

Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data

Sung C. Jun,* John S. George, Juliana Paré-Blagoev, Sergey M. Plis, Doug M. Ranken, David M. Schmidt, and C.C. Wood

Biological and Quantum Physics Group, MS D454, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 17 November 2004; revised 24 May 2005; accepted 1 June 2005
Available online 15 July 2005

Recently, we described a Bayesian inference approach to the MEG/EEG inverse problem that used numerical techniques to estimate the full posterior probability distributions of likely solutions upon which all inferences were based [Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Human Brain Mapping* 7, 195; Schmidt, D.M., George, J.S., Ranken, D.M., Wood, C.C., 2001. Spatial-temporal bayesian inference for MEG/EEG. In: Nenonen, J., Ilmoniemi, R. J., Katila, T. (Eds.), *Biomag 2000: 12th International Conference on Biomagnetism*. Espoo, Norway, p. 671]. Schmidt et al. (1999) focused on the analysis of data at a single point in time employing an extended region source model. They subsequently extended their work to a spatiotemporal Bayesian inference analysis of the full spatiotemporal MEG/EEG data set. Here, we formulate spatiotemporal Bayesian inference analysis using a multi-dipole model of neural activity. This approach is faster than the extended region model, does not require use of the subject's anatomical information, does not require prior determination of the number of dipoles, and yields quantitative probabilistic inferences. In addition, we have incorporated the ability to handle much more complex and realistic estimates of the background noise, which may be represented as a sum of Kronecker products of temporal and spatial noise covariance components. This reduces the effects of undermodeling noise. In order to reduce the rigidity of the multi-dipole formulation which commonly causes problems due to multiple local minima, we treat the given covariance of the background as uncertain and marginalize over it in the analysis. Markov Chain Monte Carlo (MCMC) was used to sample the many possible likely solutions. The spatiotemporal Bayesian dipole analysis is demonstrated using simulated and empirical whole-head MEG data.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Bayesian inference; Source localization; Spatiotemporal analysis; Dipole analysis; Magnetoencephalography; Markov Chain Monte Carlo

Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) are non-invasive techniques. These methods measure direct physical consequences of neuronal currents and are capable of resolving temporal patterns of neural activity in the millisecond range. The MEG/EEG source localization problem, which identifies active brain regions from measurements on or outside of the human head, has been important in medical diagnosis of conditions like epilepsy, in surgical planning, and in neuroscience research. However, the MEG/EEG source localization inverse problem is mathematically ill-posed, that is, it has no unique solution.

For several decades, researchers have worked to develop MEG/EEG source localization methods to try to overcome the inherent ill-posed nature of the inverse problem. A number of localization methods which assume a dipolar source or an extended source have been developed (see Hämäläinen et al. (1993) for review). Most existing approaches fall into two broad categories: (1) few-parameter models having $N_p \ll N_s$ and (2) many-parameter models having $N_p \gg N_s$, where N_p is the number of parameters to be estimated in the model and N_s is the number of measurements, typically the number of sensors in MEG/EEG system. In general, few-parameter models are solved by finding a best-fitting solution through various nonlinear optimization techniques (Hämäläinen et al., 1993; Mosher et al., 1992; Huang et al., 1998; Uutela et al., 1998a; Jun et al., 2002). Many-parameter models are usually solved by the minimum norm method, or variants of the same, that selects the one solution minimizing a specified norm from the many solutions that fit the data equally well (Hämäläinen and Ilmoniemi, 1994; Gorodnitsky et al., 1995; Robinson and Vrba, 1999; Pascual-Marqui et al., 1994).

Recently, new probabilistic approaches to the MEG/EEG source localization problem based on Bayesian inference using Markov Chain Monte Carlo (MCMC) have been reported by Schmidt et al. (1999), Bertrand et al. (2001a,b), and Kincses et al. (2003). Unlike other probabilistic approaches (Baillet and Garnero, 1997; Phillips et al., 1997), the Bayesian inference

* Corresponding author.

E-mail address: jschan@lanl.gov (S.C. Jun).

Available online on ScienceDirect (www.sciencedirect.com).

approach does not result in a single best solution to the problem but produces a large number of likely solutions that fit both the data and any prior information. From the many sampled likely solutions, we can characterize some statistical information on any feature of solutions. This provides an effective means for quantifying uncertainty that is distinct from the other approaches to quantify uncertainty in inverse algorithms (Medvick et al., 1989; Singh and Harding, 2000; Darvas et al., 2005). Schmidt and Bertrand focused on the analysis of data at a single point in time and demonstrated the utility of Bayesian inference both for including pertinent prior information (anatomical location and orientation, sparseness of regions of activity, limitations on current strength, and spatial correlation) and for yielding robust results in spite of the under-determined inverse problem. Schmidt et al. (1999) used an extended region model for neural activity and Reversible Jump (RJ) MCMC method, while Bertrand et al. (2001a,b) used a multi-dipole model and combined RJ-MCMC and Parallel Tempering (PT) MCMC method. Schmidt et al. (2001) extended their work to a Bayesian inference analysis of the full spatiotemporal MEG/EEG data set, using their extended region model for neural activity.

Here, we present a spatiotemporal Bayesian inference technique for multi-dipole analysis. Compared to the full spatiotemporal analysis for extended regions, it is faster and does not require the use of the subject's anatomical information. Furthermore, in distinction to most other dipole analyses, it does not require the prior determination of the number of dipoles.

We begin with an overview of the general techniques of Bayesian inference. Then, we formulate the posterior probability distribution by incorporating the relevant priors into the Bayesian framework. To reduce computation costs and to improve MCMC performance, the posterior probability distribution is simplified by a marginalization technique over current time courses and a noise covariance matrix. A speed-up strategy for computing the posterior probability distribution is proposed, the MCMC sampling technique is briefly introduced, and then noise covariance approximation is discussed. Finally, results from experiments on simulated and empirical data are presented.

Formulation of Bayesian inference

Bayesian inference is a general procedure for constructing a posterior probability distribution for quantities of interest from the measurements and the given prior probability distributions for all uncertain parameters. The method is conceptually simple and relatively straightforward for even complicated problems.

The starting point for Bayesian inference is Bayes' rule of probability:

$$P(\theta, \mathbf{B}) = P(\theta|\mathbf{B})P(\mathbf{B}) = P(\mathbf{B}|\theta)P(\theta),$$

If θ represents parameters of interest and \mathbf{B} represents data depending on θ , then the probability of θ given \mathbf{B} is

$$P(\theta|\mathbf{B}) = \frac{P(\theta, \mathbf{B})}{P(\mathbf{B})} = \frac{P(\mathbf{B}|\theta)P(\theta)}{P(\mathbf{B})}.$$

Here, $P(\theta, \mathbf{B})$ is the joint probability distribution for θ and \mathbf{B} , $P(\theta|\mathbf{B})$ is the conditional probability distribution of θ given \mathbf{B} , $P(\mathbf{B})$ is the marginal probability distribution of \mathbf{B} , and $P(\theta)$ is the

prior probability distribution of θ , which represents one's information of θ before measurement. $P(\mathbf{B}|\theta)$ is the likelihood function which modifies the prior $P(\theta)$ to produce the posterior probability distribution $P(\theta|\mathbf{B})$. Since $P(\mathbf{B})$ is independent of θ , it is constant and can be omitted from the posterior density:

$$P(\theta|\mathbf{B}) \propto P(\mathbf{B}|\theta)P(\theta).$$

Bayes' rule of probability formulates how prior information and measurements can be combined and encoded in the posterior distribution. Commonly, the obtainable posterior distribution is complex and in such cases is numerically sampled using MCMC techniques (Chen et al., 2000; Gilks et al., 1995).

In this work, we propose a spatiotemporal MEG/EEG dipole analysis based on Bayesian inference. This analysis is formulated in the following way: assuming a localized effective dipole nature of the neuromagnetic sources that can explain the spatiotemporal data, we construct a current model that assumes a variable number of current dipoles of brain activity that are composed of their locations within a sphere of some radius R_0 , dipole orientations, and current time courses representing dipole magnitudes over time. Furthermore, we assume a fixed dipole model, where dipole locations and orientations are fixed over time, but dipole magnitudes vary over time. There can be any number N of active current dipoles from minimum N_{\min} up to some maximum N_{\max} . We used a spherical head model and the Sarvas forward model (Sarvas, 1987), but our analysis could employ other forward models as well.

Given the spatiotemporal measurement set, the Bayesian formulation is as follows:

$$P(N, \mathbf{X}, \Theta, \mathbf{J}|\mathbf{B}) \propto P(\mathbf{B}|N, \mathbf{X}, \Theta, \mathbf{J}) P(\Theta|N) P(\mathbf{J}|N) P(N)$$

\mathbf{B}	$T \times L$ matrix representing observed spatiotemporal data. L and T represent the number of sensors and the number of time samples in measurements, respectively.
N	A priori unknown number of dipole sources.
$\mathbf{X} = (X_1, X_2, \dots, X_N)$	Vector of N dipole sources, with each $X_i = (x_i, y_i, z_i)$ representing the location of the i -th dipole.
$\mathbf{J} = (J_1, J_2, \dots, J_N)$	Vector of N current time courses, with each $J_i = (j_i^1, j_i^2, \dots, j_i^T)$ representing signed dipole moment magnitude over time of i -th dipole. Negative sign means that dipole moment orientation is reversed.
$\Theta = (\theta_1, \theta_2, \dots, \theta_N)$	Vector of N dipole moment orientations, with each θ_i representing a unit tangential direction of i -th dipole.

The prior distributions are constructed as follows:

- The dipole current time course prior distribution is chosen as a Gaussian distribution:

$$P(\mathbf{J}|N) = \frac{1}{\prod_{\alpha=1}^N [(2\pi\sigma_\alpha^2)^{T/2} |\mathbf{C}_{cu}|^{1/2}]} e^{-\frac{1}{2} \sum_{\alpha=1}^N \frac{1}{\sigma_\alpha^2} J_\alpha^T \mathbf{C}_{cu}^{-1} J_\alpha}. \quad (1)$$

Here, $|\cdot|$ denotes the determinant. \mathbf{C}_{cu} is the temporal correlation matrix of one time point with another, which allows us to include the temporal correlation at nearby latencies. σ_α represents the prior standard deviation of time varying current magnitudes of each dipole. Both \mathbf{C}_{cu} and σ_α are predetermined based on spatiotem-

poral data to be analyzed and its expected active source properties. In the examples described later, \mathbf{C}_{cu} is parameterized as

$$\mathbf{C}_{cu}(i,j) = e^{-\frac{1}{2}(i-j)^2/\beta^2}, \quad i,j = 1,2,\dots,T. \quad (2)$$

$\beta \geq 0$ is a parameter which controls correlation time scale. The larger β yields the stronger correlation, while the smaller β yields the weaker correlation. The prior standard deviation of current magnitude σ_α can be predetermined depending on signal strength of spatiotemporal data. For example, for the very early response signal from a somatosensory median nerve stimulus experiment, if it is expected that time varying current magnitudes can be reached up to around 80 nAm at the peak, σ_α can be set to some value between 0 and 80 nAm.

- The dipole location prior distribution is chosen to be uniform over the region of the whole head model:

$$P(\mathbf{X}|N) = \begin{cases} \left(\frac{3}{4\pi R_0^3}\right)^N & \mathbf{X}_\alpha \in \text{Ball}(R_0) \quad \forall \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

$\text{Ball}(R_0)$ is a spherical head model with a radius of R_0 . We consider the spherical head model in this work. However, if MRI data are available, it is also possible to confine the region of our interest to brain cortex with a uniform distribution. Another possible approach is to give a normal distribution centered on brain cortex since most active dipole sources are on or close to the cortex area.

- The dipole moment orientation prior distribution is chosen to be uniform on $[0, 2\pi]$:

$$P(\Theta|\mathbf{X}, N) = \left(\frac{1}{2\pi}\right)^N.$$

We take into account only the unit orientation vectors which are orthogonal to radial vectors, i.e., dipole location vectors, since radial directions of dipoles have no effect on MEG signal in a spherical forward model. Orientation vector can be parameterized with just one angular variable θ . This restriction would be relaxed if a non-spherical forward model were used.

