

# Hierarchical Bayesian Inference for the EEG Inverse Problem using Realistic FE Head Models: Depth Localization and Source Separation for Focal Primary Currents

Felix Lucka<sup>a,b,\*</sup>, Sampsa Pursiainen<sup>c</sup>, Martin Burger<sup>a</sup>, Carsten H. Wolters<sup>b</sup>

<sup>a</sup>*Institute for Computational and Applied Mathematics, University of Muenster, Germany*

<sup>b</sup>*Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Germany*

<sup>c</sup>*Institute of Mathematics, Aalto University, Finland*

---

## Abstract

The estimation of the activity-related ion currents by measuring the induced electromagnetic fields at the head surface is a challenging, severely ill-posed inverse problem. Especially the recovery of brain networks involving deep-lying sources by means of EEG/MEG recordings 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 right depth localization, a well known systematic error of many CDR methods, and the separation of single sources in multiple-source scenarios. Both aspects are very important in clinical applications, e.g., in presurgical epilepsy diagnosis. For these tasks, HBM provides a promising framework, which is able to improve upon established CDR methods like 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. In addition, we introduce Wasserstein distances as performance measures for the validation of inverse methods in complex source scenarios.

**Keywords:** EEG, inverse problem, source localization, current density reconstruction, hierarchical Bayesian modeling, Full-MAP, Full-CM, depth localization, MCMC, Wasserstein distance, realistic head model

---

## 1. Introduction

*Electroencephalography (EEG)* and *magnetoencephalography (MEG)* recordings are used in a wide range of applications today, ranging from clinical routine to cognitive science (Niedermeyer and da Silva, 2004). One aim in EEG and MEG is to reconstruct brain activity by means of non-invasive measurements of the associated *bioelectromagnetic* fields. This task poses 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., Sarvas, 1987; Hämäläinen et al., 1993). The reconstruction of the so-called *primary* or *impressed* currents (a simplified source model, see Sarvas, 1987; de Munck et al., 1988; Hämäläinen et al., 1993) is called the *inverse problem of EEG/MEG*. In its generic formulation, the inverse problem lacks a unique solution: Infinitely many source configurations - often with extremely different properties - can explain the measured fields. All inverse methods rely on the usage of *a-priori information* on the source activity to choose a particular solution from the set of possible solutions. This a-priori information can reflect computational constraints as well as neurological considerations. Nevertheless, since the problem is heavily under-determined, the

results of the different methods for one and the same measurement data differ considerably. Up to date, there is no universal inverse method available: Most methods work well for certain source-configurations while failing to recover others. Therefore, a careful examination of the performance of the methods for different source scenarios is still 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 clinical applications like presurgical epilepsy diagnosis (Boon et al., 2002; Stefan et al., 2003; Rampp and Stefan, 2007) or the analysis of evoked potentials (Pantev et al., 1989; Buchner et al., 1997), the framework easily extends to recover spatially distributed sources encountered, e.g., in cognitive neuroscience. This work comprises results from a diploma thesis, Lucka (2011). In the following, 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.

### 1.1. Inverse Methods for EEG/MEG

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

---

\*Corresponding author. Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Germany. Tel.: +49/(0)251/83-57958

Email address: felix.lucka@uni-muenster.de (Felix Lucka)

struct solutions reflecting certain *a-priori information*. A first classification can be made into *focal current modeling*, *spatial scanning/beamforming* and *distributed current modeling*. Focal current modeling tries to reconstruct the real current by a small number of *equivalent current dipoles* having arbitrary location and orientation (Scherg and Cramon, 1985; Mosher et al., 1992; Jun et al., 2008). 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 (Sekihara and Nagarajan, 2008; Dalal et al., 2008). In distributed current models, the current is discretized by a large number of focal elementary sources having a fixed location and orientation, 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 CDR methods, two main concepts dominated 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).

### 1.2. Brain Networks Involving Deep-lying Sources

The location of the source space nodes is a crucial choice for CDRs: First, high resolution structural imaging scans (CT or MRI) from the cortex where the neural generators of the EEG/MEG signal are located (Nunez and Srinivasan, 2005) have to be taken. Due to its deep but thin sulci and strong folding, sophisticated segmentation algorithms are needed to process this data. Instead, often only flattened and smoothed representation of the cortical surface are used which do not include 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 (Lütkenhöner et al., 2000; Parkkonen et al., 2009; Santiuste et al., 2008; Dalal et al., 2010). One example are networks involving the *hippocampus* which plays an important role in memory and navigation (Duvernoy, 2005; Andersen, 2007). Concerning its pathology, it is often the focus of epileptic seizures: Hippocampal sclerosis is the most commonly visible type of tissue damage in temporal lobe epilepsy (Chang and Lowenstein, 2003; Stefan et al., 2009). To recover networks involving the hippocampus, a complete representation of the gray matter compartment by source space nodes is mandatory.

Accounting for the complete gray matter, many more deep-lying locations form the source space and a phenomenon called *depth bias* gains fundamental importance: Many inverse methods fail to reconstruct deep-lying sources in the right depth, reconstructing them too close to the skull (see, e.g., Figure 2(e)). This is a well known systematic error (e.g., Ahlfors et al., 1992; Wang et al., 1992; Gencer and Williamson, 1998) and was subject to many studies (e.g., Ioannides et al., 1990; Pascual-Marqui, 1999; Fuchs et al., 1999; Pascual-Marqui, 2002; Wagner et al., 2004; Greenblatt et al., 2005; Sekihara et al., 2005; Lin et al., 2006; Grave de Peralta et al., 2009). The depth bias can be a crucial error, e.g., in the presurgical diagnosis for epilepsy patients, where the task is to determine the right location of the resection volume (Boon et al., 2002; Stefan et al., 2003).

Another effect related to the depth bias is the *masking* of deep-lying sources by superficial ones: If the real source configuration consists of multiple, 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 can lead to crucial errors in the presurgical diagnosis for epilepsy patients suffering from *multi focal epileptiform discharges*: This form of epilepsy is correlated to a worse postoperative outcome regarding seizure freedom and complicates the presurgical diagnosis (Chang and Lowenstein, 2003). Often, an operation is not possible at all. The correct detection and separation of multiple sources is hence of greatest importance to guide the presurgical diagnosis and operation planning.

### 1.3. Contributions and Structure of this Study

This article examines fully-Bayesian inference for HBM for the source scenarios described above in a systematic way. In Section 2.1, we will outline CDR approaches from the perspective of Bayesian inference. This will then lead us to the hierarchical Bayesian modeling in Section 2.2, for which we will describe the *fully-Bayesian* inference methods (which we will call *CM* and *MAP1*) and propose improved Full-MAP estimation methods in Sections 2.3-2.4, which we will call *MAP2* and *MAP3*. In Section 2.5, a new performance measure, *earth mover's distance (EMD)*, will be introduced that is needed for an appropriate validation of inverse methods in complex source scenarios. Section 3 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 Section 4 results, limitations and future directions of research are discussed and Section 5 contains the final conclusions.

## 2. Methods

### 2.1. Bayesian Formulation of the Inverse Problem

We will briefly introduce the Bayesian formulation of the inverse problem, revisit some commonly known inverse methods and introduce the hierarchical model we will examine. 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  $r_i$ ,  $i = 1, \dots, k$  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 sensors caused by  $s$  can be calculated via:

$$b = Ls, \quad (1)$$

where  $L \in \mathbb{R}^{m \times n}$  denotes the *lead-field* or *gain* matrix (see Hämäläinen and Ilmoniemi, 1984; Sarvas, 1987; Hämäläinen et al., 1993). For calculating its entries, one needs to solve the forward problem, which includes head and source modeling and an appropriate (numerical) solution scheme (see Section 3.1.1). The ill-posedness of the inverse problem is reflected in  $L$ : Since  $m \ll n$ , (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, but rather reflects our *lack of information* about its concrete value. In our situation, we first model the (additive) measurement noise by a Gaussian random variable  $\mathcal{E} \sim \mathcal{N}(0, \Sigma_{\mathcal{E}})$ . For simplicity, we assume  $\Sigma_{\mathcal{E}} = \sigma^2 I_m$  here, where  $I_m$  is the  $m$ -dim. identity matrix. This leads to the *likelihood model*

$$B = LS + \mathcal{E} \quad (2)$$

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

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

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

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

The conditional density of  $S$  given  $B$  is called *posterior*. In Bayesian inference this density is the complete solution to the inverse problem. The term  $p(b)$  is called *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); Henson et al. (2009b,a, 2010). The common way to exploit the information contained in the posterior is to infer a point estimate for the value of  $S$  out of it. There are two popular choices. The first, called *maximum a-posteriori-estimate* (*MAP*), is the highest *mode* of the posterior, whereas the second, called *conditional mean-estimate* (*CM*), is the mean or expected value of

the posterior:

$$\hat{s}_{\text{MAP}} := \underset{s \in \mathbb{R}^n}{\operatorname{argmax}} p_{post}(s|b) \quad (5)$$

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

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:

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

Here,  $\mathcal{P}(s)$  is an energy functional penalizing unwanted features of  $s$ . Now, after suppressing terms not dependent on  $s$ , the MAP estimate is given by

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

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

This is a *Tikhonov-type* regularization of equation (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  $\Sigma_s$  and yields the *weighted minimum norm estimate* (*WMNE*, Dale and Sereno, 1993). Multiple *depth-weighting* matrices have been introduced chosen to reduce the depth-bias of the MNE (Ioannides et al., 1990; Fuchs et al., 1999). We will examine  $\ell_2$  weighting (Fuchs et al., 1999) and regularized  $\ell_\infty$  weighting (Fuchs et al., 1999):

$$\Sigma_s^{\ell_2} = \operatorname{diag} \left( (\|L_{(\cdot,i)}\|_2^2)^{-1} \right);$$

$$\Sigma_s^{\ell_\infty, \text{reg}} = \operatorname{diag} \left( \frac{\chi_i^2}{(\chi_i^2 + \beta^2)^2} \right),$$

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

The well known sLORETA method (Pascual-Marqui, 2002) relies on computing a non-diagonal weighted norm of a MNE and will be examined as well. More methods relying on formulation (9) are listed on page 9 in Lucka (2011).

