Systematic Regularization of Linear Inverse Solutions of the EEG Source Localization Problem

Christophe Phillips,*,†,‡ Michael D. Rugg,* and Karl J. Friston†

*Institute of Cognitive Neuroscience and †Wellcome Department of Cognitive Neurology, Institute of Neurology, University College London, London, United Kingdom; and ‡Centre de Recherches du Cyclotron, Université de Liège, Liège, Belgium

Distributed linear solutions of the EEG source localization problem are used routinely. Here we describe an approach based on the weighted minimum norm method that imposes constraints using anatomical and physiological information derived from other imaging modalities to regularize the solution. In this approach the hyperparameters controlling the degree of regularization are estimated using restricted maximum likelihood (ReML). EEG data are always contaminated by noise, e.g., exogenous noise and background brain activity. The conditional expectation of the source distribution, given the data, is attained by carefully balancing the minimization of the residuals induced by noise and the improbability of the estimates as determined by their priors. This balance is specified by hyperparameters that control the relative importance of fitting and conforming to prior constraints. Here we introduce a systematic approach to this regularization problem, in the context of a linear observation model we have described previously. In this model, basis functions are extracted to reduce the solution space *a priori* in the spatial and temporal domains. The basis sets are motivated by knowledge of the evoked EEG response and information theory. In this paper we focus on an iterative "expectation-maximization" procedure to jointly estimate the conditional expectation of the source distribution and the ReML hyperparameters on which this solution rests. We used simulated data mixed with real EEG noise to explore the behavior of the approach with various source locations, priors, and noise levels. The results enabled us to conclude: (i) Solutions in the space of informed basis functions have a high face and construct validity, in relation to conventional analyses. (ii) The hyperparameters controlling the degree of regularization vary largely with source geometry and noise. The second conclusion speaks to the usefulness of using adaptive ReML hyperparameter estimates. © 2002 Elsevier Science (USA)

1. INTRODUCTION

Estimating the sources of an electromagnetic field (EEG/MEG) recorded on the scalp surface is a major problem and the subject of a large literature. Various solutions have been proposed that attain varying degrees of success. These can be divided into two main categories: "equivalent current dipole" approaches, in which the EEG/MEG signals are assumed to be generated by a relatively small number of focal sources (Miltner *et al.,* 1994; Scherg and Ebersole, 1994; Scherg *et al.,* 1999; Aine *et al.,* 2000), and the "linear distributed" approach, in which "all" possible source locations are considered simultaneously (Backus and Gilbert, 1970; Sarvas, 1987; Hämäläinen and Ilmoniemi, 1994; Grave de Peralta Menendez and Gonzalez Andino, 1999; Pascual-Marqui, 1999; Uutela *et al.,* 1999).

Here we present a linear distributed approach, based on the weighted minimum L2-norm (WMN) method, that employs constraints using anatomical and physiological information derived from other imaging modalities. This approach has already been introduced in Phillips *et al.* (2002) and rests upon three assumptions (Nunez, 1981; Hämäläinen *et al.,* 1993; Dale and Sereno, 1993): sources are located in gray matter, they are oriented orthogonal to the cortical sheet, and, for a sufficiently dense dipole distribution, they are locally coherent (i.e., their activity changes smoothly along the cortical sheet). These constraints are used to reduce the solution space *a priori* by modeling the spatial source distribution with a set of basis functions. By analogy to the spatial basis functions, the construction of a temporal basis set can be motivated by knowledge about the evoked EEG response: namely, the "window" of activity and its temporal coherence. This set of tem-

poral basis functions can be used to further reduce the size of the solution space.

Spatial and temporal basis functions are used in this paper to illustrate how priors, which can be specified with high precision, can be employed to form basis sets. The reduced problem can then be solved using, for example, a classical WMN method. Further "soft" constraints can be introduced in the weighting of the WMN solution using, for example, hemodynamic measures of brain activity as spatial priors (Rugg, 1998; Liu *et al.,* 1998; Dale *et al.,* 2000). The focus of this paper is the relative weighting of these priors in the context of observation noise.

In the context of noisy data, regularization is necessary to control the relative influence of the likelihood of, and priors on, the solution. Within the WMN approach regularization is achieved by carefully balancing, through some hyperparameters λ_i, the fitting of the data and the strength of priors, according to the level of noise, and the precision of the priors that are considered. An iterative procedure for estimating hyperparameters, using "restricted maximum likelihood" (ReML), is presented. An "expectation-maximization" (EM) algorithm is used to jointly estimate the parameters (source distribution) and the hyperparameters (that parameterize the variance of the priors and noise).

In this paper we fix the relative contribution of different priors so that the effective number of hyperparameters reduces to 1. This hyperparameter can be regarded as the standard deviation of the noise relative to that expected for the underlying sources. The reason for doing this is to show how the Bayesian formulation can be developed from standard regularization, in which the regularization parameter is exactly the same as our single hyperparameter.

This paper is divided into two subsequent sections. In the first (Section 2) the theoretical background and operational details of our approach are described. The first part (Sections 2.1–2.3) is a brief synopsis of the technique presented in full in Phillips *et al.* (2002), while the second part (Sections 2.4–2.5) introduces the regularization procedure. In the last section (Section 3), we use simulated signals mixed with real EEG noise to explore the behavior of the approach over a range of variables and noise levels. To assess the construct validity of our method we compare it with two established approaches to the source localization problem, simple weighted minimum norm and a maximum smoothness (LORETA-like) solution.

In the rest of this paper, a, $\vec{a}$, $\mathbf{a}$, and $\mathbf{A}$ will represent, respectively, a scalar, a vector of size 3×1, a vector of any size $N_a \times 1$ and a matrix; $\mathbf{A}^t$ will designate the transpose of $\mathbf{A}$.

2. THEORY

The instantaneous source localization problem in EEG can be summarized by the equation

$$\mathbf{v} = \mathscr{F}(\vec{r}, \vec{j}) + \epsilon, \tag{1}$$

where $\mathbf{v}$, a vector of size $N_e \times 1$, is the potential at the N_e electrodes, $\vec{r}$ and $\vec{j}$ are the source location and moment, ϵ is the additive noise, and $\mathscr{F}$ is the function linking the sources $(\vec{r}, \vec{j})$ and the potential $\vec{r}$. The function $\mathscr{F}$ is the solution of the forward problem and depends only on the head model adopted.

For multiple sources defined by $\vec{r}_i$ and $\vec{j}_i$ (with $i = 1, \ldots, N_i$) the source localization problem is written as

$$\mathbf{v} = \sum_{i=1}^{N_i} \mathscr{F}(\vec{r}_i, \vec{j}_i) + \epsilon. \tag{2}$$

2.1. Head and Source Model

$\mathscr{F}$ was calculated using the "boundary element method" (BEM) for a three-shell realistic head model (Geselowitz, 1967; Hämäläinen and Sarvas, 1989; Meijs *et al.,* 1989; de Munck, 1992; Schlitt *et al.,* 1995; Ferguson and Stroink, 1997; Buchner *et al.,* 1997). A structural MR image of the head was segmented (Ashburner and Friston, 1997) and divided into three volumes with homogeneous isotropic conductivity: the brain, the skull, and the scalp volume. Then the potential at the electrode sites on the scalp surface was computed for any current source located in the brain volume.

In this paper, the sources of the EEG signal are modeled by a fixed, uniform, three-dimensional grid of current dipoles throughout the entire brain volume, conforming to the "distributed linear solution" approach. Because the location $\vec{r}$ of each current source is now fixed, Eq. (1) becomes an underdetermined but linear problem,

$$\mathbf{v} = \mathbf{L}\mathbf{j} + \epsilon, \tag{3}$$

where $\mathbf{j}$ represents the current dipoles at all the N_d locations simultaneously, and $\mathbf{L}$ is the lead field matrix linking the current sources $\mathbf{j}$ to the electrical potential $\mathbf{v}$. If the source orientation is left free, then $\mathbf{j} = [\vec{j}_1^t, \vec{j}_2^t, \ldots, \vec{j}_{N_d}^t]^t$, where $\vec{j}_i = [j_{x,i}, j_{y,i}, j_{z,i}]^t$ encodes both orientation and amplitude of the ith current dipole. Otherwise the orientated sources, $\mathbf{j} = [j_i, j_2, \ldots, j_{N_d}]^t$, where each j_i specifies only the amplitude of the ith current dipole.

For discrete data time series with N_t time bins, Eq. (3) can be expressed as a multivariate linear model,

$$\mathbf{V} = \mathbf{L}\mathbf{J} + \mathscr{E}, \tag{4}$$

with $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_t}]$, $\mathbf{J} = [\mathbf{j}_1, \mathbf{j}_2, \ldots, \mathbf{j}_{N_t}]$, and $\mathscr{E} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_{N_t}]$ where $\mathbf{v}_l$, $\mathbf{j}_l$, and ϵ_l are the potential, current dipoles, and additive noise at the lth time.

For the simulations presented in Section 3, the realistic BEM head model was generated from the template T1-weighted MR image of the SPM99 software (Wellcome Department of Cognitive Neurology, 1999). The lead field matrix $\mathbf{L}$ was calculated for a set of $N_e = 61$ approximately equidistant electrodes and $N_d = 12{,}300$ dipoles.

2.2. Weighted Minimum Norm Solution

As Eq. (3) is underdetermined, some constraints are necessary to estimate a unique solution. These constraints are generally based on *a priori* knowledge about the sources. One common approach to solve this problem is the WMN solution or Tikhonov regularization method (Tikhonov and Arsenin, 1977), in which the *a priori* constraint can be interpreted in a Bayesian sense.

