



ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg  
NeuroImage xx (2007) xxx–xxx

## Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG

Stefan J. Kiebel,<sup>a,\*</sup> Jean Daunizeau,<sup>a</sup> Christophe Phillips,<sup>b</sup> and Karl J. Friston<sup>a</sup>

<sup>a</sup>The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3AR, UK

<sup>b</sup>Centre de Recherches du Cyclotron, Université de Liège, Liège, Belgium

Received 1 May 2007; revised 7 August 2007; accepted 3 September 2007

In magneto- and electroencephalography (M/EEG), spatial modelling of sensor data is necessary to make inferences about underlying brain activity. Most source reconstruction techniques belong to one of two approaches: point source models, which explain the data with a small number of equivalent current dipoles and distributed source or imaging models, which use thousands of dipoles. Much methodological research has been devoted to developing sophisticated Bayesian source imaging inversion schemes, while dipoles have received less such attention. Dipole models have their advantages; they are often appropriate summaries of evoked responses or helpful first approximations. Here, we propose a variational Bayesian algorithm that enables the fast Bayesian inversion of dipole models. The approach allows for specification of priors on all the model parameters. The posterior distributions can be used to form Bayesian confidence intervals for interesting parameters, like dipole locations. Furthermore, competing models (*e.g.*, models with different numbers of dipoles) can be compared using their evidence or marginal likelihood. Using synthetic data, we found the scheme provides accurate dipole localizations. We illustrate the advantage of our Bayesian scheme, using a multi-subject EEG auditory study, where we compare competing models for the generation of the N100 component.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** EEG; MEG; Equivalent current dipole; Variational Bayes

### Introduction

The analysis of evoked responses using magneto- and electroencephalography (M/EEG) can proceed in several ways. If one is interested in inferring the locations of M/EEG generators within brain space, one has to solve the inverse spatial problem (Baillet *et al.*, 2001). There are two main approaches to estimating sources from observed sensor data. The first assumes that sensor

data can be explained by a small set of equivalent current dipoles. The inversion of this model amounts to a nonlinear optimization problem, because the forward model is nonlinear in dipole location (Mosher *et al.*, 1992). Recently, the source reconstruction problem has been addressed by placing many dipoles in brain space, and using constraints on the solution to make it unique; for example (Baillet and Garnero, 1997; Mattout *et al.*, 2006; Phillips *et al.*, 2005). This approach is attractive, because it produces images of brain activity comparable to other imaging modalities and it eschews subjective constraints on the inversion. For imaging solutions, most constraints can be motivated by anatomical and physiological arguments, *e.g.*, smoothness constraints and approximate location priors, based on regional activity in functional magnetic resonance imaging (fMRI). Traditional few-dipole solutions, however, are usually regarded as depending too much on user-specified modelling decisions; like the number of dipoles and their initial locations. Mathematically, it can be argued that the inversion of dipole models is a harder problem than inversion of distributed models, because the inverse problem of distributed source imaging is basically linear. These reasons might explain why much methodological research has been devoted to developing sophisticated Bayesian source imaging inversion schemes, while dipole models have received less such attention.

However, models with a few dipoles are useful, because they represent a direct mapping from scalp topography to a small set of parameters. Dipole solutions usually lend themselves to simple interpretations and provide an informative way to explain the observed data. Furthermore, it is easy to report the sufficient statistics of dipole parameters, over subjects. Operationally, summarising distributed activity with a small number of sources simplifies analyses of connectivity among those sources (*e.g.*, dynamic casual modelling of evoked or induced responses (Kiebel *et al.*, 2006)). Critically, in a Bayesian context, different models can be compared using their evidence or marginal likelihood. This model comparison is superior to classical goodness-of-fit measures, because it takes into account the complexity of the models (*e.g.*, the number of dipoles) and, implicitly, uncertainty about the model parameters. For this reason, classical schemes have adopted

\* Corresponding author. Fax: +44 20 7813 1420.

E-mail address: skiebel@fil.ion.ucl.ac.uk (S.J. Kiebel).

Available online on ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)).

other measures for model comparison (e.g., the Akaike Information Criterion (AIC); see also Supek and Aine (1993) for an example using classical model comparison). For most models, the AIC and its cousin, the Bayesian Information Criterion (BIC) are a rough approximation to the model evidence (Beal, 2003; Penny et al., 2004), and are less accurate than the negative free energy. In this paper, we provide some examples of the usefulness of model comparison, with dipole models.

When the model comprises only one or two dipoles, the best solution can usually be found without using any constraints and a Bayesian framework appears to be superfluous. For three or more dipoles, model inversion is more difficult because many local minima of the high-dimensional objective function exist. In this situation, it is practically infeasible to visit all local minima and select the best solution. Rather, one can introduce constraints that preclude certain un-physiological solutions, and guide the inversion towards favoured solutions. Such constraints are implemented naturally using Bayesian techniques, but they invite criticism that using informative priors imposes a pre-selected solution. This criticism can be countered by observing that Bayesian model comparison allows one to assess several solutions objectively and assert that there is strong evidence in favour of a particular solution (Penny et al., 2004). Usually, in M/EEG, candidate models already exist, based on cognitive theories and preceding studies. These predictions motivate the use of informed priors, and the subsequent comparison of competing models. Therefore, Bayesian model comparison is a useful way to decide which theory explains the observed data best and informative priors are central to this strategy. Even inconclusive model comparison (i.e., all models explain the data equally well) tells us the data do not provide enough evidence in favour of one theory over the other. These procedures and inferences could not proceed in a classical (i.e., non-Bayesian) framework.

In short, fast Bayesian inversion for dipole models seems to be a useful addition to the toolbox for M/EEG analysis. In the present paper, we propose a variational Bayes (VB) inversion scheme for a single time point. Only a few Bayesian inversion schemes for (spatial or spatiotemporal) dipole models have been described in the literature (Auranen et al., 2007; Jun et al., 2005, 2006; Schmidt et al., 1999). These approaches are based on Monte Carlo–Markov chain techniques, which use time-consuming sampling procedures to compute the posterior distributions. Variational Bayes provides a fast and efficient approximation to the necessary integrals and has been applied successfully to source imaging in M/EEG (Daunizeau et al., 2007; Sato et al., 2004) and other problems in functional imaging (Flandin and Penny, 2007; Penny et al., 2003; Woolrich and Behrens, 2006).

There are other approximate optimization schemes that we could have used for implementing a Bayesian approach to dipole models. Among them are ‘iterative conditional modes’ (ICM) or conditional expectation–maximization algorithms. However, it is known that these techniques are not invariant under re-parameterizations while VB is. Furthermore, ICM does not per se compute the model evidence, which is easy to do with VB.

In the following, we first describe the equivalent current dipole model and derive the VB algorithm. In the second section, we use the VB and conventional scheme on synthetic and real data. We provide some examples of using informed priors and compare the two schemes. In the discussion, we address advantages, disadvantages and potential extensions of the approach.

## Theory

### Equivalent current dipole model

It is generally assumed that the bulk of remotely detected M/EEG signal is generated by synchronous depolarization of pyramidal populations, where the current flows between synapses proximate and distal to the cell bodies. The relationship between scalp data  $y$  and primary current density is linear and instantaneous so that

$$y = G(s)w \quad (1)$$

where  $G(s)$  is the  $(N_c \times 3N_s)$  lead-field matrix.  $N_c$  is the number of channels or sensors and  $N_s$  is the number of sources. The  $(3N_s \times 1)$  vector location  $s$  forms the input argument for the nonlinear lead-field function,  $G(s)$ , whose output is multiplied by the  $(3N_s \times 1)$  moment vector  $w$  to form the observed data.<sup>1</sup> The lead-field accounts for passive propagation of the electromagnetic field from the sources to the sensors (Mosher et al., 1999). Note that although the relationship between the data and primary current density is linear, it is non-linear in the dipole locations.

For EEG, a popular head model is based on four concentric spheres, each with homogeneous and isotropic conductivity. The four spheres approximate the brain, skull, cerebrospinal fluid (CSF) and scalp. The parameters of the model are the radii and conductivities for each layer. Here, we use radii of; 71, 72, 79 and 85 mm, with conductivities 0.33, 1.0, 0.0042 and 0.33 S/m, respectively. For MEG, one can use a single sphere as a good approximation. The potential or magnetic field at the sensors requires an evaluation of an infinite series, which can be approximated using fast algorithms (Mosher et al., 1999; Zhang, 1995). For the ECD forward model, we used a Matlab (MathWorks) routine that is freely available as part of the FieldTrip package (<http://www2.ru.nl/fcdonders/fieldtrip/>, see also Oostenveld, 2003) under the GNU general public license.

### The observation model

We transform Eq. (1) into an observation model by adding an error term.

$$y = G(s)w + \varepsilon \quad (2)$$

We assume an independent and identically distributed (i.i.d.) normal error, which is parameterised by a precision parameter  $\gamma_\varepsilon$ . This specifies a likelihood model for the data given the model parameters. The probabilistic generative model is completed by the specification of priors: The normally distributed parameter vectors,  $w$  and  $s$ , have gamma-distributed prior precisions  $\gamma_w$  and  $\gamma_s$ , which are scale parameters for prior covariance matrices  $\Sigma_{w_0}$  and  $\Sigma_{s_0}$  of the location and moment vectors. These do not need to be diagonal and can encode user-specified prior constraints (see below for illustrative examples). Fig. 1 shows the graphical model for this equivalent current dipole forward model, which will guide us in the subsequent derivation of update rules. We assume that the location and moment parameters are *a priori* independent of each other; that is, they are drawn independently of each other to generate the data. This precludes any prior correlation between location and moment, but such correlations are not used generally in ECD solutions.