## 2.2. 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 needed 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 used for the inversion of this data: By extending the

model by a new level of inference, the prior on  $S$  is not fixed but random, determined by the values of additional parameters  $\boldsymbol{\gamma} \in \mathbb{R}^h$ , called *hyperparameters*. These parameters follow an a-priori assumed distribution (the *hyperprior*) and are subject to estimation schemes, too. As this construction follows a top-down scheme, it is called *hierarchical* modeling:

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

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

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

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 called a *Gaussian scale mixture* or *conditionally Gaussian hypermodel* (Wipf and Nagarajan, 2009; Calvetti et al., 2009):  $p_{pr}(s|\boldsymbol{\gamma})$  is a Gaussian density with zero mean and a covariance determined by  $\boldsymbol{\gamma}$ :

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

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

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

The first important choice is choosing an appropriate set  $C$ . A variety of approaches encoding different a-priori information on the spatial source covariance pattern have been proposed, e.g., spatial smoothness components (Phillips et al., 2005; Mattout et al., 2006) 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):

$$C = \{\mathbf{e}_i \mathbf{e}_i^t \otimes I_d, i = 1, \dots, k\},$$

Thus 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 to the minimum norm estimate (which corresponds to  $C = \{I_n\}$ ), 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  $\gamma_i$ . Furthermore, since we do not want to bias our model to certain source locations a-priori, all single hyperparameters should be distributed equally. Finally, since we are looking for focal solutions, hyperpriors leading to *sparse* estimates of  $\boldsymbol{\gamma}$  will be used, i.e., hyperpriors forcing most hyperparameters to be (nearly) zero, while few are allowed to have

a large amplitude. Our particular choice for this purpose is the *inverse-gamma* distribution:

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

The parameters  $\alpha > 0$  and  $\beta > 0$  determine *shape* and *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); Wipf and Nagarajan (2009). Due to the diagonal shape of  $\Sigma_s$  the full posterior for this model reads (cf. (3), (14) and (15)):

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

where we abbreviated the sum of the  $\ell_2$ -norms of the  $d$  sources at location  $i$  with  $\|s_{i*}\|_2^2$ . 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 (16) is quadratic with respect to  $s$  and the  $\gamma_i$ 's are mutually independent. This simplifies and accelerates many practical computations with this model.

### 2.3. Inference for Hierarchical Models

Note that the posterior (16) depends on two kinds of parameters, the ones of main interest,  $s$ , and the hyperparameters,  $\boldsymbol{\gamma}$ . This offers more ways of inference than the simple CM and MAP estimation scheme introduced in Section 2.1. Five main approaches are established:

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

In the traditional Bayesian framework, all kinds of parameters should be treated equally that is why the first two schemes are also referred to as *fully-Bayesian* methods. Still, practically, the hyperparameters have been 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  $\boldsymbol{\gamma}$ -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.

#### 2.4. 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 computing the full-MAP and full-CM estimate numerically. 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. 2.1) is intractable by means of traditional quadratures. Integration by *Monte Carlo* methods can avoid these difficulties: A sequence of points  $(s_i, \boldsymbol{\gamma}_i)$ ,  $i = 1, \dots, M$  is constructed that is distributed like the posterior. Optimally, they should be drawn independently, because in this case, the *law of large numbers* would guarantee that

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

almost surely and in  $\ell_1$  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, since it 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, due 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  $p_{\text{post}}(s, \boldsymbol{\gamma}|b)$  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 *blocked Gibbs sampling* scheme (MacKay, 2003; Gelman et al., 2003) proposed in Nummenmaa et al. (2007a); Calvetti et al. (2009). It exploits the special structure of (16) by drawing from the posterior either conditioned on  $s$  or on  $\boldsymbol{\gamma}$  at a time:

##### Algorithm 1 (Blocked-Gibbs-Sampling algorithm).

Initialize  $\boldsymbol{\gamma}$  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  $j = 1, \dots, M + Q$  do:

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

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

This sampling technique is a very simple, but also very powerful one. A main advantage over other MCMC schemes is that it does not need any manual tuning of sampling parameters. The sampling problem in step 1 is solved by a reformulation into a least-squares problem, 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 will be a cyclic algorithm taking advantage of the special form of the posterior, called *Iterative Alternating Sequential (IAS)*. The form we use was introduced by Calvetti and Somersalo (Calvetti and Somersalo, 2007a, 2008a; Calvetti et al., 2009), inspired by a similar, more general algorithm called *half quadratic minimization* (Aubert and Kornprobst, 2006):

##### Algorithm 2 (Iterative Alternating Sequential).

Initialize  $\boldsymbol{\gamma}$  by  $\boldsymbol{\gamma}_0$  and set  $j = 1$ . Define the desired iteration number  $T$ ; For  $j = 1, \dots, T$  do:

1. Update  $s$  by  $s^{[j]} = \operatorname{argmax}\{p_{\text{post}}(s|\boldsymbol{\gamma}^{[j-1]}, b)\} = \operatorname{argmax}\{p_{\text{post}}(s, \boldsymbol{\gamma}^{[j-1]}|b)\}$ ;
2. Update  $\boldsymbol{\gamma}$  by  $\boldsymbol{\gamma}^{[j]} = \operatorname{argmax}\{p_{\text{post}}(\boldsymbol{\gamma}|s^{[j]}, b)\} = \operatorname{argmax}\{p_{\text{post}}(\boldsymbol{\gamma}, s^{[j]}|b)\}$ ;

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

(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 by 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 specify an initialization rule for  $\boldsymbol{\gamma}$ , yet. 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 end up in one of the local minima around the initialization point. While 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, a potential problem in our setting is the *multimodality* of the posterior (16): It results from the non-convexity of the energy of the inverse gamma hyperprior, i.e., the negative natural logarithm of its 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  $\alpha$  and  $\beta$  and the interplay with the under-determinedness of the likelihood equation (2) determine to what extent it affects the estimation process practically.

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  $\gamma^{[0]} = \gamma_{\text{CM}}$ .
- **MAP3:**  $U$  very rough approximations to the CM estimate  $(\hat{s}, \hat{\gamma})_{\text{CM}}^i, i = 1, \dots, U$  are computed first by using very small sample sizes  $M$ . Then they are used as seeds for the IAS algorithm:  $\gamma^{[0],i} = \gamma_{\text{CM}}^i, i = 1, \dots, U$ . The results  $(\hat{s}, \hat{\gamma})_{\text{MAP}}^i, i = 1, \dots, U$  are compared with respect to their posterior probability, and the result achieving the highest probability is chosen as a final MAP estimate.

Choosing CM estimates as initialization for MAP estimation seems unmotivated at this point, yet the methods MAP2 and MAP3 will turn out to perform best in all simulation studies. Especially, they are able to improve upon the performance of the CM estimate they rely on. We will outline reasons for this in the discussion section.

### 2.5. Validation Means and Inverse Crimes

While subsequent work will perform validation of the fully-Bayesian methods by means of real data, this paper focuses on extensive simulation studies to work out their basic properties. When using synthetic data produced by an invented source-configuration, it is crucial to avoid an *inverse crime*, i.e., 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 but rather place them independent of them. As a number of commonly used measures do rely on an inverse crime, as they assume that reference and 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, works in 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): It measures the minimal amount of (physical) work to transfer the mass of one density into the other. Illustratively, one can think of one density as a pile of sand, and of the other as a bunch of holes. Then the EMD is the minimal amount of work one needs 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 them, i.e., it reduces to the DLE). Mathematical details and a closer examination of its 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 1.2) we define the *depth* of a location in the head model as the minimal distance to one of the sensors.

## 3. Results

### 3.1. Setting for the Studies

#### 3.1.1. 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 a realistic modeling of the head volume conductor and its computational speed. 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 (csf, gray matter and white matter) because we want to focus on the effect of depth bias separate from others, 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). In addition, to facilitate the interpretation of the results, we need 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 placed 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. Especially 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 an artificial sensor configuration consisting of 134 EEG sensors distributed uniformly over the surface of the head model (see Figure D.8).

In the following, we will outline the model generation pipeline, which is also depicted in Figure 1: T1- and T2- weighted magnetic resonance images (MRI) of a healthy subject were measured on a 3T 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  $1, 17 \times 1, 17 \times 1, 17$

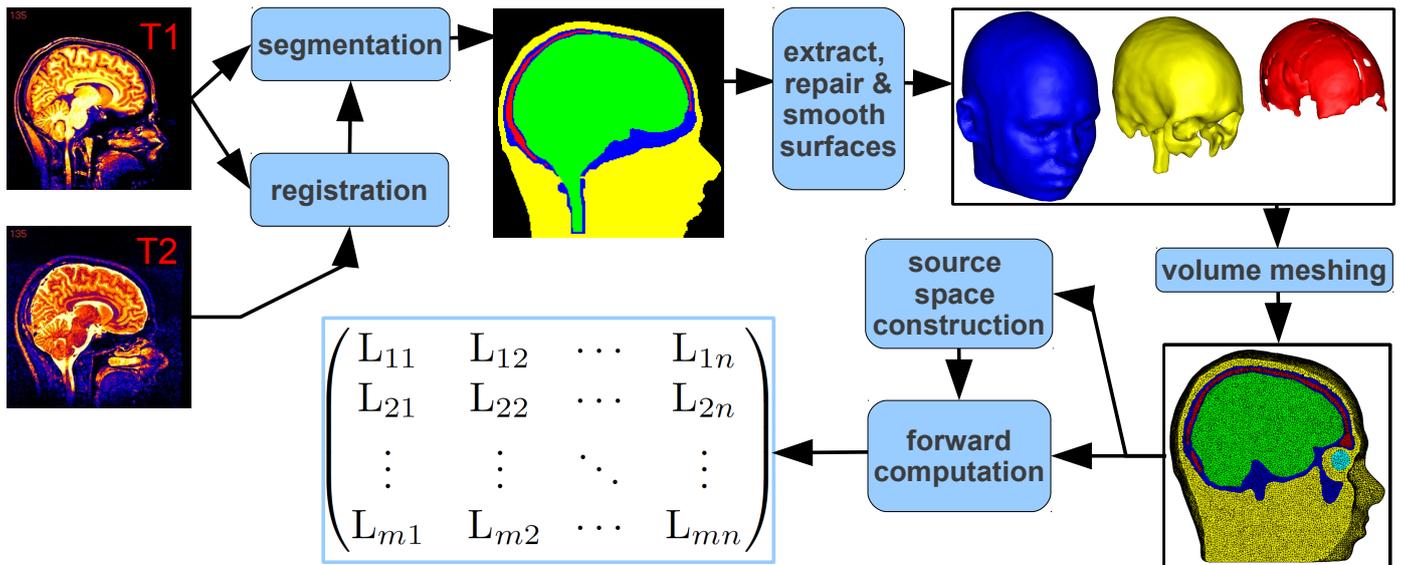


Figure 1: Model generation pipeline.

