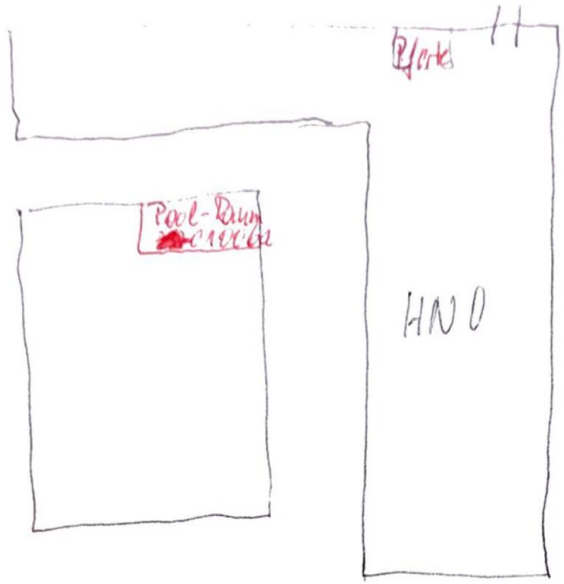
EEG and MEG inverse problem, Part II: Dipole scan, Current Density Reconstruction (CDR) and Hierarchical Bayesian Modeling (HBM)

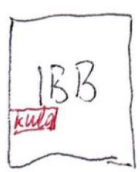
Institute for Biomagnetism and Biosignalanalysis

Carsten Wolters

Institut für Biomagnetismus und Biosignalanalyse, Westfälische Wilhelms-Universität Münster

Lectures, May 13, 2025





Schlüssel für Pool-Raum und CURRY9-Dongle

(1) Ausleihen bei Jasmin Kula im IBB

(2) Rückgabe " " " " "

(3) Falls (2) nicht geht, bitte im Briefumschlag mit Adresse "An Jasmin Kula, IBB" an der HNO-Pforte abgeben.

Carsten.wolters@uni-münster.de

Dipole scan

Dipole scans in different inner product spaces

Definition 1 (Hilbert norm). Let C be a positive definite operator on some Hilbert space. Then we can define an inner product via

$$\langle x, y \rangle_C = \langle C \cdot x, y \rangle$$

and a corresponding norm via

$$\|x\|_C = \sqrt{\langle C \cdot x, x \rangle}.$$

We call a norm arising in this way a *Hilbert norm.*

Note that the choice $C = \mathrm{Id}$ reproduces the original inner product and norm.

In the following, let $x_1, ..., x_M \in \mathbb{R}^3$ denote a set of source positions with corresponding leadfields $L_1, ..., L_M \in \mathbb{R}^{N \times 3}$. Furthermore, let $d \in \mathbb{R}^N$ denote a measurement vector.

Definition 2 (Generalized dipole scan). Let C be a positive definite operator defining a Hilbert norm $\| \cdot \|_C$. For each $1 \leq i \leq M$ we then estimate a dipole moment j_i via

$$j_i = \arg\min_{j \in \mathbb{R}^3} \|d - L_i \cdot j\|_C^2.$$

We then define the *goodness of fit (GOF)* of the source at position x_i to be

$$\mathrm{GOF}(i, C) = 1 - \frac{\|d - L_i \cdot j_i\|_C^2}{\|d\|_C^2}.$$

The *dipole scan* is then given by searching for the index i which maximizes the goodness of fit.

The standard GOF-scan now corresponds to the choice $C = \mathrm{Id}$ on $\mathbb{R}^N$ equipped with the standard inner product.

Standardized LOw Resolution Electromagnetic TomogrAphie (sLORETA)

[Pascual-Marqui, *Meth & Find. In Exp. & Clin. Pharmacol.*, 2002]
[Höltershinken, Erdbrügger & Wolters, arXiv:2409.07459, *2024*]

sLORETA: Dipole scan in a different inner product space

Definition 3 (generalized sLORETA, vector case). Let $C \in \mathbb{R}^{N \times N}$ be a symmetric matrix defining a positive definite operator on a space containing the range of the leadfields. Then define for $1 \leq i \leq M$ the vector

$$j_{i,C}^{\text{sLORETA}} = (L_i^\intercal C L_i)^{-\frac{1}{2}} L_i^\intercal C d \in \mathbb{R}^3.$$

As defined in [1], the generalized sLORETA method now consists of searching for large values of $\|j_{i,C}^{\text{sLORETA}}\|^2$.

The classical sLORETA approach now corresponds to the choice $C = (LL^\intercal + \alpha H)^+$, where $L = (L_1, ..., L_M)$ is the complete leadfield, H is the orthogonal projection onto $\{1\}^\perp$, and $\alpha \geq 0$ (as defined by Pascual-Marqui)

Theorem 1 (Equivalence of sLORETA and dipole scanning). Let C be a symmetric matrix defining a positive definite operator on a space containing the range of the leadfields. Let $\|\cdot\|_C$ denote the corresponding Hilbert norm. Then we have

$$\|j_{i,C}^{\text{sLORETA}}\|^2 = \|d\|_C^2 \cdot \mathrm{GOF}(i, C).$$

Hence we see that the sLORETA estimate at a source position is, up to a scalar multiple, given by the goodness of fit of the source position to the data. In particular we see that visualizing the sLORETA estimate is the same as visualizing a goodness of fit distribution.

[Lucka, *PhD thesis in Mathematics*, 2015]
[Lucka, Burger, Pursiainen & Wolters, *NeuroImage*, 2012]
[Lucka, Burger, Pursiainen & Wolters, *Biomag2012, 2012*]
[Lucka, *Diploma thesis in Mathematics*, March 2011]

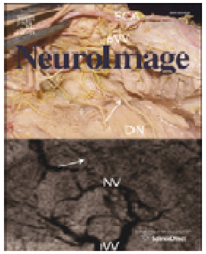
NeuroImage 61 (2012) 1364–1382

ELSEVIER

Contents lists available at SciVerse ScienceDirect

NeuroImage

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



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

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

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

[Lucka, *Diploma thesis in Mathematics*, March 2011:
http://www.sci.utah.edu/~wolters/PaperWolters/2011/LuckaDiplom.pdf]

Westfälische Wilhelms-Universität Münster

Diplomarbeit in Mathematik

Hierarchical Bayesian Approaches to the Inverse Problem of EEG/MEG Current Density Reconstruction

eingereicht von

Felix Lucka

Münster, 10. März, 2011

Fachbereich 10 Mathematik und Informatik

IBB Institute for Biomagnetism and Biosignalanalysis

Gutachter:

Prof. Dr. Martin Burger

Institut für Numerische und Angewandte Mathematik

Priv.-Doz. Dr. Carsten Wolters

Institut für Biomagnetismus und Biosignalanalyse

[Lucka, *PhD thesis in Mathematics*, 2015
http://www.sci.utah.edu/~wolters/PaperWolters/2015/LuckaDissertation.pdf]

WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Fach: Mathematik

Bayesian Inversion in Biomedical Imaging

Inaugural Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

– Dr. rer. nat. –

im Fachbereich Mathematik und Informatik

der Mathematisch-Naturwissenschaftlichen Fakultät

der Westfälischen Wilhelms-Universität Münster

eingereicht von

Felix Lucka

aus Hannover

– 2014 –

Current density reconstruction methods (CDR)

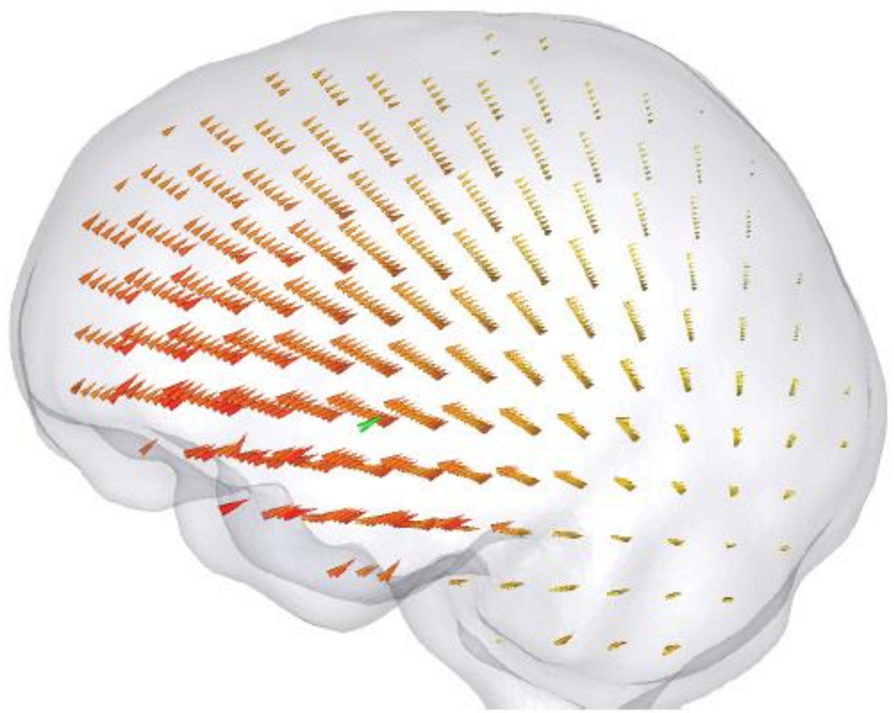


Figure A.18.: MNE result (red-yellow cones) for a single dipole (green cone).

