Nummenmaa, A., Auranen, T., Vanni, S., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Vehtari, A., and Sams, M. (2007). Sparse MEG inverse solutions via hierarchical Bayesian modeling: evaluation with a parallel fMRI study. Helsinki University of Technology, Laboratory of Computational Engineering Publications, Report B65, ISBN 978-951-22-9141-0.

Helsinki University of Technology Laboratory of Computational Engineering Publications
Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja
Espoo 2007 Report B65

SPARSE MEG INVERSE SOLUTIONS VIA HIERARCHICAL BAYESIAN MODELING: EVALUATION WITH A PARALLEL fMRI STUDY

Aapo Nummenmaa Toni Auranen Simo Vanni Matti S. Hämäläinen
Iiro P. Jääskeläinen Jouko Lampinen Aki Vehtari Mikko Sams

Kaija Virolainen, PhD (Editor)

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Laboratory of Computational Engineering Publications
Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja
Espoo 2007 Report B65

SPARSE MEG INVERSE SOLUTIONS VIA HIERARCHICAL BAYESIAN MODELING: EVALUATION WITH A PARALLEL fMRI STUDY

Aapo Nummenmaa Toni Auranen Simo Vanni Matti S. Hämäläinen
Iiro P. Jääskeläinen Jouko Lampinen Aki Vehtari Mikko Sams

Kaija Virolainen, PhD (Editor)

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Computational Engineering

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Laskennallisen tekniikan laboratorio

Distribution:
Helsinki University of Technology
Laboratory of Computational Engineering
http://www.lce.hut.fi
P.O. Box 9203
FI-02015 TKK
Tel. +358-9-451 4826
Fax. +358-9-451 4830
http://www.lce.hut.fi

E-mail: Aapo.Nummenmaa@hut.fi

ISBN 978-951-22-9141-0
ISSN 1455-0474

PicaSet Oy
Helsinki 2007

Sparse MEG inverse solutions via hierarchical Bayesian modeling: evaluation with a parallel fMRI study

Aapo Nummenmaa,[1,2*] Toni Auranen,[1,2] Simo Vanni,[2,3] Matti S. Hämäläinen,[4,5] Iiro P. Jääskeläinen,[1,2] Jouko Lampinen,[1], Aki Vehtari[1], and Mikko Sams[1,2,3]

[1]*Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland*

[2]*Advanced Magnetic Imaging Centre, Helsinki University of Technology, Espoo, Finland*

[3]*Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, Espoo, Finland*

[4]*Massachusetts General Hospital – Massachusetts Institute of Technology – Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02139, USA*

[5]*Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

6th September 2007

Technical Report B65
Helsinki University of Technology, Laboratory of Computational Engineering

* Corresponding author. Fax: +358 9 451 4830 E-mail address: Aapo.Nummenmaa@hut.fi (A. Nummenmaa).

Keywords: fMRI, MEG, Inverse problem, Hierarchical Bayesian modeling, Sparse solutions

Abstract: Here we demonstrate how sparse cortically constrained MEG inverse solutions can be obtained via the hierarchical Bayesian minimum-norm estimation approach, wherein Gaussian priors with individual precisions (inverse variances) are first assumed for the distributed current amplitudes at each cortical location. A common second-level prior (hyperprior) of specific form is then imposed on the prior precisions to cause most of the current amplitudes to vanish, thus resulting in sparse inverse solutions. We validated the approach using an empirical dataset, wherein identical visual stimulation experiments were carried out in MEG and fMRI. The physiologically highly feasible fMRI statistical parametric maps were used as a reference for the MEG solutions. The results show that the hierarchical Bayesian approach is capable of producing solutions concordant with the fMRI data in a rather automated fashion, despite the characteristic complexity of the visual evoked magnetic fields. The proposed method of selecting the hyperprior to obtain sparsity provides an effective and straightforward way to regularise the solutions. We also discuss the possibility of utilising fMRI information in the MEG source reconstruction.

Introduction

During the past decades many different solution strategies have been suggested to the inverse problem in magnetoencephalography (MEG) and electroencephalography (EEG) (for a review of MEG/EEG and most common inverse algorithms, see, *e.g.*, Baillet et al., 2001). The multidipole approach aims at explaining the observed MEG fields by a small number of equivalent current dipoles, whereas an estimate of the currents throughout the brain is obtained in distributed source modeling. Both approaches have their virtues and limitations; in the multidipole modeling, the (usually unknown) number of dipoles plays an important role and the performance of the algorithms usually deteriorate quickly with increasing number of dipoles. The distributed models face the underdetermined problem of estimating thousands of unknowns based on a few hundred measurements, necessitating some prior assumptions on the distribution of the currents, which then significantly constrain the solutions typically with little neurophysiological justification. Statistical formulation of the inverse problem and Bayesian data analysis (Kaipio and Somersalo, 2005; Bernardo and Smith, 2000; Gelman et al., 2003) have gained increasing popularity within both multidipole (Schmidt et al., 1999; Bertrand et al., 2001a,b; Jun et al., 2005, 2006; Auranen et al., 2007) and distributed inverse analysis (Phillips et al., 2002; Trujillo-Barreto et al., 2004; Phillips et al., 2005; Auranen et al., 2005; Mattout et al., 2006).

The classical Minimum-Norm Estimate (MNE) (Hämäläinen and Ilmoniemi, 1984; Dale and Sereno, 1993) can be interpreted to arise from a Bayesian model, where the *a priori* probability distribution of the current amplitudes at each location of the discretised cortex is assumed to be a Gaussian with zero mean and fixed variance. The prior variance parameter can be used to regularise the solutions, smaller variance setting the currents closer to zero. The virtue of this approach is that with fixed prior variance, the maximum *a posteriori* (MAP) estimate is unique, and a simple inverse operator matrix can be constructed to calculate the solutions. Because the symmetric spatially homogeneous prior constrains the currents to be of similar magnitude throughout the brain, the resulting MNE's are very diffuse. Consequently, such solutions as obtained with multidipole modeling, where few locations contain large currents while others are set to zero, can not arise from classical MNE model. Because a vast number of current configurations, including diffuse and highly dipolar, can produce an identical set of MEG measurements, they can all be justified and correspond to different prior assumptions. Which one is the "true" can not be deduced from the MEG data alone.

An interesting hierarchical Bayesian framework was presented by Sato et al. (2004) for distributed source estimation, which generalises the basic MNE model so that Gaussians with individual precision (inverse variance) parameters are assumed for each point of the solution space. The hierarchy then emerges from imposing a common second-level prior, here called hyperprior, on the prior precisions. The prior precisions are often called hyperparameters as they are parameters of the prior – in a sense, the current amplitudes can be thought of being the first level parameters and their prior precisions the second level. The hierarchical approach has the virtue that now some of the prior precisions can have small values (large prior variance), leading to large current amplitudes for those source locations, while the others keep their respective currents close to zero and considerably less diffuse solutions than the MNE emerge.

The hierarchical prior has the desirable feature of being more flexible, but the joint posterior of the parameters and hyperparameters is not of tractable form. However, the currents and their prior precisions can be estimated from MEG data by using a Variational Bayesian (VB) method (for an introduction to VB-methods and graphical models, see, *e.g.*, Ghahramani and Beal, 2001). The VB-method is closely related to the Expectation Maximisation (EM) algorithm and has also interesting information-theoretic interpretations (Honkela and Valpola, 2004). Close relatives of the hierarchical prior model studied here have been used to perform Automatic Relevance Determination (ARD) for artificial neural network input selection (Neal,

1996) and sparse Bayesian learning (SBL) (Tipping, 2001).