The WMN solution constrains the reconstructed source distribution by minimizing a linear mixture of some weighted norm $\|\mathbf{Hj}\|$ of the current sources $\mathbf{j}$ and the residuals of the fit. Assuming the noise ϵ is Gaussian $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{C}_\epsilon)$ with a covariance that can be factorized into a scalar variance σ^2 and a known matrix $\mathbf{C}_\epsilon$,

$$\hat{\mathbf{j}} = \arg \min_{\mathbf{j}} \{\|\mathbf{C}_\epsilon^{-1/2}(\mathbf{Lj} - \mathbf{v})\|^2 + \lambda^2\|\mathbf{Hj}\|^2\}, \qquad (5)$$

where the hyperparameter λ expresses the balance between fitting the model $\|\mathbf{C}_\epsilon^{-1/2}(\mathbf{Lj} - \mathbf{v})\|$ and minimizing the *a priori* constraint $\|\mathbf{Hj}\|$. We will refer to $\mathbf{C}_\epsilon$ as a correlation matrix, but note this does not imply $\mathrm{diag}(\mathbf{C}_\epsilon) = 1$. In this paper a correlation matrix plays the same role as a covariance matrix but is scaled so that $\mathrm{trace}(\mathbf{C}_\epsilon) = \mathrm{rank}(\mathbf{C}_\epsilon)$.

Equation (5) can also be expressed as an overdetermined least-square problem,

$$\begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{v} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{L} \\ \lambda\mathbf{H} \end{bmatrix}\mathbf{j} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \qquad (6)$$

where the solution is obtained by minimizing the norm of the residuals $[\epsilon_1^t\ \epsilon_2^t]^t$. The solution of Eq. (5) or (6) for a given λ is

$$\hat{\mathbf{j}} = \mathbf{Tv}, \qquad (7)$$

where

$$\mathbf{T} = [\mathbf{L}^t\mathbf{C}_\epsilon^{-1}\mathbf{L} + \lambda^2(\mathbf{H}^t\mathbf{H})]^{-1}\mathbf{L}^t\mathbf{C}_\epsilon^{-1} \qquad (8a)$$

$$= (\mathbf{H}^t\mathbf{H})^{-1}\mathbf{L}^t[\mathbf{L}(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{L}^t + \lambda^2\mathbf{C}_\epsilon]^{-1} \qquad (8b)$$

using the matrix inversion Lemma.

For the continuous case of multiple observations like Eq. (4), the least-square problem is formally identical to Eq. (6),

$$\begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{V} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{L} \\ \lambda\mathbf{H} \end{bmatrix}\mathbf{J} + \begin{bmatrix} \mathscr{E}_1 \\ \mathscr{E}_2 \end{bmatrix}, \qquad (9)$$

and has the corresponding solution Eq. (7).

The important and useful connection with Bayesian estimates of the sources rests on Gaussian assumptions, when the conditional expectation or posterior mean of the sources $\mathbf{j}$ is given by

$$E(\mathbf{j}|\mathbf{v}) = [\mathbf{L}^t\mathbf{C}_\epsilon^{-1}\mathbf{L} + \sigma^2\mathbf{C}_\mathbf{j}^{-1}]^{-1}\mathbf{L}^t\mathbf{C}_\epsilon^{-1}\mathbf{v} \qquad (10a)$$

$$= \mathbf{C}_\mathbf{j}\mathbf{L}^t[\mathbf{L}\mathbf{C}_\mathbf{j}\mathbf{L}^t + \sigma^2\mathbf{C}_\epsilon]^{-1}\mathbf{v}, \qquad (10b)$$

where $\mathbf{C}_\mathbf{j}$ is the prior covariance of the sources and σ^2 is the noise variance. Comparing Eqs. (10) with Eqs. (8) provides the motivation for choosing forms of $\mathbf{H}$ where

$$\left(\frac{\lambda}{\sigma}\right)^2(\mathbf{H}^t\mathbf{H}) = \mathbf{C}_\mathbf{j}^{-1}. \qquad (11)$$

As will be shown later (Section 2.4.1), λ can be thought of as the ratio σ/η where $\mathbf{C}_\mathbf{j} = \eta^2(\mathbf{H}^t\mathbf{H})^{-1}$ and $(\mathbf{H}^t\mathbf{H})^{-1}$ is the prior correlation matrix of $\mathbf{j}$ and η^2 is the prior variance.

One could of course embrace a whole series of priors or constraints,

$$\eta_1^{-2}(\mathbf{H}_1^t\mathbf{H}_1) + \eta_2^{-2}(\mathbf{H}_2^t\mathbf{H}_2) + \ldots = \mathbf{C}_\mathbf{j}^{-1}. \qquad (12)$$

This would correspond to expanding the prior dispersion $\mathbf{C}_\mathbf{j}^{-1}$ in terms of a linear basis set, specified by $\mathbf{H}_i$. Although the EM algorithm below can easily handle multiple hyperparameters (see Section 2.4), for simplicity we will deal with a single set of constraints.

In some instances the priors can be so precise that they preclude the solution from spanning certain subspaces of the solution space (precision is the inverse of variance); i.e., if $\mathbf{C}_\mathbf{j}$ was 0 along its leading diagonal the regularization or penalty would be infinite at that location and the conditional estimate Eqs. (10) would be zero. For example, the prior variance of source activity in white matter can be set to 0. In these situations it is computationally more efficient to remove these "impossible" subspaces before computing the WMN solution. This involves projecting the solution onto a subspace that has nonzero prior variance. An example of this approach is the use of spatial or temporal basis functions to reduce the solution space as described in the next section. The use of basis functions can therefore be viewed as implementing priors with infinite precision.

2.3. Constraints and Priors

Anatomical and physiological information derived from other imaging modalities are often used to constrain the solution in the spatial domain. A two-step approach was presented in Phillips *et al.* (2002): First the source distribution is modeled by a set of spatial and temporal basis functions ("hard" constraints specified with infinite precision), then further "soft" constraints are introduced through **H** in the WMN solution in the usual way.

2.3.1. Basis Functions

As presented in Phillips *et al.* (2002), prior knowledge about the sources **J** can be embodied in matrices summarizing the prior constraints on source location (in space or time), say **G**, and dispersion, say **D**. This prior on the signal induces spatiotemporal coherence. **G** is a leading diagonal matrix encoding the prior variance in signal at different points in space or peristimulus time. **D** will be a (generally nonstationary) convolution matrix that controls the degree of coherence in the prior covariance matrix **C**, where

$$\mathbf{C} = \mathbf{G}^{t/2}\mathbf{D}^t\mathbf{D}\mathbf{G}^{1/2}. \qquad (13)$$

Informed basis functions **B** are then obtained from the eigenvector solution of **C** or, equivalently, by using the singular value decomposition of $\mathbf{C}^{1/2} = \mathbf{DG}^{1/2}$:

$$\mathbf{USW}^t = \text{svd}(\mathbf{DG}^{1/2}). \qquad (14)$$

Columns of **W** corresponding to normalized eigenvalues $\mathbf{S}^2$ greater than unity are retained to form the basis set **B**. Selecting these eigenvectors, either in space or time, ensures a high mutual information between the signal before and that after projection onto the basis set (under prior assumptions). In terms of the Bayesian formulation this is equivalent to setting the prior variance of spatial or temporal modes (corresponding to the "minor" eigenvectors) to 0. The major modes, retained to form the basis set, could be entered into Eq. (12) but again, for simplicity we will discount them by effectively setting their associated eigenvalues to infinity. This corresponds to making the prior density (in the subspace spanned by the major modes) "flat" or uniform with zero precision (i.e., an infinitely wide Gaussian density function in this subspace is essentially "flat"). Because their precision is 0 they do not enter into Eq. (12) (i.e., their associated hyperparameter is 0).

We now describe the constraints used in our simulations, starting with spatial and moving to temporal constraints.

2.3.2. Spatial Constraints

The orientation and location of the dipoles are embodied in the lead field matrix **L**. The gray matter density at each dipole location and the spatial coherence imposed among the dipoles can be represented by a diagonal matrix $\mathbf{G}_s$ and a covariance matrix $\mathbf{D}_s$. With these matrices $\mathbf{G}_s$ and $\mathbf{D}_s$ a set of "spatially informed basis functions" (sIBF) $\mathbf{B}_s$ can be constructed (as in Section 2.3.1 above), such that

$$\mathbf{B}_s^t\mathbf{j} = \mathbf{k}_s \Rightarrow \mathbf{j} = \mathbf{B}_s\mathbf{k}_s, \qquad (15)$$

where $\mathbf{k}_s$ is the vector of size $N_{k_s} \times 1$ with $N_{k_s} \ll N_d$. By substituting Eq. (15) into Eq. (5), the unknowns become $\mathbf{k}_s$ and the size of the solution space is reduced from N_d to N_{k_s}.

In the simulations presented below 2 degrees of spatial coherence between the sources were modeled in $\mathbf{D}_s$ by using two different Gaussians, $\sigma = 10$ mm and $\sigma = 5$ mm, of the geodesic distance between the dipoles as described in Phillips *et al.* (2002). The solutions obtained using Eqs. (8) and (15) will be referred to as the "informed basis functions" (IBF) solutions of kernel 10 mm (IBF10) or 5 mm (IBF5), with $N_{k_s} = 621$ and $N_{k_s} = 1903$, respectively.

Further soft constraints are embodied in the matrix **H** of the WMN solution, Eq. (5). Through relation (11), it is possible to render some sources more (or less) probable by scaling their *a priori* variance. First, in order to ensure that sources are likely to influence the electrical potential equally at the electrodes irrespective of their depth (Ioannides *et al.,* 1990; Grave de Peralta Menendez and Gonzalez Andino, 1998; Pascual-Marqui, 1999), deeper sources are given a larger *a priori* variance than superficial sources. The depth is indexed by the norm of the lead field for each source and the covariance component of this constraint is the diagonal matrix $(\text{diag}(\mathbf{L}^t\mathbf{L}))^{-1}$. Using this form for $\mathbf{C_j}$ is equivalent to normalizing the column of the lead field matrix in an Euclidian sense.