- The prior distribution for the number of dipoles is chosen to be uniform:

$$P(N) = \frac{1}{N_{\max} - N_{\min} + 1}.$$

Usually, $N_{\min} = 0$ in this work. Poisson distribution is also usable.

- Assuming an LT -dimensional Gaussian noise model of mean zero and noise covariance matrix \mathbf{C}_0 , the likelihood $P(\mathbf{B}|N, \mathbf{X}, \Theta, \mathbf{J})$ of the observed measurements \mathbf{B} for given parameters N , \mathbf{J} , \mathbf{X} , and Θ is described as:

$$P(\mathbf{B}|N, \mathbf{X}, \Theta, \mathbf{J}) = \frac{1}{(2\pi)^{LT/2} |\mathbf{C}_0|^{1/2}} e^{-\frac{1}{2} \text{vec}(\mathbf{B} - \mathbf{B}_c)' \mathbf{C}_0^{-1} \text{vec}(\mathbf{B} - \mathbf{B}_c)}, \quad (3)$$

$$\mathbf{B}_c = \mathbf{J}\mathbf{A},$$

where \mathbf{B}_c is a calculated measurement through the forward model. \mathbf{A} is an $N \times L$ matrix representing lead field and depends on dipole location \mathbf{X} and dipole orientation Θ . $\text{vec}(\mathbf{E})$ denotes a vector stacked in all the columns of \mathbf{E} . The symbol $'$ denotes transposition and $^{-1}$ inversion.

Finally, we obtain the Bayesian formulation of spatiotemporal MEG localization problem through Bayes' rule:

$$P(N, \mathbf{X}, \Theta, \mathbf{J}|\mathbf{B}) \propto \frac{1}{\prod_{\alpha=1}^N [(2\pi\sigma_\alpha^2)^{T/2} |\mathbf{C}_{cu}|^{1/2}]} e^{-\frac{1}{2} (\mathbf{B}'_c \mathbf{C}_0^{-1} \mathbf{B}_c + \sum_{\alpha=1}^N \frac{1}{\sigma_\alpha^2} \mathbf{C}_{cu}^{-1} j_\alpha)}. \quad (4)$$

Here, $\mathbf{B}_c = \text{vec}(\mathbf{B} - \mathbf{J}\mathbf{A})$. For simplicity, the constant was dropped in Eq. (4).

Sampling issues

Having first formulated a complete probabilistic description of the problem and then constructed the posterior, the next step in Bayesian inference is to extract a representative sample of likely solutions from the posterior distribution using an MCMC sampling technique. Our MEG source localization problem can be categorized into a very high dimensional problem. For example, if $N = 3$ dipoles, $L = 121$ sensors, and $T = 70$ time samples, then the dimension of this problem (i.e., number of total unknown parameters we should sample) amounts to 223 for Eq. (4). In our experience, a straightforward implementation of MCMC technique for problem Eq. (4) was likely to fail to extract well-distributed samples. This was determined by generating multiple MCMC runs with different starting points and random seeds. A good test of whether an MCMC sampling has converged and is well sampled is if the results from such multiple chains are consistent. We discovered that this posterior distribution (4) has numerous valleys and peaks over the parameter region of interest, and this formulation of the posterior had the following problems for sampling:

- High complexity: MCMC has more chance to be trapped in a local maximum. This is analogous to the local minima problem encountered with other spatiotemporal multi-dipole analyses (Huang et al., 1998).

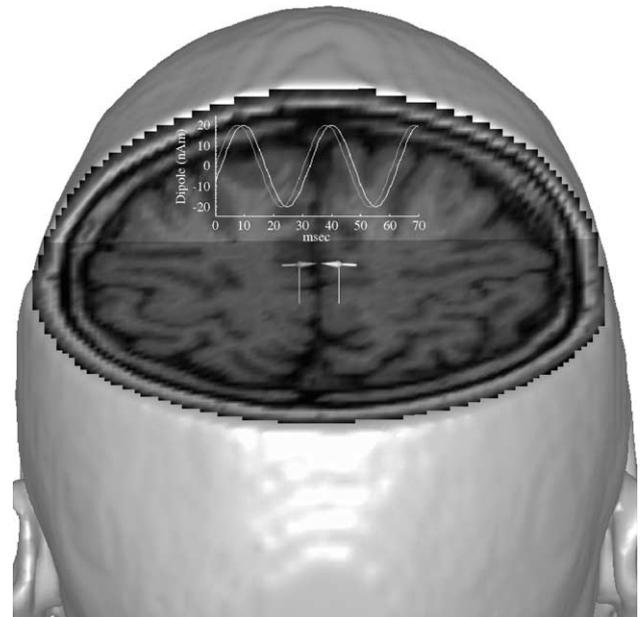


Fig. 1. Two dipole sources problem.

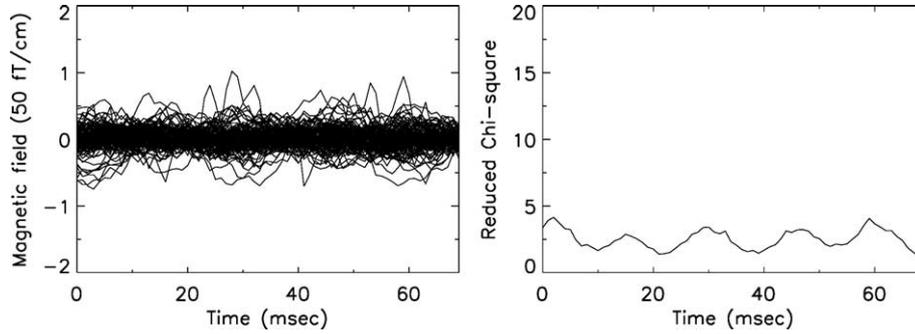


Fig. 2. Two dipole sources problem. Left: simulated spatiotemporal data. Right: reduced Chi-square time course (RCTC) for simulated data.

- Slow convergence: MCMC requires tremendous time to explore the whole complex range.

These observations indicated that we had formulated the problem in a way that was too rigid. In order to overcome high complexity and slow convergence, smoothing of this high dimensional posterior distribution function as well as reduction of dimension of unknown parameters can be attempted in a way which critical information of distribution function will not be lost. We know that the inverse problem has a number of uncertainties, and we can include those to help smooth out the posterior. In reality, noise covariance is complex and is only approximated and thus has inherent uncertainty. Incorporation of noise covariance \mathbf{C} as an uncertain parameter into the Bayesian inference frame yields:

$$P(N, \mathbf{X}, \Theta, \mathbf{J}, \mathbf{C} | \mathbf{B}) \propto P(\mathbf{B} | N, \mathbf{X}, \Theta, \mathbf{J}, \mathbf{C}) P(\Theta | \mathbf{X}, N) P(\mathbf{J} | N) P(\mathbf{X} | N) P(\mathbf{C}) P(N).$$

The prior distribution of the noise covariance matrix is chosen as the k -th inverse Wishart distribution¹ (Triantafyllopoulos, 2002):

$$P(\mathbf{C}) = \frac{1}{Z(k)} |r_0 \mathbf{C}_0|^{(k-r_0-1)/2} |\mathbf{C}|^{-k/2} e^{-\frac{1}{2} \text{Tr}(r_0 \mathbf{C}_0 \mathbf{C}^{-1})},$$

$$Z(k) = 2^{(k-r_0-1)r_0/2} \pi^{r_0(r_0-1)/4} \prod_{j=1}^{r_0} \Gamma\left(\frac{k-r_0-j}{2}\right),$$

$$r_0 = LT. \quad (5)$$

Here, k is a degree of freedom which should be greater than $2r_0$, i.e., twice the dimension of \mathbf{C} . \mathbf{C}_0 is an estimation of a noise covariance matrix, which can be given an identity matrix or any rough estimation of a noise covariance matrix. $\Gamma(\cdot)$ is a gamma function, and $\text{Tr}(\mathbf{A})$ denotes a trace of a matrix \mathbf{A} . It is known that expectation $E[\mathbf{C}]$ with respect to the above distribution is \mathbf{C}_0 (Triantafyllopoulos, 2002).

¹ If X_i for $i = 1, \dots, m$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix Σ , and \mathbf{X} denotes the $m \times p$ matrix composed of the row vectors X_i , then the $p \times p$ matrix $\mathbf{X}^T \mathbf{X}$ has a Wishart distribution with scale matrix Σ and degrees of freedom parameter m . The Wishart distribution is most typically used when describing the covariance matrix of multi-normal samples. In this work, we are interested in an inverse noise covariance, so the inverse Wishart distribution is chosen.

Replacing \mathbf{C}_0 in Eq. (3) with \mathbf{C} yields $P(\mathbf{B} | N, \mathbf{X}, \Theta, \mathbf{J}, \mathbf{C})$, and the final smooth Bayesian formulation of spatiotemporal MEG localization problem is obtained as follows:

$$P(N, \mathbf{X}, \Theta, \mathbf{J}, \mathbf{C} | \mathbf{B}) \propto \frac{|r_0 \mathbf{C}_0|^{(k-r_0-1)/2} |\mathbf{C}|^{-(k+1)/2}}{Z_0 \prod_x^N \left[(2\pi\sigma_x^2)^{T/2} |\mathbf{C}_{cu}|^{1/2} \right]} \times e^{-\frac{1}{2} \left[\mathbf{B}_x^T \mathbf{C}^{-1} \mathbf{B}_x + \sum_{\alpha=1}^N \frac{1}{\sigma_x^2} \mathbf{C}_{cu}^{-1} j_{\alpha} + \text{Tr}(r_0 \mathbf{C}_0 \mathbf{C}^{-1}) \right]},$$

where $Z_0 = Z\left(\frac{4T(T-1)\pi^2 R_0^3}{3}\right)^N (2\pi)^{r_0/2} (N_{\max} - N_{\min} + 1)$. (6)

In order to reduce the dimension of unknown parameters and to make the distribution that is to be sampled more smooth, it makes sense to marginalize the distribution function over some redundant or less interesting parameters. The noise covariance matrix has a large number of parameters that are typically not of interest. In addition, for an implementation of MCMC for the posterior distribution, where every parameter of current time courses was sampled, it was very difficult to generate a time course at random that would have a decent chance of being accepted in the MCMC. We found that marginalization over time courses \mathbf{J} could avoid this problem. Therefore, we have good reasons to marginalize the posterior over a noise covariance as well as time courses.

Marginalization analysis

Marginalization of noise covariance \mathbf{C} and current time courses \mathbf{J}

Here, we describe marginalization over both \mathbf{C} and \mathbf{J} , and the other significant parameters are sampled from the marginalized

Table 1

Distribution	Number of MCMC runs stuck in a local maximum	Average localization error (per dipole)
Marginal posterior over \mathbf{J}	11	3.98 mm
Marginal posterior over \mathbf{J} and \mathbf{C}	1	4.89 mm

Two dipole sources problem: number of MCMC runs that were stuck in a local maximum and the average localization error per dipole for two marginal posterior distributions. A total of 50 MCMC runs were tested. Using the posterior that marginalizes over both \mathbf{J} and \mathbf{C} results in a significant reduction of local maxima errors with only a marginal increase in localization error.

