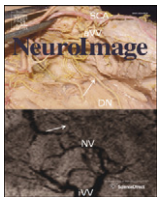
NeuroImage 61 (2012) 1364–1382

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: Depth localization and source separation for focal primary currents

Felix Lucka[a,b,*], Sampsa Pursiainen[c], Martin Burger[a], Carsten H. Wolters[b]

[a] Institute for Computational and Applied Mathematics, University of Muenster, Germany
[b] Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Germany
[c] Department of Mathematics, Tampere University of Technology, Finland

ARTICLE INFO

Article history:
Accepted 7 April 2012
Available online 17 April 2012

Keywords:
EEG inverse problem
Current density reconstruction
Hierarchical Bayesian modeling
Fully-Bayesian inference
Depth localization
Wasserstein distance

ABSTRACT

The estimation of the activity-related ion currents by measuring the induced electromagnetic fields at the head surface is a challenging and severely ill-posed inverse problem. This is especially true in the recovery of brain networks involving deep-lying sources by means of EEG/MEG recordings which is still a challenging task for any inverse method. Recently, hierarchical Bayesian modeling (HBM) emerged as a unifying framework for current density reconstruction (CDR) approaches comprising most established methods as well as offering promising new methods. Our work examines the performance of fully-Bayesian inference methods for HBM for source configurations consisting of few, focal sources when used with realistic, high-resolution finite element (FE) head models. The main foci of interest are the correct depth localization, a well-known source of systematic error of many CDR methods, and the separation of single sources in multiple-source scenarios. Both aspects are very important in the analysis of neurophysiological data and in clinical applications. For these tasks, HBM provides a promising framework and is able to improve upon established CDR methods such as minimum norm estimation (MNE) or sLORETA in many aspects. For challenging multiple-source scenarios where the established methods show crucial errors, promising results are attained. Additionally, we introduce Wasserstein distances as performance measures for the validation of inverse methods in complex source scenarios.

Introduction

Electroencephalography (*EEG*) and *magnetoencephalography* (*MEG*) recordings are used in a wide range of applications today and range from clinical testing to cognitive science (Niedermeyer and Lopez da Silva, 2004). One aim in using EEG and MEG is to reconstruct brain activity by means of non-invasive measurements of the associated *bioelectromagnetic* fields. This task involves challenging mathematical problems. Simulating the field distribution on the head surface for a given current source in the brain is called the *EEG/MEG forward problem* (e.g., Hämäläinen et al., 1993; Sarvas, 1987). The reconstruction of the so-called *primary* or *impressed* currents (a simplified source model, see de Munck et al., 1988; Hämäläinen et al., 1993; Sarvas, 1987) is called the *EEG/MEG inverse problem*. In its generic formulation, the inverse problem lacks a unique solution, and infinitely many source configurations, often with extremely different properties, can explain the measured fields. All inverse methods rely on the use of a priori *information* on the source activity to choose a particular solution from the set of likely solutions. This a priori information can reflect computational constraints as well as neurological considerations. Nevertheless, because the problem is heavily under-determined, the results from different methods for the same measurement data can still differ considerably. Consequently, most methods work well for certain source configurations while failing to recover other configurations. Therefore, a careful examination of the performance of the methods for different source scenarios is mandatory. This article focuses on the results of estimation methods based on a certain class of inference strategies called *hierarchical Bayesian modeling* (*HBM*). While we investigate source scenarios including multiple focal primary currents that occur, e.g., in the analysis of evoked potentials (Scherg and Buchner, 1993; Parkkonen et al., 2009) and specific scenarios encountered in presurgical epilepsy diagnosis (Rampp and Stefan, 2007; Rullmann et al., 2009; Stefan et al., 2003), the framework easily extends to recover spatially more distributed sources encountered, e.g., in cognitive neuroscience (Hämäläinen et al., 1993) or in other presurgical epilepsy diagnosis scenarios (Ebersole and Ebersole, 2010; Rampp and Stefan, 2007; Stefan et al., 2003; Tao et al., 2005). This work comprises the results from a diploma thesis, Lucka (2011). In the following sections, we will outline the development of HBM for EEG/MEG *current density reconstruction* (*CDR*) and motivate our interest in scenarios where the source activity results from networks of few and focal sources.

Inverse methods for EEG/MEG

From a mathematical point of view, the EEG/MEG inverse problem is *severely-ill-posed* (Engl et al., 1996; Hämäläinen et al., 1993; Lucka, 2011). As a practical consequence, a variety of different approaches

* Corresponding author at: Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Malmedyweg 15 D-48149 Münster, Germany.
E-mail address: felix.lucka@uni-muenster.de (F. Lucka).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.04.017

exist that aim to reconstruct solutions that reflect certain a priori *information*. First, a classification can be made in *focal current modeling, spatial scanning/beamforming* and *distributed current modeling*. Focal current modeling attempts to reconstruct the real current using a small number of *equivalent current dipoles* with arbitrary locations and orientations (Jun et al., 2008; Mosher et al., 1992; Scherg and Cramon, 1985). When the number of sources is unknown or the current distribution might have a larger spatial extent, focal current models are not suitable. Spatial scanning methods/beamforming repeatedly optimize the estimate at a single location or a small region while suppressing crosstalk from other areas (Dalal et al., 2008; Sekihara and Nagarajan, 2008). In distributed current models, the current is discretized using a large number of focal elementary sources with fixed locations and orientations, which is called *current density reconstruction* (*CDR*). Then, a priori information on the global properties of the solution is incorporated (e.g., *minimum norm estimation*, Hämäläinen and Ilmoniemi, 1994).

Concerning the CDR methods, two main concepts dominate the formulation of how the a priori information is introduced: *Regularization* (see Sarvas, 1987 for the introduction to EEG/MEG and Engl et al., 1996 for a general reference) and *Bayesian inference* (see Hämäläinen et al., 1987 for the introduction to EEG/MEG and Kaipio and Somersalo, 2005 for a general reference).

The focus of our work is on the most recent branch of Bayesian inference for CDR called hierarchical Bayesian modeling, for which we will examine *fully-Bayesian* inference methods (in contrast to, e.g., *variational* or *semi-Bayesian* inference methods, see Wipf and Nagarajan, 2009).

Brain networks involving deep-lying sources

The location of the source space nodes is a crucial choice for CDRs. First, high-resolution structural MRI scans have to be taken of the cortex where the neural generators of the EEG/MEG signal are located (Nunez and Srinivasan, 2005). Due to the deep but thin sulci and strong folding of the cortex, sophisticated segmentation algorithms are required to process these data. Instead, often, only a flattened and smoothed representation of the cortical surface is used, which does not include the deep-lying gray matter areas or areas encased by white matter, e.g., the insular, the cingulate cortex, the hippocampus or the thalamus. Working with such surface representations is reasonable and advantageous for a wide range of experimental designs. Nevertheless, active brain networks often involve deep-lying sources as well (Dalal et al., 2010; Parkkonen et al., 2009; Sander et al., 2010; Santiuste et al., 2008; Scherg and Buchner, 1993). One example is the different components active in the response to somatosensory stimuli (Scherg and Buchner, 1993). Another example is given by the potentials and fields generated by the auditory pathway (Parkkonen et al., 2009; Sander et al., 2010). The analysis of both networks is a common clinical application of EEG/MEG. To recover such networks, a complete representation of the gray matter compartment by source space nodes is mandatory.

By accounting for the complete gray matter, many more deep-lying locations form the source space and a phenomenon called *depth bias* is of fundamental importance: Many inverse methods fail to reconstruct deep-lying sources at the correct depth; rather, the sources are reconstructed too close to the skull. This is a well-known systematic error (e.g., Ahlfors et al., 1992; Gencer and Williamson, 1998; Wang et al., 1992) and was subject of many studies (e.g., Fuchs et al., 1999; Grave de Peralta et al., 2009; Greenblatt et al., 2005; Ioannides et al., 1990; Lin et al., 2006; Pascual-Marqui, 1999, 2002; Sekihara et al., 2005; Wagner et al., 2004). The depth bias can be a crucial error, e.g., in the presurgical functional mapping of the eloquent cortex (Schiffbauer et al., 2002). Another effect related to the depth bias is the *masking* of deep-lying sources by superficial sources. If the true source configuration consists of multiple and spatially separated sources with different depths, many inverse methods only recover the sources close to the skull (see, e.g., Wagner et al., 2004). This effect complicates the analysis of networks of interacting brain areas which is a recent topic of interest in brain imaging (Kiebel et al., 2009). Furthermore, several clinical applications require a correct detection and separation of multiple sources, e.g., the reconstruction of the auditory pathway (Parkkonen et al., 2009) or specific cases of epileptiform discharges (Hufnagel et al., 1994; Janszky et al., 2000).

Contributions and structure of this study

This article examines, in a systematic way, fully-Bayesian inference for HBM for a single-time-point analysis of the source scenarios described above (the reasons for not using temporal information will be explained later). In Bayesian formulation of the static inverse problem section, we will outline CDR approaches from the perspective of Bayesian inference. This will then lead us to the hierarchical Bayesian modeling in Hierarchical modeling in EEG/MEG section, in which we will describe the *fully-Bayesian* inference methods (which we will call *CM* and *MAP1*) and propose improved full-MAP estimation methods in Inference for Hierarchical models and Algorithms for fully-Bayesian inversion sections, which we will call *MAP2* and *MAP3*. In Validation means and inverse crimes section, a new performance measure called the *earth mover's distance* (*EMD*) will be introduced, which is required for an appropriate validation of the inverse methods in complex source scenarios. Results section describes the setting and results of the simulation studies. For the forward computation, we will use a realistic, high-resolution finite element (FE) head model. In Discussion section, the results, limitations and future directions of our research are discussed, and Conclusions section contains the final conclusions.

Methods

Bayesian formulation of the static inverse problem

We will briefly introduce the Bayesian formulation of the static inverse problem, revisit some commonly known inverse methods and introduce the hierarchical model that we will study here. More details on the concepts of Bayesian modeling can be found in Kaipio and Somersalo (2005) and Lucka (2011), Chapter 2. Subsequently, all random variables are denoted by upper case letters (e.g., X), their corresponding concrete realizations by lower case letters (e.g., $X = x$) and their probability density functions by $p(x)$. Assume that we have k locations r_i, $i = 1,\ldots,k$ within the brain and place d focal elementary sources with different orientations at each of these locations. A current distribution can be described as a linear combination of the elementary sources and the corresponding coefficients $s \in \mathbb{R}^n$ (where $n := d \cdot k$) will become the main parameters of interest in the following (also called *sources*). The measurements $b \in \mathbb{R}^m$ at the m sensors caused by s can be calculated via:

$$b = L\,s, \tag{1}$$

where $L \in \mathbb{R}^{m \times n}$ denotes the *lead-field* or *gain* matrix (see Hämäläinen and Ilmoniemi, 1984; Hämäläinen et al., 1993; Sarvas, 1987). For calculating the entries of the lead-field matrix, one needs to solve the forward problem, which includes head and source modeling and an appropriate (numerical) solution scheme (see Head model and source space section). The ill-posed nature of the inverse problem is reflected in L. Because $m \ll n$, Eq. (1) is under-determined, and furthermore, L is ill-conditioned.

Central to the Bayesian approach is to account for every uncertainty concerning the value of a variable explicitly. The variable is modeled as a *random* variable, but this randomness is *not* a property of the objects itself; rather, it reflects our *lack of information* about its concrete value. In our situation, we first model the (additive) measurement noise using a Gaussian random variable $\varepsilon \sim N(0, \Sigma_\varepsilon)$. For

simplicity, we assume $\Sigma_\varepsilon = \sigma^2 I_m$ here, where $\sigma > 0$ is the standard deviation of the single-channel Gaussian noise and I_m is the m-dim identity matrix. This leads to the following *likelihood model*:

$$B = LS + \varepsilon \tag{2}$$

Note that we changed b and s to the random variables B and S, respectively, as well. The conditional probability density of B given S is determined by Eq. (2) and is, thus, called the *likelihood* density:

$$p_{li}(b|s) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} \exp\left(-\frac{1}{2\,\sigma^2}||b - L\,s||_2^2\right) \tag{3}$$

Due to the ill-posed nature of Eq. (1), inference about S given B is not feasible by Eq. (3). We need to encode a priori information about S in its density $p_{pr}(s)$ which is hence called the *prior*. Then, the model can be inverted via Bayes' rule:

$$p_{post}(s|b) = \frac{p_{li}(b|s)p_{pr}(s)}{p(b)}. \tag{4}$$

The conditional density of S given B is called the *posterior*. In Bayesian inference, this density is the complete solution to the inverse problem. The term $p(b)$ is called the *model evidence* and has its own importance, as it can be used to perform *model averaging* or *model selection*. For interesting applications in EEG/MEG see Sato et al. (2004), Trujillo-Barreto et al. (2004), and Henson et al. (2009b,a, 2010). For our investigations, $p(b)$ is just a normalizing constant that is not important for the inference strategies presented in this study. A common way of exploiting the information contained in the posterior is to infer a point estimate for the value of S. There are two popular choices. The first method, called the *maximum a posteriori* estimate (*MAP*), is the highest *mode* of the posterior, whereas the second method, called the *conditional mean* estimate (*CM*), is the mean or expected value of the posterior:

$$\hat{s}_{MAP} := \underset{s\in\mathbb{R}^n}{\arg\max}\, p_{post}(s|b) \tag{5}$$

$$\hat{s}_{CM} := \mathbb{E}[s|b] = \int s\; p_{post}(s|b)ds. \tag{6}$$

Practically, computing the MAP estimate is a high-dimensional *optimization* problem, whereas the CM estimate is a high-dimensional *integration* problem.

To revisit some commonly known inverse methods, we consider *Gibbs distributions* as priors:

$$p_{pr}(s) \propto \exp\left(-\frac{\lambda}{2\sigma^2}\mathcal{P}(s)\right) \tag{7}$$

Here, $\mathcal{P}(s)$ is an energy functional that penalizes unwanted features of s, and $\lambda > 0$ is a scaling parameter that is called the *regularization parameter*. Now, after suppressing terms not dependent on s, the MAP estimate is given by

$$\hat{s}_{MAP} := \underset{s\in\mathbb{R}^n}{\arg\max}\left\{\exp\left(-\frac{1}{2\sigma^2}||b - Ls||_2^2 + \frac{\lambda}{2\sigma^2}\mathcal{P}(s)\right)\right\} \tag{8}$$

$$= \underset{s\in\mathbb{R}^n}{\arg\min}\left\{||b - Ls||_2^2 + \lambda\mathcal{P}(s)\right\} \tag{9}$$

This is a *Tikhonov-type* regularization of Eq. (1) (Engl et al., 1996). For EEG/MEG, the choice of $\mathcal{P}(s) = ||s||_2^2$, which corresponds to a white noise Gaussian prior, yields the *minimum norm estimate* (*MNE*, Hämäläinen and Ilmoniemi, 1994). $\mathcal{P}(s) = ||\Sigma_s^{-1/2}s||_2^2$ corresponds to a general Gaussian prior with covariance Σ_s and yields the *weighted minimum norm estimate* (*WMNE*, Dale and Sereno, 1993). Multiple *depth-weighting* matrices have been introduced and chosen to reduce the depth bias of the MNE (Fuchs et al., 1999; Ioannides et al., 1990). We will examine ℓ_2 weighting (Fuchs et al., 1999) and regularized ℓ_∞ weighting (Fuchs et al., 1999):

$$\Sigma_s^{\ell_2} = \underset{i=1,\dots,n}{diag}\left(\left(||L_{(\cdot,i)}||_2^2\right)^{-1}\right);$$
$$\Sigma_s^{\ell_{\infty,reg}} = \underset{i=1,\dots,n}{diag}\left(\frac{\chi_i^2}{\left(\chi_i^2 + \beta^2\right)^2}\right),$$

$$\text{with}\,\chi_i = ||L_{(\cdot,i)}||_\infty;\quad \beta = \max(\chi)\cdot\frac{m\sigma^2}{||b||_2^2}.$$

The well known sLORETA method (Pascual-Marqui, 2002) relies on the same prior model as the MNE: The MAP estimate (which is the MNE) is standardized by the posterior covariance, which yields a pseudo statistic of F-type for the source amplitude at a source space node (Pascual-Marqui, 2002). More methods relying on the formulation of Eq. (9) are listed on page 9 in Lucka (2011).

