

Spatiotemporal noise covariance estimation from limited empirical magnetoencephalographic data

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2006 Phys. Med. Biol. 51 5549

(<http://iopscience.iop.org/0031-9155/51/21/011>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 128.231.213.25

The article was downloaded on 30/10/2012 at 21:04

Please note that [terms and conditions apply](#).

Spatiotemporal noise covariance estimation from limited empirical magnetoencephalographic data

Sung C Jun¹, Sergey M Plis^{1,2}, Doug M Ranken¹ and David M Schmidt¹

¹ MS-D454, Applied Modern Physics Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

² Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA

E-mail: jschan@lanl.gov

Received 5 June 2006, in final form 7 August 2006

Published 9 October 2006

Online at stacks.iop.org/PMB/51/5549

Abstract

The performance of parametric magnetoencephalography (MEG) and electroencephalography (EEG) source localization approaches can be degraded by the use of poor background noise covariance estimates. In general, estimation of the noise covariance for spatiotemporal analysis is difficult mainly due to the limited noise information available. Furthermore, its estimation requires a large amount of storage and a one-time but very large (and sometimes intractable) calculation or its inverse. To overcome these difficulties, noise covariance models consisting of one pair or a sum of multi-pairs of Kronecker products of spatial covariance and temporal covariance have been proposed. However, these approaches cannot be applied when the noise information is very limited, i.e., the amount of noise information is less than the degrees of freedom of the noise covariance models. A common example of this is when only averaged noise data are available for a limited prestimulus region (typically at most a few hundred milliseconds duration). For such cases, a diagonal spatiotemporal noise covariance model consisting of sensor variances with no spatial or temporal correlation has been the common choice for spatiotemporal analysis. In this work, we propose a different noise covariance model which consists of diagonal spatial noise covariance and Toeplitz temporal noise covariance. It can easily be estimated from limited noise information, and no time-consuming optimization and data-processing are required. Thus, it can be used as an alternative choice when one-pair or multi-pair noise covariance models cannot be estimated due to lack of noise information. To verify its capability we used Bayesian inference dipole analysis and a number of simulated and empirical datasets. We compared this covariance model with other existing covariance models such as conventional diagonal covariance, one-pair and multi-pair noise covariance models, when noise information is sufficient to estimate them. We found that our proposed noise covariance model yields better localization performance than a diagonal noise covariance, while it performs slightly worse than one-pair or multi-pair noise covariance

models—although these require much more noise information. Finally, we present some localization results on median nerve stimulus empirical MEG data for our proposed noise covariance model.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Brain imaging technologies have been developing for nearly a century. About a decade ago, MEG/EEG technologies emerged commercially and opened a new era for brain research. Unlike the widely used functional magnetic resonance imaging (fMRI), they can detect neural electrical current directly and map brain temporal functional activities at up to a fraction of a millisecond resolution. The source localization problem, i.e., the inverse problem of reconstructing the neural electrical current that produced the MEG/EEG measurements, is inherently illposed³ and confounds the spatial resolution of MEG/EEG systems.

Since the inception of commercial MEG/EEG systems, a number of MEG/EEG source localization methods have been developed and commercialized; see Hämäläinen *et al* (1993). Most methods reconstruct optimized current distributions or dipole sources under certain constraints. These methods can be categorized into parametric (Hämäläinen *et al* 1993, Mosher *et al* 1992, Huang *et al* 1998, Uutela *et al* 1998, Schmidt *et al* 1999, Jun *et al* 2002) and non-parametric approaches (Pantazis *et al* 2005, Darvas *et al* 2005). Parametric approaches aim to find the optimal solution or solutions (in some sense) which fit well both the measurements and any other prior information. Non-parametric approaches aim to estimate source/sources by scanning source space and detecting statistically significant source configurations.

Due to the central limit theorem, the noise of averaged data is expected to be Gaussian distributed and to be parameterized with a noise covariance matrix. This Gaussian noise model is a basic component of most statistically based inverse approaches and the proper use of its noise covariance is of great importance. It has been reported that noise covariance estimation has a significant effect on localization performance in both spatial-only analyses (Jun *et al* 2002) and spatiotemporal analyses (De Munck *et al* 2002, Plis *et al* 2006).

The most widely used noise covariance model consists of sensor variances with no spatial or temporal correlation. This is a very simple model compared to the most general form for the spatial–temporal noise covariance matrix. The most general form has too many parameters to estimate in practice, while the simple and most commonly used noise model ignores the correlation that is present in the background noise. Recently, Huizenga *et al* (2002), De Munck *et al* (2002), Plis *et al* (2006) proposed new noise covariance models that use a pair (or sum of pairs) of Kronecker products of spatial covariances and temporal covariances under the assumption that spatial and temporal noise structures are separable. These models have many fewer free parameters than the most general covariance models and may be estimated by prestimulus regions of typical non-averaged, multiple-epoch evoked response data. Nevertheless, they still require more noise data than are usually available in the prestimulus regions of averaged evoked response data.

In this work, we propose a different noise covariance model that can be estimated from the prestimulus regions of commonly available averaged evoked response data, using simple statistics without the need for a more complex and time-consuming procedure-like optimization. Our proposed noise model's capability was tested on a number of simulated

³ Many different solutions exist from the same set of measurements.

datasets with a spatiotemporal, multi-dipole analysis (Jun *et al* 2005). These simulated data were generated by adding simulated dipolar sources to empirical noise data. The performance of our proposed noise model was compared with that of other existing noise covariance models, including the diagonal model, one-pair and multi-pair Kronecker products models. Finally we present and compare localization results from the multiple noise models using empirical median nerve stimulus data.

2. Noise covariance models

Due to the central limit theorem of statistics, it is reasonable to expect that brain noise of averaged data is, in general, Gaussian distributed with a covariance explaining its noise structure. For parametric source analysis approaches, it is common to optimize how well a given source model of neural current fits the measurements using a log-likelihood function. For Gaussian, zero-mean averaged background noise, the log-likelihood function is as follows:

$$-\frac{1}{2} \sum_{ktk't'} \left[B_{kt} - \int \mathcal{L}_k(x) J(x, t) dx \right] \mathbf{COV}_{kt;k't'}^{-1} \left[B_{k't'} - \int \mathcal{L}_{k'}(x') J(x', t') dx' \right]. \quad (1)$$

Here B_{kt} are the measurements (the data being analysed) at channel k and time t ; $J(x, t)$ is the neural current source at location x and time t ; $\mathcal{L}_k(x)$ is the linear operator (lead field) projecting source space (source at x) into sensor space (channel k). \mathbf{COV} is the sample covariance of the averaged background noise, which is of our ultimate interest to estimate.

