Data-driven forward model inference for EEG brain imaging

Hansen, Sofie Therese; Hauberg, Søren; Hansen, Lars Kai

Published in:
NeuroImage

Link to article, DOI:
10.1016/j.neuroimage.2016.06.017

Publication date:
2016

Document Version
Peer reviewed version

Data-driven forward model inference for EEG brain imaging

Sofie Therese Hansen[a,*], Søren Hauberg[a], Lars Kai Hansen[a]

[a]*Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby*

Abstract

Electroencephalography (EEG) is a flexible and accessible tool with excellent temporal resolution but with a spatial resolution hampered by volume conduction. Reconstruction of the cortical sources of measured EEG activity partly alleviates this problem and effectively turns EEG into a brain imaging device. The quality of the source reconstruction depends on the forward model which details head geometry and conductivities of different head compartments. These person-specific factors are complex to determine, requiring detailed knowledge of the subject's anatomy and physiology. In this proof-of-concept study, we show that, even when anatomical knowledge is unavailable, a suitable forward model can be estimated directly from the EEG. We propose a data-driven approach that provides a low-dimensional parametrization of head geometry and compartment conductivities, built using a corpus of forward models. Combined with only a recorded EEG signal, we are able to estimate both the brain sources and a person-specific forward model by optimizing this parametrization. We thus not only solve an inverse problem, but also optimize over its specification. Our work demonstrates that personalized EEG brain imaging is possible, even when the head geometry and conductivities are unknown.

Keywords: Forward model, Inverse problem, Free energy, Principal component analysis, EEG

1. Introduction

Functional brain imaging is an important tool for understanding the computational architectures underlying behavior and for guiding possible therapies for neurological diseases [1]. While EEG is growing increasingly popular for these tasks due to its experimental flexibility and excellent temporal resolution [2, 3, 4, 5], a direct interpretation of the EEG signal based on the native scalp electrode measurements is hampered by the confounding effects of volume conduction [6, 7]. However, the macroscopic EEG signal is generally believed to originate from well-localized gray matter sources [8, 9], and therefore makes full 3D spatial reconstruction of the dipole source distribution a valuable imaging modality. The source reconstruction process has been shown to reduce non-brain artifact signal components [10]; it allows incorporation of spatial a priori information from functional activation databases [11]; and it generally leads to improved interpretability [12, 13] by reducing the blurring effects of volume conduction.

The EEG scalp electrodes measure the aggregate activity of a large number of synchronously active neurons [8, 9]. At the relevant frequencies for EEG, the signal propagation from cortical sources to scalp can be considered linear and instantaneous, hence implying a linear relationship between neural activity, represented by the set of discrete dipolar sources, and the scalp measurements [6]. This linear relation can be represented by a so-called 'forward model'. Source inference is fundamentally ill-posed, as we generally have many fewer electrodes than potential locations of activated dipoles [14]. Inference is therefore highly dependent on a priori information to succeed. With a few notable exceptions to be discussed below, current research almost exclusively focuses on managing a priori information with respect to the source distributions, while considering the forward model 'known' [15]. Here, we challenge the assumption of the forward model being known and instead suggest learning the forward model from the actual EEG data, using a new data-driven representation of the set of feasible forward models.

The forward model summarizes the geometry and conductances of the various tissue compartments (skull, scalp, etc.) and is therefore inherently person-

*Corresponding author

Email addresses: sofha@dtu.dk (Sofie Therese Hansen), sohau@dtu.dk (Søren Hauberg), lkai@dtu.dk (Lars Kai Hansen)

dependent. Estimation of the forward model currently depends on access to anatomical information, e.g. in the form of computerized tomography (CT) or magnetic resonance imaging (MRI) scans of the person's head [15, 16, 17]. Such scans are segmented to produce an anatomical model consisting of nested compartments [18, 19] and a forward model is then established, essentially by solving Poisson's equation in the so-defined geometry [6]. Obtaining a high-quality model of the head geometry further demands inspection of the segmented head compartments and human intervention to correct for possible mistakes [20] and thus introduces variability and complicates the procedure. Knowing the exact head geometry must be combined with the correct conductivity values of the head compartments to yield accurate EEG source localization. Most often, these values are taken to be population averages or stem from the experimental findings of e.g. Rush et al. [21] and Cohen et al. [22]. However, it is known that the skull:brain conductivity ratio in particular varies greatly between people, and additionally that a correct specification of this ratio is important for accurate EEG imaging [23, 24]. Akalin Acar et al. [25] suggest to optimize the skull:brain conductivity ratio based on the compactness and focality of the reconstructed sources. The technique, however, is reliant on the subject's MRI data.

The lack of a well-specified forward model has led to an interest in the factors that contribute to its uncertainty, and the skull shape in particular has been found to be an important factor [24]. Statistically, the uncertainty can be represented by treating the forward model as a stochastic variable to be estimated as part of the source reconstruction problem. Bayesian evidence can be used to choose the most likely forward model among a small set of pre-defined candidates, for example [26, 27]. This does not, however, allow interpolation of forward models, i.e. a new subject is handled by a forward model from a subject in the candidate-set.

In the more general setup [28], the forward model uncertainty was represented by a multivariate Gaussian distribution, for which the mean is the conventional anatomically based estimate of the forward model. Bayesian inference then allowed for source reconstruction, where the forward model can be mildly adapted to the EEG recordings. In practice, the forward models attained were similar to the anatomically based mean, and limited flexibility was gained. Thus, there is a need for a flexible prior over forward models that allows generation of forward models tailored to new subjects.

When structural scans are unavailable, template or average models can be used. Studies have demonstrated the usefulness of spherical harmonics to describe the head anatomy and to generate approximate head models [29, 30, 31, 32]. In the noise-free case, approximate boundary element method (BEM) head models based on population averages showed relatively low localization errors [29]. The averages were suggested to be either surface-based, where the head geometry was decomposed using spherical harmonics in order to provide inter-subject correspondence, or based on averaging lead field matrices. The approximate head models were further investigated by Valdés-Hernández et al. who performed Bayesian model averaging (BMA) based on the recorded EEG to estimate a weighted average over database head models [30]. López et al. used spherical harmonics and BMA to infer the cortical surface based on optimization of the model evidence only given the M/EEG [31].

We propose to extend and combine the previous literature using a data-driven approach in which a forward model corpus is used as a prior for new subjects. Combined with the EEG of a new subject, the prior is optimized to provide an individualized forward model. In this proof-of-concept-study, we show for synthetic data that the inferred forward models for unseen subjects provide more accurate source distributions than a template forward model. We invoke the so-called Variational Garrote [33, 34]; a Bayesian framework that conveniently allows us to integrate a priori information and in recent work has shown promise for spatio-temporal source reconstruction [35, 36]. For synthetic and real EEG data, we further show that the inferred forward models lead to source reconstructions of similar quality to those obtained via the unused MRI scan of the subject. This is evidence that adequate forward models can be estimated without access to subject-specific anatomical or biophysical information. As the proposed method does not require structural scans of the new subject or the skull:brain conductivity ratio, we believe that the technique will enable a wider applicability of EEG-based imaging.

2. Methods

In the following section, we describe the first step towards a completely data-driven approach for forward model inference, also visualized in Fig. 1.