The VB-estimated values of the currents depend directly on their corresponding estimated prior precisions, as is apparent from the VB-algorithm (summarised in the Materials and methods section). Furthermore, our previous results (Nummenmaa et al., 2007a,b) show that the estimates of the prior precisions depend in turn strongly on their prior distribution, the hyperprior (assumed to be a Gamma-distribution). For instance, the ARD-approach of Tipping (2001) to Bayesian regression involves many of the prior precisions being driven to infinity using a noninformative hyperprior, thus eliminating these variables from the model and leading to a sparse representation of the data. The sparsity assumption follows, loosely speaking, the logic of Occam's razor preferring a model which has the smallest (effective) number of parameters needed to explain the observed data. This is also our motivation here – if few locations with large currents can explain the observed fields, we prefer this *a priori* to the more diffuse MNE.

With the severely underdetermined MEG inverse problem, the ARD-prior in fact does not automatically lead to sparse solutions even when the noninformative Jeffrey's hyperprior is used (this improper hyperprior has the additional disadvantage of leading also to an improper posterior, Nummenmaa et al., 2007a,b). Here we provide a detailed analysis of how to obtain sparse solutions to the MEG inverse problem within the ARD-framework, based on the VB-EM algorithm with a suitable choice of the hyperprior. The flexibility of the hierarchical framework is utilised in forcing most of the currents to extremely small values, while still enabling some of the currents to have large values. The sparsity assumption also provides a clearcut resolution to the thresholding problem of the solutions raised in (Nummenmaa et al., 2007b), as the locations of the large currents can be robustly separated from the (effectively) vanishing ones.

Here, to obtain a realistic view on the properties of the sparse solutions, we used an empirical dataset for evaluation of the method. Visual evoked MEG responses provide a good test bench for an automatic inverse algorithm, as the responses are generated by several physiologically a priori relatively well known sources with adjacent locations and temporally overlapping activation patterns, which gives rise to lots of manual work when the inverse modeling is done using traditional dipole-fitting methods (Vanni et al., 2004b,a). Three subjects participated in identical experiments carried out in MEG and fMRI, with a drifting grating visual stimulus. The fMRI data provide a qualitative reference for the locations of the MEG sources. We also briefly discuss the possibility of utilising the fMRI information in the hierarchical MEG inverse estimation.

Materials and methods

The dataset

The same dataset has been previously analysed by Auranen et al. (2007) with a multidipole approach. Here we present only a synopsis of the experimental design and the data analysis.

Three male subjects with normal or corrected-to-normal vision participated in the study. The study was approved by the Ethics committee of the Hospital District of Helsinki and Uusimaa. A local drifting grating activated retinotopic visual areas, and visual motion sensitive area V5. To confirm the functional identification of the active areas, the borders of retinotopic areas were mapped with multifocal fMRI (Vanni et al., 2005). The stimuli are shown in Fig. 1 and a multifocal mapping based retinotopy verification of drifting grating activation is demonstrated in Fig. 2.

The grating was located in the middle of the lower left quadrant of the visual field with 2-7 degree eccentricity covering a 60 degree polar sector (spatial frequency of 1.3 cycles/degree at 4.5 degree eccentricity). Movement of one cycle was presented as four consecutive images. Due to different refreshment rates of data projectors at MEG and fMRI, the respective stimulus durations were 134 ms and 107 ms, corresponding drifting speeds being 7.46 and 9.35 cycles/second. Stimuli were presented with

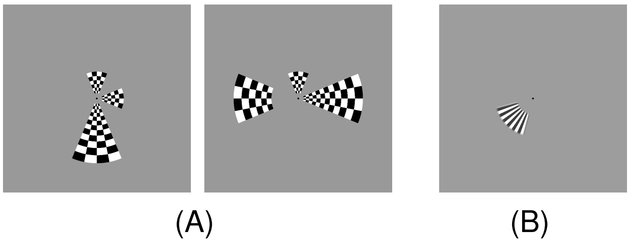

Figure 1. (A): Example frames of the multifocal design. Contrast reversing checkerboard pattern activated the vertical and horizontal meridians, mapping the borders of the low-level retinotopic areas. (B): Drifting grating stimulus for the MEG and fMRI experiment.

a varying ISI of 0.8-1.2 seconds in a block design, alternating with rest in period of 50 seconds. Task was to fixate at the (marked) center of the images and to passively view the stimuli.

We aimed to good signal-to-noise -ratio in both fMRI and MEG to minimize any ambiguity in signal location. A total of 308 functional volumes were collected in three separate runs, one volume consisting of 27 slices with slice thickness 3.0 mm, the stack aligned perpendicular to the parieto-occipital sulcus. Parameters of the EPI sequence were TR=2000 ms, TE=30 ms, flip angle=60 deg, FOV=19 cm, matrix size 64 $\times$ 64. The data of the identical MEG experiment was recorded at 600 Hz sampling frequency, downsampled to 300 Hz, notch-filtered to remove 50-Hz noise, and high-pass filtered (Butterworth, corner frequency 0.2 Hz) to remove slow drifts. Over 800 artefact free trials were used to calculate the average evoked fields.

The general linear model of FSL was used to analyse the fMRI data (Smith et al., 2004). As an exception to the standard analysis setup, neither spatial smoothing nor intensity normalisation was applied in order to project the data to the FreeSurfer-reconstructed (Dale et al., 1999; Fischl et al., 1999) cortical surfaces with maximal resolution. The resulting Z-statistic maps were thresholded by $Z > 5$ and clusterwise significance level of $p = 0.05$. The evoked MEG fields and fMRI statistical parametric maps (SPM) are shown in Figs. 3 and 4.

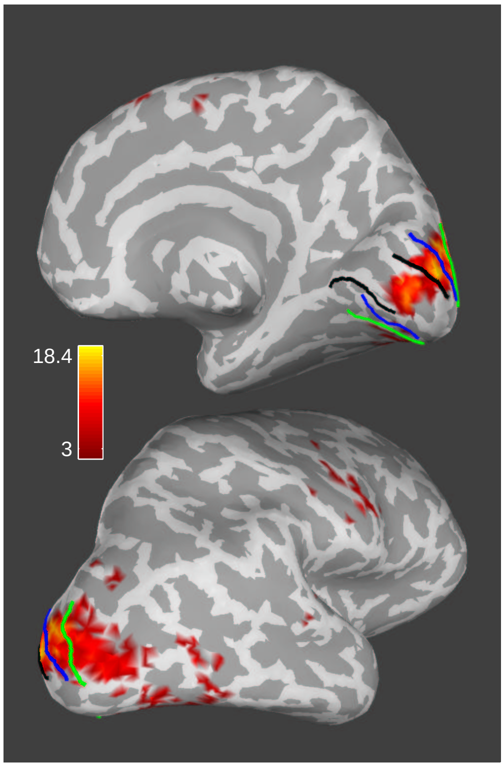

Figure 2. FMRI activation for the drifting grating stimulus for Subject 3. The area borders for V1, V2, and V3/VP were mapped with the multifocal design. The black lines bound V1, and then both in dorsal and ventral directions V2 is bounded by the black and blue and V3/VP by the blue and green lines. The colourbar shows the scale of the thresholded Z-statistic SPM (for details, see text).

Considerable intersubject variability in both MEG and fMRI data is evident. Subject 1 has the best signal to noise ratio (SNR) for the MEG data, and Subject 3 the worst. On the contrary, Subject 3 has the largest Z-scores in the SPM's and Subject 1 the smallest. As the final projection of the voxel fMRI SPM to the cortical surface involves smoothing, leading to diminished Z-scores, the threshold was lowered to value 3 for the surface visualisation.