2.1. Diagonal model

The simplest noise covariance model is the diagonal model. This model assumes that the noise is uncorrelated, and is parameterized by diagonal elements consisting of sensor variances. This model is easily calculated even when noise information is very limited and is the most commonly used noise model.

2.2. Kronecker product models

Based on the assumption that temporal (\mathbf{T}) and spatial (\mathbf{S}) covariances of background noise are separable and independent, a single pair Kronecker product model was developed with its parameters estimated through a maximum likelihood method (De Munck *et al* 2002):

$$\mathbf{COV}_{\text{one-pair}} = \mathbf{T} \otimes \mathbf{S}. \quad (2)$$

Here \otimes denotes the Kronecker product. In addition, Huizenga *et al* (2002) described a simpler one-pair model consisting of parameterized spatial covariance and Toeplitz temporal covariance whose parameters were determined through minimizing the difference between an estimated full noise covariance (sample noise covariance) and the given model.

To better estimate and capture noise structure, a model consisting of a sum of Kronecker products was introduced (Plis *et al* 2006):

$$\mathbf{COV}_{\text{multi-pair}} = \sum_{l=1}^L \mathbf{T}_l \otimes \mathbf{S}_l. \quad (3)$$

In this model spatial components \mathbf{S}^l of rank 1 are paired with their corresponding full temporal covariance matrices \mathbf{T}^l . In Plis *et al* (2006), this model and its inversion were estimated as

follows:

$$\left(\sum_{l=1}^L \hat{\mathbf{T}}_l \otimes \hat{\mathbf{S}}_l \right)^{-1} = \sum_{l=1}^L \hat{\mathbf{T}}_l^{-1} \otimes \hat{\mathbf{S}}_l,$$

$$\hat{\mathbf{S}}_l = v_l v_l', \quad \hat{\mathbf{T}}_l = \lambda_l^2 \frac{1}{M(M-1)} \sum_{m=1}^M (u_{m,l} - \bar{u}_l)(u_{m,l} - \bar{u}_l)'$$

Here $\hat{\mathbf{T}}_l$ and $\hat{\mathbf{S}}_l$ are estimators of the model \mathbf{T}_l and \mathbf{S}_l . v_l is the l th column vector of the orthogonal matrix \mathbf{V} estimated by singular value decomposition (SVD)⁴. We remark that the spatial components $\hat{\mathbf{S}}_l$ are orthogonal. This orthogonality ensures that the inversion of this covariance model is practical.

These one-pair and multi-pair Kronecker product models tremendously reduce the degree of freedom (DOF) of spatiotemporal noise covariance, thereby enabling one to estimate noise covariance in an efficient way as long as reasonably sufficient noise information is available⁵. However, these models are inadequate when available noise information is less than the noise model's DOFs or even slightly more than these DOFs. This is the case, for example when only averaged noise (prestimulus) information for a limited time duration (about a few hundred milliseconds) is available. Furthermore, these models are difficult to apply to real-time analysis due to their required preprocessing time.

2.3. Our proposed noise covariance model and its estimation method

Our primary interest is to show an improvement on the estimation of noise structure while using very limited noise information compared to that achieved with the widely used diagonal covariance. If the total available noise information is extremely limited, i.e., comparable to the DOF of diagonal noise covariance, the noise covariance estimation cannot be improved further. However, we have room to develop a better noise covariance estimation if the total available noise information is much greater than that used by the DOF of diagonal noise covariance but it is not enough to estimate Kronecker product noise models. This is the situation we focus on presently.

Our proposed model retains the one-pair Kronecker product model and constrains it into a diagonal spatial covariance matrix and a Toeplitz temporal covariance matrix:

$$\mathbf{COV}_{\text{one-pair-limited-data}} = \mathbf{T}_{\text{toeplitz}} \otimes \mathbf{S}_{\text{diag}}. \quad (5)$$

⁴ Assuming that we have M times single trial noise data (T (number of time points) $\times L$ (number of sensors) matrix), $\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_M\}$, singular value decomposition is applied to all stacked background noise data \mathbf{A} :

$$\mathbf{A} = \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_m \\ \vdots \\ \mathbf{E}_M \end{pmatrix} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_m \\ \vdots \\ \mathbf{U}_M \end{pmatrix} \mathbf{\Sigma} \mathbf{V}', \quad (4)$$

where \mathbf{V} is an $L \times L$ orthogonal matrix consisting of spatial column vectors $\{v_1, v_2, \dots, v_L\}$, the $T \times L$ matrices \mathbf{U}_i form an $MT \times L$ orthogonal matrix \mathbf{U} , and $\mathbf{\Sigma}$ is an $L \times L$ diagonal matrix with diagonal elements $\{\lambda_1, \lambda_2, \dots, \lambda_L\}$ being singular values of \mathbf{A} . Here $u_{m,l}$ and \bar{u}_l denote l th column vectors of matrices \mathbf{U}_m and $\bar{\mathbf{U}}$, which is an averaged matrix of \mathbf{U}_m over $m = 1, \dots, M$. In this derivation, cross covariance between orthogonal spatial component and time point was assumed to be negligibly small. For more detail, refer to Plis *et al* (2006).

⁵ In general, single trial noise collection without any stimuli for a certain time is required.

It was reported in Bijma *et al* (2003) that spatiotemporal empirical noise for MEG/EEG has temporal stationarity. This allows the temporal noise covariance to be modelled as a Toeplitz matrix. Interestingly, this model amounts to a special case (the parameters β in Huizenga's work, determining spatial correlations, is set to zero) of the model proposed by Huizenga *et al* (2002). Huizenga's model should be determined through minimizing the difference between the model and the estimated full spatiotemporal noise covariance. It requires a great deal of noise information for estimating the full noise covariance as well as a time-consuming (in general) optimization procedure.