We generate a corpus of forward models from structural scans of 16 participants combined with different skull:brain conductivity ratios to produce multiple forward models for each subject. We represent the information of this forward model corpus in a low-dimensional subspace using principal component analysis (PCA) [37]. PCA representation is a generative

Postprint. Final version: http://www.sciencedirect.com/science/article/pii/S105381191630252X

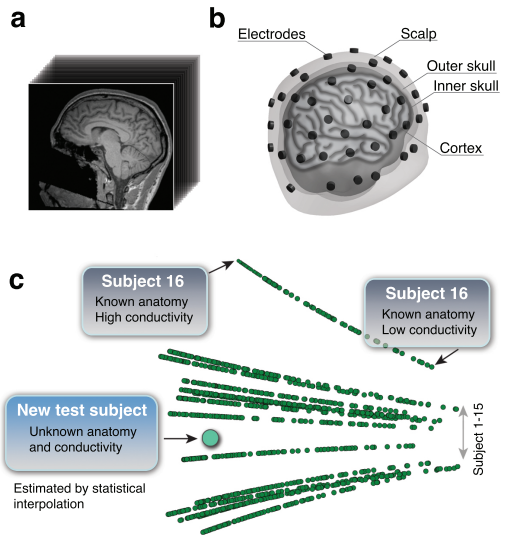

Figure 1: The process of creating forward models and their projection to PCA space. (**a**) For each of the 16 subjects, a T1-weighted image is used to construct a forward model. (**b**) The forward model is here constructed using a three-layered BEM head model (scalp-skull-brain). For each subject, 100 forward models are created with varying skull:brain conductivity; from 1:250 to 1:15. (**c**) 2D PCA projection of the forward models. The test subject is withheld from the PCA representation.

model (a probability density function [38]) and it can therefore be used to simulate or predict new forward models, effectively interpolating in the corpus of forward models. Based on the suggested data-driven representation, it is possible to propose or actively search for a potential forward model for a person not included in the database. We suggest to infer a forward model for a new subject by using this person's recorded EEG signal to optimize an estimate of the model evidence, visualized in Fig. 2 and expressed in equation A.6. The main steps in the proposed forward model inference pipeline thus include:

- Generate a corpus of P forward models representative of variations in head geometry and conductivities. The corpus is contained in a matrix of size $P \times (N \cdot K)$ when defining each forward model to map N cortical sources to K electrodes.
- Decompose the forward model corpus using PCA and create a low-dimensional representation of forward models.
- For a new subject, search for a forward model in the PCA representation which optimizes the free energy given the EEG data and the source model. The result is a personalized forward model and a source distribution for the new subject.

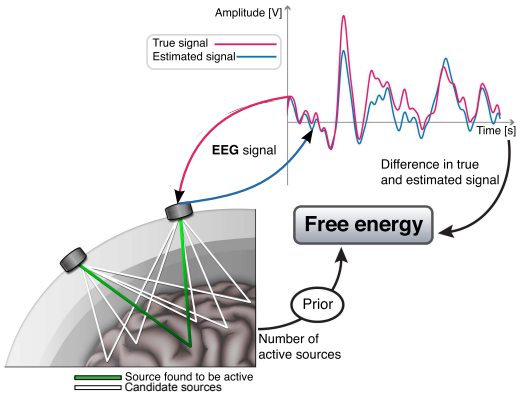

Figure 2: The free energy summarizing the ability of a model to describe the data and its complexity. We calculate the free energy based on the inference scheme proposed in the Variational Garrote (VG). Formally, the free energy in VG expresses the Bayesian combination of data fit (difference in true and estimated signal) which considers the estimated source density and the forward model, and complexity through a sparsity-promoting prior on the source density.

These steps are more carefully described below together with the data used for validating the method.

2.1. Neuroimaging Data

We apply the EEG recordings and structural MRI data from 16 healthy subjects (F=7, M=9, age=23-31 years) from the multimodal dataset acquired by R. Henson and D. Wakeman [39, 40]. Functional MRI (fMRI) and MEG were also recorded but not applied in this study. The EEG was recorded with 70 electrodes and the structural MRIs are T1-weighted images recorded on a Siemens 3T Trio. The study was originally conceived and carried out to investigate the mechanisms of face perception [41]. We preprocess the EEG data following the SPM8 (`http://www.fil.ion.ucl.ac.uk/spm`) [18] framework through MATLAB (Mathworks Inc.) scripts provided by R. Henson. The data are thus filtered and averaged across epochs within conditions. Finally we create the differential event-related potential (ERP) contrasting 'faces' versus 'scrambled faces' for one test subject. We thus follow the common approach in investigating the face-evoked response, i.e. by creating the differential response and thereby strengthening the face-sensitive signal [42, 43].

2.2. Forward Modeling

We employ the widely used software packages, SPM8 [18] and FieldTrip (`http://www.fieldtriptoolbox.org/`) [19] to create a database

of BEM forward models. The structural MRI of each subject is thus spatially normalized to a template (MNI) brain. The inverse of this transformation is used in warping a template/canonical mesh into a subject-specific mesh on which a forward model is generated [44]. Each of the 16 participants' anatomical MRI scans (Fig. 1a) are thus segmented into scalp, skull and brain (Fig. 1b). The BEM, in the 'bemcp' implementation [45], is used to create the forward models with scalp, skull and brain conductivities corresponding to $[1, c, 1] \cdot 0.33$, where c is drawn from a uniform distribution between 1/250 and 1/15. For each subject, 100 samples are drawn and combined with the subject's segmented skull layer thus generating in total 1,600 forward models covering the relevant range of skull:brain conductivity ratios [28]. In contrast, SPM8 fixes the conductivities to $[1, 1/80, 1] \cdot 0.33$. Co-registration to the EEG electrodes is obtained through fiducials placed on the nasion and the left and right pre-auricular, and through headshape points. The cortex mesh is set to consist of 8196 vertices.

Although commonly used, the applied procedure to generate forward models impose several simplifying assumptions. For example, the head is modeled as consisting of only three head layers and each of these have isotropic conductivity. According to several studies, a layer modeling the cerebrospinal fluid (CSF), for example, should be included in the head model to obtain accurate EEG imaging [24, 46]. However, a recent study shows that the omission of a CSF layer can be partly compensated for by adjusting the skull conductivity appropriately [20]. As we do not fix the skull:brain conductivity ratio in our study, but instead approximate it based on the EEG data, the influence of the missing CSF is expected to be reduced. Anatomical compartments with anisotropic conductivities can be achieved by replacing the BEM head model with finite element method (FEM) estimations [6]. The BEM head model is, however, often applied because of its low complexity and high accessibility. A further simplification is the assumption that the cortical folding of a subject can be accurately described by a nonlinear warping of a template model, as implemented in SPM8. The benefit of this method is the existence of a direct one-to-one correspondence of brain locations between subjects. Akalin Acar et al. furthermore showed that a subject-specific warping of a template head model provides reasonable source recovery of scalp maps generated by a more realistic BEM forward model [24]. Note that these forward models did not assume fixed dipole orientations, as we do in this study. As we are aiming at generating forward models personalized to subjects for whom the head geometry is unknown, these simplifications are considered reasonable. We finally note that it is indeed possible to implement more realistic forward models in the proposed framework.

2.3. Forward Model Representation

Using PCA [37], we obtain a low-dimensional representation of the corpus of forward models (Fig. 1c). Each forward model is a 70×8196 matrix, which we reshape to produce vectors with 573,720 elements. Forward models are removed from the corpus if their l_2-norm deviates by more than two standard deviations from the average l_2-norm. Of the 1,600 forward models, 49 are excluded and the matrix $\mathbf{L} \in \mathbb{R}^{1{,}551 \times 573{,}720}$ thus contains the forward models used for the PCA analysis. Eigendecomposition is applied to the inner product of the corpus (where the average forward model has been subtracted), i.e.

$$\mathbf{\Sigma_L} = \mathbf{L}\mathbf{L}^\top = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top, \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{1551 \times 1551}$ contains the eigenvectors and $\mathbf{\Lambda} \in \mathbb{R}^{1551 \times 1551}$ contains the eigenvalues in the diagonal. $\mathbf{L}$ can be decomposed by $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^\top$ meaning that $\mathbf{V}^\top = (\mathbf{U}\mathbf{\Lambda}^{1/2})^{-1}\mathbf{L} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^\top\mathbf{L}$, where $\mathbf{V} \in \mathbb{R}^{573{,}720 \times 1551}$. A new lead field is given by $\mathbf{A}_{new} = \mathbf{w}\mathbf{V}^\top$, where $\mathbf{w}$ is a row vector containing the position of a forward model in the PCA space. For visualization purposes, we create a two-dimensional PCA representation. The new basis $\bar{\mathbf{V}} \in \mathbb{R}^{573{,}720 \times 2}$ is thus formed by the two principal components explaining most variance, corresponding to the two first columns of the eigenvalue sorted matrix $\mathbf{V}$. A forward model in PCA position $\mathbf{w} \in \mathbb{R}^{1 \times 2}$ can therefore be generated by $\mathbf{A}_{new} = \mathbf{w}\bar{\mathbf{V}}^\top$. To establish an unbiased estimate of the goodness, we invoke a leave-one-out cross-validation setup, i.e. we estimate the forward model PCA representation on all but one test subject.