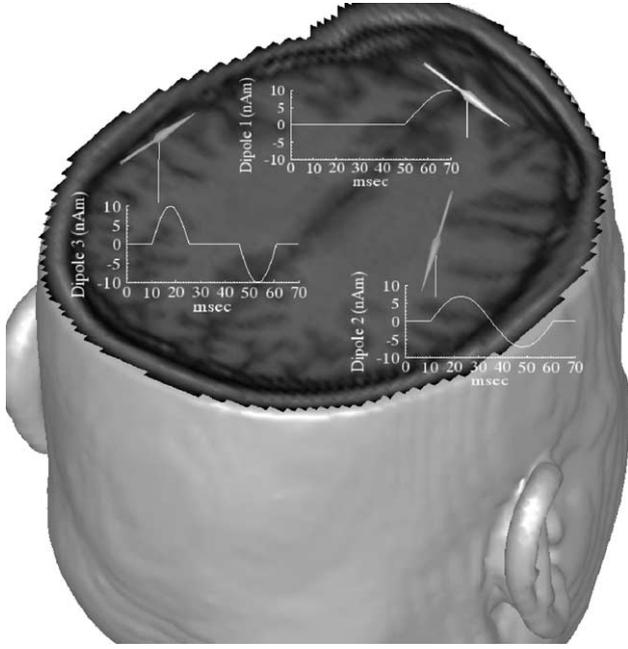


Fig. 3. The location, orientation and time courses of the three dipole sources used to generate simulated data as described in Three dipole sources problem: results and discussions.

posterior distribution function by a MCMC technique. Current time courses \mathbf{J} are separately sampled from the approximate normal distribution, which is obtained by substituting the other sampled parameters into Eq. (15). We first attempt marginalization over noise covariance \mathbf{C} and then marginalize over current time course \mathbf{J} . Assuming σ_z is independent of α for simplicity, we rewrite Eq. (6) in terms of \mathbf{C} :

$$\frac{|r_0 \mathbf{C}_0|^{(k-r_0-1)/2} |\mathbf{C}|^{-(k+1)/2}}{Z_0 (2\pi\sigma^2)^{NT/2} |\mathbf{C}_{cu}|^{N/2}} e^{-\frac{1}{2} \text{Tr}([r_0 \mathbf{C}_0 + \mathbf{B}_n \mathbf{B}_n'] \mathbf{C}^{-1}) + \frac{1}{\sigma^2} \text{Tr}(\mathbf{J}' \mathbf{C}_{cu}^{-1} \mathbf{J})}. \quad (7)$$

Here, we used the identity $\mathbf{v}' \mathbf{A} \mathbf{v} = \text{Tr}(\mathbf{v} \mathbf{v}' \mathbf{A})$ for any vector \mathbf{v} and any matrix \mathbf{A} . Integration of Eq. (7) over \mathbf{C} yields (see Appendix A for more detail):

$$\frac{Z(k+1) |r_0 \mathbf{C}_0|^{-1/2}}{Z_0 (2\pi\sigma^2)^{NT/2} |\mathbf{C}_{cu}|^{N/2}} \left(1 + \frac{1}{r_0} \mathbf{B}_n' \mathbf{C}_0^{-1} \mathbf{B}_n\right)^{-(k-r_0)/2} e^{-\frac{1}{2\sigma^2} \text{Tr}(\mathbf{J}' \mathbf{C}_{cu}^{-1} \mathbf{J})}. \quad (8)$$

Now, we are in a position to integrate over \mathbf{J} . By using the approximation $(1+x)^{-p} \approx e^{-px}$ for $0 < x \ll 1$ and $p \gg 0$, we

obtain the final approximated marginal posterior distribution (see Appendix B for more detail):

$$P(N, \mathbf{X}, \Theta | \mathbf{B}) \propto \frac{(1+r)^{-(k-r_0)/2}}{|2\sigma^2 \mathbf{C}_2|^{1/2}} e^{-\frac{(k-r_0)}{\sigma^2(1+r)} q' \mathbf{C}_2^{-1} \mathbf{C}_1 q},$$

$$\mathbf{F} = (\mathbf{A}' \otimes \mathbf{I}),$$

$$\mathbf{C}_1 = \mathbf{F}' \mathbf{C}_0^{-1} \mathbf{F},$$

$$q = \mathbf{C}_1^{-1} \mathbf{F}' \mathbf{C}_0^{-1} \text{vec}(\mathbf{B}),$$

$$r = \frac{1}{r_0} \text{vec}(\mathbf{B})' [\mathbf{C}_0^{-1} - \mathbf{C}_0^{-1} \mathbf{F} \mathbf{C}_1^{-1} \mathbf{F}' \mathbf{C}_0^{-1}] \text{vec}(\mathbf{B}),$$

$$\mathbf{C}_2 = \frac{(k-r_0)}{2(1+r)r_0} \mathbf{C}_1 + \frac{1}{2\sigma^2} \mathbf{I}. \quad (9)$$

Speed-up strategy

Because Bayesian inference requires many likely solutions, we need an efficient way to compute the marginal posterior. For every computation of the marginal posterior probability distribution Eq. (9), we need to compute inversions of matrix \mathbf{C}_1 and matrix \mathbf{C}_2 of size $NT \times NT$. Problems having a small number of dipoles and a small number of time samples are easily dealt with. However, for any problem having tens of dipoles and hundreds of time samples, the inversion of a sizeable matrix (thousands by thousands) is required for every evaluation of the marginal posterior distribution, which would be time-consuming. For this reason, the following speed-up strategy was used:

First, we look at the temporal current correlation matrix \mathbf{C}_{cu} of size $T \times T$. Due to its symmetry and its positive definiteness (inherently given to correlation matrix), it can be eigenvalue decomposed as below:

$$\mathbf{C}_{cu} = \mathbf{P}_{cu} \mathbf{\Sigma}_{cu} \mathbf{P}_{cu}'$$

Looking into eigenvalues of \mathbf{C}_{cu} , we may see that there are some number (about 65% among its eigenvalues in example in Example: simulated data section) of almost vanishing eigenvalues. The condition number of matrix \mathbf{C}_{cu} might be large enough and its inversion might be very sensitive due to some vanishing eigenvalues. To avoid this numerical problem, it is conventional to eliminate vanishing eigenvalues within a certain threshold by setting them to zero and to remove their components from the computational domain. Thus, the above decomposition can be approximated by

$$\mathbf{C}_{cu} \approx \hat{\mathbf{P}}_{cu} \hat{\mathbf{\Sigma}}_{cu} \hat{\mathbf{P}}_{cu}' \quad \mathbf{C}_{cu}^{-1} \approx \hat{\mathbf{P}}_{cu} \hat{\mathbf{\Sigma}}_{cu}^{-1} \hat{\mathbf{P}}_{cu}',$$

where $\hat{\mathbf{\Sigma}}_{cu}$ is a $T' \times T'$ diagonal matrix consisting of T' (should be $\ll T$) significant eigenvalues, and $\hat{\mathbf{P}}_{cu}$ is a $T \times T'$ matrix consisting of eigenvectors corresponding to significant eigenvalues.

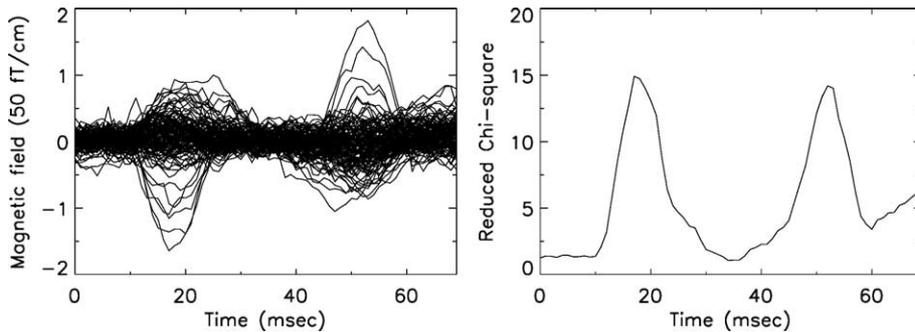


Fig. 4. Three dipole sources problem. Left: simulated spatiotemporal data. Right: reduced Chi-square time course (RCTC) for simulated data.

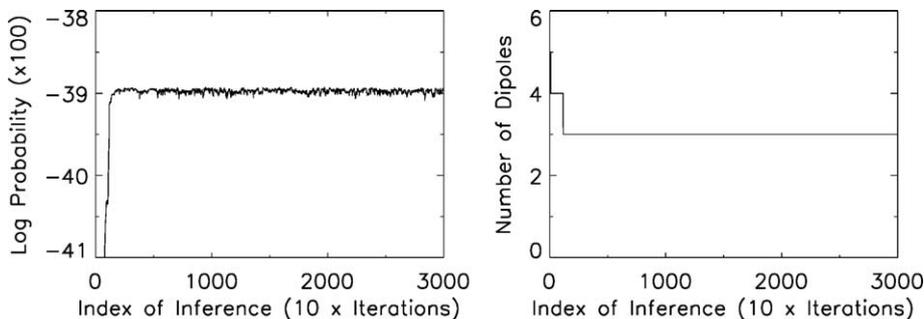


Fig. 5. Three dipole sources problem. Log probability (left) and number of dipoles (right) as a function of index of MCMC sample for the diagonal one pair approximation.

In order to use the above approximated decomposition in the Bayesian frame, the current time course prior $\tilde{\mathbf{J}} = \hat{\Sigma}_{\text{cu}}^{-1/2} \hat{\mathbf{P}}_{\text{cu}} \mathbf{J}$ in eigenspace is given below:

$$P(\tilde{\mathbf{J}}|N) = \prod_{\alpha=1}^N \left(\frac{1}{(2\pi\sigma_{\alpha}^2)^{T/2}} \right) e^{-\frac{1}{2} \sum_{x=1}^T \frac{\tilde{J}_{\alpha}^2}{\sigma_{\alpha}^2}}. \quad (10)$$

Replacing Eq. (1) into Eq. (10) in the posterior distribution Eq. (6) and marginalizing it over both \mathbf{C} and $\tilde{\mathbf{J}}$ yields

$$P_{\tilde{\mathbf{J}}}(N, \mathbf{X}, \Theta|\mathbf{B}) \propto \frac{(1+r)^{-(k-r_0)/2}}{|2\sigma^2 \mathbf{C}_2|^{1/2}} e^{-\frac{(k-r_0)}{\sigma^2(1+r)} q' \mathbf{C}_2^{-1} \mathbf{C}_1 q}, \quad (11)$$

$$\tilde{\mathbf{F}} = \mathbf{A}' \otimes (\hat{\mathbf{P}}_{\text{cu}} \hat{\Sigma}_{\text{cu}}^{1/2}).$$

The other parameters \mathbf{C}_1 , q , r , and \mathbf{C}_2 are the same as those in Eq. (9) if \mathbf{F} is replaced into $\tilde{\mathbf{F}}$. Since $\tilde{\mathbf{F}}$ is an $LT \times NT'$ matrix, the dimensions of \mathbf{C}_1 and \mathbf{C}_2 becomes smaller NT' than those of \mathbf{C}_1 and \mathbf{C}_2 in Eq. (9). This plays an important role in speeding up the MCMC technique.

In addition, the inversion of matrix \mathbf{C}_0 of size $LT \times LT$ should be computed once. This usually requires a large amount of memory storage as well as significantly large computation cost. Our solution to this computation and storage issue will be discussed later in Noise covariance approximation.

The formulation is transformed from original time variables to eigentime variables, keeping only the significant ones. Thus, the solutions in the MCMC are calculated using this truncated space, and then the formulation is transformed back into original time variables for recording. Finally, we use Eq. (11) as the final marginal posterior distribution which is to be sampled using MCMC.