Hierarchical modeling in EEG/MEG

Brain activity is a complex process comprising many different spatial patterns. No fixed prior can model all of these phenomena without becoming *uninformative*, i.e., not able to deliver the required additional a priori information. This problem can be solved by introducing an adaptive, data-driven element into the estimation process. The idea of *hierarchical Bayesian models* (*HBM*) is to let the same data determine the appropriate model for the inversion of these data. By extending the model by a new level of inference, the prior on S is not fixed but is random and is determined by the values of additional parameters $\gamma \in \mathbb{R}^h$, called *hyperparameters*. These parameters follow an a priori assumed distribution (the *hyperprior*) and are subject to estimation schemes as well. As this whole construction follows a top-down scheme, it is called *hierarchical* modeling:

$$p(s,\gamma) = p_{pr}(s|\gamma)\;\; p_{hpr}(\gamma) \tag{10}$$

$$\Rightarrow p_{pr}(s) = \int p_{pr}(s|\gamma)p_{hpr}(\gamma)d\gamma \tag{11}$$

$$\Rightarrow p_{post}(s,\gamma|b) \propto p_{li}(b|s)p_{pr}(s|\gamma)p_{hpr}(\gamma). \tag{12}$$

We refer to MacKay (2003) and Gelman et al. (2003) for a general reference on hierarchical Bayesian modeling. The hierarchical model used in most methods for EEG/MEG relies on a special construction of the prior, and this construction is called a *Gaussian scale mixture*, or a *conditionally Gaussian hypermodel* (Calvetti et al., 2009; Wipf and Nagarajan, 2009). $p_{pr}(s|\gamma)$ is a Gaussian density with a mean of zero and a covariance determined by γ:

$$S|\gamma \sim \mathcal{N}(0, \Sigma_s(\gamma)) \tag{13}$$

The total *source covariance* Σ_s is a weighted sum of the *covariance components* C_i belonging to a predefined set $\mathcal{C} \subset \mathbb{R}^{n\times n}$ of symmetric, positive, semi-definite matrices. The weighting between them is controlled by a (positive) hyperparameter $\gamma \in \mathbb{R}^h$:

$$\begin{aligned}&\Sigma_s(\gamma) = \sum_{i=1}^{h}\gamma_i C_i \qquad \text{where} \qquad C_i \in \mathcal{C}\\ &\Rightarrow p_{pr}(s|\gamma) = (2\pi)^{-n/2}|\Sigma_s|^{-1/2}\exp\left(-\frac{1}{2}\left(s\Sigma_s^{-1}s^t\right)\right).\end{aligned} \tag{14}$$

The first important choice is in choosing an appropriate set $\mathcal{C}$. A variety of approaches that encode different a priori information on the spatial source covariance pattern have been proposed, e.g., spatial smoothness components (Mattout et al., 2006; Phillips et al., 2005) or *multiple sparse priors* (Friston et al., 2008). A recent overview is given

in Lucka (2011), page 19. In this study, we will rely on single-location priors (I_d denotes the identity matrix in d dimensions):

$$\mathcal{C} = \left\{ e_i e_i^t \otimes I_d, i = 1, \ldots, k \right\},$$

Therefore the number of hyperparameters equals the number of source locations: $h = k$. For instance, if $d = 3$, C_i is a matrix where only the entries (i, i), $(k+i, k+i)$ and $(2k+i, 2k+i)$ have the value 1 whereas all others are 0. As compared with the minimum norm estimate (which corresponds to $\mathcal{C} = \{I_n\}$), each source location is given an individual variance in this approach.

The second crucial point is the choice of the hyperprior. For a general discussion, see Lucka (2011), page 20. For our studies, we only consider hyperpriors that factorize over the single hyperparameters γ_i. Furthermore, because we do not want to bias our model to certain source locations a priori, all single hyperparameters should be identically distributed. Finally, because we are searching for focal solutions, hyperpriors leading to *sparse* estimates of γ will be used, i.e., hyperpriors forcing most hyperparameters to be (nearly) zero, while few hyperparameters are allowed to have a large amplitude. Our particular choice for this purpose is the *inverse-gamma* distribution:

$$p_{hpr}(\gamma) = \prod_{i=1}^{k} p_{hpr}^{i}(\gamma_i) = \prod_{i=1}^{k} \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma_i^{-\alpha-1} \exp\left(-\frac{\beta}{\gamma_i}\right). \tag{15}$$

The parameters $\alpha > 0$ and $\beta > 0$ determine the *shape* and the *scale* of the distribution, whereas $\Gamma(x)$ denotes the Gamma function.

This choice of prior and hyperprior was also used in Sato et al. (2004), Nummenmaa et al. (2007a,b), Calvetti et al. (2009) and Wipf and Nagarajan (2009). Due to the diagonal shape of Σ_s, the full posterior for this model becomes (cf., Eqs. (3), (14) and (15)):

$$p_{post}(s, \gamma | b) \propto \exp\left(-\frac{1}{2} \left(\frac{1}{\sigma^2} \|b - Ls\|_2^2 + \sum_{i=1}^{k} \frac{\|s_{i*}\|^2}{\gamma_i} + 2\sum_{i=1}^{k} \left(\frac{\beta}{\gamma_i}\right) + 2\left(\alpha + \frac{5}{2}\right) \sum_{i=1}^{k} \ln \gamma_i \right) \right), \tag{16}$$

where we abbreviated the sum of the ℓ_2-norms of the d sources at location i with $\|s_{i*}\|^2$. For a more detailed derivation of this formula, we refer to Lucka (2011), page 41. The analytical advantage of such a model over other possible approaches is that the expression within the brackets in Eq. (16) is quadratic with respect to s and the γ_i's are mutually independent. This advantage simplifies and speeds up many practical computations with this model.

Inference for hierarchical models

Note that the posterior (Eq. (16)) depends on two types of parameters: the parameters of main interest, s, and the hyperparameters, γ. This situation offers more methods of inference than the simple CMand MAP estimation scheme introduced in Bayesian formulation of the static inverse problem section. Five main approaches are established:

- Full-MAP: Maximize $p_{post}(s, \gamma | b)$ w.r.t. s and γ;
- Full-CM: Compute the expectation of $p_{post}(s, \gamma | b)$ w.r.t. s and γ;
- S-MAP: Compute the expectation of $p_{post}(s, \gamma | b)$ w.r.t. γ, and maximize over s (*Type I approach*);
- γ-MAP: Compute the expectation of $p_{post}(s, \gamma | b)$ w.r.t. s, and maximize over γ, first; then, use $p_{post}(s, \hat{\gamma}(b) | b)$ to infer s (*Type II approach*, *Hyperparameter MAP*, *Empirical Bayes*);
- VB: Assume an approximative factorization of $p_{post}(s, \gamma | b) \approx \hat{p}_{post}(s | b)\, \hat{p}_{post}(\gamma | b)$; approximate both with distributions that are analytically tractable (VB = *Variational Bayes*);

In the traditional Bayesian framework, all parameter types should be treated equally, which is why the first two schemes are also referred to as *fully-Bayesian* methods. However, in practical applications, hyperparameters are introduced with the explicit intention that they have a different meaning than the normal parameters; hence, a different treatment can be justified from the methodical point of view. The corresponding schemes, S-MAP and γ-MAP, are usually classified as *semi-Bayesian* methods (see Wipf and Nagarajan, 2009 for a comprehensive discussion). Variational Bayesian techniques (often referred to as *approximate-Bayesian* methods) actually rely on more advanced considerations than a simple approximation, but this cannot be pursued in detail here (Friston et al., 2007; Nummenmaa et al., 2007a; Wipf and Nagarajan, 2009). The focus of our work lies on the fully-Bayesian methods.

Algorithms for fully-Bayesian inversion

None of the estimates mentioned in the last section can be computed explicitly. In this section, we outline the ideas behind the algorithms we utilize for numerically computing the full-MAP and full-CM estimates. Details, especially concerning a fast and stable implementation, are presented in the appendix.

CM estimation

Due to the high dimension of the source space, the integration (cf., Bayesian formulation of the static inverse problem) is intractable by means of traditional quadratures. Integration using *Monte Carlo* methods can circumvent these difficulties. A sequence of points $(s_i, \gamma_i), i = 1, \ldots, M$ is constructed that is distributed like the posterior. Optimally, these points should be drawn independently because in this case, the *law of large numbers* would guarantee that

$$\frac{1}{M} \sum_{i=1}^{M} (s_i, \gamma_i) \overset{M \to \infty}{\to} (s, \gamma)_{CM} = \int_{\mathbb{R}^n \times \mathbb{R}^k} (s, \gamma) p_{post}(s, \gamma | b) ds\, d\gamma$$

almost surely and in ℓ_1 with rate $O(M^{-1/2})$, i.e., the empirical mean of the sequence converges to the expected value of the posterior (Klenke, 2008). A difficulty in our setting is that the posterior is not given in a form that allows for drawing independent samples because the posterior is only known up to a normalizing constant (the model evidence) and does not belong to a class of distributions for which such sampling schemes are known. However, by to the *strong ergodic theorem*, the above convergence and its rate still hold if the sequence is dependent but originates from an *ergodic Markov chain* that has $p_{post}(s, \gamma | b)$ as its *equilibrium distribution* (Klenke, 2008). Techniques to construct such chains are called *Markov chain Monte Carlo* (*MCMC*) methods. For our application, we rely on a MCMC scheme that is called *blocked Gibbs sampling* (Gelman et al., 2003; MacKay, 2003) with the concrete form proposed in Nummenmaa et al. (2007a); Calvetti et al. (2009). This scheme exploits the special structure of Eq. (16) by drawing from the posterior either conditioned on s or on γ at a time:

Algorithm 1. Blocked-Gibbs-sampling algorithm
Initialize γ by $\gamma_i^{[0]} = \beta/\alpha$ for all i and set $j = 1$. Define the desired sample size M and burn-in size Q;
For $j = 1, \ldots, M + Q$ do:

1. Draw $s^{[j]}$ from $p_{post}(s | \gamma^{[j-1]}, b) \propto p_{post}(s, \gamma^{[j-1]} | b)$ using the conditional normality of p_{post};
2. Draw $\gamma^{[j]}$ component-wise from $p_{post}(\gamma | s^{[j]}, b) \propto p_{post}(\gamma, s^{[j]} | b)$ using the factorization over γ_i;

Approximate $(\hat{s}, \hat{\gamma})_{CM}$ by the empirical mean of the samples $j = Q+1, \ldots, Q+M$.

This sampling technique is very simple, but also very powerful. A main advantage over other MCMC schemes is that it does not require any manual tuning of sampling parameters. The sampling problem in step 1 is solved by a reformulation into a least-squares problem, and the sampling problem in step 2 can be solved efficiently utilizing the *conjugacy* of the inverse gamma hyperprior and the factorization. For readers interested in the technical details, references are given in Appendix A.

MAP estimation

Our main tool for the MAP estimation is a cyclic algorithm that takes advantage of the special form of the posterior and is called the *Iterative Alternating Sequential* (*IAS*). The form we use was introduced by Calvetti and Somersalo (Calvetti and Somersalo, 2007a, 2008a; Calvetti et al., 2009) and is inspired by a similar, more general, algorithm called *half quadratic minimization* (Aubert and Kornprobst, 2006):

Algorithm 2. Iterative alternating sequential
Initialize γ by γ_0 and set $j=1$. Define the desired iteration number T; For $j=1,\ldots,T$ do:

1. Update s by $s^{[j]} = \arg\max\{p_{post}(s|\gamma^{[j-1]},b)\} = \arg\max\{p_{post}(s,\gamma^{[j-1]}|b)\}$;
2. Update γ by $\gamma^{[j]} = \arg\max\{p_{post}(\gamma|s^{[j]},b)\} = \arg\max\{p_{post}(\gamma,s^{[j]}|b)\}$;

Approximate $(\hat{s},\hat{\gamma})_{MAP}$ by the last sample $(s^{[T]},\gamma^{[T]})$.

(Note that the conditional densities are always proportional to the corresponding joint density by a factor only dependent on the conditioned parameter).

As for the CM estimation, step 1 is solved using a reformulation into a least-squares problem, and step 2 can be solved explicitly by utilizing the *conjugacy* of the inverse gamma hyperprior and the factorization. Further details are given in Appendix A.

Note that we did not yet specify an initialization rule for γ. This choice will turn out to be the crucial point due to the following difficulty. The IAS algorithm as a component-wise gradient-based optimization method is only locally convergent, i.e., it will terminate in one of the local minima around the initialization point. This is not a problem as long as the posterior energy (i.e., the negative natural logarithm of its density function) is convex, and thus, a unique minimum exists. However, a potential problem in our setting is the *multimodality* of the posterior Eq. (16). This problem results from the non-convexity of the *energy* of the inverse gamma hyperprior, which is the negative natural logarithm of the density function. Details and illustration of this phenomenon are given in Nummenmaa et al. (2007a) and Lucka (2011), Section 4.4.2. The multimodality is always present to some extent; however, the concrete choice of the parameters α and β and the interplay with the underdeterminedness of the likelihood Eq. (2) determine to what extent the multimodality practically affects the estimation process.

For these reasons, we examine three different initialization schemes:

- MAP1: A uniform initialization by $\gamma_i^{[0]} = \beta/\alpha$ for all i. This corresponds to the method used in Calvetti et al. (2009) and yields a very fast MAP estimation method;
- MAP2: A CM estimate is computed first, and $\gamma^{[0]} = \gamma_{CM}$;
- MAP3: U very rough approximations to the CM estimate $(\hat{s},\hat{\gamma})^i_{CM}$, $i=1,\ldots,U$ are first computed using very small sample sizes M. Then they are used as seeds for the IAS algorithm: $\gamma^{[0],i} = \gamma^i_{CM}$, $i=1,\ldots,U$. The results $(\hat{s},\hat{\gamma})^i_{MAP}$, $i=1,\ldots,U$ are compared with respect to their posterior probability, and the result with the highest probability is chosen as the final MAP estimate.

Choosing CM estimates as initializations for MAP estimation seems unmotivated at this point; however, the MAP2 and MAP3 methods will yield good performances in all the simulation studies. Specifically, these methods are often able to improve upon the performance of the CM estimate on which they rely. We will outline the reasons for this phenomenon in the Discussion section.