Here we propose an alternative estimation of the one-pair Kronecker product model (5) for limited data. If collected data of continuous spatiotemporal noise information in a matrix \mathbf{n} of $L \times M$

$$\mathbf{n} = \{n(i, j) | 1 \leq i (\text{channel}) \leq L, 1 \leq j (\text{time sample}) \leq M\} \quad (6)$$

is given, and an $L \times L$ spatial covariance matrix estimator $\hat{\mathbf{S}}_{\text{diag}}$ and $P \times P$ ($P \leq M$) temporal covariance matrix estimator $\hat{\mathbf{T}}_{\text{toeplitz}}$ should be estimated, they are simply calculated by

$$\hat{s}(i, i) = \left(\sum_{j=1}^M (n(i, j) - \bar{n}_{i,\cdot})^2 \right) / M \quad (\text{sensor variance}) \quad (7)$$

$$\hat{t}(i, j) = \left(\sum_{l=1}^L a_{\text{corl}}(i - j) \right) / L \quad (\text{averaged autocorrelation}) \quad (8)$$

$$a_{\text{corl}}(p) = \begin{cases} \frac{\sum_{m=1}^{M-|p|} (n(l, m) - \bar{n}_{l,\cdot})(n(l, m + |p|) - \bar{n}_{l,\cdot})}{\sum_{m=1}^M (n(l, m) - \bar{n}_{l,\cdot})^2}, & |p| < M \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$\bar{n}_{i,\cdot} = \left(\sum_{j=1}^M n(i, j) \right) / M. \quad (10)$$

Here $\hat{s}(i, j)$ and $\hat{t}(i, j)$ are elements of matrices $\hat{\mathbf{S}}_{\text{diag}}$ and $\hat{\mathbf{T}}_{\text{toeplitz}}$, respectively.

This estimation procedure comes from physical intuition rather than optimization procedure. The optimized estimators based on the model (5) could be sought, but we experienced that lack of enough noise information could cause the divergence of an optimization procedure. Our intention is to develop a reasonable (it may not optimal, but should be efficient in terms of computation speed and accuracy) noise covariance. Thus the above estimator (7)–(10) is good enough in a sense that it is calculated in real time from simple statistics and it captures more temporal noise structure than a diagonal noise covariance. Studying how our estimator can be related to the optimized estimators is another interesting topic that can be investigated.

This estimation requires more time samples than the dimension of the temporal covariance matrix in order to get reasonable statistics. In section 4, we will discuss the cases in which the number of time samples is less. We remark that our proposed noise covariance estimation can be computed in real-time and thus could be applicable for real-time single trial analysis under the assumption of Gaussianity of single trial noise.

3. Experiments

The main reason one seeks to develop good noise covariance models and their estimators is to improve source localization performance. In this section, we apply our proposed noise covariance for both simulated and empirical MEG data, and investigate how our proposed noise covariance model could improve localization performance. As a localization method, we use the spatiotemporal Bayesian inference dipole analysis (Jun *et al* 2005, 2006) which was recently developed by our team and is briefly outlined below.

3.1. Bayesian inference dipole analysis

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, and begins with the standard Bayes' rule of probability for known information \mathbf{B} and unknown information θ :

$$P(\theta|\mathbf{B}) \propto P(\mathbf{B}|\theta)P(\theta). \quad (11)$$

Here $P(\theta|\mathbf{B})$, $P(\mathbf{B}|\theta)$ and $P(\theta)$ mean conditional posterior of θ for given \mathbf{B} , conditional probability of \mathbf{B} for given θ (likelihood distribution), and prior information of θ , respectively. This enables one to combine any additional information straightforwardly through the Bayes' rule. The posterior distribution $P(\theta|\mathbf{B})$ contains the information—how well a state θ can explain the measurements \mathbf{B} in the given physical model. Commonly, the obtainable posterior distribution is numerically sampled using Markov Chain Monte Carlo (MCMC) techniques (Jun *et al* 2005, 2006, Chen *et al* 2000, Gilks *et al* 1995).

Formulation. Assuming that the sources of neuromagnetic fields are localized, we employ a fixed dipole source model with a variable number (from 0 to some maximum) of current dipoles where the dipole locations and orientations are fixed over time. We used a spherical head model and the Sarvas forward model (Sarvas 1987) in this work. Considering this configuration, θ and its prior distribution in (11) can be given by

$$\begin{aligned} \theta &= \{N, \mathbf{X}, \mathbf{O}, \mathbf{J}, \bar{\mathbf{t}}, \mathbf{C}\} \\ P(\theta) &= P(\mathbf{O}|\mathbf{X}, N)P(\mathbf{J}|N)P(\mathbf{X}|N)P(\bar{\mathbf{t}}|N)P(\mathbf{C})P(N). \end{aligned} \quad (12)$$

Here \mathbf{B} , N , \mathbf{X} , \mathbf{J} , \mathbf{O} , $\bar{\mathbf{t}}$ and \mathbf{C} mean a spatiotemporal measurement, the number of sources, a location matrix, a current time course matrix, an orientation matrix, active time range information, and a noise covariance matrix, respectively.

We use uniform priors for the orientation \mathbf{O} , the location \mathbf{X} , the active time range $\bar{\mathbf{t}}$ and the number of sources N . Particularly, we use the prior of covariance \mathbf{C} as the k th-order inverse Wishart distribution with mean \mathbf{COV} . \mathbf{COV} would be estimated here as our proposed noise covariance model. k is $2 \dim(\mathbf{COV}) + 1$. Prior for current time course \mathbf{J} is chosen as a Gaussian distribution $N(\mathbf{0}, \mathbf{C})$. A covariance \mathbf{C} was given as the temporal correlation matrix of one time point with another, which allows us to include the temporal correlation at nearby latencies.

