Technical Note

Electromagnetic source reconstruction for group studies

Vladimir Litvak *, Karl Friston

The Wellcome Trust Centre for Neuroimaging, University College London, UK

ARTICLE INFO

ABSTRACT

The aim of this paper is to describe a simple procedure for electromagnetic (EEG or MEG) source reconstruction, in the context of group studies. This entails a simple extension of existing source reconstruction techniques based upon the inversion of hierarchical models. The extension ensures that evoked or induced responses are reconstructed in the same subset of sources, over subjects. Effectively, the procedure aligns the deployment of reconstructed activity over subjects and increases, substantially, the detection of differences between evoked or induced responses at the group or between-subject level.

© 2008 Elsevier Inc. All rights reserved.

Introduction

There has been considerable progress over the past few years in reconstructing electromagnetic source activity from channel data (*e.g.*, Baillet and Garnero, 1997; Phillips et al., 2002; Phillips et al. , 2005; Sato et al., 2004; Serinagaoglu et al., 2005; Trujillo-Barreto et al., 2004; Jun et al., 2006; Mattout et al., 2006; Nummenmaa et al., 2007; Wipf and Nagarajan, in press). These advances rest upon reformulating classical inversion approaches within a Bayesian framework and then using hierarchical or empirical Bayes (Kass and Steffey, 1989; Phillips et al., 2005; Wipf and Nagarajan, in press) to optimise both the estimated sources *and the constraints or priors themselves*. This optimisation uses standard sampling or variational schemes to provide the most likely distribution of source activity, given observed data. While these reconstructions are optimal for data from a single subject or session, they are not optimal for detecting group differences. This is because the hierarchical model does not include a between-subject level and is therefore unable to exploit the fact that the same paradigm was repeated over subjects or sessions. The analysis described in this paper is based on two established approaches: the summary-statistic approach to hierarchical models and the use of a canonical mesh in source reconstruction.

The summary-statistic approach

The problem of making inferences about population responses is resolved by hierarchical models. In classical statistics, these are known as random or mixed-effects models that include random effects at the within and between-subject level. In Bayesian inference these are known as parametric empirical Bayes models, where between-subject variations in parameter estimates furnish constraints on within-subject estimation. Usually, these random effects or hierarchical models are inverted using data from all the subjects analysed. An alternative approach is to optimise the parameters of subject-specific models separately and use these optimised parameters as summary-statistics, which are then analysed in a second-level or between-subject model. This can be repeated until the top level of the hierarchical model is reached. Provided the models for each subject are the same, the summary-statistic approach gives, in expectation, exactly the same results as a full mixed-effects analysis. Furthermore, even if there are differences between the models at the single-subject level; the differences between the summary-statistic approach and a mixed-effects analysis are trivial (see Friston et al., 2005 for example in fMRI). In neuroimaging people generally use the summary-statistic approach because, computationally and conceptually, it is much simpler to implement; and because it avoids handling all the data from all the subjects at the same time. This is important in neuroimaging, where the size of datasets can be very large.

---

* Corresponding author. The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London, WC1N 3BG, UK. Fax: +44 207 813 1445.
E-mail address: v.litvak@ion.ucl.ac.uk (V. Litvak).

In the analysis of group EEG and MEG studies, we anticipate that people will want to use a summary-statistic approach, in which a summary of evoked responses (within some time and frequency window) for each subject is optimised at the within-subject level and then passed to a simple ANOVA model at the second level, to produce statistical parametric maps (SPM) of regionally significant differences among trials. In this particular application, one would use Bayesian estimators as the summary-statistic at the within-subject level and classical statistics for inference at the between-subject level. However, for this two-stage procedure to work, the reconstructed activity at a particular source in one subject should correspond to the same source in the other subjects. This brings us to the notion of a canonical mesh.

The canonical mesh

To facilitate group-level analyses, we introduced the idea of a canonical mesh (Mattout et al., 2007), which ensures that activity is reconstructed in the same source space over subjects (Talairach and Tournoux, 1988). Briefly, the forward model for each subject starts with a canonical mesh, defining a lattice of sources on the cortical surface. This mesh is then warped using an inverse spatial normalization so that the canonical mesh is roughly in the same place as the subject's cortical sheet. This means that, after inversion of the ensuing forward model, reconstructed activity can be assigned to the same sources over subjects. One can then project this two-dimensional manifold into a three-dimensional image and proceed in the usual way to make inferences about regionally-specific effects using SPM. However, for this to be efficient, one needs to suppress inter-subject variability in the spatial deployment of reconstructed activity. This can arise from the underdetermined aspect of the inverse problem and calls for formal constraints on single-subject reconstructions. These constraints or hyperpriors are the subject of this technical note.

Hyperpriors

The importance of hierarchical or empirical models for single-subject data rests upon optimising not just the parameters of the model (like source activity) but also the priors themselves. This optimisation proceeds under hyperpriors, which can be regarded as priors or constraints on the prior expectations about the pattern of source expression. Critically, the hyperpriors can also be optimised to select the best hyperprior or models (*c.f.*, model selection). A simple example of different models would be the distinction between a conventional minimum norm prior and multiple sparse priors. In the former, the prior constraint is that activity has exactly the same variance at every source and, furthermore, is independent. This is encoded by a prior covariance on source space that corresponds to the identity matrix; *i.e.*, $Q_1^U = I$. Conversely, in multiple sparse priors the hyperprior is that activity can be expressed in multiple patches or covariance components, $Q_1^U, ..., Q_m^U$, each of which has an associated hyperparameter (Friston et al., 2008a). The difference between minimum norm and multiple sparse prior models rests on the form or number of prior covariance components or, equivalently, the number of hyperparameters. These hyperparameters can be optimised using a variety of techniques. Recently we introduced a modelling approach that is

formally identical to restricted maximum likelihood for optimising hundreds of hyperparameters, under multiple sparse priors. The sparse aspect speaks to the fact that many of these prior covariance components are unnecessary or irrelevant and can be switched off (*c.f.*, automatic relevance determination; ARD; Neal, 1998). This automatic hyperprior or model selection depends on choosing the right objective function (a variational bound on the model's log-evidence that includes weakly informative hyperpriors). An alternative approach to optimising the hyperpriors is to use a greedy search as described in Friston et al. (2008b). The distinction between ARD and greedy search optimisation of models is relatively simple: in ARD one assumes that every hyperparameter has a different value. These hyperparameters are then optimised so that many fall to zero, leaving a small number of relevant or important empirical priors (Tipping, 2001). In a greedy search, one starts by assuming all the hyperparameters have the same value and then splits the set of hyperparameters recursively, until the objective function stops increasing (Friston et al., 2008b). We will see examples of minimum norm, ARD and greedy search schemes later.