<sup>1</sup> The moment vector can also be expressed as two angles and amplitude.

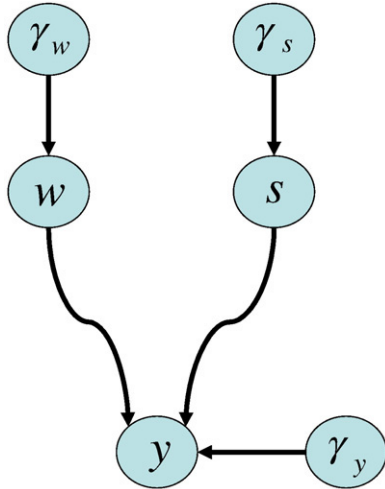


Fig. 1. Directed Bayesian graph for the equivalent current dipole forward model. This summary of the forward or generative model shows the conditional dependencies of the variables responsible for generating data. Dipole locations  $s$  and dipole moments  $w$  generated data using the equality  $y=G(s)w+\varepsilon$ , where  $G$  is the lead-field function and  $\varepsilon$  is white observation noise with precision  $\gamma_y$ . The locations  $s$  and moments  $w$  are drawn from normal distributions with precisions  $\gamma_y$  and  $\gamma_w$ . These gamma variables are themselves drawn from prior distributions and scale the covariance matrices for the location and moment parameters. Equivalently, the precision  $\gamma_y$  is drawn from a prior gamma distribution. See Eq. (5) for a full specification of these distributions.

Using the Markov properties (*i.e.*, conditional independences) of the graphical model, the joint posterior distribution of model  $m$  is

$$p(y, w, s, \gamma_y, \gamma_w, \gamma_s | m) = p(y | w, s, \gamma_y, m) p(\gamma_y | m) p(w | \gamma_w, m) p(s | \gamma_s, m) p(\gamma_s | m) \quad (3)$$

The likelihood is given by

$$p(y | w, s, \gamma_y, m) = N(G(s)w, \gamma_y^{-1} I_{N_c}). \quad (4)$$

As noted above, the prior distributions of the parameters are assumed to be normal with Gamma hyperpriors on the precisions. The Gamma hyperpriors were chosen for conjugacy purposes. See below for a detailed description of how to choose the sufficient statistics of these priors and hyperpriors.

$$p(w | \gamma_w, m) = N(\mu_{w_0}, \gamma_w^{-1} \Sigma_{w_0})$$

$$p(s | \gamma_s, m) = N(\mu_{s_0}, \gamma_s^{-1} \Sigma_{s_0})$$

$$p(\gamma_y | m) = Ga(a_{y_0}, b_{y_0})$$

$$p(\gamma_w | m) = Ga(a_{w_0}, b_{w_0})$$

$$p(\gamma_s | m) = Ga(a_{s_0}, b_{s_0}) \quad (5)$$

This concludes the specification of the generative model. Next, we look at its variational inversion to obtain the posterior or conditional density.

#### The variational Bayesian (VB) scheme

The assumption of conditional independence of the parameters; the locations, moments and precisions,  $\theta = \{w, s, \gamma_y, \gamma_w, \gamma_s\}$  is at the heart of the variational approach and corresponds to a mean-field approximation: the true marginal posteriors of the parameters are replaced by variational approximations, which are computed using the sufficient statistics of the other variables. Under the mean-field approximation, the posterior distributions on all parameters factorizes into the marginals

$$q(\theta) = q(w)q(s)q(\gamma_y)q(\gamma_w)q(\gamma_s) \quad (6)$$

where  $q(\bullet)$  is the variational approximation of any marginal posterior density,  $p(\bullet | y, m)$ . With this approximation, we can express the log-marginal likelihood or evidence as

$$\ln p(y | m) = \underbrace{\langle \ln p(\theta, y | m) \rangle_{\prod q(\bullet)} + \sum s(q(\bullet))}_{F(q)} + D_{\text{KL}}(q(\theta), p(\theta | y, m)) \quad (7)$$

where the first two terms are the negative free energy,  $F(q)$ , and  $S(q(\bullet))$  denotes the Shannon entropy of each marginal. Maximising the negative free energy with respect to  $q(\theta)$  means minimising the Kullback–Leibler divergence term; *e.g.*, Beal (2003); at which point the free energy becomes a lower-bound approximation to the log-evidence and  $q(\theta)$  approximates the true posterior. This means the sufficient statistics and functional form of the marginals can be found by maximising the free energy. The functional form is given by setting the variation of the free energy, with respect to each marginal, to zero

$$\partial_{q(\bullet)} F(Q) = 0 \Rightarrow$$

$$\ln q(\bullet) = \langle \ln p(\theta, y | m) \rangle_{\prod \tilde{q}(\bullet)} + \ln Z. \Rightarrow$$

$$q(w) = N(\mu_w, \Sigma_w)$$

$$q(s) = N(\mu_s, \Sigma_s)$$

$$q(\gamma_y) = Ga(a_y, b_y)$$

$$q(\gamma_w) = Ga(a_w, b_w)$$

$$q(\gamma_s) = Ga(a_s, b_s) \quad (8)$$

where  $Z$ , is a normalization constant known as the partition function and  $\tilde{q}(\bullet)$  is the Markov blanket of  $q(\bullet)$  (*i.e.*, marginals that are the children, parents or parents of the children of each marginal in the dependency graph in Fig. 1). By using conjugate priors (a gamma prior for a normal distribution) we can derive analytic expressions for the expectations above, in terms of the sufficient statistics of the marginals,  $q(\bullet)$ ; these closed forms specify relatively simple update rules for the sufficient statistics, which are summarised in Fig. 2. (These rules are derived in detail in

## Update Rules

### Data precision

$$b_y = \frac{1}{2} \left( y^T y - 2\mu_w^T G(\mu_s)^T y + g(\mu_s)^T (D + E) g(\mu_s) + \text{tr} \left( \Sigma_s \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) \right) + b_{y0}$$

$$a_y = \frac{1}{2} N_c + a_{y0}$$

### Moments

$$\mu_w = \Sigma_w \left( \frac{a_y}{b_y} G(\mu_s)^T y + \frac{a_w}{b_w} \Sigma_{w_0}^{-1} \mu_{w0} \right)$$

$$\Sigma_w = \left( \frac{a_y}{b_y} (G(\mu_s)^T G(\mu_s) + B) + \frac{a_w}{b_w} \Sigma_{w_0}^{-1} \right)^{-1}$$

### Parameter precisions

$$b_w = \frac{1}{2} (r(w)_w^T \Sigma_{w_0}^{-1} r(w)_w + \text{tr}(\Sigma_{w_0}^{-1} \Sigma_w)) + b_{w0}$$

$$a_w = \frac{3}{2} N_s + a_{w0}$$

$$b_s = \frac{1}{2} (r(s)_s^T \Sigma_{s_0}^{-1} r(s)_s + \text{tr}(\Sigma_{s_0}^{-1} \Sigma_s)) + b_{s0}$$

$$a_s = \frac{3}{2} N_s + a_{s0}$$

### Locations

$$\mu_s = \Sigma_s \left( \frac{a_y}{b_y} \left( \frac{\partial g^T}{\partial s} \left( (\mu_w \otimes y) - (D + E) \left( g(\mu_s) - \frac{\partial g}{\partial s} \mu_s \right) \right) \right) + \frac{a_s}{b_s} \Sigma_{s_0}^{-1} \mu_{s0} \right)$$

$$\Sigma_s = \left( \frac{a_y}{b_y} \left( \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) + \frac{a_s}{b_s} \Sigma_{s_0}^{-1} \right)^{-1}$$

$$E = (\mu_w \mu_w^T \otimes I)$$

$$D = (\Sigma_w \otimes I)$$

$$B = \sum_{j=1}^{N_s} C(j + [0 : N_p - 1], j + [0 : N_p - 1])$$

$$C = \frac{\partial g}{\partial s} \Sigma_s \frac{\partial g^T}{\partial s} \quad r(w)_w = \mu_w - \mu_{w0}$$

$$r(s)_s = \mu_s - \mu_{s0}$$

Fig. 2. Variational update rules for the sufficient statistics of the approximating marginal densities.

Appendix A.) It can be seen from Eq. (8) that the sufficient statistics for any marginal depend on the other marginals (in its Markov blanket). It is therefore necessary to update these statistics iteratively, until the free energy is maximised. This entails iterating a series of variational steps, with one step for each marginal.

#### A note on priors

The use of priors, in combination with model comparison, enables one to assess the relevance of competing models and implicitly the priori assumptions. In this section, we will review briefly the use and the meaning of the priors for each marginal.