mm, were used. The T2-MRI was registered onto the T1-MRI using an affine registration approach and mutual information as a cost-function as implemented in *FSL*<sup>1</sup>. The compartments of skin, skull compacta and skull spongiosa were segmented using a grey-value based active contour model (Vese and Chan, 2002) and thresholding techniques. The segmentation was carefully checked and corrected manually. Because of the importance of skull holes on source analysis (Van den Broek et al., 1998; Oostenveld and Oostendorp, 2002), the foramen magnum and the two optic canals were correctly modeled as skull openings. Following (Bruno et al., 2003, 2004; 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*<sup>2</sup> was then used for the segmentation of the cortex surface as well as the extraction of high resolution meshes of the surfaces of skin, eyes, skull compacta, skull spongiosa and 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 skin, eyes, skull compacta and skull spongiosa were then used to create a high quality 3D Delaunay triangulation via *TetGen*<sup>3</sup>. In total, the resulting tetrahedral finite 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 skin (Dannhauer et al., 2010), 0.505 for eyes (Ramon et al., 2006), 0.0064 for skull compacta and 0.02865 for skull spongiosa (Akhtari et al., 2002; Dannhauer et al., 2010)

and 0.33 for the inner brain compartment (Dannhauer et al., 2010). Within the inner compartment, a source space consisting of 1.000 source locations based on a regular grid is chosen, the grid size is 10.99 mm (see Figures D.10). At each node,  $d = 3$  orthogonal dipoles in Cartesian directions are placed. For computing the corresponding lead-field matrix different FE approaches for modeling the source singularity are known from the literature: a subtraction approach (Bertrand et al., 1991; Schimpf et al., 2002; Wolters et al., 2007; Drechsler et al., 2009), a partial integration direct method (Weinstein et al., 2000; Schimpf et al., 2002; Vallaghé and Papadopoulou, 2010) and a Venant direct method (Buchner et al., 1997). In this study we used the Venant approach based on comparison of the performance of all three in multilayer sphere models, which suggested that for sufficiently regular meshes, it yields suitable accuracy over all realistic source locations Lew et al. (2009); Vorwerk (2011). This approach has the additional advantage of 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*<sup>4</sup>. In Figure D.9, the sum of the  $\ell_2$ -norms of the three gain-vectors is depicted.

### 3.1.2. Inverse Methods

In this section, we list the methods we use together with the choice of their internal parameters. For the hierarchical model, choosing  $\alpha$  and  $\beta$  is in fact a difficult practical and methodical question. Our choice relies on preliminary computations and considerations which can be found in Lucka (2011), Section 4.4.2. We tried to choose the parameters for each method in an optimal way to have a fair comparison of their performance. A further reference dealing with this issue is Nummenmaa et al.

<sup>1</sup>FLIRT - FMRIB's Linear Image Registration Tool, <http://www.fmrib.ox.ac.uk/fsl/flirt/index.html>

<sup>2</sup>CURrent Reconstruction and Imaging (CURRY), <http://www.neuroscan.com/>

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

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

(2007a).

The following methods will be examined in our studies:

- Full-CM estimation via the algorithm described in 2.4 for the HBM introduced in 2.2. Abbreviated by CM from now on. Parameters:  $\alpha = 0.5$  and  $\beta = 5 \cdot 10^{-8}$ .
- Full-MAP via the three methods MAP1-3 described in 2.4 for the HBM introduced in 2.2. Parameters:  $\alpha = 0.5$ ,  $\beta = 5 \cdot 10^{-6}$  for MAP1 and  $\alpha = 0.5$  and  $\beta = 5 \cdot 10^{-8}$  for MAP2 and MAP3.
- MNE as described in Section 2.1 (Hämäläinen et al., 1993).
- WMNE with  $\ell_2$  and regularized  $\ell_\infty$  weighting as described in Section 2.1 (Fuchs et al., 1999).
- sLORETA as described in Section 2.1 (Pascual-Marqui, 2002).

The regularization parameter  $\lambda$  for MNE, WMNE and sLORETA are chosen by the *discrepancy principle* (e.g., Engl et al., 1996; Kaipio and Somersalo, 2005), since we assume to know the noise variance  $\sigma^2$  (or assume to have a good estimate of it, e.g., based on pre-stimulus data).

To get an initial visual impression of the different methods, their results for a single dipole source (located in-between the source space nodes to avoid an inverse crime, cf. Section 2.5) are depicted in Figure 2.

### 3.2. Study 1: Single Dipole Reconstruction

#### 3.2.1. 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. Section 2.5). The following restriction on their depth was posed: First, the nearest sensor is searched. For that sensor, the nearest source space node is searched. The position for the dipole is only accepted if its depth (cf. Section 2.5) is larger than the depth of the source space node plus 10 mm. This way, 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 is generated using the same forward computation procedure used for the lead-field generation, and Gaussian noise at the *noise level* of 5% is added: In line with Calvetti et al. (2009) we will speak of a (relative) *noise level* of  $x$  if the standard deviation of the measurement noise (i.e.,  $\sigma$  in our notation) fulfills  $\sigma = x \cdot \|b_0\|_\infty$ , where  $b_0$  are the measurements in the noiseless case. Since we found no systematic effect of adding noise on depth bias and masking, a comparison to other noise level is omitted here. The full results can be found in Lucka (2011), Section 4.5.

Table 1: Statistics of validation measures for 1000 single unit-strength dipoles (mean  $\pm$  std)

Method	DLE	SD	EMD
CM	6.16 $\pm$ 2.37	1.3e-3 $\pm$ 1.1e-3	7.32 $\pm$ 2.31
MAP1	27.00 $\pm$ 11.90	9.8e-3 $\pm$ 5.8e-2	28.18 $\pm$ 11.54
MAP2	5.85 $\pm$ 2.16	2.2e-4 $\pm$ 3.3e-4	6.08 $\pm$ 2.22
MAP3	5.79 $\pm$ 2.13	7.1e-6 $\pm$ 4.5e-5	5.84 $\pm$ 2.21
MNE	29.46 $\pm$ 11.24	2.4e-1 $\pm$ 1.0e-1	53.20 $\pm$ 2.74
WMNE $\ell_2$	30.65 $\pm$ 13.52	2.5e-1 $\pm$ 1.1e-1	52.17 $\pm$ 2.53
WMNE $\ell_{\infty,reg}$	29.40 $\pm$ 14.81	2.2e-1 $\pm$ 8.0e-2	49.56 $\pm$ 3.64
sLORETA	6.10 $\pm$ 2.35	1.9e-1 $\pm$ 6.8e-2	40.58 $\pm$ 2.48

#### 3.2.2. 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 DLE and EMD for all methods. Table 1 shows DLE, SD and EMD, averaged over all dipoles. To give an idea of the practicality of the HBM methods, we note that our current implementations of CM, MAP2 and MAP3 in Matlab take about 5 minutes of computation time for each inverse reconstruction on a normal desktop PC.

*Depth bias.* We now focus on the first phenomenon introduced in Section 1.2: The depth bias. We will rely on a visual presentation using scatter plots. In Figures 3(a) - 3(h), the depth (cf. 2.5) of the reference source is plotted on the horizontal axis, whereas the depth of the source space node with the largest source estimate amplitude is plotted on the vertical axis. A mark within the area underneath 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. By  $q_{ab}$  we denote the percentage of 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, it suffers from depth bias (e.g., it is well known that MNE suffers from it which can be seen clearly in 3(a), 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 in the right depth but also in total.

### 3.3. Study 2: Masking of Deep-lying Sources in Two-Dipole Scenarios

#### 3.3.1. 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 by their depth. 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 at a noise level of 5% is added to the measurements.

Table 2: Statistics of validation measures for study 2 (mean  $\pm$  std)

Method	EMD	SD
CM	14.57 $\pm$ 4.98	3.0e-3 $\pm$ 1.9e-3
MAP1	42.10 $\pm$ 11.00	1.4e-3 $\pm$ 6.2e-4
MAP2	12.25 $\pm$ 6.30	8.3e-4 $\pm$ 3.0e-4
MAP3	12.41 $\pm$ 6.50	7.6e-4 $\pm$ 2.8e-4
MNE	44.63 $\pm$ 2.23	2.1e-1 $\pm$ 6.4e-2
WMNE $\ell_2$	43.75 $\pm$ 1.97	2.5e-1 $\pm$ 8.7e-2
WMNE $\ell_{\infty,reg}$	41.79 $\pm$ 2.06	2.4e-1 $\pm$ 7.6e-2
sLORETA	36.38 $\pm$ 2.51	1.9e-1 $\pm$ 5.6e-2

### 3.3.2. Results

*Initial example.* We show an initial example, where the effect of masking is very pronounced<sup>5</sup>. In Figure 4(a) the reference sources are represented by two green cones. One is very close to the sensors whereas the other one is very distant. Figure 4(b) shows the (vector) MNE result with red-yellow cones, Figure 4(c) shows the (scalar) sLORETA result as red-yellow spheres. Even a careful successive thresholding of the estimated source amplitudes does not reveal any evidence for the presence of the deep-lying source. In practice, these two results would probably not provoke a user to try out other inverse methods in addition. Hence the deep-lying source is most likely overlooked. The CM result (cf. Figure 4(d)) seems only capable of marking an ambiguous region around and in-between the support of the true sources<sup>6</sup>. The MAP3 method (cf. Figure 4(e)) is able to detect both sources (remember that the test sources are placed in between the source space grid nodes, cf. 2.5). The results of the other methods are omitted here.

*General properties.* Table 2 shows EMD and SD, averaged over all source configurations (remember that the DLE is not available in a multiple source scenario anymore, cf. 2.5).

### 3.4. Comparison between MAP approximations

We briefly compare the different MAP estimation algorithms concerning the posterior probability of their results. They all use different seed-points for their optimization, but rely on the same HBM, still only methods that rely on the same parameter set can be compared. Since MAP1 uses a different setting than MAP2 and MAP3, the results for MAP1 were recomputed using the same parameter setting as MAP2 and MAP3. These results will be named MAP1\_cmp. However, note that MAP1\_cmp performs even worse than MAP1 concerning EMD, DLE and SD. In Table 3 the average rank of the three methods within the three studies is depicted: 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. The method

<sup>5</sup>It was chosen by visual inspection after viewing the results for the first five source configurations of the study.