The particular hyperprior we propose here imposes consistency on the source reconstruction over subjects or sessions. It is a formal constraint that assumes that the prior variance at any particular source in any particular subject can be factorised into a subject-specific and source-specific term. In other words, we assume that the unknown empirical prior variance is proportionally the same over sources but the variance *per se* is scaled in a subject-specific fashion. This simple assumption allows us to estimate source-specific hyperparameters by pooling the data from all trial types and subjects to optimise the form of the prior covariance over sources. This empirical spatial prior then enters a series of single-subject inversions, which optimises the weight afforded to this prior, in relation to subject-specific measurement noise. This formal constraint ensures that the source reconstruction in each individual analysis is confined to the same sources, without biasing the parameter estimates themselves. This means that we can use the parameter estimates as summary-statistics for valid inference at the between-subject level. Note that this does not represent an inversion of a full hierarchical model. Our objective is to finesse the first-level Bayesian inversion to provide useful summary-statistics for classical inference at the second level. By pooling the data to estimate empirical priors on source space, we can meet this objective simply and effectively.

This paper comprises two sections. In the first, we present the details of the model inversion. This inversion has two stages. The first optimises the empirical prior covariance over subjects. The second entails inverting the data from each subject using the empirical prior from the pooled analysis. We will briefly describe the implicit hierarchical models and the optimisation scheme that we use. In the second section, we will present comparative analyses of between-subject or group-level inference with and without between-subject constraints. We would also take the opportunity to compare the performance of three hyperpriors: a minimum norm hyperprior with a single hyperparameter (this can be regarded as the baseline hyperprior). Second, we will look at multiple sparse priors that have been optimised using ARD. Finally, we will consider the hyperpriors furnished by a greedy search over successive partitions of multiple sparse priors. The formal details concerning these sparse priors will be found

in Friston et al. (2008a) and the technical details of the greedy search will be found in Friston et al. (2008b). In short, we will compare three hyperprior models of source activity with and without between-subject constraints. The outcome measures we use reflect the quality of each model at the within-subject level and the efficiency of detecting group effects at the between-subject level. The within-subject measure is simply the log-evidence for each of the six (2×3) models for each subject (actually five models because the minimum norm is the same with and without constraints — see below). For the between-subject analysis we use two measures based on between-subject SPMs from each of the models. These measures reflect sensitivity (peak t-values) and spatial accuracy (location of peaks). The dataset we use involves a simple sensory evoked response following median nerve stimulation. We chose these data because the location of the underlying sources is well known and allows us to interpret our comparative analyses with greater confidence.

Theory

Covariance function models

We start with a hierarchical generative model for data $Y_i = \mathfrak{R}^{c \times n}$ over c channels and n peristimulus time-bins, from $i = 1, \ldots, N$ trials or subjects[1]. We assume that random effects on the sources factorise onto subject-specific $P(\nu^P)$ and source-specific terms $U(\nu^U)$, such that their covariance is $P \otimes U$. Here, $\nu^{\cdot}$ are unknown scale parameters that control the covariance or amplitude of random effects in source space. This assumption engenders a probabilistic generative or forward model with empirical (*i.e.*, hierarchical) priors on sources J. These sources are mapped to channels by subject-specific lead-field or gain matrices, $L_1, \ldots, _N$ to give the following linear model; $Y = LJ + \varepsilon$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} L_1 & & \\ & \ddots & \\ & & L_N \end{bmatrix} \begin{bmatrix} J_1 \\ \vdots \\ J_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (1)$$

Where, under Gaussian assumptions about random effects,

$$p(Y|J, \lambda) = N(LJ, V, C)$$
$$p(J) = N(0, V, P \otimes U)$$
$$p(\lambda) = N(\eta, R)$$

$$\begin{aligned} C &= diag(v_1^C, \ldots, v_N^C) \otimes Q^C \\ U &= v_1^U Q_1^U + \ldots + v_M^U Q_M^U \\ P &= diag(v_1^P, \ldots, v_N^P) \\ v &= \exp(\lambda) \end{aligned} \quad (2)$$

Here, $p(\varepsilon) = N(0, V, C)$ denotes a zero-mean Gaussian density on a matrix with temporal correlations, V and spatial covariances among channels, C. In this model, Q^C represents a single covariance component over channels, which is scaled by a subject-specific covariance parameter or hyperparameter v_i^C to give the covariance of channel noise[2]. Similarly, Q_j^U represent covariance components in source space that are

scaled by $v_i^P v_j^U$ to give the contribution of the j-th component to the i-th subject's source activity. This scaling by a product of two scale parameters embodies the formal hyperprior that source activity factorises into source and subject-specific components.

Note that the unknown hyperparameters $\lambda_i^{\cdot} = \ln(v_i^{\cdot})$ have their own prior densities, $p(\lambda_i^{\cdot})$, which place non-negative lognormal hyperpriors on the corresponding scale parameters v_i. In what follows, we assume that V is a fixed Gaussian autocorrelation matrix with a smoothness that corresponds to any filtering applied during pre-processing. We will also assume that $Q^C = I$ and select the source components $Q_1^U, \ldots, Q_M^U$ according to the model m required (*e.g.*, multiple sparse priors, minimum norm *etc.*). We use weakly informative hyperpriors with $\eta = -16$ and $R = 32I$ that tend to eliminate unnecessary components.