An informative prior on the precision of the location parameters can be motivated by prior information; for example from previous data, or by using predictions from some theory about the functional anatomy shaping the response. A simple example is making the prior distribution of the locations tight around some prior locations,  $\mu_{s_0}$ , which is a device to constrain the source locations to some pre-defined regions or to ‘penalize’ deviations from the prior location. Note that this is not the same as fixing the source locations. The difference is that a ‘soft’ Bayesian prior will prefer the prior locations, if the data does not imply otherwise. The tightness of the prior is controlled by the prior covariance: The prior covariance matrices  $\Sigma_{s_0}$  and  $\Sigma_{w_0}$  can be used to introduce prior knowledge about the *relative* variability of the parameters, with a diagonal  $\Sigma_{s_0}$  (or  $\Sigma_{w_0}$ ) with unequal variances. Similarly, one can use the off-diagonal elements to encode knowledge about correlations among the locations or moments. This can be used for modelling

symmetric sources, as we will illustrate later. Otherwise, the prior covariance matrices  $\Sigma_{s_0}$  and  $\Sigma_{w_0}$  are set to the identity matrix in this paper. The *absolute* variability of the parameters is determined by the prior precisions, which scale these matrices:

The prior precisions  $\gamma_s$  and  $\gamma_w$  determine the importance of the prior relative to the likelihood (*i.e.*, data). By allowing the prior precisions to be optimised as free parameters, we are effectively optimising the balance between data and priors. This is an important aspect of hierarchical Bayesian models, which we have exploited in the context of parametric empirical Bayes models previously (*e.g.*, Mattout et al., 2006; Phillips et al., 2002). It was also used by Sato et al. (2004) to implement automatic relevance determination (ARD) to ‘switch off’ redundant sources in an imaging context. Unless stated otherwise, we use the same non-informative Jeffrey’s priors for the precision parameters as employed by Sato et al. (2004).

$$p(\gamma_w) \propto \frac{1}{\gamma_w}$$

$$p(\gamma_s) \propto \frac{1}{\gamma_s}$$

These are called hyperpriors because they are a prior on a sufficient statistic of a prior and represent a special case of the Gamma hyperprior above, that obtains when  $a_{w_0} = a_{s_0} \rightarrow 0$  and  $b_{w_0} = b_{s_0} \rightarrow 0$ . Strictly speaking, these are improper densities but this does not seem to confound variational schemes. Jeffrey’s priors are

uninformative about the magnitude scale of the precision parameters (they are uniform over log-precision; see also Kass and Wasserman, 1996). To make the priors more informative, we simply increase  $a_0$ , which increases the expectation of the hyperpriors and the corresponding prior precisions. The data precision hyperpriors  $a_{y_0}$  and  $b_{y_0}$  are specified in the same way.

### Initialization and convergence

In this work, we adopt a pragmatic way of dealing with potential local minima. Instead of using an informed initialization that anticipates some best solution, we run the algorithm several times (using different random initializations) and select the best solution: a similar multi-start procedure has been described by Huang et al. (1998). In our experience, to find a good solution, one should at least run four iterations for single dipole models. In our simulations, we obtained successful convergence using sixteen random initializations, for all single and multi-dipole models. Computationally, this approach is feasible, because the computing time needed for each inversion (up to 4 dipoles) is in the order of seconds. We use the multi-start approach (for example with four or more inversions) for models with more than one dipole.

The random initialization is as follows:

1. Draw location vector  $\mu_s$  repeatedly from  $N(0, 10I_{N_p})$  until all sources are inside the spherical head model
2. Draw a moment vector  $\mu_w$  from  $N(0, I_{N_p})$
3. Initialize remaining sufficient statistics

$$a_y = \frac{N_c}{2}, \quad b_y = \frac{1}{2} (y - G(\mu_s)\mu_w)^T (y - G(\mu_s)\mu_w)$$

$$a_w = \frac{N_p}{2}, \quad b_w = \frac{1}{2} (\mu_w - \mu_{w_0})^T (\mu_w - \mu_{w_0})$$

$$a_s = \frac{N_p}{2}, \quad b_s = \frac{1}{2} (\mu_s - \mu_{s_0})^T (\mu_s - \mu_{s_0})$$

$$\Sigma_w = \frac{b_w}{a_w} \Sigma_{w_0}^{-1}, \quad \Sigma_s = \frac{b_s}{a_s} \Sigma_{s_0}^{-1}$$

The implicit Gamma densities on the precisions correspond to an expected variance<sup>2</sup>  $b_\bullet/a_\bullet$  based on the sum of squared residuals, where these residuals are in measurement space or parameter space, depending on the precision in question.  $N_p$  is the number of parameters under the conditional density in question. Once these values have been initialised, they are updated until convergence. The convergence criterion is a small change in the negative free energy (see Appendix) and a maximum number of iterations. The change is computed as  $\Delta F = F^{(k)} - F^{(k-1)}$ , where  $F^{(k)}$  is the negative free energy for iteration  $k > 1$ . A typical threshold for  $\Delta F$  is  $10^{-2}$ , with a maximum of 200 iterations.

### Applications

In the following, we first establish that the variational scheme returns sensible posterior distributions. We will also show that the posterior means provide veridical estimates of the true locations

<sup>2</sup> The mean of a Gamma distribution is  $a_\bullet/b_\bullet$  and this encodes the precision or inverse variance.

and moments. The marginal posterior covariances can be used to construct Bayesian confidence intervals, which should encompass known values. For selected case studies, we will demonstrate Bayesian model comparison, when different prior assumptions are appropriate. Using real data, we will illustrate model inversion and comparison when the data are generated by more than one dipole.

### Single-dipole simulations

In this section, we show briefly that the inversion can localize single dipoles accurately. We simulated data using an extended 10–20 setup with 30 channels. This setup is used widely in evoked response potential (ERP) research (Oostenveld and Praamstra, 2001), and represents a challenge to any source localization technique, relative to high-density recordings. The 30 channels cover the scalp sparsely; the lowest channels are located just over temporal locations. For some source locations and moments, there are large posterior uncertainties in the parameters because the spatial expression of these dipoles is indistinct.

We generated EEG data (at a single time point or time-window average) caused by one dipole located randomly in head model space. The locations were drawn from a normal distribution with zero mean and a standard deviation of 50 mm. The moments were drawn randomly from a normal distribution with zero mean and standard deviation of one. Sources that were outside or close to the inner sphere of the head model (more than 65 mm from origin) were discarded. The same applies for sources with a  $z$ -coordinate less than  $-20$  mm. In this way, we generated dipoles that are within the head and are located at approximate coordinates according to the Montreal Neurological Institute (MNI) coordinate system. We then used the VB scheme of the previous section to iterate the update rules, employing uninformative priors. In these simulations, the VB scheme converged after about ten to twenty iterations. Unless otherwise stated, we used a multi-start optimization with sixteen initialisations. *Post hoc* examination of the free energies, after convergence, showed that about 80% of the initializations<sup>3</sup> converged on the same maximum, which we take to be the global maximum.

We simulated data for three different noise levels (white noise at the sensors) with typical signal-to-noise ratios (SNR) of 100, 50, and 20. We established this range experimentally by fitting dipole models to peak component data of an auditory evoked response at 100 ms, over 12 subjects (see below). We derived the SNR using the estimate of the precision parameter  $\gamma_j$ .

For each noise level, we generated 2000 data sets. In Fig. 3, we show the localization errors (*i.e.*, the distance between the true location and the posterior mean of location) as histograms. As expected, the localization error increases for higher noise levels. For an SNR of 50, effectively all sources are located within 2 cm of the true location. For all noise levels, the majority of realizations had a localization error of less than 8 mm.

The marginal posteriors can be used to derive Bayesian confidence intervals for all parameters (Friston et al., 2003). These provide a measure of certainty of the dipole location (and moment). In Table 1, we provide the percentage of simulations, in which a 95% confidence interval contained the true parameter.

Ideally, if the posterior marginal distributions were exact, the percent would be around 95%. However, for each parameter, the posterior certainty is slightly too high, *i.e.*, the confidence intervals

<sup>3</sup> The exact percentage depends on the model.

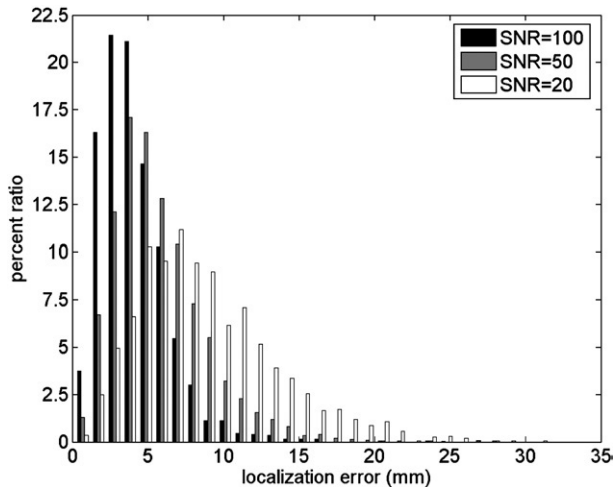


Fig. 3. Single-dipole simulations. Histogram of localization error (distance between true location and posterior mean, in mm). As expected, the localization errors are larger for low SNR.

are too tight, in particular for  $z$ -location and  $z$ -moment. This reflects the well-known overconfidence problem with mean-field approximations and is most likely a consequence of assuming that the location and moment parameters are conditionally independent (Eq. (3)); see also Discussion.

#### Two- and three-dipole models: Case studies

We could have repeated the above simulations with more than one dipole. However, this would not be very useful, because multi-dipole scenarios can be qualitatively different from each other. For example, when two dipoles are orientated in parallel, some of the parameters will be correlated, and error measures of single parameters will be inflated, relative to the case when two dipoles are orthogonal to each other. This is a fundamental issue for validation of any dipole fitting procedure and can make simulations based on random samples from multiple-dipole parameters difficult to compare. We will work through a few anecdotal but instructive cases, which are relevant in practice.

We summarise the results for each of the simulations by reporting the (i) negative free energy as an approximation to model evidence, (ii) the goodness-of-fit (Supek and Aine, 1993) and (iii) the Akaike Information Criterion (Penny et al., 2004), using their Eqs. (18) and (21).