<sup>6</sup>The CM result actually looks as if the MCMC method has not converged yet. To check this, a large sample with  $M = 20\,000\,000$  was used as well. The results look very similar, and are therefore not depicted. It is still possible that the Markov chain is not ergodic for practical reasons.

Table 3: Mean ranking of different MAP estimation algorithms in the first study.

Method	Study 1	Study 2
MAP1_cmp	2.390	2.640
MAP2	1.398	1.547
MAP3	1.002	1.093

that found the approximation with the highest probability is ranked at the first place. Methods that found an approximation with the same probability are ranked at the same place. Subsequently the mean rank of each method is computed over all 1000 dipoles.

### 3.5. Recovery of 3 Dipoles

As a last illustration, we show the results for a source configuration consisting of three dipoles: In Figures 5(a)-5(d) the turquoise cones represent three sources of which two are quite close to the sensors whereas one is very distant. Figure 5(a) shows the (vector) MNE result with red-yellow cones, Figure 5(b) shows the (scalar) sLORETA result as red-yellow spheres. Again, a careful successive thresholding of the estimated source amplitudes does not reveal any evidence for the presence of the third, deep-lying source, the sLORETA estimate even hardly recovers the second, less deep-lying one. The CM result (cf. Figure 5(c)) finds some evidence for all three sources (although this is hardly visible in the picture). The MAP2 method (cf. Figure 5(d)) is able to detect all sources and yields an EMD of 9.19, which suggests that the localization of the single sources is quite good. The results of the other methods are omitted here.

## 4. Discussion

We examined new hierarchical Bayesian inference methods (HBM) for the EEG inverse problem and compared them to the results of established current density reconstruction (CDR) methods. In particular, we compared the fully-Bayesian conditional mean (CM) and maximum a-posteriori (MAP) estimates to minimum norm estimates (MNE, Hämäläinen and Ilmoniemi, 1994), different weighted minimum norm estimates (WMNE, Fuchs et al., 1999) and sLORETA (Pascual-Marqui, 2002). For MAP estimation we examined three different approaches, MAP1, which was proposed in (Calvetti et al., 2009), and MAP2 and MAP3 which we proposed in Section 2.4.

### 4.1. Study 1 (Single Dipole Reconstruction)

*HBM methods.* The MAP2 and MAP3 methods perform well with respect to the performance measures (cf. Table 1) and further, they do not seem to suffer from depth bias (cf. Figures 3(g) and 3(h)), with the MAP3 method slightly outperforming the MAP2 method. Compared to the other MAP approximation schemes, MAP3 also clearly attains the highest posterior probability (cf. Table 3), which suggests that it should be seen as the best approximation to the real MAP estimate examined here. The CM method shows good results, however, an interesting observation is the fact that the MAP2 method, which directly relies on the CM estimate, can clearly improve upon it.

Since the additional computation time is negligible, this result suggests to always perform a subsequent optimization after an initial CM estimation. The MAP1 scheme did not show convincing results, both with regard to DLE, SD and EMD (cf. Table 1) as well as with respect to depth bias (cf. Figure 3(f)). Compared to MAP2 and MAP3, it also attains less high posterior probabilities on average (cf. Table 3), 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. Figures 1-4 on pages 893-894 in Calvetti et al., 2009). The CM approximation with an inverse gamma hyperprior (which corresponds to the CM method used here) yielded the best result, both in location and in extend of the estimated source. Moreover, it seemed to have no depth bias whereas MAP approximation by the uniformly initialized IAS algorithm (which corresponds to the MAP1 method used here) seemed to suffer from it. In our work, we confirmed the impression about the CM estimate by a study in 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 rather of the algorithm used to compute it. Due to the results in Section 3.4, we can be sure that the MAP3 result is closer to the real MAP estimate in terms of posterior probability than MAP1, and it even performs slightly better than the CM estimate with regard to depth bias ( $q_{ab} = -0.007$  to  $q_{ab} = -0.058$ , cf. Figure 3).

*Minimum norm based methods.* The WMNE schemes used in this study are modifications of the original MNE explicitly aiming to improve the depth localization. Figures 3(a) - 3(c) clearly show that they succeed in this aspect (although the Figure 3(c) and  $q_{ab} = 0.095$  suggests that the WMNE with regularized  $\ell_\infty$  slightly exaggerates this aspect). These results confirm former 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 Figure 2 do not yield a clear impression on the different characteristics of the estimates either. Hence more detailed examinations are needed. The sLORETA estimate (which is also essentially minimum norm based, as it consists of computing a non-diagonal weighted norm of a MNE, see Pascual-Marqui (2002)) performs well concerning DLE and depth bias (cf. Table 1 and Figure 3(d)). Yet, Figure 2(h) 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., (Pascual-Marqui, 2002; Sekihara et al., 2005; Wagner et al., 2004; Lin et al., 2006).

*Direct comparison.* The direct comparison for the single focal reference source scenario shows that compared to established methods like MNE and sLORETA the HBM-based methods like CM, MAP2 and MAP3 clearly show better results con-

cerning EMD and SD (cf. Table 1) and the visual impression is more convincing as well (cf. Figures 2(a) - 2(h)). However, it is important to stress that the above results were only attained for the specific source scenario examined in this study. Without further examinations, their significance might be very limited, since the ability to localize single dipoles is a rather trivial and 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 HBM clearly motivate to examine their use in more detail.

#### 4.2. Study 2 (Masking of Deep-lying Dipoles)

The initial example showed that the source scenario examined in this study is a very challenging one for inverse methods (see Figures 4(b) and 4(c), 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 best in this study (cf. Table 2 and Figure 4(e)). The comparison with the results from 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 one.

Compared to each other the MAP3 scheme still outperforms the MAP2 scheme with regard to the posterior probability (cf. Table 3), but no longer concerning the EMD. This needs to be examined in more detail. Similar to the first study, Table 2 shows that again, the MAP2 result improves upon the corresponding CM result it is based on. The results also suggest that the posterior distribution for these scenarios is more complex than for single sources.

#### 4.3. Recovery of three dipoles

Despite the fact, that only a single source configuration of three dipoles was presented here, the results confirm the impression of Study 2, that HBM is better able to detect and separate multiple sources than MNE and sLORETA (see Figure 5).

#### 4.4. The Value of the EMD as a Performance Measure

In this work, we introduced the earth mover's distance (EMD) in order to have a measure that is both sensitive to localization and spatial extent of estimate (cf. Section 2.5). Table 1 shows that the EMD fulfills these needs: Only methods attaining 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 on the right 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 for methods that produce focal estimates but mis-localize considerably (e.g., the MAP1 scheme). The big advantage of the EMD is that it is applicable to more complex source scenarios just as well. In contrast, the extension of other localization measures like the DLE is not straight forward, neither from the implementation site nor for the interpretation of the results. For the two and three sources scenarios investigated in this work, the EMD remained the only measure sensible to localization that did not rely on an inverse crime (cf. Section 2.5).

#### 4.5. Limitations and Outlook

Confirming the present results with real data provides an important future work to complement the present simulation study. Our studies especially aimed at situations that are encountered, e.g., in presurgical epilepsy diagnosis (cf. 1.2), i.e., focal source configurations that are measured at a single time instant (e.g., the time slices from averaged *inter-ictal spike activity*). A validation with such data will be carried out and reported in near future.

Motivated by 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 extended HBM for their recovery which might be of more interest for research in the area of cognitive neuroscience.

Only two of the possible estimation methods that the HBM offers (cf. Section 2.3) were examined concerning our specific questions (cf. Section 1.2). As most other publications using HBM deal with Variational Bayesian inference methods (VB, see, e.g. Sato et al. (2004); Nummenmaa et al. (2007a); Friston et al. (2008)) a direct comparison will be next topic for simulation studies.

The present results concerning MAP2 and MAP3 estimates, 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, yet 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 into this direction. Additionally, the actual cause for the depth bias, and why some methods suffer from it, has to be examined also from a theoretical perspective.

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

Only CDR methods were compared, while no comparison to dipole fitting methods and scanning/beamforming methods was carried out (cf. Section 1.1 for references). This will be an interesting direction for further studies.

## 5. 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 like MNE and sLORETA in many aspects. In particular, they show good localization properties for single dipoles and do not suffer from depth bias. As it

has been shown in this study, small localization errors for single source scenarios are not sufficient to judge about the quality of an inverse method for EEG or MEG source analysis. However, in contrast to established inverse methods like minimum norm estimation and sLORETA, HBM based methods are able to maintain 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.