Gaussian process models

This model may look complicated but it can be reduced to a very simple form, which is easy to invert. This form is a Gaussian process or covariance function model, which effectively expresses the expected sample covariance of the data as a covariance function of the hyperparameters. Note that in this form, the parameters are eliminated and we have only to optimise the hyperparameters. For a single subject, the sample covariance of spatiotemporally 'normalised' responses $S(\tilde{Y}_i)$ and its expectation $\tilde{\Sigma}(\lambda_i)$ are

$$S(\tilde{Y}_i) = \frac{1}{n} \tilde{Y}_i V^{-1} \tilde{Y}_i^T$$

$$\begin{aligned} \tilde{\Sigma}(\lambda_i) &= v_i^P \left(v_1^U \tilde{Q}_1^U + \ldots + v_M^U \tilde{Q}_M^U \right) + v_i^C \tilde{Q}_i^C \\ \tilde{Q}_k^U &= G Q_k^U G^T \\ \tilde{Q}_i^C &= A_i Q^C A_i^T \end{aligned} \quad (3)$$

Essentially, this says that that the predicted covariance of normalised channel data, $\tilde{Y}_i = A_i Y_i$ over time, is a mixture of covariance components from source space and measurement space. The normalization matrix $A_i = G L_i^T (L_i L_i^T)^{-1} \Leftarrow A_i L_i = G$ effectively re-samples the channel data to return what would have been observed if the gain-matrix was an arbitrary matrix, G. We usually use the gain-matrix from the first subject; *i.e.*, $G = L_1$. This re-sampling is useful because it allows us to specify a covariance function model for the sample covariance over subjects

$$S(\tilde{Y}) = \sum_i S(\tilde{Y}_i)$$

$$\begin{aligned} \tilde{\Sigma}(\lambda) &= \bar{v}_1^U \tilde{Q}_1^U + \ldots + \bar{v}_M^U \tilde{Q}_M^U + v_1^C \tilde{Q}_1^U + \ldots + v_N^C \tilde{Q}_N^U \\ \bar{v}_k^U &= \bar{v}^P v_k^U \\ \bar{v}^P &= v_1^P + \ldots + v_N^P \end{aligned} \quad (4)$$

Again, the covariance function $\tilde{\Sigma}(\lambda)$ is a mixture of components from source and measurement space; these are a mixture of M source-specific spatial components and N subject-specific sensor-noise components. In this model, the hyperparameters $\lambda_k^U = \ln \bar{v}_k^U$ encoding contributions from source space are the same over subjects; it is these we seek to provide an empirical prior on source space that is conserved over subjects. We now have a Gaussian process model for a group of subjects that is easy to invert or optimise.

---

[1] We will treat trials and subjects as interchangeable for simplicity. In practice, one would normally invert several trial types (*i.e.*, conditions) from several subjects by holding the subject-specific quantities constant over trials.

[2] The scheme can be extended easily to model multiple covariance components for channel noise.

Optimising Gaussian process models

Covariance function formulations of linear models are important because their hyperparameters can be optimised using standard covariance component estimation techniques to maximise the model evidence, $p(Y|m)$ for any data Y under model m[3]. Once the hyperparameters have been optimised the conditional density of the parameters is easy to compute, either for the group average or each subject (see below). In this context, optimisation uses an augmented restricted maximum likelihood scheme (Harville, 1977), as described in Friston et al. (2007). This scheme is formally equivalent to a variational inversion under the mean-field assumption that the conditional density $q(J,\lambda)=q(J)q(\lambda)$ factorises into Gaussian marginals. Here, $q(J)=N(\mu^J,\Sigma^J)$, where μ^J and Σ^J are the conditional expectation and covariance of parameters; similarly for the hyperparameters, $q(\lambda)=N(\mu^\lambda,\Sigma^\lambda)$. Under this Gaussian or Laplace assumption, the hyperparameters of any Gaussian process model maximise a variational free-energy bound on the log-evidence[4]

$$F = -\frac{D}{2}tr\left(\Sigma^{-1}S\right) - \frac{D}{2}\ln|\Sigma| + \frac{1}{2}\ln|\Sigma^\lambda\Pi| - \frac{1}{2}\left(\mu^\lambda-\eta\right)^T\Pi\left(\mu^\lambda-\eta\right) \quad (5)$$

Where, D represents the total number of samples used to evaluate the sample covariance matrix[5] and $\pi=R^{-1}$ is the prior precision of the hyperparameters from Eq. (2). This bound is a function of sample and predicted covariances, $S(Y)$ and $\Sigma(\mu^\lambda)$ respectively. Optimisation is relatively simple and involves iterating

$$F_{\lambda i} = -\frac{D}{2}tr(P_i(S-\Sigma)) - \Pi_{ii}(\mu_i^\lambda-\eta_i)$$
$$F_{\lambda\lambda ij} = -\frac{D}{2}tr(P_i\Sigma P_j\Sigma) - \Pi_{ij}$$

$$\Delta\mu^\lambda = -F_{\lambda\lambda}^{-1}F_\lambda$$
$$\Sigma^\lambda = -F_{\lambda\lambda}^{-1} \quad (6)$$

until convergence. The matrix P_i is the derivative of the precision, $\Sigma(\mu^\lambda)^{-1}$ with respect to the i-th hyperparameter. F_λ and $F_{\lambda\lambda}$ are the gradient and expected curvature of the free-energy. Note that this scheme deals with relatively small $c\times c$ matrices whose size corresponds to the number of channels. In other words, computational load does not scale with the number of time-bins, trials or subjects.

Put simply, this procedure optimises the amount of each covariance component iteratively, to minimise the difference between the sample covariance and the ensuing mixture of components. The difference is essentially the Kullback-Leibler divergence between the sample and predicted Gaussian densities encoded by the covariances (Harville, 1977).

A two-stage optimisation

The hyperparameters of our hierarchical model can be optimised in two stages. First, we estimate the hyperparameters of $\tilde{\Sigma}(\lambda)$, using the sample covariance over subjects

S ($\tilde{Y}$) to furnish the conditional estimates of source-specific scale parameters, $\bar{v}_1^U, \ldots \bar{v}_M^U$. These define an empirical prior covariance on the sources $\overline{U} = \bar{v}_1^U Q_1^U + \ldots + \bar{v}_M^U Q_M^U$ that is common to all subjects (to within a scaling factor). Second, we optimise subject-specific hyperparameters using the sample covariance of single-subject data $S(Y_i)$ and the covariance function model

$$\Sigma_i(\lambda) = v_i^P L_i\overline{U}L_i^T + v_i^C Q^C \quad (7)$$

Here, the only unknowns are the subject-specific error covariance and source-prior hyperparameters. These are optimised in exactly the same way as above to give the conditional estimates of $\lambda_i=\{v_i^P,v_i^C\}$. These specify the *maximum a-posterior* (*MAP*) estimates of source activity using the matrix inversion lemma.

$$\mu_i^J = M_iY_i$$
$$\Sigma_i^J = U_i - M_iL_iU_i. \quad (8)$$

$M_i = U_iL_i^T\Sigma_i^{-1}$ is a *MAP* projector, where the covariance function $\Sigma_i(\lambda)$ and empirical prior covariance $U_i(\lambda_i) = v_i^P\overline{U}$ are evaluated at the conditional modes of λ_i.