Validation means and inverse crimes

While subsequent work will focus on performing a validation of the fully-Bayesian methods with real data, this paper focuses on extensive simulation studies to develop the basic properties these methods. When using synthetic data produced using an invented source configuration, it is crucial to avoid an *inverse crime*, i.e., the model and reality are identified (Kaipio and Somersalo, 2005), as this usually leads to overly optimistic results. In our case, one should not produce synthetic data with the same lead-field matrix used for the inversion, which would correspond to the assumption that the real current sources are also restricted to the locations of the chosen source space nodes; they should instead be placed independent of these locations. As a number of commonly used measures do rely on an inverse crime, as they assume that the reference and the estimated source come from the same space ($\mathbb{R}^n$ in our case), we will, rather, use the following measures to evaluate our results. For single sources, the well-known *dipole localization error* (*DLE*) is the distance from the location of the reference dipole source to the source space node with the largest estimated current amplitude. We further introduce the *spatial dispersion* (*SD*) as an illustrative measure of the spatial extent of the estimated current (see Appendix B for the details of our definition, which differs from the one used in Molins et al., 2008).

While the DLE can only be used for single sources (the extension to multiple sources is not trivial) and is only sensitive to localization, the SD does not account for localization at all. Many other measures in EEG/MEG also only work for specific source scenarios, specific source forms or measure only specific aspects. To overcome these limitations, we introduced and examined a novel validation measure in Lucka (2011), Section 1.3.3 that is sensitive to localization, relative amplitude and spatial extent and works with arbitrary complex source scenarios and with arbitrary estimation formats (sLORETA, Pascual-Marqui, 2002, e.g., yields standardized activity estimates rather than real current amplitudes). The *earth mover's distance* (*EMD*) is a distance measure between probability densities. Strictly speaking, it is a type of *Wasserstein metric* originating from the theory of *optimal transport* (Ambrosio et al., 2008). The EMD measures the minimal amount of (physical) work required to transfer the mass of one density into the other density. Illustratively, one can visualize one density as a pile of sand and the other density as a bunch of holes. Then, the EMD is the minimal amount of work one needs to perform to fill up the holes with the sand. While the EMD can be computed for arbitrary complex real and estimated source scenarios, it reduces to intuitive measures in simple situations (e.g., for two dipoles, one reference and one estimated, it yields the spatial distance between the sources, i.e., it reduces to the DLE). Mathematical details and a closer examination of the EMD's features are given in Appendix B and in Lucka (2011), Section 4.7.

Finally, to examine the phenomena of depth bias in more detail (see Brain networks involving deep-lying sources section), we define the *depth* of a location in the head model as the minimal distance to one of the sensors.

Static and dynamic inverse problems

In all major inverse approaches to EEG/MEG introduced in Inverse methods for EEG/MEG section, the temporal dimension of the (already preprocessed) data can be incorporated to supplement the information given by the data at a single time point and the spatial a priori information reflected by the specific inverse method. *Single-pass* strategies extract information from one domain (time or space) to enhance the reconstruction in the other domain (see e.g., Durka et al., 2005). State-space approaches, such as *Kalman filtering*, iterate between space and time to balance both sources of information (Galka et al., 2004). Conversely, *multi-pass* strategies incorporate information from both domains simultaneously (*spatio-temporal inversion*, see, e.g., Baillet and Garnero, 1997; Gramfort et al., 2011; Trujillo-Barreto et al., 2008). This often requires sophisticated algorithms to reduce the computational load originating from the increased complexity of the problem (Schmitt et al., 2002).

For a number of reasons, we decided to not use the temporal information for our studies to examine the performance of CDR methods for source scenarios where depth bias and masking are the main challenges. First, depth bias and masking are phenomena that result from the interplay of the static likelihood (the lead-field matrix L is static) with the spatial prior information. Therefore, they are static, spatial

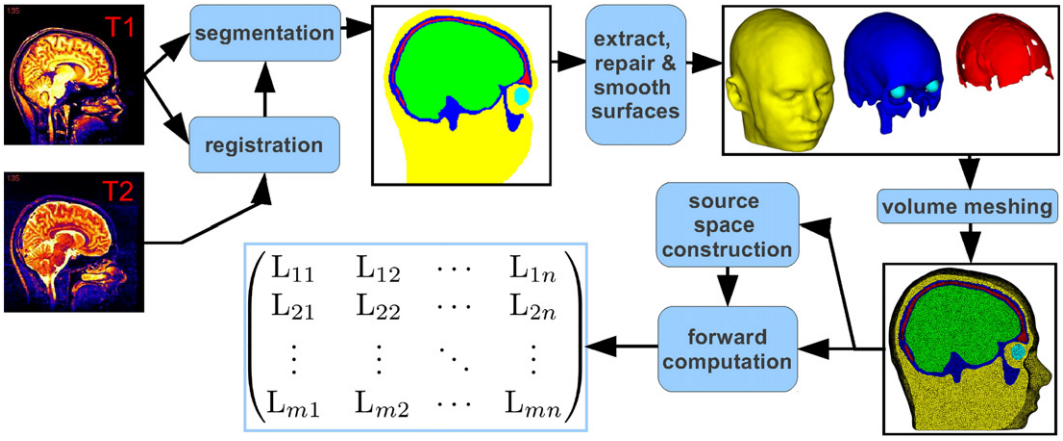


Fig. 1. Model generation pipeline.

phenomena. Incorporating temporal a priori information mainly helps by two mechanisms in this situation:

1. To mitigate the effects of noise: If the sources are temporally correlated and follow a joint dynamic that is well reflected by the temporal prior (e.g., they vary smoothly in time), the information about the underlying sources given by each measurement point is mainly the same, limited by the uncertainty caused by the independent realizations of the measurement noise. Therefore, combining the information of the whole time series primarily reduces that uncertainty. The extreme case is given by a completely static source configuration that is independently measured at many different time points (see Wipf and Nagarajan, 2009 for further discussion).
2. By simplifying source separation: If the source configuration consists of multiple sources with different time courses, the temporal measurement data can help to separate them at a time point where they are simultaneously active (see Wipf and Nagarajan, 2009 for further discussion).

In summary, the temporal extension improves the information given by the measurement while not directly counteracting the ill-conditioning of the spatial forward operator L. This suggests that one has to pay attention when comparing the results of the static and dynamic inverse methods for a source scenario in which difficulties arise from the spatial configuration of the sources. If the spatial a priori information used by the two methods differs and the performance of the dynamic is superior, it is not clear whether this is due to the improved information given by the measurements or the different spatial a priori information. For a fair and meaningful comparison one would need to consider only dynamic or static inverse methods. While dynamic formulations of equivalent current dipole and beamforming methods are established in practice, the situation for CDR is, so far, less elaborated (for a recent overview, see Gramfort et al., 2011; Ou et al., 2009). In addition, conducting simulation studies with dynamic inverse methods also incorporates the risk of committing an inverse crime by assuming time courses that are well reflected by some temporal models while inconsistent with other models.

Results

Setting for the studies

Head model and source space

For the numerical approximation of the forward problem, we use the finite element (FE) method because of its flexibility with regard to the realistic modeling of the head volume conductor at a fast computational speed with respect to this degree of modeling accuracy. Although working with a head model that is as realistic as possible is, in general, preferable (see the references in the description below), the specific aims of our studies require some simplifications. We do not want to include the inner brain compartments (the CSF, gray matter and white matter) because we want to focus on the effect of depth bias separately from other sources of error, e.g., from the effects caused by the anisotropy of the white matter (which also makes the results comparable to those obtained using *BEM* models, which cannot capture the anisotropy and normally do not differentiate between the inner brain compartments as well). Additionally, to facilitate the interpretation of the results, we require a homogeneous innermost compartment without holes and enclosures where we can place the test sources. Another important aspect for practical EEG/MEG studies is the effect of *insufficient sensor coverage*. For an optimal scan of the electromagnetic field pattern, the sensors should be uniformly distributed in every spatial direction. However, for practical reasons, this is not possible in realistic settings. The neck causes a semi-shell-like sensor distribution, which is not able to record fields in the direction of the feet. In particular deep-lying sources suffer from this insufficiency. The influence of insufficient sensor coverage should not be mixed with the effects of depth bias in this first, basic study. Therefore, we will use two sensor configurations in our studies. First, an artificial sensor configuration consisting of 134 EEG sensors distributed uniformly over the surface of the head model is created (abbreviated *f-cap* for *full cap*). From these sensor positions, a subset of 63 sensors, which represents a realistic sensor placement, are chosen as a second sensor configuration (abbreviated *r-cap* for *realistic cap*). Fig. 8 shows both configurations.

In the following section, we will outline the model generation pipeline, which is also depicted in Fig. 1.

T1- and T2-weighted magnetic resonance images (MRI) of a healthy subject were measured on a 3 T MR scanner. A T1w pulse sequence with fat suppression and a T2w pulse sequence with minimal water-fat shift, both with an isotropic resolution of $1.17 \times 1.17 \times 1.17$ mm, were used. The T2-MRI was registered onto the T1-MRI using an affine registration approach, and the mutual information as a cost function as implemented in *FSL*.[1] The compartments of the skin, the skull compacta and the skull spongiosa were segmented using a gray-value-based active contour model (Vese et al., 2002) and thresholding techniques. The segmentation was carefully checked and corrected manually. Because of the importance of skull holes on source analysis (Oostenveld and Oostendorp, 2002; Van den Broek et al., 1998), the foramen magnum and the two optic canals were correctly modeled as skull openings. Following Bruno et al. (2003, 2004) and Lanfer et al. (2010), the inferior part of the model was not directly cut below the skull but was realistically extended to avoid volume conduction modeling errors. The software *CURRY*[2] was then used for the segmentation of the cortex surface as well as the extraction of high-resolution meshes of the surfaces of the skin, the eyes,

[1] FLIRT — FMRIB's Linear Image Registration Tool, http://www.fmrib.ox.ac.uk/fsl/flirt/index.html.
[2] CURrent Reconstruction and Imaging (CURRY), http://www.neuroscan.com/.

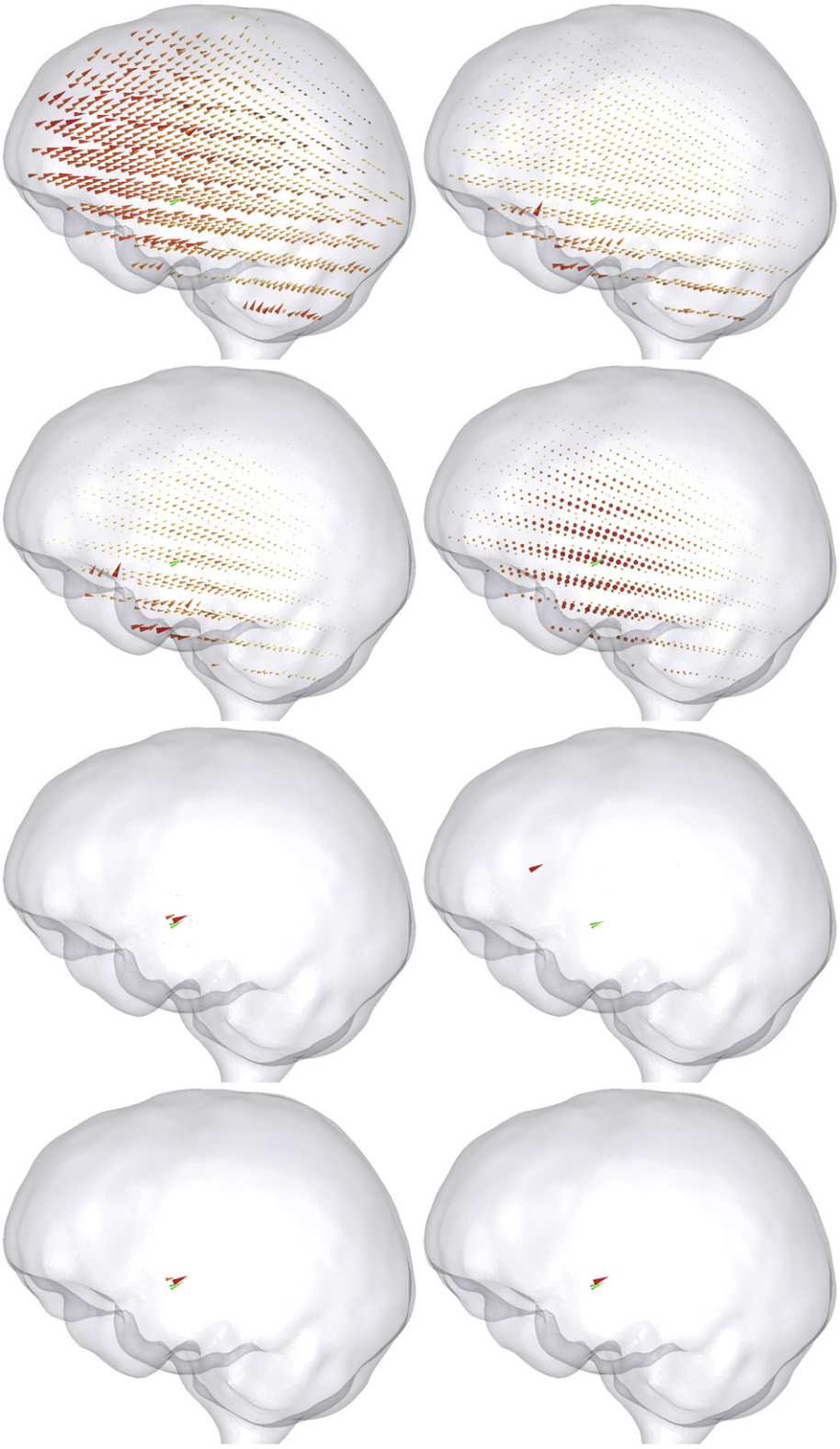


Fig. 2. Results from different inverse methods for a single reference dipole source (green cone) using the r-cap. From top to bottom and from left to right: MNE, WMNE with ℓ_2 weighting, WMNE with regularized ℓ_∞ weighting, sLORETA, CM, MAP1, MAP2 and MAP3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the skull compacta, the skull spongiosa and the brain from the voxel-based segmentation volumes. The surfaces were smoothed using Taubin smoothing (Taubin, 1995) to remove the blocky structure, which results from the fine surface sampling of the voxels. For the aims of our specific studies, only the surfaces of the skin, the eyes, the skull compacta and the skull spongiosa were then used to create a high-quality 3D Delaunay triangulation using *TetGen*.[3] In total, the resulting tetrahedral finite

[3] TetGen: A Quality Tetrahedral Mesh Generator and a 3D Delaunay Triangulator, http://tetgen.berlios.de/.

Table 1
Statistics of validation measures for study 1 for both sensor caps (mean ± std).