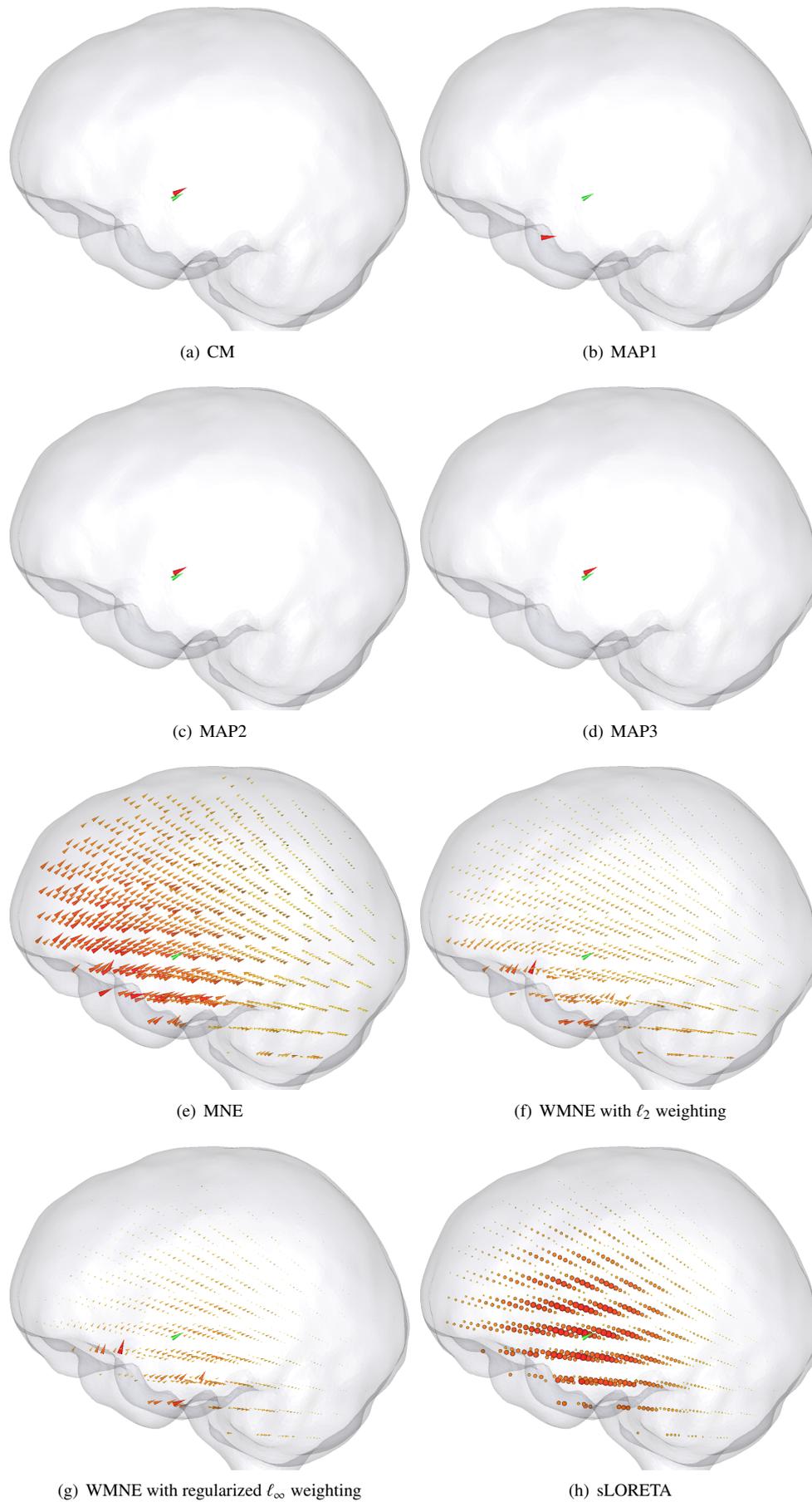


Figure 2: Results of different inverse methods for a single reference dipole source.

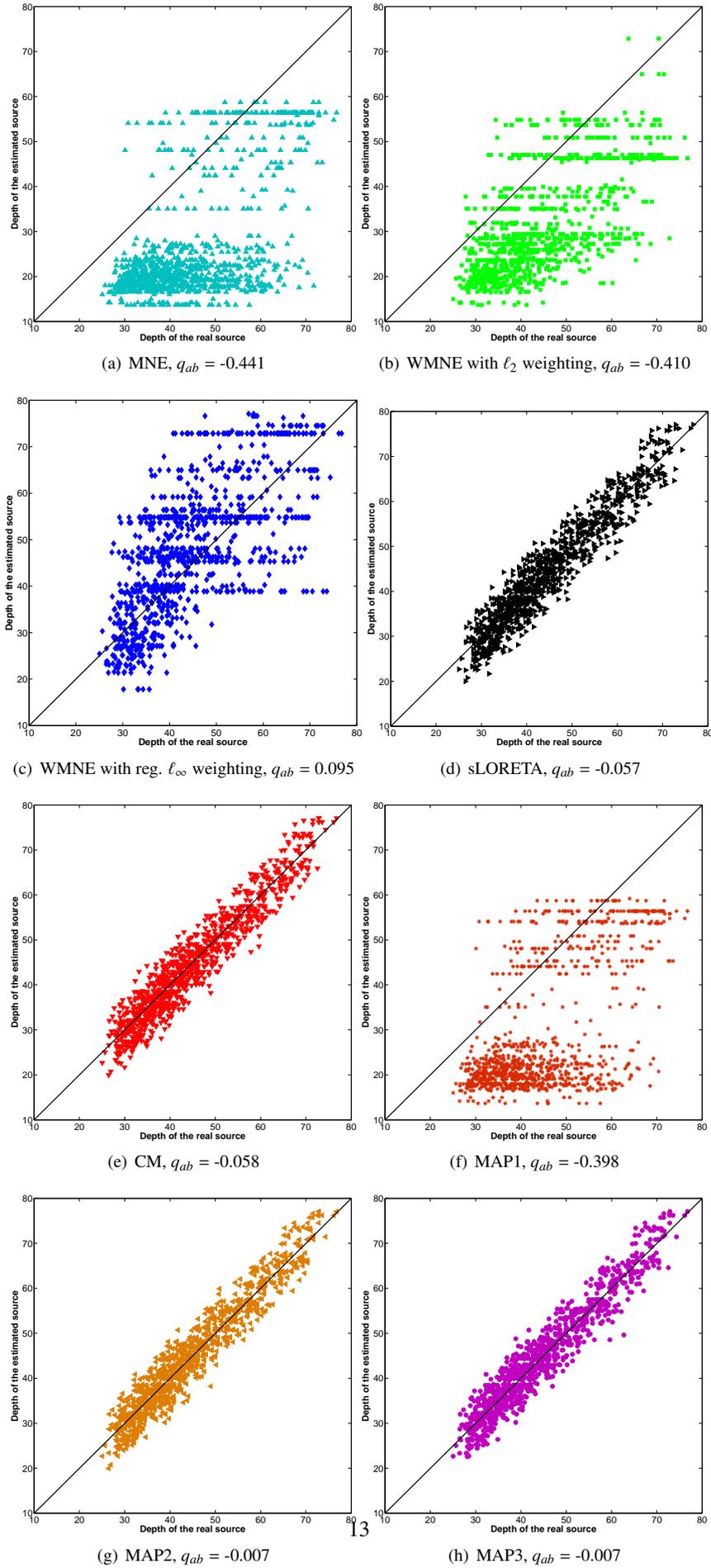
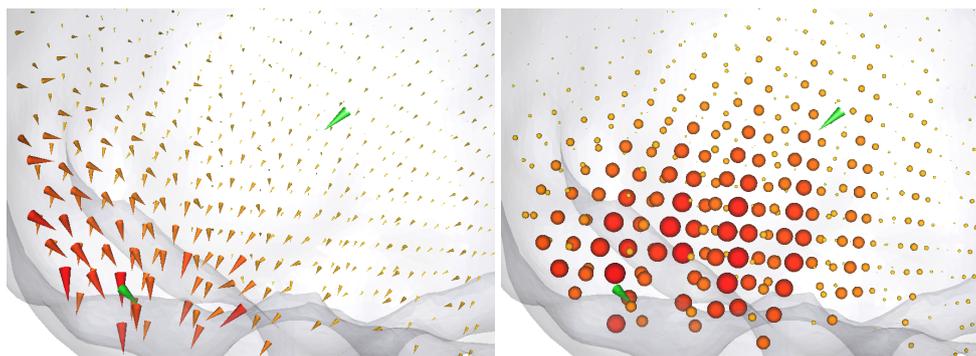


Figure 3: Scatter plots to visualize the depth bias of different inverse methods.

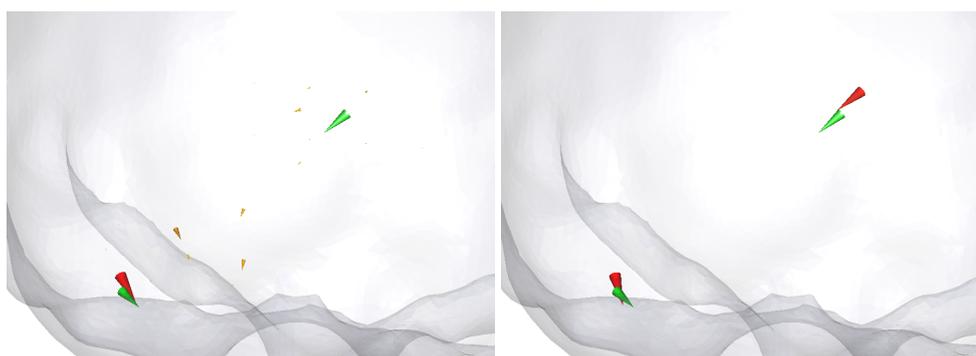


(a) Reference sources: In the left subfigure, the bottom left source is very close to the sensors, whereas the top right one is very distant.



(b) MNE

(c) sLORETA



(d) CM

(e) MAP3

Figure 4: Estimates of different inverse methods for a source configuration consisting of one near-surface and one deep-lying dipole.