#### Case 1: Symmetry priors for pair of dipoles

Often, in ERP/ERF analysis, one uses so-called symmetric dipoles. These are typically chosen to model early- to medium-latency peri-stimulus responses and assume simultaneous sub-cortical input to both hemispheres. For an example, see Deffke et al. (2007), where the authors use a pair of dipoles mirror-symmetric about the sagittal plane (separating the two hemispheres). Classical methods of fitting such dipole pairs enforce symmetry by setting  $s_x^1 = -s_x^2$ ,  $s_{y,z}^1 = s_{y,z}^2$ ,  $w_x^1 = -w_x^2$  and  $w_{y,z}^1 = -w_{y,z}^2$ , where  $[s_x^{1,2}, s_y^{1,2}, s_z^{1,2}]^T$  and  $[w_x^{1,2}, w_y^{1,2}, w_z^{1,2}]^T$  are the location and moments of the first and second dipole. Obviously, there are many variants of this parameterization; e.g., one could fix the  $x$ -axis parameters only, or allow for different overall amplitudes. Effectively, these parameterizations reduce the number of free

parameters and will give better models, if the symmetry assumption is correct. Also, one can fit models with informative priors, which cover the middle ground between the unconstrained model and fixed symmetric parameters. With Bayesian model comparison, the idea is to fit two or more models to the data and see which is best, using the free energy approximation to the model evidence. In addition, one can compare models with informative priors, which cover the middle ground between the unconstrained model and fixed symmetric parameters.

To show how our approach performs in such situations, we simulated data from three models: The first comprised a truly symmetric dipole pair; the second, a nearly symmetric pairs of dipoles, and the third, a single unilateral dipole (*i.e.*, with the second mirror-symmetric dipole missing). We show the data and the dipole parameters in Fig. 4.

We use seven models to invert each of the three data sets:

1. fully unconstrained (twelve) spatial parameters
2. prior on true location
3. prior on true moment
4. prior on true location and moment
5. prior on true location, moment and symmetry
6. hard symmetry constraint, otherwise uninformative priors
7. hard symmetry constraint, with priors on true location and moment

Although, for real data, one would never know the true parameters to specify these priors, we use them to simulate well-informed beliefs about the parameters. All priors were implemented by increasing the prior expectations of the prior precisions. This was implemented by making  $a_{w_0}$  or  $a_{s_0}$  (the scale parameters of the Gamma hyperpriors) equal to half the number of free parameters under the prior density (see Eq. (5)). For example, for two dipoles and ‘soft’ priors, the number of moment parameters is six, so  $a_{w_0} = 3$ . The hard symmetry constraints of models 6 and 7 encode fixed symmetry as described above; *i.e.*, twelve free parameters are reduced to six. We implemented this constraint using a projector matrix to remove the unwanted spatial degrees of freedom (see Appendix C). The symmetry prior of model 5 was induced with strong correlations ( $\pm 0.99$ ) in the prior covariance matrices  $\Sigma_{w_0}$  and  $\Sigma_{s_0}$ . This is similar to using classical hard constraints, but uses Bayesian priors.

We simulated data using observation noise with an SNR of 100. Assuming a uniform prior over the seven models, the log-evidences in Table 2 can be used directly for model comparison.

Table 1

Single-dipole simulations: percent of realisations in which the 95% posterior confidence interval includes the true parameter; for each parameter and noise level

	SNR=100	SNR=50	SNR=20
$x$ -location	87.90 (2.95)	86.60 (3.68)	86.40 (3.73)
$y$ -location	90.05 (2.87)	88.45 (2.78)	88.60 (3.63)
$z$ -location	79.20 (3.50)	79.15 (4.02)	78.50 (5.03)
$x$ -moment	90.40 (2.87)	89.30 (3.53)	89.10 (3.77)
$y$ -moment	87.95 (2.28)	88.00 (4.09)	88.00 (2.96)
$z$ -moment	77.50 (4.71)	78.25 (3.89)	80.25 (4.44)

The numbers in parentheses are the standard deviations of the percent ratios. It can be seen that the posterior confidence intervals are too tight, in particular for  $z$ -location, and  $z$ -moment.

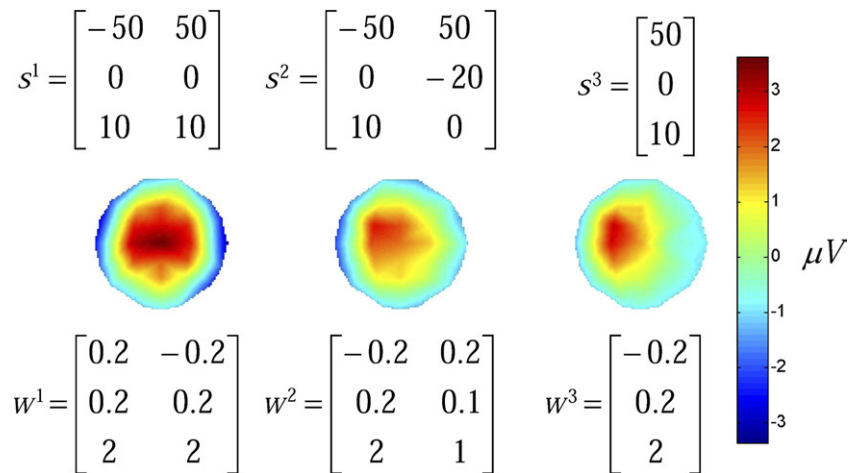


Fig. 4. First case study: Three scalp topographies (nose is up) and their generating dipole parameters. Left: truly symmetric pair of dipoles, middle: nearly symmetric pair of dipoles, right: unilateral dipole.

For truly symmetric dipoles, the models which incorporate informative priors and symmetry constraints are the best. It does not seem to matter whether symmetry is enforced using hard constraints, or whether one uses the prior covariance matrix to enforce strong correlations. The unconstrained model is the worst for truly symmetric data, because it is over-parameterised and too complex. For nearly symmetric dipoles, the best models are the ones that use moment or location priors only. For the single dipole data, the models with the least assumptions about symmetry are the best. Inspection of the posterior means show that the moment parameters of the second (non-existent) dipole have been estimated as being close to zero (the 95% confidence interval contains the zero moments in each direction). For single-subject analyses, model comparison can be used to identify the best model; for group studies, one can select the best model over subjects, using the sum of log-evidences. At the group level, one can add subject-wise log posterior model probabilities (see below Eq. (9)). Given our demonstration above, the best model will most likely be either a symmetric model (model 5 or 7), or the model with informative priors (models 2 and 3).

#### Case 2: Two or three dipoles?

In this example, we will use model comparison to decide if there are two or three dipoles generating the data. We simulated

two sets of data. The first was generated by a symmetric pair of dipoles. The second data set was generated by the same symmetric dipoles and a third dipole in a medial frontal location. The parameters and resulting scalp topography are shown in Fig. 5.

The first scalp topography looks distinctively symmetric, while the second shows clearly the effect of, at least, one asymmetric dipole. The question is, whether one can retrieve the generating models from each data set, given some prior knowledge about the sources. We assume that some cognitive theory predicts a symmetric pair of dipoles and a frontal source, for which we know the approximate location from other M/EEG or fMRI studies. Given our priors about the generating sources, we can compare five models or hypotheses:

1. One dipole, uninformed priors
2. Two dipoles, uninformed priors
3. Three dipoles, uninformed priors
4. Two dipoles with informative location priors
5. Three dipoles with informative location priors

We inverted these models using the two data sets. For the informative priors, we used the true location parameters. To check that our results did not depend on knowing exactly where the true sources were, we repeated the analysis with location priors that

Table 2  
First case study

	Symmetric dipoles	Nearly symmetric	Unilateral dipole
Model 1: Unconstrained	244.85 (95.49%, 290.93)	253.03 (98.54%, 318.83)	266.98 (98.62%, 328.41)
Model 2: Prior on location	262.99 (98.82%, 310.83)	267.75 (97.86%, 314.37)	281.00 (98.72%, 329.99)
Model 3: Prior on moment	269.99 (99.11%, 315.18)	268.69 (97.51%, 312.10)	202.72 (98.57%, 327.72)
Model 4: Prior on location and moment	280.63 (98.98%, 313.37)	230.08 (97.48%, 311.98)	197.39 (98.68%, 328.70)
Model 5: Prior on location, moment and symmetry	286.12 (98.66%, 309.09)	169.54 (92.04%, 294.96)	159.23 (77.60%, 287.37)
Model 6: Hard symmetry constraint	256.68 (98.84%, 317.09)	249.05 (88.70%, 295.67)	247.05 (76.40%, 292.58)
Model 7: Hard symmetry constraint and priors on true location and moment	286.11 (98.66%, 315.25)	217.92 (92.17%, 301.20)	219.50 (79.33%, 294.55)

Log-evidences for seven models, computed for three different data sets (see text for description). In parentheses: The goodness-of-fit; *i.e.*, percent of total variance explained by model, and the Akaike Information Criterion (AIC). The best models are highlighted in yellow. For the first, truly symmetric data set, models that incorporate symmetry constraints are the best. For non-symmetric data, informed but non-symmetric priors lead to the best two dipole models. Note the failure of the goodness-of-fit and AIC to select the appropriate model (highlighted in blue).