A second important constraint can be derived from fMRI indices of activation that enter as $\boldsymbol{\alpha}$, a leading diagonal matrix with elements that reflect the prior probability of whether the source is active (here we allow only values of 0, the variance is left unchanged, or 1, the variance is increased according to the value of a hyperparameter). Combining these components, we can form the following expression:

$$\mathbf{C_j}^{-1} = \eta^2(\mathbf{H}^t\mathbf{H}) = \eta^2\text{diag}(\mathbf{L}^t\mathbf{L})(\mathbf{I}_{N_j} + \beta\boldsymbol{\alpha})^{-1}. \quad (16)$$

In principle we could estimate both hyperparameters η and β (or their ratio with σ). Generally one would take

any arbitrary expression like Eq. (16) for the prior precision and reduce it to the form of Eq. (12) using a Taylor expansion with respect to the hyperparameters. Algorithms for finding ReML estimates of multiple parameters are readily available. However, to simplify the ReML demonstration we will assume fixed values for β and focus on estimating η (or $\lambda = \sigma/\eta$).

We chose three values for β, 0, 3, and 24, corresponding to "no," "weak," and "strong" (fMRI) location priors. By taking β equal to 3 or 24, we are assuming that the *a priori* variance of the activity is 4 or 25 times larger than the other locations (if the depth constraint was discounted).

2.3.3. Temporal Constraints

The temporal constraints are derived in the same way as the spatial basis functions and are based on knowledge of the evoked EEG response. The time window that encompassed the evoked response was modeled by the leading diagonal matrix $\mathbf{G}_t$ (size $N_t \times N_t$). The ith diagonal element of $\mathbf{G}_t$ takes a value between 0, zero probability of evoked activity at the ith time instant, and 1, possible activity at the ith time instant. Temporal coherence is modeled by the matrix $\mathbf{D}_t$ of size $N_t \times N_t$. The matrix $\mathbf{D}_t$ can be defined, for example, as a convolution matrix comprising Gaussian kernels. By choosing kernels of different sizes throughout the epoch, it would be possible, for example, to allow higher frequencies at the beginning of the epoch than at the end.

As for the sIBF, a set of "temporally informed basis functions" (tIBF) $\mathbf{B}_t$ can be constructed (Section 2.3.1), such that the "temporal domain" is reduced from N_t to N_{k_t}. With $\mathbf{B}_s$ and $\mathbf{B}_t$, the $N_d \times N_t$ source matrix $\mathbf{J}$ can be modeled as

$$\mathbf{B}_s^t \mathbf{J} \mathbf{B}_t = \mathbf{K}_{st} \Rightarrow \tilde{\mathbf{J}} = \mathbf{K}_t \mathbf{B}_t^t = \mathbf{B}_s \mathbf{K}_{st} \mathbf{B}_t^t, \qquad (17)$$

where $\mathbf{K}_{st}$ is a $N_{k_s} \times N_{k_t}$ matrix and $\mathbf{B}_t$ is a $N_t \times N_{k_t}$ matrix.

In this paper the prestimulus period was chosen to last 400 ms from -400 to 0 ms (100 time samples), thus the window of stimulus-related activity ranged from 0 to 1600 ms (400 time samples). Rise and fall periods, 40 ms each, at the extremities of the window of activity were modeled as the ascending and descending part of a Hanning window to give the constraint matrix $\mathbf{G}_t$. The temporal coherence matrix $\mathbf{D}_t$ was modeled by a Gaussian convolution matrix, the kernel of the Gaussian function was $\sigma_t = 16$ ms wide (4 points) and was kept constant throughout the whole epoch.

2.4. Noise Regularization Procedure

The WMN solution, presented in Section 2.2, depends on the hyperparameter λ. The hyperparameter λ balances the relative contribution of fitting the model $\mathbf{C}_\epsilon^{-1/2}(\mathbf{L}\mathbf{j} - \mathbf{v})$ and the priors on the solution $\mathbf{H}\mathbf{j}$. As λ varies, the regularized solution $\mathbf{j}_\lambda$ changes in a way that depends strongly on λ. Therefore the choice of λ is crucial. As a general rule, the degree of regularization (λ) should increase with the level of noise in the data, i.e., the importance of the priors should increase as the model fit decreases, but this rule is not sufficient to estimate λ.

A heuristic way to display and understand the properties of $\mathbf{j}_\lambda$ is to plot the (weighted) norm of the regularized solution $\|\mathbf{H}\mathbf{j}_\lambda\|^2$ versus the norm of the residual vector $\|\mathbf{C}_\epsilon^{-1/2}(\mathbf{L}\mathbf{J}_\lambda - \mathbf{v})\|^2$ for different values of λ. The curve obtained usually has an L shape (in ordinary or double logarithmic scale), hence its name "L-curve." A satisfactory λ would lie close to the inflection of the L-curve (Hansen, 1992). But Engl and Grever (1994) showed that hyperparameter choice strategies, based on this plot alone, e.g., on the corner of the L-curve, are independent of the norm of the difference between the true noise-free signal and the measured (and noise contaminated) signal. Another disadvantage of the L-curve approach is that the solution must be calculated for a large number of values of λ to find an appropriate regularization level.

Fortunately there are other ways to simultaneously estimate the hyperparameter λ and its associated solution $\mathbf{j}_\lambda$ in an optimal way. The ReML approach is a variation of the classic "maximum likelihood" solution, in which the hyperparameters are estimated in alternation with the (parameters of the) solution itself. The mathematical details of the approach are presented in the Appendix.

The least-square expressions of the WMN problem, see Eqs. (6) and (9), are formally identical to those used by the "simple ReML" solution, Eqs. (29) and (35) in the Appendix, with the following substitutions:

$$\mathbf{A}_1 = \mathbf{C}_\epsilon^{-1/2} \mathbf{L_B}, \qquad (18a)$$

$$\mathbf{A}_2 = \mathbf{H_B}, \qquad (18b)$$

$$\mathbf{x} = \mathbf{k} \ (\text{or } \mathbf{X} = \mathbf{K}_{st}), \qquad (18c)$$

$$\mathbf{b}_1 = \mathbf{C}_\epsilon^{-1/2}\mathbf{v} \ (\text{or } \mathbf{B}_1 = \mathbf{C}_\epsilon^{-1/2}\mathbf{V_B}), \qquad (18d)$$

$$\mathbf{b}_2 = 0 \ (\text{or } \mathbf{B}_2 = 0), \qquad (18e)$$

$$\theta_1 = \sigma, \qquad (18f)$$

$$\theta_2 = \eta, \qquad (18g)$$

where relations (15) and (17) are used and $\mathbf{B}_s$ and $\mathbf{B}_t$ are combined with $\mathbf{L}$, $\mathbf{H}$, and $\mathbf{V}$: $\mathbf{L_B} = \mathbf{LB}_s$, $\mathbf{H_B} = \mathbf{HB}_s$, and $\mathbf{V_B} = \mathbf{VB}_t$.

By considering λ^2 as the ratio between the variance σ^2 of the whitened residuals of the model $\mathbf{C}_\epsilon^{-1/2}(\mathbf{L_B}\mathbf{k} - \mathbf{v})$ and the variance η^2 of the weighted solution $\mathbf{H_B}\mathbf{k}$, then it is possible to apply the iterative procedure described in the Appendix to estimate both $\mathbf{k}$ and $\lambda = \sigma/\eta$ in

$$\hat{\mathbf{k}} = \arg\min_{\mathbf{k}} \left\| \begin{bmatrix} \sigma^{-1}\mathbf{I}_{N_e} & 0 \\ 0 & \eta^{-1}\mathbf{I}_{N_j} \end{bmatrix} \right.$$

$$\left. \times \left(\begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{L_B} \\ \mathbf{H_B} \end{bmatrix} \mathbf{k} - \begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{v} \\ 0 \end{bmatrix} \right) \right\|^2. \tag{19}$$

This approach can be understood as (i) reformulating a linear Bayesian (or WMN) estimation as an augmented least-squares problem and (ii) the hyperparameters of the variance components of this problem are estimated using an EM algorithm to find their ReML estimate. The use of scalar variance hyperparameter σ^2 for the residuals of the model and η^2 for the constraint is appropriate. This is because the residuals of the model $\mathbf{L_B}\mathbf{k} - \mathbf{v}$ are premultiplied by $\mathbf{C}_\epsilon^{-1/2}$, the inverse of the square root of its correlation (or its estimator); therefore the residuals are "whitened" and the resulting covariance can be approximated by a scaled identity matrix $\sigma^2\mathbf{I}_{N_e}$. Similarly, the matrix $\mathbf{H_B}$ can be interpreted as the inverse of the square root of the scaled *a priori* correlation of the sources, as shown in Eq. (11), thus the variance of the weighted solution $\mathbf{H_B}\mathbf{k}$ can be modeled by $\eta^2\mathbf{I}_{N_j}$.

It can be seen that the ReML estimation actually estimates two hyperparameters, the variance of the noise σ^2 and the prior variance of the parameters η^2. The WMN perspective requires only their ratio (λ^2), which is generally specified directly. The reason we framed the introduction in terms of a single hyperparameter was to establish the connection between classical regularization approaches in linear estimation theory and variance component estimation using ReML in the context of conditional or Bayesian estimation. As will be shown later, simply assuming a fixed and apparently sensible value for λ may be suboptimal, especially when the influence of noise can change markedly with source geometry and other measurement configurations.