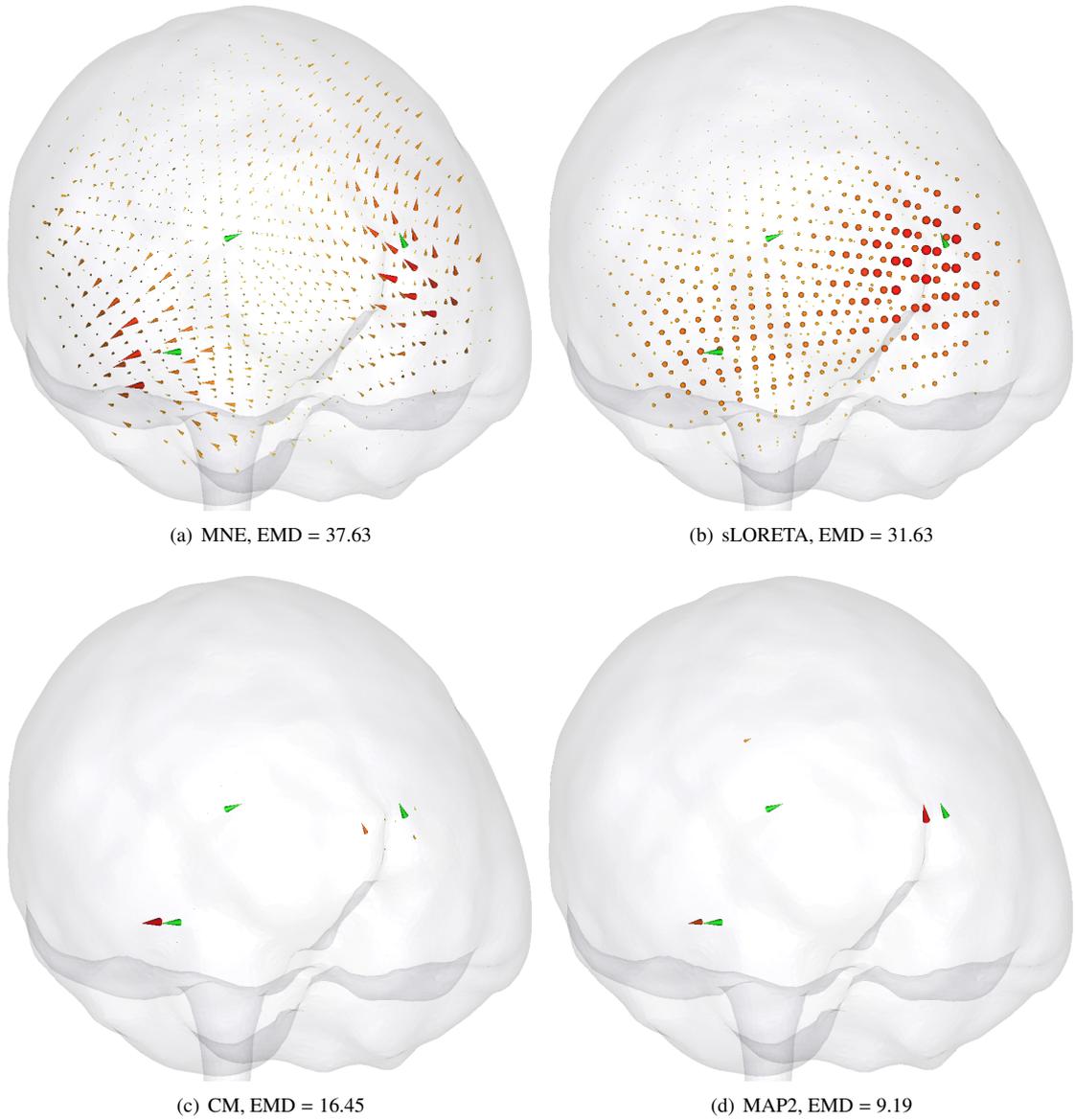


Figure 5: Estimates of different inverse methods for a source configuration consisting of three dipoles of different depth.

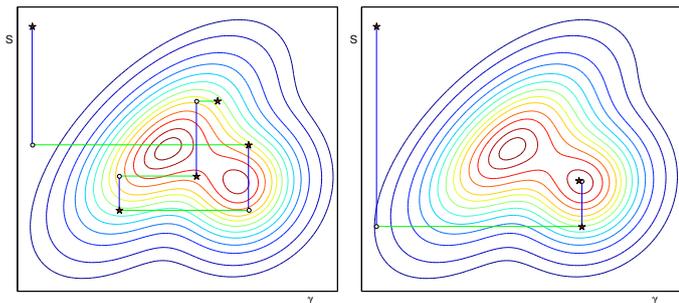


Figure A.6: Sketch of alternated conditional moves for a multimodal posterior (plotted via contour lines). Red stars mark subsequent states, circles mark half steps. Left: Algorithm 1; the blue lines correspond to step 1, the green lines to step 2. Right: Algorithm 2; the blue lines correspond to step 1, the green lines to step 2.

## 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  $(s_i, \gamma_i) \in \mathbb{R}^n \times \mathbb{R}^k$ ,  $i = 1, \dots, t$ : In a first half step (step 1 in both algorithms) the value of  $s$  is changed keeping  $\gamma$  fixed, while in the second half step (step 2 in both algorithms), the value of  $\gamma$  is changed while keeping  $s$  fixed. This is sketched in Figure A.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 Figure A.6, it is apparent why the IAS algorithm might get 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  $\gamma$  are strongly correlated).

As a consequence of this similar foundation, steps 1 and 2 in both algorithms can be solved in a surprisingly similar fashion.

### Step 1

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

$$\begin{aligned} \mathbb{E}_{p(s|\gamma,b)}(s) &= \Sigma_s L^t (L \Sigma_s L^t + \sigma^2 \text{Id}_m)^{-1} b \\ \text{Cov}_{p(s|\gamma,b)}(s) &= \Sigma_s - \Sigma_s L^t (L \Sigma_s L^t + \sigma^2 \text{Id}_m)^{-1} L \Sigma_s \\ &= \left( \Sigma_s^{-1} + \frac{1}{\sigma^2} L^t L \right)^{-1}, \end{aligned}$$

has to be solved (a derivation is given in Kaipio and Somersalo, 2005 and Lucka (2011), Section A.1.4). Remind that  $\Sigma_s = \Sigma_s(\gamma)$  changes every step  $j$ , so 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:

$$\begin{bmatrix} L \\ \sigma \Sigma_s^{-1/2} \end{bmatrix} s^{[j]} \stackrel{!}{=} \begin{bmatrix} b \\ 0 \end{bmatrix} + \sigma \begin{bmatrix} \omega_m \\ \omega_n \end{bmatrix}, \quad (\text{A.1})$$

where we set  $\omega_m = 0$ ,  $\omega_n = 0$  to attain the conditional mode and draw  $\omega_m$  and  $\omega_n$  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 on 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 easily be transferred to other fields of inverse problems, where the forward mapping is not given in explicit matrix form (Kaipio and Somersalo, 2005; Calvetti and Somersalo, 2007a,b, 2008a,b). In addition 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 EEG (we usually use  $m < 150$ ), 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:

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

This formula can be implemented in a straight forward manner:

### Algorithm 3 (Explicit Step 1 Solution).

1. Set  $r = \left( L^t (\sigma^{-2} b + \sigma^{-1} \omega_m) + \Sigma_s^{-1/2} \omega_n \right)$ ;
2. Set  $s_1 = \Sigma_s r$ ;
3. Set  $t = L s_1$ ;
4. Set  $\tilde{\Sigma}_b = \left( L \Sigma_s^{-1/2} \right) \left( L \Sigma_s^{-1/2} \right)^t + \sigma^2 \text{Id}_m$ ;
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  $\Sigma_s$  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, 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  $\gamma_i$  (cf. (16)), optimization and sampling can be performed componentwise. The hyperparameter dependent single component part of the posterior reads (cf. (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 a maximum of this expression shows that the update rule is given by:

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

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  $\tilde{\beta} = \frac{1}{2}\|s_{i*}\|^2 + \beta$  and  $\tilde{\alpha} = (\alpha + 3/2)$  (cf. (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.4.

## 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 whose amplitude is above  $q$  times the maximal source amplitude  $\max\|s_{i*}\|_2$ . We will call this measure  $\hat{f}(s, q)$ . However,  $\hat{f}(s, q)$  is not continuous, and involves some arbitrariness, since  $q$  has to be chosen ad hoc. In Figure B.7 three plots of  $\hat{f}(s, q)$  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:

Table A.4: Parameters used in the simulation studies.

Method	Parameter	Study1	Study 2
CM	$Q$	1 000	1 000
	$M$	50 000	200 000
MAP1	$T$	50	50
	$Q$	1 000	1 000
MAP2	$M$	50 000	200 000
	$T$	50	50
MAP3	$U$	128	256
	$Q$	25	25
	$M$	200	200
	$T$	50	50

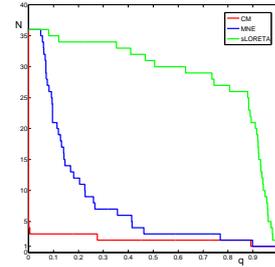


Figure B.7: The curves of  $N = \hat{f}(s, q)$  for  $s = \hat{s}_{\text{CM}}$  (red),  $s = \hat{s}_{\text{MNE}}$  (blue) and  $s = \varphi_{\text{SLORETA}}$  (green) for a simplified model.

**Definition 1** (Spatial dispersion, SD).

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

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

*Earth Mover's Distance (EMD).* Supplementary to the text, we give the mathematical definition of the EMD and some comments on their practical computation. The EMD is a Wasserstein metric, which are distance measures between probability distributions (Ambrosio et al., 2008):

**Definition 2** (Wasserstein metric). *Let  $\mu$  and  $\nu$  be two probability measures on a Radon space  $(\Omega, d)$  that have a finite  $p^{\text{th}}$  moment for some  $p \geq 1$ . Then the  $p^{\text{th}}$  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  $\mu$  and  $\nu$ .

In our study, we are looking at 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 it in 1781 as an optimal transport problem: The idea is to think of the first probability measure as an amount of sand piled on a space  $\Omega$ , and of the second 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 found (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 looks like a rather abstract concept for the practical task we are aiming at, but the lack of a more simple 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 with the reference source activity. To compute the EMD between reference and estimated source activity, both are transferred into discrete probability distributions: In our setting, the reference source activity  $j_{ref}$  was composed of single current dipoles at locations  $\bar{r}_i, i = 1 \dots, \tau$ :

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

Now define a discrete *signature*  $P$  by:

$$P = \{(p_1, w_{p_1}), \dots, (p_\tau, w_{p_\tau})\}$$

with  $p_i := \bar{r}_i; \quad w_{p_i} := \frac{|M_i|}{M_{tot}}; \quad M_{tot} = \sum_{i=1}^{\tau} |M_i|$

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

$$Q = \{(q_1, w_{q_1}), \dots, (q_l, w_{q_l})\}$$

with  $q_i := r_i; \quad w_{q_i} := \frac{\|s_{i*}\|_2}{a_{tot}}; \quad a_{tot} = \sum_{i=1}^k \|s_{i*}\|_2$

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*

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

subject to the following constraints:

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

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

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

The minimal work resulting from this computation is the EMD between  $P$  and  $Q$ . The constraints (B.2) - (B.4) ensure that  $\Gamma$  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 position.

(B.4) determines the amount of mass that has to be transferred into one position.