ABSTRACT

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

Bayesian formulation of the static inverse problem

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

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

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

Hierarchical Bayesian Modeling (HBM): Mathematics: The likelihood model

- Central to Bayesian approach: Accounting for each uncertainty concerning the value of a variable explicitly: The variable is modeled as a random variable
- In this study, we model the additive measurement noise by a Gaussian random variable

$$\mathcal{E} \sim \mathcal{N}(0, \Sigma_\varepsilon) \qquad \Sigma_\varepsilon = \sigma^2 I_m$$

- For EEG/MEG, this leads to the following likelihood model:

$$B = LS + \mathcal{E}$$

Carsten.wolters@uni-münster.de

[Lucka, Burger, Pursiainen & Wolters, *NeuroImage*, in revision]
[Lucka, Burger, Pursiainen & Wolters, *Biomed.Eng., 2011*]
[Lucka, *Diploma thesis in Mathematics*, March 2011]

Hierarchical Bayesian Modeling (HBM): Mathematics: The likelihood model

• The conditional probability density of B given S is called likelihood density, in our (Gaussian) case, it is thus:

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

Hierarchical Bayesian Modeling (HBM): Mathematics: Prior and Bayes rule

- Due to the ill-posedness, inference about S given B is not feasible like that, we need to encode a-priori information about S in its density $p_{pr}(s)$, which is called prior
- We call the conditional density of S given B the posterior: $p_{post}(s|b)$
- Then, the model can be inverted via Bayes rule:

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

- The term p(b) is called model evidence, (see **Sato et al., 2004; Trujillo-Barreto et al., 2004; Henson et al., 2009, 2010**). Here, it is just a normalizing constant and not important for the inference presented now

Hierarchical Bayesian Modeling (HBM): Mathematics: MAP and CM

• 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 Maximum A-Posteriori (MAP, the highest mode of the posterior) and the Conditional Mean (CM, the expected value of the posterior):

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

$$\hat{s}_{\mathrm{CM}} := \mathbb{E}\left[s|b\right] = \int s \; p_{post}(s|b) \, ds$$

• Practically, the MAP is a high-dimensional optimization problem and the CM is a high-dimensional integration problem

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Mathematics: Specific priors used in EEG/MEG

- To revisit some commonly known inverse methods, we consider Gibbs distribution as prior:

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

- Here, P(s) is an energy functional penalizing unwanted features of s
- The MAP-estimate is then given by:

$$\begin{aligned}\hat{s}_{\mathrm{MAP}} &:= \underset{s\in\mathbb{R}^n}{\operatorname{argmax}} \left\{\exp\left(-\frac{1}{2\,\sigma^2}\|b - L\,s\|_2^2 - \frac{\lambda}{2\sigma^2}\mathcal{P}(s)\right)\right\} \\ &= \underset{s\in\mathbb{R}^n}{\operatorname{argmin}} \left\{\|b - L\,s\|_2^2 + \lambda\,\mathcal{P}(s)\right\}\end{aligned}$$

Hierarchical Bayesian Modeling (HBM): Mathematics: Some choices for P(s) used in EEG/MEG

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

$$\mathcal{P}(s) = \|s\|_2^2$$

- Minimum Norm Estimation (MNE), see **Hämäläinen and Ilmoniemi, 1984**

Hierarchical Bayesian Modeling (HBM): Mathematics: Some choices for P(s) used in EEG/MEG

$$\hat{s}_{\mathrm{MAP}} := \operatorname*{argmax}_{s \in \mathbb{R}^n} \left\{ \exp\left(-\frac{1}{2\,\sigma^2} \|b - L\,s\|_2^2 + \frac{\lambda}{2\sigma^2} \mathcal{P}(s) \right) \right\}$$

$$= \operatorname*{argmin}_{s \in \mathbb{R}^n} \left\{ \|b - L\,s\|_2^2 + \lambda\,\mathcal{P}(s) \right\}$$

$$\mathcal{P}(s) = \|\Sigma_s^{-1/2} s\|_2^2$$

- Weighted Minimum Norm Estimation (WMNE), see **Dale and Sereno, 1993**

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

- Specific choices for WMNE: **Fuchs et al., 1999**

$$\Sigma_s^{\ell_\infty, reg} = \operatorname*{diag}_{i=1,\ldots,n} \left(\frac{\chi_i^2}{(\chi_i^2 + \beta^2)^2} \right)$$

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

Hierarchical Bayesian Modeling (HBM): Mathematics:

- Brain activity is a complex process comprising many different spatial patterns
- No fixed prior can model all of these phenomena without becoming uninformative, that is, 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

Hierarchical Bayesian Modeling (HBM): Mathematics:

- The idea of Hierarchical Bayesian Modeling (HBM) is to let the same data determine the appropriate model used for the inversion of these data by extending the model by a new level of inference: The prior on S is not fixed but random, determined by values of additional parameters called hyperparameters
- The hyperparameters γ follow an a-priori assumed distribution (the so-called hyperprior $p_{hpr}(\gamma)$) and are subject to estimation schemes, too.
- As this construction follows a top-down scheme, it is called hierarchical modeling:

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

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

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

Hierarchical Bayesian Modeling (HBM): Mathematics for EEG/MEG application

• The hierarchical model used in most methods for EEG/MEG relies on a special construction of the prior called Gaussian scale mixture or conditionally Gaussian hypermodel (Calvetti et al., 2009; Wipf and Nagarajan, 2009)

• $p_{pr}(s|\gamma)$ is a Gaussian density with zero mean and a covariance determined by γ :

$$S|\gamma \sim N(0, \Sigma_s(\gamma))$$

$$\Sigma_s(\gamma) = \sum_{i=1}^{h} \gamma_i C_i \qquad \text{where} \qquad C_i \in C$$

$$\Rightarrow \; p_{pr}(s|\gamma) = (2\pi)^{-n/2} |\Sigma_s|^{-1/2} \exp\left(-\frac{1}{2}\left(s\,\Sigma_s^{-1}\,s^t\right)\right)$$

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Mathematics for EEG/MEG application

The first important choice is in choosing an appropriate set $\mathcal{C}$. A variety of approaches that encode different a priori information on the spatial source covariance pattern have been proposed, e.g., spatial smoothness components (Mattout et al., 2006; Phillips et al., 2005) or *multiple sparse priors* (Friston et al., 2008). A recent overview is given in Lucka (2011), page 19. In this study, we will rely on single-location priors (I_d denotes the identity matrix in d dimensions):

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

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

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Mathematics for EEG/MEG application

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

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

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

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

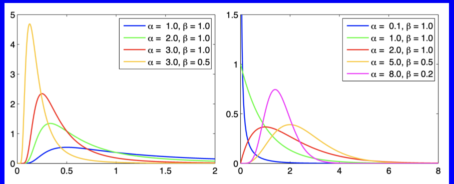


Figure A.1: Plots of the pdfs of inverse gamma (left) and gamma distribution (right).

Carsten.wolters@uni-münster.de

[Lucka, *Diploma thesis in Mathematics*, March 2011]

Hierarchical Bayesian Modeling (HBM): Mathematics for EEG/MEG application

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

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

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

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

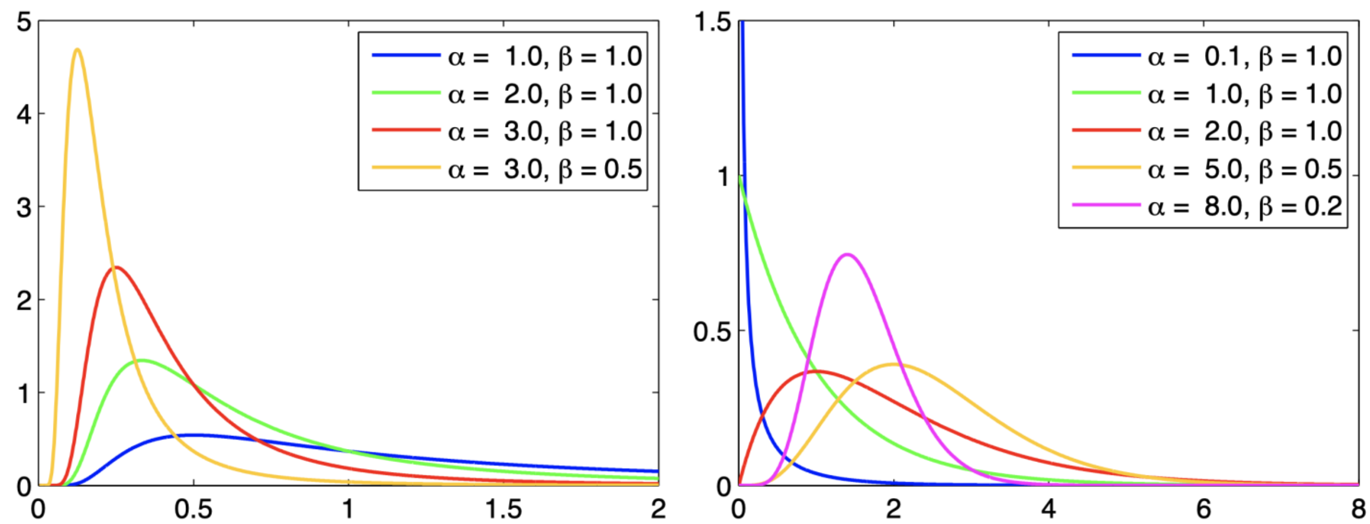


Figure A.1: Plots of the pdfs of inverse gamma (left) and gamma distribution (right).