Assuming a Gaussian noise model of mean zero and noise covariance matrix \mathbf{C} , the likelihood $P(\mathbf{B}|\theta)$ can be described:

$$P(\mathbf{B}|\theta) \propto \frac{1}{|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\bar{\mathbf{B}} - \bar{\mathbf{B}}_c)' \mathbf{C}^{-1} (\bar{\mathbf{B}} - \bar{\mathbf{B}}_c)\right), \quad (13)$$

where $\bar{\mathbf{B}}_c$ is a calculated measurement through the forward model and is depending on the unknown parameters. Finally, combining the likelihood distribution as well as all prior

distributions yields the following posterior distribution:

$$P(\theta|\mathbf{B}) \propto \frac{|\mathbf{COV}|^{(k-1)/4} |\mathbf{C}|^{-(k+1)/2}}{|\bar{\mathbf{C}}|^{1/2}} \times \exp\left(-\frac{1}{2} \left[\bar{\mathbf{B}}_n' \mathbf{C}^{-1} \bar{\mathbf{B}}_n + \bar{\mathbf{J}}' \bar{\mathbf{C}}^{-1} \bar{\mathbf{J}} + \text{Tr}\left(\frac{k-1}{2} [\mathbf{COV}] \mathbf{C}^{-1}\right) \right]\right),$$

where $\bar{\mathbf{B}}_n = \bar{\mathbf{B}} - \bar{\mathbf{B}}_c$. $\bar{\mathbf{A}}$ means the vector stacking all column vectors of the given matrix \mathbf{A} . Tr , $|\cdot|$ and $'$ denote trace, determinant and transposition, respectively.

In order to lessen local minima problems as well as to compute the posterior distribution efficiently, the posterior can be simplified by eliminating insignificant eigenvalues of temporal correlation matrix $\bar{\mathbf{C}}$ and by marginalizing the posterior over both current time course $\bar{\mathbf{J}}$ and noise covariance \mathbf{C} (Jun *et al* 2005, 2006).

Sampling. Next, we produce a sampling of many likely solutions from the obtained posterior distribution. We use Markov Chain Monte Carlo (MCMC) to sample the posterior probability distribution $P(\theta|\mathbf{B})$ on the parameter space. All MCMC methods are designed to construct a Markov chain $(\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots)$ and to choose the transition probabilities $P_T(\theta^{(p+1)}|\theta^{(p)})$ in such a way that the probability distribution of the p th realization converges to targeted distribution as p goes to infinity. After discarding samples during a burn-in period (MCMC is usually required to do a convergence process for a while after its initialization), drawing realizations of the Markov chain gives us a random sample of the probability distribution. In particular in order to assure convergence of MCMC, it is sufficient that the following ‘detailed-balance’ condition is satisfied:

$$P(\theta^{(p)}|\mathbf{B}) P_T(\theta^{(p+1)}|\theta^{(p)}) = P(\theta^{(p+1)}|\mathbf{B}) P_T(\theta^{(p)}|\theta^{(p+1)}). \quad (14)$$

We use the reversible jump (RJ) MCMC technique (Green 1995) allowing movement between different parameter spaces and satisfying the detailed balance condition. In our RJ-MCMC procedure, a candidate sample (θ^*) is chosen from two categorized proposal distributions.

- Trans-dimensional proposal
 - birth move: a new dipole and its parameters are proposed;
 - 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;
 - orientation update move: a dipole is randomly chosen and its new orientation is proposed;
 - active time range update move: a dipole is randomly chosen and its new active time range is proposed.

For more details, refer to (Jun *et al* 2005, 2006).

3.2. Comparative localization performance for various noise covariances

We collected ten kinds of empirical noise datasets from a 4D Neuroimaging 122 Neuromag gradiometer system (four kinds from two different experimental paradigms) and a VSM CTF 275-Channel MEG system (six kinds from three different experimental paradigms). More detailed information on empirical noise datasets are described in the appendix. For each empirical noise dataset, we generated a total of 20 kinds of three-dipole problems as follows.

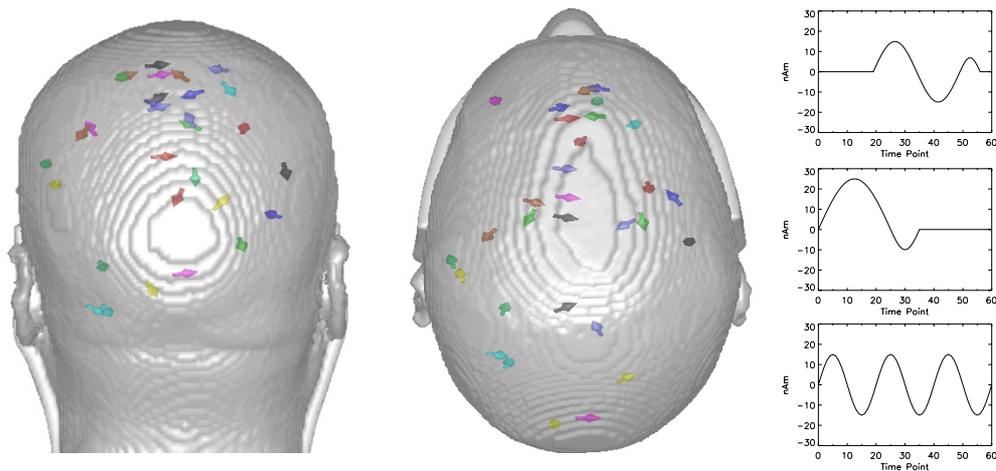


Figure 1. Left and middle: ten configurations for the three-dipole sources problem, which were randomly chosen to generate ten simulated MEG data for one empirical datum (MN1-NEURO122-LEFT). The same coloured three dipoles denote one {three-dipole sources problem}. Right: time courses for each dipole source.

- D1. We randomly chose ten kinds of three dipole locations on the cortex.
- D2. We chose corresponding tangential orientations (the tangential part of voxel orientation at the dipole location) of chosen dipoles.
- D3. We generated three kinds of time courses, each of which corresponds with one of three dipoles. In other words, we used the same three time courses for all three-dipole problems.
- D4. We calculated the corresponding spatiotemporal measurement through the forward model (spherical head model, dipole current model) for each three-dipole problem.
- D5. We extracted two different spatiotemporal empirical noise realizations (60 time sample points) from the given empirical noise dataset.
- D6. We added each empirical noise realization into the calculated measurement, and finally generated two kinds of simulated data for each three-dipole problem.

Figure 1 illustrates two configurations (red and blue) of the three-dipole sources problem.

Regarding noise covariance, for each empirical noise dataset we estimated up to four kinds of noise covariances based on different models, their estimation methods and noise data availability:

- conventional diagonal estimation (DIAG);
- one-pair Kronecker product through maximum likelihood method (ONE-ML);
- multi-pair Kronecker product by singular value decomposition (MUL-SVD);
- our proposed one-pair Kronecker product by simple statistics (ONE-TOEP).