The 2D projections of the forward models using the two first principal components are seen in Fig. 1c. While the horizontal dimension in Fig. 1c is clearly dominated by the skull conductivity value (decreasing from left to right) the interpretation of the vertical dimension is less clear. It therefore appears to be a composite of both inter-individual anatomical differences and the conductivity ratio. Subject 16, for example, has a bigger sized brain than the other subjects, as seen in Supplementary Fig. 1, and is also seen to be something of an outlier in the vertical dimension of the PCA representation. However, brain size alone does not explain the subjects' locations in the PCA space in Fig. 1c. The

Postprint. Final version: http://www.sciencedirect.com/science/article/pii/S105381191630252X

two first principal components explain 73% of the forward model variance, while 99% of the variance can be explained by the first 18 principal components.

We base our forward model inference on a measure of statistical goodness. Here, we use the free energy (see Fig. 2), which provides a bound on the evidence in Bayesian modeling [47]. When optimized, the free energy can thus be used to quantify the evidence. As seen in Fig. 2, the free energy provides optimal data fit while penalizing complexity. We optimize the free energy with respect to both the source configuration as well as the forward model representation, as similarly done in [48]. We apply a source localization procedure based on a statistical model whose prior favors sparse solutions; the so-called Variational Garrote [33, 34, 35, 36], described in Appendix A. The source localization procedure is contingent on a single regularization parameter: the prior sparsity level. Sparsity is a common assumption, employed in estimating ill-posed inverse solutions, and widely applied in EEG imaging [12, 39, 28, 42]. In EEG, the sparsity assumption is motivated by the apparently sparse focal nature of brain activation [16].

Cross-validation is a general technique used to estimate how well a model generalizes to new data, and for independently sampled data, the performance estimator is unbiased [49]. Here, we apply cross-validation at two levels: At the forward model level to optimize statistical regularization parameters (the sparsity level which determines the number of active dipoles), and at the corpus level to infer the forward model for a hold-out subject, as previously mentioned. For the first level of cross-validation, we split the EEG data into four folds by partitioning the 70 EEG electrodes (Fig. 3a). Each fold contains 17-18 electrodes and covers the surface of the scalp. While, importantly, the overall performance estimator is unbiased, the correlations among electrode signals imply that our parameter estimation step may be suboptimal.

2.4. Sparsity Estimation

In our analysis, the optimal forward model is the one which yields lowest free energy, as defined in eq. (A.6). The free energy, however, depends on both the unknown forward model as well as on the sparsity parameter γ. The latter is estimated using four-fold cross-validation at 250 randomly selected training forward models and interpolated across the PCA space using kernel regression [50] with a Gaussian kernel. The bandwidth of the kernel specifies the smoothness with which γ changes. In order not to depend on a particular choice of the bandwidth, we consider a uniform prior on this parameter,

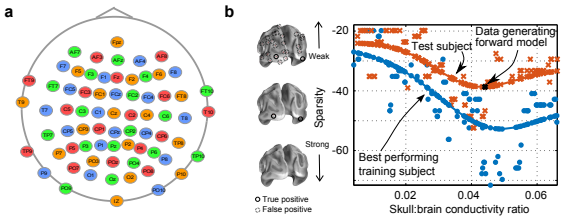

Figure 3: Estimation of optimal sparsity. (a) Partitioning of the 70 EEG electrodes into four folds. Each fold is represented by one color. (b) The sparsity levels obtained by cross-validation using the mean squared error (MSE) between the true and predicted signal. For visualization purposes, we only show the results for the test subject (red crosses) and the non-test subject who obtained best F_1-measure (blue dots), see also Fig. 5. The black cross indicates the data-generating forward model. The estimated sparsity levels were smoothed (full lines) with respect to conductivity ratio within each subject.

which is marginalized numerically such that the free energy is evaluated and then averaged across a discrete set of bandwidth values (from 0.25 to 3 in 12 steps). However, for the simulation studies we investigate only one bandwidth. For the database forward model prediction, we smooth the sparsity within each subject across conductivity ratio, see Fig. 3b. The smoothing is in general performed to reduce the noise introduced by the coarse four-fold cross-validation procedure.

2.5. Synthetic EEG Data

In the first experiment we construct synthetic data by positioning two bilateral sources in the occipital lobes (Fig. 4a-b). In the second simulation study we additionally plant two frontal sources, (Fig. 6a-b). The sources in the two studies have the temporal dynamics of one or two pairs, respectively, synchronous sine waves across 25 time samples. Assuming a sampling frequency of 200 Hz, the simulated sine waves have a frequency of approximately 15 Hz. The created source distribution is projected to 70 electrodes through the forward model of a test subject with a specific skull:brain conductivity ratio. We add noise to yield a signal-to-noise ratio (SNR) of 5 dB.

3. Results

Our analysis of predictive forward model representations was based in part on simulated data and in part on real EEG data. For both simulated and real EEG data, we used the forward models previously described.

To validate the predicted 2D PCA forward models, we calculated selected summary data; the matrix coherence [51] and the condition number [52], see Table 1.

	Real	Predicted
1−coh [$\times 10^{-4}$]	2.7 (0.95 - 5.5)	2.6 (1.0 - 4.1)
κ	97.3 (59.2 - 795.6)	109.4 (61.5 -281.1)

Table 1: Matrix properties of the real and PCA-predicted forward models. Median (and full interval) of the coherence (coh) and condition number (κ) are shown. As all forward models approach a coherence of 1, we show 1 minus the coherence. Note that by excluding the 'outlier subject' (subject 16) the maximum condition number among the real forward models was 151.

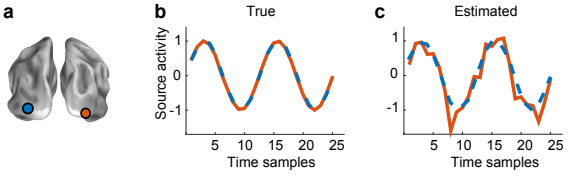

Figure 4: Source distributions of the planted activity and as estimated using the free energy-optimal forward model. (a) Posterior view of the inflated brain showing the locations of the two planted sources. The locations of the estimated sources were identical to these. (b) The real and (c) estimated time courses of the two sources.

High correspondence between actual and predicted forward models was found for both measures. In further studies we found that, while increasing the number of principal components yielded higher similarity between the predicted and real forward models, it did not necessarily increase source recovery accuracy (Supplementary Fig. 2-4).

3.1. *Performance of database forward models - Simulations*

As a validation step, we investigated whether, from all of the 1,600 corpus forward models, the free energy was able to recover an adequate forward model. The simulated EEG signal for this study arose from two active sources, see Fig. 4a. In Fig. 3b, we show the estimated sparsity levels and the smoothed values for the test subject and the best-performing non-test subject across skull:brain conductivity ratios.