Forward model

FreeSurfer based model was used to constrain the possible locations and orientations of the sources according to individual cortical geometry. The number of discretisation points in the reduced surface used in the inverse estimation

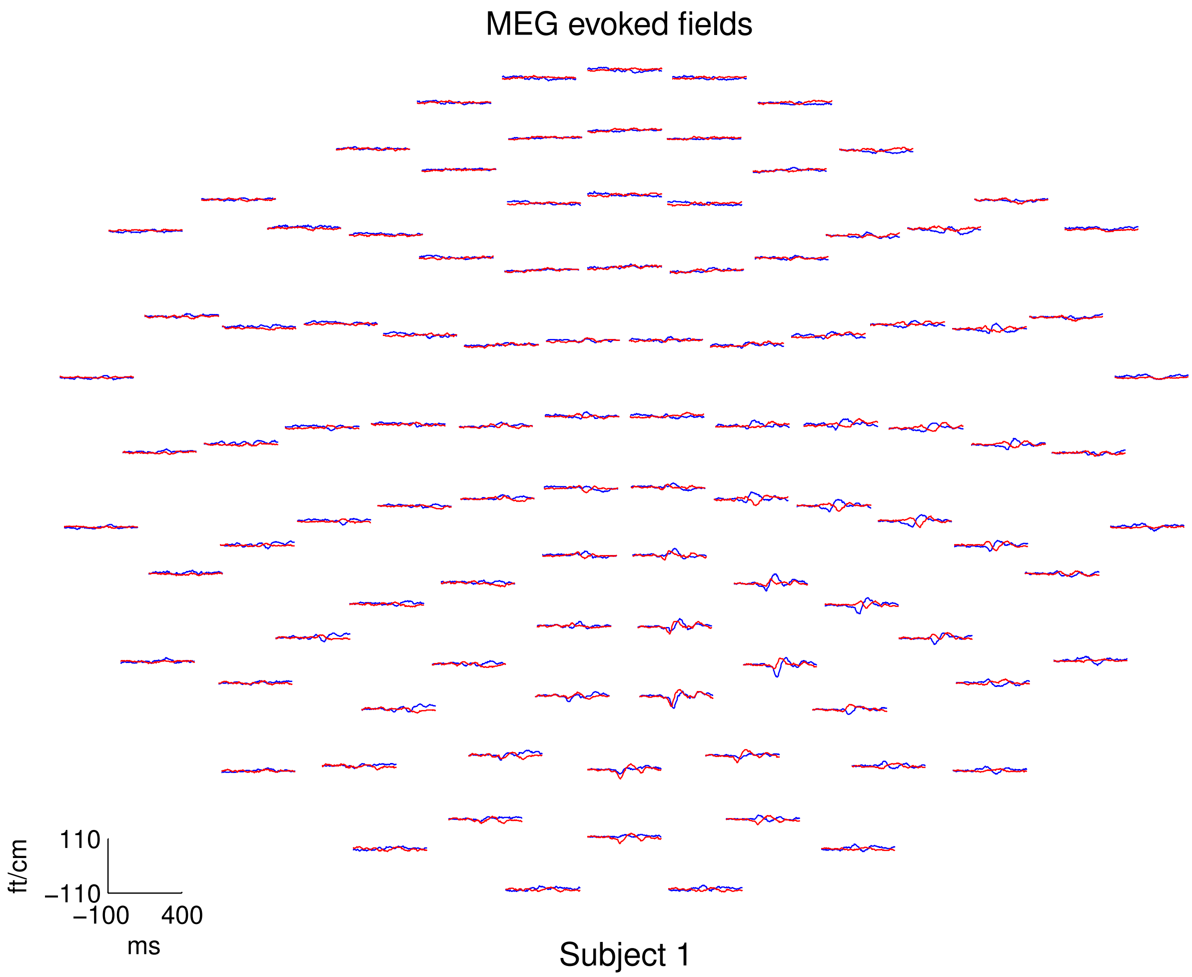

Figure 3. The MEG data of Subject 1 is plotted on top of the sensor grid, nose pointing up. The two planar gradiometers are shown, with red and blue lines.

was approximately 8000. One-layer boundary element method was used in the forward computations (see, *e.g.*, Mosher et al., 1999), resulting in the linear equation relating the current amplitudes $\boldsymbol{J}(t)$ to the observed MEG fields $\boldsymbol{B}(t)$:

$$\boldsymbol{B}(t) = \boldsymbol{G}\boldsymbol{J}(t) + \boldsymbol{N}(t), \quad t = 1, \ldots, T, \quad (1)$$

where $\boldsymbol{G}$ is the gain matrix, T is the number of timepoints (151), and the measurement noise $\boldsymbol{N}(t)$ is assumed to be independent of time and to have a Gaussian distribution with zero mean and inverse covariance $\boldsymbol{\Sigma_G}$:

$$\boldsymbol{N}(t) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{G}}^{-1}). \quad (2)$$

Here we deviate from (Sato et al., 2004) and (Nummenmaa et al., 2007a,b) in assuming that the estimated inverse noise covariance $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{G}}$ is not subject to any uncertainty; hence the previously used scale parameter β quantifying this is set to be equal to unity. This simplifies the analyses and can be justified with the large number of trials available for calculating the average evoked response.

The hierarchical prior

Gaussian prior with individual precision (inverse variance) parameter is imposed on the currents at location i of the source space:

$$\begin{aligned} \boldsymbol{J}(t)_i &\sim \mathrm{N}(0, \alpha_i^{-1}), \\ &\quad \text{or with vector notation} \\ \boldsymbol{J}(t) &\sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{A}^{-1}). \end{aligned} \quad (3)$$