Among ten kinds of empirical datasets (four from Neuromag 122 and six from CTF 275), both single trial noise datasets and averaged noise datasets for Neuromag 122 are available, but single trial datasets for CTF 275 are not available. While DIAG and ONE-TOEP could be estimated for all ten kinds of datasets, ONE-ML and MUL-SVD could not be estimated for CTF 275 datasets due to this limited data availability. As a matter of fact, such cases motivated us to develop our proposed noise covariance model. For comparison purposes, DIAG and ONE-TOEP were estimated from only averaged noise datasets for all cases. We believe

that DIAG and ONE-TOEP estimations from single trial noise datasets would yield better performance than those from averaged noise datasets because single trial datasets include more noise structure information than averaged datasets. We note that for the Neuromag 122 system, roughly 1200 spatiotemporal samples with 60 time sample points were obtained to use for ONE-ML and MUL-SVD and continuous (up to 100 time points) averaged prestimulus data were obtained to use for DIAG and ONE-TOEP. For the CTF 275 system, continuous (up to 200 time points) averaged prestimulus time sample points were obtained.

The Bayesian inference dipole analysis is allowed to vary the number of dipole sources, but for our purpose here we fixed the maximum number of dipole sources at three (the number of the exact dipole sources) in order to be a representative of other multi-dipole analyses where it is common practice to fix the number of dipole sources. The procedure used to collect interesting (high posterior probability) samples is as follows:

- S1. For each simulated datum we ran two MCMCs with different random seeds. Each run consisted of 50 000 iterations, of which 5000 samples were collected (one sample randomly chosen among ten iterations).
- S2. We discarded the first 3000 samples as a burn-in period and stored the remaining 2000 samples for analysis.
- S3. Finally, we collected a total of 4000 samples (2000 samples from each run) for each problem. Among these 4000 samples we chose 100 samples with the highest posterior probability and calculated the average location and time course from these. This approach was taken in order to avoid local minima effects and to produce a good estimate of the most likely solution.

Localization performance was checked by comparing source location error and time course error. Source location error per dipole (LOC-ERR) and time course error per dipole per time point (TC-ERR) were computed as follows:

$$\text{LOC-ERR} = \left[\sum_{k=1}^3 \text{distance}(\mathbf{X}_{k,\text{avg}}, \mathbf{X}_{k,\text{exact}}) \right] / 3 \quad (15)$$

$$\text{TC-ERR} = \left[\sum_{k=1}^3 \sqrt{\sum_{t=1}^{60} (J_{k,\text{avg}}(t) - J_{k,\text{exact}}(t))^2} \right] / 60/3. \quad (16)$$

Here $\mathbf{X}_{\cdot,\text{avg}}$ and $\mathbf{X}_{\cdot,\text{exact}}$ denote the averaged source location vector of collected samples and the exact source location vector, respectively. $J_{\cdot,\text{avg}}$ and $J_{\cdot,\text{exact}}$ denote averaged time course of collected samples and the exact time course, respectively. LOC-ERR and CUR-ERR mean averaged distance from the exact source location over three dipole sources and averaged root square error over three dipole sources and the whole time window.

Tables 1 and 2 show the comparative performances over four kinds of noise covariance models. A total of 20 three-dipole problems for each empirical noise dataset were tested and averaged. As expected, DIAG showed the worst performance for all of the different empirical datasets while MUL-SVD showed the best performance. Our proposed method, ONE-TOEP was between DIAG and ONE-ML, thus it was superior to DIAG and was inferior to ONE-ML and MUL-SVD in most cases.

To see how the models performed over various signal-to-noise ratio (SNR) data, we chose 16 three-dipole problems among the previously generated 40 problems for two empirical datasets (MN1-NEURO122-LEFT and MN1-NEURO122-RIGHT). For each problem we regenerated five kinds of problems with different SNRs (4.0, 2.0, 1.0, 0.8, 0.7) by controlling

Table 1. The location error (LOC-ERR) comparison over four kinds of noise covariances. A total of 20 simulated data were tested and averaged. All units are in mm. ‘–’ denotes unavailability of noise covariance. The first four datasets were acquired from the 4D Neuroimaging Neuromag 122 system during (left and right hand) median nerve stimulation experiments for each different subject. The next six datasets were collected from the CTF 275 system during index finger (left and right) stimulation experiments (fifth, sixth, ninth and tenth) for the same subject for each different day, and a median nerve stimulation experiment (seventh and eighth).

Empirical data/covariance	DIAG	ONE-ML	MUL-SVD	ONE-TOEP
MN1-NEURO122-LEFT	1.700	1.052	1.040	1.193
MN1-NEURO122-RIGHT	1.957	1.540	1.281	1.687
MN2-NEURO122-LEFT	2.060	1.431	1.343	1.640
MN2-NEURO122-RIGHT	1.714	1.367	1.070	1.407
IF1-CTF275-LEFT	1.465	–	–	1.109
IF1-CTF275-RIGHT	1.558	–	–	1.069
MN1-CTF275-LEFT	2.723	–	–	1.353
MN1-CTF275-RIGHT	1.857	–	–	1.003
IF2-CTF275-LEFT	1.451	–	–	0.971
IF2-CTF275-RIGHT	1.573	–	–	1.077

Table 2. The time course error (TC-ERR) comparison over four kinds of noise covariances. A total of 20 simulated data were tested and averaged. All units are in nA m. ‘–’ denotes unavailability of noise covariance. Data are described as in table 1.