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Mathematics: Our chosen posterior

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

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

Hierarchical Bayesian Modeling (HBM): Mathematics:

Inference for hierarchical models

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

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

Hierarchical Bayesian Modeling (HBM): Mathematics:

- CM estimation: Blocked Gibbs sampling, a Markov chain Monte Carlo (MCMC) scheme (**Nummenmaa et al., 2007; Calvetti et al., 2009**)
- MAP estimation: Iterative alternating sequential (IAS) (**Calvetti et al., 2009**)

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

Hierarchical Bayesian Modeling (HBM): Mathematics:

Table 4.4: EMD of IAS result for different parameters of the inverse gamma hyperprior, averaged over 1 000 single unit-strength dipole sources.

$\beta \downarrow$ / $\alpha \rightarrow$	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**	**1.2**	**1,5**	**2.0**
10^{-4}	38.08	38.31	38.48	38.58	38.73	38.81	39.19	39.80	40.75
$5 \cdot 10^{-5}$	34.17	34.31	34.46	34.69	34.92	35.03	35.28	35.78	36.75
10^{-5}	28.82	28.94	29.04	29.17	29.22	29.35	29.66	30.08	31.21
$5 \cdot 10^{-6}$	28.18	28.28	28.36	28.49	28.66	28.88	29.29	29.98	31.67
10^{-6}	32.90	33.50	33.90	34.44	35.03	35.70	37.34	39.36	41.54

The parameters which we will use later (i.e., $\alpha = 0.5$, $\beta = 5 \cdot 10^{-8}$) yield the best results while exhibiting moderate computation times.

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Mathematics:

For these reasons, we examine three different initialization schemes:

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

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

Hierarchical Bayesian Modeling (HBM): Goal of our study

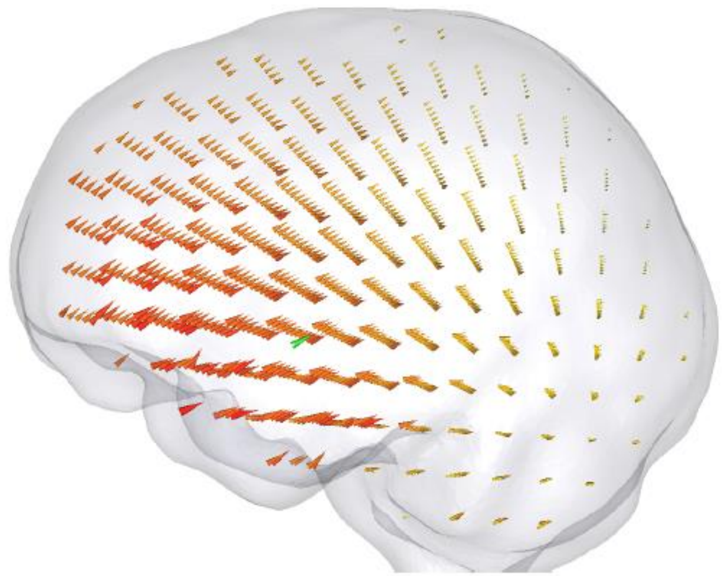


Figure A.18.: MNE result (red-yellow cones) for a single dipole (green cone).

- Step 1: Computed forward EEG for reference source (green dipole)
- Step 2: Computed HBM inverse solution without indicating the number of sources (yellow-orange-red current density distribution on source space)

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Validation means: DLE and SP

Validation means and inverse crimes

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

Hierarchical Bayesian Modeling (HBM): Validation means: EMD

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

Hierarchical Bayesian Modeling (HBM): Validation means: Source depth

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

Hierarchical Bayesian Modeling (HBM): Methods: Head model

Head model and source space

For the numerical approximation of the forward problem, we use the finite element (FE) method because of its flexibility with regard to the realistic modeling of the head volume conductor at a fast computational speed with respect to this degree of modeling accuracy. Although working with a head model that is as realistic as possible is, in general, preferable (see the references in the description below), the specific aims of our studies require some simplifications. We do not want to include the inner brain compartments (the CSF, gray matter and white matter) because we want to focus on the effect of depth bias separately from other sources of error, e.g., from the effects caused by the anisotropy of the white matter (which also makes the results comparable to those obtained using *BEM* models, which cannot capture the anisotropy and normally do not differentiate between the inner brain compartments as well). Additionally, to facilitate the interpretation of the results, we require a homogeneous innermost compartment without holes and enclosures where we can place the test sources.

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling (HBM): Methods: Head model

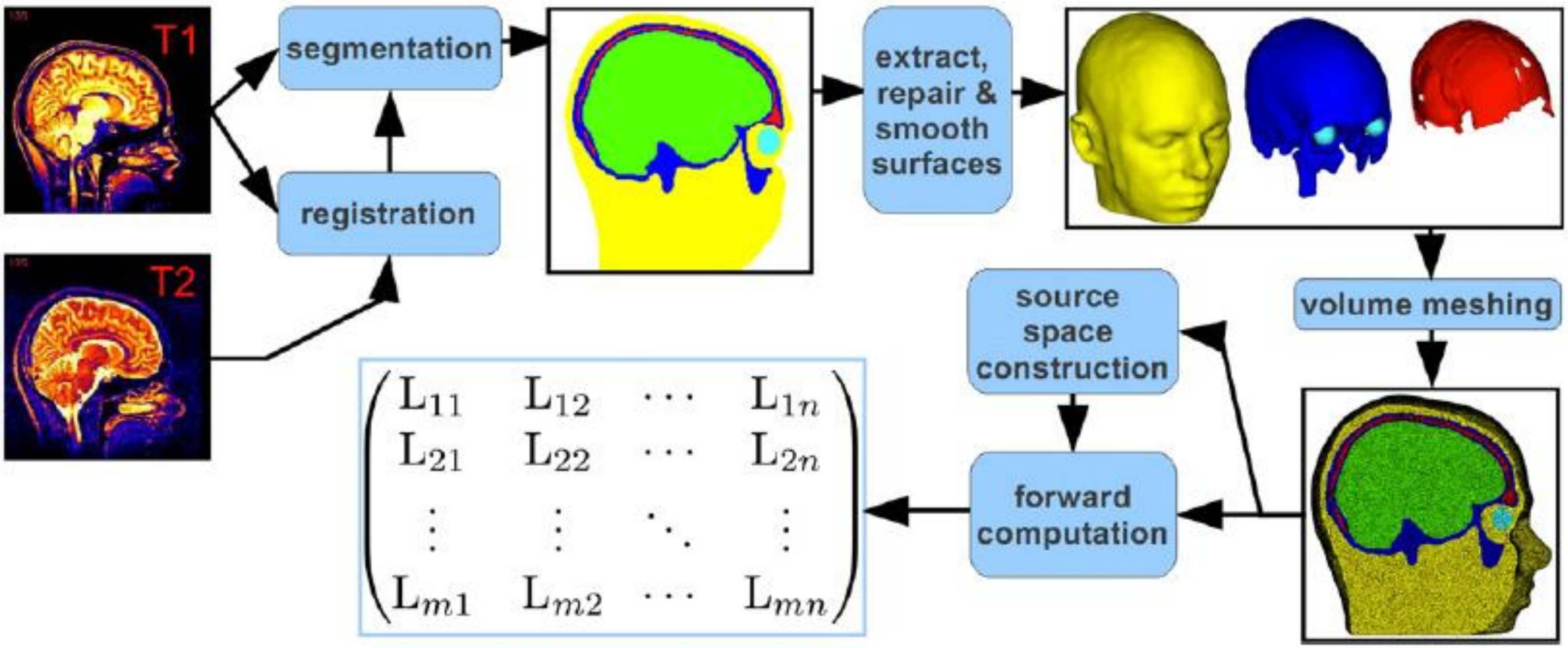
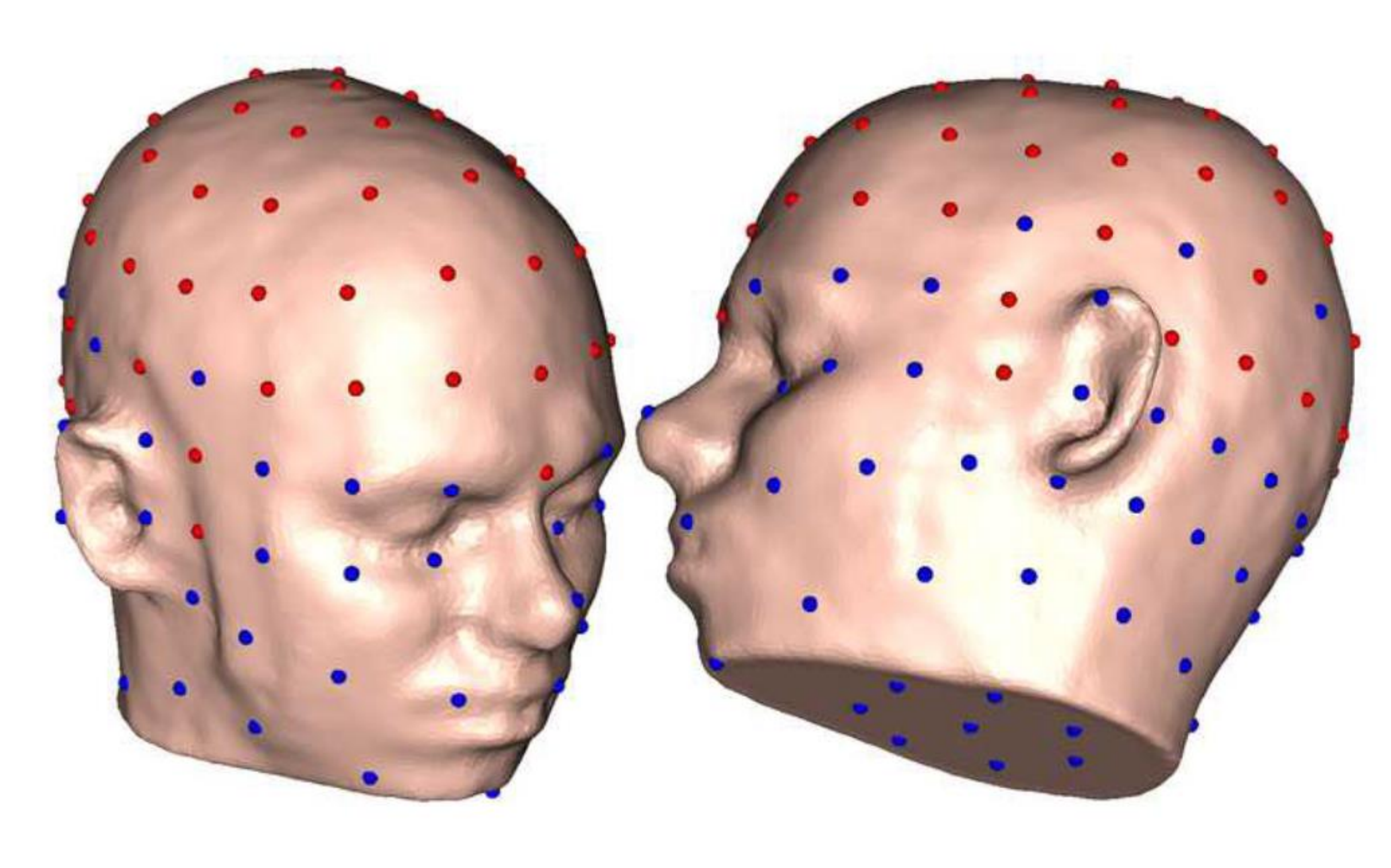