To incorporate multiple priors (see Eq. (12)), one simply augments Eq. (19) to

$$\hat{\mathbf{k}} = \arg\min_{\mathbf{k}} \left\| \begin{bmatrix} \sigma^{-1}\mathbf{I}_{N_e} & 0 & \cdots & 0 \\ 0 & \eta_1^{-1}\mathbf{I}_{N_j} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \eta_N^{-1}\mathbf{I}_{N_j} \end{bmatrix} \right.$$

$$\left. \times \left(\begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{L_B} \\ \mathbf{H}_{1,\mathbf{B}} \\ \vdots \\ \mathbf{H}_{N,\mathbf{B}} \end{bmatrix} \mathbf{k} - \begin{bmatrix} \mathbf{C}_\epsilon^{-1/2}\mathbf{v} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) \right\|^2, \tag{20}$$

which can be solved through a minimization of Eq. (28) in the Appendix. This generalization of simple "noise" regularization to balance different classes of constraints will be the subject of a subsequent paper.

2.5. Comparison Methods

The IBF approach described here was compared with two other commonly employed approaches: a WMN solution and a maximum smoothness (MS) solution.

The WMN solution is simply the solution of the problem formulated in Eq. (5) given by Eq. (8) discounting the sIBF $\mathbf{B}_s$ (*i.e.*, $\mathbf{B}_s = \mathbf{I}_{N_d}$). In the version of the method employed here, the orientation of the dipoles was fixed as with the IBF method, so that only the amplitudes were unknown. The weighting matrix $\mathbf{H}$ was defined as in Eq. (16): it incorporated depth weighting and prior knowledge (or not) of the location of active sources. The WMN was used here to calculate three different solutions with the same range of β: without priors, with weak priors, or with strong priors.

The MS solution is also a particular case of Eqs. (5) and (8). No sIBF are used and the orientations of the dipoles are left free. The weighting matrix $\mathbf{H}$ was defined as a weighted three-dimensional Laplacian, $\mathbf{H} = \mathbf{MW}$. $\mathbf{W}$ is a leading diagonal matrix defined by $\mathbf{w} \otimes [1\ 1\ 1]^t$, where $\mathbf{w} = [(\mathbf{I}_{N_j} \otimes [1\ 1\ 1])\text{diag}(\mathbf{L}^t\mathbf{L})]^{1/2}$, which is equivalent to the depth weighting used for the IBF and WMN solutions. The operator $\otimes$ denotes the Kronecker product and $\mathbf{L}$ is the $N_e \times N_{3N_j}$ lead field matrix corresponding to the orientation-free sources vector $\mathbf{j}$, as described in Section 2.1. The Laplacian matrix $\mathbf{M}$ is a regularized discrete three-dimensional second-order derivative operator defined as in Pascual-Marqui (1999). A single value j_i for the amplitude of the electrical activity at each source location i was then obtained by calculating the norm of the resulting dipole $j_i = \sqrt{j_{x,i}^2 + j_{y,i}^2 + j_{z,i}^2}$. This "classic" implementation of the MS solution does not provide means to indepen-

dently include different priors other than the maximum smoothness constraint.

No temporal basis functions were used for the simulations with the WMN and MS solutions but the hyperparameter λ was still obtained through the ReML algorithm. The main objective of these simulations was to establish the face and construct validity of the ReML hyperparameter estimates. We hoped to show that the ReML estimate was sensitive to different levels of noise variance and behaved adaptively and appropriately. Furthermore we wanted to establish this face validity using a series of qualitatively different priors. Differences among priors are implicit in the IBF, WMN, and MS approaches. A secondary aim was to ensure our IBF still performed well, in relation to WMN and MS, when using empirical estimates of λ.

3. SIMULATIONS

3.1. Methods

Spontaneous brain activity was recorded from a volunteer with 61 approximately equidistant electrodes. The head model adopted in Section 2.1 used the same distribution of electrodes. The EEG background activity (and noise) was sampled at 250 Hz over epochs of 2 s, thus containing $N_t = 500$ time samples. Each epoch was bandpass filtered between 0.05 and 20 Hz by a third-order digital Chebychev filter. Finally 150 epochs were averaged to produce a realistic noise time series $\mathscr{E}$ at each electrode site.

A set of 100 locations were randomly selected to assess the efficiency of the methods presented in Sections 2.3 and 2.5. At each of these locations, an instantaneous distributed source set $\mathbf{j}_o$ was generated as a set of anatomically connected dipoles within a 7-mm radius of a "central" dipole. Simulated fMRI information about prior location was provided as a sphere of 12 mm radius around this "central" dipole, and dipoles within this volume were defined as being *a priori* active sources, i.e., the corresponding diagonal elements of α (see Section 2.3.2) were set to 1.

Each source $\mathbf{j}_o$ was modulated over time to generate a time-extended $N_d \times N_t$ data set $\mathbf{J}_o$. The time course adopted for all the sources is shown in Fig. 1. The data with realistic noise were obtained using Eq. (4). The averaged spontaneous brain activity ϵ was scaled to achieve two different levels of signal-to-noise ratio (SNR). The SNR was defined as the ratio between the power of the potential $\mathbf{V}$ over the scalp, i.e., at the electrodes, and the average power of ϵ over the scalp. In one case, the SNR was low at 1.5, and in the other, it was high at 6. The potentials at one electrode, over time, with these two SNRs are shown in Fig. 1.

The IBF10 and IBF5 solutions were calculated for each data set $\mathbf{V}$ with the tIBF and the ReML hyperparameter estimate. For the WMN and MS solutions no
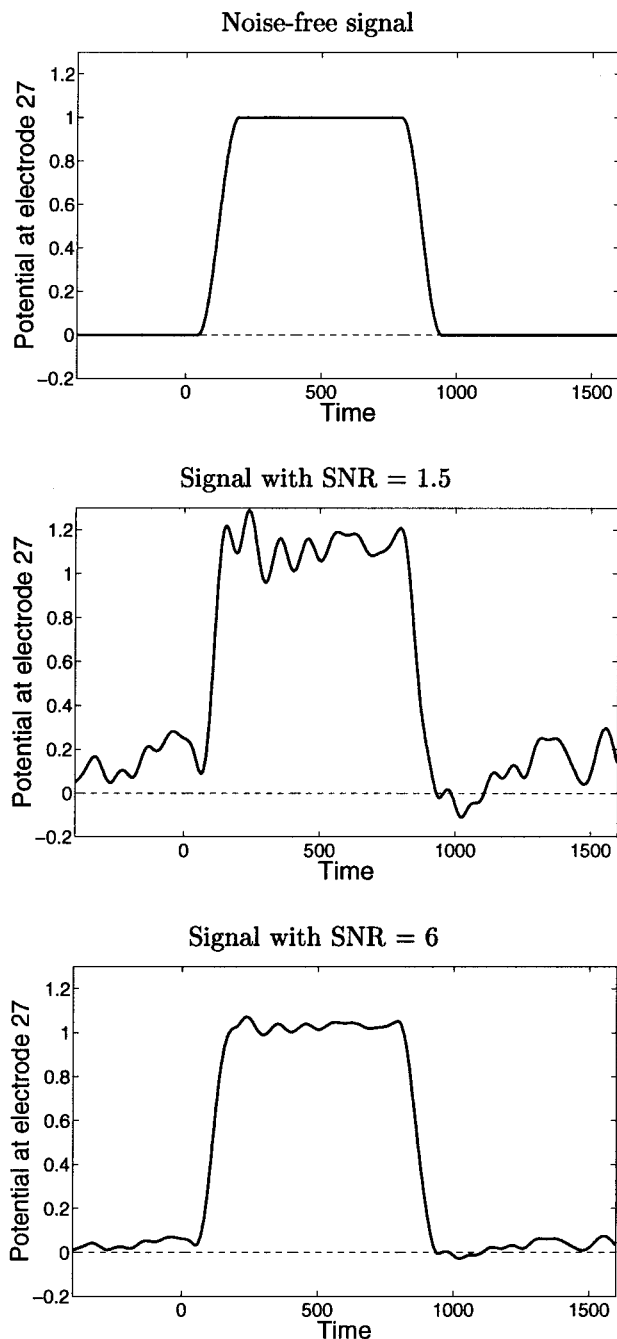

FIG. 1. Time course of the sources and example of data with realistic noise. Free brain activity was added to the noise-free signal at one electrode (top), and all the sources had the same time course. The noise was scaled at two different levels in order to produce two data sets with signal-to-noise ratios (SNRs): SNR = 1.5 (middle) and SNR = 6 (bottom).