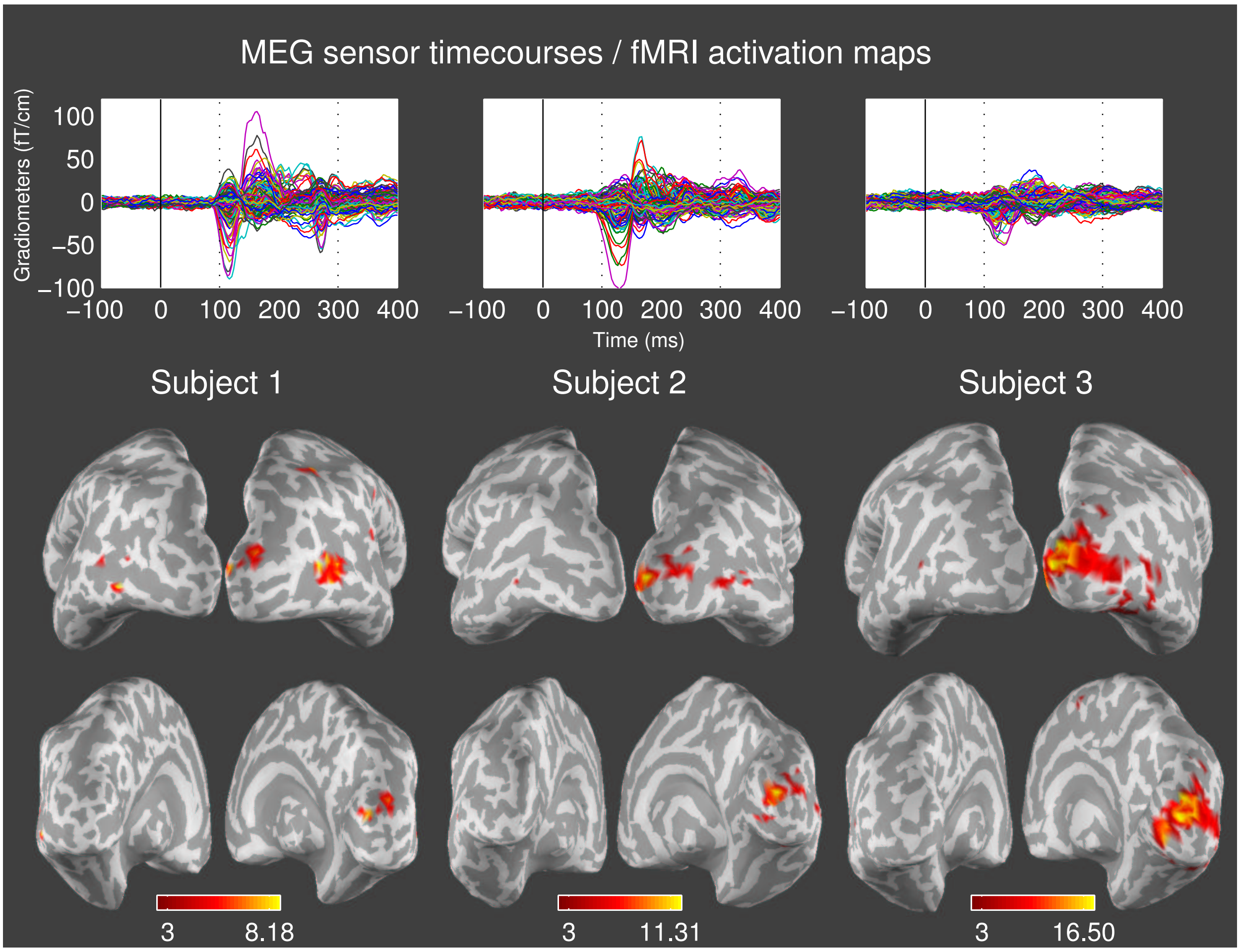

Figure 4. The columns 1-3 show the MEG data and fMRI statistical parametric (Z) maps for the corresponding subjects. Note the different scales in the fMRI activation images. Sensor timecourses of all MEG channels are collapsed on a single plot for each subject to facilitate comparison with those calculated from the inverse solutions.

A common Gamma-distribution hyperprior is then imposed on the hyperparameters α_i

$$\alpha_i \sim \mathrm{Gamma}(\alpha_0, \gamma_0), \tag{4}$$

where the Gamma-distribution is parameterised as

$$\mathrm{Gamma}(\alpha_i|\alpha_0, \gamma_0) = \frac{1}{\alpha_i}\left(\frac{\alpha_i \gamma_0}{\alpha_0}\right)^{\gamma_0} \Gamma(\gamma_0)^{-1} \exp\left(-\frac{\alpha_i \gamma_0}{\alpha_0}\right), \tag{5}$$

and $\Gamma(\cdot)$ being the Euler Gamma function.

Estimating the currents and the hyperparameters

If the hyperparameters α_i were known or fixed to values $\hat{\alpha}_i$, the maximum *a posteriori* (MAP) estimate for the currents would be the linear MNE:

$$\begin{aligned} \hat{\boldsymbol{J}}(t) &= \hat{\boldsymbol{A}}^{-1} \boldsymbol{G}' \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}} \boldsymbol{B}(t), \\ \text{where} \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{-1} &= \boldsymbol{G} \hat{\boldsymbol{A}}^{-1} \boldsymbol{G}' + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{G}}^{-1}. \end{aligned} \tag{6}$$

Typically this is not the case, but the α_i's are unknown and thus preferably estimated from the data as well. The joint posterior of the currents and their precisions is not of tractable form, and hence the VB-method is developed in (Sato et al., 2004) to obtain an approximate factored posterior. Operationally, this leads to an EM-type update for re-estimating the precision parameters α_i based on the previously computed MNE:

$$\hat{\alpha}_i^{new} = \frac{\gamma_0 + \frac{T}{2}}{\frac{\gamma_0}{\alpha_0} + \frac{1}{2}\sum_{t=1}^{T} \hat{\boldsymbol{J}}(t)_i^2 + \frac{T}{2}\left(\hat{\boldsymbol{A}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{A}}^{-1}\boldsymbol{G}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\boldsymbol{G})\right)_{ii}}. \quad (7)$$

The equations (6) and (7) are iterated until convergence. Along with the number of time-points T available for inferring the values of the hyperparameters α_i, the updated estimates depend also on the parameters of their prior, α_0 and γ_0. In practice these must be manually fixed to some reasonable values (Nummenmaa et al., 2007b). The resulting estimates depend strongly on the selection of these parameters – for instance, from equation (7) one can see that by setting $\alpha_0 = 10$ and $\gamma_0 = 10^{10}$, the hyperprior will dominate the hyperparameter estimation process resulting in $\alpha_i \approx \alpha_0$ for all i and a very classical MNE-like solution (Nummenmaa et al., 2007b).

Regularisation, initialisation, sparsity and thresholding

The parameters α_0, γ_0 describe the mean and the degrees of freedom (or shape) of the Gamma-hyperprior, respectively. These both have a regularising effect, the α_0 setting the overall magnitude of the precisions and γ_0 how much the precisions can vary from their mean *a priori*. The variance of the Gamma-distribution is α_0^2/γ_0. Thus, the lower the value of γ_0, the more the precisions can deviate from α_0, and the less "informative" the hyperprior is. Lower values of γ_0 also cause the convergence of the VB-algorithm to slow down. In the limit $\gamma_0 \to 0$, the Gamma-distribution becomes the improper Jeffrey's prior, leading to the posterior becoming also improper and consequently to some theoretical and practical difficulties. The resulting estimates are found to be more sensitive to the specific value of γ_0 (Nummenmaa et al., 2007a). The general impact of the hyperprior, like all priors, is to bias the estimates while reducing variance. Making the hyperprior more informative or strict (choosing large values for γ_0 and/or α_0) puts more weight to the prior over the likelihood. As we are solving an ill-posed, underdetermined problem, introducing some bias is most likely necessary to obtain robust estimates with empirical data.

The estimates also depend on the initialisation of the algorithm (the hyperparameters α_i), and the true joint posterior of the parameters and hyperparameters is thus multimodal (Nummenmaa et al., 2007a). However, let us assume that all of the α_i's are initialised to a common value α_s. At the first iteration of the VB-algorithm, the current estimate is a MNE with regularisation parameter α_s. The larger the (at this stage common) precision parameter α_s is, the more regularised MNE the VB-algorithm begins and also most likely ends with. To obtain estimates with an intermediate regularisation also in this aspect, we initialise the algorithm in the region of the hyperparameter space where the hyperprior is relatively flat. We also study the utility of initialising the algorithm with fMRI data; in this case at the first iteration the current estimate is an fMRI-weighted MNE (for fMRI-weighted MNE, see, *e.g.*, Liu et al., 2002).

Effective sparsity can be realised by setting α_0 to a large value to make the average value of currents very small, and γ_0 to a sufficiently large value to obtain a suitable regularising effect and robust convergence of the algorithm. We speak of "effective" sparsity, since the currents are not set to exactly zero but to very small values.

In the previous study, we raised the question of thresholding the MNE-type inverse estimates. Usually, the thresholding is done for better visualization of the "real" activations and the small current ripples are set to zero. In principle, this causes problems for models such as the basic MNE, where the prior variance is assumed to be uniform throughout the source space, each location consequently explaining roughly equal proportions of the data (Nummenmaa et al., 2007b). The small subthreshold currents may actually be responsible for explaining a larger part of the variability in the data than the actual