Fig. 1. Model generation pipeline.

Hierarchical Bayesian Modeling (HBM): Methods: EEG sensors

Another important aspect for practical EEG/MEG studies is the effect of *insufficient sensor coverage*. For an optimal scan of the electromagnetic field pattern, the sensors should be uniformly distributed in every spatial direction. However, for practical reasons, this is not possible in realistic settings. The neck causes a semi-shell-like sensor distribution, which is not able to record fields in the direction of the feet. In particular deep-lying sources suffer from this insufficiency. The influence of insufficient sensor coverage should not be mixed with the effects of depth bias in this first, basic study. Therefore, we will use two sensor configurations in our studies. First, an artificial sensor configuration consisting of 134 EEG sensors distributed uniformly over the surface of the head model is created (abbreviated *f-cap* for *full cap*). From these sensor positions, a subset of 63 sensors, which represents a realistic sensor placement, are chosen as a second sensor configuration (abbreviated *r-cap* for *realistic cap*). Fig. 8 shows both configurations.

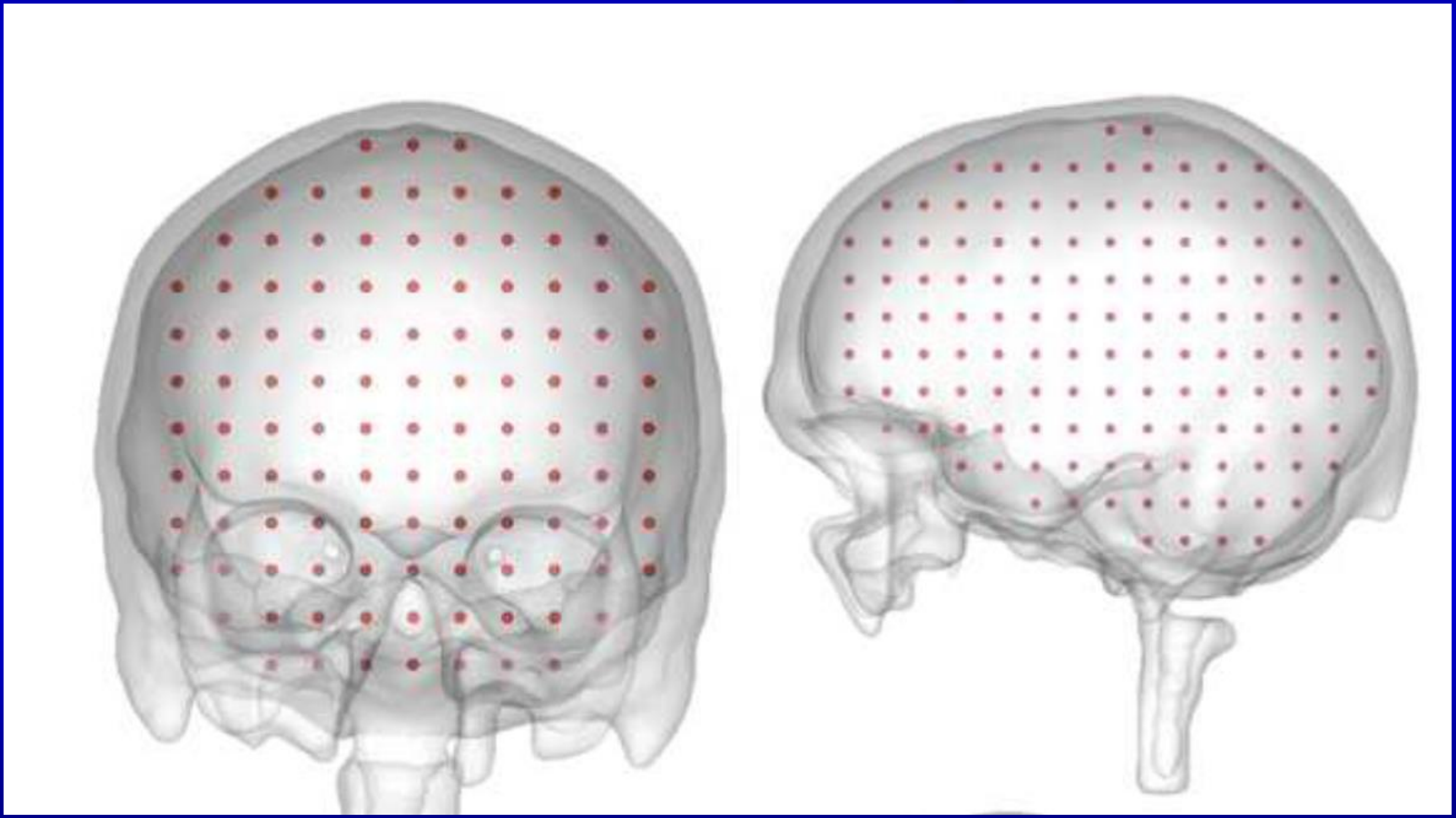
Hierarchical Bayesian Modeling (HBM): Methods: Full-cap (f-cap), realistic cap (r-cap)



Hierarchical Bayesian Modeling (HBM): Methods: Source space and EEG lead field

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

Hierarchical Bayesian Modeling (HBM): Methods: Source space



Hierarchical Bayesian Modeling (HBM): Study 1: Single dipole reconstruction

Study 1: single dipole reconstruction

Setting

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

Hierarchical Bayesian Modeling (HBM): Methods: Generation of noisy measurement data

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

Hierarchical Bayesian Modeling (HBM): Results: Single focal source scenario

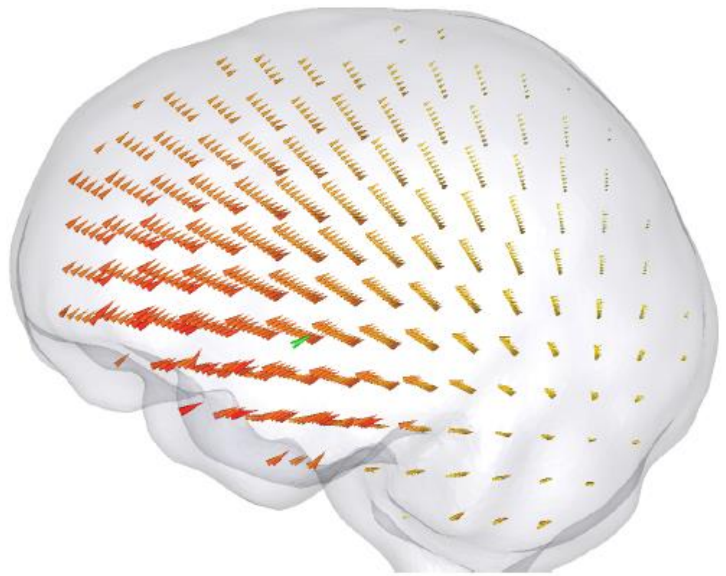


Figure A.18.: MNE result (red-yellow cones) for a single dipole (green cone).