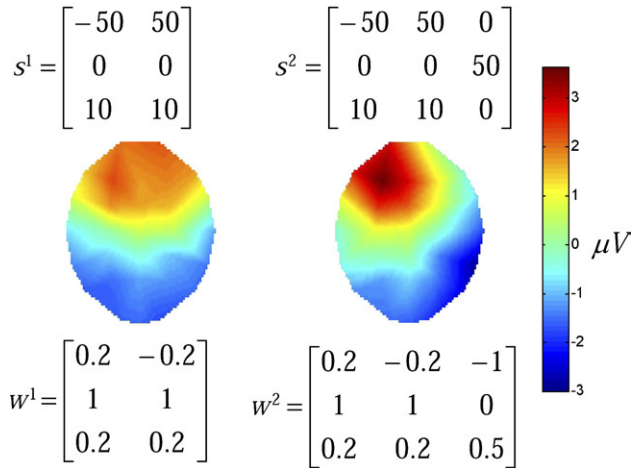


Fig. 5. Second case study. Two scalp topographies and their generating dipole parameters. Left: symmetric pair of dipoles, right: the same pair with an additional frontal source.

were 14 mm from the true locations; the results (not shown), remained qualitatively the same (see also Discussion). The informative location prior was implemented with  $a_{s_0} = 3$ .

The resulting log-evidences are listed in Table 3. As can be seen, model comparison selects the veridical model. For both data sets, the best model has a log-evidence that was three or more above those of competing models. This indicates ‘strong’ evidence in favour of the best model (Penny et al., 2004). With informed (location) priors, the posterior location is more accurate (as indicated by the improved goodness-of-fit; e.g., model 4 vs. model 2), and the model-evidence clearly identifies a superior model.

### Case 3: Two pairs of symmetric dipoles

As illustrated in the first case study, it is typical in M/EEG research to fit a single pair of symmetric dipoles to symmetric scalp topographies. However, it is possible that symmetry results from multiple pairs of symmetric dipoles, which overlap in sensor space. In this case study, we will show that one can disambiguate between one or two pairs of dipoles, given informed location priors. Again, we assume that priors about the sources exist; e.g., derived from fMRI studies and cognitive theories. We simulated three sets of data, with an SNR of 100. The first data set was generated using a single pair of symmetric dipoles (location:

$[\pm 20, -40, 10]$ , moment:  $[\pm 0.2, 1, 0.2]$ ). The second data set was generated using another pair of symmetric dipoles (location:  $[\pm 40, -20, 0]$ , moment:  $[\pm 0.1, 0.7, 0.4]$ ). The third data set was generated using both pairs of dipoles. See Fig. 6 for the three resulting scalp topographies.

We used six competing models:

1. One dipole, uninformed priors
2. Two dipoles, uninformed priors
3. Three dipoles, uninformed priors
4. Single pair of dipoles with informative priors on true location of 1st pair of dipoles
5. Single pair of dipoles with informative priors on true location of 2nd pair of dipoles
6. Two pairs of dipoles with informative priors on true locations of two pairs of dipoles

The log-evidences in Table 4 show that model comparison succeeded in selecting the true model, for all three data sets.

### Real data: an auditory evoked potential study

#### Experimental design

We studied a group of fourteen healthy volunteers aged 24–35 (5 females), see also (Garrido et al., 2007). Each subject gave signed informed consent before the study, conducted under local ethical committee guidelines. Subjects sat on a comfortable chair in front of a desk in a dimly illuminated room. Electroencephalograph activity was measured during an auditory ‘oddball’ paradigm; subjects heard “standard” (1000 Hz) and “deviant” tones (2000 Hz), occurring 80% (480 trials) and 20% (120 trials) of the time, respectively, in a pseudo-random sequence. The stimuli were presented binaurally via headphones for 15 min, every 2 s. The duration of each tone was 70 ms with 5 ms rise and fall times. The subjects were instructed not to move, to keep their eyes closed and to count the deviant tones.

#### Acquisition and pre-processing

EEG data were recorded with a Biosemi system and 128 scalp electrodes at a sampling rate of 512 Hz. Vertical and horizontal eye movements were monitored using EOG (electro-oculogram) electrodes. Data were epoched offline, with a peri-stimulus window of  $-100$  to  $400$  ms, down-sampled to 200 Hz, band-pass filtered between 0.5 and 40 Hz and re-referenced to the average

Table 3  
Second case study

	1st data set	2nd data set
	Symmetric pair	Symmetric pair and frontal dipole
Model 1: Single dipole, unconstrained	218.24 (95.01%, 310.73)	214.37 (96.20%, 309.58)
Model 2: Two dipoles, unconstrained	258.13 (94.81%, 304.13)	253.19 (96.10%, 303.02)
Model 3: Three dipoles, unconstrained	257.35 (94.62%, 297.58)	251.68 (96.08%, 296.78)
Model 4: Two dipoles, location prior	277.13 (98.56%, 323.26)	260.11 (96.12%, 303.26)
Model 5: Three dipoles, location prior	274.08 (98.65%, 317.86)	265.99 (98.21%, 308.46)

Log-evidences for five different models, computed using two data sets (see text for description).

In parentheses: The goodness-of-fit; i.e., percent of total variance explained by model, and the Akaike Information Criterion (AIC). The best models are highlighted in yellow. For both data sets, model comparison selected the generating (true) model. The best models selected by goodness-of-fit and AIC are highlighted in blue.

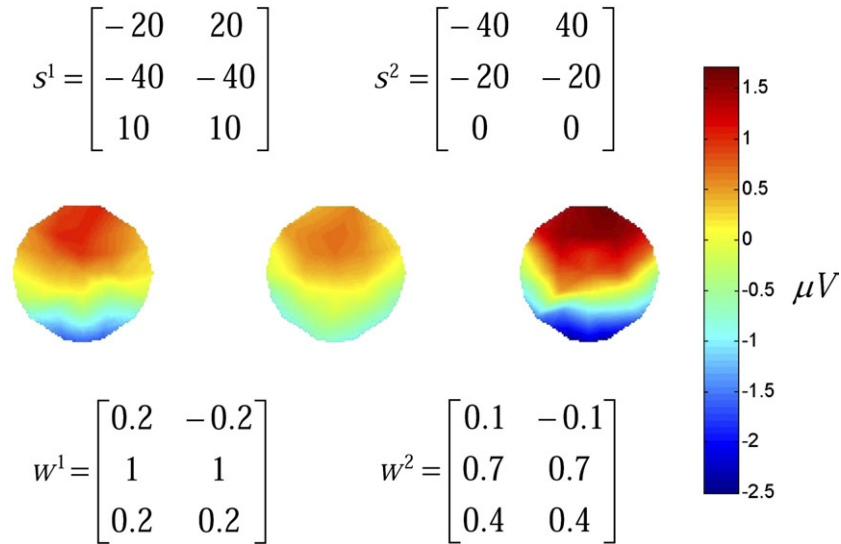


Fig. 6. Third case study. Three scalp topographies and their generating dipole parameters. Left: first symmetric pair of dipoles, middle: second symmetric pair of dipoles, right: both pairs of dipoles (using the same parameters).

over channels. Trials in which the absolute amplitude of the signal exceeded 100  $\mu\text{V}$  were excluded. Two subjects were eliminated from further analysis due to excessive numbers of trials containing artefacts. In the remaining twelve subjects, an average 18% of trials were excluded.

#### Equivalent current dipole analysis

To illustrate the application of the VB scheme to real data, we model the N100 component of the oddball condition. The data, for each subject, consist of the average for peri-stimulus times 95 to 105 ms. We use four different models which were motivated by the literature; *e.g.*, Lutkenhoner and Steinstrater (1998):

1. Single dipole, uninformed priors
2. Two dipoles with symmetry priors (as in case study 1), prior location of  $[\pm 44, -30, 12]$  (Planum Temporale), with  $a_{s_0} = 1$
3. Two dipoles, uninformative prior location of  $[\pm 44, -30, 12]$  (Planum Temporale), with  $a_{s_0} \rightarrow 0$
4. Two pairs of dipoles with symmetry priors, prior locations at  $[\pm 44, -30, 12]$  (Planum Temporale), and an anterior location  $[\pm 50, -8, 8]$ ; informed with  $a_s = 6$ .

The first (simple) model would be considered as an inappropriate model for activity generated within both hemispheres. Models 2 and 3 incorporate standard assumptions about the main generators of the N100 component. The difference is that the second uses (weak) symmetry priors, whereas the third has a priori uncorrelated sources. The fourth model is a possibly over-parameterised, but conceivable model.

To summarise the model comparison over subjects we evaluated the log conditional probability of each model, for each subject. This obtains, under flat priors on models, by normalising the marginal likelihoods to unity; *i.e.*,

$$\ln p(m_i|y) = \ln p(y|m_i) - \ln \sum_j p(y|m_j) \approx F_i - \ln \sum_j \exp(F_j) \quad (9)$$

where  $F_i$  is the free energy of model  $m_i$ . The results are shown in Fig. 7; where we plot, for each subject, the log posterior model probabilities adjusted to a minimum of  $-32$  and shifted by 8, for visualization purposes. It can be seen that the worst model is the single dipole model. The third model, a pair of dipoles without

Table 4  
Third case study

	1st pair of dipoles	2nd pair of dipoles	Both pair of dipoles
Model 1: Single dipole, uninformed	263.36 (98.11%, 320.56)	268.50 (95.73%, 320.38)	247.48 (97.48%, 303.32)
Model 2: Two dipoles, uninformed	259.01 (98.11%, 314.27)	266.05 (95.58%, 313.79)	245.25 (98.26%, 302.10)
Model 3: Three dipoles, uninformed	256.05 (98.10%, 307.86)	270.26 (95.33%, 306.97)	242.72 (96.23%, 295.50)
Model 4: 1st single pair of dipoles, location prior	275.14 (98.58%, 318.88)	282.82 (95.47%, 313.41)	253.58 (97.27%, 296.19)
Model 5: 2nd single pair of dipoles, location prior	263.39 (96.90%, 307.19)	279.02 (98.28%, 327.89)	251.97 (97.27%, 296.13)
Model 6: Two pairs of dipoles, location prior	271.50 (98.57%, 306.50)	281.72 (98.31%, 315.72)	256.12 (98.37%, 291.53)

Log-evidences for six models, computed using three data sets (see text for description).