basis functions (spatial or temporal) were employed but the hyperparameter λ was still obtained through the ReML algorithm. In the literature (Pascual-Marqui *et al.,* 1994; Pascual-Marqui, 1999), the MS solution is usually implemented in the LORETA software (Pascual-Marqui, 1998) for the ideal noise-free case, using
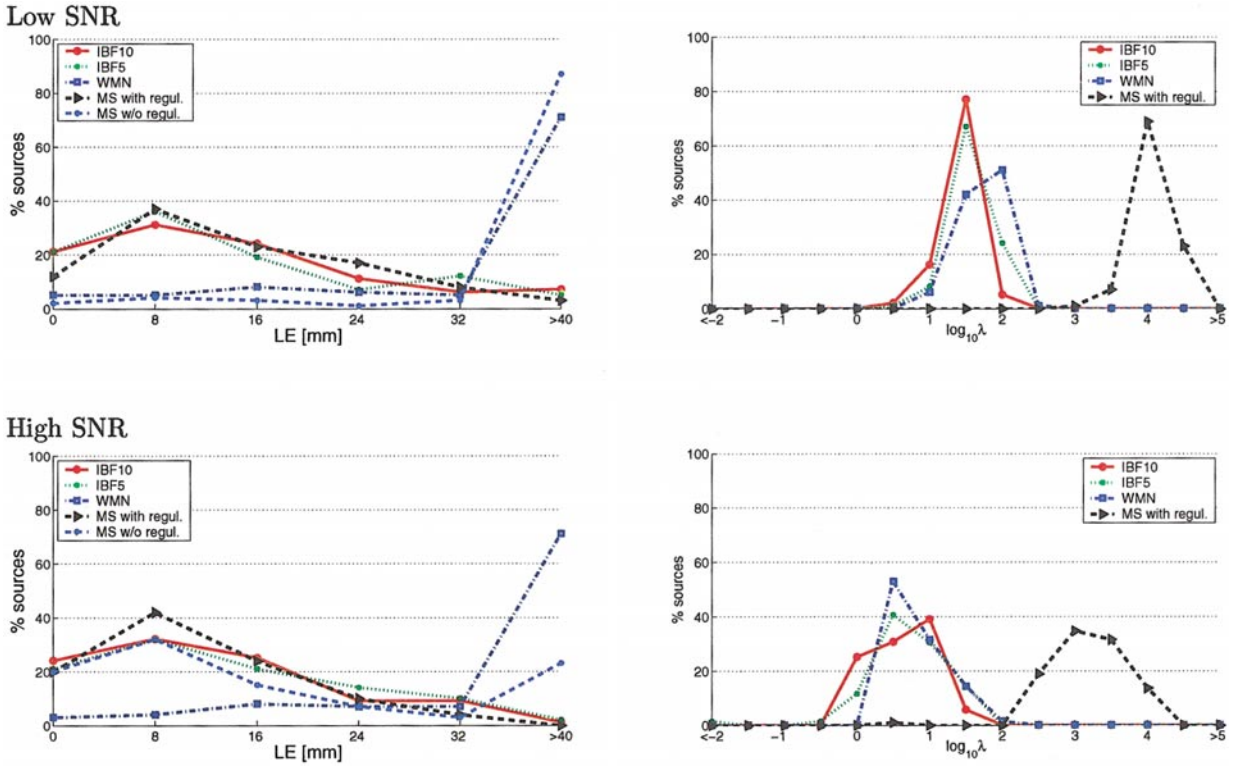
FIG. 2. Localization error (LE), left, and $\log_{10}\lambda$, right, for the four solutions (IBF10, IBF5, WMN, and MS with and without noise regularization) for two levels of signal-to-noise ratio (SNR), low SNR (top), and high SNR (bottom), applied to the simple source simulated data with realistic noise.

Eq. (8) with $\lambda \to 0$ (although Pascual-Marqui (1995) has suggested employing noise regularization by taking $\lambda \neq 0$). We reproduced this $\lambda = 0$ analysis in our simulations and refer to it as "MS with no regularization." The IBF10, IBF5, and WMN were tested without location priors and with weak and strong location priors.

Two criteria were used to assess and compare the performance of the different methods (IBF10, IBF5, WMN, and MS): "locational error" (LE) and "root mean square error" (RMSE) (see also Phillips *et al.,* 2002). The LE is defined as the distance between the location $\vec{r}_r$ of the maximum (absolute value) of the reconstructed source $\mathbf{j}_r$ and the location $\vec{r}_o$ of the original source set $\mathbf{j}_o$. In order to obtain a single value of LE for the estimated and actual distributed source ($\mathbf{J}_o$, $\mathbf{J}_r$), the following procedure was adopted: Given $j_{r,\max}$, the amplitude of the largest (absolute value) dipole in $\mathbf{J}_r$, only reconstructed sources $\mathbf{j}_{r,i}$ with at least one dipole of amplitude (absolute value) greater than 85% of $j_{r,\max}$ were considered. For this subset of reconstructions $\mathbf{j}_{r,i}$, the LE_i was calculated. The largest value of LE_i was retained as the "worst case LE" for ($\mathbf{J}_o$, $\mathbf{J}_r$). The RMSE is defined as the Frobenius norm of the difference of the scaled distributions,
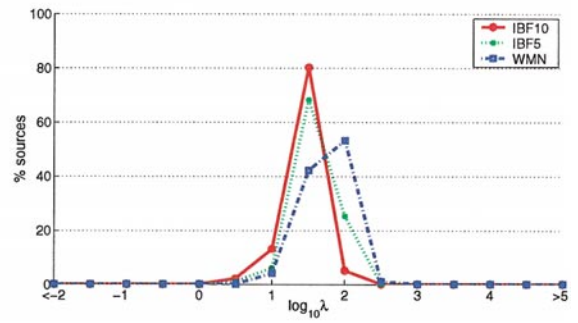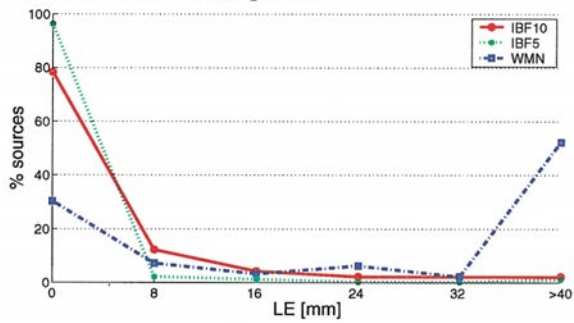
$$\text{RMSE} = \|\mathbf{J}_r^* - \mathbf{J}_o^*\|_{\text{Fro}}, \qquad (21)$$

where $\mathbf{J}_r^* = \mathbf{J}_r/j_{r,\max}$ and $\mathbf{J}_o^* = \mathbf{J}_o/j_{o,\max}$. This ensures that the root mean square error measures the discrepancy between the original and the reconstructed source distributions without any the effect of global scaling effect.
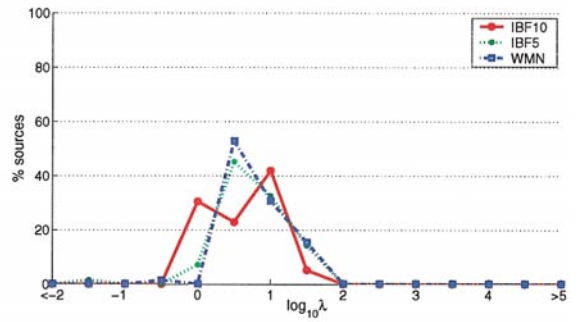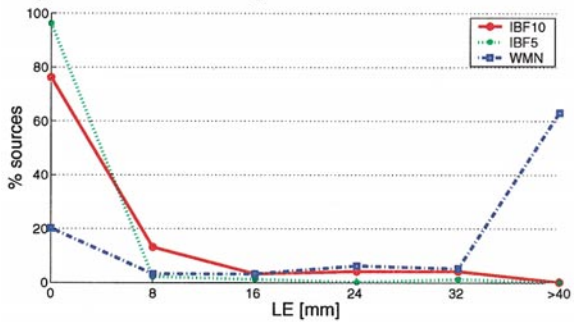
The LE provides a measure of the localization accuracy of the reconstruction method; a small value of LE indicates that the location of the original source was recovered well. The RMSE measures the "goodness of fit" of the reconstruction. A small value of RMSE indicates a small discrepancy between the original and the reconstructed source distributions. The RMSE is useful only to further compare two solutions that have approximately the same LE. Indeed, if both solutions have almost the same LE, the one with the smaller RMSE would be preferred, as the reconstructed source is then more focal. A very focal reconstructed source with a large LE will have a smaller RMSE than a blurred reconstructed source with a small LE, but the latter solution, although oversmoothed, provides at least some location information.

To assess the influence of the level of noise, the LE and RMSE were also calculated with all the methods in the ideal case in which the data are noise free, i.e., the SNR was infinite. These results have been presented in Phillips *et al.* (2002).
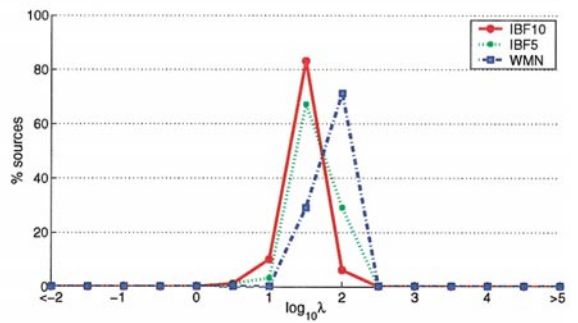
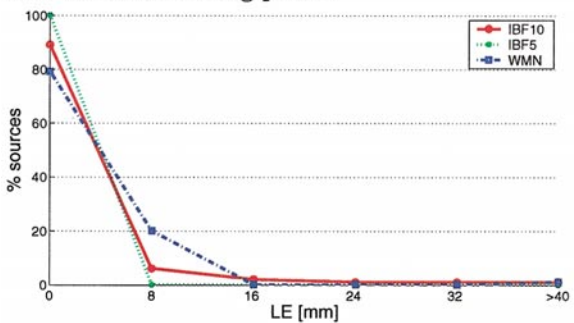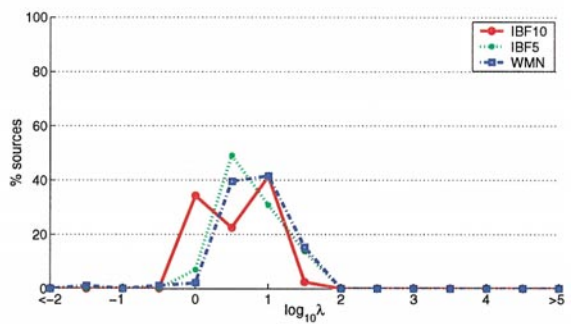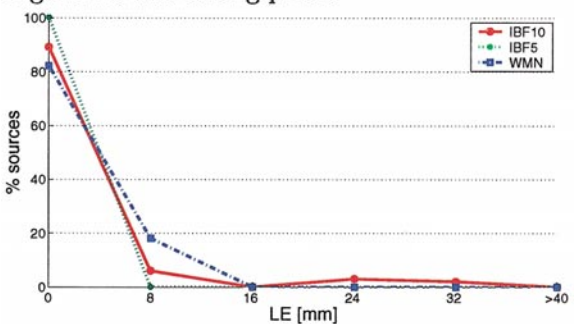**FIG. 3.** Location error (LE) and $\log_{10} \lambda$ for the three solutions in which location priors can be incorporated (IBF10, IBF5, WMN) for two levels of SNR and two levels of location priors, applied to the simple source simulated data with realistic noise.