Empirical data/covariance	DIAG	ONE-ML	MUL-SVD	ONE-TOEP
MN1-NEURO122-LEFT	0.126	0.100	0.090	0.101
MN1-NEURO122-RIGHT	0.216	0.122	0.112	0.145
MN2-NEURO122-LEFT	0.139	0.106	0.095	0.111
MN2-NEURO122-RIGHT	0.111	0.094	0.080	0.092
IF1-CTF275-LEFT	0.139	–	–	0.109
IF1-CTF275-RIGHT	0.153	–	–	0.112
MN1-CTF275-LEFT	0.225	–	–	0.180
MN1-CTF275-RIGHT	0.190	–	–	0.134
IF2-CTF275-LEFT	0.165	–	–	0.125
IF2-CTF275-RIGHT	0.140	–	–	0.154

peak magnitudes of the three current time courses together (the same fraction is multiplied into the original time course). Here, we calculate SNR using

$$\text{SNR} = \frac{\sqrt{\sum_{k,t} (\text{signal at channel } k, \text{ time } t)^2}}{\sqrt{\sum_{k,t} (\text{noise at channel } k, \text{ time } t)^2}}. \quad (17)$$

As with the previous experiment, we used the same procedure to get interesting samples. Figure 2 shows localization performance over various SNRs for four kinds of noise covariances. Each point is an averaged result over well-localized⁶ problems. For LOC-ERR, our proposed covariance method (ONE-TOEP) consistently yielded the intermediate performance between ONE-ML and DIAG, while it was almost comparable to ONE-ML and MUL-SVD for TC-ERR. This simulation study shows that compared to DIAG, our proposed covariance model

⁶ As SNR grows smaller, i.e., spatiotemporal signal is dominated by noise, we often saw weakest sources undetected for ONE-ML and MUL-SVD covariances. On averaging, we excluded undetected cases.

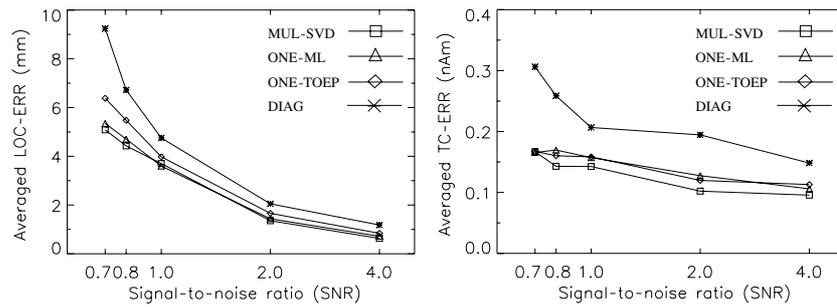


Figure 2. Location error (LOC-ERR) and current time course error (TC-ERR) as SNR varies.

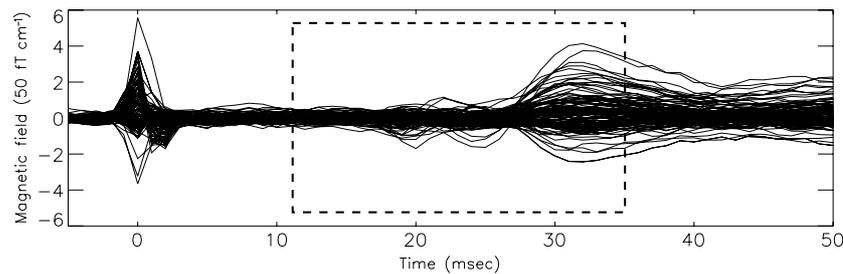


Figure 3. 122 channel overplotted empirical MEG data for left hand median nerve stimulation. MEG data in dashed box were analysed.

ONE-TOEP can improve localization performance by up to 3 mm (per dipole) in location accuracy and up to about 0.15 nA m (per time point per dipole) in time course accuracy.

3.3. Empirical experiments

In this experiment, we applied our proposed noise covariance model (ONE-TOEP), conventional diagonal noise covariance (DIAG), and multi-pair Kronecker product model (MUL-SVD) to empirical MEG data for comparison. MN1-NEURO122-LEFT empirical data were used for our purpose. In particular, we analysed left hand median nerve stimulation data (25 ms window 11 ms after stimulus onset), which is illustrated in figure 3. As with the simulations in the previous subsection, we did MCMC runs with different random seeds and obtained consistent results. Initially, we allowed the number of dipoles to vary up to a maximum of six, on the assumption that the true model order would be between one and six. For ONE-TOEP and MUL-SVD, two dipole sources were consistently seen⁷, while five sources showed up for the DIAG analyses. Looking at source locations and time course shapes in the DIAG results, two sources correspond to the sources found in the ONE-TOEP and MUL-SVD results. The extra three sources in the DIAG analysis are most likely spurious sources modelling correlated noise in the data. Based on the results, the reasonable model order may be two.

For our next analyses, we restricted the number of allowable dipole sources to two and ran MCMCs for three covariances. Interestingly, for the DIAG covariance model, the two source clusters localize differently from those for the ONE-TOEP and MUL-SVD models, as shown in figure 4. The blue clusters for all cases are at the same location, even though

⁷ Two interesting sources on median nerve stimulation empirical data were recently reported in Huang *et al* (2000).