The forward model with the lowest free energy was found to be from the test subject and had a conductivity ratio very similar to the data-generating forward model (Fig. 5a). This choice of conductivity ratio was further supported by also having low cross-validation error (Fig. 5b). However the cross-validation error seemed to be more unspecific and did not have an as clearly defined minimum as the free energy. The geometric localization error (Fig. 5c) and a source retrieval index, viz. the F_1-measure [53], balancing the source localization precision and recall scores (Fig. 5d), attained their optimal values at the conductivity ratio with lowest free energy. Furthermore, the best-performing training subject

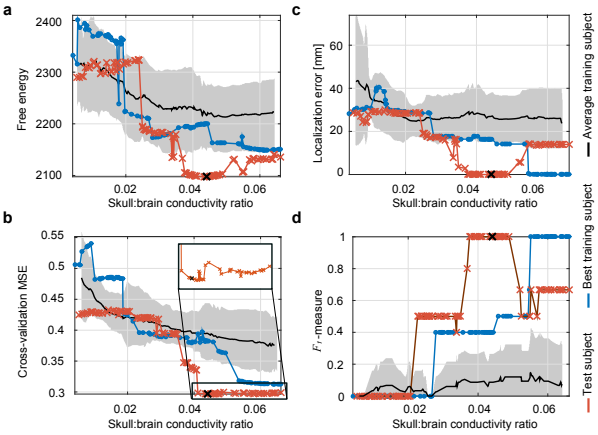

Figure 5: Forward model prediction among the database test and training subjects for simulated data. The source signal in Fig. 4a-b was projected to sensor space with the forward model indicated by the black cross. The smoothed sparsity levels in Fig. 3b were applied to the Variational Garrote in combination with the forward models of the test subject (red) and the training subjects (averaged in black, s.d. in gray, training subject with highest F_1-measure in blue). (a) The free energy computed on all electrodes. (b) The normalized cross-validation MSE including zoom inset. (c) The Euclidean localization error summed across the two sources. (d) The F_1-measure.

also had a subset of forward models leading to perfect source reconstruction. Using the free energy-optimal forward model from the set of test and training subjects, we obtained a source distribution with the correct source locations, and temporal dynamics very similar to the true activity (Fig. 4).

In our example, the consequence of choosing a wrong conductivity ratio when having a forward model based on the subject's structural scan is a summed localization error of up to 30 mm (Fig. 5c), i.e. an average error of 15 mm. This result is in line with previous studies [24].

3.2. *Performance of 2D PCA-generated forward models - Simulations*

Next, we assessed the ability of the free energy to optimize over the PCA-predicted set of forward models, i.e., not restricting ourselves to the actual database of forward models. The simulated source activity is seen in (Fig. 6a-b). The PCA-projected forward models (Fig. 1c) of the training subjects spanned our search space (Fig. 7 and Supplementary Fig. 5). Note that we withheld the forward models of the test subject from the PCA decomposition.

The lowest free energy (Fig. 7a) matched the optimal region of the localization error (Fig. 7b) and F_1-measure (Fig. 7c). The sources localized with the forward model having the lowest free energy were thus accurately placed and additionally had similar temporal

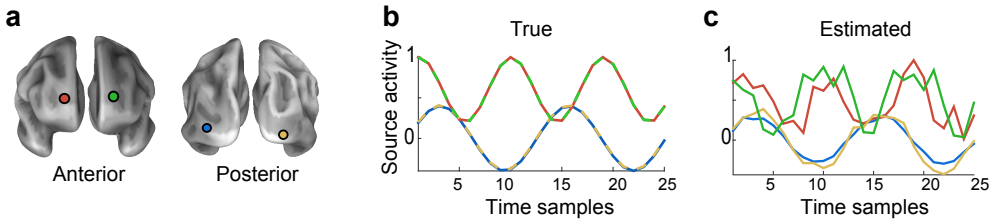

Figure 6: Source distributions of the planted activity and as estimated using the free energy-optimal forward model. (a) Posterior view of the inflated brain showing the locations of the four planted sources. The locations of the estimated sources were identical to these. (b) The real and (c) estimated time courses of the four sources.

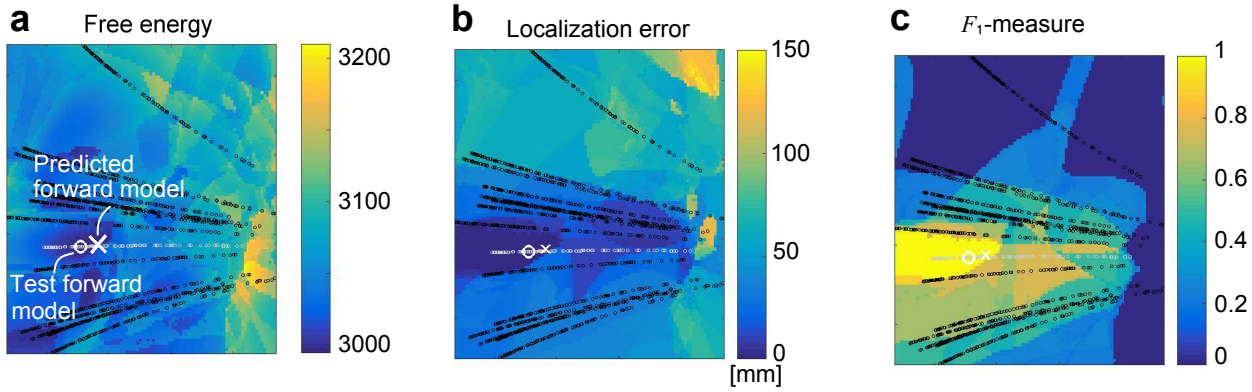

Figure 7: Search for the optimal forward model in the 2D PCA space created by the training subjects; simulated EEG data (Fig. 6). A test forward model (white circle) from the withheld test subject was used to generate the test EEG data. The 2D PCA space of forward models wherein (a) the free energy, (b) the localization error and (c) the F_1-measure were calculated. The forward models of the training subjects (black) and test subject (gray) are overlayed. Minimum free energy was found at the white cross.

dynamics to the truth (Fig. 6c). In Table 2, we compare the performance of the recovered forward model with that of template and subject-specific forward models. The forward model having the correct anatomy and conductivity is seen to perform similarly to the inferred forward model. Assuming template conductivity ratio performed reasonably well, while also using template anatomy severely impaired the performance.

The inference pipeline was investigated on five more subjects. In general, we found that a reasonable forward model could be inferred when the requested test forward model was in the span of the training forward models (Supplementary Figs. 6-10). For three subjects, the source densities estimated with the inferred forward models were of similar high performance as the true forward models (Supplementary Tables 2-4). For the fourth subject, one of the sources was not retrieved by the predicted forward model (Supplementary Table 5). Finally, when using the 'outlier subject' as the test subject, we were only able to recover one of the simulated sources (Supplementary Table 6).

3.3. Performance of 2D PCA-generated forward models - Real EEG data

Finally, we demonstrate the forward model inference pipeline on a real EEG dataset. We used the differential EEG response of seeing faces versus scrambled faces stemming from EEG data recorded from the same test subject as used in the previous experiments. Again we created a 2D PCA space using the remaining 15 subjects on which we investigated the free energy, cross-validation error profile and the sparsity profile (Fig. 8a-c; see additionally Supplementary Fig. 11).