Method	DLE, f-cap	DLE, r-cap	SD, f-cap	SD, r-cap	EMD, f-cap	EMD, r-cap
MNE	29.46 ± 11.24	33.07 ± 12.65	2.4e−1 ± 1.0e−1	2.5e−1 ± 1.0e−1	53.20 ± 2.74	54.90 ± 4.50
WMNE ℓ_2	30.65 ± 13.52	35.08 ± 15.96	2.5e−1 ± 1.1e−1	2.5e−1 ± 9.7e−2	52.17 ± 2.53	53.64 ± 3.35
WMNE $\ell_{\infty,reg}$	29.40 ± 14.81	35.38 ± 17.42	2.2e−1 ± 8.0e−2	2.2e−1 ± 7.0e−2	49.56 ± 3.64	51.08 ± 3.82
sLORETA	6.10 ± 2.35	6.60 ± 2.83	1.9e−1 ± 6.8e−2	2.2e−1 ± 7.1e−2	40.58 ± 2.48	43.43 ± 3.42
CM	6.16 ± 2.37	6.94 ± 3.14	1.3e−3 ± 1.1e−3	2.0e−3 ± 1.9e−3	7.32 ± 2.31	8.85 ± 3.33
MAP1	27.00 ± 11.90	32.77 ± 14.32	9.8e−3 ± 5.8e−2	2.7e−2 ± 9.8e−2	28.18 ± 11.54	33.76 ± 13.70
MAP2	5.85 ± 2.16	6.39 ± 2.74	2.2e−4 ± 3.3e−4	1.1e−4 ± 2.6e−4	6.08 ± 2.22	6.45 ± 2.74
MAP3	5.79 ± 2.13	6.14 ± 2.48	7.1e−6 ± 4.5e−5	2.8e−5 ± 1.2e−4	5.84 ± 2.21	6.15 ± 2.49

element (FE) model consists of 512.394 nodes and 3.176.162 tetrahedral elements. The conductivity values (denoted in S/m) for the different compartments were chosen to be 0.43 for the skin (Dannhauer et al., 2011), 0.505 for the eyes (Ramon et al., 2006), 0.0064 for the skull compacta, 0.02865 for the skull spongiosa (Akhtari et al., 2002; Dannhauer et al., 2011) and 0.33 for the inner brain compartment (Dannhauer et al., 2011).

Within the inner compartment, a source space consisting of 1.000 source locations based on a regular grid is chosen, and the grid size is 10.99 mm (see Fig. 10). At each node, $d=3$ orthogonal dipoles in Cartesian directions are placed. For computing the corresponding lead-field matrices, different FE approaches for modeling the source singularity are known from the literature: The subtraction approach (Bertrand et al., 1991; Drechsler et al., 2009; Schimpf et al., 2002; Wolters et al., 2007b), the partial integration direct method (Schimpf et al., 2002; Vallaghé and Papadopoulo, 2010; Weinstein et al., 2000) and the Venant direct method (Buchner et al., 1997). In this study, we used the Venant approach based on a comparison of the performance of all three in multilayer sphere models, which suggested that for sufficiently regular meshes, the Venant approach yields suitable accuracy over all realistic source locations (Lew et al., 2009; Vorwerk, 2011). This approach has the additional advantage that the resulting FEM approach has a high computational efficiency when used in combination with the FE transfer matrix approach (Wolters et al., 2004). Standard piecewise linear basis functions were used. The computations were performed with *SimBio*.[4] In Fig. 9, the sum of the ℓ_2-norms of the three gain vectors for each location is depicted for both sensor caps.

Another benefit of the Venant approach is that the sources are not modeled as point-like sources, but patch-like, which is more realistic (Buchner et al., 1997; Hämäläinen et al., 1993; Niedermeyer and Lopez da Silva, 2004; Wolters et al., 2007a). We also stress that the term "focal" in our investigations is to be understood relative to the grid resolution used. The activity at a source space node represents the cumulative activity of a volume of gray matter tissue, which is dependent on the grid resolution. Representing reference sources by single dipoles should also be regarded as modeling with respect to this resolution rather than a definite representation independent of the resolution. Simulating the corresponding reference data using dipoles is the current "gold standard" in EEG/MEG. Therefore, we chose to use this source model in this first, elementary study, especially because important features of the inverse problem, such as depth bias and masking, are already present using this source model (and can, thus, be addressed within our studies).

Inverse methods

In this section, we list the methods we use together with the choice of internal parameters. For the hierarchical model, choosing α and β is, in fact, a difficult practical and methodical task. Our choice relies on preliminary computations and considerations that can be found in Lucka (2011), Section 4.4.2. We chose the parameters for each method in an optimal way for a fair comparison of their performance. A further reference addressing with this issue is Nummenmaa et al. (2007a).

[4] SimBio: A generic environment for bio-numerical simulations, https://www.mrt.uni-jena.de/simbio.

The following methods will be examined in our studies:

- MNE as described in Bayesian formulation of the static inverse problem section (Hämäläinen et al., 1993);
- WMNE with ℓ_2 and regularized ℓ_∞ weighting as described in Bayesian formulation of the static inverse problem section (Fuchs et al., 1999);
- sLORETA as described in Bayesian formulation of the static inverse problem section (Pascual-Marqui, 2002);
- Full-CM estimation using the algorithm described in Algorithms for fully-Bayesian inversion section for the HBM introduced in Hierarchical modeling in EEG/MEG section (this methods will be abbreviated by CM from this point forward). Parameters: $\alpha=0.5$ and $\beta=5\cdot10^{-8}$;
- Full-MAP using the three methods, MAP1-3, described in Algorithms for fully-Bayesian inversion section for the HBM introduced in Hierarchical modeling in EEG/MEG section. Parameters: $\alpha=0.5$, $\beta=5\cdot10^{-6}$ for MAP1 and $\alpha=0.5$ and $\beta=5\cdot10^{-8}$ for MAP2 and MAP3;

The regularization parameter λ for MNE, WMNE and sLORETA are chosen using the *discrepancy principle* (e.g., Engl et al., 1996; Kaipio and Somersalo, 2005) because we assume that the noise variance σ^2 is known (or assume that we have a good estimate of it, e.g., based on pre-stimulus data).

To obtain an initial visual impression of the different methods, the reconstruction results for a single dipole source using the r-cap are shown in Fig. 2 (the source is located between the source space nodes to avoid an inverse crime, cf., Validation means and inverse crimes section).

Study 1: single dipole reconstruction

Setting

For the first study, 1000 single unit-strength source dipoles with random location and orientation were placed in the inner compartment (not necessarily on the source space nodes to avoid an obvious inverse crime, cf. Validation means and inverse crimes section). The following restriction on their depth (measured in the f-cap) was posed: First, the nearest sensor is searched for. For that sensor, the nearest source space node is searched for. The position of the dipole is only accepted if its depth (cf., Validation means and inverse crimes section) is larger than the depth of the source space node plus 10 mm. Using this procedure, dipoles that are closer to the sensors than any source space node are avoided, which facilitates the interpretation of the results (dipoles that are closer to the surface than any source space node cannot be reconstructed too superficial).

Measurement data are generated for both caps using the same forward computation procedure used for the lead-field generation, and Gaussian noise is added at a *noise level* of 5%. In line with Calvetti et al. (2009), we will refer to a (relative) *noise level* of x if the standard deviation of the measurement noise (i.e., σ in our notation) fulfills $\sigma = x\cdot\|b_0\|_\infty$, where b_0 are the measurements in the noiseless case. Because we did not find any systematic effect of adding noise on the depth bias and masking, a comparison to other noise levels is omitted here. The full results can be found in Lucka (2011), Section 4.5.

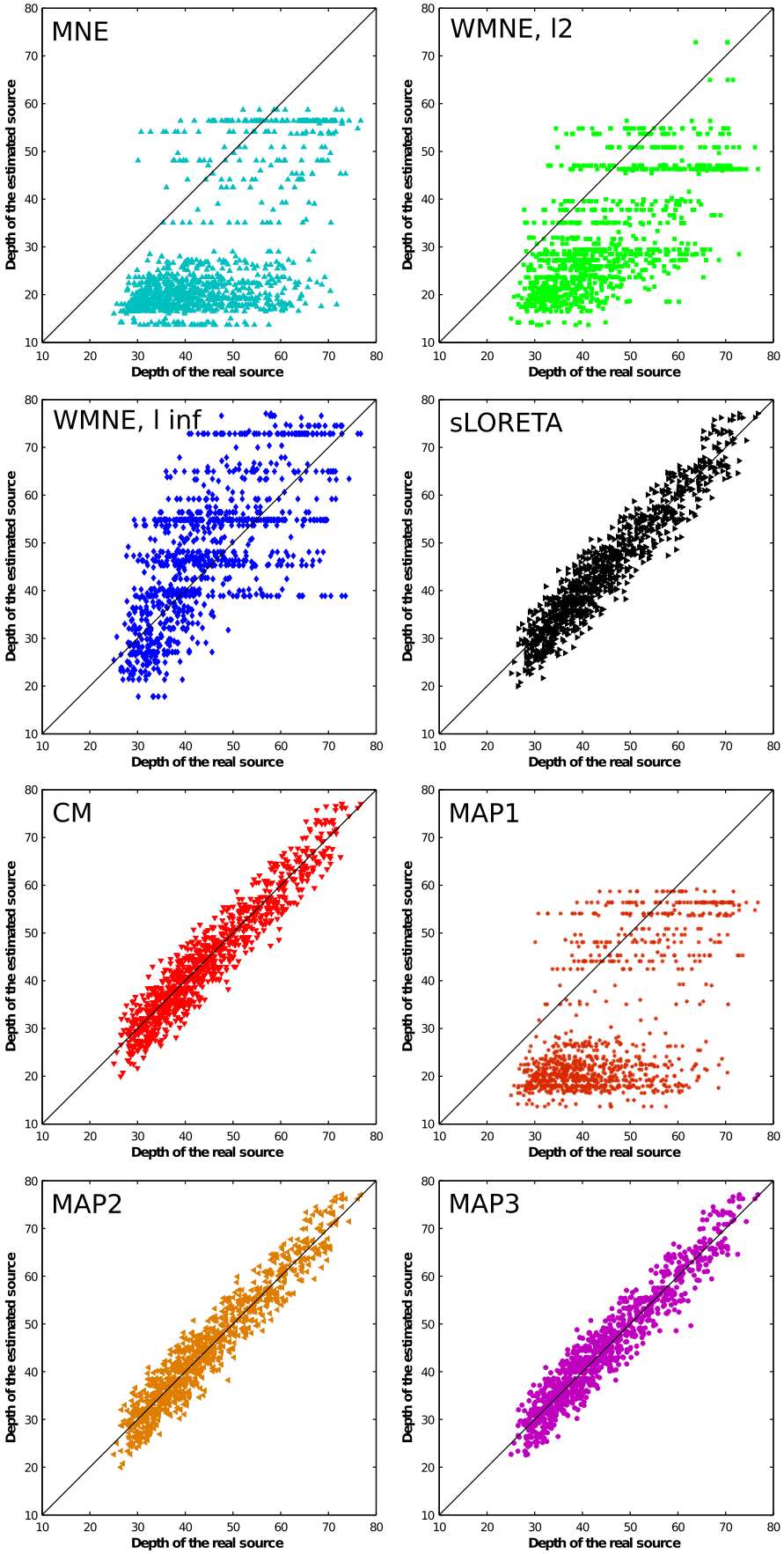


Fig. 3. Scatter plots to visualize the depth bias from different inverse methods using the f-cap. From top to bottom and from left to right: MNE, $q_{ab} = -0.441$; WMNE with ℓ_2 weighting, $q_{ab} = -0.410$; WMNE with reg. $e\ell\ell_\infty$ weighting, $q_{ab}=0.095$; sLORETA, $q_{ab}=-0.057$; CM, $q_{ab}=-0.058$; MAP1, $q_{ab}=-0.398$; MAP2, $q_{ab}=-0.007$; .MAP3, $q_{ab}=-0.007$.

Table 2
The resulting q_{ab} for 1000 single unit-strength dipoles for both sensor caps.

Method	q_{ab}, f-cap	q_{ab}, r-cap
MNE	−0.441	−0.392
WMNE ℓ_2	−0.410	−0.256
WMNE $\ell_{\infty,reg}$	0.095	0.087
sLORETA	−0.057	−0.049
CM	−0.058	−0.054
MAP1	−0.398	−0.322
MAP2	−0.007	−0.028
MAP3	−0.007	−0.019

Results

General properties. The mean distance from the reference dipoles to the next source space node was 5.27 mm, which is the lower bound for the DLE and EMD for all methods. Table 1 shows the DLE, SD and EMD averaged over all dipoles. For an idea of the practicality of the HBM methods, we note that our current implementations of the CM, MAP2 and MAP3 in Matlab take approximately 5 min of computation time for the f-cap and 3 min for the r-cap for each inverse reconstruction on a normal desktop PC.

Depth bias. We now focus on the first phenomenon introduced in Brain networks involving deep-lying sources section: the depth bias. To separate this phenomenon from the effects of insufficient sensor coverage, we will mainly use the reconstructions based on the f-cap measurements (see Setting for the studies section). We rely on a visual presentation using scatter plots. In Fig. 3, the depth (cf. Validation means and inverse crimes section) of the reference source is plotted on the horizontal axis, whereas the depth of the source space node with the largest estimated source amplitude is plotted on the vertical axis. A mark within the area below the $y = x$ line indicates that the dipole has been reconstructed too close to the surface, whereas a mark above the line indicates the opposite. q_{ab} denotes the percentage of the marks above the $y = x$ line minus 0.5. If a method shows a clear tendency to favor the lower area and q_{ab} is considerably below 0, the method suffers from depth bias (e.g., it is well known that MNE suffers from depth bias which can be seen clearly in Fig. 3, and is reflected in a q_{ab} of −0.441). A method performs well if its marks in this type of scatter plot are tightly distributed around the $y = x$ line, as this does usually not only indicate a localization at the correct depth but also a good overall localization. Table 2 lists q_{ab} for both the f-cap and the r-cap (note that the measure of depth relies on the individual caps, and therefore, a direct comparison between both q_{ab} is difficult).

Study 2: masking of deep-lying sources in two-dipole scenarios

Setting

The single dipoles that we used in the first study are now combined to form source configurations consisting of a deep-lying and a near-surface dipole. The dipoles are evenly divided into three parts according to their depth (measured in the f-cap; depth ranges of the three groups: 23.76–38.25, 38.25–49.41 and 49.42–77.65). For each of the 1000 source configurations used in this study, one dipole from the part with the largest and one from the part with the smallest depth are randomly picked. Noise is added at a level of 5% is added to the measurements.

Results

Initial example. We show an initial example where the effect of masking is very pronounced.[5] In all the subfigures of Fig. 4, the reference sources are represented by two green cones. One source is very close to the sensors, whereas the other source is very distant. The different subfigures of Fig. 4 show the reconstructions of the different methods based on the r-cap data. In the results of MNE, both WMNEs and sLORETA, even with a careful successive thresholding of the estimated source amplitudes do not reveal any evidence for the presence of the deep-lying source. In practice, these results would probably not provoke a user to attempt additional inverse methods in addition. Therefore, the deep-lying source is most likely overlooked. The MAP1 estimate does not reflect the reference sources in any useful way. The CM estimate reconstructs the near-surface source and presents some evidence for the deep-lying source, although this is hardly visible in the chosen image perspective. However, this small activation is sufficient for the MAP2 estimate (which was seeded at this CM estimate) to create good-quality reconstructions of both sources (remember that the test sources are placed in between the source space grid nodes, cf. Validation means and inverse crimes section). The MAP3 estimate also reconstructs two sources, although the deep source is not reconstructed at the nearest source space node.

General properties. Table 3 shows the EMD averaged over all source configurations (the SD provides little valuable information without a corresponding localization measure, and the DLE is not available in a multiple source scenario anymore, cf. Validation means and inverse crimes section).

Study 3: masking of deep-lying sources in three-dipole scenarios

Setting

The same setting as in study 2 is used except that the source configurations consist of one deep-lying and two near-surface dipoles. A further restriction is that the minimal distance between the different sources is at least 50 mm.