In parentheses: The goodness-of-fit; *i.e.*, percent of total variance explained by model, and the Akaike Information Criterion (AIC). The best models, as indicated by the free energy, are highlighted in yellow. For all three data sets, model comparison selected the true model. The best models selected by goodness-of-fit and AIC are highlighted in blue.

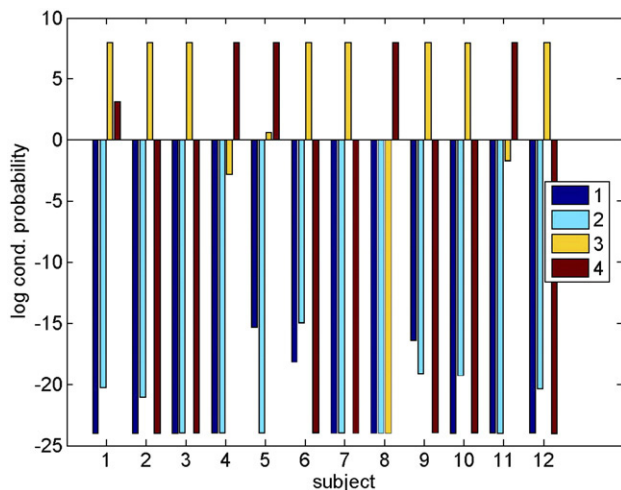


Fig. 7. Auditory evoked response data: Log-conditional probabilities of the four models, for each subject. Model 1: Single dipole, model 2: symmetric pair of dipoles in Planum Temporale (PT), model 3: pair of dipoles in PT, model 4: two pairs of symmetric dipoles in PT and slightly anterior to PT. In 8/12 subjects, the third model has the highest evidence. In the remaining four subjects, the fourth model is the best.

symmetry constraints is the best model in eight out of twelve subjects. In the remaining four subjects, the four-dipole model is the best. It is interesting to look at the latter subjects (4, 5, 8, and 11). The reason that the model evidence is higher than for the other models is that the posterior location is close to its prior. The Kullback–Leibler divergence between the approximate posterior and the prior is small (this is the model complexity) and model evidence relatively high. In other words, model four wins in some subjects, because the priors are appropriate for these data and the dipoles do not incur a complexity cost, allowing the other dipole pair to increase accuracy.

For the winning model 3, the average over the posterior locations (of those subjects, where model 3 is the best) is  $[-27.4, -9.6, 19.2]$  and  $[36.6, 1.9, 18.2]$ . Their standard deviations are large:  $[17.6, 19.1, 12.1]$  and  $[9.7, 14.7, 14.8]$ . These posterior locations point to a more anterior location than the Planum Temporale prior. There are two explanations for this result: the head model is too simple, or/and there are multiple generators involved; in particular, anterior to Planum Temporale. We cannot discard the first explanation, although to our knowledge, the 4-sphere head model should be appropriate for the locations in question. The second explanation is made more plausible by the results from subjects 4, 5, 8, and 11, where a second pair of dipoles anterior to PT turned out to be the best model.

## Discussion

We have described a variational Bayes (VB) scheme for source reconstruction of M/EEG data using a spatial model with a small number of dipoles. We have shown that the approach finds veridical solutions for synthetic EEG data and, for real EEG data, the results are sensible. There are several advantages of the present scheme over conventional approaches, which all follow from its Bayesian formulation. First, one can impose priors on the model parameters. This has advantages over using hard constraints; *i.e.*, by fixing parameters. Examples of hard constraints include the hard symmetry constraints described above, or the approach of

‘sequential dipole fitting’. The latter procedure fits single dipoles iteratively to the residuals of the previous iteration, until some goodness-of-fit criterion is reached. We do not want to argue that these approaches are not useful. Rather, we want to point out that these procedures do not take into account uncertainty about the estimates or account formally for differences in model complexity. As we have shown, a Bayesian approach can incorporate this ‘fixing parameters’ approach, but also allows for ‘soft’ prior constraints, which are mediated by information in the data. These priors are useful, because they allow a principled balance between prior knowledge, and letting the data speak for themselves. Conventional approaches can only choose between imposing prior knowledge, by fixing parameters, or not. A subtle but important aspect of the scheme described above is that the hierarchical nature of the generative model means that the relative importance attached to various priors is itself optimised. This means that priors that have been improperly specified will, in principle, not be used. Specifically, the use of hyperpriors means that prior constraints can be switched off and on, depending on whether they enable a better explanation of the data at hand.

The second advantage of the VB scheme is that it supports model comparison. With conventional approaches, the goodness-of-fit (GOF) falls into the trap of over-fitting, *i.e.*, attaining a high GOF by selecting an overly complex model. There are other measures, in the literature, that either use classical model selection, *e.g.*, an  $F$ -test (Supek and Aine, 1993), or use a simple approximation to the model evidence, *e.g.*, the Bayesian or Akaike Information Criterion (Beal, 2003; Penny et al., 2004). Although these measures are in widespread use, they do not work well when different models can have the same number of parameters but different informative priors. For AIC, this can be seen exemplarily in case studies one to three (Tables 2–4) where AIC does not point to the best model but either seems to prefer simple models, or complex models with a high goodness-of-fit. The negative free energy is an accurate approximation to model evidence and therefore is an appropriate criterion to perform model selection; among models with different priors and number of dipoles. This is an important result, because an often-encountered issue in M/EEG source localization is whether multiple sources are involved. Without priors there is no compelling way of disentangling these sources. Once priors are brought into the game, it becomes feasible to distinguish the best model from other good but less likely models. One might argue that our simulations are not relevant to real data scenarios; either because we do not include confounds or other noise sources in the simulations, or because one never has good location priors. This might be true. However, our main point is that, in principle, Bayesian model comparison is a useful way of interrogating data, when competing hypotheses can be formulated as different prior constraints that induce different models.

Furthermore, our scheme returns posterior distributions, not just point estimates. These can be used to form confidence intervals about each parameter, in particular dipole locations. This issue has been addressed by a number of techniques (Braun et al., 1997; Fuchs et al., 2004; Jun et al., 2005); approaches like ours provide a practical solution by approximating the true marginal posteriors. As we have shown empirically, for single dipoles with uninformed priors, there is a slight overconfidence about location and moment parameters. This effect is most notable for the parameters in  $z$ -direction (up and down). For these, one expects higher uncertainties than for the other parameters, because, after removing the

average reference, the sensor data are less informative than for  $x$ - and  $y$ -directions. This reflects the lack of sensors in the lower half of the head, which leads to correlations between  $z$ -locations and  $z$ -moments. The mean-field factorization between locations and moments may explain the overconfidence, which increases with dependencies between the two mean-field subsets (location and moments). Overconfidence can be reduced by assigning informed priors to one of the parameter subsets; this can be shown using simple theoretical arguments. From the update rules (Fig. 2), one can see that the uncertainty about one subset of parameters is expressed in another through its first two moments only; this means the posteriors do not account fully for dependencies between the two subsets of parameters. In practice, the certainty intervals for  $x$ - and  $y$ -parameter can be used, although they are slightly too tight. We recommend using the  $z$ -parameter intervals, with uninformed priors, only in a cautious manner.

Other Bayesian algorithms for dipole models have been described previously, most notably (Jun et al., 2005). Their algorithm uses Markov chain–Monte Carlo (MCMC) to sample the posteriors. The two key differences between MCMC and VB techniques lie in their speed and accuracy. VB is an approximation, but fast as compared to MCMC. MCMC is slow and for interesting models, convergence criteria; *i.e.*, when to stop the algorithm, are based on heuristics. In practice, our algorithm converges after seconds (using an AMD Opteron processor, 2.39 GHz), and the approximation seems to be sufficient in terms of face validity. In M/EEG source reconstruction, MCMC has been used to perform quality control on VB approximations (Nummenmaa et al., 2007; Sato et al., 2004).

Although we have illustrated our scheme using EEG data only, it can be applied to MEG after exchanging the lead-field function  $G$ . Moreover, there is no constraint on the type of the forward model. It would be feasible to use realistic head models; *e.g.*, boundary element models (BEM). Because these models take a long time to compute given location parameters; in practice, one might compute the lead-field function using a lookup table, where lead-fields are pre-computed for dipoles on a grid. The lead-fields for locations between grid points can then be interpolated, as shown in Yvert et al. (2001). In a similar vein, it is also possible to incorporate parameters of the forward model into the full model (Fig. 1). For example, it has been shown that the standard head model parameters; *i.e.*, conductivities and parameters, vary over subjects, *e.g.*, Radich and Buckley (1995). This is important for EEG because in typical dipole fitting (including ours), these parameters are fixed to some standard values. However, a variation in, for example, skin conductivity can have a large effect on how a dipole appears at the sensors; *i.e.*, it is more focused or more spread out. In practice, for a single dipole model, this means that location parameters, in particular their distance from the spheres, co-vary with skin conductivity (and other forward model parameters). A potential and simple extension of our scheme would incorporate selected head model parameters, or their ratios, into the model, and use informed priors to allow the model to optimise them. In principle, this approach would allow for better localization, if the standard head model parameters are not appropriate for a given subject.