The free energy-optimal forward model for the left-out-subject's EEG data provided a source distribution (Fig. 8d) with a maximal response at 160 ms, corresponding to the N170 face-related EEG component [41]. The estimated sources were bilaterally located and four of the dominating sources were in the vicinity of the O/FFA (Fig. 8d, upper panel). The recovered face perception locations were thus consistent with previous EEG/MEG [43], as well as fMRI [54] studies. We further compared our results to the source densities obtained when applying a forward model built

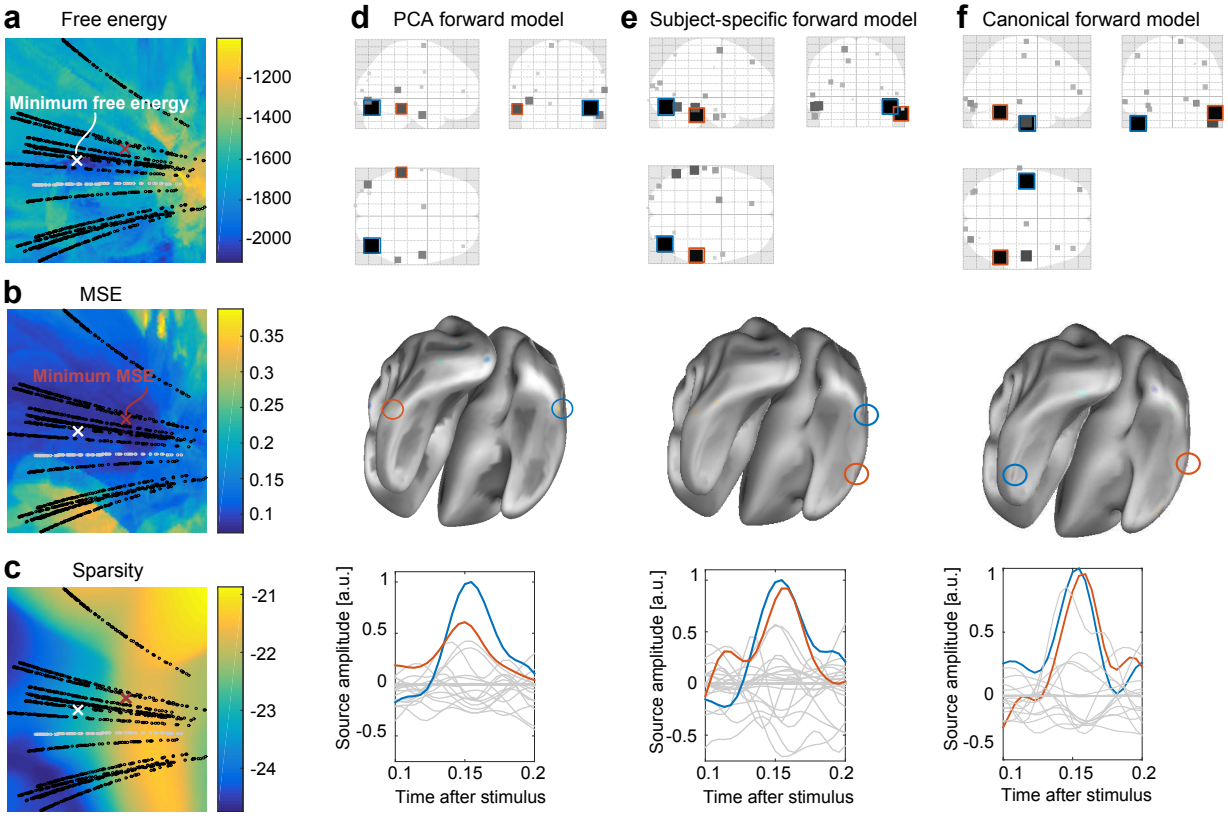

Figure 8: Search for the optimal forward model; real EEG data recorded from the test subject. (**a-c**) The 2D PCA space created by 15 training subjects, whose projections are seen in black. The test subject is seen in gray. The minimum free energy is indicated by a white cross and minimum MSE by a red cross. Due to uncertainty concerning the bandwidth controlling the smoothing of the sparsity, several bandwidths were applied; the averages across these are shown. (**d-f**) The source densities as estimated by the forward model yielding minimum free energy in the PCA space and the canonical forward model with and without adaptation to the test subject's sMRI. Upper panel: Glass brain representation of the recovered sources; spatial extension and color intensity reflect the relative activity strength. Mid panel: 3D representation of the estimated source localization visualized on an inflated cortex. The two strongest sources are indicated in blue and red. Lower panel: The temporal dynamics of the two strongest sources; the remaining sources are shown in gray.

Table 2: Performance of the forward model inferred by lowest free energy (white cross in Fig. 7) and forward models constructed from template/subject-specific head geometry and skull:brain conductivity ratio (σ). Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior.
[†]Calculated as the Euclidean distance between a true source and the strongest estimated source in the same half hemisphere.
[‡] Calculated as the sum of the variational mean, **m** (see Appendix A).

	Optimized forward model	Template head, template σ	Subject head, template σ	Subject head, true σ
Free energy	2994	3192	3057	2956
MSE	0.63	0.55	0.93	0.61
F_1-measure	1	0	0.44	0.5
Localization error [†]				
Left posterior	0 mm	16.7 mm	15.1 mm	0 mm
Right posterior	0 mm	18.4 mm	6.0 mm	6.0 mm
Left anterior	0 mm	19.7 mm	0 mm	5.7 mm
Right anterior	0 mm	23.7 mm	0 mm	0 mm
Sum	**0 mm**	**78.5 mm**	**21.1 mm**	**11.7 mm**
Estimated number of active sources[‡]	4.0	4.0	5.1	4.0

using the MRI scan of the subject and the SPM8 default skull:brain conductivity ratio 1:80 (Fig. 8e), and obtained when using the canonical/template forward model (Fig. 8f). The personalized canonical forward model provided a source distribution similar to the one of the PCA forward model, however, the estimated sources were less symmetrically located (Fig. 8e, upper panel). The template forward model yielded even less hemispheric symmetry and more anteriorly located activity (Fig. 8f, upper panel). The typical face-related N170 peak was recovered with all three forward models (Fig. 8d-f, lower panel).

4. Discussion

Functional brain imaging by EEG source localization poses a highly ill-posed inverse problem due to the low spatial resolution of the sensors and the high number of potential locations of the cortical sources. There is a broad consensus that forward model uncertainty is an important limiting factor for EEG imaging by source reconstruction [17, 24, 55, 56, 57, 58, 59]. Our results add quantitative evidence to this view, both in terms of the tissue conductivity ratio, which is the main source of uncertainty if the brain topography is correct, and more broadly when both anatomy and conductivities are unknown. This evidence is our main motivation for proposing a data-driven inference scheme for the forward model: Is there a way to reduce the uncertainties inherent in conventional electrophysiological tools for estimating forward models, i.e. the uncertainty of brain topography and conductivity distributions?

Previously, attempts have been made to achieve point estimates of the conductivity ratio or to model it as a random variable establishing a posterior distribution that encodes the uncertainty of the forward models. However, in the former case the ratio is estimated from a discrete set of specific values [55]. In the latter study validation was found to be challenging in real data, and it was suggested that in future work, the validation could be assisted by active conductance mapping using electrical impedance tomography (EIT) [56]. These techniques, however, introduce a new set of highly ill-posed inverse problems. While we here focus on forward model inference in the setting of EEG, we note that the methods developed may also assist other important tools, such as transcranial magnetic stimulation, direct-current stimulation [57], and indeed EIT.

As a route of reducing forward model uncertainty, we proposed a data-driven mechanism for building a representation of forward models based on the variability expressed in a large corpus of models. This approach represents the database as a relatively low-dimensional manifold, here chosen to be a two-dimensional linear subspace. Equipped with an appropriate probability density function, the representation allowed us to simulate new forward models and search for the best-suited forward model for a specific EEG dataset, without involving the subject's anatomical data. We showed that the predicted forward models based on the new representation share important characteristics with the database models. We opted for a rather simple, two-dimensional representation, for the sake of visualization. However, the complexity of the forward model representation can be inferred by statistical means: the more data, the more complex the forward model representation [38].