Results

Initial example. Similar to study 2, we show an initial example, where the effect of masking is very pronounced. The different subfigures of Fig. 5 show the reconstructions of the different methods relying on the r-cap data of a three-dipole source configuration (green cones). Again, with careful successive thresholding of the estimated source amplitudes of MNE, both WMNEs and sLORETA do not reveal any evidence for the presence of the third, deep-lying source. The WMNE with ℓ_2 weighting and sLORETA barely reveal the presence of the two near-surface sources, but they reconstruct one large cluster of activity between them. Once again, the MAP1 estimate does not reflect the reference sources in any useful way. The CM estimate presents evidence for all three sources, although it does not reconstruct them with the same amplitude. The results of the MAP2 and MAP3 are both accurate, and the MAP2 estimate outperforms the MAP3 estimate.

General properties. Table 4 shows the EMD averaged over all source configurations.

Comparison between MAP approximations

We briefly compare the different MAP estimation algorithms focusing on the posterior probability of their results. These algorithms all use different seed points for their optimization but rely on the same HBM. However, only methods that rely on the same parameter set can be compared. Because MAP1 uses a different setting than MAP2 and MAP3, the results for MAP1 were recomputed using the same parameter setting as that used in MAP2 and MAP3. These results will be named MAP1_cmp. However, note that MAP1_cmp performs even worse than MAP1 in terms of the EMD, DLE and SD. In Table 5, the average rank of the three methods within the three studies is shown. For each source configuration in a study, a ranking of the methods is computed by comparing the (rounded) probabilities of the MAP approximations found by the different methods.

[5] It was chosen by visual inspection after viewing the results for the study's first five source configurations.

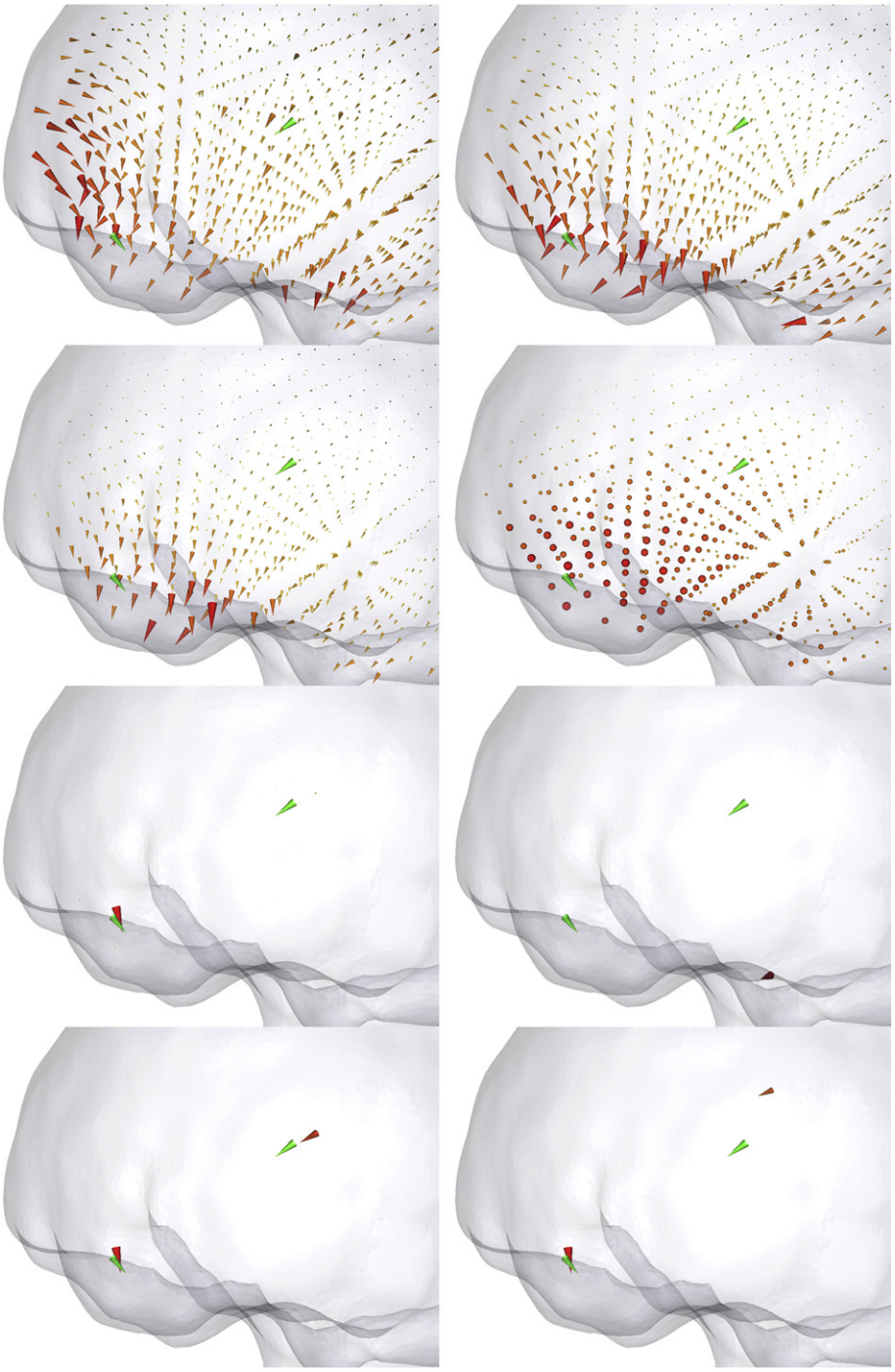


Fig. 4. Estimates of different inverse methods for a source configuration consisting of one near-surface and one deep-lying dipole (green cones) using the r-cap. The bottom left source is very close to the sensors, whereas the top right source is very distant. From top to bottom and from left to right: MNE, WMNE with ℓ_2 weighting, WMNE with regularized ℓ_∞ weighting, sLORETA, CM, MAP1, MAP2 and MAP3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Statistics of the EMD for study 2 for both sensor caps (mean ± std).

Method	EMD, f-cap	EMD, r-cap
MNE	44.63 ± 2.23	45.75 ± 3.06
WMNE ℓ_2	43.75 ± 1.97	44.62 ± 2.28
WMNE $\ell_{\infty,reg}$	41.79 ± 2.06	42.78 ± 2.20
sLORETA	36.38 ± 2.51	38.07 ± 2.70
CM	14.57 ± 4.98	18.21 ± 6.05
MAP1	42.10 ± 11.00	47.97 ± 10.98
MAP2	12.25 ± 6.30	16.53 ± 9.47
MAP3	12.41 ± 6.50	15.83 ± 9.32

The method that yielded the approximation with the highest probability is ranked highest. Methods that yielded an approximation with the same probability are ranked at similar levels. Subsequently, the mean rank of each method is computed over all 1000 source configurations.

Discussion

We examined new hierarchical Bayesian inference methods (HBM) for the EEG inverse problem and compared these methods to the results of established current density reconstruction (CDR) methods. In particular, we compared the fully-Bayesian conditional mean (CM) and

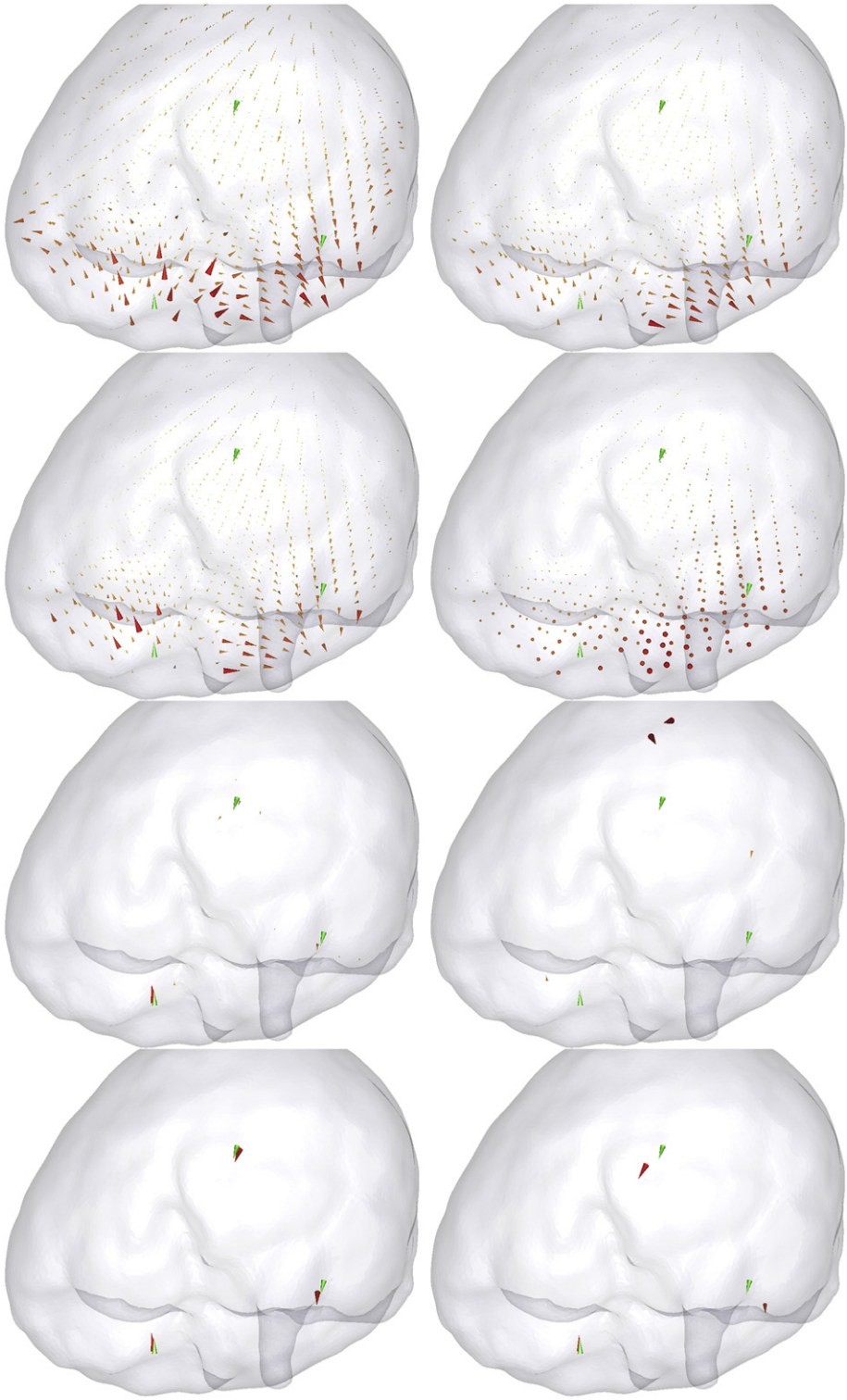


Fig. 5. Estimates of different inverse methods for a source configuration consisting of two near-surface and one deep-lying dipole (green cones) using the r-cap. The bottom sources are very close to the sensors, whereas the top source is very distant. From top to bottom and from left to right: MNE, WMNE with ℓ_2 weighting, WMNE with regularized ℓ_∞ weighting, sLORETA, CM, MAP1, MAP2 and MAP3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

maximum a-posteriori (MAP) estimates to the minimum norm estimates (MNE, Hämäläinen and Ilmoniemi, 1994), two weighted minimum norm estimates (WMNE, Fuchs et al., 1999) and sLORETA (Pascual-Marqui, 2002). For the MAP estimation we examined three different approaches: MAP1, which was proposed in (Calvetti et al., 2009), MAP2 and MAP3, which we proposed in Algorithms for fully-Bayesian inversion section.

Study 1 (single dipole reconstruction)

HBM methods

The MAP2 and MAP3 methods perform well with respect to the performance measures (cf. Table 1), and furthermore, they do not seem to suffer from depth bias (cf. Fig. 3) The MAP3 method slightly outperforms the MAP2 method. Compared to the other MAP approximation schemes,

Table 4
Statistics of the EMD for study 3 for both sensor caps (mean ± std).

Method	EMD, f-cap	EMD, r-cap
MNE	39.59 ± 1.72	40.57 ± 2.37
WMNE ℓ_2	39.02 ± 1.56	39.78 ± 1.83
WMNE $\ell_{\infty,reg}$	37.97 ± 1.52	38.76 ± 1.69
sLORETA	34.59 ± 2.18	36.05 ± 2.48
CM	17.60 ± 5.14	22.35 ± 5.89
MAP1	50.04 ± 13.43	57.30 ± 13.12
MAP2	17.10 ± 7.64	24.25 ± 9.22
MAP3	18.89 ± 7.88	25.72 ± 8.92

MAP3 also clearly attains the highest posterior probability (cf. Table 5), which suggests that it is the best approximation of the real MAP estimate. The CM method shows good result; however, it is interesting that the MAP2 method, which directly relies on the CM estimate, can clearly improve upon the CM estimate. Because the additional computation time is negligible, this result suggests that a subsequent optimization after an initial CM estimation should always be performed. The MAP1 scheme did not show convincing results both with regard to the DLE, SD and EMD (cf. Table 1) as well as with respect to depth bias (cf. Fig. 3). Compared to MAP2 and MAP3, it also attains less high posterior probabilities on average (cf. Table 5), which suggests that it might often only find a minor mode of the posterior and might, thus, not yield a reliable representation of the MAP estimate.

Our work was partly motivated by the results of Calvetti et al. (2009): Within a simplified geometry, a single deep-lying source was reconstructed (cf. Figs. 1–4 on pages 893–894 in Calvetti et al., 2009). The CM estimate with an inverse gamma hyperprior (which corresponds to the CM method used here) yielded the best result, both in location and in the extent of the estimated source. Moreover, the CM estimate seemed to have no depth bias, whereas the MAP estimate, by use of the uniformly initialized IAS algorithm (which corresponds to the MAP1 method used here), seemed to suffer from depth bias. In our work, we confirmed this impression about the CM estimate by performing a study on a realistic 3D head model over a larger number of single dipoles and by assessing performance measures. However, we also found that depth bias is not a feature of the MAP estimate itself, as suggested in the Discussion section in Calvetti et al. (2009), but the depth bias is, rather, a feature of the algorithm used to compute it. Due to the results in Comparison between MAP approximations section, we can be sure that the MAP3 result is closer to the real MAP estimate in terms of posterior probability as compared with MAP1, and MAP3 even performs slightly better than the CM estimate with regard to depth bias ($q_{ab} = -0.007$ to $q_{ab} = -0.058$, cf. Table 2).

Minimum norm based methods

The WMNE schemes used in this study are modifications of the original MNE explicitly aiming to improve the depth localization. Fig. 3 clearly show that they succeed in this aspect (although the scatterplot for the WMNE with regularized ℓ_∞ weighting and $q_{ab} = 0.095$ suggests that this weighting slightly exaggerates this aspect). These results confirm earlier studies on this topic, see, e.g., Fuchs et al. (1999). Concerning EMD, DLE and SD, the conclusion is less clear (cf. Table 1). The visualizations in Fig. 2 do not yield clear impressions of the different estimate characteristics as well. Therefore, more detailed examinations are required. The sLORETA estimate (which is also essentially based on the minimum norm, as it consists of computing a non-diagonal weighted norm of a MNE, see Pascual-Marqui (2002)) performs well concerning the DLE and depth bias (cf. Table 1 and Fig. 3). However, Fig. 2 suggests that the sLORETA result overestimates the spatial extent of the reference source scenario considerably. The average EMD and SP of sLORETA clearly confirm this impression (cf. Table 1). These results are in line with several other theoretical and numerical studies, see, e.g., (Lin et al., 2006; Pascual-Marqui, 2002; Sekihara et al., 2005; Wagner et al., 2004).