TABLE 1

Maximum Localization Error (LE) of the IBF10, IBF5, WMN, and MS (with and without Noise Regularization) Solutions

fMRI priors	SNR	IBF10	IBF5	WMN	MS with regul.	MS w/o regul.
None	Infinite	20 (86%)	20 (85%)	44 (85%)	—	20 (91%)
	High	20 (81%)	28 (88%)	116 (82%)	20 (86%)	52 (81%)
	Low	28 (87%)	28 (83%)	116 (84%)	28 (89%)	116 (82%)
Weak	Infinite	4 (90%)	4 (95%)	28 (85%)	—	—
	High	12 (89%)	4 (96%)	116 (83%)	—	—
	Low	12 (90%)	4 (96%)	100 (82%)	—	—
Strong	Infinite	4 (97%)	4 (100%)	4 (82%)	—	—
	High	4 (89%)	4 (100%)	4 (82%)	—	—
	Low	4 (89%)	4 (100%)	112 (99%)	—	—

Note. The simulations were performed with two levels of SNR (low and high) and two levels of priors (weak and strong). The LE is expressed as the maximum LE allowed to recover at least 80% of the sources within this bound; the actual percentage of sources recovered within this "max LE" is in parentheses.

3.1.1. Solutions without Location Priors

Here the solutions were calculated without prior knowledge about the location of active sources. For each distributed source, the hyperparameter λ was estimated for the two levels of noise (low and high SNR). The LE and the estimated λ are shown in Fig. 2. As the values of λ vary over a large range, its logarithm in base 10 ($\log_{10} \lambda$) is displayed. For some simulations, when the SNR was high, the value of λ tended to 0. These values of λ were not included in the histogram.

A "maximum LE" (maxLE) was calculated from the LE of all the simulations, such that at least 80% of the sources were recovered within this bound. The maxLE, mean value of RMSE, and mean value of $\log_{10} \lambda$ are summarized in the top (Priors: None) of Tables 1, 2, and 3 for the two levels of SNR. When the value of λ tended to 0, these values were not entered into the mean of $\log_{10} \lambda$. In Table 3, the percentage of cases in which λ was greater than 0 is given in brackets.

Considering the LE, the IBF10, IBF5, and MS solutions, with noise regularization, gave estimates that were only slightly worse than in the noise-free case. The noise could thus be appropriately accommodated.

In contrast, the MS solution without regularization was clearly inefficient for low SNR and performed only slightly better for high SNR. The WMN solution, without location priors, remained sensitive to noise even with the inclusion of noise regularization and its performance was much worse than that in the noise-free simulations. The WMN solution without location priors was unable to localize the source of the signal.

The hyperparameter λ was on average smaller for the high SNR than the low SNR. As would be expected, with higher SNR, the constraints have less influence on the solution. At high SNR, in 26 to 30% of cases no noise regularization was required ($\lambda \rightarrow 0$) and the data were treated as effectively noise free. The main difference between the IBF10, the IBF5, and the WMN solutions is the relative level of spatial coherence imposed *a priori*: large, small, and none, respectively. On average the estimated hyperparameter λ was smaller for the IBF5 solution than for the WMN solution and smaller for the IBF10 solution than for the IBF5 solution. With its more restricted solution subspace, the IBF10 is less sensitive to noise, therefore less noise regularization was required. The values of λ obtained

TABLE 2

Mean Values of the Root Mean Square Error of the IBF10, IBF5, WMN, and MS
(with and without Noise Regularization) Solutions

fMRI priors	SNR	IBF10	IBF5	WMN	MS with regul.	MS w/o regul.
None	Infinite	146.8	128.1	48.0	—	216.2
	High	167.9	150.3	54.3	241.8	250.9
	Low	191.6	177.3	62.7	284.9	288.4
Weak	Infinite	109.4	66.7	62.7	—	—
	High	121.1	75.6	52.8	—	—
	Low	133.6	85.7	60.1	—	—
Strong	Infinite	86.8	36.0	50.7	—	—
	High	93.4	37.9	48.9	—	—
	Low	98.2	39.4	49.3	—	—

Note. The simulations were performed with two levels of SNR (low and high) and two levels of priors (weak and strong).

TABLE 3

Mean Values of the $\text{Log}_{10}\ \lambda$ of the IBF10, IBF5, WMN, and MS (with Noise Regularization) Solutions

fMRI priors	SNR	IBF10	IBF5	WMN	MS
None	High	0.53 (74%)	0.65 (70%)	0.81 (71%)	2.75 (100%)
	Low	1.43 (100%)	1.57 (100%)	1.71 (100%)	4.08 (100%)
Weak	High	0.54 (80%)	0.66 (72%)	0.76 (73%)	—
	Low	1.44 (100%)	1.59 (100%)	1.73 (100%)	—
Strong	High	0.52 (86%)	0.69 (89%)	0.82 (99%)	—
	Low	1.46 (100%)	1.64 (100%)	1.84 (100%)	—

Note. The simulations were performed with two levels of SNR (low and high) and two levels of priors (weak and strong). Simulations with no noise correspond to $\lambda = 0$. The mean of $\log_{10} \lambda$ takes into account only nonzero values of λ, the percentage of cases in which $\lambda \neq 0$ is in parentheses.

for the MS solution cannot be directly compared to those of the IBF and WMN solutions because the constraints imposed on the solutions are defined in completely different ways.

The mean RMSE was larger for the low SNR than the high SNR, indicating a more accurate reconstruction with higher SNR. As for the noise-free case, the RMSE was proportional to the level of smoothness imposed on the solution.

3.1.2. Solution with Accurate Location Priors

In this section, the solutions were calculated with accurate fMRI priors about the location of active sources. The IBF10, IBF5, and WMN solutions were assessed for three levels of noise, low, high, and infinite SNR, and two level of priors, weak and strong priors. The values obtained for LE and $\log_{10} \lambda$ are shown in Fig. 3. The maxLE and mean RMSE and $\log_{10} \lambda$ for each solution and each condition are summarized in the bottom (Priors: Weak and Strong) of Tables 1, 2, and 3.

The inclusion of location priors improved the accuracy of the reconstruction greatly. With strong priors, the LE of the IBF10, IBF5, and WMN solutions was approximately equivalent to the LE observed with the noise-free simulations. However, with weak priors, only the IBF5 solution performed as well as in the noise-free simulations. With weak priors, the LE of the IBF10 solution was reduced compared to the results obtained without priors but remained larger than that of the noise-free case. In the case of the WMN solution, the inclusion of the weak priors did not improve the accuracy of the solution.

The values of the hyperparameter λ were on average similar to (or slightly larger than) those obtained without priors and followed the same proportions: larger values for low levels of *a priori* smoothness and/or low SNR and smaller values with large amounts of *a priori* smoothness and/or high SNR. The percentage of cases in which the data were treated as noise free ($\lambda \rightarrow 0$) was reduced by the introduction of priors. This indicates that, even at high SNR, when location priors are included the constraint has a stronger influence on the solution than when no location priors are employed: when location priors are included, a larger component of the data **V** can be considered noise and the solution relies more on the constraints.

With the inclusion of priors, the RMSE was reduced in all cases. Nevertheless, the RMSE remained larger for the data with low SNR than with high SNR. The IBF5 solution with strong location priors and high SNR data provided the most accurate and focal solution.

3.2. Summary of Results

With realistic noise, the performance of the IBF solutions deteriorated slightly relative to simulations with noise-free data. The noise regularization adapted to the SNR of the data, i.e., the hyperparameter λ increased with noise level. As in the noise-free simulations, when no location priors were provided, the best localization (on average) was achieved with relatively large *a priori* smoothness constraints. Less smoothness was required, however, if location priors were provided.

With noise regularization the MS solution behaved like the IBF solutions without location priors: its performance was slightly reduced by the noise component and the noise regularization adapted to the SNR. On the other hand, if no noise regularization was applied, the LE of the MS solution increased dramatically, even for a high SNR. The WMN was also strongly affected by noise and capable of localizing sources only when strong location priors were embodied in the solution.

3.3. Discussion

One way to overcome the intrinsic limitations (on temporal or spatial resolution) of individual brain imaging modalities is to combine data obtained from different techniques within the same mathematical framework. Such integration should provide an optimal solution that harnesses the strengths of each technique. In this paper we have outlined a way in which structural and functional data are used as priors in the

estimation of EEG sources. Crucially, we have illustrated the role of ReML hyperparameter estimates in modeling the relative contributions of EEG residuals and MRI-based priors to the estimation.

The approach adopted in this work does not strictly follow the scheme of a distributed linear reconstruction. The spatial IBF, obtained from the anatomical information, and temporal IBF, specified by our knowledge of EEG ERP, allow constraints on the source localization to enter in two separable steps. First the dimensionality of the problem is reduced by projecting the solution space onto a subspace spanned by the spatial and temporal IBF. Second the subspace problem is solved using a regularized or Bayesian estimator.

The extraction of spatial and temporal IBFs is a key element in the approach presented here. Although an important step (see below), it can be time consuming and computationally demanding. Fortunately, the spatial IBFs need be calculated only once for a particular head and source model. Afterward, various hypotheses about the prior location of activation using different EEG time series can be studied with the same spatial IBF set but different "soft" constraints. The introduction of the spatial IBF offers a clear separation between the "hard" constraints relying on anatomy and other "soft" constraints that have intermediate precision. Temporal IBFs allow a more efficient characterization of the dynamics of the EEG over time and provide a way to constrain this estimate.