To evaluate the goodness of a given forward model for a specific EEG dataset, we applied the Variational Garrote [34]; a Bayesian sparsity-promoting source reconstruction approach producing two mea-

Postprint. Final version: http://www.sciencedirect.com/science/article/pii/S105381191630252X

sures of goodness: the 'free energy', a measure of the model evidence, and cross-validation error based on the scalp electrode measurements. These measurements were themselves validated in simulation experiments in which we showed that the free energy identifies forward models with small source localization errors and general high accuracy. Future work will investigate whether the conceptual approach can also be used with other inferential frameworks. The inverse solvers implemented in SPM, for example, also provide estimates of the model evidence [60].

The possibility of effectively recovering important aspects of the forward model directly from EEG data using a data-driven approach is the main novelty of our method. In the state-of-the-art approach [26], the optimal model is selected within a limited set of candidate models, all based on the given subject's anatomical data, i.e. requiring an MRI or CT scan. We presented evidence that our approach can infer the forward model for a test subject not included in the database. The simulation study indicated that, by optimizing the free energy, we can identify a set of forward models that have optimal source retrieval. Thus, our results have immediate consequences for studies for which the brain topography is not available, e.g. because MRI or CT scans are not recorded, or because available scans do not provide enough detail. Our method also has potential to be beneficial for specific patient groups for which an MRI or CT scan is practically/ethically unobtainable, e.g. for patients in pain, with claustrophobia or other factors making it difficult for the subject to remain immobile. Furthermore, the EEG is often recorded with the subject being in a different position than when the structural scans were recorded, and this could misrepresent the actual propagation paths from source-to-scalp measures [61]. This could potentially be remedied by adapting the forward model using the free energy, as similarly suggested for inferring the head position in MEG acquisition [31]. Finally, one may speculate whether the approach can be generalized to a dynamic scenario in which the subject is in motion and hence the brain position relative to the skull and scalp varies, calling for a dynamic forward model.

The dataset [40] from where we obtained EEG and the anatomical MRI scans additionally contains MEG and fMRI datasets for all 16 subjects for the face-recognition task. This paradigm has previously been used to test EEG and MEG source reconstruction methods [26, 59, 42] and thus allowed us to test the forward model inference hypothesis. The functional data were acquired to identify the networks involved in human-face processing, and consist of randomized presentations of human faces and scrambled faces. On the differential EEG response, i.e. the signal mean difference for the two conditions, we found activation located in the vicinity of the left and right O/FFAs, showing the face-related N170 component [41]. The most direct comparison can be made with a multi-modal fusion study [43], which compared and fused MEG and EEG data to investigate the spatial location of sources and the response dynamics. When analyzing MEG data alone, activations in the vicinity of the left and right FFAs were found, while when analyzing the EEG data, activations in the vicinity of the OFAs were found. Combining both MEG and EEG modalities made it possible to reproduce the activation in all four face areas, as also found in fMRI studies [54]. Our results are thus consistent with the EEG/MEG-combined findings. It is our experience that assuming spatial coherency improves source reconstruction further, e.g. by using spatial basis functions [42].

While the present study gives evidence that it is indeed possible to infer forward models based on a subject's EEG data and an external database of general anatomical information, it should be extended in several directions. First, we aim at making the manifold description richer by using more realistic head models [57] and representing the information in higher dimensions. The latter would hinder a grid search for the optimal forward model due to the 'curse of dimensionality' and optimization techniques such as Bayesian schemes, e.g. BayesOpt [62] or Metropolis search combined with BMA [48] would become necessary. The database can be further extended by adding head geometries for more subjects using large anatomical scan databases such as the Biomedical Informatics Research Network [63]. The applied database contains healthy subjects of similar age and the generated forward model representation is therefore not expected to generalize directly to very dissimilar subject groups. However, through the creation of a large and comprehensive database, we would potentially be able to infer the diverse and complex head geometry that influences the measured EEG signal and thereby obtain better source localization results for a wide group of subjects. By expanding the ability of EEG to act as a stand-alone brain imaging device, the presented strategy therefore has potential to play a key role in understanding the mechanisms of cognitive processes.

Acknowledgements

We thank D. Wakeman and R. Henson for providing the dataset used. We thank the reviewers for their

detailed and constructive comments that significantly improved the manuscript. The work was supported in part by the Novo Nordisk Foundation Interdisciplinary Synergy Program 2014 ['Biophysically adjusted state-informed cortex stimulation (BASICS)'] (STH), the Danish Research Council for Natural Sciences (SH), and the Danish Innovation Foundation (LKH).

References

[1] Soekadar, S. R., Witkowski, M., Cossio, E. G., Birbaumer, N., Robinson, S. E., Cohen, L. G. In vivo assessment of human brain oscillations during application of transcranial electric currents. *Nature Comm* **4**, 2032 (2013).

[2] Kouider S., Stahlhut, C., Gelskov, S.V., Barbosa, L.S., Dutat, M., De Gardelle, V., Christophe, A., Dehaene, S., Dehaene-Lambertz, G. A neural marker of perceptual consciousness in infants. *Science* **340**, 376-380 (2013).

[3] Hulbert, S., Adeli. H. EEG/MEG-and imaging-based diagnosis of Alzheimer's disease. *Reviews in the neurosciences* **24**, 563-576 (2013).

[4] De Ciantis, A., Lemieux, L. Localisation of epileptic foci using novel imaging modalities. *Current opinion in neurology* **26**, 368:373 (2013).

[5] Stopczynski, A., Stahlhut, C., Larsen, J.E., Petersen, M.K., Hansen, L.K. The smartphone brain scanner: A portable real-time neuroimaging system. *PloS one* **9**, e86733 (2014).

[6] Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K.P., Fabri, SG., Van Huffel, S., Lemahieu, I. Review on solving the forward problem in EEG source analysis. *Journal of neuroengineering and rehabilitation* **4**, 1-29 (2007).

[7] Aydin, U., Vorwerk, J., Kupper, P., Heers, M., Kugel, H., Galka, A., Hamid, L. Combining EEG and MEG for the Reconstruction of Epileptic Activity Using a Calibrated Realistic Volume Conductor Model. *PloS one* **9**, e93154 (2014).

[8] Nunez, P.L., Srinivasan, R. *Electric fields of the brain: the neurophysics of EEG.* Oxford university press (2006).

[9] Nunez, P.L., Srinivasan, R., Fields, R.D. EEG functional connectivity, axon delays and white matter disease. *Clinical Neurophysiology* **126**, 110–120 (2015).

[10] Besserve, M., Martinerie, J., Garnero. L. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage* **55**, 1536-1547 (2011).

[11] Baillet, S., Garnero, L. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Transactions on Biomedical Engineering* **44**, 374-385 (1997).

[12] Ahn, M., Hong, J.H., Jun, S.C. Feasibility of approaches combining sensor and source features in brain–computer interface. *Journal of neuroscience methods* **204**, 168-178 (2012).

[13] Edelman, B. J., Baxter, B., He, B. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Transactions on Bio-Medical Engineering* **63(1)**, 4–14 (2016). doi:10.1109/TBME.2015.2467312

[14] von Helmholtz, H.L.F. Some laws concerning the distribution of electric currents in volume conductors with applications to experiments on animal electricity (translated). *Proceedings of the IEEE* **92**, 868-870 (2004).

[15] Hämäläinen, M., Hari, R.,Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics* **65**, 413-460 (1993).

[16] Baillet, S., Mosher, J.C., Leahy, R.M. Electromagnetic brain mapping. *Signal Processing Magazine, IEEE* **18**, 14-30 (2001).

[17] Oostenveld, R., Oostendorp, T.F. Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human brain mapping* **17**, 179-192 (2002).