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

### Appendix C. Acknowledgements

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 measurement of the MRI and A. Janssen, S. Rampersad (both 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 D. Additional Figures

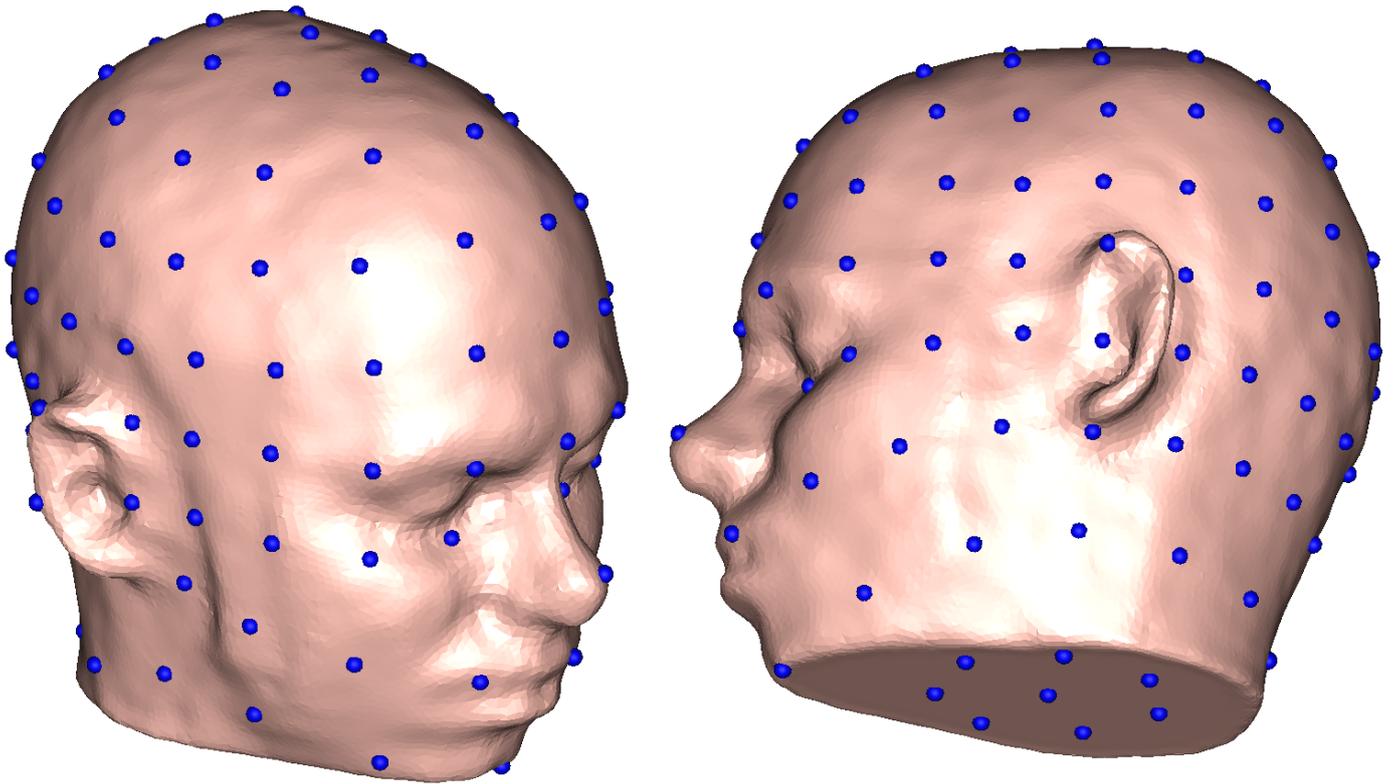


Figure D.8: Artificial full coverage EEG sensor configuration consisting of 134 EEG sensors.

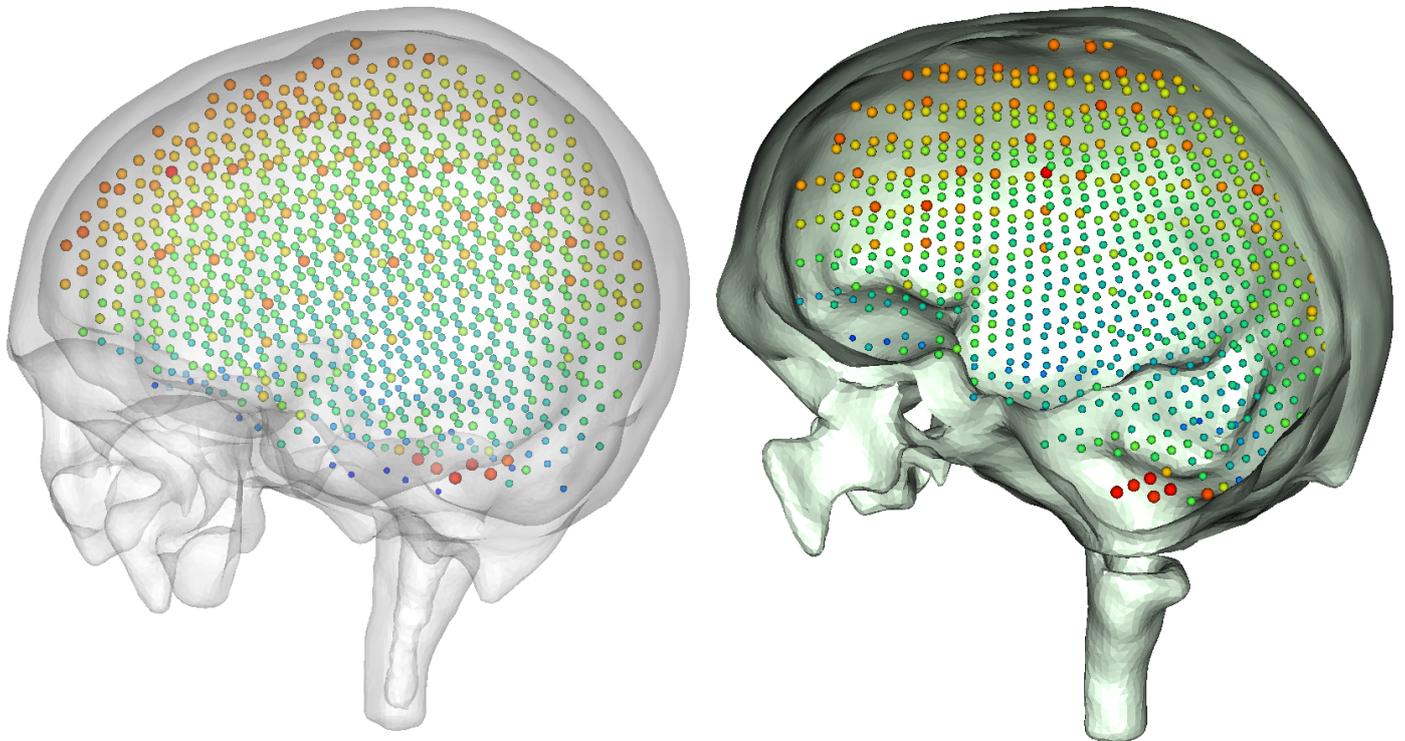


Figure D.9: The sum of the  $\ell_2$  norms of the three gain-vectors at a given position is depicted. The influence of the hole at the base of the skull (foramen magnum) on the magnitudes of the deep-lying sources is noticeable (we checked that this feature is not due to our artificial sensor configurations but occurs with realistic ones as well)