Hard constraints correspond here to a bipartition of the solution space into a subspace (spanned by the basis set) with "flat priors" (i.e., infinite variance) and its complement that has infinite precision (i.e., prior variance is 0). Soft constraints are specified, in the linear estimation scheme above, by prior densities in terms of their covariances or precisions that lie between zero and infinity.

General approaches to noise are usually empirical and proceed on a trial-and-error basis: The level of regularization is manually adapted such that the resulting solution and assumed noise component seem reasonable. Here, in contrast to these approaches, the ReML procedure was successfully applied to control the noise regularization by systematically estimating a regularization hyperparameter λ. With noise regularization, the IBF, MS, and MWN solutions behaved almost as they did with noise-free data. If no noise regularization was applied, i.e., $\lambda \to 0$, the MS solution was strongly affected by noise and was unable to provide any proper localization, even for a high SNR. The maximum smoothing constraint alone is thus not enough to control the effect of the noise.

Even at constant SNR and for the same levels of location priors, the values of λ vary over a wide range. As the noise component $\mathscr{E}$ was the same for all the simulations, the value of λ depends on the source con-

figuration and the distribution of potentials it generates over the scalp, therefore any fixed value of λ can lead to suboptimal solutions. For example, some sources may arise in cortical regions where priors may be specified very precisely leading to high values of λ. In other regions priors may be less informative, rendering a smaller value of λ more appropriate. The flexibility afforded by parameterizing the priors in terms of hyperparameters lies in being able to specify the form of the covariances ($\mathbf{C}_\epsilon$ and $(\mathbf{H}_i^t\mathbf{H}_i)^{-1}$) without fixing the exact values. These forms are scaled by the hyperparameters, which we estimated by ReML. The advantage of this approach is that the relative importance of the likelihood of, and priors on, the solution can be determined empirically. This affords the prior constraints the latitude to shape themselves in relation to observation error and each other.

It may seem intuitively implausible that both the model's parameters and the hyperparameters can be estimated from the data. A heuristic understanding is suggested by the iterative approach adopted in most ReML procedures. If one knew the hyperparameters, the posterior or conditional distribution of the parameters is readily estimated. Conversely, if the distribution of the model parameters was known, the hyperparameters could be estimated using the variance component estimation after augmenting the model as described in the Appendix. This is equivalent to taking the average value of the hyperparameters, integrated over the conditional distribution of the parameters. By estimating the parameters given the current value of the hyperparameters and then the hyperparameters given the new conditional density of the parameters, one converges on the ReML estimate for the hyperparameters and the conditional distribution of the parameters of interest. This a powerful approach that calls on an iterative scheme. A fundamental advantage of using the spatial and temporal IBF is that the iterative ReML algorithm is rendered much more tractable. Indeed, at each iteration a problem of size $N_{k_s} \times N_{k_t}$ is solved instead of $N_d \times N_t$, where $N_{k_s} \ll N_d$ and $N_{k_t} \ll N_t$. This requires much less memory and computational time.

APPENDIX: RESTRICTED MAXIMUM LIKELIHOOD ESTIMATES

Maximum Likelihood Solution

Consider a linear stochastic model of the form

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{r}, \qquad (22)$$

where $\mathbf{A}$, the model or design matrix, is of size $m \times n$; $\mathbf{x}$, the unknown vector, is of size $n \times 1$; $\mathbf{b}$, the data vector, and $\mathbf{r}$, the residual or error vector, are of size

$m \times 1$; and $\mathbf{C}$, the covariance matrix of $\mathbf{r} = \mathcal{N}(0, \mathbf{C})$, is of size $m \times m$. The normal equations of this model are

$$(\mathbf{A}^t\mathbf{C}^{-1}\mathbf{A})\mathbf{x} = \mathbf{A}^t\mathbf{C}^{-1}\mathbf{b}. \tag{23}$$

By solving the normal equations, the "best linear unbiased estimate" is obtained,

$$\hat{\mathbf{x}} = (\mathbf{A}^t\mathbf{C}^{-1}\mathbf{A})^-\mathbf{A}^t\mathbf{C}^{-1}\mathbf{b}, \tag{24}$$

where, for any matrix B, B^- denotes an arbitrary generalized inverse of B, i.e., any solution to $BB^-B = B$.

This is equivalent to maximizing the objective likelihood function (Patterson and Thompson, 1971)

$$p(\mathbf{b}|\mathbf{x}) \propto |\mathbf{C}|^{-1/2}e^{-1/2(\mathbf{b}-\mathbf{Ax})^t\mathbf{C}^{-1}(\mathbf{b}-\mathbf{Ax})}, \tag{25}$$

which is equivalent to minimizing the function

$$\begin{aligned}F(\mathbf{b}; \mathbf{x}) &= -2\log(p(\mathbf{b}|\mathbf{x})) \\ &= \log|\mathbf{C}| + (\mathbf{b}-\mathbf{Ax})^t\mathbf{C}^{-1}(\mathbf{b}-\mathbf{Ax}) + \text{const.}\end{aligned} \tag{26}$$

The variance–covariance matrix $\mathbf{C}$ is necessary to weight the observations $\mathbf{b}$ according to their variance and to account for their covariance. There may be a model for this matrix, depending on hyperparameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots]$, but its exact value is not necessarily known *a priori* and needs to be estimated as well as $\mathbf{x}$.

ReML Solution

Assume that the variance–covariance matrix $\mathbf{C}$ is a function of the unknown hyperparameters θ_i, $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta})$ and $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots]$. There is a scheme that allows the simultaneous estimation of $\mathbf{x}$ and $\mathbf{C}(\boldsymbol{\theta})$, which properly takes into account the loss of degrees of freedom in the model incurred from estimating $\mathbf{x}$ (Patterson and Thompson, 1971) when $\mathbf{C}(\boldsymbol{\theta})$ is calculated.

Starting estimates are initially assigned to $\boldsymbol{\theta}$. With the current estimate $\hat{\boldsymbol{\theta}}$, $\hat{\mathbf{x}}$ is estimated by maximizing the first likelihood function (Eq. (25)). Then an updated estimate of $\boldsymbol{\theta}$ is calculated from the current value $\hat{\mathbf{x}}$. The procedure is repeated until convergence of both $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\theta}}$.

Harville (1974) showed that this iterative procedure maximizes the following function, which has subsequently been named restricted maximum likelihood objective function:

$$p(\mathbf{b}|\mathbf{x}, \boldsymbol{\theta}) = \sqrt{\frac{|\mathbf{A}^t\mathbf{A}|}{(2\pi)^{m-n}|\mathbf{C}(\boldsymbol{\theta})|\,|\mathbf{A}^t\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{A}|}} \\ \times e^{-1/2(\mathbf{b}-\mathbf{Ax})^t\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{b}-\mathbf{Ax})}. \tag{27}$$

Maximizing Eq. (27) is equivalent to minimizing

$$\begin{aligned}F(\mathbf{b}; \mathbf{x}, \boldsymbol{\theta}) &= -2\log(p(\mathbf{b}|\mathbf{x}, \boldsymbol{\theta})) \\ &= \log|\mathbf{A}^t\mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{A}| + \log|\mathbf{C}(\boldsymbol{\theta})| \\ &\quad + (\mathbf{b}-\mathbf{Ax})^t\mathbf{C}(\boldsymbol{\theta})^{-1}(\mathbf{b}-\mathbf{Ax}) + \text{const.}\end{aligned} \tag{28}$$

There exists a general iterative approach for estimating $\mathbf{x}$ and $\boldsymbol{\theta}$ but a linear parameterization of $\mathbf{C}$, i.e., $\mathbf{C}(\boldsymbol{\theta}) = \sum \theta_i\mathbf{G}_i$ where $\mathbf{G}_i$ are $n \times n$ symmetric matrices whose elements are known, leads to a much simpler and less computationally demanding approach (Harville, 1977).

Simple ReML Solution

One such case of linear parameterization of $\mathbf{C}$ involves a diagonal covariance matrix with each element parameterized by only one element of $\boldsymbol{\theta}$. With only two hyperparameters $\boldsymbol{\theta} = [\theta_1^2\ \theta_2^2]$, the matrices and vectors of Eq. (22) can be separated into two parts,

$$\begin{bmatrix}\mathbf{b}_1 \\ \mathbf{b}_2\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1 \\ \mathbf{A}_2\end{bmatrix}\mathbf{x} + \begin{bmatrix}\mathbf{r}_1 \\ \mathbf{r}_2\end{bmatrix} \tag{29}$$

and

$$\text{var}\left(\begin{bmatrix}\mathbf{r}_1 \\ \mathbf{r}_2\end{bmatrix}\right) = \begin{bmatrix}\theta_1^2\mathbf{I}_{m_1} & 0 \\ 0 & \theta_2^2\mathbf{I}_{m_2}\end{bmatrix} = \mathbf{C}(\boldsymbol{\theta}), \tag{30}$$

with m_1, the number of rows in $\mathbf{A}_1$ and $\mathbf{b}_1$, and m_2, the number of rows in $\mathbf{A}_2$ and $\mathbf{b}_2$, such that $m_1 + m_2 = m$.