Reversible jump Markov Chain Monte Carlo technique

We can now readily use MCMC schemes to sample the posterior probability distribution $P_{\tilde{\mathbf{J}}}(N, \psi|\mathbf{B})$ in Eq. (11) on the parameter space $\cup_{N \in \{N_{\min}, \dots, N_{\max}\}} \{N\} \times \Psi^N$, where Ψ is a collection of all possible state ψ representing all parameters except for N to be sampled from Eq. (11). MCMC methods are primarily used to construct a Markov chain $(\{N^{(0)}, \psi^{(0)}\}, \{N^{(1)}, \psi^{(1)}\}, \{N^{(2)}, \psi^{(2)}\}, \dots)$ and to choose the transition probabilities $P(\{N^{(k+1)}, \psi^{(k+1)}\}|\{N^{(k)}, \psi^{(k)}\})$ in such a way that the probability distribution of k -th realization converges to targeted distribution as k goes to infinity, i.e.,

$$\lim_{k \rightarrow \infty} P(\{N^{(k)}, \psi^{(k)}\}) = P_{\tilde{\mathbf{J}}}(N, \psi|\mathbf{B}).$$

After discarding samples during a burn-in period (in other words, using a convergence process), drawing realizations of the Markov chain gives us a random sample of the probability distribution. In order to assure convergence of MCMC, it is sufficient that the following ‘‘detailed-balance’’ condition is satisfied:

$$\begin{aligned} P_{\tilde{\mathbf{J}}}(N^{(k)}, \psi^{(k)}|\mathbf{B}) P(N^{(k+1)}, \psi^{(k+1)}|N^{(k)}, \psi^{(k)}) \\ = P_{\tilde{\mathbf{J}}}(N^{(k+1)}, \psi^{(k+1)}|\mathbf{B}) P(N^{(k)}, \psi^{(k)}|N^{(k+1)}, \psi^{(k+1)}). \end{aligned}$$

In order to sample without prior determination of the number of dipoles, we needed a trans-dimensional sampling strategy, i.e., jumps between subspaces of different dimensions. Green (1995) proposed the Reversible Jump MCMC technique which combines classical Metropolis moves with Reversible Jump (RJ) moves and allows movement between different parameter spaces that satisfies the detailed balance. The move from (N, ψ) to $(N^*,$

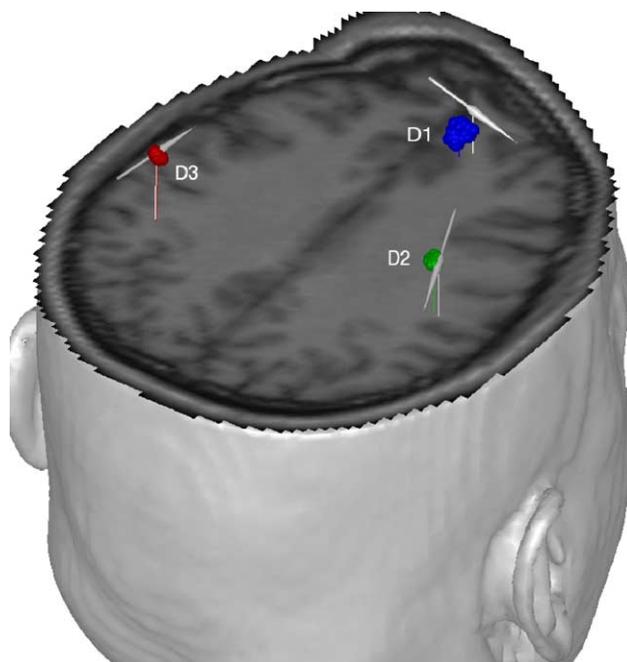


Fig. 6. Three dipole sources problem. MCMC result on the diagonal one pair approximation: 3-dimensional locations of regions containing dipoles from 200 MCMC samples (color) together with true dipole locations and orientations (gray arrows).

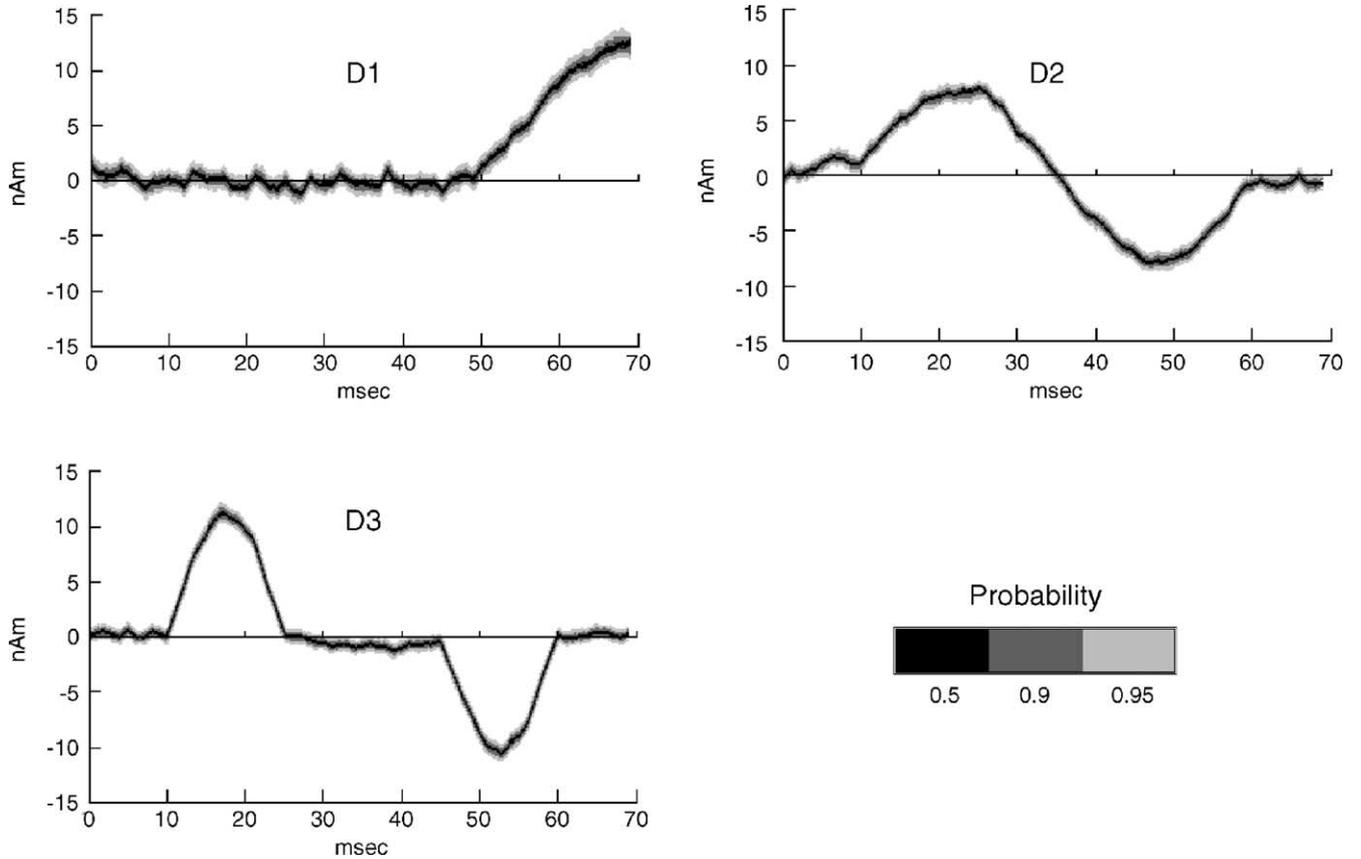


Fig. 7. Three dipole sources problem. MCMC result on the diagonal one pair approximation: histogram plots of time courses for 2000 collected inferences. Each gray intensity represents a probability level of reconstructed dipole magnitude at each time from low (black) to high (light gray)—0.50, 0.90, and 0.95. Top left to bottom left—dipole 1, dipole 2, dipole 3 (D1–D3).

ψ^*) is proposed by drawing a random vector \mathbf{u} with probability density function $g(\mathbf{u})$ and setting $\psi^* = f^*(\psi, \mathbf{u})$. The reverse move is proposed in the same way by drawing \mathbf{u}^* from a probability density function $g^*(\mathbf{u}^*)$ and setting $\psi = f(\psi^*, \mathbf{u}^*)$, which \mathbf{u} and \mathbf{u}^* satisfies $\text{Dim}(\psi) + \text{Dim}(\mathbf{u}) = \text{Dim}(\psi^*) + \text{Dim}(\mathbf{u}^*)$. In order to satisfy detailed balance, the new candidate sample (N^*, ψ^*) is accepted with probability

$$\alpha = \min \left\{ 1, \frac{P_{\mathbf{J}}(N^*, \psi^* | \mathbf{B}) j(N^* \rightarrow N) g^*(\mathbf{u}^*)}{P_{\mathbf{J}}(N, \psi | \mathbf{B}) j(N \rightarrow N^*) g(\mathbf{u})} \left| \frac{\partial(\psi^*, \mathbf{u}^*)}{\partial(\psi, \mathbf{u})} \right| \right\},$$

where $j(a \rightarrow b)$ is the probability proposing movement from a to b . In particular, for movement between adjacent dimensional subspaces, i.e., between $\{N\} \times \Psi^N$ and $\{N+1\} \times \Psi^{N+1}$,

we can simplify the acceptance ratio by removing \mathbf{u} or \mathbf{u}^* like this:

$$\alpha_{N \rightarrow N^* = N+1} = \min \left\{ 1, \frac{P_{\mathbf{J}}(N^*, \psi^* | \mathbf{B}) j(N^* \rightarrow N)}{P_{\mathbf{J}}(N, \psi | \mathbf{B}) j(N \rightarrow N^*) g(\mathbf{u})} \left| \frac{\partial \psi^*}{\partial(\psi, \mathbf{u})} \right| \right\},$$

and

$$\alpha_{N^* = N+1 \rightarrow N} = \min \left\{ 1, \frac{P_{\mathbf{J}}(N, \psi | \mathbf{B}) j(N \rightarrow N^*) g(\mathbf{u})}{P_{\mathbf{J}}(N^*, \psi^* | \mathbf{B}) j(N^* \rightarrow N)} \left| \frac{\partial(\psi, \mathbf{u})}{\partial \psi^*} \right| \right\}.$$

In our RJ-MCMC procedure, a candidate sample (N^*, ψ^*) is chosen from two categorized proposal distributions:

- Trans-dimensional proposal
 - Birth move: a new dipole and its parameters are proposed.

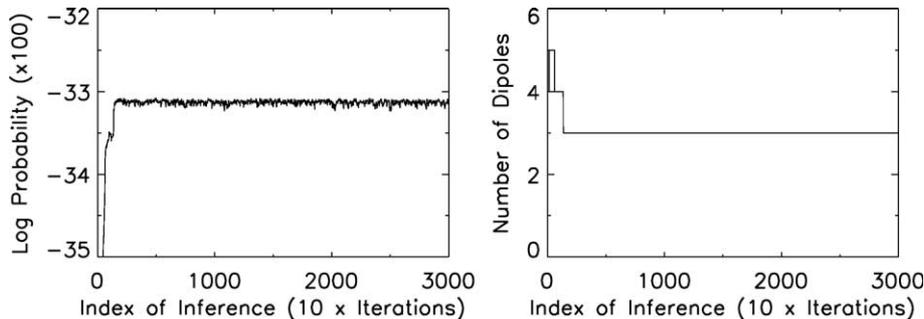


Fig. 8. Three dipole sources problem. Log probability (left) and number of dipoles (right) as a function of MCMC sample index or the multi-pair approximation.

- Death move: a randomly chosen dipole is proposed to be removed.
- Update proposal
 - Location update move: a dipole is randomly chosen and its new location is proposed. Each location component is proposed by Gaussian random generator with standard deviation $\sigma_{\text{loc}} = 1$ mm.
 - Orientation update move: a dipole is randomly chosen, and its new orientation is proposed by Gaussian random generator with standard deviation $\sigma_{\text{ori}} = 0.1$ radian.

Noise covariance approximation

The most commonly used noise covariance approximation is the uncorrelated diagonal model consisting of sensor variances. When we applied this noise model in our analysis using a nonzero temporal correlation prior on data consisting of empirical noise and simulated sources, many spurious dipoles arose to model the correlations that are typically present in empirical data (Kuriki et al., 1994). This is understandable given the fact that we had set up the analysis for uncorrelated noise and correlated signal but had applied it to data with empirical noise that had spatial and temporal correlation. Thus, the analysis tried to model these correlations using extra dipoles. We can reduce this effect by using an uncorrelated temporal prior, as is shown in the following example of Three dipole sources problem: results and discussions. Using a better model for the noise covariance allowed us to use a correlated temporal prior and improve the accuracy of the results.