- Step 1: Computed forward EEG for reference source (green dipole), add noise
- Step 2: Computed HBM inverse solution without indicating the number of

sources (yellow-orange-red current density distribution on source space)

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling: Single focal source scenario

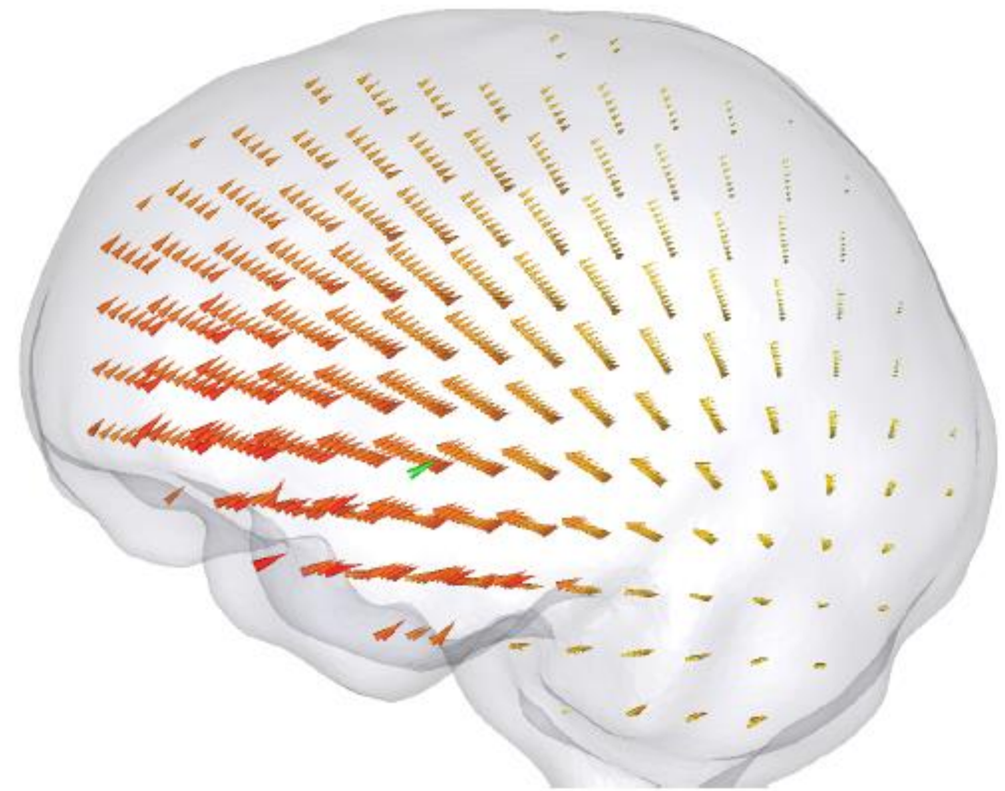


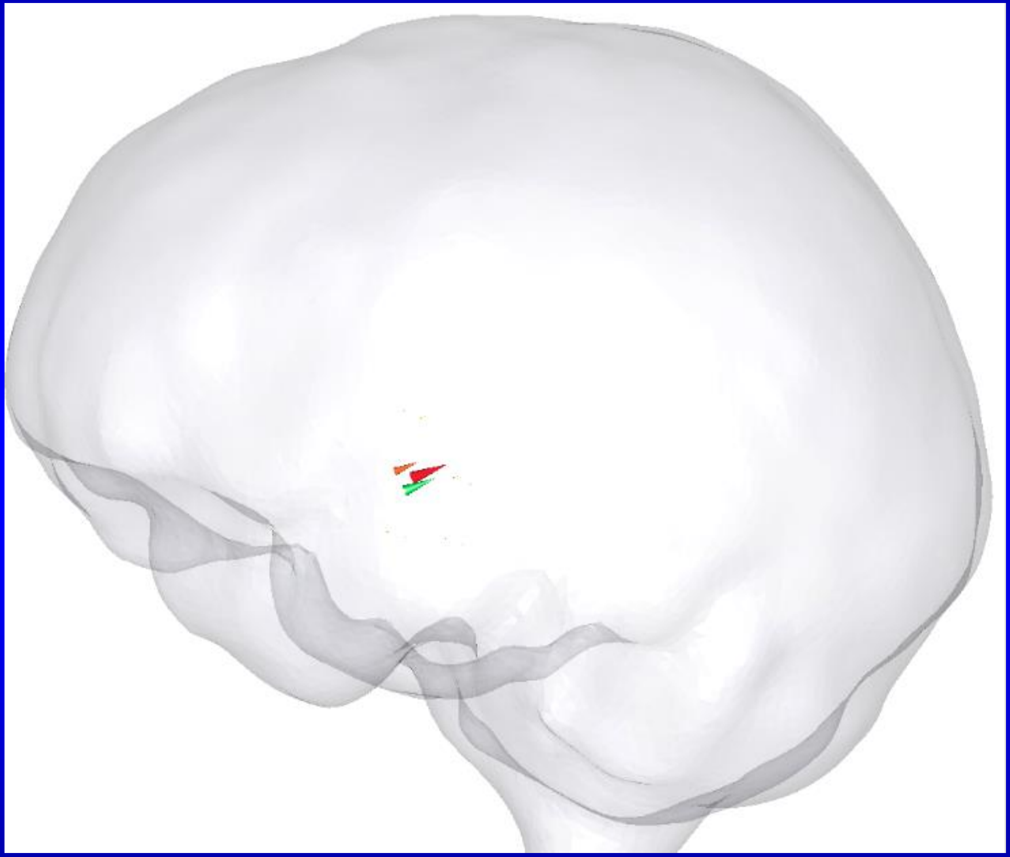
Figure A.18.: MNE result (red-yellow cones) for a single dipole (green cone).

Hierarchical Bayesian Modeling: Single focal source scenario

Figure A.17.: sLORETA result (red-yellow spheres) for a single dipole (green cone).

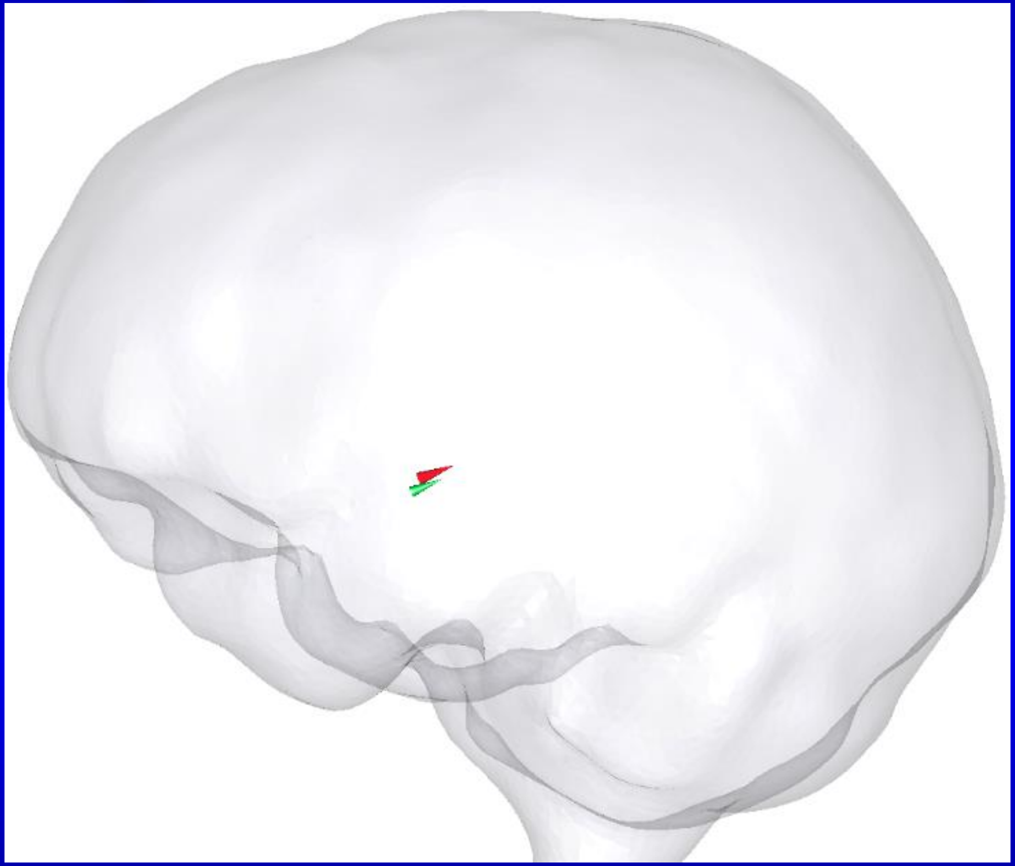
Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling: Single focal source scenario