The iterative scheme involves assigning (nonzero) starting estimates for $\boldsymbol{\theta}$ and estimating $\mathbf{x}$ with relation (24):

$$\hat{\mathbf{x}} = (\theta_1^{-2}\mathbf{A}_1^t\mathbf{A}_1 + \theta_2^{-2}\mathbf{A}_2^t\mathbf{A}_2)^{-1}(\theta_1^{-2}\mathbf{A}_1^t\mathbf{b}_1 + \theta_2^{-2}\mathbf{A}_2^t\mathbf{b}_2). \tag{31}$$

Updated values of the hyperparameters are then obtained from the residuals,

$$\mathbf{r} = \begin{bmatrix}\mathbf{r}_1 \\ \mathbf{r}_2\end{bmatrix} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}, \tag{32}$$

and take into account the loss of degrees of freedom resulting from the estimation of $\hat{\mathbf{x}}$:

$$p_1 = \text{trace}((\theta_1^{-2}\mathbf{A}_1^t\mathbf{A}_1 + \theta_2^{-2}\mathbf{A}_2^t\mathbf{A}_2)^{-1}\theta_1^{-2}\mathbf{A}_1^t\mathbf{A}_1) \tag{33a}$$

$$p_2 = \text{trace}((\theta_1^{-2}\mathbf{A}_1^t\mathbf{A}_1 + \theta_2^{-2}\mathbf{A}_2^t\mathbf{A}_2)^{-1}\theta_2^{-2}\mathbf{A}_2^t\mathbf{A}_2) = n - p_1. \tag{33b}$$

The new estimates of θ are

$$\hat{\theta}_1^2 = \frac{\mathbf{r}_1^t \mathbf{r}_1}{m_1 - p_1}, \tag{34a}$$

$$\hat{\theta}_2^2 = \frac{\mathbf{r}_2^t \mathbf{r}_2}{m_2 - p_2}. \tag{34b}$$

The iterative procedure continues by using these new hyperparameter estimates to reestimate $\mathbf{x}$ and then from the updated $\hat{\mathbf{x}}$, a new set of θ. Eventually the algorithm converges to a stable solution.

The hyperparameters θ_1^2 and θ_2^2 are the variance components of each subproblem. If the algorithm is started with strictly positive values, then at no point can the values of θ even become negative and, in fact, will never reach 0. Obviously the algorithm should not be started with 0 or negative starting estimates.

Simple ReML Solution in a Noninstantaneous Problem

Consider the case of the linear stochastic model Eq. (22) with noninstantaneous data, i.e., the $m \times 1$ data vector $\mathbf{b}$ (resp. residual vector $\mathbf{r}$) is replaced by a $m \times q$ matrix $\mathbf{B}$ (resp. $\mathbf{R}$) and the $n \times 1$ vector $\mathbf{x}$ by a $n \times q$ matrix $\mathbf{X}$:

$$\mathbf{B} = \mathbf{AX} + \mathbf{R} \leftrightarrow \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{X} + \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix}. \tag{35}$$

If the covariance matrix $\mathbf{C}$ of the residuals $\mathbf{R}$ still depends linearly on two hyperparameters $\theta = [\theta_1^2 \ \theta_2^2]$ as in Eq. (30), then a similar iterative scheme can be applied to estimate both $\mathbf{X}$ and θ. With starting estimates for θ, a first solution for $\mathbf{X}$ is calculated with Eq. (31):

$$\hat{\mathbf{X}} = (\theta_1^{-2} \mathbf{A}_1^t \mathbf{A}_1 + \theta_2^{-2} \mathbf{A}_2^t \mathbf{A}_2)^{-1} (\theta_1^{-2} \mathbf{A}_1^t \mathbf{B}_1 + \theta_2^{-2} \mathbf{A}_2^t \mathbf{B}_2). \tag{36}$$

The values of θ are then updated from the residuals,

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{1,1} & \mathbf{r}_{1,2} & \cdots & \mathbf{r}_{1,q} \\ \mathbf{r}_{2,1} & \mathbf{r}_{2,2} & \cdots & \mathbf{r}_{2,q} \end{bmatrix} = \mathbf{B} - \mathbf{A}\hat{\mathbf{X}}, \tag{37}$$

using relations similar to Eq. (34),

$$\hat{\theta}_1^2 = \frac{\sum_{i=1}^q \mathbf{r}_{1,i}^t \mathbf{r}_{1,i}}{q(m_1 - p_1)} \tag{38a}$$

$$\hat{\theta}_2^2 = \frac{\sum_{i=1}^q \mathbf{r}_{2,i}^t \mathbf{r}_{2,i}}{q(m_2 - p_2)}, \tag{38b}$$

where p_1 and p_2 are calculated with Eqs. (33). The iterative procedure continues, as under Simple ReML Solution, by reestimating $\mathbf{X}$ with these new values of θ. The process is repeated until convergence of the values of $\mathbf{X}$ and θ.

REFERENCES

Aine, C., Huang, M., Stephen, J., and Christner, R. 2000. Multistart algorithms for MEG empirical data analysis reliably characterize locations and time courses of multiple sources. *NeuroImage* **12:** 159–172, doi:10.1006/nimg.2000.0616.

Ashburner, J., and Friston, K. J. 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* **6:** 209–217.

Backus, G. E., and Gilbert, J. F. 1970. Uniqueness in the inversion of inaccurate gross earth data. *Philos. Trans. R. Soc.* **266:** 123–192.

Buchner, H., Knoll, G., Fuchs, M., Rienäcker, A., Beckmann, R., Wagner, M., Silny, J., and Pesch, J. 1997. Inverse localization of electric dipole current sources in finite element models of the human head. *Electroencephalogr. Clin. Neurophysiol.* **102:** 267–278.

Dale, A. M., and Sereno, M. I. 1993. Improved localisation of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J. Cognit. Neurosci.* **5:** 162–176.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., and Halgren, E. 2000. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neurone* **26:** 55–67.

de Munck, J. C. 1992. A linear discretization of the volume conductor boundary integral equation using analytically integrated elements. *IEEE Trans. Biomed. Eng.* **39:** 986–990.

Engl, H. W., and Grever, W. 1994. Using the L-curve for determining optimal regularization parameters. *Numerische Math.* **69:** 25–31.

Ferguson, A. S., and Stroink, G. 1997. Factors affecting the accuracy of the boundary element method in the forward problem. I. Calculating surface potential. *IEEE Trans. Biomed. Eng.* **44:** 1139–1155.

Geselowitz, D. B. 1967. On bioelectric potentials in an inhomogeneous volume conductor. *Biophys. J.* **7:** 1–11.

Grave de Peralta Menendez, R., and Gonzalez Andino, S. 1998. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans. Biomed. Eng.* **45:** 440–448.

Grave de Peralta Menendez, R., and Gonzalez Andino, S. 1999. Backus and Gilbert method for vector fields. *Hum. Brain Mapping* **7:** 161–165.

Hämäläinen, M. S., and Ilmoniemi, R. J. 1994. Interpreting magnetic fields of the brain: Minimum norm estimates. *Med. Biol. Eng. Comput.* **32:** 35–42.

Hämäläinen, M. S., and Sarvas, J. 1989. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Trans. Biomed. Eng.* **36:** 165–171.

Hämäläinen, M. S., 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:** 413–497.

Hansen, P. C. 1992. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* **34:** 561–580.

Harville, D. A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika* **61:** 383–385.

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

Ioannides, A. A., Bolton, J. P. R., and Clarke, C. J. S. 1990. Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Probl.* **6:** 523–543.

Liu, A. K., Belliveau, J. W., and Dale, A. M. 1998. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA* **95:** 8945–8950.

Meijs, J. W. H., Weier, O. W., Peters, M. J., and van Oosterom, A. 1989. On the numerical accuracy of the boundary element method. *IEEE Trans. Biomed. Eng.* **36:** 1038–1049.

Miltner, W., Braun, C., Johnson, R., Simpson, G. V., and Ruchkni, D. S. 1994. A test of brain electrical source analysis (BESA): A simulation study. *Electroencephalogr. Clin. Neurophysiol.* **91:** 295–310.

Nunez, P. L. 1981. *Electric Fields of the Brain: The Neurophysics of EEG.* Oxford Univ. Press, New York.

Pascual-Marqui, R. D. 1995. Reply to comments by M. Hämäläinen, R. Ilmoniemi, and P. Nunez. In *Source Localization: Continuing Discussion of the Inverse Problem* (W. Skrandies, Ed.), ISBET Newsletter, No. 6, pp. 16–28.

Pascual-Marqui, R. D. 1998. *Low Resolution Brain Electromagnetic Tomography (LORETA).* http://www.unizh.ch/keyinst/NewLORETA/LORETA01.htm.

Pascual-Marqui, R. D. 1999. Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.* **1:** 75–86.

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. 1994. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* **18:** 49–65.

Patterson, H. D., and Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58:** 545–554.

Phillips, C., Rugg, M. D., and Friston, K. J. 2002. Anatomically informed basis functions for EEG source localisation: Combining functional and anatomical constraints. *NeuroImage,* in press.

Rugg, M. D. 1998. Convergent approaches to electrophysiological and hæmodynamics investigations of memory. *Hum. Brain Mapping* **6:** 394–398.

Sarvas, J. 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* **32:** 11–22.

Scherg, M., and Ebersole, J. S. 1994. Brain source imaging of focal and multifocal epileptiform EEG activity. *Clin. Neurophysiol.* **24:** 51–60.

Scherg, M., Bast, T., and Berg, P. 1999. Multiple source analysis of interictal spikes: Goals, requirements, and clinical value. *J. Clin. Neurophysiol.* **16:** 214–224.

Schlitt, H. A., Heller, L., Aaron, R., Best, E., and Ranken, D. M. 1995. Evaluation of boundary element methods for the EEG forward problem: Effect of linear interpolation. *IEEE Trans. Biomed. Eng.* **42:** 52–58.

Tikhonov, A. N., and Arsenin, V. Y. 1977. *Solutions of Ill-Posed Problems.* Wiley, New York.

Uutela, K., Hämäläinen, M. S., and Somersalo, E. 1999. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10:** 173–180, doi:10.1006/nimg.1999.0454.

Wellcome Department of Cognitive Neurology. 1999. *Statistical Parametric Mapping, SPM'99.* http://www.fil.ion.ucl.ac.uk/spm/spm99.html.