The averaged (assuming M_1 epochs²) noise covariance matrix \mathbf{C}_0 in its most general form can be estimated from all collected real brain noise sets $\{\mathbf{n}^1, \mathbf{n}^2, \dots, \mathbf{n}^{M_2}\}$ (assuming M_2 sets³ like this (usually $M_2 \gg M_1$)⁴:

$$\mathbf{C}_0 = \frac{1}{M_1(M_2 - 1)} \times \sum_{i=1}^{M_2} (\text{vec}(\mathbf{n}^i) - \text{vec}(\bar{\mathbf{n}})), (\text{vec}(\mathbf{n}^i) - \text{vec}(\bar{\mathbf{n}}))' \quad (12)$$

$$\bar{\mathbf{n}} = \frac{1}{M_2} \sum_{i=1}^{M_2} \mathbf{n}^i. \quad (13)$$

Looking into Eq. (11), we find that the inversion of \mathbf{C}_0 is needed to calculate the posterior distribution. However, this matrix can be very large. For the problem in the following example ($L = 121$ sensors, $T = 70$ time samples), the inversion of an 8470×8470 matrix is required. This would be severely time consuming

² To get reasonable (in terms of S/N) evoked responses, many replicates of stimulus-locked post-stimulus signal are averaged. In our experiment, 602 (M_1) and 328 (M_1) replicates were averaged in Example: simulated data and in Example: empirical data, respectively.

³ To estimate noise covariance, we need as many real noise sets as possible. In our experiment, we collected 1992 (M_2) and 1250 (M_2) single trial real noise sets from off-stimulus region in Example: simulated data and in Example: empirical data, respectively.

⁴ This assumption is not required in our analysis. In practice, because of subject fatigue, it is hard to collect as many replicates for the same stimulus as are needed. Practical alternatives include using large ISIs (interstimulus intervals) or collecting data during subject rest time for several minutes because thousands of off-stimulus signals can be obtained.

and would require a significant amount of memory to store after pre-computing its inverse. Under the reasoning that spatial and temporal covariances are almost independent and separable, the approximation of \mathbf{C}_0 by a Kronecker product of a spatial covariance matrix \mathbf{S} and a temporal covariance matrix \mathbf{T} has been recently proposed as a solution to these difficulties (Huizenga et al., 2002; De Munck et al., 2002), i.e.,

$$\mathbf{C}_0 \approx \mathbf{S} \otimes \mathbf{T}.$$

Huizenga et al. (2002) proposed a parameterized spatial model and Toeplitz temporal model and found an optimized one pair approximation in the sense of Frobenius matrix norm. De Munck et al. (2002) developed an unparameterized one pair approximation by maximum likelihood method. This approximation has the valuable property that its inversion is easily computed and the memory required for approximate \mathbf{C}_0 to be stored is tremendously small. To obtain a better approximation of spatiotemporal noise structure, we developed a multi-pair approximation by using the sums of pairs (the number of pairs is the same as the number of sensors) of Kronecker product. In this way, the inversion is still easily computed:

$$\mathbf{C}_0 \approx \sum_{l=1}^L \mathbf{S}_l \otimes \mathbf{T}_l, \mathbf{C}_0^{-1} \approx \sum_{l=1}^L \mathbf{S}_l \otimes \mathbf{T}_l^{-1},$$

where \mathbf{S}_l and \mathbf{T}_l represent l -th the spatial covariance component and temporal covariance component, respectively (Plis et al., 2005). \mathbf{S}_l are obtained by using singular value decomposition (SVD) in space, so they are mutually orthonormal in matrix product sense. After transforming the data into an orthonormal spatial basis, the \mathbf{T}_l are estimated. In the following examples, we used this multi-pair noise covariance approximation together with a correlated temporal prior ($\beta > 0$ in Eq. (2)) as well as a diagonal

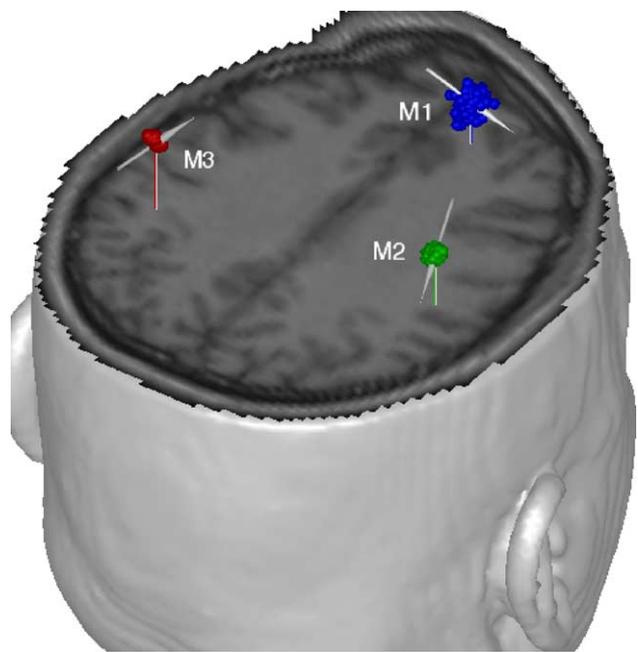


Fig. 9. Three dipole sources problem. MCMC result on the multi-pair approximation: 3-dimensional locations of regions containing dipoles from 200 MCMC samples (color) together with true dipole locations and orientations (gray arrows).

noise covariance approximation with an uncorrelated temporal prior ($\beta = 0$) for experiments.

Example: simulated data

Data

Note that although we present applications of this Bayesian inference analysis to MEG data only, this analysis can be applied to EEG data as well. Empirical MEG noise (for generating simulated MEG data and for estimating a noise covariance) was acquired from the following experiment:

Median nerve stimulation at the motor twitch threshold was applied using a block design of 30 s on, 30 s off for a total of 10 blocks for each of 8 runs. Data were acquired during both stimulation ‘on’ and ‘off’ epochs, the latter being used to construct the present noise data set. Stimulus alternated across runs, with four runs total of left side stimulation and four runs total of right side stimulation. The ISI (interstimulus interval) was randomized from between 0.25 and 0.75 s. Data were collected with 1 kHz sampling from a male subject, age 38, on a 4D Neuroimaging Neuromag-122 whole-head gradiometer system with 122 channels (Ahonen et al., 1993). One of 122 channels was discarded due to its malfunctioning. Structural MR data on the same subject were collected on a 1.5 T Picker scanner for registration purposes. Data were (1-median)⁵ filtered. 60 Hz noise and its harmonics were filtered out by removal of peaks in the spectrum and interpolation between adjacent spectrum points.

To generate simulated data, we added the simulated signal from the given dipole source information to the noise data.⁶ As a time window size, we used 70 (T) time samples and averaged 602 (M_1) stimulation ‘off’ epochs to produce our sample noise data. We used reduced Chi-square time course (RCTC) as our measure of the signal to noise ratio (SNR) as follows:

$$\text{RCTC}(t) = \sum_{s=1}^L \left(B_{t,s}^2 / \text{var}_s \right) / L.$$

Here, $B_{t,s}$ is a spatiotemporal signal at time t and channel s , and var_s denotes s -th channel noise variance. L is the number of channels. This is a statistical measure that will have a value around 1 for pure noise. Values above 1 indicate increasingly significant evidence for the presence of signal.

For estimation of noise covariance, we collected empirical single trial spatiotemporal noise data sets ($M_2 = 1992^7$ for 70 time window size) from this experiment during off-stimulus periods and approximated a noise covariance as a multi-pair Kronecker product approximation. In addition, a commonly used noise

covariance was estimated as a pair Kronecker product of a diagonal spatial covariance consisting of 121 (L) channel noise variances and an identity temporal covariance. Both models had the same sensor variances and only differed in their correlation structure.

Convergence of MCMC: local maxima issue and discussions

Most MEG/EEG source localization methods have encountered local maxima (or minima) problems. To overcome them, heuristic methods such as simulated annealing (SA) and genetic algorithms (GA) have been developed (Uutela et al., 1998b). Even though these methods mathematically guarantee that global maximum can be obtained, they often require enormous computation time. As alternatives, multi-start iterative methods (Huang et al., 1998) (choosing best one among results from many initial starts) and hybrid iterative methods (Jun et al., 2002; Jun and Pearlmuter, 2005) (combining automatic or semi-automatic fast initializer into conventional iterative method) have been proposed.

In our analysis, a marginalization technique over both noise covariance and time courses has been applied to overcome the local maxima problem in Marginalization analysis. In this section, we investigate how our proposed analysis works in the special MEG localization problem we generated, which is a complicated two dipole sources problem presumably having many local maxima. Fig. 1 illustrates this two dipole sources problem—two dipoles are located closely and almost oppositely oriented; their time courses have the same shapes, and phases are slightly different. The spatiotemporal signal from two dipoles was added to real noise data, which is the same as the one in Example: simulated data. Fig. 2 describes the simulated signal data (overlapped plot) over 121 channels and its RCTC. For the comparison, conventional MCMC procedure was applied for two distribution functions—marginal posterior distribution⁸ over time courses \mathbf{J} and marginal posterior distribution over time courses \mathbf{J} and noise covariance \mathbf{C} (our proposed analysis: Eq. (11)). The diagonal noise covariance approximation we estimated in Example: simulated data with an uncorrelated temporal prior was used for both.

For each distribution function, we ran 50 MCMC⁹ runs. Each run started with a randomly chosen initial start configuration near the true dipole location and was allowed to sample a total of 2000 inferences¹⁰ (20,000 iterations here). For each run, the first 1500 inferences were automatically discarded to allow a burn-in-period and the remaining 500 sampled inferences were used to check whether that run’s results were in a local maximum. An MCMC run was labeled as being stuck in a local maximum if the locations of that run’s reconstructed dipoles did not encompass the true dipole locations and the maximum posterior probability of these solutions was orders of magnitude lower than the posterior probability of the true solution. This latter condition was included because it is possible for there to exist a local maximum with a posterior probability

⁵ We subtracted median filtered data from unfiltered original data.

⁶ The same forward model as in our analysis was applied.

⁷ All 602 single trial noise data sets for generating simulated data were collected from uniformly between 10 and 80 ms before stimulus onset. Single trial noise data sets for noise covariance approximation were collected as many as possible. This was determined by the duration of the off-stimulus period, which varied over the epoch. In this way, we ensured that each of single trial noise data sets for simulated data would be at least partly and in some cases totally used for noise covariance approximation.

⁸ This marginal distribution can be obtained by marginalizing Eq. (4) over time courses \mathbf{J} only.

⁹ $N_{\max} = 2$ and $N_{\min} = 0$. The standard deviation of the dipole magnitude prior σ for all dipoles was set to 10 nA.

¹⁰ One inference was randomly chosen and recorded among every 10 inferences during sampling procedure. Thus, one inference required 10 iterations.