[18] Ashburner, J., Chen, C.-C., Moran, R., Henson, R.N., Glauche, V., Phillips, C. SPM8 manual. *The FIL Methods Group*, (2012).

[19] Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* (2011). doi:10.1155/2011/156869

[20] Stenroos, M., Nummenmaa, A. Incorporating and compensating cerebrospinal fluid in surface-based forward models of magneto- and electroencephalography. *bioRxiv preprint* (2016). doi = http://dx.doi.org/10.1101/037788

[21] Rush, S., Driscoll, D. Current distribution in the brain from surface electrodes. *Anesthesia & Analgesia* **47(6)**, 717–723 (1968).

[22] Cohen, D., Cuffin, B. N.Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalography and Clinical Neurophysiology* **56(1)**, 38–51 (1983). doi:10.1016/0013-4694(83)90005-6

[23] Gonçalves, S., de Munck. J,, Verbunt. J., Bijma, F., Heethaar, R., Lopes da Silva, F. In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head. *IEEE Transactions on Biomedical Engineering* **50**, 754–767 (2003).

[24] Akalin Acar, Z., Makeig, S. Effects of forward model errors on EEG source localization. *Brain topography* **26**, 378-396 (2013).

[25] Akalin Acar, Z., Acar, C. E., Makeig, S. Simultaneous head tissue conductivity and EEG source location estimation. *NeuroImage* **124**, 168–180 (2016). doi:10.1016/j.neuroimage.2015.08.032

[26] Henson, R.N., Mattout, J., Phillips, C., Friston, K.J. Selecting forward models for MEG source-reconstruction using model-evidence. *Neuroimage* **46**, 168-176 (2009).

[27] Strobbe, G., van Mierlo, P., De Vos, M., Mijović, B., Hallez, H., Van Huffel, S., Lopez, J, Vandenberghe, S. Bayesian model selection of template forward models for EEG source reconstruction. *NeuroImage* **93**, 11–22 (2014). doi:10.1016/j.neuroimage.2014.02.022

[28] Stahlhut, C., Mørup, M., Winther, O., Hansen, L.K. Simultaneous EEG source and forward model reconstruction (sofomore) using a hierarchical bayesian approach. *Journal of Signal Processing Systems* **65**, 431-444 (2011).

[29] Valdés-Hernández, P. A., von Ellenrieder, N., Ojeda-Gonzalez, A., Kochen, S., Alemán-Gómez, Y., Muravchik, C., Valdés–Sosa, P. A. Approximate average head models for EEG source imaging. *Journal of Neuroscience Methods* **185(1)**, 125–132 (2009). doi:10.1016/j.jneumeth.2009.09.005

[30] Valdés-Hernández, P. A., Trujillo-Barreto, N., Valdes-Sosa, P. A. Fast Electrical Source Imaging without the subject's MRI: Bayesian Modal Averaging across heads. Abstract in the conference proceedings of the *21st Annual Meeting of the Organization for Human Brain Mapping* (2015).

[31] López, J. D., Troebinger, L., Penny, W., Espinosa, J. J., Barnes, G. R. Cortical surface reconstruction based on MEG data and spherical harmonics. *In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 6449–6452 (2013). doi:10.1109/EMBC.2013.6611031

[32] Stevenson, C., Brookes, M., López, J. D., Troebinger, L., Mattout, J., Penny, W., ... Barnes, G. Does function fit structure?

A ground truth for non-invasive neuroimaging. *NeuroImage* **94**, 89–95 (2014). doi:10.1016/j.neuroimage.2014.02.033

[33] Kappen, H. The Variational Garrote. *arXiv Preprint arXiv*1109.0486 (2011). Retrieved from http://arxiv.org/abs/1109.0486

[34] Kappen, H.J., Gómez, V. The Variational Garrote. *Machine Learning*, 1-16 (2013).

[35] Hansen, S.T. and Stahlhut, C., Hansen, L.K. Sparse Source EEG Imaging with the Variational Garrote. *Pattern Recognition in Neuroimaging (PRNI), IEEE 2013 International Workshop on*, 106-109 (2013)

[36] Hansen, S.T., Stahlhut, C., Hansen, L.K. Expansion of the Variational Garrote to a Multiple Measurement Vectors Model. *Twelfth Scandinavian Conference on Artificial Intelligence* **257**, 105-114 (2013).

[37] Jolliffe, I. *Principal component analysis*. Springer series in statistics (2002).

[38] Hansen, L.K., Larsen, J., Nielsen, F.AA., Strother, S.C., Rostrup, E., Savoy, R., Lange, N., Sidtis, J.J., Svarer, C., Paulson, O.B. Generalizable patterns in neuroimaging: How many principal components? *NeuroImage* **9**, 534-544 (1999).

[39] Henson, R.N., Wakeman, D.G., Litvak, V., Friston, K.J. A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration. *Frontiers in human neuroscience* **5**, 1-16 (2011).

[40] Wakeman, D. G., Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data* **2**, 150001. doi:10.1038/sdata.2015.1

[41] Henson, R.N., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., Rugg, M.D. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral cortex* **13**, 793-805 (2003).

[42] Friston, K.J., Harrison, L., Daunizeau, J., Kiebel, S.J., Phillips, C., Trujillo-Barreto, N., Henson, R.N., Flandin, G., Mattout, J. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage* **39**, 1104-1120 (2008).

[43] Henson, R.N., Mouchlianitis, E., Friston, K.J. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage* **47**, 581-589 (2009).

[44] Mattout, J., Henson, R. N., Friston, K. J. (2007). Canonical source reconstruction for MEG. *Computational Intelligence and Neuroscience*, 67613 (2007). doi:10.1155/2007/67613

[45] Phillips, C. Source estimation in EEG. *University de Liege, Belgium* (2000).

[46] Vorwerk, J., Cho, J.-H., Rampp, S., Hamer, H., Knösche, T. R., Wolters, C. H. A guideline for head volume conductor modeling in EEG and MEG. *NeuroImage* **100**, 590–607 (2014). doi:10.1016/j.neuroimage.2014.06.040

[47] Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W. Variational free energy and the Laplace approximation. *NeuroImage* **34(1)**, 220–234 (2007). doi:10.1016/j.neuroimage.2006.08.035

[48] López, J. D., Penny, W. D., Espinosa, J. J., Barnes, G. R. A general Bayesian treatment for MEG source reconstruction incorporating lead field uncertainty. *NeuroImage* **60(2)**, 1194–1204 (2012). doi:10.1016/j.neuroimage.2012.01.077

[49] Toussaint, G.T. Bibliography on estimation of misclassification. *IEEE Transactions on information Theory* **20**, 472-479 (1974).

[50] Nadaraya, E.A. On Estimating Regression. *Theory of Probability and its Applications* **9**, 141–142 (1964).

[51] Donoho, D. L., Elad, M., Temlyakov, V. N. Stable recovery of sparse overcomplete representations in the presence of noise. *Transactions on Information Theory, IEEE* **52(1)**, 6–18 (2006). doi:10.1109/TIT.2005.860430

[52] Belsley, D. A., Kuh, E., Welsch, R. E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. *John Wiley & Sons* (2005).

[53] Rijsbergen, C. J. Van. (1979). Information Retrieval (2nd ed.). *Butterworth-Heinemann*.

[54] Kanwisher, N., Yovel, G. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **361(1476)**, 2109–2128 (2006). doi:10.1098/rstb.2006.1934

[55] Lew, S., Wolters, C., Anwander, A., Makeig, S., MacLeod, R. Low resolution conductivity estimation to improve source localization. *International Congress Series* **1300**, 149-152 (2007).

[56] Plis, S.M., George, J.S., Jun, S.C., Ranken, D.M., Volegov, P.L., Schmidt, D. Probabilistic forward model for electroencephalography source analysis. *Physics in medicine and biology* **52**, 5309-5327 (2007).