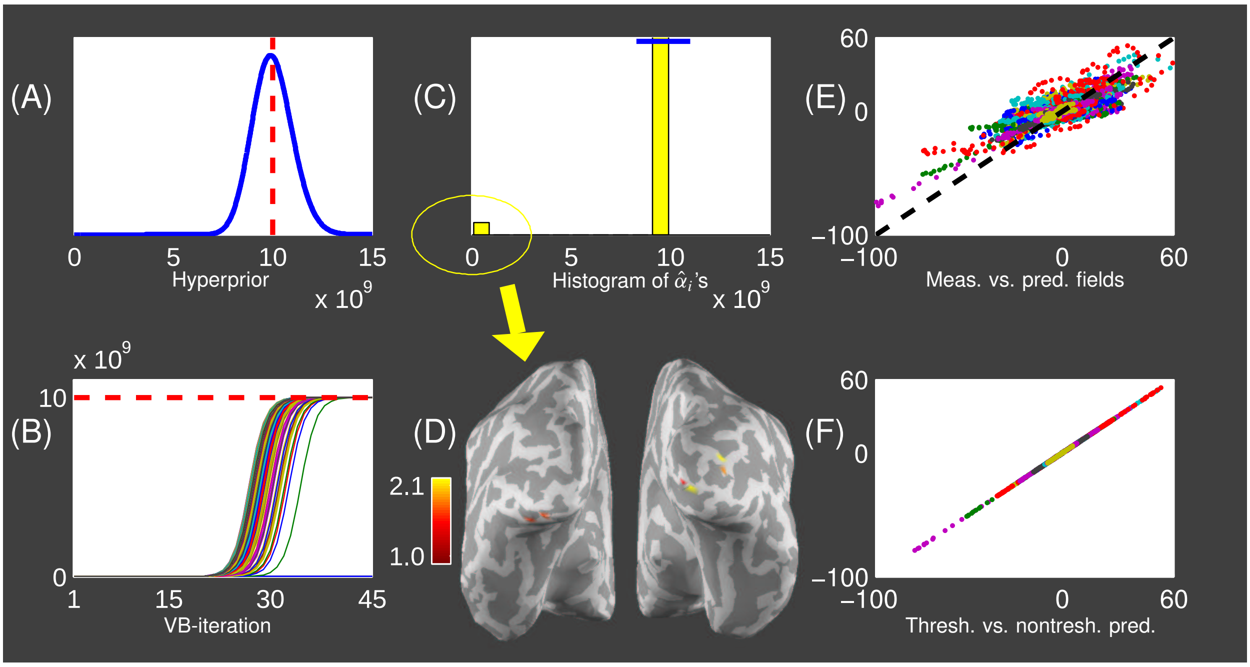

Figure 5. (A): The gamma-hyperprior with parameters $\gamma_0 = 100, \alpha_0 = 10^{10}$. (B): Trends of the prior precisions in the course of the VB-algorithm. (C): Histogram of the estimated prior precisions, showing that most of the precicions end up close to α_0, and hence the solutions are effectively sparse. The blue line indicates that the histogram continues beyond the plotted range. (D): The cluster of the small prior precisions visualised on the cortical surface. The shown values are deviation quantities (inverse square roots of the precisions). (E): The forward computed fields and the empirical data shown as a scatterplot. The black broken line shows the case of "perfect fit". (F): The forward computed fields with only the large currents (thresholded) and with all currents shown as a scatterplot, demonstrating that small currents have a vanishing effect on the predicted fields.

visualised activations. Within the hierarchical framework, if the parameter α_0 is set to a sufficiently large value, the contributions of the small currents to the data fit can be ensured to vanish, thus circumventing the thresholding issue.

Figure 5 illustrates how the sparse solutions emerge from the hierarchical model, by using the data of Subject 2. The hyperprior with parameters $\gamma_0 = 100, \alpha_0 = 10^{10}$ is shown in (A), and it is seen that most of the prior probability mass of the α_i concentrated around α_0. As the VB-algorithm proceeds, most of the hyperprior estimates $\hat{\alpha}_i$ are driven close to α_0. The histogram of the $\hat{\alpha}_i$ is then shown in (C), which illustrates that there are two clearly separated clusters, those with small estimated prior precision (large variance, inside the yellow circle), and those for which the prior precisions are extremely large. The locations with the small prior precisions can be plotted on the inflated cortical surface, which is shown in (D). The forward calculated measurements versus the actual observed MEG fields are then shown in (E) as a scatterplot. To ensure that the small currents do not actually contribute anything to the forward-calculated fields, the calculations were made with and without the contributions of these, the scatterplot of (F) showing the high degree of similarity of the two cases.

Incorporating fMRI information

There are several ways in which fMRI statistical maps can be incorporated to the hierarchical inverse estimation. We could, for instance, choose different hyperprior parameters α_{0f}, γ_{0f} to locations which are likely activated based on fMRI information, resulting in a rather "soft"

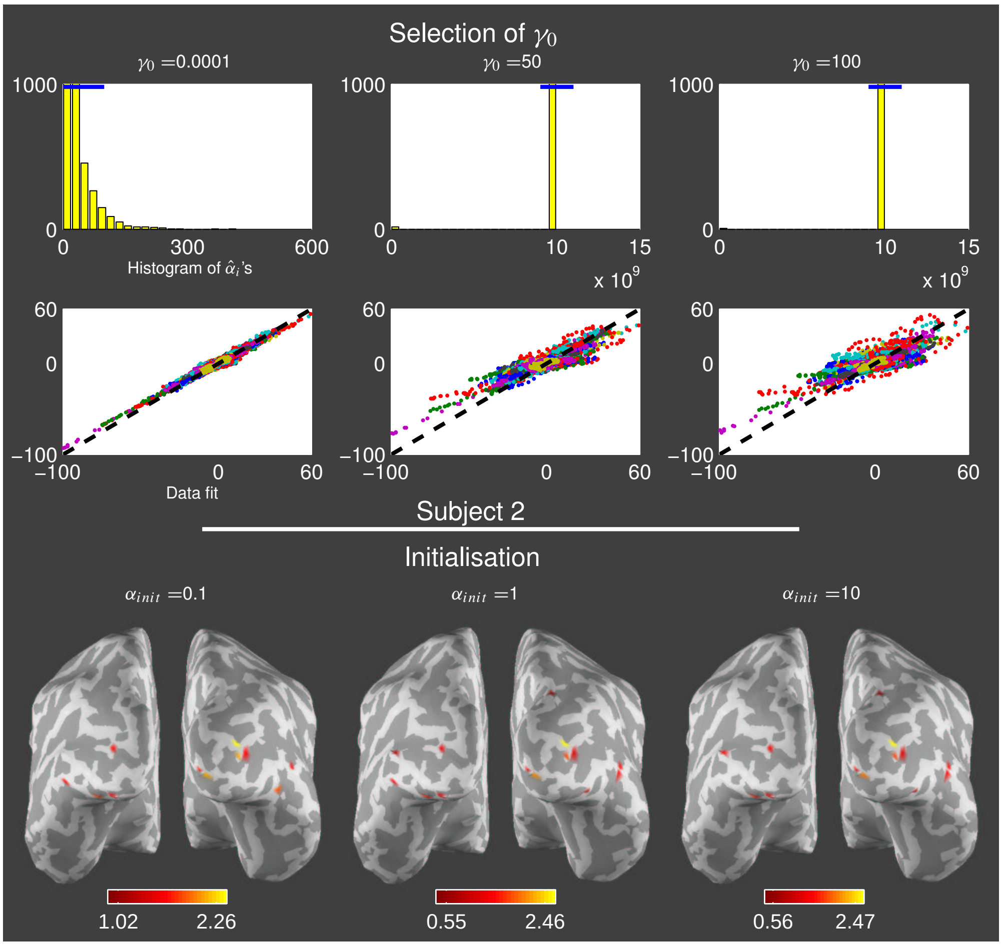

Figure 6. The two upmost rows demonstrate the emergence of sparse solutions and the regularising effect of γ_0, respectively. Below, the solutions (estimated prior deviations) corresponding to different initialisations of the VB-algorithm are depicted. Again, the blue line indicates that the histogram continues beyond the plotted range.

constraint. Or, we could directly scale the α_i's according to some fMRI weights. It turns out, that the above procedures are easily either too soft or too hard, producing little effect or leading to very biased estimates completely determined by the fMRI prior. This behaviour is not surprising, as the central idea of the method is that the prior variances are estimated from the MEG data. Since the prior precision parameters are assumed not to change in time, there are T datapoints (number of timepoints in the sensor data) from which these are estimated. As more data means less weight for the prior (see, Eq. (7)), the fMRI component of the prior would have to be rather sharp. Since the sparsity assumption is already quite strong, the priors can consequently also be conflicting. The initialisation, on the other hand, has clearly an effect to which the solutions are not overly sensitive, and here we study the utility of incorporating the fMRI information into this aspect. As stated before, initialising with spatially homogeneous prior variance corresponds to the first iteration being an MNE, initialising with fMRI leads to an fMRI-weighted MNE at the first VB-step.