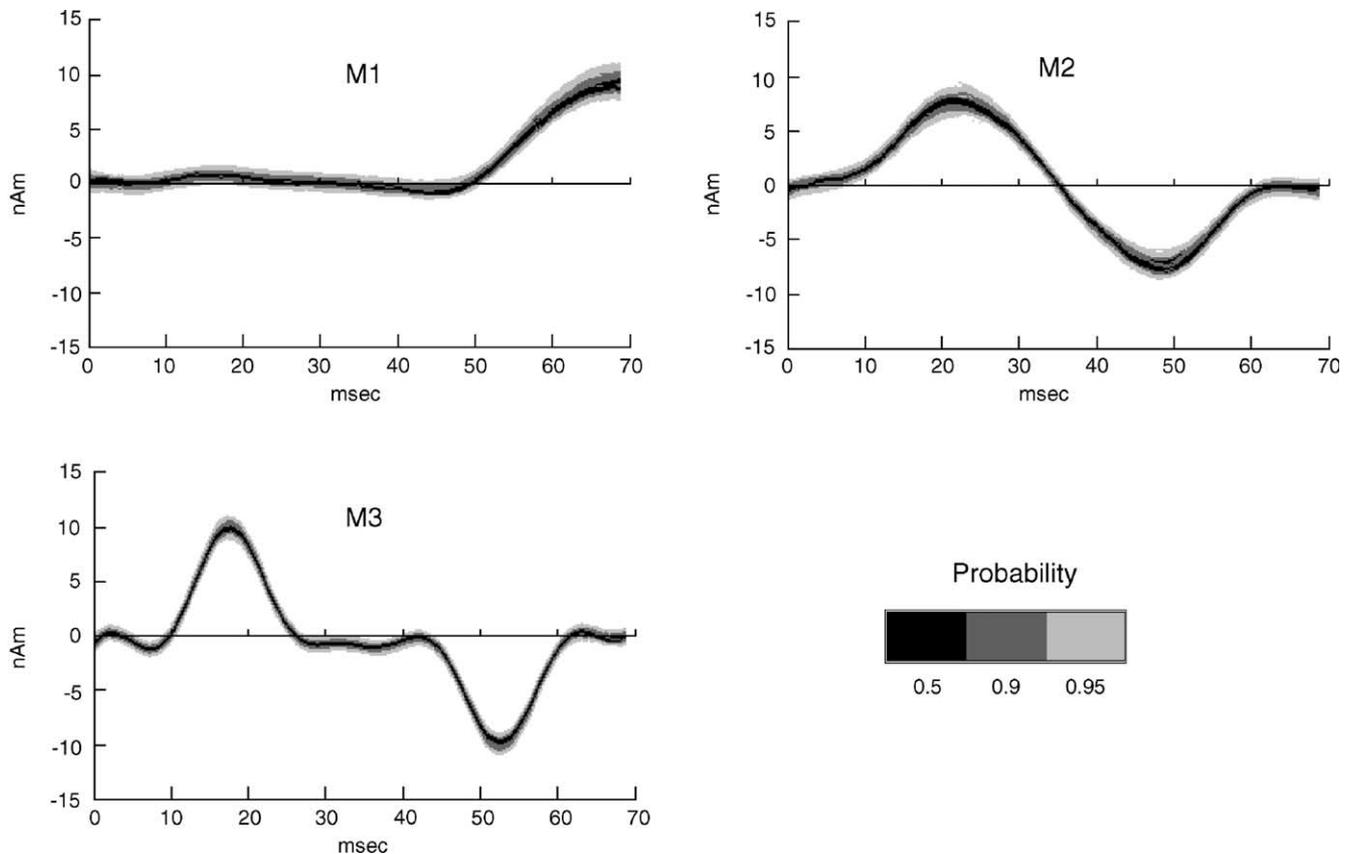


Fig. 10. Three dipole sources problem. MCMC result on the multi-pair approximation: histogram plots of time courses for 2000 collected inferences. Each gray intensity represents a probability level of reconstructed dipole magnitude at each time from low (black) to high (light gray)—0.50, 0.90, and 0.95. Top left to bottom left—dipole 1, dipole 2, dipole 3 (M1–M3).

near that of the true solution that a sampling algorithm should in fact sample.

Table 1 shows the number of MCMC runs that were stuck in a local maximum for each distribution function. For our proposed analysis, only 1 among 50 MCMC runs was stuck in a local maximum, while 11 out of 50 runs were stuck in a local maximum for the posterior distribution that marginalized only over \mathbf{J} . Of these 11, there were many different local maxima, and all had maximum posterior probabilities that were significantly smaller than that of the true solution. For those runs that were not stuck in a local maximum, we calculated the average localization error per dipole for both types of posterior distribution. These results are shown in the right column of Table 1 and indicate that there is about a millimeter more error for the distribution that marginalized over both \mathbf{J} and \mathbf{C} . This reflects the extra uncertainty that was added by treating the noise covariance as uncertain. The magnitude of this increased error is much smaller than the error that would be caused by mistakenly using the results from one of the local maxima so that it is much better to use the posterior that marginalized over both \mathbf{J} and \mathbf{C} . We use this posterior for the remainder of the analyses presented in this paper.

Finally, we note that for sampling algorithms all local maxima should be visited in proportion to their posterior probability. Indeed, MCMC jumping rules are constructed to mathematically guarantee that long iterations yield perfect sampling. However, the length of the MCMC chain needed to achieve this objective is

not limited and becomes a very important practical issue. There is also the possibility that we could construct a more efficient MCMC algorithm that would be less prone to get stuck in a local maximum (Green and Mira, 2001; Haario et al., 1999). This will be examined in future work.

Three dipole sources problem: results and discussions

To further investigate the feasibility of our proposed Bayesian inference dipole analysis, we generated a three dipole source problem. Their locations, orientations, and current time courses are illustrated in Fig. 3, and the simulated signal data over 121 channels and a measure of its signal to noise are described in Fig. 4. The same empirical noise as was used in the previous section was used here. In this section, we examine how different noise covariance approximations (diagonal or multi-pair) affect the performance of our analysis.

The posterior distribution function Eq. (11) for each noise covariance approximation was sampled using MCMC with $N_{\min} = 0$, $N_{\max} = 9$ and the standard deviation of the dipole magnitude prior σ for all dipoles was set to 20 nAm. The MCMC runs generated 3000 inferences.¹¹ After discarding the first 1000

¹¹ Actually, one inference was randomly chosen and recorded among every 10 inferences during sampling procedure. Thus, one inference required 10 iterations.

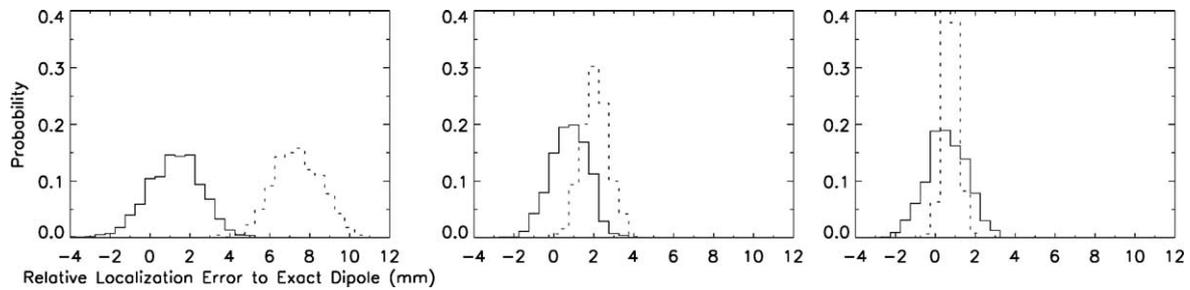


Fig. 11. Three dipole sources problem. Relative location error histograms for the diagonal one pair approximation (dotted line) and the multi-pair approximation (solid line). We used 2000 collected inferences for each. Left to right—(D1 vs. M1), (D2 vs. M2), (D3 vs. M3).

inferences due to burn-in-period, the remaining 2000 inferences were chosen for analysis and comparison.

Diagonal noise covariance

For the diagonal noise covariance case, we used a current time course prior which is temporally uncorrelated ($\beta = 0$ in Eq. (2)). As has already been stated, the use of a correlated temporal prior with a diagonal (uncorrelated) noise covariance yields extra spurious dipoles that attempt to model the correlated noise. The use of an uncorrelated temporal prior eliminates this problem as is shown in this example. Fig. 5 shows the number of dipoles and log probability of the targeted posterior probability distribution as a function of collected MCMC sample for the diagonal one pair approximation. Three dipole sources were finally sampled after approximately 200 inferences, and we consistently found three sources for even multiple MCMC runs for different random seeds and initial starts after some amount of burn-in-periods.

Fig. 6 shows 3-dimensional locations of regions that contained the reconstructed dipoles from randomly chosen 200 among a total of 2000 MCMC samples together with the true dipole locations. The region of the weakest dipole (D1) looks more broadly distributed than those of other dipoles, which is due to the well known fact that the higher signal-to-noise (S/N) signals can be more focused and better localized than the lower S/N signals. Furthermore, this region for D1 appears to not encompass the true location. This will be examined more closely in comparison to the results with the multi-pair noise covariance.

Any feature or inference (such as dipole location) can be quantified probabilistically using the MCMC samples. For example, the radius of the sphere that contains each dipole at a 90% level is estimated by finding the radius of the sphere that contains a dipole in 90% of the MCMC samples. For the three dipole sources, this was found to be 2.63 mm, 1.36 mm, and 1.31 mm (from dipole D1 to dipole D3), respectively. We can also quantify the probability for time courses. Fig. 7 shows the posterior distribution of each dipole's time course. Here, we have displayed it using a few distinct probability levels. These were constructed by forming a two-dimensional histogram of the time courses from a total of 2000 MCMC samples. These distributions are consistent with the true time courses.

Multi-pair noise covariance

With the multi-pair noise covariance, we ran the MCMC sampling code with a correlated time course prior ($\beta = 8$ ms in Eq. (2)). Figs. 8–10 display the results in the same manner as for the diagonal noise covariance case. Here, too, three dipoles were found, and the time course distributions are consistent with the true time courses, although they are smoother, reflecting the use of a

correlated time course prior. However, for this case, the locations of all three reconstructed dipoles clearly encompass the true locations. The radii of the spheres that contain dipoles at a 90% level were found to be 4.36 mm, 2.36 mm, and 1.80 mm (from dipole M1 to dipole M3), respectively.

In order to investigate in more detail the possible localization error problem of D1 in the diagonal noise covariance case, we examined the posterior distribution of the relative localization error to the exact dipole location. This was constructed by histogramming the reconstructed dipole locations across the MCMC samples projected along the axis going through the exact dipole location (origin) and the peak of the posterior location distribution. These results are shown in Fig. 11 for each dipole. Clearly, D1 in the diagonal noise covariance is not correctly localized as its posterior distribution does not encompass the true location (at 0 in these plots). We believe this is due to using a noise covariance that does not adequately capture the correlation structure of the noise, given that the more complex multi-pair noise covariance results do not have this problem. The width of the posterior distributions are also slightly smaller for the diagonal noise approximation than for the multi-pair approximation. This is because the lack of positive correlation in the diagonal approximation increases the effective degrees of freedom in the data yielding an erroneous decrease in the width of the posterior location distributions.

These results suggest that the use of the diagonal (uncorrelated) noise covariance model in our analysis with empirical data, which typically has a complex spatial and temporal correlation structure, can lead to location errors and or spurious dipoles. However, this needs to be investigated in more detail, using other data sets and a wider range of source configurations. We also intend to investigate the effects of prior distribution parameters.

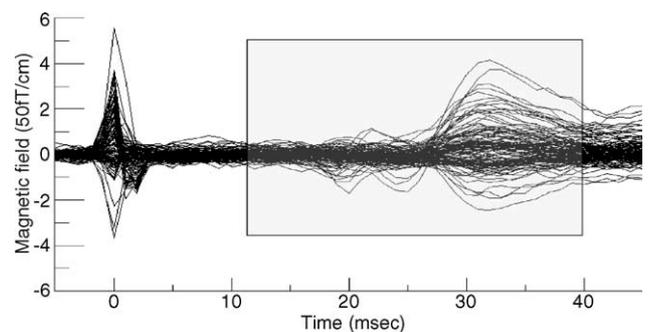


Fig. 12. Empirical Data. Left hand median nerve stimulation spatiotemporal data from 122 channels are overlapped. The shaded region shows the time window that was analyzed.

Example: empirical data

Data

Here, we demonstrate the performance of our analysis with an empirical median nerve stimulation MEG data set that is from a different experiment and a different subject than in the previous sections. We make use of all the features of our analysis method that were described in the previous sections. These include the use of the multi-pair noise covariance model with temporally correlated time course prior and treating the noise as uncertain and sampling from the posterior distribution that has been marginalized over both time courses and noise covariance.