HBM: Conditional Mean (CM) estimate

Hierarchical Bayesian Modeling: Single focal source scenario



HBM: CM followed by Maximum A-Posteriori estimate (MAP)

Hierarchical Bayesian Modeling: Study 1: Single focal source scenario

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

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

Hierarchical Bayesian Modeling: Study 1: Single focal source scenario

Method	DLE, f-cap	DLE, r-cap	EMD, f-cap	EMD, r-cap
MNE	29.46	33.07	53.20	54.90
WMNE ℓ_2	30.65	35.08	52.17	53.64
WMNE $\ell_{\infty,reg}$	29.40	35.38	49.56	51.08
sLORETA	6.10	6.60	40.58	43.43
CM	6.16	6.94	7.32	8.85
MAP	5.85	6.39	6.08	6.45

Hierarchical Bayesian Modeling : Study 1: Single focal source scenario

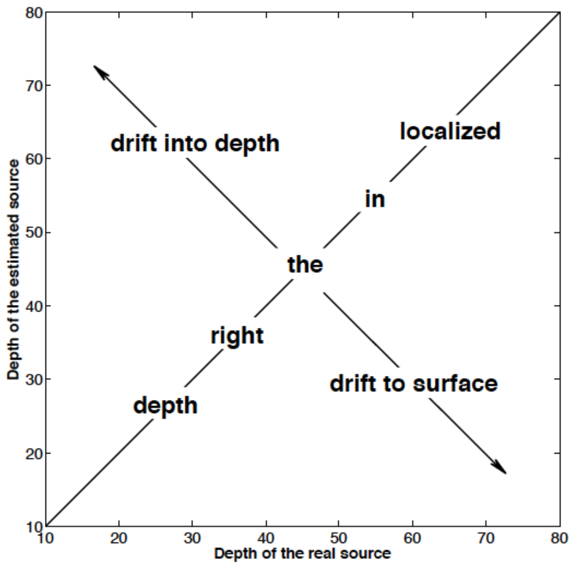


Figure 4.9.: Explanation of the scatter plots

- A mark within the area underneath the y=x line indicates that the dipole has been reconstructed too close to the surface
- A mark above the line indicates the opposite
- q_{ab} denotes the percentage of marks above the y=x line minus 0.5 (optimally: $q_{ab}= 0$)

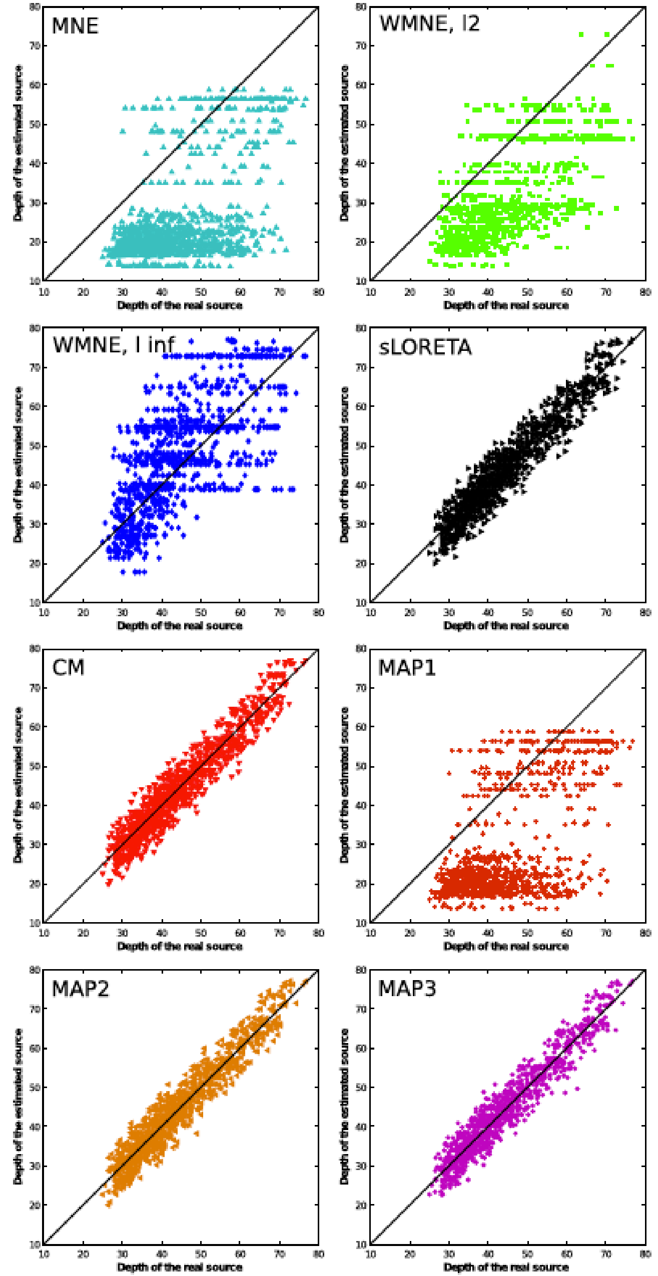
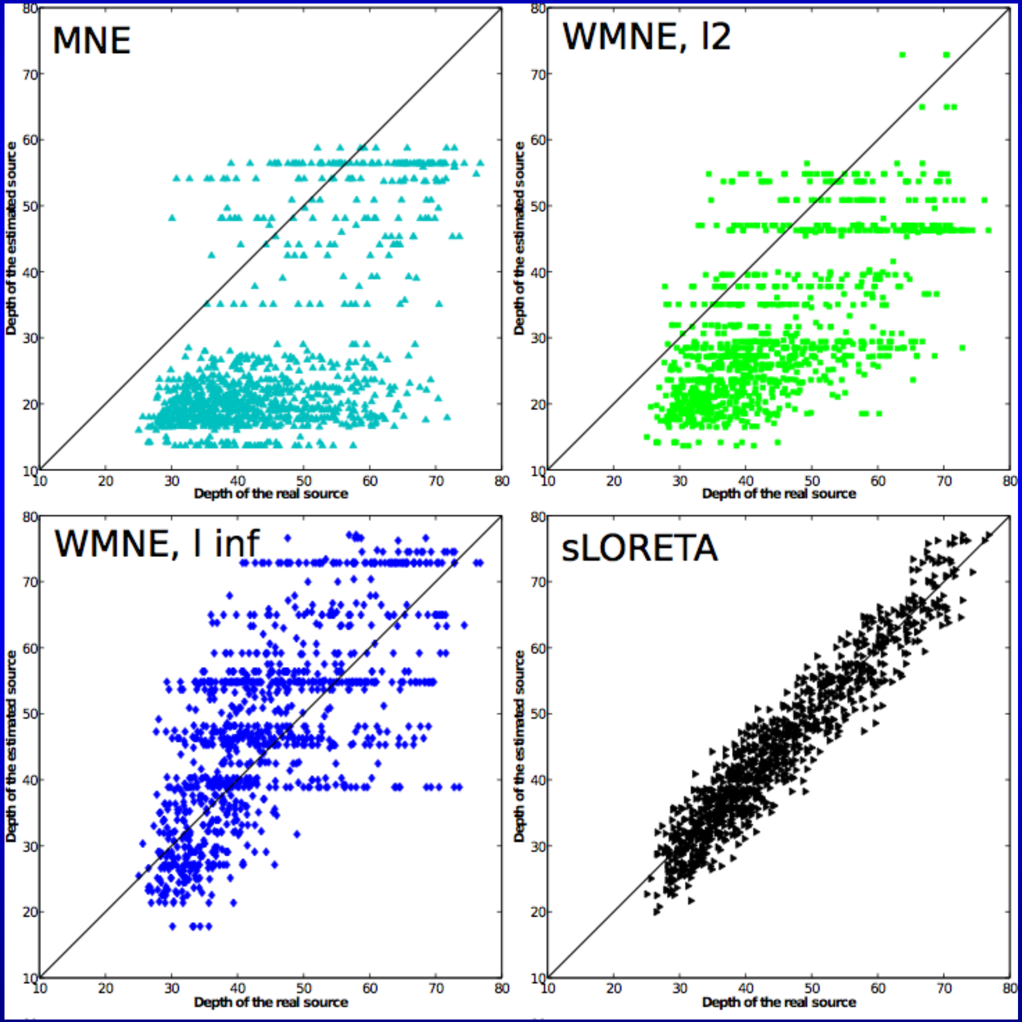