[57] Windhoff, M., Opitz, A., Thielscher, A. Electric field calculations in brain stimulation based on finite elements: an optimized processing pipeline for the generation and usage of accurate individual head models. *Human brain mapping* **34**, 923-935 (2013).

[58] Gençer, N.G., Acar, C.E. Sensitivity of EEG and MEG measurements to tissue conductivity. *Physics in medicine and biology* **49**, 701 (2004).

[59] Stahlhut, C., Attias, H.T., Stopczynski, A., Petersen, M.K., Larsen, J.E., Hansen, L.K. An evaluation of EEG scanner's dependence on the imaging technique, forward model computation method, and array dimensionality. *34th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, 1538-1541 (2012).

[60] López, J. D., Litvak, V., Espinosa, J. J., Friston, K., Barnes, G. R. Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. *NeuroImage* **84**, 476–487 (2013). doi:10.1016/j.neuroimage.2013.09.002

[61] Rice, J.K., Rorden, C., Little, J.S., Parra, L.C. Subject position affects EEG magnitudes. *NeuroImage* **64**, 476-484 (2013).

[62] Martinez-Cantin, R. BayesOpt: A Bayesian optimization library for nonlinear optimization, experimental design and bandits. *The Journal of Machine Learning Research* **15(1)**, 3735–3739 (2014). Retrieved from http://dl.acm.org/citation.cfm?id=2627435.2750364

[63] Keator, D.B., Grethe, J. S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., ..., Papadopoulos, P. A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). *IEEE Transactions on Information Technology in Biomedicine* **12**, 162-172 (2008).

[64] Dale, A.M., Sereno, M.I. Improved Localizadon of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience* **5(2)**, 162–176 (1993). doi:10.1162/jocn.1993.5.2.162

[65] Ishwaran, Hemant, J. Sunil Rao. Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics*: 730-773 (2005).

[66] Hansen, S. T., Hansen, L. K. EEG source reconstruction performance as a function of skull conductance contrast. *In Acoustics, Speech and Signal Processing, IEEE International Conference on (ICASSP)* (2015).

Postprint. Final version: `http://www.sciencedirect.com/science/article/pii/S105381191630252X`

Appendix A. The Variational Garrote

As also described earlier, there exists a linear relationship between the cortical sources and scalp EEG [64]. The mathematical relation is given by the forward model $\mathbf{A} \in \mathbb{R}^{K \times N}$ which maps N dipolar sources ($\mathbf{X}$) to K EEG electrode signals ($\mathbf{Y}$) in T time samples, i.e.,

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \epsilon, \tag{A.1}$$

where ϵ is noise.

We chose to perform source reconstruction using a modified version of a sparsity-inducing Bayesian-inference scheme; the Variational Garrote (VG) [33, 34]. VG has been adapted to solve the EEG inverse problem in a multiple-measurement vectors framework, called time-expanded VG (teVG) [36]. The teVG (and VG) enforces a 'spike-and-slab'-like representation [65] by including a binary variable for each potential source, thus encoding whether the source is active or not. The problem to solve is now

$$Y_{kt} = \sum_{n=1}^{N} A_{kn} s_n X_{nt} + \epsilon_{nt}, \quad \epsilon_{nt} \sim N(0, \beta^{-1}), \tag{A.2}$$

where $s_n \in \{0, 1\}$ which is assigned the prior $p(\mathbf{s}|\gamma) = \prod_{n=1}^{N} p(s_n|\gamma)$ where $p(s_n|\gamma) = \dfrac{\exp(\gamma s_n)}{1 + \exp(\gamma)}$ [34]. The hyperparameter γ controls the sparsity level. Note that making $\mathbf{s}$ independent of time samples corresponds to an assumed fixed support for all time samples. The solution scheme proposed by Kappen et al. [34] is based on Bayesian inference by maximizing the following posterior probability

$$p(\mathbf{s}, \mathbf{X}, \beta|\mathbf{D}, \gamma) \propto p(\mathbf{s}|\gamma) p(\mathbf{D}|\mathbf{s}, \mathbf{X}, \beta), \tag{A.3}$$

where $\mathbf{D} = \{\mathbf{A}, \mathbf{Y}\}$ and $p(\mathbf{X}, \beta)$ is assumed flat. The solution is non-trivial and Kappen et al. suggest marginalizing over $\mathbf{s}$ and employing a variational approximation. When taking the logarithm

$$\log \sum_{\mathbf{s}} p(\mathbf{s}|\gamma) p(D|\mathbf{s}, \mathbf{X}, \beta) = \log \sum_{\mathbf{s}} \frac{q(\mathbf{s})}{q(\mathbf{s})} p(\mathbf{s}|\gamma) p(D|\mathbf{s}, \mathbf{X}, \beta) \tag{A.4}$$

and using Jensen's inequality, a bound on the approximation is recovered (reproduced from [34])

$$\log \sum_{\mathbf{s}} \frac{q(\mathbf{s})}{q(\mathbf{s})} p(\mathbf{s}|\gamma) p(D|\mathbf{s}, \mathbf{X}, \beta) \geq - \sum_{\mathbf{s}} q(\mathbf{s}) \log \frac{q(\mathbf{s})}{p(\mathbf{s}|\gamma) p(D|\mathbf{s}, \mathbf{X}, \beta)} = -F(q, \mathbf{X}, \beta), \tag{A.5}$$

where $F(q, \mathbf{X}, \beta)$ is the variational free energy. The variational approximation is defined as $q(\mathbf{s}) = \prod_{n=1}^{N} q_n(s_n)$, here $q_n(s_n) = m_n s_n + (1 - m_n)(1 - s_n)$ [34]. The parameter m_n is the variational mean and can be interpreted as the probability of s_n being active, and therefore has values between 0 and 1. In order to obtain a tight bound $-F(q, \mathbf{X}, \beta)$ should be maximized or equivalently $F(q, \mathbf{X}, \beta)$ minimized. Posed in a 'dual formulation' the free energy is

$$\begin{aligned} F(\mathbf{m}, \mathbf{X}, \beta, \mathbf{Z}, \lambda) = & -\frac{TK}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t=1}^{T} \sum_{k=1}^{K} (Z_{kt} - Y_{kt})^2 \\ & + \frac{K\beta}{2} \sum_{t=1}^{T} \sum_{n=1}^{N} m_n (1 - m_n) X_{nt}^2 \chi_{nn} - \gamma \sum_{n=1}^{N} m_n + N \log(1 + \exp(\gamma)) \\ & + \sum_{n=1}^{N} \left(m_n \log(m_n) + (1 - m_n) \log(1 - m_n) \right) + \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_{kt} \left(Z_{kt} - \sum_{n=1}^{N} m_n X_{nt} A_{kn} \right). \end{aligned} \tag{A.6}$$

Here, $\boldsymbol{\chi}$ is the covariance matrix of the forward model $\mathbf{A}$. The terms $\boldsymbol{\lambda}$ (Lagrange multipliers) and $Z_{kt} = \sum_{n=1}^{N} m_n X_{nt} A_{kn}$ both stem from the dual formulation. Finally γ controls the sparsity level and is found through cross-validation [34]. The remaining parameters are found by equating the partial derivatives with zero and solving the resulting equation

set. The free energy is seen to be determined by the data fit between the observations and their expected values, and the model priors. The free energy thus considers both the proposed forward model and the source distribution.

Originally, the solution was obtained through fixed-point iterations, which had a computational complexity scaling quadratically with the number of electrodes and linearly with the number of sources; thus computation time was relatively low [34]. Parameter updating was further improved by using gradient descent [66]; we adopted the same scheme in this work. MATLAB code implementing teVG is available at `https://github.com/STherese/VG_inverse_solvers`.

Postprint. Final version: `http://www.sciencedirect.com/science/article/pii/S105381191630252X`