Table 5
Mean ranking of different MAP estimation algorithms in the three studies.

Method	Study 1, f-cap	Study 1, r-cap	Study 2, f-cap	Study 2, r-cap	Study 3, f-cap	Study 3, r-cap
MAP1_cmp	2.390	2.313	2.640	2.7480	2.928	2.9340
MAP2	1.398	1.275	1.547	1.6460	1.561	1.7150
MAP3	1.002	1.049	1.093	1.1020	1.367	1.2190

Direct comparison

The direct comparison in the single focal reference source scenario shows that compared to established methods such as the MNE and sLORETA, the HBM-based methods such as the CM, MAP2 and MAP3 clearly yield better results concerning the EMD and SD (cf. Table 1), and the visual impression is more convincing as well (cf. Fig. 2). However, it is important to stress that the above results were only attained for the specific source scenario examined in this study. Without further examination, their significance might be very limited because the ability to localize single dipoles is a rather trivial and a largely uninformative property, as shown by Grave de Peralta et al. (2009). Nevertheless, reconstructing single dipoles is a starting test for every inverse method for CDR, and the results for the methods based on the HBM clearly motivate the examination of their use in more detail.

Study 2 (masking of deep-lying dipole by near-surface dipole)

The initial example showed that the source scenario examined in this study is a very challenging scenario for inverse methods (see Fig. 4, and the studies in Wagner et al., 2004). The methods that performed best in the first study, i.e., the MAP2 and MAP3 scheme, also performed the best in this study (cf. Table 3 and Fig. 4). The comparison with the results from the MNE and sLORETA shows that HBM is able to improve upon established inverse methods in this source scenario by detecting the deep-lying source despite the presence of the near-surface source.

Compared to each other, the MAP3 scheme still outperforms the MAP2 scheme with regard to the posterior probability (cf. Table 5), but the situation concerning the EMD is less clear (see Table 3). This needs to be examined in more detail. Similar to the first study, Table 3 shows that, again, the MAP2 result improves upon the corresponding CM result upon which it is based. The results also suggest that the posterior distribution for these scenarios is more complex than for single sources.

Study 3 (masking of deep-lying dipole by two near-surface dipoles)

The results are very similar to the results of study 2. However, Tables 3 and 5 show that the relationship between the CM, MAP2 and MAP3 becomes more complicated and is to be examined in even more detail. Part of the reason that the MAP2 outperforms MAP3 in this scenario might be that the parameters Q and M for MAP3 have been optimized in a single-dipole recovery study.

The value of the EMD as a performance measure

In this work, we introduced the earth mover's distance (EMD) as measure that is both sensitive to localization and spatial extent of an estimate (cf. Validation means and inverse crimes section). Table 1 shows that the EMD fulfills these requirements. Only methods that attain a low DLE and SD will produce a low EMD. However, with regard to the sLORETA estimate, it would be preferable if more weight is put on the correct localization. Even though the sLORETA method has a small DLE and is commonly

used due to its localization properties, its EMD is much larger than that of other methods that produce focal estimates but considerably mis-localize the sources (e.g., the MAP1 scheme). The big advantage of the EMD is that it is applicable to more complex source scenarios as well. Conversely, the extension of other localization measures such as the DLE is not straightforward neither in terms of the implementation nor in the interpretation of the results. For the two- and three-source scenarios investigated in this work, the EMD was the only measure sensitive to localization that did not rely on an inverse crime (cf. Validation means and inverse crimes section).

Limitations and outlook

Confirming the present results with real data provides an important future work to complement the present simulation study. Our study specifically aimed at situations that are normally encountered, e.g., in the analysis of the early components of evoked responses (cf. Brain networks involving deep-lying sources section), i.e., focal source configurations that are measured at a single time instant. A validation with real data will be performed and reported in the near future.

Now that we have a profound knowledge on how the different methods behave with regards to the difficulties of the static inverse problem for the source scenarios investigated, an extension into the temporal domain can be investigated on this basis (cf. Static and dynamic inverse problems section). For the HBM, different possibilities to incorporate temporal information have been proposed or tested (Trujillo-Barreto et al., 2008; Wipf and Nagarajan, 2009), while a detailed comparison between these approaches has not yet been undertaken (to the best of our knowledge). This direction for future studies is extremely important to fully exploit the high temporal resolution offered by EEG/MEG recordings.

Motivated by the analysis of the early components of evoked responses and specific cases in presurgical epilepsy diagnosis, our current focus was on focal source scenarios incorporating up to three active focal sources. The HBM we used was tailored for such situations. In the future, we will examine extended source scenarios and the extended HBM for such scenarios, which should be of more interest for research in the area of cognitive neuroscience and other scenarios in presurgical epilepsy diagnosis with more extended underlying source patches.

Only two of the possible estimation methods that the HBM offers (cf. Inference for Hierarchical models section) were examined concerning our specific questions (cf. Brain networks involving deep-lying sources section). As most other publications using HBM address Variational Bayesian inference methods (VB, see, e.g., Friston et al., 2008; Nummenmaa et al., 2007a; Sato et al., 2004), a direct comparison between fully and variational Bayesian inference for HBM will be the next topic for simulation studies.

The present results concerning MAP2 and MAP3 estimates, which were introduced in this article, clearly show that superior results concerning performance measures and visual impression can be achieved as compared to the approach of Calvetti et al. (2009). To further improve the MAP estimation performance, alternative, non-convex optimization schemes for finding the true (global) MAP will be considered. MAP2 and MAP3 rely on searching for the MAP estimate in the vicinity of the CM estimate, and the present results clearly motivate research in this direction. Additionally, the actual cause for the depth bias, and why some methods suffer from it, has to be examined from a theoretical perspective as well.

For this first, elementary study, we simplified the brain volume conduction properties as homogeneous and isotropic, as is often performed in source analysis (see e.g., Acar and Makeig, 2010; Fuchs et al., 1998; Kybic et al., 2005). Future studies will investigate the interplay of HBM and more realistic head modeling, e.g., by incorporating the inner brain compartments and white matter anisotropy (Hallez, 2008; Haueisen et al., 2002).

Using only point estimates such as MAP and CM results on the neglect of a huge amount of the information that is contained in the abstract, $n+k$-dimensional posterior distribution (cf. Bayesian Formulation of the static inverse problem section). The exploitation and translation of this information to allow statements about certain reconstruction qualities (like the uncertainty of the estimated source location or extent) or to test a hypothesis is a very exciting topic for future research.

Only CDR methods were compared, while no comparison to dipole-fitting methods and scanning/beamforming methods was performed (cf. Inverse methods for EEG/MEG section for references). This will be an interesting direction for further studies.

Conclusions

HBM is a promising framework for EEG source localization. For the important source scenarios we examined, fully-Bayesian inference methods for HBM are able to improve upon established CDR methods such as MNE and sLORETA, in many aspects. In particular, these methods show good localization properties for single dipoles and do not suffer from a depth bias. As has been shown in this study, small localization errors for single source scenarios are not sufficient to judge the quality of an inverse method for EEG or MEG source analysis. However, in contrast to established inverse methods, such as minimum norm estimation and sLORETA, HBM-based methods are able to yield good reconstructions in the presence of two or three focal sources. Wasserstein metrics, in particular the earth mover's distance (EMD), are promising validation tools for future research on more complex source scenarios with multiple sources.

Acknowledgments

We thank the anonymous reviewers for their helpful critiques and comments that significantly improved our manuscript.

This research was supported by the German Research Foundation (DFG), projects WO1425/1-1, WO1425/2-1 and JU445/5-1 and by the Academy of Finland, project no 136412.

This work was made possible in part by software from the NIH/NCRR Center for Integrative Biomedical Computing, 2P41 RR0112553-12.

The authors would like to thank H. Kugel (Department of Clinical Radiology, University of Münster, Germany) for the MRI measurements of the MRI and A. Janssen, S. Rampersad (both from the Department of Neurology and Clinical Neurophysiology, Radboud University Nijmegen, the Netherlands) and B. Lanfer (Institute for Biomagnetism and Biosignalanalysis, University of Münster, Germany) for their help in setting up the realistic head model.

Appendix A. Algorithmical details

The blocked Gibbs sampler, Algorithm 1, and the IAS, Algorithm 2, are based on sampling or optimizing conditional densities. In more abstract words, they rely on *alternated conditional moves* through the parameter space $\mathbb{R}^n \times \mathbb{R}^k$ to construct a sequence of points $(s_i, \gamma_i) \in \mathbb{R}^n \times \mathbb{R}^k, i = 1, \ldots, t$. In the first half step (step 1 in both algorithms) the value of s is changed keeping γ fixed, while in the second half step (step 2 in both algorithms), the value of γ is changed while keeping s fixed. This is sketched in Fig. 6. While the CM approximation is inferred from that sequence by computing its empirical mean, the MAP approximation is given by the last point of the sequence. From Fig. 6, it is apparent why the IAS algorithm might become stuck in local minima when used with a multimodal posterior. However, Gibbs samplers are known to exhibit problems with multimodality as well (especially if s and γ are strongly correlated).

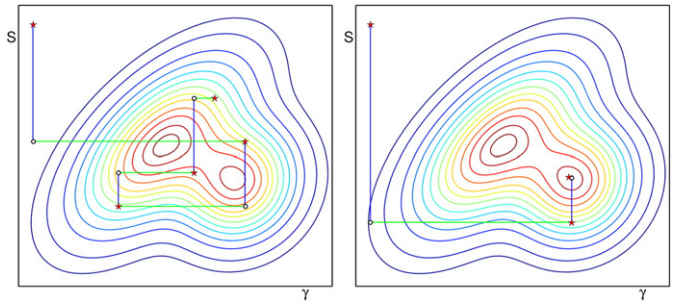


Fig. 6. Sketch of alternated conditional moves for a multimodal posterior (plotted via contour lines). Red stars mark subsequent states and circles mark half steps. Left: Algorithm 1; Right: Algorithm 2; The blue lines correspond to step 1 and the green lines correspond to step 2 in the respective algorithms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As a consequence, steps 1 and 2 in both algorithms can be solved using surprisingly similar approaches.

Step 1

In step 1, the sampling and optimization of a conditional Gaussian density with expectation and covariance given by

$$\mathbb{E}_{p(s|\gamma,b)}(s) = \Sigma_s L^t \left(L\Sigma_s L^t + \sigma^2 Id_m\right)^{-1} b$$
$$\mathbb{C}ov_{p(s|\gamma,b)}(s) = \Sigma_s - \Sigma_s L^t \left(L\Sigma_s L^t + \sigma^2 Id_m\right)^{-1} L\Sigma_s = \left(\Sigma_s^{-1} + \frac{1}{\sigma^2} L^t L\right)^{-1},$$

has to be solved (a derivation is given in Kaipio and Somersalo, 2005 and Lucka (2011), Section A.1.4). Remember that $\Sigma_s = \Sigma_s(\gamma)$ changes every step j, and therefore, a direct computation of the above quantities is not preferable with respect to computation time (and with respect to stability for the covariance matrix). Instead, both optimization and sampling can be realized by solving a relaxed weighted least squares problem:

$$\begin{bmatrix} L \\ \sigma\ \Sigma_s^{-1/2} \end{bmatrix} s^{[j]} \overset{ls}{=} \begin{bmatrix} b \\ 0 \end{bmatrix} + \sigma \begin{bmatrix} \omega_m \\ \omega_n \end{bmatrix}, \tag{A.1}$$

where we set $\omega_m = 0$, $\omega_n = 0$ to attain the conditional mode and draw ω_m and ω_n from standard normal distributions of dimension m and n to attain a sample from the conditional distribution (the details and derivation of this reformulation can be found in Lucka (2011), Section A.1.4).

Iterative solvers

Solving (A.1) can be done by using *Krylov subspace methods* such as the *conjugate gradient least squares* method (*CGLS*) with a preconditioning by $\Sigma_s^{-1/2}(\gamma)$ as proposed in Calvetti et al. (2009). Applied to iterative solvers for inverse problems, this technique is called *priorconditioning* (Calvetti and Somersalo, 2007b). In our hierarchical framework, the prior covariance itself is not fixed but relies on the fixation of the hyperparameters to their current values. The idea of using this present state of information, updated in every step of composite conditional walks is referred to as a *hyperpriorconditioning* (Calvetti et al., 2009).

Using preconditioned iterative solvers for problem (A.1) was proposed in Calvetti et al. (2009) and seems to be a canonical choice with regard to the high dimension of the problem. The advantage is that these schemes can be easily transferred to other fields of inverse problems, where the forward mapping is not provided in explicit matrix form (Calvetti and Somersalo, 2007a,b, 2008a,b; Kaipio and Somersalo, 2005). Additionally, the CGLS solver allows for the construction of *blocked inversion* schemes, where multiple right hand sides are inverted simultaneously which results in a considerable gain in speed (details on this can be found in Lucka, 2011, Section 3.6).

Explicit solution

Due to the small number of sensors in the EEG (we use $m = 63,134$ in our studies), we found a very simple alternative implementation that is competitive to the iterative approaches in terms of computation speed can be found. Using some matrix identities, the explicit solution of the systems can be computed very efficiently:

$$s^{[j]} = \left(\Sigma_s - \Sigma_s L^t \left(L\Sigma_s L^t + \sigma^2 Id_m\right)^{-1} L\Sigma_s\right) \cdot \left(L^t\left(\sigma^{-2} b + \sigma^{-1}\omega_m\right) + \Sigma_s^{-1/2}\omega_n\right)$$

This formula can be implemented in a straight forward manner:

Algorithm 3. Explicit Step 1 solution
1. Set $r = (L^t(\sigma^{-2}b + \sigma^{-1}\omega_m) + \Sigma_s^{-1/2}\omega_n)$;
2. Set $s_1 = \Sigma_s r$;
3. Set $t = Ls_1$;
4. Set $\tilde{\Sigma}_b = \left(L\Sigma_s^{1/2}\right)\left(L\Sigma_s^{1/2}\right)^t + \sigma^2 Id_m$;
5. Solve $\tilde{\Sigma}_b x = t$;
6. Set $s_2 = \Sigma_s L^t x$;
7. The solution is given by $s^{[j]} = s_1 - s_2$.

Remember that the multiplication with Σ_s can be performed componentwise. The computation of the projected source covariance $L\Sigma_s L^t$ within step 4 is the most computationally intensive part of the algorithm, and solving the linear system in step 5 is far less demanding. The system is only of size $m \times m$ and is symmetric positive definite. A solution via Cholesky decomposition is still fast enough to be negligible in comparison to the matrix–matrix multiplication in step 4. The solution of (A.1) with this algorithm is considerably faster than with iterative solvers (see Section A.1.10. in Lucka, 2011), and finding an optimal implementation is less demanding. Furthermore, it yields the exact solution of (A.1) within the bounds posed by ill-condition and finite precision, and no stopping criteria have to be chosen ad hoc. Another advantage is that the computation time is effectively independent of the right hand side, which is not the case for the iterative solvers we applied. Empirically, it was observed that more complex source configurations also result in a slower convergence of the CGLS algorithm.