Time–frequency contrasts of these subject and trial-specific source reconstructions can then be used to summarise the subject-specific responses to each trial type. These contrasts generally test for specific time–frequency components by defining a temporal subspace of interest (*e.g.*, gamma oscillations between 300 and 400 ms after stimulus onset). The contrast matrix can be a simple vector; for example a Gaussian window $W\in\mathfrak{R}^{n\times1}$ over a short period of peristimulus time or cover specified frequency ranges (with one frequency per column) over extended periods of peristimulus time (Kiebel and Friston, 2004). The conditional expectation of the energy in a contrast is

$$E\left(J_iWW^TJ_i^T\right) = M_iY_iWW^TY_i^T\tilde{M}_i^T + \Sigma_i^J tr\left(W^TVW\right) \quad (9)$$

See Friston et al. (2006) for details. The conditional estimates of contrast energy can then be used as summaries of condition-specific responses for each subject and entered into statistical models of between-subject responses in the usual way. Eq. (9) provides more accurate estimates than classical beam-forming approaches because uncertainty about the activity enters the estimate of power or energy in a contrast. In other words, the expected energy is not just the energy of expected activity (as assumed in beam-forming) but involves an extra term that accounts for variability in the expected activity (the second term above). In the next section, we will use the equations above to estimate summary-statistics for evoked responses using a series of different models that are defined by the prior covariance components employed.

Prior covariance components

For the minimum norm model there is only one prior covariance component that corresponds to the identity matrix, $Q_1^U=I$. In this simplest case, the between-subject constraint should make no difference to the within-subject estimates of activity. However, for the ARD and greedy search schemes, which use multiple prior components, the effects of between-subject constraints could be quite profound. The

---

multiple sparse priors we use in this paper are described in Friston et al. (2008a) Briefly, they comprise 256 small patches with compact spatial support in each hemisphere and a further 256 components corresponding to a bilateral deployment of homologous patches. Each patch is formed by diffusing a point source on the cortical mesh, using a graph Laplacian with a diffusion coefficient of 0.6. This coefficient is basically a smoothness parameter and can vary between zero and one. In practice, we do not use the data from all peristimulus times but project the responses onto a temporal subspace using a modified Kaiser criterion. This subspace is defined by singular value decomposition of the temporal responses over trials and subjects. This usually identified about four to eight temporal modes that span more than 95% of the data variance. This temporal subspace also precludes frequencies lower than 1 Hz and higher than 64 Hz. This concludes the theoretical background. In the next section, we evaluate the between-subject constraint under the three different hyperprior models detailed above.

Comparative analyses

In this section, we present comparative analyses of the source reconstructions at the within and between-subject level. Data were acquired as a part of previously published study (Litvak et al., 2007). The original study examined the effect of combined somatosensory and transcranial magnetic stimulation (TMS) on median nerve somatosensory evoked potentials (MN-SSEP). The MN-SSEP were recorded before and after TMS. Each of eleven healthy volunteers (eight men and three women), aged 20.9 to 44.5 (mean 26.3±7.2 years) participated in five experiments, which differed in the TMS parameters. For the present analysis, we use only MN-SSEP recorded before TMS. Thus, our dataset consisted of five MN-SSEP recordings per subject.

Data acquisition and pre-processing

The recordings were performed under identical conditions with intervals of several days to one year between sessions. Median nerve stimulation was performed using an electrical stimulator (Digitimer D7AH, Digitimer, Welwyn Garden, Hertfordshire, UK) with a standard stimulation block (cathode proximal), pulse width 200 μs at a frequency of 3 Hz and a stimulation intensity of 300% of the individual perceptual threshold. Eighteen hundred stimuli were delivered. Electroencephalographic (EEG) signals were recorded from the scalp with a 64-channel Quickamp system (Brain Products GmbH, Munich, Germany). The electrodes were placed at extended 10–20 system locations and fixed on the subject's head with an elastic cap (EASYCAP GmbH, Herrsching-Breitbrunn, Germany). The electrode impedance was maintained below 5 kΩ. Electrode positions and anatomic reference points were measured for each experiment with a 3D navigation system (Brainsight, Rogue research, Montreal, Canada). The brain signals were average referenced in hardware, filtered between 0 and 560 Hz and sampled at 2000 Hz.

The data were processed with the SPM software package (http://www.fil.ion.ucl.ac.uk/spm). The continuous EEG data were high-pass filtered (above a cut-off frequency 20 Hz) and epoched between −50 and 50 ms relative to the median nerve stimulus. Trials were rejected if they contained deflections exceeding 50 μV. Channels were marked as bad if they contained such deflections in more than 20% of the trials. The number of trials retained for analysis was 1545±242. No channels were excluded in 34 out of 55 experiments, 1 channel in 12 cases, 2 channels in 5 cases, 3 channels in 2 cases and 6 channels in 2 cases. Retained trials were averaged to obtain MN-SSEP waveforms and used, in combination with measured positions of the retained sensors and the fiducials, as input for SPM source reconstruction.

We chose this dataset for the present study for two reasons. The first is that MN-SSEP sources have been studied extensively with invasive and non-invasive recordings in humans and animals; they are known to reside in the hand area of the primary somatosensory cortex (S1), principally in Brodmann area 3b, which is located at the posterior wall of the central sulcus; with a weaker and slightly delayed source in adjacent Brodmann area 1, which is located on the top of the postcentral gyrus (Allison et al., 1991). The second reason is that this particular dataset is especially suitable for testing source reconstruction with group constraints. By selecting one of the five experiments from each of the eleven subjects, 5^{11} different group MN-SSEP datasets can be formed. Ideally, source reconstructions and statistical tests performed on these group-data should yield consistent results. By looking at how variable the results are, under different hyperpriors, the utility of these models can be assessed. In what follows, an 'experiment' means a single recoding session subtending a subject-specific MN-SSEP and 'group' refers to collection of experiments from eleven different subjects.