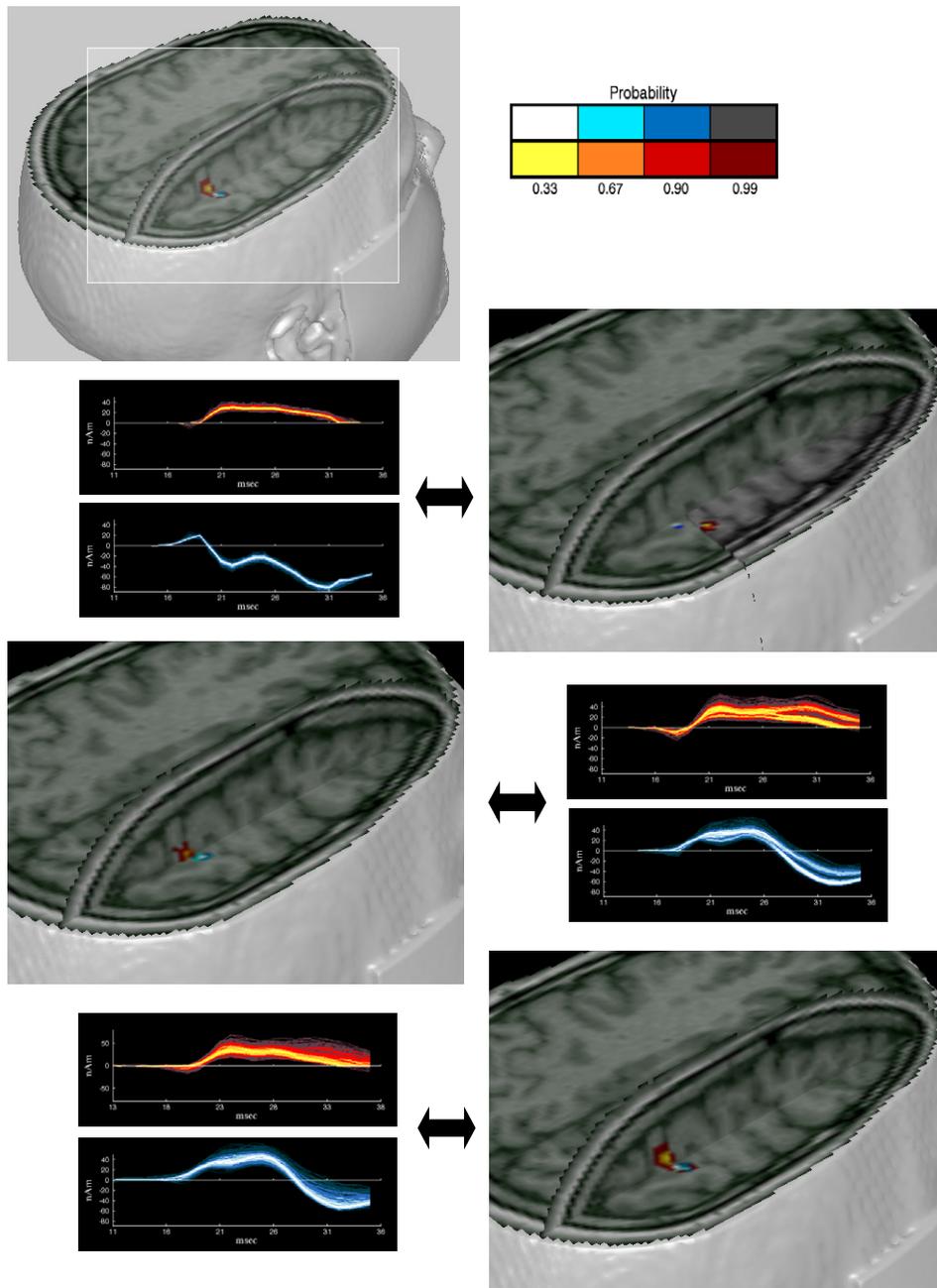


Figure 4. Left hand median nerve stimulation experiment for analysis using DIAG, ONE-TOEP and MUL-SVD noise covariance models. Overlaid source distribution MRVIEW plots and their corresponding time course histogram plots, respectively. Top: figure (left) showing the orientation of the whole head. The highlighted rectangle in this figure shows the view area for the source location figures. Confidence level colour bar (right). Second row: DIAG. Third row: ONE-TOEP. Fourth row: MUL-SVD. In the second row, the top time course histogram plot on the left is for the outer source distribution blob on the right and the bottom one is for the inner source distribution blob. However, in the third and fourth rows the top time course histogram plots are for inner source distribution blobs and the bottom ones are for outer source distribution blobs.

the time course histogram plot for the DIAG analysis is very different from those for the ONE-TOEP and MUL-SVD analyses. The red cluster for the DIAG case is clustered more densely and is more superficial than the blue cluster, while those for the ONE-TOEP and MUL-SVD are located slightly inward and above the associated blue clusters. As a whole, DIAG yielded a different result from ONE-TOEP and MUL-SVD, whose results are consistent with those reported previously (Allison *et al* 1991a, 1991b, Wood *et al* 1985, 1988, Huang *et al* 2000, Jun *et al* 2005).

Our results suggest that our Toeplitz-based noise covariance model is more effective in modelling noise than a diagonal covariance model, when used in analyses of empirical MEG data.

4. Discussion

In general, most parametric approaches using noise covariance estimation require a predetermined number of sources, which can be roughly estimated through a singular value decomposition of the MEG measurements. One advantage of the Bayesian inference dipole analysis used in this work is that it does not require a preset number of sources. As reported in section 3.3, when the number of sources is allowed to be higher than the expected or exact number of sources, the DIAG covariance model produced more sources than expected for most simulated and empirical MEG data. Due to inadequate noise structure modelling in the DIAG case, these extra sources are most likely modelling noise in the data. With our proposed ONE-TOEP we rarely see these extra sources.

As seen in equation (9), our proposed ONE-TOEP model can be easily computed from continuous spatiotemporal noise data. Temporally discontinuous noise data can be used to estimate sensor variances, but its usefulness in estimating the temporal covariance matrix would depend on characteristics of the temporal discontinuities. We would like to investigate how more general noise information can be used in our proposed covariance estimation method. For less or even far less temporal noise information than the dimension of the temporal covariance matrix, the noise covariance can be estimated roughly by setting the temporal covariance to zero for more delayed time points than the available time points in the noise information. However, one simulation study (not shown) indicates that very rough estimation of the temporal covariance matrix is likely to reduce localization performance, making the conventional diagonal covariance-based analysis a better choice.

Our proposed model consists of a diagonal matrix and a Toeplitz matrix. Its invertibility relies on the invertibility of each matrix. Even though we cannot assure the invertibility of the estimated Toeplitz matrix, we have seen empirically that the Toeplitz temporal covariance matrix is non-singular. Regarding simulation studies, we tested our noise covariance model for three-dipole problems. Single-dipole problems and two-dipole problems were tested, but we could not see any significant differences. Four or more dipole problems would yield a more complex posterior distribution, and could cause even small errors in the noise modelling to severely degrade localization performance. Therefore, we expect our covariance model to be more advantageous for such cases than a conventional diagonal covariance model.

5. Conclusion

We proposed a different spatiotemporal noise covariance model consisting of a Kronecker product between a diagonal spatial matrix and a Toeplitz temporal covariance matrix. We also proposed a simple estimation method for this noise covariance model using intrinsic statistics

of the noise information. This estimation is simple to implement and can be computed in real time. We have shown through a simulation study that the proposed Toeplitz-based model is a better choice in terms of localization performance than the conventional diagonal noise covariance model, and can be used in the cases where noise information is not sufficiently available to use more accurate noise models. In inverse analyses of empirical MEG data, the proposed Toeplitz-based model outperformed a conventional diagonal covariance model.

Appendix

We describe detailed experimental paradigms of empirical datasets acquired from the Neuromag 122 system and the CTF 275 system.