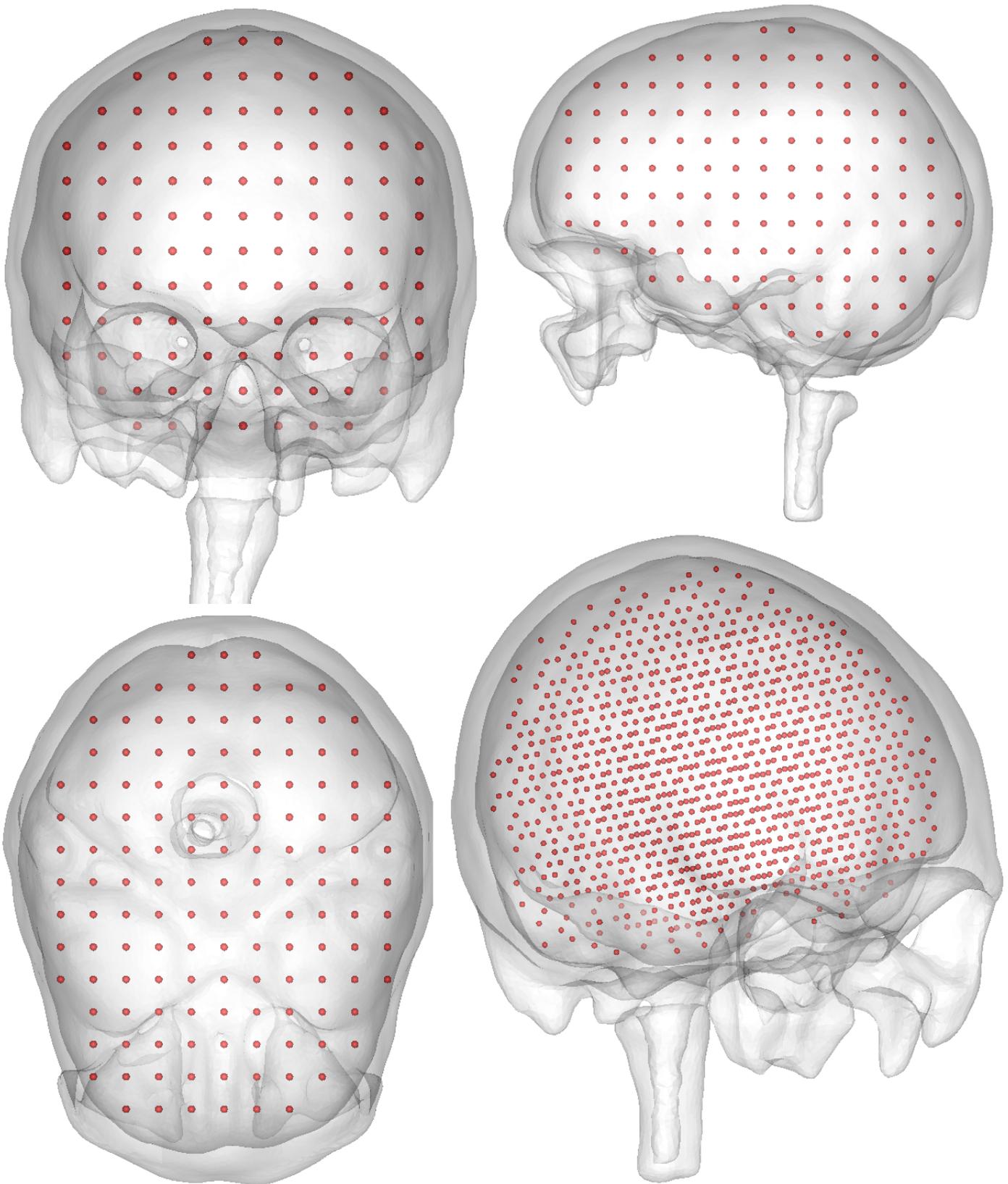


Figure D.10: The locations of the 1000 source space nodes used in the studies

## References

- Acar, Z.A., Makeig, S., 2010. Neuroelectromagnetic Forward Head Modeling Toolbox. *Journal of Neuroscience Methods* 190, 258–270.
- Ahlfors, S.P., Ilmoniemi, R.J., Hämmäläinen, M., 1992. Estimates of visually evoked cortical currents. *Electroencephalogr Clin Neurophysiol* 82, 225–36.
- Akhtari, M., Bryant, H., Mamelak, A., Flynn, E., Heller, L., Shih, J., Mandelkern, M., Matlachov, A., Ranken, D.M., Best, E., et al., 2002. Conductivities of three-layer live human skull. *Brain topography* 14, 151–167.
- Ambrosio, L., Gigli, N., Savaré, G., 2008. *Gradient Flows in Metric Spaces and in the Spaces of Probability Measures*. Birkhauser.
- Andersen, P., 2007. *The Hippocampus Book*. Oxford University Press, USA.
- Aubert, G., Kornprobst, P., 2006. *Mathematical Problems in Image Processing*. volume 147 of *Applied Mathematical Sciences*. Springer. 2nd edition.
- 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*, Report of the Dep. of Tech.Physics, Helsinki University of Technology. pp. 154–171.
- Boon, P., D’Have, M., Vanrumste, B., Van Hoey, G., Vonck, K., Van Wallegghem, P., Caemart, J., Achten, E., De Reuck, J., 2002. Ictal source localization in presurgical patients with refractory epilepsy. *J. Clin. Neurophysiol.* 19, 461–468.
- Bruno, P., Vatta, F., Minimel, S., Inchingolo, P., 2004. Referenced EEG and Head Volume Conductor Model: Geometry and Parametrical Setting., in: *Proc. of the 26th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, San Francisco, USA, Sep. 1-5*, <http://www.ucsfresno.edu/embs2004>.
- Bruno, P., Vatta, F., Minimel, S., Inchingolo, P., 2003. Head model extension for the study of bioelectric phenomena. *Biomedical Sciences Instrumentation* 39, 59–64.
- 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 102, 267–278.
- Calvetti, D., Hakula, H., Pursiainen, S., Somersalo, E., 2009. Conditionally Gaussian hypermodels for cerebral source localization. *SIAM J. Imaging Sci.* 2, 879–909.
- Calvetti, D., Somersalo, E., 2007a. A Gaussian hypermodel to recover blocky objects. *Inverse Problems* 23, 733–754.
- Calvetti, D., Somersalo, E., 2007b. Introduction to Bayesian Scientific Computing, volume 2 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer New York.
- Calvetti, D., Somersalo, E., 2008a. Hypermodels in the Bayesian imaging framework. *Inverse Problems* 24, 034013 (20pp).
- Calvetti, D., Somersalo, E., 2008b. Recovery of shapes: Hypermodels and Bayesian learning, in: *Journal of Physics: Conference Series*, IOP Publishing. p. 012014.
- Chang, B.S., Lowenstein, D.H., 2003. Epilepsy. *N Engl J Med* 349, 1257–66.
- 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, in: *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.H., Knösche, T.R., 2010. Modeling of the human skull in EEG source analysis. *Human Brain Mapping* 32, 1383–1399. DOI: 10.1002/hbm.21114, PMID: 20690140.
- 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, 960–6.
- 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.
- Duvernoy, H., 2005. *The Human Hippocampus: Functional Anatomy, Vascularization, and Serial Sections with MRI*. Springer Verlag.
- Engl, H., Hanke-Bourgeois, M., Neubauer, A., 1996. Regularization of Inverse Problems. *Mathematics and its applications*, Kluwer Acad. Publ.
- 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, 1104–20.
- Friston, K.J., Mattout, J., Trujillo-Barreto, N.J., Ashburner, J., Penny, W.D., 2007. Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–34.
- Fuchs, M., Drenckhahn, R., Wischmann, H., Wagner, M., 1998. An improved boundary element method for realistical volume conductor modeling 45, 980–997.
- Fuchs, M., Wagner, M., Köhler, T., Wischmann, H.A., 1999. Linear and non-linear current density reconstructions. *Journal of clinical Neurophysiology* 16, 267.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. *Bayesian Data Analysis. Texts in Statistical Science*, Chapman and Hall/CRC. 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, 827–38.
- Grave de Peralta, R., Hauk, O., Gonzalez, S.L., 2009. The neuroelectromagnetic inverse problem and the zero dipole localization error. *Comput Intell Neurosci* , 659247.
- Greenblatt, R.E., Ossadtschi, A., Pflieger, M.E., 2005. Local Linear Estimators for the Bioelectromagnetic Inverse Problem. *IEEE Transactions on Signal Processing* 53, 3403–3412.
- Hallez, H., 2008. Incorporation of Anisotropic Conductivities in EEG Source Analysis. Ph.D. thesis. Faculteit Ingenieurswetenschappen, Universiteit Gent, Belgium.
- Hämäläinen, M., Haario, H., Lehtinen, M., 1987. Inference about sources of neuromagnetic fields using Bayesian parameter estimation. Preprint, 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.

- 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.
- Hauelsen, 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., Flandin, G., Friston, K.J., Mattout, J., 2010. A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. *Hum Brain Mapp* .
- Henson, R.N., Mattout, J., Phillips, C., Friston, K.J., 2009a. Selecting forward models for MEG source-reconstruction using model-evidence. *Neuroimage* 46, 168–76.
- Henson, R.N., Mouchlianitis, E., Friston, K.J., 2009b. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage* 47, 581–9.
- Ioannides, A.A., Bolton, J.P.R., Clarke, C.J.S., 1990. Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Problems* 6, 523.
- 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, 1581–94.
- Kaipio, J.P., Somersalo, E., 2005. Statistical and Computational Inverse Problems. volume 160 of *Applied Mathematical Sciences*. Springer New York.
- Kantorovich, L., 1942. On the translocation of masses, CR (Doklady) Acad. Sci. URSS (NS) 37, 199–201.
- 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, Publication of the Akademii Nauk SSSR, Moscow-Leningrad , 110–138.
- Klenke, A., 2008. Probability Theory: A Comprehensive Course. Springer Verlag.
- Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., Papadopoulou, 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., in: Proc. of the 16th Annual Meeting of the Organization for Human Brain Mapping, Barcelona, Spain, June 6-10. <http://www.humanbrainmapping.org/barcelona2010/>.
- 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, 160–71.
- Lucka, F., 2011. Hierarchical Bayesian Approaches to the Inverse Problem of EEG/MEG Current Density Reconstruction. Master's thesis. University of Muenster, Germany. <http://www.math.uni-muenster.de/num/burger/teaching/diplomanden.html>.
- Lütkenhöner, B., Lammertmann, C., Ross, B., Pantev, C., 2000. Brain stem auditory evoked fields in response to clicks. *Neuroreport* 11, 913.
- MacKay, D., 2003. Information Theory, Inference, and Learning Algorithms. Cambridge Univ Pr.
- 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, 753–67.
- 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, 1069–77.
- 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, 541–57.
- Niedermeyer, E., da Silva, F., 2004. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincott Williams & Wilkins.
- 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, 669–85.
- 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, 876–89.
- Nunez, P.L., Srinivasan, R., 2005. Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, USA. 2nd edition.
- 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 Map* 17, 179–192.
- Pantev, C., Hoke, M., Lutkenhoner, B., Lehnertz, K., 1989. Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science* 246, 486–8.
- Parkkonen, L., Fujiki, N., Mäkelä, J., 2009. Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Human brain mapping* 30, 1772–1782.
- Pascual-Marqui, R.D., 1999. Review of methods for solving the EEG inverse problem. *International Journal of Bioelectromagnetism* 1, 75–86.
- Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol* 24 Suppl D, 5–12.
- 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., Hauelsen, J., 2006. Influence of head models on EEG simulations and inverse source localizations. *Biomed Eng Online* 5, 10.
- Rampp, S., Stefan, H., 2007. Magnetoencephalography in presurgical epilepsy diagnosis. *Expert Rev.Med.Devices* 4, 335–347.
- 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. *Journal of Clinical Neurophysiology* 25, 331.
- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology* 32, 11.
- Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage* 23, 806–26.
- Scherg, M., Cramon, D.V., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 62, 32–44.
- Schimpf, P., Ramon, C., Hauelsen, J., 2002. Dipole Models for the EEG and MEG 49, 409–418.
- Sekihara, K., Nagarajan, S.S., 2008. Adaptive Spatial Filters for Electromagnetic Brain Imaging (Series in Biomedical Engineering). Springer. 1st edition.
- 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, 1056–67.
- Stefan, H., Hildebrandt, M., Kerling, F., Kasper, B.S., Hammen, T., Dörfler, A., Weigel, D., Buchfelder, M., Blümcke, I., Pauli, E., 2009. Clinical prediction of postoperative seizure control: structural, functional findings and disease histories. *J Neurol Neurosurg Psychiatry* 80, 196–200.
- 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.
- Taubin, G., 1995. A signal processing approach to fair surface design, in: 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.
- Vallaghé, S., Papadopoulou, 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. *Electroencephalography and clinical neurophysiology* 106, 522–534.
- Vese, L., Chan, T., 2002. "A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model.". *International Journal of Computer Vision* 50, 271–293.
- Vorwerk, J., 2011. Comparison of Numerical Approaches to the EEG Forward Problem. Master's thesis. University of Muenster, Germany.

- [Http://wwwmath.uni-muenster.de/num/burger/teaching/diplomanden.html](http://wwwmath.uni-muenster.de/num/burger/teaching/diplomanden.html).
- Wagner, M., Fuchs, M., Kastner, J., 2004. Evaluation of sLORETA in the presence of noise and multiple sources. *Brain Topogr* 16, 277–80.
- 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, 665–75.
- Weinstein, D., Zhukov, L., Johnson, C., 2000. Lead-field bases for Electroencephalography source imaging. *Annals of Biomed.Eng.* 28, 1059–1066.
- Wipf, D., Nagarajan, S.S., 2009. A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage* 44, 947–66.
- 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 Problems* 20, 1099.
- Wolters, C.H., Köstler, H., Möller, C., Härtlein, J., Grasedyck, L., Hackbusch, W., 2007. Numerical mathematics of the subtraction method for the modeling of a current dipole in EEG source reconstruction using finite element head models. *SIAM J. on Scientific Computing* .