Source reconstruction

The somatosensory evoked activity was reconstructed from 10 to 40 ms after stimulation. This reconstruction proceeded under the three hyperprior models: minimum norm (MN), multiple sparse priors optimised with ARD and multiple sparse prior partitions optimised with a greedy search (GS). Note that the ARD and GS implicitly optimise both the empirical priors and the model itself. This is because when a prior component is switched on or off, the form of the model changes. Conversely, minimum norm reconstructions with and without group constraints are identical. Thus ARD and GS reconstruction schemes were repeated with and without the formal constraint that imposes the same empirical spatial prior over subjects. In total, we examined five source reconstruction schemes: MN, ARD, GS, ARD with group constraints (gARD) and GS with group constraints (gGS).

Reconstruction without group constraints was performed once for each experiment and each hyperprior model. For reconstruction with group constraints, 100 groups of experiments (each with N = 11 subjects) were randomly sampled from the 5^{11} possible groups. Thus, each experiment was used in approximately one fifth of group reconstructions (i.e., ~20 reconstructions per hyperprior model). The same 100 groups were used for all reconstruction schemes and t-tests (see below). The reconstructions with and without group constraints used exactly the same two-stage procedure but the unconstrained reconstruction used only one subject per group (our software implementation uses just one routine that can be called with one or more subjects). Critically, this means that the final reconstruction (stage 2; in Fig. 1) used just two covariance components (an empirical source prior and sensor noise). This means that models, with and without constraints had the same number of unknown parameters and hyperparameters.
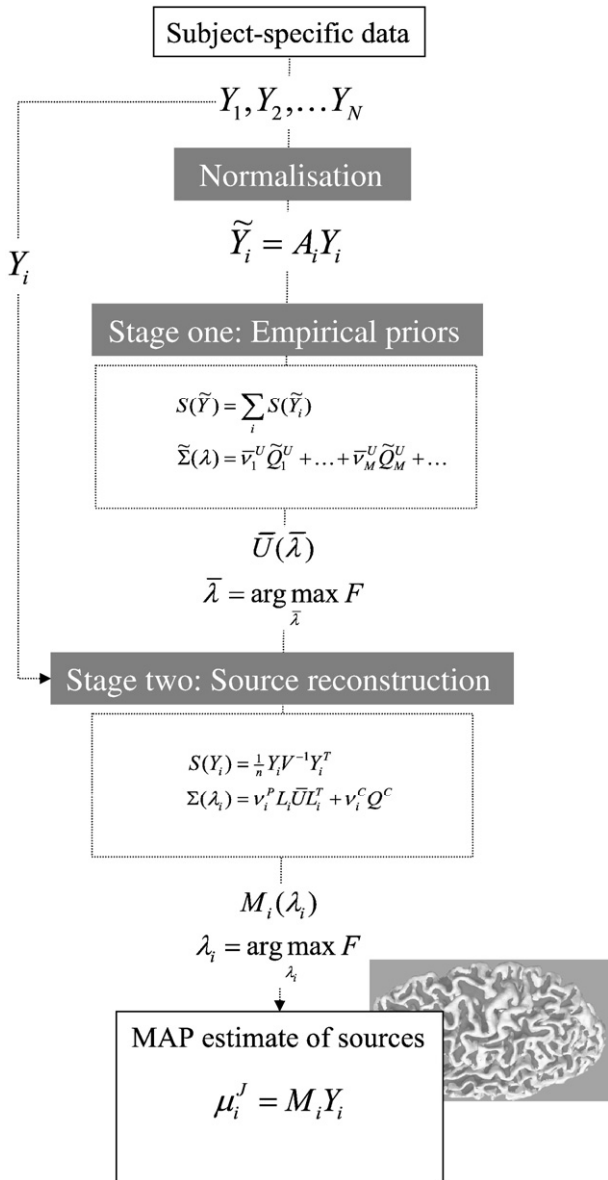
Subject-specific data

$$Y_1, Y_2, \ldots Y_N$$

Normalisation

$$\widetilde{Y}_i = A_i Y_i$$

Y_i

Stage one: Empirical priors

$$S(\widetilde{Y}) = \sum_i S(\widetilde{Y}_i)$$

$$\widetilde{\Sigma}(\lambda) = \bar{v}_1^U \widetilde{Q}_1^U + \ldots + \bar{v}_M^U \widetilde{Q}_M^U + \ldots$$

$$\bar{U}(\bar{\lambda})$$

$$\bar{\lambda} = \arg\max_{\bar{\lambda}} F$$

Stage two: Source reconstruction

$$S(Y_i) = \tfrac{1}{n} Y_i V^{-1} Y_i^T$$

$$\Sigma(\lambda_i) = v_i^P L_i \bar{U} L_i^T + v_i^C Q^C$$

$$M_i(\lambda_i)$$

$$\lambda_i = \arg\max_{\lambda_i} F$$

MAP estimate of sources

$$\mu_i^J = M_i Y_i$$

Fig. 1. Schematic of the two-stage source reconstruction scheme for group studies. See main text for a detailed explanation of the variables.

Model comparison at the within-subject level

The reconstructed activities per experiment were summarized with a Gaussian time-window centred at 20 ms (with a standard deviation of 8 ms), which is about the time the maximum sensory evoked response is expressed in somatosensory cortex. For each reconstruction, we recorded the free-energy bound on log-evidence (F). In the case of reconstructions with group constraints, the values of F for each experiment were averaged over all reconstructions in which that experiment was used. The log-evidence provides a measure of the model quality that accommodates both accuracy and complexity (*i.e.*, number of priors). Critically, it pertains to the ability of models to explain the data from each experiment but does not reflect those models' generalisability over different datasets. We expected that ARD and GS schemes would provide much better explanations of data, relative to the rather unlikely priors entailed by the MN model. Furthermore, we anticipated that the group constants would *reduce* the log-evidence for ARD and GS, because they constrain the optimisa-

tion of subject-specific parameters, when trying to explain a particular experiment. In other words, the advantage of group constraints should only be evident at the between-subject or group level and may even compromise the optimisation at the within-subject or experiment level.