Step 2

As the posterior factorizes over the single hyperparameters γ_i (cf. Eq. (16)), optimization and sampling can be performed componentwise. The hyperparameter-dependent single-component part of the posterior is (cf. Eq. (16)):

$$p_{post}(\gamma_i|s,b) \propto \exp\left(-\frac{1}{2}\left(\frac{\|s_{i*}\|^2}{\gamma_i} + 2\left(\frac{\beta}{\gamma_i}\right) + 2\left(\alpha + \frac{5}{2}\right)\ln\gamma_i\right)\right)$$

Computing the first- and second-order conditions for the maximum of this expression shows that the update rule is given by:

$$\gamma_i^{[j]} = \frac{\frac{1}{2}\|s_{i*}\|^2 + \beta}{\kappa}, \quad \text{with} \quad \kappa = \alpha + 3/2$$

Concerning the sampling, the conditional distribution $p_{post}(\gamma_i, s|b)$ can be rearranged to:

$$p_{post}(\gamma_i, s|b) \propto \exp\left(\frac{-\frac{1}{2}\|s_{i*}\|^2 + \beta}{\gamma_i} + (-(\alpha + 3/2) - 1)\ln(\gamma_i)\right).$$

This is also an inverse gamma distribution, with parameters $\bar{\beta} = \frac{1}{2}\|s_{i*}\|^2 + \beta$ and $\bar{\alpha} = (\alpha + 3/2)$ (cf. Eq. (15)). This invariance property is called *conditional conjugacy* and simplifies the sampling scheme considerably, as standard sampling routines can be used.

Parameter setting

The values for the parameters Q, M, T and U used in the studies for the HBM-based methods are listed in Table A.6.

Parameters used in the simulation studies.

Method	Parameter	Study1	Study 2
CM	Q	1000	1000
	M	50 000	200 000
MAP1	T	50	50
MAP2	Q	1000	1000
	M	50 000	200 000
	T	50	50
MAP3	U	128	256
	Q	25	25
	M	200	200
	T	50	50

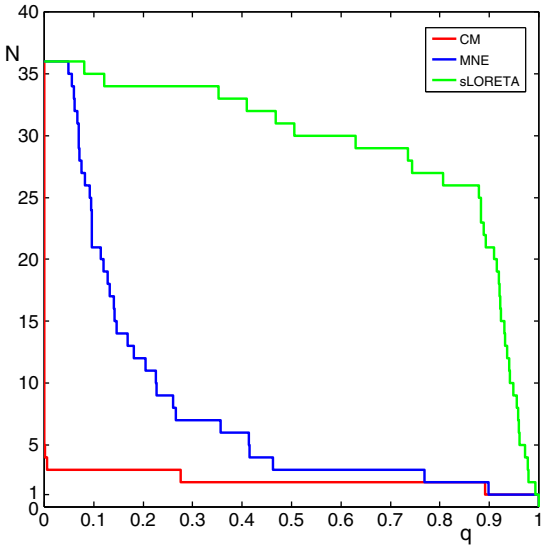


Fig. 7. The curves of $N = \mathring{f}(s,q)$ for $s = \hat{s}_{\mathrm{CM}}$ (red), $s = \hat{s}_{\mathrm{MNE}}$ (blue) and $s = \varphi_{\mathrm{sLORETA}}$ (green) for a simplified model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix B. Validation measures

Spatial dispersion (SD)

A standard approach to measure the spatial spread of an estimated current distribution would be to define a threshold q, and count the percentage of sources of which the amplitude is above q times the maximal source amplitude $\max\|s_i\|_2$. We will call this measure $\mathring{f}(s,q)$. However, $\mathring{f}(s,q)$ is not continuous, and involves some arbitrariness, because q has to be chosen ad hoc. In Fig. 7 three plots of $\mathring{f}(s,q)$ as a function of q are depicted for a simplified model geometry. The curves for focal and widespread CDRs show quite obvious differences.

We, therefore, propose to use a normalized version of the area below the curve as a measure for the spatial dispersion:

Definition 1. Spatial dispersion, SD

$$\Gamma_{SP} := \frac{1}{(k-1)}\left(\int_0^1 \mathring{f}(s,q)dq - 1\right) = \frac{1}{(k-1)}\left(\frac{\sum_{i=1}^{k}\|s_{i*}\|_2}{a_{\star,\infty}} - 1\right), \quad \text{with} \quad a_{\star,\infty} = \max_j \|s_{j*}\|_2 .$$

Note that this measure does not compare the spatial spread of the real and estimated source but only yields information on the estimate.

Earth mover's distance (EMD)

Supplementary to the text, we provide the mathematical definition of the EMD and some comments on its practical computation. The EMD is a Wasserstein metric which are distance measures between probability distributions (Ambrosio et al., 2008):

Definition 2. Wasserstein metric

Let μ and ν be two probability measures on a Radon space (Ω, d) that have a finite pth moment for some $p \geq 1$. Then the pth Wasserstein distance $W_p(\mu,\nu)$ is defined as:

$$W_p(\mu,\nu) = \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\Omega\times\Omega} d(x,y)^p \; d\gamma(x,y)\right)^{1/p},$$

where $\Gamma(\mu,\nu)$ denotes the class of all transport maps, i.e., measures on $\Omega\times\Omega$ with marginals μ and ν.

In our study, we examine the $p=1$ Wasserstein distance for the 3D-Euclidean distance $d(x,y) = \|x-y\|_2$, which is also called earth mover's distance due to the following analogy. The intuitive explanation behind this quantity dates back to Monge who published its formulation in 1781 as an optimal transport problem: The first probability measure is considered as an amount of sand piled on a space Ω and the second measure as a hole with the same size. For a given distance function d, the minimum-cost transport of the sand into the holes has to be determined (where the cost of a single assignment is understood as classical physical work in terms of distance times amount of sand). This minimal cost is the Wasserstein distance between the two measures.

The definition appears rather abstract for our practical task, but the lack of a simpler measure that is commonly accepted may be rooted in the fact that the task is not that simple after all. A good measure has to mimic the way source estimates from inverse methods are interpreted by the user and compare this interpretation with the reference source activity.

To compute the EMD between the reference and estimated source activity, both activities are transferred into discrete probability distributions. In our setting, the reference source activity j_{ref} was composed of single current dipoles at locations $\bar{r}_i, i = 1\ldots,\tau$:

$$j_{ref}(r) = \sum_{i=1}^{\tau} M_i \cdot \delta_i(\bar{r}_i - r) \qquad \forall r \in \Omega$$

Now define a discrete *signature* P by:

$$P = \left\{\left(p_1, w_{p_1}\right), \ldots, \left(p_\tau, w_{p_\tau}\right)\right\}$$
$$\text{with} \quad p_i := \bar{r}_i; \quad w_{p_i} := \frac{|M_i|}{M_{tot}}; \quad M_{tot} = \sum_{i=1}^{\tau}|M_i|$$

For the estimated CDR, we define a signature Q by:

$$Q = \left\{\left(q_1, w_{q_1}\right), \ldots, \left(q_l, w_{q_k}\right)\right\}$$
$$\text{with} \quad q_i := r_i; \quad w_{q_i} := \frac{\|s_{i*}\|_2}{a_{tot}}; \quad a_{tot} = \sum_{i=1}^{k}\|s_{i*}\|_2$$

Finally, define the distance matrix D by letting $D_{(i,j)}$ be the 3D-Euclidean distance between p_i and q_j. Now, we are ready to recast the computation of the EMD between P and Q into a linear programming problem as formulated by Kantorovich (Kantorovich, 1942; Kantorovich and Gavurin, 1949):

Definition 3. Reformulation of the EMD

With the above definitions, find a transport plan $\Gamma \in \mathbb{R}^{\tau \times k}$ that minimizes the work

$$\mathcal{W}(P, Q, \Gamma) = \sum_{i=1}^{\tau} \sum_{j=1}^{k} D_{(i,j)} \cdot \Gamma_{i,j} \tag{B.1}$$

subject to the following constraints:

$$\Gamma_{i,j} \geq 0, \quad 1 \leq i \leq \tau, 1 \leq j \leq k \tag{B.2}$$

$$\sum_{j=1}^{k} \Gamma_{i,j} = w_{p_i}, \quad 1 \leq i \leq \tau \tag{B.3}$$

$$\sum_{i=1}^{\tau} \Gamma_{i,j} = w_{q_j}, \quad 1 \leq j \leq l \tag{B.4}$$

The minimal work resulting from this computation is the EMD between P and Q. The constraints (B.2)–(B.4) ensure that Γ is a valid transport plan:

(B.2) ensures that the mass is transferred from P to Q and not vice versa.
(B.3) determines the amount of mass that has to be transferred from one place.
(B.4) determines the amount of mass that has to be transferred into one place.

In the studies we performed, the size of P is usually very small, and the problem can be solved using standard linear programming toolboxes. The transformation of (B.1) into the standard form can be found in (Lucka, 2011), Section A.1.6.

Appendix C. Additional figures

Figs. 8, 9 and 10.

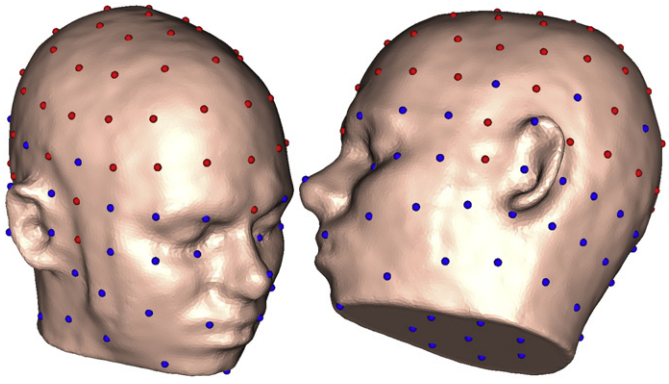


Fig. 8. Sensor positions used in the studies. The realistic r-cap configuration consists of 63 electrodes at the positions marked with the red spheres. The artificial f-cap configuration consists of 134 electrodes at the positions marked with both the red and blue spheres. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

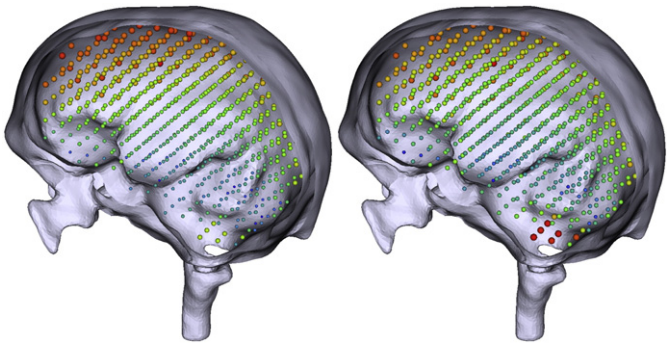


Fig. 9. The sum of the ℓ_2 norms of the three gain-vectors at a given position is depicted. Left: realistic sensor configuration (r-cap); Right: artificial sensor configuration (f-cap). The influence of the hole at the base of the skull (foramen magnum) on the magnitudes of the deep-lying sources is noticeable for both sensor configurations.

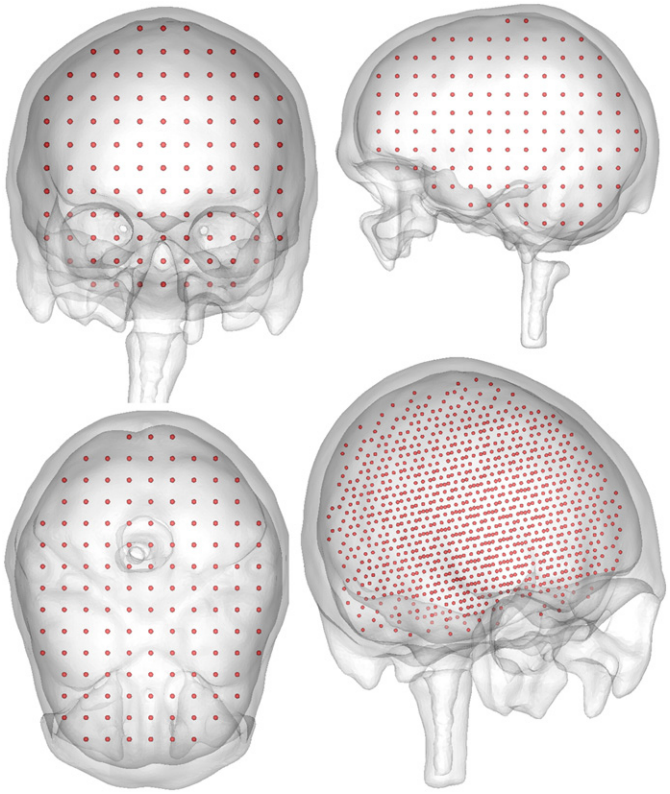


Fig. 10. The locations of the 1000 source space nodes used in the studies.

References

Acar, Z.A., Makeig, S., 2010. Neuroelectromagnetic forward head modeling toolbox. J. Neurosci. Methods 190, 258–270.
Ahlfors, S.P., Ilmoniemi, R.J., Hämäläinen, M., 1992. Estimates of visually evoked cortical currents. Electroencephalogr. Clin. Neurophysiol. 82, 225–236.
Akhtari, M., Bryant, H., Mamelak, A., Flynn, E., Heller, L., Shih, J., Mandelkem, M., Matlachov, A., Ranken, D.M., Best, E., et al., 2002. Conductivities of three-layer live human skull. Brain Topogr. 14, 151–167.
Ambrosio, L., Gigli, N., Savaré, G., 2008. Gradient Flows in Metric Spaces and in the Spaces of Probability Measures. Birkhäuser, Basel. 2nd edition.
Aubert, G., Kornprobst, P., 2006. Mathematical problems in image processing, 2nd edition. Applied Mathematical Sciences, vol. 147. Springer.
Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans. Biomed. Eng. 44, 374–385.
Bertrand, O., Thévenet, M., Perrin, F., 1991. 3D finite element method in brain electrical activity studies. In: Nenonen, J., Rajala, H., Katila, T. (Eds.), Biomagnetic Localization and 3D Modelling. Helsinki University of Technology, Helsinki, pp. 154–171.
Bruno, P., Vatta, F., Mininel, S., Inchingolo, P., 2003. Head model extension for the study of bioelectric phenomena. Biomed. Sci. Instrum. 39, 59–64.