There are occasions when the algorithm updates a dipole location to outside the head. This is of course an un-physiological solution and can happen because of two reasons. First, the algorithm uses a first-order Taylor approximation, which precludes strong non-linearity constraints at the head boundary. We therefore

check for inadmissible updates and iteratively half the step-size (up to a maximum of 16 after which the algorithm reports its failure to find a solution inside the head; cf., Levenberg–Marquardt regularisation). Secondly, updates to outside the head generally indicate that the priors are too lenient. For example, if there are too many dipoles and uninformed priors, some of the dipoles will be used to model noise, which might entail a source outside the head. This issue is also encountered with classical dipole fitting techniques. However, with our approach, one can resolve it with more informed priors.

It has been pointed out that a dipole model for a single time slice, or the average over several time slices, is not optimal for localization (Mosher et al., 1993). Given that sources are active over time, seeing the full spatiotemporal data would provide for more certainty about parameters and potentially more veridical estimates. Our model is simple, and one can think of several extensions, including a spatiotemporal model, where the temporal dynamics are described by some neural mass model (David et al., 2006; Kiebel et al., 2006). Here, our main intentions were to present a practical Bayesian routine and its application to routine evoked responses analysis. Furthermore, variational Bayes approximates the model evidence. It can be shown that this implies, for model comparison, a bias towards simple models (Beal, 2003). In this paper, we make no attempt at quantifying this bias, or using MCMC for cross-validation, but observe that for the synthetic and real data, our results are sensible and any systematic bias seems to be small.

## Conclusion

We have presented a variational Bayes approach to source localization using models consisting of a few dipoles. The inversion routines are computationally efficient and return veridical posterior distributions. We have shown, in several examples on synthetic and real data, the usefulness of a Bayesian approach using informed priors and model comparison.

## Software note

The scheme described in this note has been implemented as Matlab (MathWorks) code. The source code is available in the Statistical Parametric Mapping package (SPM5) from <http://www.fil.ion.ucl.ac.uk/spm/> as an academic freeware.

## Acknowledgments

The Wellcome Trust funded this work. CP is funded by the FNRS, Belgium. We thank Robert Oostenveld for helpful discussions. We also would like to thank Marta Garrido and James Kilner for providing the EEG data.

## Appendix A. Derivation of update rules

To derive the update rules we found the following helpful:  $a^T b = \text{tr}(ab^T)$  and  $\text{tr}(C^T B C) = \text{vec}(C^T)^T (B \otimes I) \text{vec}(C^T)$ , where  $B$  is a symmetric matrix. For terms involving the non-linear lead-field function  $G(s)$ , we use the vectorised first-order Taylor expansion

$$g(s) \approx g(\mu_s) + \frac{\partial g}{\partial s}(s - \mu_s)$$

This use of this first-order expansion is motivated by theoretical convergence results for any Gauss–Newton optimization scheme. It exploits the same simplifications afforded by the Laplace approximation in fixed-form variational schemes (e.g., Friston et al., 2007) but, in this instance, is used only to finesse the optimization of free energy.

To make the notation more compact, we use the following Kronecker tensor products:

$$E = (\mu_w \mu_w^T \otimes I)$$

$$D = (\Sigma_w \otimes I)$$

and auxiliary residual variables

$$r(w)_w = \mu_w - \mu_{w_0}$$

$$r(s)_s = \mu_s - \mu_{s_0}$$

In what follows we work though the update rules for the sufficient statistics of the conditional marginal densities, based on the equality in Eq. (8). Unless otherwise stated we omit constants (i.e., terms that are not functions of the parameter of interest) and drop the conditional dependency on the model for clarity.

#### Data precision

From Eq. (8) we have

$$\ln q(\gamma_y) = \langle \ln p(y|\gamma_y, w, s) + \ln p(\gamma_y) \rangle_{q(w)q(s)}$$

$$\ln p(y|\gamma_y, w, s) = \frac{1}{2} N_c \ln \gamma_y - \frac{1}{2} (y - G(s)w)^T \Sigma_y^{-1} (y - G(s)w)$$

$$\ln p(\gamma_y) = (a_{y_0} - 1) \ln \gamma_y - b_{y_0} \gamma_y$$

It follows that

$$\begin{aligned} \ln q(\gamma_y) = & \left( \frac{1}{2} N_c + a_{y_0} - 1 \right) \ln \gamma_y - \gamma_y \left( \frac{1}{2} (y^T y - 2\mu_w^T G(\mu_s)^T y \right. \\ & \left. + g(\mu_s)^T (D + E) g(\mu_s) + \text{tr} \left( \Sigma_s \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) \right) + b_{y_0} \end{aligned}$$

This equality has the form  $\ln q(\gamma) = (a-1)\ln\gamma - b\gamma + c$ ; from this, we conclude that  $q(\gamma_y) = Ga(\gamma_y; a_y, b_y)$  is a Gamma density, where

$$\begin{aligned} b_y = & \frac{1}{2} (y^T y - 2\mu_w^T G(\mu_s)^T y + g(\mu_s)^T (D + E) g(\mu_s) \\ & + \text{tr} \left( \Sigma_s \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right)) + b_{y_0} \end{aligned}$$

$$a_y = \frac{1}{2} N_c + a_{y_0}$$

#### Parameter precisions

We can follow a similar derivation for the prior precisions on the parameters

$$\ln q(\gamma_w) = \langle \ln p(w|\gamma_w) + \ln p(\gamma_w) \rangle_{q(w)}$$

$$\ln p(w|\gamma_w) = \frac{3}{2} N_s \ln \gamma_w - \frac{1}{2} \gamma_w r(w)_w^T \Sigma_{w_0}^{-1} r(w)_w$$

$$\ln p(\gamma_w) = (a_{w_0} - 1) \ln \gamma_w - b_{w_0} \gamma_w$$

$$\begin{aligned} \ln q(\gamma_w) = & \left( \frac{3}{2} N_s + a_{w_0} - 1 \right) \ln \gamma_w \\ & - \gamma_w \left( \frac{1}{2} (r(\mu_w)_w^T \Sigma_{w_0}^{-1} r(\mu_w)_w + \text{tr}(\Sigma_{w_0}^{-1} \Sigma_w)) \right) + b_{w_0} \end{aligned}$$

Where  $r(w)_w = \mu_w - \mu_{w_0}$ , such that  $q(\gamma_w) = Ga(\gamma_w; a_w, b_w)$  where

$$b_w = \frac{1}{2} (r_w^T \Sigma_{w_0}^{-1} r_w + \text{tr}(\Sigma_{w_0}^{-1} \Sigma_w)) + b_{w_0}$$

$$a_w = \frac{3}{2} N_s + a_{w_0}$$

Equivalently, for the precision of the location parameters;  $q(\gamma_s) = Ga(\gamma_s; a_s, b_s)$ , where

$$b_s = \frac{1}{2} (r_s^T \Sigma_{s_0}^{-1} r_s + \text{tr}(\Sigma_{s_0}^{-1} \Sigma_s)) + b_{s_0}$$

$$a_s = \frac{3}{2} N_s + a_{s_0}$$

#### Moments

For the moments the Markov blanket has three subsets, such that

$$\ln q(w) = \langle \ln p(y|w, s, \gamma_y) + \ln p(w|\gamma_w) \rangle_{q(s)q(\gamma_y)q(\gamma_s)}$$

$$\ln p(w|\gamma_w) = -\frac{\gamma_w}{2} r(w)_w^T \Sigma_{w_0}^{-1} r(w)_w$$

$$\ln p(y|w, s, \gamma_y) = -\frac{1}{2} (y - G(s)w)^T \Sigma_y^{-1} (y - G(s)w) \Rightarrow$$

$$\begin{aligned} \ln q(w) = & -\frac{a_w}{2b_w} r(w)_w^T \Sigma_{w_0}^{-1} r(w)_w - \frac{a_y}{2b_y} \left( -2w^T G(\mu_s)^T y \right. \\ & \left. + w^T G(\mu_s)^T G(\mu_s) w + \text{tr} \left( \Sigma_s \frac{\partial g^T}{\partial s} (I \otimes w w^T) \frac{\partial g}{\partial s} \right) \right) \\ = & -\frac{1}{2} w^T \left( \frac{a_w}{b_w} \Sigma_{w_0}^{-1} + \frac{a_y}{b_y} (G(\mu_s)^T G(\mu_s) + B) \right) w \\ & + w^T \left( \frac{a_y}{b_y} G(\mu_s)^T y + \frac{a_w}{b_w} \Sigma_{w_0}^{-1} \mu_{w_0} \right) \end{aligned}$$

Where, in Matlab notation, matrix  $B = \sum_{j=1}^{N_c} C(j + [0:N_p - 1]j + [0:N_p - 1])$  with  $C = \frac{\partial g}{\partial s} \Sigma_s \frac{\partial g^T}{\partial s}$ . The above expression has a quadratic form in which  $w$  means  $q(w) = N(w; \mu_w, \Sigma_w)$  is a Gaussian density, where

$$\mu_w = \Sigma_w \left( \frac{a_y}{b_y} G(\mu_s)^T y + \frac{a_w}{b_w} \Sigma_{w_0}^{-1} \mu_{w_0} \right)$$

$$\Sigma_w = \left( \frac{a_w}{b_w} \Sigma_{w_0}^{-1} + \frac{a_y}{b_y} (G(\mu_s)^T G(\mu_s) + B) \right)^{-1}$$