We were interested in relative, rather than absolute values of the free-energy and primarily in the comparison between the GS and ARD schemes, with and without group constraints. Therefore, we analysed the differences between the *F*-values for minimum norm and for the other four schemes for each experiment. We also subjected the *F*-values obtained with GS and ARD to an ANOVA with two factors (hyperprior model; ARD *vs.* GS and between-subject constraint; with *vs.* without). This allowed us to assess the overall effect of different models, group constraints and how the model-effect depends on group constraints.

Model comparison at the between-subject level

To assess the quality of reconstructed source activity we performed paired samples *t*-tests using conventional statistical parametric mapping (SPM) procedures to compare the original contrast images and the same images flipped across the midsagittal plane. This is a device to assess the lateralisation of responses evoked by unilateral median nerve stimulation. The contrast images were constructed after diffusion on the canonical mesh and projection into three-dimensional anatomical space.

For each reconstruction scheme 100 *t*-tests were performed using the 100 group-datasets. The statistical images produced by SPM were assessed with respect to two criteria: the maximal *t*-value in each SPM and the distance of this maximum from the maximum in the average SPM, over the 100 replications of the same reconstruction scheme (we will refer to this as the mean location). These two measures, the height of the maximum *t*-value and its distance from the
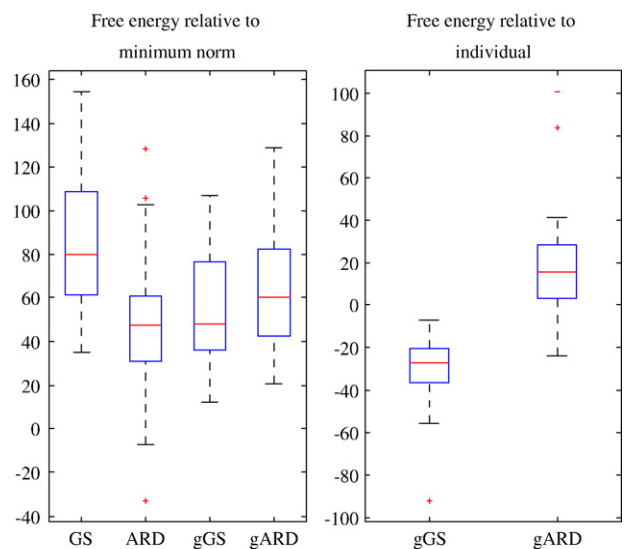
Fig. 2. (A) Distributions of the values of free-energy bound to the log-evidence for the four reconstruction methods compared in the present study relative to minimum norm. The distributions are presented as box-plots with lines at the lower quartile, median, and upper quartile values. Values beyond the 1.5 interquartile range from the median are shown as outliers. For gGS and gARD the value for each experiment is an average of ~20 values obtained when that experiment participated in reconstructions with group constraints. (B) Distributions of differences between the average log-evidences and the value obtained for source reconstruction without group constraints.

mean location, quantify the sensitivity of the between-subject analysis and its spatial precision. We anticipated that the best model would co-localise reconstructed activity in a consistent way over subjects and produce the largest *t*-values with the least dispersion across repetitions in unilateral somatosensory cortex. To exploit the fact that we had 100 realisations of group data, we present the distribution of these two metrics in terms of their cumulative sample density.

Results

Model comparison at the within-subject level

As expected, the log-evidence for the reconstruction schemes using multiple sparse priors, significantly exceeded the evidence for conventional minimum norm models (mean difference±SEM GS: 85.4±7.8, ARD: 47.0±7.4, gGS: 56.0±6.4,
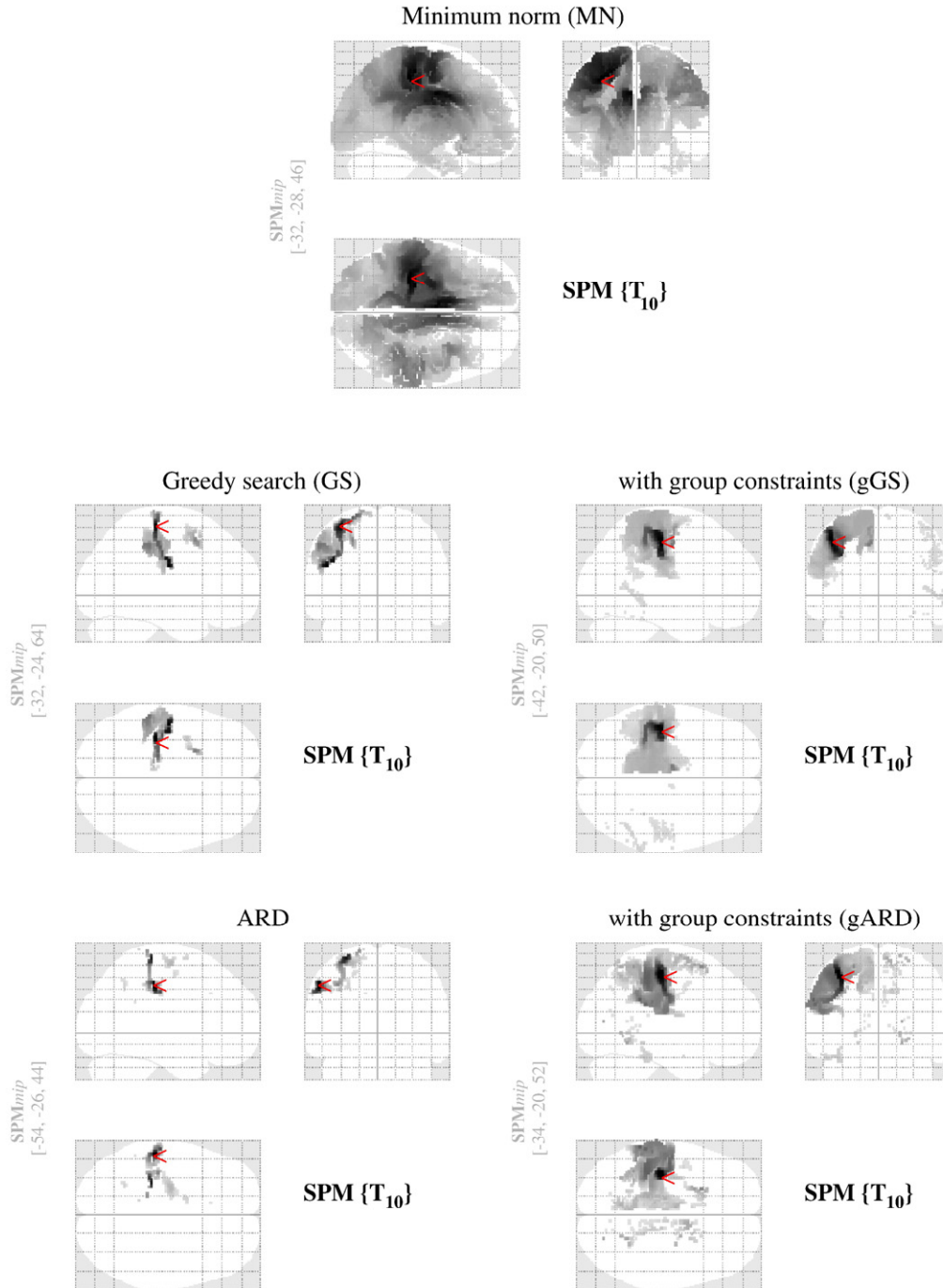