Results

Regularisation: Selection of γ_0 and initialisation

For studying the regularising effect of the hyperparameter γ_0 we set $\alpha_0 = 10^{10}$ and ran the VB-

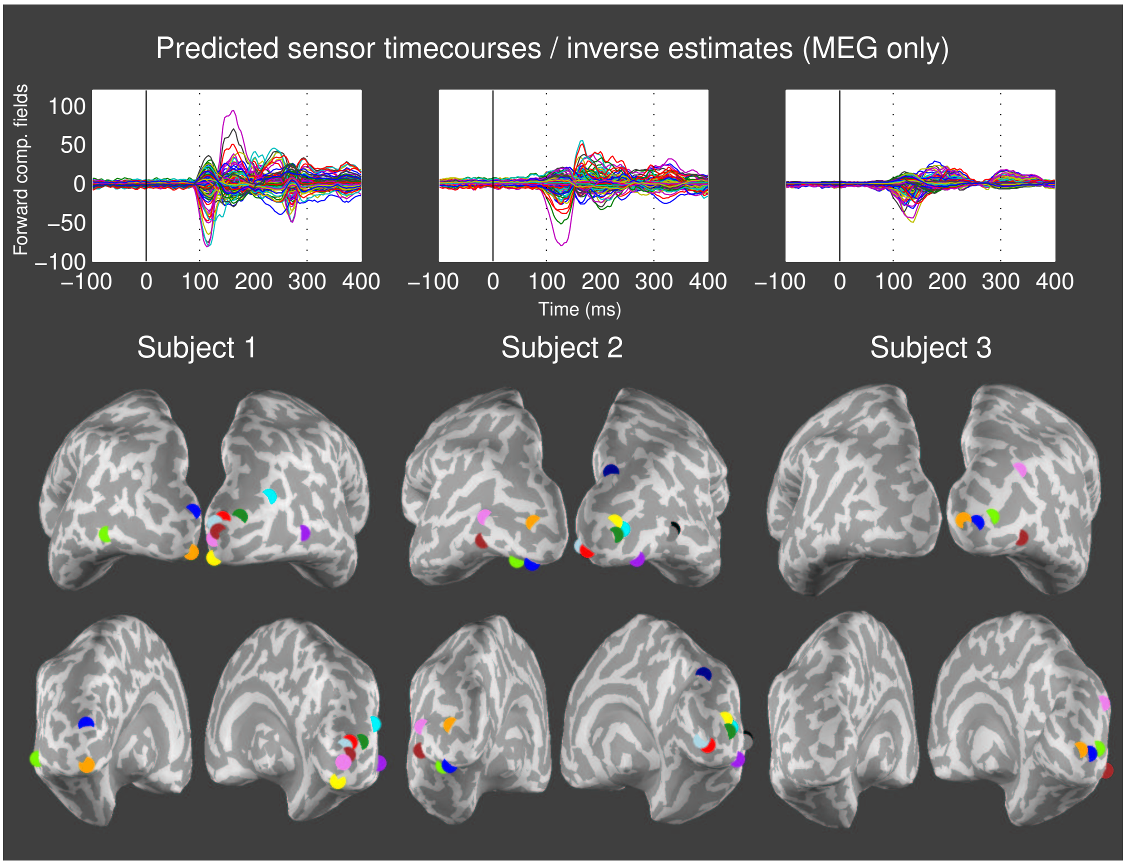

Figure 7. The columns 1-3 show the forward computed MEG sensor timecourses and the cortical locations containing large currents for the corresponding subjects.

algorithm with $\gamma_0 = 0.0001, 50$ and 100, with initialisation $\hat{\alpha}_i = 1$. The value $\gamma_0 = 50$ was chosen for investigating the estimates obtained by different initialisations, which were chosen to be $\hat{\alpha}_i = 0.1$, 1 and 10. The data of Subject 2 were used and the results are shown in Figure 6.

In the case of very low value of γ_0 and noninformative hyperprior, the data fit corresponding to the estimate is very good, indicating little bias. From the histogram of estimated $\hat{\alpha}_i$'s we see that the estimate is not sparse and no obvious criterion/scale for thresholding or sparsifying the estimate is evident. The solutions corresponding to very small γ_0 may also visually look "overfitted" (Nummenmaa et al., 2007b). With increasing γ_0 the data fit diminishes, as the hyperprior constrains most of the currents close to zero, which can also be seen from the corresponding histogram. Providing a sufficient degree of regularisation and quick convergence of the VB-algorithm, the value of $\gamma_0 = 50$ is used for the rest of the analyses and for all subjects. The lower part of Figure 6 shows the estimated prior deviations ($\hat{\sigma}_i = \hat{\alpha}_i^{-1/2}$) obtained with different initialisations. It is apparent that (with the rather informative hyperprior) the algorithm is not too sensitive for the initialisation, and we select the intermediate value of 1 for this purpose in what follows.

Localisation with MEG only

The hierarchical VB-estimation was done with hyperprior parameters $\alpha_0 = 10^{10}, \gamma_0 = 50$ and initialisation $\hat{\alpha}_{i,init} = 1$ for all subjects, based on the analysis of previous section. The localisation results and the forward computed fields are shown in Figure 7.

The number of locations which were estimated to contain large currents were 11, 14, and 5 for Subjects 1,2, and 3, respectively. This is roughly in line with the different SNR's of the datasets. Estimates for Subjects 1 and 2, whose data have significantly higher SNR's, contain also significantly more sources, as all estimates are calculated with same degrees of regularisation. The forward computed fields are in satisfactory agreement with the actual measured fields (see, Figure 4). For all subjects, the overall amplitude is somewhat smaller, and the sensor timecourses are altered in places, due to the regularisation. The locations of the estimated MEG sources show close resemblance to the fMRI activation maps. For Subject 1, the cluster of sources close to the primary visual areas V1-V3, seems to be slightly too posterior. For Subject 2, there is a "unoccupied" space between clusters containing supposedly the sources of the primary visual areas, caused by some of the sources being perhaps localised to other bank of a more appropriate sulcus. For Subject 3, the cluster of the primary sources is a bit too superficial in comparison with the fMRI activations. For all subjects, the activation of motion sensitive area (V5) of the right hemisphere seems to be quite accurately localised (for Subject 2 there are actually two nearby sources). On the left hemisphere, no sources are estimated for Subject 3 most likely due to the poor SNR, while for Subjects 1 and 2 the locations are reasonable in comparison with each other and with fMRI data, which shows contralateral activity only at the level of V5. All subjects also have an MEG source in the posterior parietal cortex not clearly visible in the fMRI, indicating the possible activation of also higher levels of the dorsal visual stream.

The timecourses of the sources are shown in Figure 8. As expected from properties of the basic MNE, largest current amplitudes are estimated for sources which are most superficial. The earliest activation onset latencies do not show any clear visual processing hierarchy (such as V1 $\rightarrow$ V2 $\rightarrow$ V3) which is observed, *e.g.*, by Vanni et al. (2004b,a). Also, in the present dataset, the first onset latencies which can be safely distinguished from noise are slightly before 100 ms, whereas the earliest onset latencies of V1 activations seen in (Vanni et al., 2004a) are already at about 50 ms. These differences could be explained by the different stimuli used in these experiments (the drifting grating versus checkerboard reversals). In a recent EEG study by Delon-Martin et al. (2006) activity patterns elicited by different types of visual motion stimuli were compared, and their earliest onset latencies and ERP waveforms resemble those of the present MEG data. This is especially true for the rotation stimulus, being closest in nature to the drifting grating.