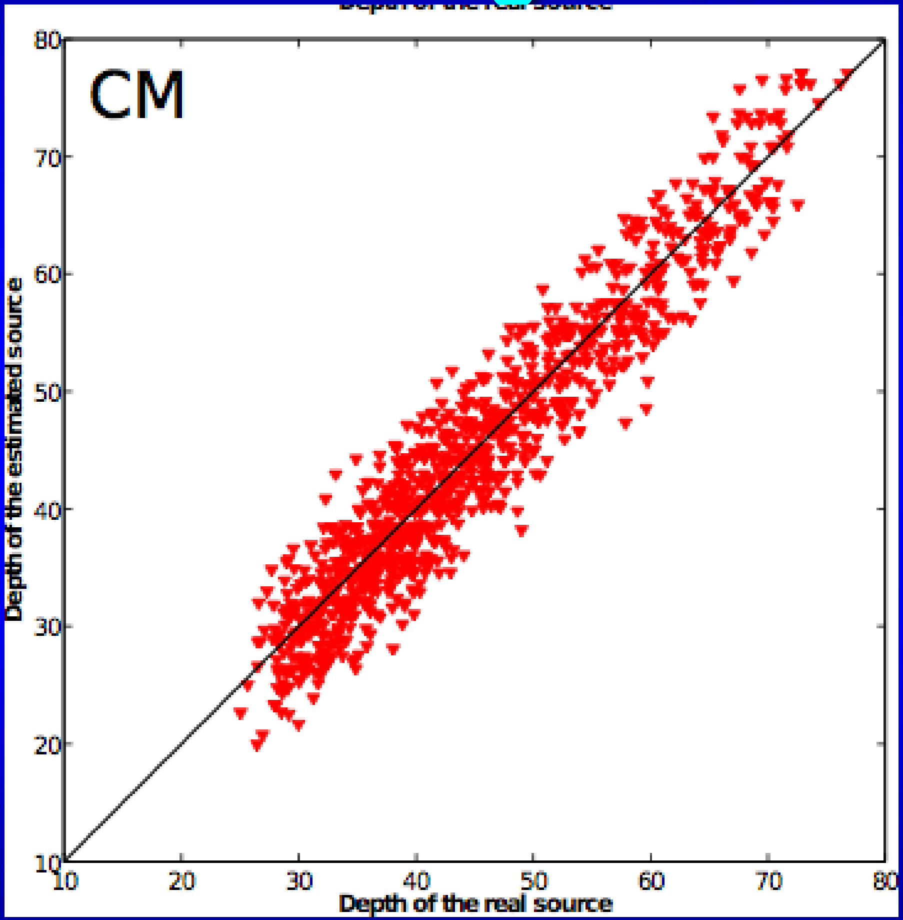
Fig. 3. Scatter plots to visualize the depth bias from different inverse methods using the f-cap. From top to bottom and from left to right: MNE, $q_{ab} = -0.441$; WMNE with ℓ_2 weighting, $q_{ab} = -0.410$; WMNE with reg. $\ell\ell_\infty$ weighting, q_{ab}=0.095; sLORETA, q_{ab}=−0.057; CM, q_{ab}=−0.058;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

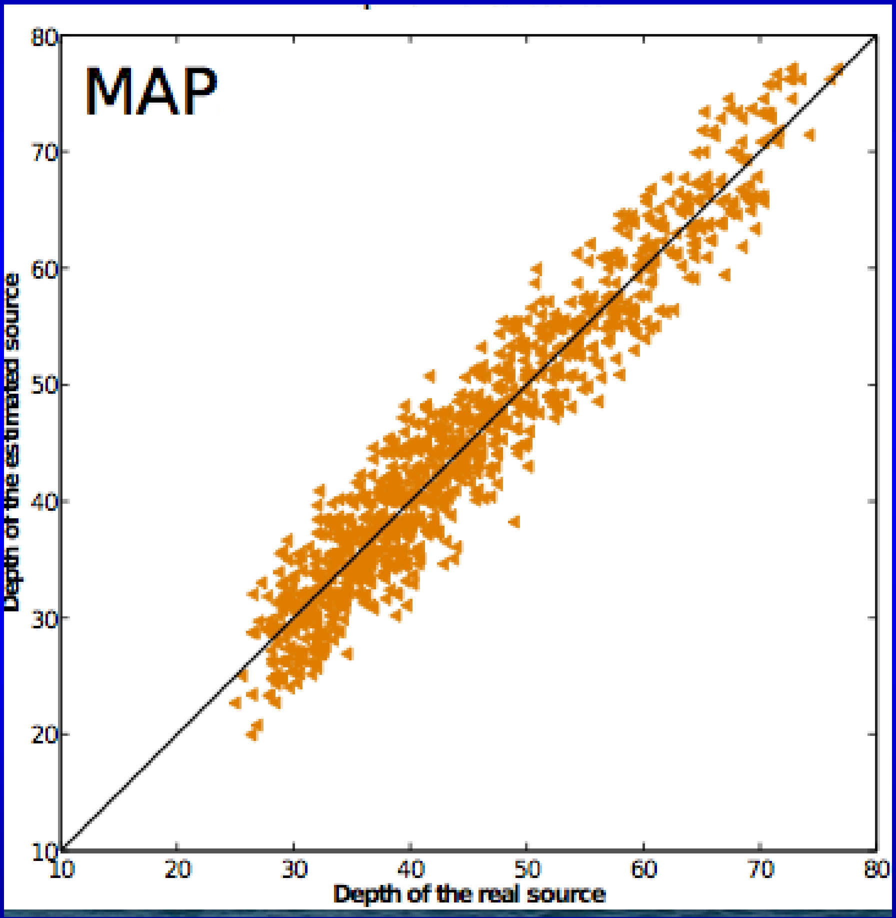
Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling : Single focal source scenario



Hierarchical Bayesian Modeling : Single focal source scenario





Hierarchical Bayesian Modeling: Study 1: Single focal source scenario

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

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

Carsten.wolters@uni-münster.de

Hierarchical Bayesian Modeling: Study 1: Single focal source scenario

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

Hierarchical Bayesian Modeling (HBM): Study 2: Two sources scenario

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

Setting

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

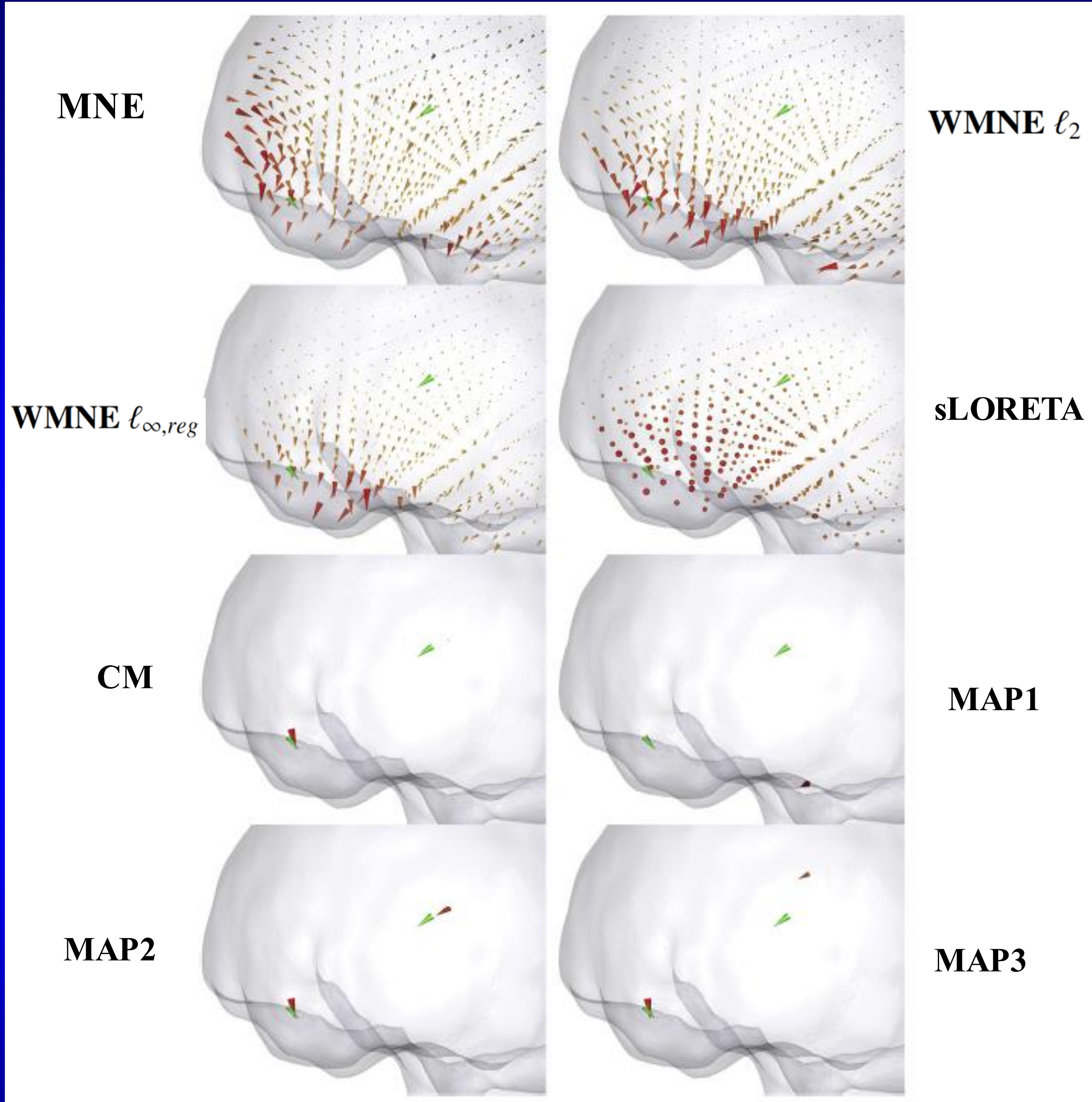
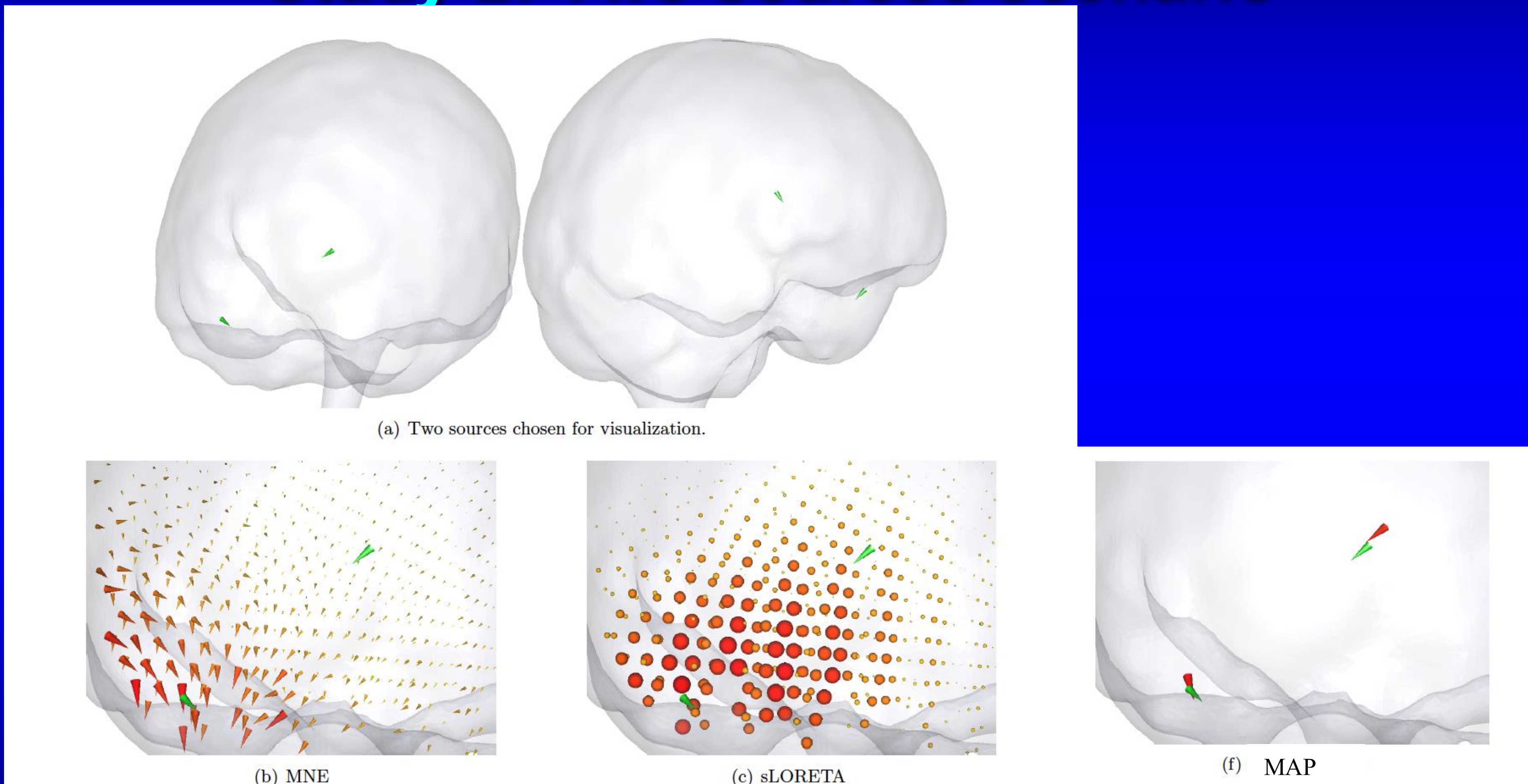