Fig. 3. Results of paired *t*-tests comparing the reconstructed images with the midsagittal-flipped version of the same images. All models show maximal activation around the left sensorimotor area. However they differ greatly with respect to the number of suprathreshold voxels. These SPMs have been thresholded at the same level ($p < 0.001$; uncorrected) and the location of the maximum *t*-value is marked by a red pointer in each SPM.

gARD: 64.2±6.8; Fig. 2A). An analysis of variance of the log-evidences from the experiment-specific inversions showed a significant cross-over interaction between the model and the effect of group constraints ($F_{(1, 54)}$=258.78, $p<0.001$). For GS F was significantly higher without group constraints ($p<0.001$) whereas for ARD, F was higher with group constraints ($p<0.001$) (Fig. 2B). On average, the model with the largest evidence was returned by the greedy search without group constraints. For the GS models, this is exactly what we had expected, in that the group constraints compromise the optimisation of the parameters and hyperparameters, in relation to any one dataset. Furthermore, in the absence of group constraints, the superiority of the greedy search over the ARD was not surprising, in that the ARD model is very over parameterised and can easily get stuck in local maxima (see Wipf and Nagarajan, in press for an analysis of convexity of these sorts of objective functions).

What was surprising is the fact that group constraints *improved the evidence for ARD models*. Strictly speaking, this is impossible because any empirical source priors, furnished by group constraints, should have been discovered during the ARD. The only explanation is that ARD finds local maxima in its very high-dimensional search space and that these local maxima are suppressed by group constraints. In other words, the group constraints remove local maxima and allow the ARD scheme to find better solutions. This is encouraging because it means the group constraints must be partly veridical; if they are not, there would have been no improvement in the evidence for ARD models.

Finally, the interaction between model and group constraints makes sense from this perspective; in that ARD benefits much more from constraints than the GS, because it has a much higher dimensional search space to navigate. We were slightly surprised that ARD outperformed GS with group constraints, given our anecdotal experience with these schemes. However, it may be the case that the focal responses evoked by median nerve stimulation do not require the models of distributed responses afforded by the partitions (sets of ARD priors) optimised during the greedy search.

Model comparison at the between-subject level

Five one-sample *t*-test SPMs testing for a sensory evoked response from one of the randomly selected groups are shown in Fig. 3. These illustrate the similarities and differences across the five reconstruction schemes we evaluated. Fig. 4 shows the cumulative distributions of maximal *t*-values (4A) and the distribution of distance of the voxel with the maximal t-value from the mean location for the same reconstruction scheme (4B).

Interestingly, on average, the maximum *t*-values for the ARD and GS schemes were lower than for minimum norm, but for GS with group constraints the difference was not significant (GS: $p<0.001$, ARD: $p<0.001$, GS group: $p=0.07$, ARD group: $p=0.001$). Repeated measures two-way ANOVA[6]

---

[6] Strictly speaking the *t*-value summary statistics cannot be treated as independent summaries, over the 100 realisations of the group analyses, because these realisations borrowed data from the same five experiments in each subject. However, we appeal to the sphericity of mildly correlated but uniform correlations, among the ensuing errors over groups, to eschew a correction for non-sphericity. Heuristically, each of the 100 distinct combinations of the five experiments brings a degree of freedom to the inference about effects over groups.
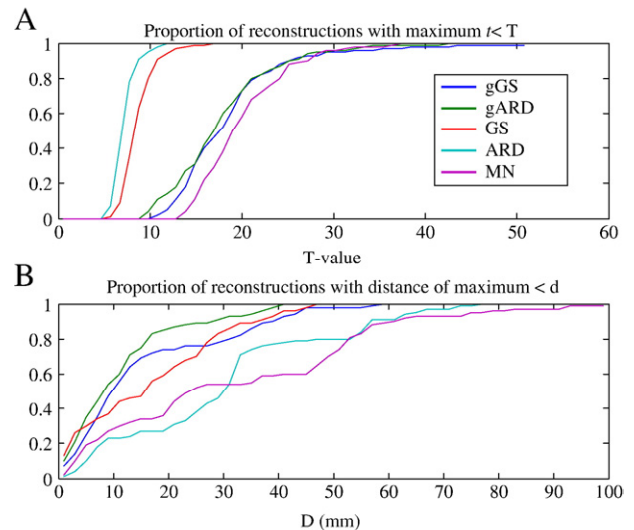

Fig. 4. Cumulative distributions of the 'quality measures' for the *t*-test SPMs, such as shown in Fig. 2, across 100 repetitions using randomly selected groups of 11 experiments with the same 11 subjects. (A) The maximal *t*-value in the image. (B) The distance of the voxel with the maximal *t*-value from the 'mean location' over the 100 group analyses, with the same reconstruction scheme.

showed no interaction between the hyperprior model and the effect of group constraints ($F_{(1, 99)}$=0.52, $p=0.47$). Main effects of both factors were significant (hyperprior type: $F_{(1, 99)}$=7.78, $p=0.006$, group constraints: $F_{(1, 99)}$=512.5, $p<0.001$). Maximal *t*-values were higher for GS, than for ARD and higher with group constraints, than without. The highest *t*-values (except for minimum norm) were attained with GS with group constraints. However, the difference between GS with group constraints and ARD with group constraints was not significant ($p=0.29$).