- Neuromag 122 system

- MN1-NEURO122-LEFT and MN1-NEURO122-RIGHT

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 (interstimulus interval). Data were digitized at 1 kHz with the online filters set to 0.03–330 Hz. An interval of 0.1 s prestimulus and 0.5 s poststimulus was collected. Data were $(1 - \text{median})^8$ filtered to remove low frequency drifts but were not filtered for 60 Hz noise and its harmonics because their effects were negligibly small.

- MN2-NEURO122-LEFT and MN2-NEURO122-RIGHT

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 ten blocks for each of eight 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 was randomized between 0.25 and 0.75 s. Data were collected with 1000 Hz sampling from a male subject, age 38. One of 122 channels was discarded due to it malfunctioning. Data were $(1 - \text{median})$ filtered. 60 Hz noise and its harmonics were filtered out by removal of peaks in the spectrum and interpolation between adjacent spectrum points.

- CTF 275 system

All data were collected with 2400 Hz sampling from the same female, age 30. The ISI for all datasets is 1.5 s on average, but varies between 1.25 and 1.75 s. Noise cancellation procedure was off.

- IF1-CTF275-LEFT and IF1-CTF275-RIGHT

Index finger stimulation was applied. 2 μ s pulses were generated with a Grass S88 stimulator.

- MN1-CTF275-LEFT and MN1-CTF275-RIGHT

The median nerve stimulation was applied. 2 μ s pulses were generated with a Grass S88 stimulator.

- IF2-CTF275-LEFT and IF2-CTF275-RIGHT

The same experiment procedure as IF1-CTF275-LEFT and IF1-CTF275-RIGHT was applied on a different day.

⁸ Median filtered data were subtracted from unfiltered original data.

Acknowledgments

This work was supported by NIH grant 2 R01 EB000310-07 and the Mental Illness and Neuroscience Discovery (MIND) Institute. For the first two 4D Neuroimaging Neuromag 122 MEG empirical datasets 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 also thank Cheryl Aine at the MIND Institute for giving access to six VSM CTF 275-Channel MEG empirical datasets as well as the two 4D Neuroimaging Neuromag 122 MEG empirical datasets. Specifically, six VSM CTF 275 MEG empirical datasets were obtained through a MEG consortium project (DOE grant DE-FG02-99ER62764 to the MIND institute) between MIND Institute partners (MIND/UNM, UMN/VA, MGH/Harvard and LANL). Some figures showing the results of the Bayesian dipole analysis were generated using MRVIEW (Ranken *et al* 2002). Finally, we are thankful to Juliana Paré-Blagoev at the MIND Institute for helping us to improve our manuscript.

References

- Allison T, McCarthy G, Wood C C and 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** 2465–503
- Allison T, Wood C C and 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
- Bijma F, de Munck J C, Huizenga H M and Heethaar R M 2003 A mathematical approach to the temporal stationarity of background noise in MEG/EEG measurements *NeuroImage* **20** 233–43
- Chen M, Shao Q and Ibrahim J G 2000 *Monte Carlo Methods in Bayesian Computation* (Berlin: Springer)
- Darvas F, Rautiainen M, Pantazis D, Baillet S, Benali H, Mosher J C, Garnero L and Leahy R M 2005 Investigations of dipole localization accuracy in MEG using the bootstrap *NeuroImage* **25** 355–68
- De Munck J C, Huizenga H M, Waldorp L J and Heethaar R 2002 Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise *IEEE Trans. Biomed. Eng.* **50** 1565–72
- Gilks W R, Richardson S and Spiegelhalter D J 1995 *Markov Chain Monte Carlo in Practice* (London/Boca Raton, FL: Chapman and Hall/CRC)
- Green P J 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination *Biometrika* **82** 711–32
- Hämäläinen M, Hari R, Ilmoniemi R J, Knuutila J and Lounasmaa O V 1993 Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain *Rev. Mod. Phys.* **65** 413–97
- Huang M *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 *Human Brain Mapp.* **11** 59–76
- Huang M, Aine C J, Supek S, Best E, Ranken D and Flynn E R 1998 Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography *Electroenceph. Clin. Neurophysiol.* **108** 32–44
- Huizenga H M, de Munck J C, Waldorp L J and Grasman R P P 2002 Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model *IEEE Trans. Biomed. Eng.* **49** 533–9
- 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** 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 spatiotemporal Bayesian inference dipole analysis *Phys. Med. Biol.* **51** 2395–414
- Jun S C, Pearlmutter B A and Nolte G 2002 Fast accurate MEG source localization using a multilayer perceptron trained with real brain noise *Phys. Med. Biol.* **47** 2547–60
- Mosher J C, Lewis P S and Leahy R M 1992 Multiple dipole modeling and localization from spatiotemporal MEG data *IEEE Trans. Biomed. Eng.* **39** 541–57
- Pantazis D, Nichols T E, Baillet S and Leahy R M 2005 A comparison of random field theory and permutation methods for the statistical analysis of MEG data *NeuroImage* **25** 383–94
- Plis S M, George J S, Jun S C, Paré-Blagoev J, Ranken D M, Schmidt D M and Wood C C 2006 Modeling spatiotemporal covariance for MEG/EEG source analysis. <http://arxiv.org/abs/physics/0503063> submitted

- Ranken D M, Best E, Stephen J M, Schmidt D M, George J S, Wood C C and Huang M 2002 MEG/EEG forward and inverse modeling using mriview *Biomag 2002: 13th Int. Conf. on Biomagnetism (Jena, Germany)* pp 785–7, <http://www.lanl.gov/p/p21/mriview.shtml>
- Sarvas J 1987 Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem *Phys. Med. Biol.* **32** 11–22
- Schmidt D M, George J S and Wood C C 1999 Bayesian inference applied to the electromagnetic inverse problem *Human Brain Mapp.* **7** 195–212
- Uutela K, Hämäläinen M and Salmelin R 1998 Global optimization in the localization of neuromagnetic sources *IEEE Trans. Biomed. Eng.* **45** 716–23
- Wood C C, Cohen D, Cuffin B N and Allison T 1985 Electrical sources in human somatosensory cortex: identification by combined magnetic and potential recordings *Science* **227** 1051–3
- Wood C C, Spencer D D, Allison T, McCarthy G, Williamson P D and Goff W R 1988 Localization of human sensorimotor cortex during surgery by cortical surface recordings of somatosensory evoked potentials *J. Neurosurg.* **68** 99–111