Bruno, P., Vatta, F., Minimel, S., Inchingolo, P., 2004. Referenced EEG and head volume conductor model: geometry and parametrical setting. Conf Proc IEEE Eng Med Biol Soc, San Francisco, USA.
Buchner, H., Knoll, G., Fuchs, M., Rienäcker, A., Beckmann, R., Wagner, M., Silny, J., Pesch, J., 1997. Inverse localization of electric dipole current sources in finite element models of the human head. Electroencephalogr. Clin. Neurophysiol. 102, 267–278.
Calvetti, D., Somersalo, E., 2007a. A Gaussian hypermodel to recover blocky objects. Inverse Probl. 23, 733–754.
Calvetti, D., Somersalo, E., 2007b. Introduction to Bayesian scientific computing. Surveys and Tutorials in the Applied Mathematical Sciences., vol. 2. Springer, New York.
Calvetti, D., Somersalo, E., 2008a. Hypermodels in the Bayesian imaging framework. Inverse Probl. 24, 034013 (20 pp.).
Calvetti, D., Somersalo, E., 2008b. Recovery of shapes: hypermodels and Bayesian learning. J. Phys. Conf. Ser. IOP Publishing, p. 012014.
Calvetti, D., Hakula, H., Pursiainen, S., Somersalo, E., 2009. Conditionally Gaussian hypermodels for cerebral source localization. SIAM J. Imaging Sci. 2, 879–909.
Dalal, S., Guggisberg, A., Edwards, E., Sekihara, K., Findlay, A., Canolty, R., Berger, M., Knight, R., Barbaro, N., Kirsch, H., Nagarajan, S., 2008. Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. NeuroImage 40, 1686–1700.
Dalal, S., Jerbi, K., Bertrand, O., Adam, C., Ducorps, A., Schwartz, D., Garnero, L., Baillet, S., Martinerie, J., Lachaux, J., 2010. Insights from simultaneous recording of MEG and intracranial EEG. Front. Neurosci. Conference Abstract: Biomag 2010 — 17th International Conference on Biomagnetism.
Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.
Dannhauer, M., Lanfer, B., Wolters, C., Knösche, T., 2011. Modeling of the human skull in EEG source analysis. Hum. Brain Mapp. 32, 1383–1399.
de Munck, J.C., van Dijk, B.W., Spekreijse, H., 1988. Mathematical dipoles are adequate to describe realistic generators of human brain activity. IEEE Trans. Biomed. Eng. 35, 960–966.
Drechsler, F., Wolters, C.H., Dierkes, T., Si, H., Grasedyck, L., 2009. A full subtraction approach for finite element method based source analysis using constrained Delaunay tetrahedralisation. NeuroImage 46, 1055–1065.
Durka, P.J., Matysiak, A., Montes, E.M., Sosa, P.V., Blinowska, K.J., 2005. Multichannel matching pursuit and EEG inverse solutions. J. Neurosci. Methods 148, 49–59.
Ebersole, J.S., Ebersole, S.M., 2010. Combining MEG and EEG source modeling in epilepsy evaluations. J. Clin. Neurophysiol. 27, 360–371.
Engl, H., Hanke-Bourgeois, M., Neubauer, A., 1996. Regularization of inverse problems. Mathematics and Its Applications. Springer, Netherland, Berlin.
Friston, K.J., Mattout, J., Trujillo-Barreto, N.J., Ashburner, J., Penny, W.D., 2007. Variational free energy and the Laplace approximation. NeuroImage 34, 220–234.
Friston, K.J., Harrison, L., Daunizeau, J., Kiebel, S.J., Phillips, C., Trujillo-Barreto, N.J., Henson, R.N., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. NeuroImage 39, 1104–1120.
Fuchs, M., Drenckhahn, R., Wischmann, H., Wagner, M., 1998. An improved boundary element method for realistical volume conductor modeling. IEEE Trans. Biomed. Eng. 45, 980–997.
Fuchs, M., Wagner, M., Köhler, T., Wischmann, H.A., 1999. Linear and nonlinear current density reconstructions. J. Clin. Neurophysiol. 16, 267.
Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdés-Sosa, P., 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. NeuroImage 23, 435–453.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian data analysis, Chapman and Hall/CRC Texts in Statistical Science2nd edition. CRC Press. 2nd edition.
Gencer, N.G., Williamson, S.J., 1998. Differential characterization of neural sources with the bimodal truncated SVD pseudo-inverse for EEG and MEG measurements. IEEE Trans. Biomed. Eng. 45, 827–838.
Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M., Kowalski, M., 2011. Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries. Inf. Process. Med. Imaging. Springer, pp. 600–611.
Grave de Peralta, R., Hauk, O., Gonzalez, S.L., 2009. The neuroelectromagnetic inverse problem and the zero dipole localization error. Comput. Intell. Neurosci. 2009, 11 pages. http://dx.doi.org/10.1155/2009/659247 (Article ID 659247).
Greenblatt, R.E., Ossadtchi, A., Pflieger, M.E., 2005. Local linear estimators for the bioelectromagnetic inverse problem. IEEE Trans. Signal Process. 53, 3403–3412.
Hallez, H., 2008. Incorporation of Anisotropic Conductivities in EEG Source Analysis. Ph.D. thesis. Faculteit Ingenieurswetenschappen, Universiteit Gent, Belgium.
Hämäläinen, M., Ilmoniemi, R.J., 1984. Interpreting measured magnetic fields of the brain: minimum norm estimates of current distributions. Helsinki University of Technology, Technical Report TKK-F-A559.
Hämäläinen, M., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.
Hämäläinen, M., Haario, H., Lehtinen, M., 1987. Inference about sources of neuromagnetic fields using Bayesian parameter estimation. Helsinki University of Technology, Technical Report TKK-F-A620.
Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497.
Haueisen, J., Tuch, D.S., Ramon, C., Schimpf, P.H., Wedeen, V.J., George, J.S., Belliveau, J.W., 2002. The influence of brain tissue anisotropy on human EEG and MEG. NeuroImage 15, 159–166.
Henson, R.N., Mattout, J., Phillips, C., Friston, K.J., 2009a. Selecting forward models for MEG source-reconstruction using model-evidence. NeuroImage 46, 168–176.
Henson, R.N., Mouchlianitis, E., Friston, K.J., 2009b. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. NeuroImage 47, 581–589.
Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Hum. Brain Mapp. 31, 1512–1531.
Hufnagel, A., Elger, C.E., Pels, H., Zentner, J., Wolf, H.K., Schramm, J., Wiestler, O.D., 1994. Prognostic significance of ictal and interictal epileptiform activity in temporal lobe epilepsy. Epilepsia 35, 1146–1153.
Ioannides, A.A., Bolton, J.P.R., Clarke, C.J.S., 1990. Continuous probabilistic solutions to the biomagnetic inverse problem. Inverse Probl. 6, 523–542.
Janszky, J., Jokeit, H., Schulz, R., Hoppe, M., Ebner, A., 2000. EEG predicts surgical outcome in lesional frontal lobe epilepsy. Neurology 54, 1470–1476.
Jun, S.C., George, J.S., Kim, W., Paré-Blagoev, J., Plis, S.M., Ranken, D.M., Schmidt, D.M., 2008. Bayesian brain source imaging based on combined MEG/EEG and fMRI using MCMC. NeuroImage 40, 1581–1594.
Kaipio, J.P., Somersalo, E., 2005. Statistical and computational inverse problems. Applied Mathematical Sciences, vol. 160. Springer, New York.
Kantorovich, L., 1942. On the translocation of masses. Dokl. Akad. Nauk SSSR 37, 227–229.
Kantorovich, L., Gavurin, M., 1949. The application of mathematical methods in problems of freight flow analysis. Collection of Problems Concerned with Increasing the Effectiveness of Transports, pp. 110–138.
Kiebel, S.J., Garrido, M.I., Moran, R., Chen, C.C., Friston, K.J., 2009. Dynamic causal modeling for EEG and MEG. Hum. Brain Mapp. 30, 1866–1876.
Klenke, A., 2008. Probability Theory: A Comprehensive Course, 1st edition. Springer, London.
Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., Papadopoulo, T., 2005. A common formalism for the integral formulations of the forward EEG problem. IEEE Trans. Med. Imaging 24, 12–28.
Lanfer, B., Scherg, M., Dannhauer, M., Knösche, T., Wolters, C., 2010. Influence of deficiencies in segmenting the skull on EEG source modeling. Proc. of the 16th Annual Meeting of the Organization for Human Brain Mapping, Barcelona, Spain, June 6–10.
Lew, S., Wolters, C.H., Dierkes, T., Röer, C., MacLeod, R.S., 2009. Accuracy and run-time comparison for different potential approaches and iterative solvers in finite element method based EEG source analysis. Appl. Numer. Math. 59, 1970–1988.
Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. NeuroImage 31, 160–171.
Lucka, F., 2011. Hierarchical Bayesian Approaches to the Inverse Problem of EEG/MEG Current Density Reconstruction. Master's thesis. University of Muenster, Germany.
MacKay, D., 2003. Information Theory, Inference, and Learning Algorithms, 1st edition. Cambridge University Press.
Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. NeuroImage 30, 753–767.
Molins, A., Stufflebeam, S.M., Brown, E.N., Hämäläinen, M., 2008. Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation. NeuroImage 42, 1069–1077.
Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling and localization from spatio-temporal MEG data. IEEE Trans. Biomed. Eng. 39, 541–557.
Niedermeyer, E., Lopez da Silva, F.L., 2004. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 5th edition. Lippincot Williams & Wilkins, Philadelphia.
Nummenmaa, A., Auranen, T., Hämäläinen, M., Jääskeläinen, I.P., Lampinen, J., Sams, M., Vehtari, A., 2007a. Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. NeuroImage 35, 669–685.
Nummenmaa, A., Auranen, T., Hämäläinen, M., Jääskeläinen, I.P., Sams, M., Vehtari, A., Lampinen, J., 2007b. Automatic relevance determination based hierarchical Bayesian MEG inversion in practice. NeuroImage 37, 876–889.
Nunez, P.L., Srinivasan, R., 2005. Electric Fields of the Brain: The Neurophysics of EEG, 2nd edition. Oxford University Press, USA.
Oostenveld, R., Oostendorp, T., 2002. Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. Hum. Brain Mapp. 17, 179–192.
Ou, W., Hämäläinen, M.S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG inverse solver. NeuroImage 44, 932–946.
Parkkonen, L., Fujiki, N., Mäkelä, J., 2009. Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. Hum. Brain Mapp. 30, 1772–1782.
Pascual-Marqui, R.D., 1999. Review of methods for solving the EEG inverse problem. Int. J. Bioelectromagn. 1, 75–86.
Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find. Exp. Clin. Pharmacol. 24, 5–12 Suppl. D.
Phillips, C., Mattout, J., Rugg, M.D., Maquet, P., Friston, K.J., 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. NeuroImage 24, 997–1011.
Ramon, C., Schimpf, P.H., Haueisen, J., 2006. Influence of head models on EEG simulations and inverse source localizations. Biomed. Eng. Online 5, 10 (13 pp.).
Rampp, S., Stefan, H., 2007. Magnetoencephalography in presurgical epilepsy diagnosis. Expert Rev. Med. Devices 4, 335–347.
Rullmann, M., Anwander, A., Dannhauer, M., Warfield, S.K., Duffy, F.H., Wolters, C.H., 2009. EEG source analysis of epileptiform activity using a 1 mm anisotropic hexahedra finite element head model. NeuroImage 44, 399–410.
Sander, T.H., K.T., S.A., K.F., W.C., H.J., T.L., 2010. Recent advances in modeling and analysis of bioelectric and biomagnetic sources. Biomed. Tech. (Berl) 55, 65–76.
Santiuste, M., Nowak, R., Russi, A., Tarancon, T., Oliver, B., Ayats, E., Scheler, G., Graetz, G., 2008. Simultaneous magnetoencephalography and intracranial EEG registration: technical and clinical aspects. J. Clin. Neurophysiol. 25, 331–339.

Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys. Med. Biol. 32, 11–22.
Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. NeuroImage 23, 806–826.
Scherg, M., Buchner, H., 1993. Somatosensory evoked potentials and magnetic fields: separation of multiple source activities. Physiol. Meas. 14, A35.
Scherg, M., Cramon, D.V., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62, 32–44.
Schiffbauer, H., Berger, M., Ferrari, P., Freudenstein, D., Rowley, H., Roberts, T., 2002. Preoperative magnetic source imaging for brain tumor surgery: a quantitative comparison with intraoperative sensory and motor mapping. J. Neurosurg. 97, 1333–1342.
Schimpf, P., Ramon, C., Haueisen, J., 2002. Dipole models for the EEG and MEG. IEEE Trans. Biomed. Eng. 49, 409–418.
Schmitt, U., Louis, A.K., Wolters, C.H., Vauhkonen, M., 2002. Efficient algorithms for the regularization of dynamic inverse problems: II. Applications. Inverse Probl. 18, 659–676.
Sekihara, K., Nagarajan, S.S., 2008. Adaptive spatial filters for electromagnetic brain imaging, 1st edition. Series in Biomedical EngineeringSpringer.
Sekihara, K., Sahani, M., Nagarajan, S.S., 2005. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. NeuroImage 25, 1056–1067.
Stefan, H., Hummel, C., Scheler, G., Genow, A., Druschky, K., Tilz, C., Kaltenhauser, M., Hopfengartner, R., Buchfelder, M., Romstock, J., 2003. Magnetic brain source imaging of focal epileptic activity: a synopsis of 455 cases. Brain 126, 2396–2405.
Tao, J.X., Ray, A., Hawes-Ebersole, S., Ebersole, J.S., 2005. Intracranial EEG substrates of scalp EEG interictal spikes. Epilepsia 46, 669–676.
Taubin, G., 1995. A signal processing approach to fair surface design. Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM, pp. 351–358.
Trujillo-Barreto, N.J., Aubert-Vázquez, E., Valdés-Sosa, P.A., 2004. Bayesian model averaging in EEG/MEG imaging. NeuroImage 21, 1300–1319.
Trujillo-Barreto, N.J., Aubert-Vázquez, E., Penny, W.D., 2008. Bayesian M/EEG source reconstruction with spatio-temporal priors. NeuroImage 39, 318–335.
Vallaghé, S., Papadopoulo, T., 2010. A trilinear immersed finite element method for solving the electroencephalography forward problem. SIAM J. Sci. Comput. 32, 2379–2394.
Van den Broek, S., Reinders, F., Donderwinkel, M., Peters, M., 1998. Volume conduction effects in EEG and MEG. Electroencephalogr. Clin. Neurophysiol. 106, 522–534.
Vese, L.A., Chan, T.F., 2002. A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vis. 50, 271–293.
Vorwerk, J., 2011. Comparison of Numerical Approaches to the EEG Forward Problem. Master's thesis. University of Muenster, Germany.
Wagner, M., Fuchs, M., Kastner, J., 2004. Evaluation of sLORETA in the presence of noise and multiple sources. Brain Topogr. 16, 277–280.
Wang, J.Z., Williamson, S.J., Kaufman, L., 1992. Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. IEEE Trans. Biomed. Eng. 39, 665–675.
Weinstein, D., Zhukov, L., Johnson, C., 2000. Lead-field bases for electroencephalography source imaging. Ann. Biomed. Eng. 28, 1059–1065. http://dx.doi.org/10.1114/1.1310220.
Wipf, D., Nagarajan, S.S., 2009. A unified Bayesian framework for MEG/EEG source imaging. NeuroImage 44, 947–966.
Wolters, C.H., Grasedyck, L., Hackbusch, W., 2004. Efficient computation of lead field bases and influence matrix for the FEM-based EEG and MEG inverse problem. Inverse Probl. 20, 1099–1116.
Wolters, C.H., Köstler, H., Möller, C., Härdtlein, J., Anwander, A., 2007a. Numerical approaches for dipole modeling in finite element method based source analysis. Int. Congr. Ser. 1300, 189–192 ISBN-13:978-0-444-52885-8. http://dx.doi.org/10.1016/j.ics.2007.02.014.
Wolters, C.H., Köstler, H., Möller, C., Härtlein, J., Grasedyck, L., Hackbusch, W., 2007b. Numerical mathematics of the subtraction method for the modeling of a current dipole in EEG source reconstruction using finite element head models. SIAM J. Sci. Comput. 24–45.