The empirical data were acquired from the following experiment:

The median nerve was stimulated using two surface electrodes placed on the forearm. A 0.5 ms current pulse was applied using a Grass Constant Current Stimulator. The electrodes and voltage were adjusted until a thumb twitch was obtained in each hand. If the maximum voltage was reached without a thumb twitch, the subject was run with the maximum voltage. The right and left median nerves were stimulated randomly with a 0.5 s ISI. Data were digitized at 1 kHz with the online filters set to 0.03–330 Hz. An interval of 0.1 s pre-stimulus and 0.5 s

post-stimulus were collected. Data were collected from a single subject on a 4D Neuroimaging Neuromag-122 whole-head gradiometer system with 122 channels. Structural MR data on the same subject were collected on a 1.5 T Picker scanner for registration purposes. Data were (1-median) filtered to remove low frequency drifts but were not filtered for 60 Hz noise and its harmonics because their effects were negligibly small.

A total of 328 (M_1) epochs of left hand stimulation data were averaged to produce the data set to be analyzed. Sensor waveform overplots for this data set are shown in Fig. 12. The data from a total of 30 ms (T) time samples starting 11 ms post-stimulus onset were analyzed. The multi-pair noise covariance parameters were estimated from $M_2 = 1250$ single trial spatiotemporal noise epochs that were far in time from a stimulus or a stimulus evoked neural response.

Results and discussions

This data set was analyzed by applying our MCMC sampling algorithm using the temporal time course prior distribution shown in Eq. (2) with $\beta = 3$ ms, and the standard deviation of the dipole magnitude prior σ for all dipoles was set to 20 nAm. The prior for the number of dipoles was uniform between 0 and 9.

Multiple MCMC runs were generated using different random number seeds and starting configurations. Each run generated

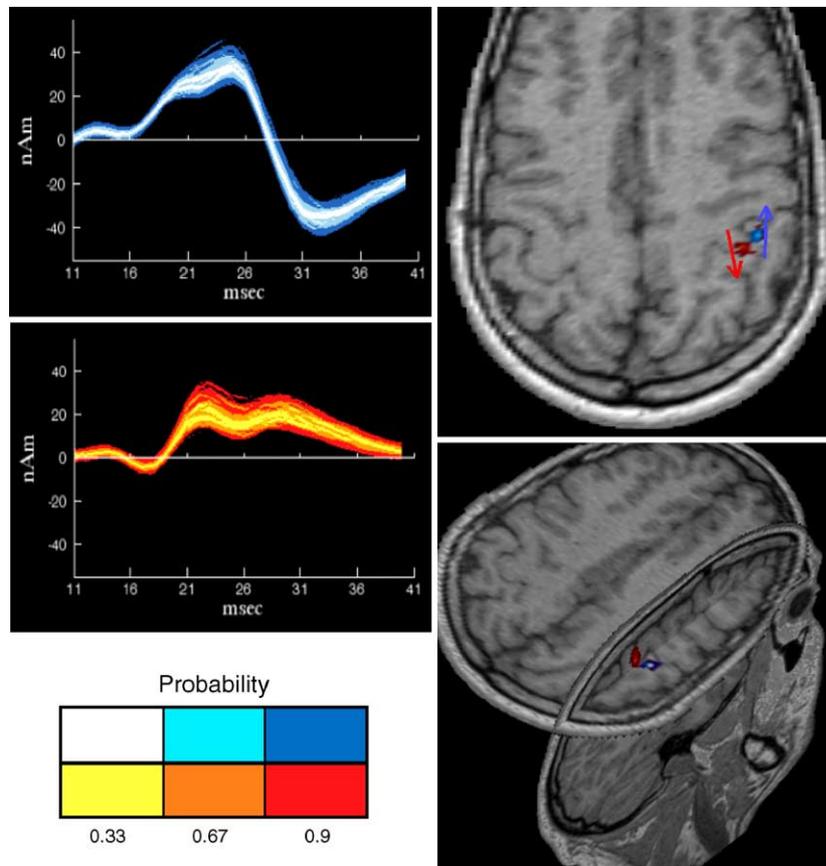


Fig. 13. Empirical data. Results from our analysis showing the posterior probability distributions for the time courses (left column) and locations (right column) of the 2 reconstructed dipoles. The probability has been quantitated into distinct levels as shown in the color bar in the lower left panel. Both spatial and temporal distributions use the same probability levels.

5000 samples, of which the final 3000 were used in making inferences about the results. The results from all runs were found to be consistent. Here, we present the results from one MCMC run.

Even though the number of dipoles was allowed to vary, only two dipoles were consistently obtained across the MCMC samples, both located on the postcentral gyrus near the central sulcus, one somewhat more medial and posterior than the other. Fig. 13 shows the posterior probability distributions for the locations and time courses of the two dipole sources, superimposed on the subject's anatomical MRI. These distributions are quantitative and have been quantitated into distinct posterior probability levels, as shown in the color bar in the lower left panel of the figure. Both spatial and temporal distributions are shown with the same probability levels.

These spatial and temporal distributions are strongly reminiscent of previous reports of electrical and magnetic recordings of short-latency somatosensory activity in humans and monkeys as reported by Allison et al. (1991a,b); McCarthy et al. (1991); Wood et al. (1985, 1988); Huang et al. (2000). The more lateral source (shown in shades of blue) was deeper on the postcentral gyrus, adjacent to the central sulcus. The more medial source (shown in shades of red) was located more superficially on the crown of the postcentral gyrus, straddling the sulcus itself so that there was nonzero probability both anterior and posterior to the sulcus. These patterns are similar to those seen in invasive recordings in monkeys (McCarthy et al., 1991) and have been attributed by Allison and colleagues to activity in cytoarchitectonic areas 3b and 1, respectively, on the postcentral gyrus (Allison et al., 1991a,b). The issue of whether the precentral gyrus (cytoarchitectonic area 4) contributes to any of this activity remains controversial (e.g., Huang et al., 2000), and the present results are consistent with either interpretation. We also analyzed the data from the right hand stimulation and found very similar results in terms of the number, locations relative to the contralateral central sulcus, and time courses of the reconstructed dipoles.

Conclusion

We have demonstrated a Bayesian inference dipole analysis for spatiotemporal MEG/EEG data that has several positive and innovative features. It does not require the prior determination of the number of dipoles, significantly reduces the common local minima problem of multi-dipole analyses by adding uncertainty for the noise covariance, and yields quantifiable probabilistic inferences. We have demonstrated the ability to handle complex and realistic estimates of the background noise and have shown that this can reduce the effects of undermodeling the complex correlation structure that is typically present in noise from empirical MEG/EEG data. Finally, experiments on both simulated and empirical data have shown the value and capability of our Bayesian inference dipole analysis.

Acknowledgments

This work was supported by NIH grant 2 R01 EB000310-05 and the Mental Illness and Neuroscience Discovery (MIND) Institute. For the first empirical data set, we thank Daniel

Sheltraw for help with conducting the empirical MEG experiment at the Albuquerque VA Medical Center and for help using the hard-real-time stimulation software, RTstim (<http://www.unm.edu/sheltraw/rtstim/>). We thank Cheryl Aine, Julia Stephen, and Sonja Kovacevic for access to the second empirical data set. We thank Elaine Best for help in preprocessing both empirical data sets using MEGAN (<http://www.lanl.gov/p/p21/megan.shtml>). Some figures showing the results of the Bayesian dipole analysis were generated using MRVIEW (Ranken et al., 2002).

Appendix A. Derivation of Eq. (8)

Let us start by taking into account \mathbf{C} of Eq. (7):

$$|r_0 \mathbf{C}_0|^{(k-r_0-1)/2} |\mathbf{C}|^{-(k+1)/2} e^{-\frac{1}{2}(\text{Tr}(r_0 \mathbf{C}_0 + \mathbf{B}_n \mathbf{B}_n' \mathbf{C}^{-1}))}. \quad (14)$$

We can rearrange it into $(k+1)$ -th inverse Wishart distribution having $\tilde{\mathbf{C}}_0 = \mathbf{C}_0 + 1/r_0 \mathbf{B}_n \mathbf{B}_n'$ in place of \mathbf{C}_0 :

$$|r_0 \mathbf{C}_0|^{(k-r_0-1)/2} Z(k+1) |r_0 \tilde{\mathbf{C}}_0|^{-(k-r_0)/2} \frac{|r_0 \tilde{\mathbf{C}}_0|^{(k-r_0)/2} |\mathbf{C}|^{-(k+1)/2}}{Z(k+1)} \\ \times e^{-\frac{1}{2}(\text{Tr}(r_0 \tilde{\mathbf{C}}_0 \mathbf{C}^{-1}))}.$$

Integration over \mathbf{C} using normalization property of probability distribution function yields

$$|r_0 \mathbf{C}_0|^{-1/2} Z(k+1) |\mathbf{C}_0^{-1} \tilde{\mathbf{C}}_0|^{-(k-r_0)/2}.$$

Substitution of $\mathbf{C}_0 + 1/r_0 \mathbf{B}_n \mathbf{B}_n'$ for $\tilde{\mathbf{C}}_0$ and properties (symmetry, positive definiteness) of \mathbf{C}_0 yields

$$|r_0 \mathbf{C}_0|^{-1/2} Z(k+1) \left| \mathbf{I} + \frac{1}{r_0} (\mathbf{C}_0^{-1/2} \mathbf{B}_n) (\mathbf{C}_0^{-1/2} \mathbf{B}_n)' \right|^{-(k-r_0)/2}.$$

Let us take into account

$$\left| \mathbf{I} + \frac{1}{r_0} (\mathbf{C}_0^{-1/2} \mathbf{B}_n) (\mathbf{C}_0^{-1/2} \mathbf{B}_n)' \right|.$$

Letting $\tilde{\mathbf{B}}_n = \mathbf{C}_0^{-1/2} \mathbf{B}_n$ and eigenvalue decomposition of $\tilde{\mathbf{B}}_n \tilde{\mathbf{B}}_n'$ yields

$$\tilde{\mathbf{B}}_n \tilde{\mathbf{B}}_n' = \mathbf{P} \mathbf{D} \mathbf{P}' \\ = \left(\frac{\tilde{\mathbf{B}}_n}{|\tilde{\mathbf{B}}_n|} \phi_2 \cdots \phi_{r_0} \right) \begin{pmatrix} \tilde{\mathbf{B}}_n' \tilde{\mathbf{B}}_n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{B}}_n' \\ |\tilde{\mathbf{B}}_n| \phi_2 \cdots \phi_{r_0} \end{pmatrix},$$

where ϕ_i , $i = 2, \dots, LT$ are orthonormal eigenvectors of $\tilde{\mathbf{B}}_n \tilde{\mathbf{B}}_n'$ which are orthogonal to $\tilde{\mathbf{B}}_n$. Computing $|\mathbf{I} + 1/r_0 \tilde{\mathbf{B}}_n \tilde{\mathbf{B}}_n'|$ gives

$$\left| \mathbf{I} + \frac{1}{r_0} \tilde{\mathbf{B}}_n \tilde{\mathbf{B}}_n' \right| = \left| \mathbf{I} + \frac{1}{r_0} \mathbf{P} \mathbf{D} \mathbf{P}' \right| = \left| \mathbf{I} + \frac{1}{r_0} \mathbf{D} \right| \\ = 1 + \frac{1}{r_0} \tilde{\mathbf{B}}_n' \tilde{\mathbf{B}}_n = 1 + \frac{1}{r_0} \mathbf{B}_n' \mathbf{C}_0^{-1} \mathbf{B}_n.$$

Finally, the integration of Eq. (14) over \mathbf{C} gives

$$|r_0 \mathbf{C}_0|^{-1/2} Z(k+1) \left(1 + \frac{1}{r_0} \mathbf{B}'_n \mathbf{C}_0^{-1} \mathbf{B}_n \right)^{-(k-r_0)/2}$$

Eq. (8) follows from this result.