#### Locations

Similarly, for the locations we have

$$\ln q(s) = \langle \ln p(y|w, s, \gamma_y) + \ln p(s|\gamma_s) \rangle_{q(w)q(\gamma_y)q(\gamma_s)}$$

$$\ln p(s|\gamma_s) = -\frac{\gamma_s}{2} r(s)_s^T \Sigma_{s_0}^{-1} r(s)_s$$

$$\ln p(y|w, s, \gamma_y) = -\frac{1}{2} (y - G(s)w)^T \Sigma_y^{-1} (y - G(s)w)$$

⇒

$$\begin{aligned} \ln q(s) = & -\frac{a_s}{2b_s} r(s)_s^T \Sigma_{s_0}^{-1} r(s)_s - \frac{a_y}{2b_y} \left( -2s^T \frac{\partial g^T}{\partial s} (\mu_w \otimes y) \right. \\ & + 2s^T \frac{\partial g^T}{\partial s} (D + E)g(\mu_s) - 2s^T \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \mu_s \\ & \left. + s^T \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) \end{aligned}$$

From this, we conclude,  $q(s) = N(s; \mu_s, \Sigma_s)$ , where

$$\begin{aligned} \mu_s = & \Sigma_s \left( \frac{a_s}{b_s} \Sigma_{s_0}^{-1} \mu_{s_0} + \frac{a_y}{b_y} \left( \frac{\partial g^T}{\partial s} \left( (\mu_w \otimes y) \right. \right. \right. \\ & \left. \left. \left. - (D + E) \left( g(\mu_s) - \frac{\partial g}{\partial s} \mu_s \right) \right) \right) \right) \end{aligned}$$

$$\Sigma_s = \left( \frac{a_s}{b_s} \Sigma_{s_0}^{-1} + \frac{a_y}{b_y} \left( \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) \right)^{-1}$$

This completes the derivation of the update rules, which are summarized in Fig. 2.

## Appendix B. Derivation of the negative free energy

The [negative] free energy is given by

$$F = \int q(\theta|y) \ln \frac{p(y, \theta)}{q(\theta|y)} d\theta$$

$$= L_{av} - KL_{prior}$$

With accuracy and complexity terms given by

$$L_{av} = \int q(\theta|y) \ln p(y|\theta) d\theta$$

$$KL_{prior} = \int q(\theta|y) \ln \frac{q(\theta|y)}{p(\theta)} d\theta$$

The prior and the approximate posterior are

$$p(\theta) = p(w|\gamma_w) p(\gamma_w) p(s|\gamma_s) p(\gamma_s) p(\gamma_y)$$

$$q(\theta) = q(w) q(\gamma_w) q(s) q(\gamma_s) q(\gamma_y)$$

Giving

$$KL_{prior} = KL(w) + KL(s) + KL(\gamma_w) + KL(\gamma_s) + KL(\gamma_y)$$

$$L_{av} = \int q(w) q(s) q(\gamma_y) \ln p(y|\theta) dw ds d\gamma_y$$

After similar calculations to those presented in Appendix A, we get

$$\begin{aligned} F = & -\frac{N_c}{2} \ln 2\pi + \frac{N_c}{2} (\Psi(a_y) - \ln(b_y)) \\ & - \frac{a_y}{2b_y} \left( y^T y - 2y^T G(\mu_s) \mu_w + g(\mu_s)^T (D + E) g(\mu_s) \right. \\ & \left. + \text{tr} \left( \Sigma_s \frac{\partial g^T}{\partial s} (D + E) \frac{\partial g}{\partial s} \right) \right) \end{aligned}$$

where  $\psi(\bullet)$  is the digamma function. Note that the free energy bound on the log-evidence is not an explicit function of the hyperpriors; this is because their effect is mediated through the conditional density on the parameters.

## Appendix C. Projection matrices and hard constraints

To enforce  $N$  symmetric dipoles, one can define the following matrices for location and moments:

$$T_s = T_w = \frac{1}{2} I_N \otimes T T^T \quad T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$T_s$  is an idempotent rank-deficit projector matrix (*i.e.*,  $T_s = T_s T_s$ ) that effectively removes unwanted degrees of freedom from the location and moment vectors. One can also think of this constraint as a re-parameterization; in which one dipole comes to encode the spatial deployment of two symmetric dipoles; *i.e.*,  $(I_N \otimes I_3)w \rightarrow (\frac{1}{2} I_N \otimes T T^T)w$ . Practically, the updates of the posterior means are pre-multiplied by  $T_s$ ; *i.e.*,  $\mu_s \rightarrow T_s \mu_s$  and the covariances are sandwiched to give  $\Sigma_s \rightarrow T_s \Sigma_s T_s^T$  (similarly for the moment parameters). Suitable matrix inversions (*i.e.*, pseudo-inverses) are required during the updates, if rank-deficient covariance matrices are induced in this way. This procedure can also be applied to other hard constraints, implemented by a user-specified  $T$ .

## References

- Auranen, T., Nummenmaa, A., Hamalainen, M.S., Jaaskelainen, I.P., Lampinen, J., Vehtari, A., Sams, M., 2007. Related Articles, Links Abstract Bayesian inverse analysis of neuromagnetic data using cortically constrained multiple dipoles. *Hum. Brain Mapp.* 28 (10), 979–994 (Oct).
- Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomofunctional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* 44, 374–385.
- Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 18, 14–30.
- Beal, M.J., 2003. Variational Algorithms for Approximate Bayesian Inference. University College, London.
- Braun, C., Kaiser, S., Kincses, W.E., Elbert, T., 1997. Abstract Confidence interval of single dipole locations based on EEG data. *Brain Topogr.* 10 (1), 31–39 (Fall).
- Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pelegrini-Issac, M., Lina, J.M., Benali, H., 2007. Free Full Text Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. *NeuroImage* 36 (1), 69–87 (May 15).
- David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Abstract Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage* 30 (4), 1255–1272 (May 1).
- Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., Lueschow, A., 2007. Abstract MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *NeuroImage* 35 (4), 1495–1501 (May 1).
- Flandin, G., Penny, W.D., 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage* 34, 1108–1125.
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *NeuroImage* 19, 1273–1302.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007.

- Variational free energy and the Laplace approximation. *NeuroImage* 34, 220–234.
- Fuchs, M., Wagner, M., Kastner, J., 2004. Confidence limits of dipole source reconstruction results. *Clin. Neurophysiol.* 115, 1442–1451.
- Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. *NeuroImage* 36, 571–580.
- Huang, M., Aine, C.J., Supek, S., Best, E., Ranken, D., Flynn, E.R., 1998. Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography. *Electroencephalogr. Clin. Neurophysiol.* 108, 32–44.
- Jun, S.C., George, J.S., Pare-Blagoev, J., Plis, S.M., Ranken, D.M., Schmidt, D.M., Wood, C.C., 2005. Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data. *NeuroImage* 28, 84–98.
- Jun, S.C., George, J.S., Plis, S.M., Ranken, D.M., Schmidt, D.M., Wood, C.C., 2006. Improving source detection and separation in a spatiotemporal Bayesian inference dipole analysis. *Phys. Med. Biol.* 51, 2395–2414.
- Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* 91, 1343–1370.
- Kiebel, S.J., David, O., Friston, K.J., 2006. Abstract Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage* 30 (4), 1273–1284 (May 1).
- Lutkenhoner, B., Steinstrater, O., 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neuro-otol.* 3, 191–213.
- Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. *NeuroImage* 30, 753–767.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39, 541–557.
- Mosher, J.C., Spencer, M.E., Leahy, R.M., Lewis, P.S., 1993. Error bounds for EEG and MEG dipole source localization. *Electroencephalogr. Clin. Neurophysiol.* 86, 303–321.
- Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. *IEEE Trans. Biomed. Eng.* 46, 245–259.
- Nummenmaa, A., Auranen, T., Hamalainen, M.S., Jaaskelainen, I.P., Lampinen, J., Sams, M., Vehtari, A., 2007. Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *NeuroImage* 35, 669–685.
- Oostenveld, R., 2003. Improving EEG Source Analysis using Prior Knowledge. Thesis, Katholieke Universiteit Nijmegen.
- Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719.
- Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. *NeuroImage* 19, 727–741.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. *NeuroImage* 22, 1157–1172.
- Phillips, C., Rugg, M.D., Friston, K.J., 2002. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. *NeuroImage* 16, 678–695.
- Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J., 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. *NeuroImage* 24, 997–1011.
- Radich, B.M., Buckley, K.M., 1995. EEG dipole localization bounds and MAP algorithms for head models with parameter uncertainties. *IEEE Trans. Biomed. Eng.* 42, 233–241.
- Sato, M.A., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage* 23, 806–826.
- Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Supek, S., Aine, C.J., 1993. Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution. *IEEE Trans. Biomed. Eng.* 40, 529–540.
- Woolrich, M.W., Behrens, T.E., 2006. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Trans. Med. Imag.* 25, 1380–1391.
- Yvert, B., Crouzeix-Cheylus, A., Pernier, J., 2001. Fast realistic modeling in bioelectromagnetism using lead-field interpolation. *Hum. Brain Mapp.* 14, 48–63.
- Zhang, Z., 1995. A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres. *Phys. Med. Biol.* 40, 335–349.