Fig. 4. Estimates of different inverse methods for a source configuration consisting of one near-surface and one deep-lying dipole (green cones) using the r-cap. The bottom left source is very close to the sensors, whereas the top right source is very distant. From top to bottom and from left to right: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

Hierarchical Bayesian Modeling: Study 2: Two sources scenario



(a) Two sources chosen for visualization.

(b) MNE

(c) sLORETA

(f) MAP

Hierarchical Bayesian Modeling: Study 2: Two sources scenario

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

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

Hierarchical Bayesian Modeling: Study 2: Two sources scenario

Method	EMD, f-cap	EMD, r-cap
MNE	44.63	45.75
WMNE ℓ_2	43.75	44.62
WMNE $\ell_{\infty,reg}$	41.79	42.78
sLORETA	36.38	38.07
CM	14.57	18.21
MAP	12.25	16.53

Hierarchical Bayesian Modeling (HBM): Study 3: Three sources scenario

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

Setting

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

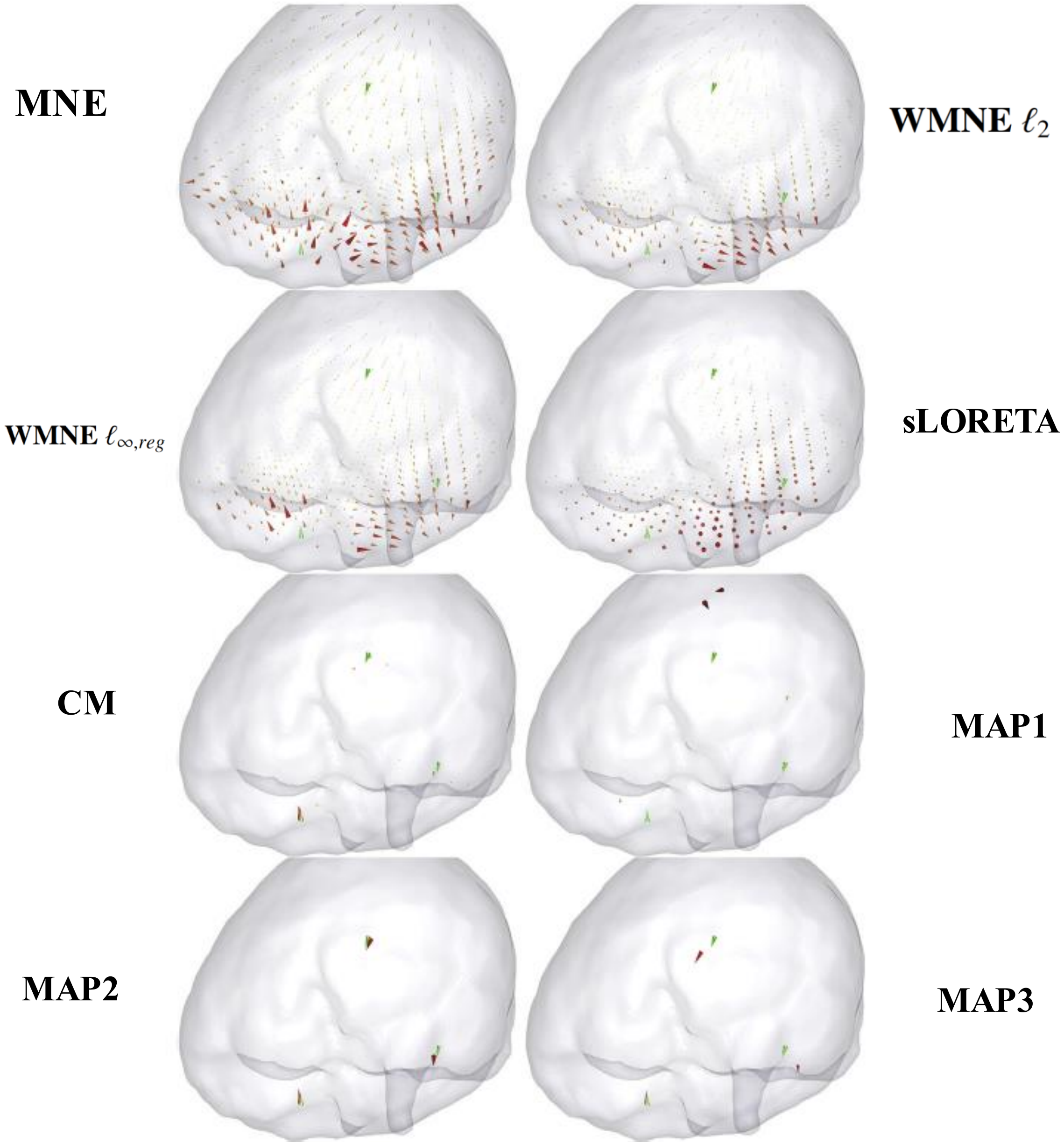


Fig. 5. Estimates of different inverse methods for a source configuration consisting of two near-surface and one deep-lying dipole (green cones) using the r-cap. The bottom sources are very close to the sensors, whereas the top source is very distant. From top to bottom and from left to right: MNE, WMNE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

Hierarchical Bayesian Modeling: Study 3: Three sources scenario

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

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

Hierarchical Bayesian Modeling: Study 3: Three sources scenario

Method	EMD, f-cap	EMD, r-cap
MNE	39.59	40.57
WMNE ℓ_2	39.02	39.78
WMNE $\ell_{\infty,reg}$	37.97	38.76
sLORETA	34.59	36.05
CM	17.60	22.35
MAP	17.10	24.25

Conclusions

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

Thank you for your attention!

Carsten.wolters@uni-münster.de