Appendix B. Derivation of Eq. (9)

We start from Eq. (8):

$$\frac{Z(k+1) |r_0 \mathbf{C}_0|^{-1/2}}{Z_0 (2\pi\sigma^2)^{NT/2} |\mathbf{C}_{cu}|^{N/2}} \left(1 + \frac{1}{r_0} \mathbf{B}'_n \mathbf{C}_0^{-1} \mathbf{B}_n \right)^{-(k-r_0)/2} e^{-\frac{1}{2\sigma^2} \text{Tr}(\mathbf{J}' \tilde{\mathbf{C}}_0^{-1} \mathbf{J})}$$

Dropping constant of the above equation and rearranging yields

$$|1 + a + (\text{vec}(\mathbf{J}) - \mathbf{b})' \tilde{\mathbf{C}} (\text{vec}(\mathbf{J}) - \mathbf{b})|^{-p} e^{-\frac{1}{2\sigma^2} \text{vec}(\mathbf{J})' \tilde{\mathbf{C}}_{cu}^{-1} \text{vec}(\mathbf{J})},$$

where

$$a = \text{vec}(\mathbf{B})' \left(\frac{1}{r_0} [\mathbf{C}_0^{-1} - \mathbf{C}_0^{-1} \mathbf{X}' (\mathbf{X}' \mathbf{C}_0^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{C}_0^{-1}] \right) \times \text{vec}(\mathbf{B}),$$

$$\mathbf{b} = \tilde{\mathbf{C}}^{-1} \mathbf{X}' \mathbf{C}_0^{-1} \text{vec}(\mathbf{B})$$

$$\tilde{\mathbf{C}} = \mathbf{X}' \mathbf{C}_0^{-1} \mathbf{X},$$

$$\mathbf{X} = (\mathbf{A}' \otimes \mathbf{I}), p = (k - r_0)/2.$$

By using the approximation $(1+x)^{-p} \approx e^{-px}$, we obtain

$$(1+a)^{-p} e^{-\frac{p}{1+a} (\text{vec}(\mathbf{J}) - \mathbf{b})' \tilde{\mathbf{C}} (\text{vec}(\mathbf{J}) - \mathbf{b}) - \frac{1}{2\sigma^2} \text{vec}(\mathbf{J})' \tilde{\mathbf{C}}_{cu}^{-1} \text{vec}(\mathbf{J})}.$$

Rearranging in terms of $\text{vec}(\mathbf{J})$ yields

$$(1+a)^{-p} e^{-\left[\text{vec}(\mathbf{J})' \left(\frac{p}{1+a} \tilde{\mathbf{C}} + \frac{1}{2\sigma^2} \tilde{\mathbf{C}}_{cu}^{-1} \right) \text{vec}(\mathbf{J}) - \frac{2p}{1+a} \mathbf{b}' \tilde{\mathbf{C}} \text{vec}(\mathbf{J}) + \frac{p}{1+a} \mathbf{b}' \tilde{\mathbf{C}} \mathbf{b} \right]}.$$

Squaring the exponent part yields

$$(1+a)^{-p} e^{-\left(\text{vec}(\mathbf{J}) - \tilde{\mathbf{b}} \right)' \Sigma (\text{vec}(\mathbf{J}) - \tilde{\mathbf{b}}) - \frac{p^2}{1+a} \mathbf{b}' \tilde{\mathbf{C}} \mathbf{b} + \frac{p^2}{(1+a)^2} \mathbf{b}' \tilde{\mathbf{C}} \Sigma^{-1} \tilde{\mathbf{C}} \mathbf{b}}. \quad (15)$$

Here

$$\tilde{\mathbf{b}} = \frac{p}{1+a} \Sigma^{-1} \tilde{\mathbf{C}} \mathbf{b},$$

$$\Sigma = \frac{p}{1+a} \tilde{\mathbf{C}} + \frac{1}{2\sigma^2} \tilde{\mathbf{C}}_{cu}^{-1}.$$

Finally, integration over \mathbf{J} yields

$$\left(\frac{\pi^{NT}}{|\Sigma|} \right)^{1/2} (1+a)^{-p} e^{-\frac{p}{1+a} \mathbf{b}' \tilde{\mathbf{C}} \mathbf{b} + \frac{p^2}{(1+a)^2} \mathbf{b}' \tilde{\mathbf{C}} \Sigma^{-1} \tilde{\mathbf{C}} \mathbf{b}}.$$

Eq. (9) follows from this result, and realizations of $\text{vec}(\mathbf{J})$ can be drawn from the normal distribution $N(\tilde{\mathbf{b}}, \Sigma^{-1})$ during MCMC procedure.

References

- Ahonen, A.I., Hämäläinen, M.S., Knuutila, J.E.T., Kajola, M.J., Laine, P.P., Lounasmaa, O.V., Parkkonen, L.T., Simola, J.T., Tesche, C.D., 1993. 122-channel SQUID instrument for investigating the magnetic signals from the human brain. *Phys. Scr.* T49, 198–205.
- Allison, T., McCarthy, G., Wood, C.C., Jones, S.J., 1991a. Potentials

- evoked in human and monkey cerebral cortex by stimulation of the median nerve: a review of scalp and intracranial potentials. *Brain* 114, 2465–2503.
- Allison, T., Wood, C.C., McCarthy, G., 1991b. Cortical somatosensory evoked potentials: II. Effects of excision of somatosensory or motor cortex in humans and monkeys. *J. Neurophysiol.* 66, 64–82.
- Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomic-functional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* 44, 374–385.
- Bertrand, C., Hamada, Y., Kado, H., 2001a. MRI prior computation and parallel tempering algorithm: a probabilistic resolution of the MEG/EEG inverse problem. *Brain Topogr.* 14, 57–68.
- Bertrand, C., Ohmi, M., Suzuki, R., Kado, H., 2001b. A probabilistic solution to the MEG inverse problem via MCMC methods: the reversible jump and parallel tempering algorithms. *IEEE Trans. Biomed. Eng.* 48, 533–542.
- Chen, M., Shao, Q., Ibrahim, J.G., 2000. *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag.
- Darvas, F., Rautiainen, M., Pantazis, D., Baillet, S., Benali, H., Mosher, J.C., Garnero, L., Leahy, R.M., in press. Investigations of dipole localization accuracy in MEG using the bootstrap. *Neuroimage*.
- De Munck, J.C., Huizenga, H.M., Waldorp, L.J., Heethaar, R., 2002. Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise. *IEEE Trans. Biomed. Eng.* 50, 1565–1572.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1995. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC.
- Gordnitsky, I., George, J., Rao, B., 1995. Neuromagnetic source imaging with focus: a recursive weighted minimum norm algorithm. *Electroencephalogr. Clin. Neurophysiol.* 95, 231–251.
- Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
- Green, P.J., Mira, A., 2001. Delayed rejection in reversible jump metropolis hastings. *Biometrika* 88, 1035–1053.
- Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk metropolis algorithm. *Comput. Stat.* 14, 375–395.
- Hämäläinen, M., Ilmoniemi, R., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42.
- Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Phys.* 65, 413–497.
- Huang, M., Aine, C.J., Supek, S., Best, E., Ranken, D., Flynn, E.R., 1998. Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography. *Electroencephalogr. Clin. Neurophysiol.* 108, 32–44.
- Huang, M., Aine, C.J., et al., 2000. Sources on the anterior and posterior banks of the central sulcus identified from magnetic somatosensory evoked responses using multi-start spatio-temporal localization. *Hum. Brain Mapp.* 11, 59–76.
- Huizenga, H.M., de Munck, J.C., Waldorp, L.J., Grasman, R.P.P.P., 2002. Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Trans. Biomed. Eng.* 49, 533–539.
- Jun, S.C., Pearlmutter, B.A., 2005. Fast robust subject-independent magnetoencephalographic source localization using an artificial neural network. *Hum. Brain Mapp.* 24, 21–34.
- Jun, S.C., Pearlmutter, B.A., Nolte, G., 2002. Fast accurate MEG source localization using a multilayer perceptron trained with real brain noise. *Phys. Med. Biol.* 47 (14), 2547–2560.
- Kincaes, W.E., Braun, C., Kaiser, S., et al., 2003. Reconstruction of extended cortical sources for EEG and MEG based on a Monte-Carlo-Markov-Chain estimator. *Hum. Brain Mapp.* 18, 100–110.

- Kuriki, S., Takeuchi, F., Kobayashi, T., 1994. Characteristics of the background fields in multichannel-recorded magnetic field responses. *Electroencephalogr. Clin. Neurophysiol., Evoked potentials* 92, 56–63.
- McCarthy, G., Wood, C.C., Allison, T., 1991. Cortical somatosensory evoked potentials: I. recordings in the monkey *Macaca fascicularis*. *J. Neurophysiol.* 66, 53–63.
- Medvick, P.A., Lewis, P.S., Aine, C., Flynn, E.R., 1989. Monte-Carlo analysis of localization errors in magnetoencephalography. In: Williamson, S.J., Hoke, M., Stroink, G., Kotani, M. (Eds.), *Advances in Biomagnetism*. Plenum, New York, pp. 543–546.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling and localization from spatiotemporal MEG data. *IEEE Trans. Biomed. Eng.* 39 (6), 541–557.
- Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65.
- Phillips, J.W., Leahy, R.M., Mosher, J.C., 1997. MEG-based imaging of focal neuronal current sources. *IEEE Trans. Med. Imag.* 16, 338–348.
- Plis, S.M., George, J.S., Jun, S.C., Pare-Blagoev, J., Ranken, D.M., Schmidt, D.M., Wood, C.C., 2005. Spatiotemporal noise covariance model for MEG/EEG data source analysis. Tech Report: LAUR-043643. URL: http://arxiv.org/find/physics/1/au:+Plis_S/0/1/0/all/0/%1.
- Ranken, D.M., Best, E., Stephen, J.M., Schmidt, D.M., George, J.S., Wood, C.C., Huang, M., 2002. MEG/EEG forward and inverse modeling using mriview. In: Haueisen, J., Nowak, H. (Eds.), *Biomag 2002: 13th International Conference on Biomagnetism*. Jena, Germany, pp. 785–787 (URL: <http://www.lanl.gov/p/p21/mriview.shtml>).
- Robinson, S.E., Vrba, J., et al., 1999. Functional neuroimaging by synthetic aperture magnetometry (SAM). In: Yoshimoto, T., et al., (Eds.), *Recent Advances in Biomagnetism*. Tohoku Univ. Press, Sendai, Japan, pp. 302–305.
- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* 32 (1), 11–22.
- Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Schmidt, D.M., George, J.S., Ranken, D.M., Wood, C.C., 2001. Spatial-temporal Bayesian inference for MEG/EEG. In: Nenonen, J., Ilmoniemi, R.J., Katila, T. (Eds.), *Biomag 2000: 12th International Conference on Biomagnetism*. Espoo, Norway, pp. 671–673.
- Singh, K.D., Harding, G.F.A., 2000. Monte-Carlo analysis and confidence region ellipsoids for equivalent current dipole solutions to EEG/MEG data. In: Aine, C.J., Okada, Y., Stroink, G., Swithenby, S.J., Wood, C.C. (Eds.), *BIOMAG 96: Proceedings of the Tenth International Conference on Biomagnetism*, vols I and II. Springer-Verlag, Santa Fe, USA, pp. 346–349.
- Triantafyllopoulos, K., 2002. On observational variance learning for multivariate Bayesian time series and related models. Ph.D. Dissertation, University of Warwick.
- Uutela, K., Hämäläinen, M., Salmelin, R., 1998a. Global optimization in the localization of neuromagnetic sources. *IEEE Trans. Biomed. Eng.* 45, 716–723.
- Uutela, K., Hämäläinen, M., Salmelin, R., 1998b. Global optimization in the localization of neuromagnetic sources. *IEEE Trans. Biomed. Eng.* 45 (6), 716–723.
- Wood, C.C., Cohen, D., Cuffin, B.N., Allison, T., 1985. Electrical sources in human somatosensory cortex: identification by combined magnetic and potential recordings. *Science* 227, 1051–1053.
- Wood, C.C., Spencer, D.D., Allison, T., McCarthy, G., Williamson, P.D., Goff, W.R., 1988. Localization of human sensorimotor cortex during surgery by cortical surface recordings of somatosensory evoked potentials. *J. Neurosurg.* 68, 99–111.