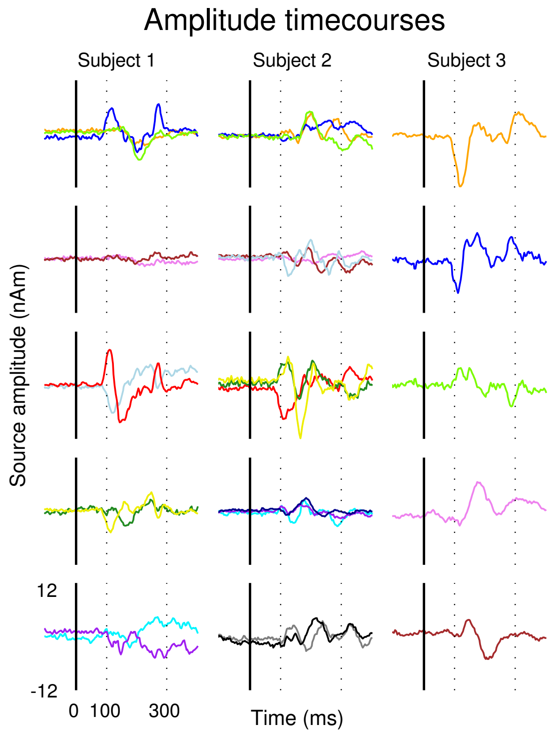

Figure 8. The columns 1-3 show the timecourses of the sources for the corresponding subjects. Colours of the plots refer to the source locations in the Fig. 7.

All in all, the locations and the timecourses of the sources seem to be plausible given the type of stimuli used. Many peculiarities and intersubject differences also emerge. For example, the two nearby V5 sources of Subject 2 have rather similar waveforms, the more anterior source being slightly ahead. Whether these are really two different sources or a modeling artefact remains an open question because of the limited number of subjects.

Initialising with fMRI

We examine the possibility of utilising the fMRI information in the source localisation. The $\hat{\alpha}_i$'s of the "fMRI-active" locations above the threshold $Z > 3$ are assumed to have value 0.25. The $\hat{\alpha}_i$'s of the remaining locations are set to 1 (recall that smaller prior precisions correspond to larger estimated currents). Same hyperprior parameter values $\alpha_0 = 10^{10}, \gamma_0 = 50$ are assumed for the VB-estimation. The initial fMRI-derived prior precisions and the resulting inverse estimates are shown in Figure 9.

Comparing to the estimates obtained with the MEG data alone (Figure 7), the initialisation has some influence on the source locations. Most notably, the posterior parietal source is missing for Subject 2, as there is no corresponding activation in the fMRI (a potential "fMRI false negative"). On the other hand, the "fMRI false positives" surviving the thresholding for Subject 1 have no drastic effects on the solutions. Overall, there are minor differences in the locations and number of sources, but the general characteristics of the solutions remain the same. This is to be expected, since the estimation of the currents and their prior precisions is done based on the MEG data, and the sparsity assumption and implementation with the hyperprior is rather strong as such.

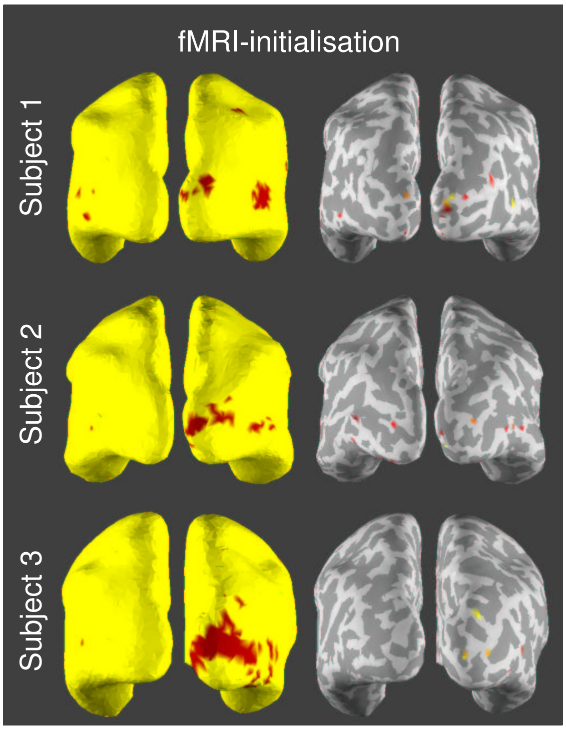

Figure 9. The left column shows the fMRI-based initialisations for each subject. The right column shows the resulting inverse estimates.

Discussion

We proposed an algorithm to produce effectively sparse solutions to the distributed MEG source reconstruction problem, by using the hierarchical VB-method introduced by Sato et al. (2004) and subsequently further analysed by Nummenmaa et al. (2007a,b). Experiments with identical visual stimuli were carried out in MEG and fMRI, the activity maps of which were used for validation of the inverse method. The locations of the estimated sources showed considerable agreement with the fMRI data. We also studied the possibility of using the fMRI information in the inverse estimation.

Our method bears similarities to the FOCUSS algorithm (Gorodnitsky et al., 1995) which is also based on recursive minimum-norm estimation. In our approach, both the sparsity (or focality) and the regularisation come through the Bayesian hierarchical model hyperprior structure. In a recent article, Wipf et al. (2007) present an interesting theoretical analysis of ARD-based models and hyperparameter estimation for the MEG inversion. A speed-up strategy relative to that of Tipping (2001) is also presented for estimating the prior precisions. It will be interesting to see how these different estimation methods behave and perform in empirical data analysis. One virtue of our approach is also the guaranteed convergence, as the basic VB-EM estimation is used, in contrast to the heuristic speed-ups. From a practical point of view, the hyperprior selection (and the hyperparameter γ_0) regularises the estimates quite explicitly, simply and effectively. The estimates are not too sensitive to the initialisation of the algorithm and the thresholding issue is conveniently circumvented by the sparsity assumption.

There are also limitations. Firstly, the true joint posterior distribution of the current amplitudes and their prior precisions is multimodal meaning that different initialisations of the VB-algorithm may lead to different inverse solutions. The presently *ad hoc* selection of the hyperprior parameters may have a significant effect on the posterior mass proportions, which can be in principle estimated with the VB-method (Nummenmaa et al., 2007b). It can be questioned how Bayesian the inverse solutions actually are, as the uncertainties involved in these estimates are not properly quantified. Secondly, typically such priors or regularisation methods which somehow constrain magnitudes of the currents also produce bias towards those source locations which produce larger gain to the sensors based on their physical location and orientation (for MEG, these are usually the superficial and/or tangential). This bias has been sometimes compensated by using "depth weighting" (for the effect of this on MNE, see, *e.g.* Lin et al., 2006). From viewpoint of general statistical inference, it is not obvious how meaningful this compensation is. The depth weighting corresponds to a prior assumption that deeper/weakly visible sources are larger in amplitude and/or extent, which indeed they must be for producing comparable MEG fields above the noise level. Now the problem is that if good data fits are also obtained by placing smaller sources to the well-visible locations, based on what information the deeper and larger sources should be favoured? With a "minimum-something" based prior source model, it is most likely necessary to somehow enforce the sources deeper. In order not to confuse the model with conflicting priors, the superficial counterparts of the deep sources (*i.e.*, superficial sources or combinations of such, which produce similar field patterns to the deep sources) should possibly be somehow eliminated from the space of possible sources to render the estimates reliable.