It might seem paradoxical for the minimum norm solutions to yield the highest *t*-values, given their models had the lowest evidence on a per subject basis. The reason for this is that the MN solutions deploy reconstructed activity in a non-focal fashion, with very little spatial specificity. This is evident when we consider the evaluation of spatial precision: The dispersion of maxima under GS ($p<0.001$), gGS ($p<0.001$) and gARD ($p<0.001$) was much smaller than for minimum norm models. ARD without group constraints was not significantly different from minimum norm, with respect to this measure ($p=0.55$). The cumulative densities in Fig. 4B shows that 50% of the SPMs obtained with GS under group constraints furnished a maximum within about 10 mm of the mean location. This contrasts with the MN solutions, where 50% of maxima were only within about 30 mm of their mean location. ANOVA showed a significant interaction between the model and the effect of group constraints ($F_{(1, 99)}$=42.01, $p<0.001$). For GS there was no significant effect of the group constraints on dispersion ($p=0.6$). For ARD, however, the effect of group constraints was highly significant ($p<0.001$). The best results (lowest distance) were attained using ARD with group constraints. These results were significantly better than for GS with group constraints ($p=0.008$). Again, ARD benefits significantly more from group constraints, relative to GS. In fact, in terms of spatial precision it outperformed GS (but not in terms of sensitivity), provided it has access to group constraints. As above this may reflect the focal nature of sensory evoked potentials in this paradigm.

Discussion

Source reconstructions with multiple sparse priors (Friston et al., 2008a) yield solutions which are much more focal than those provided by classical minimum norm models. Furthermore, they have significantly higher model evidences in relation to single-subject data. However, the regional specificity of these solutions turns into a disadvantage when contrast images are pooled across different experiments. This is because inter-subject variability in regional effects can lead to low *t*-values, when sources do not overlap. One way to solve this problem is through spatial smoothing of reconstructed images. However this compromises spatial resolution and subverts the main advantage of multiple sparse priors. Our present study shows that introduction of group constraints makes it possible to attain *t*-values close to those of minimum norm (Fig. 4A); with solutions that remain focal and consistent across subjects (Fig. 4B).

Group constraints may lead to a reduction of model evidence for individual subject data. In the present study, however, this was observed only for GS. For ARD the evidence was actually higher with group constraints (Fig. 2B), suggesting exploration of hyperparameter or model space was somehow finessed by information from other subjects. On comparing the two methods with group constraints, GS and ARD, we conclude that both yield similar *t*-values (close to those of minimum norm case). The largest *t*-values were provided by GS with group constraints; however, for ARD with group constraints the solutions were, anatomically, more consistent across replications. ARD benefited much more than GS from group constraints and this may reflect the large number of hyperparameters it has to optimise. Without group constraints GS is clearly better according to all the three criteria: model evidence, *t*-values and spatial precision.

Conclusion

We have described a simple device that allows one to place constraints on reconstructed source activity in the context of group studies. This is not a full hierarchical inversion but finesses single-subject source reconstruction; as a prelude to summarizing subject-specific responses for classical inference at the group level. Our analyses suggest that this between-subject constraint markedly increases the reliability of detecting systematic responses over subjects, in terms of their functional anatomy. The inversion procedure is based upon generic variational techniques and yet proceeds quickly and efficiently, using only second-order statistics in channel space. This means the entire inversion for several subjects takes seconds (as opposed to minutes).

Acknowledgments

Software note

The algorithm and schemes introduced in this paper are available as academic freeware as part of the SPM software package (http://www.fil.ion.ucl.ac.uk/spm). The group inversion scheme can be accessed through the 3D-source reconstruction by selecting "group inversion". Forward models will then be built automatically and the group inversion described in this paper implemented. All the user has to supply is the names of the data files and the locations of the sensors and fiducials.

References

Allison, T., McCarthy, G., Wood, C.C., Jones, S.J., 1991. Potentials evoked in human and monkey cerebral cortex by stimulation of the median nerve. A review of scalp and intracranial recordings. Brain 114, 2465–2503.

Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans. Biomed. Eng. 44 (5), 374–385.

Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., Kiebel, S., 2005. Mixed-effects and fMRI studies. NeuroImage 24 (1), 244–252 (Jan 1).

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

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. NeuroImage 34 (1), 220–234.

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008a. Multiple sparse priors for the M/EEG inverse problem. NeuroImage 39 (3), 1104–1120 (Feb 1).

Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., Ashburner, J., 2008b. Bayesian decoding of brain images. NeuroImage 39 (1), 181–205 (Jan 1).

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72, 320–338.

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 (10), 2395–2414.

Kass, R.E., Steffey, D., 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). J. Am. Stat. Assoc. 407, 717–726.

Kiebel, S.J., Friston, K.J., 2004. Statistical parametric mapping for event-related potentials (II): a hierarchical temporal model. NeuroImage 22 (2), 503–520.

Litvak, V., Zeller, D., Oostenveld, R., Maris, E., Cohen, A., Schramm, A., Gentner, R., Zaaroor, M., Pratt, H., Classen, J., 2007. LTP-like changes induced by paired associative stimulation of the primary somatosensory cortex in humans: source analysis and associated changes in behaviour. Eur. J. Neurosci. 25, 2862–2874.

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.

Mattout, J., Henson, R.N., Friston, K.J., 2007. Canonical source reconstruction for MEG. Comput. Intell. Neurosci. 67613.

Neal, R.M., 1998. Assessing relevance determination methods using DELVE. In Neural Networks and Machine Learning. Springer, pp. 97–129.

Nummenmaa, A., Auranen, T., Hamalainen, M.S., Jääskeläinen, 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 (2), 669–685.

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

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

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 (3), 806–826.

Serinagaoglu, Y., Brooks, D.H., MacLeod, R.S., 2005. Bayesian solutions and performance analysis in bioelectric inverse problems. IEEE Trans. Biomed. Eng. 52 (6), 1009–1020.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System — An Approach to Cerebral Imaging. Thieme Medical Publishers, New York, NY.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.

Trujillo-Barreto, N., Aubert-Vazquez, E., Valdes-Sosa, P., 2004. Bayesian model averaging. NeuroImage 21, 1300–1319.

Wipf, D. Nagarajan, S., in press. A Unified Bayesian Framework for MEG/EEG Source Imaging. NeuroImage.