard, B., Lina, J.M., Benali, H., 2005. Assessing the relevance of fMRI-based prior in the EEG inverse problem: a Bayesian Model Comparison approach. IEEE Trans. Signal Process. 53, 3461–3472.

Daunizeau, J., Mattout, J., Clonda, D., Goulard, B., Benali, H., Lina, J.M., 2006. Bayesian spatio-temporal approach for EEG source reconstruction: conciliating ECD and distributed models. IEEE Trans. Biomed. Eng. 3, 503–516.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., pelegrini Issac, M., Lina, J.M., Benali, H., in press. Symmetrical event-related EEG/fMRI information fusion in a variational bayesian framework. NeuroImage.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. NeuroImage 30, 1255–1272.

de Munck, J.C., 1988. The potential distribution in a layered spheroidal volume conductor. J. Appl. Phys. 64, 464–470.

Ebersole, J.S., 1997. Defining epileptogenic foci: past, present and future. J. Clin. Neurophys. 14, 470–483.

Egan, J.P., 1975. Signal Detection Theory and ROC Analysis. Academic Press, New York.

Freeman, W.J., Kozma, R., Werbos, P.J., 2001. Biocomplexity: adaptive behavior in complex stochastic dynamical systems. Biosyst. J. Biol. Inf. Process. Sci. 59, 109–123.

Friston, K.J., Penny, W., 2003. Posterior probability maps and spms. NeuroImage 19, 1240–1249.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477.

Friston, K.J., Penny, W., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465–483.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Friston, K.J., Mattout, J., Penny, W., in preparation. Variational free energy and the Laplace approximation.

Ghahramani, Z., Beal, M.J., 2000. Advanced Mean Field Methods—Theory and Practice, Chapter Graphical Models and Variational Methods. MIT Press.

Grova, C., Daunizeau, J., Bénar, C.G., Benali, H., Gotman, J., 2006. Evaluation of EEG localization methods using realistic simulations of interictal spikes. NeuroImage 29, 734.

Harrison, L.M., David, O., Friston, K.J., 2005. Stochastic models of neuronal dynamics. Philos. Trans. R. Soc. 360, 1075–1091.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain-minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.

Horwitz, B., Tagamets, M.A., McIntosh, A.R., 1999. Neural modeling, functional brain imaging and cognition. Trends Cogn. Sci. 3, 91–98.

Kiebel, S.J., David, O., Friston, K.J., submitted for publication. Dynamic causal modeling of evoked responses in EEG and MEG with lead-field parametrization. NeuroImage, 30.

Kilner, J.M., Mattout, J., Henson, R., Friston, K.J., 2005. Hemodynamic correlates of EEG: a heuristic. NeuroImage 28, 280–286.

Kohler, T., Wagner, M., Fuchs, M., Wischmann, H.A., Drenckhahn, R., Theifien, A., 1996. Depth normalization in MEG/EEG current density imaging. In 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam 1996; 3.5.2: EEG/MEG—Inverse Problem and Mapping.

Mattout, J., Philipps, C., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. NeuroImage 30, 753–767.

Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. IEEE Trans. Biomed. Eng. 46, 245–259.

Nunez, P.L., 1981. Electric Fields of the Brain. New York Press, Oxford.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. 18, 49–65.

Penny, W., Friston, K., 1999. Mixtures of general linear models for functional neuroimaging. IEEE Trans. Med. Imag. 22, 504–514.

Penny, W., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. Neural Netw. 22, 1157–1172.

Phillips, C., Rugg, M.D., Friston, K.J., 2002. Systematic regularization of linear inverse solutions of the EEG source localization problem. NeuroImage 17, 287–301.

Riera, J., Aubert, E., Iwata, K., Kawashima, R., Wan, X., Ozaki, T., 2005. Fusing EEG and fMRI based on a bottom-up model: Inferring activation and effective connectivity in neural masses. Philos. Trans. R. Soc. Lond., B Biol. Sci. 29, 1025–1041.

Riera, J., Wan, X., Jimenez, J.C., Kawashima, R., 2006. Nonlinear local electrovascular coupling: I. A theoretical model. Hum. Brain Mapp. 25.

Scherg, M., von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. Electroencephalogr. Clin. Neurophysiol. 65, 344–360.