Because there evidently is a correspondence with hemodynamic and electromagnetic measures of brain activity Logothetis et al. (2001), one might very well want to favour some cortical locations as generators of the MEG/EEG data if corresponding fMRI data is available. We incorporated the spatial fMRI information to the estimates as initialisation of the prior precisions. The manner in which this was done is still somewhat crude, and the results showed little improvement in quality. Kilner et al. (2005) classify attempts to combine fMRI with MEG/EEG as integration through 1) prediction 2) constraints 3) fusion (forward models). Our approach is tangential to the integration through constraints, but more of an fMRI "guide" (Ahlfors and Simpson, 2004; Auranen et al., 2007). It seems that only integration through the forward models would properly oblige the inverse method to explain missing MEG sources and/or deeper source locations. The way in which such aspects could be encoded to the hierarchical Bayesian scheme studied here remains a challenging research problem.

The presented method is quite automatic, as only the parameter γ_0 needs to be tuned for suitable regularisation (generic initialisation such as $\hat{\alpha}_i = 1$ should be expected to work reasonably well with various datasets). It would be also interesting to see how the hierarchical method performs in a real user test similar to the one done by Stenbacka et al. (2002). Simulated data are typically too simple to reveal the real utility of a model, and more empirical data must be collected. For example, suitable retinotopic stimuli could be used to obtain MEG measurements from sources with varying depth and extent for studying what type of assumptions and constraints are needed to correctly estimate these aspects.

Acknowledgments

This research was supported in part by Academy of Finland (projects: 105682, 200521, 206368, 210347, Centre of Excellence: 202871, 213464, 213470), The Center for Functional Neuroimaging Technologies (NIH grant P41 RR14075), NIH grants 5R01HD040712-04 and 5R01NS044319-04, as well as Department of Energy under Award Number DE-FG02-99ER62764 to The MIND Institute. A. N. would like to thank David Wipf for providing a manuscript of his thesis along with helpful discussion.

References

Ahlfors, S. P. and Simpson, G. V. (2004). Geometrical interpretation of fMRI-guided MEG/EEG inverse estimates. *NeuroImage*, 22:323–332.

Auranen, T., Nummenmaa, A., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Vehtari, A., and Sams, M. (2005). Bayesian analysis of the neuromagnetic inverse problem with ℓ^p-norm priors. *NeuroImage*, 26(3):870–884.

Auranen, T., Nummenmaa, A., Vanni, S., Vehtari, A., Hämäläinen, M. S., Lampinen, J., and Jääskeläinen, I. P. (2007). Automatic fMRI-guided MEG multidipole localization for visual responses. *Submitted*.

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

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. John Wiley & Sons, Ltd.

Bertrand, C., Hamada, Y., and Kado, H. (2001a). MRI prior computation and parallel tempering algorithm: A probabilistic resolution of the MEG/EEG inverse problem. *Brain Topography*, 14(1).

Bertrand, C., Ohmi, M., Suzuki, R., and Kado, H. (2001b). A probabilistic solution to the MEG inverse problem via MCMC methods: The reversible jump and parallel tempering algorithms. *IEEE Transactions on Biomedical Engineering*, 48(5).

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage*, 9:179–194.

Dale, A. M. and Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, 5(2):162–176.

Delon-Martin, C., Gobbelé, R., Buchner, H., Haug, B. A., Antal, A., Darvas, F., and Paulus, W. (2006). Temporal pattern of source activities evoked by different types of motion onset stimuli. *NeuroImage*, 31:1567–1579.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9:195–207.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman & Hall/CRC, second edition.

Ghahramani, Z. and Beal, M. (2001). *Graphical Models and Variational Methods*. In *Advanced Mean Field Methods — Theory and Practice*, eds. M. Opper and D. Saad. MIT Press.

Gorodnitsky, I. F., George, J. S., and Rao, B. D. (1995). Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical Neurophysiology*, 95:231–251.

Hämäläinen, M. S. and Ilmoniemi, R. J. (1984). Interpreting measured magnetic fields of the brain: Estimates of current distributions. Technical Report TKK-F-A559, Helsinki University of Technology, Department of Technical Physics.

Honkela, A. and Valpola, H. (2004). Variational learning and bits-back coding: An information-theoretic view to Bayesian learning. *IEEE Transactions on Neural Networks*, 15(4):800–810.

Jun, S. C., George, J. S., Paré-Blagoev, J., Plis, S. M., Ranken, D. M., Schmidt, D. M., and Wood, C. C. (2005). Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data. *NeuroImage*, 28(1):84–98.

Jun, S. C., George, J. S., Plis, S. M., Ranken, D. M., Schmidt, D. M., and Wood, C. C. (2006). Improving source detection and separation in a spatiotemporal Bayesian inference dipole analysis. *Physics in Medicine and Biology*, 51:2395–2414.

Kaipio, J. P. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer.

Kilner, J., Mattout, J., Henson, R., and Friston, K. (2005). Hemodynamic correlates of EEG: A heuristic. *NeuroImage*, 28:280–286.

Lin, F.-H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S.

(2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*.

Liu, A. K., Dale, A. M., and Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*, 16:47–62.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157.

Mattout, J., Phillips, C., Penny, W. D., Rugg, M. D., and Friston, K. J. (2006). MEG source localization under multiple constraints: An extended Bayesian framework. *NeuroImage*, 30:753–767.

Mosher, J. C., Leahy, R. M., and Lewis, P. S. (1999). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46(3):245–259.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag.

Nummenmaa, A., Auranen, T., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Sams, M., and Vehtari, A. (2007a). Hierarchical Bayesian estimates of distributed MEG sources: Theoretical aspects and comparison of variational and MCMC methods. *NeuroImage*, 35(2):669–685.

Nummenmaa, A., Auranen, T., Hämäläinen, M. S., Jääskeläinen, I. P., Sams, M., Vehtari, A., and Lampinen, J. (2007b). Automatic relevance determination based hierarchical Bayesian MEG inversion in practice. *NeuroImage*, 37(3):876–889.

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

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

Sato, M.-A., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., and Kawato, M. (2004). Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage*, 23:806–826.

Schmidt, D. M., George, J. S., and Wood, C. C. (1999). Bayesian inference applied to the electromagnetic inverse problem. *Human Brain Mapping*, 7:195–212.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., Luca, M. D., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N. D., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219.

Stenbacka, L., Vanni, S., Uutela, K., and Hari, R. (2002). Comparison of minimum current estimate and dipole modeling in the analysis of simulated activity in the human visual cortices. *NeuroImage*, 16:936–943.

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

Trujillo-Barreto, N. J., Aubert-Vázquez, E., and Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21:1300–1319.

Vanni, S., Dojat, M., Warnking, J., Delon-Martin, C., Segebarth, C., and Bullier, J. (2004a). Timing of interactions across the visual field in the human cortex. *NeuroImage*, 21:818–828.

Vanni, S., Henriksson, L., and James, A. C. (2005). Multifocal fMRI mapping of visual cortical areas. *NeuroImage*, 27:95–105.

Vanni, S., Warnking, J., Dojat, M., Delon-Martin, C., Bullier, J., and Segebarth, C. (2004b). Sequence of pattern onset responses in the human visual areas: An fMRI constrained VEP source analysis. *NeuroImage*, (21):801–817.

Wipf, D., Ramírez, R., Palmer, J., Makeig, S., and Rao, B. (2007). Analysis of empirical Bayesian methods for neuroelectromagnetic source localization. In Schölkopf, B., Platt, J., and Hofmann, T., editors, *Advances in Neural Information Processing Systems 19*, Cambridge, MA. MIT Press.

ISBN 978-951-22-9141-0
ISSN 1455-0474
PicaSet Oy, Helsinki 2007