# EEG Based Inference of Spatio-Temporal Brain Dynamics

**Hansen, Sofie Therese**

*Publication date:*
2017

*Document Version*
Publisher's PDF, also known as Version of record

# EEG Based Inference of Spatio-Temporal Brain Dynamics

Sofie Therese Hansen

Kongens Lyngby 2016
PHD-2016-410

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk
PHD: ISSN 0909-3192

# Summary (English)

Electroencephalography (EEG) provides a measure of brain activity and has improved our understanding of the brain immensely. However, there is still much to be learned and the full potential of EEG is yet to be realized. In this thesis we suggest to improve the information gain of EEG using three different approaches; 1) by recovery of the EEG sources, 2) by representing and inferring the propagation path of EEG sources, and 3) by combining EEG with functional magnetic resonance imaging (fMRI). The common goal of the methods, and thus of this thesis, is to improve the spatial dimension of EEG.

The main topic of this thesis is the localization of the EEG generators. This entails solving both a forward and an inverse problem. The inverse problem maps the EEG signal recorded on the scalp to its origin in the brain. It is a highly ill-posed problem which we tackle by employing a sparsity promoting 'spike and slab' like method augmented with physiologically relevant source priors. The incorporated temporal and spatial priors exploit coherence between neighboring time samples and between neighboring source locations, respectively. We show that these augmentations effectively increase the source recovery ability.

The forward problem describes the propagation of neuronal activity in the brain to the EEG electrodes on the scalp. The geometry and conductivity of the head layers are normally required to model this path. We propose a framework for inferring forward models which is based on the EEG signal and a low dimensional representation of forward models. The representation is built by principal component analysis of a corpus of forward models. The method can be used to recover subject-specific forward models when structural scans and/or conductivity estimations are not available.

Finally, we investigate the extraction of EEG components having bandpower dynamics correlated with fMRI components. We show that adding anatomical information to the inference scheme improves the recovery of correlated components compared to only using functional information. The anatomical information is incorporated through the EEG forward model and assumes that the activity of the fMRI component overlaps spatially with the origin of the coupled EEG component.

# Summary (Danish)

Elektroencefalografi (EEG) er et mål for hjerneaktivitet og har været med til at forbedre vores forståelse af hjernen. Meget er dog stadig uvist, og det fulde potentiale af EEG er endnu ikke realiseret. I denne afhandling foreslår vi at forbedre informationsudbyttet fra EEG ved tre fremgangsmåder; 1) ved rekonstruktion af EEG kilderne, 2) ved repræsentation og inferens af EEG kildernes propagering gennem hovedet, og 3) ved at kombinere EEG og *funktionel magnetisk resonans-billeddannelse* (fMRI). Det fælles mål for metoderne, og således også for denne afhandling, er at forbedre den rumlige opløsning af EEG.

Hovedemnet for denne afhandling er lokalisering af EEG generatorerne, hvilket indbefatter at løse både et *forward* problem og et inverst problem. Det inverse problem beskriver estimeringen af EEG aktivitetens oprindelse i hjernen. Det er et meget dårligt specificeret problem, hvilket vi håndtere ved at bruge priors, der fremmer *sparsity* ved hjælp af en *spike and slab*-lignende repræsentation, samt via relevante fysiologiske priors. De inkorporerede temporale og rumlige priors udnytter henholdsvis afhængighed mellem nabotidspunkter og afhængighed mellem nabokildeplaceringer. Vi viser at disse augmentationer forbedrer kilderekonstruktionen.

Forward problemet beskriver propageringen af neuronal aktivitet i hjernen til EEG elektroderne på hovedbunden. Hovedgeometrien og -konduktiviteten er normalt påkrævet for at kunne modellere denne sti. Vi foreslår en alternativ fremgangsmåde, hvor nye forward modeller udledes på baggrund af EEG signalet samt en lav-dimensional repræsentation af forward modeller. Repræsentationen er konstrueret ved hjælp af principal komponent analyse af en samling af forward modeller. Metoden kan således bruges til at finde person-specifikke

forward modeller, når personens hovedgeometri og konduktivitet ikke er tilgængelige.

Endelig beskriver vi en metode til at finde koblede EEG og fMRI komponenter. Koblingen søger efter EEG komponenter hvis energi er korreleret med dynamikken i fMRI komponenterne. Vi viser at evnen til at finde korrelerede komponenter øges ved at tilføje anatomisk information sammenlignet med kun at lede efter en funktionel kobling. Den anatomiske information inkorporeres via EEG forward modellen, og derved antages det, at EEG komponenten er opstået samme sted i hjernen som fMRI komponenten.

# Preface

This thesis was prepared at DTU Compute in the section of Cognitive Systems in fulfillment of the requirements for acquiring a PhD degree in computer science.

The thesis includes of a summary report of the studied theory, proposed methods, key findings and perspectives on these. The thesis, furthermore, contains eight papers of which six are published and two are submitted for publication (at the time of writing). The thesis work was performed in the period March 15 2013 to March 14 2016.

Lyngby, 14-March-2016

Sofie Therese Hansen

# Acknowledgements

The guidance of my supervisor, Lars Kai Hansen, and my initial co-supervisor, Carsten Stahlhut, has been essential for this thesis. Our meetings have been motivating, inspiring, challenging and encouraging.

I would also like to thank the CogSys group, including fellow PhD students and co-workers for inspiring collaborations, interesting reading groups and general support throughout my PhD.

I am grateful for the Machine Learning group at the Technical University of Berlin for hosting me in the winter of 2014/2015. I would especially like to thank Klaus-Robert Müeller, Sven Dähne, Irene Winkler, and Daniel Miklody for fruitful discussions and collaboration and thanks also goes to the rest of the group for their heartfelt welcoming.

I am thankful for the financial support I received for my external stay in Berlin from the Stibo Foundation's IT travel scholarship, IDA and Berg-Nielsen's Study and Support Fund and the Trane Foundation. I additionally wish to thank the Otto Mønsted Foundation and the Thomas B. Thrige Foundation for supporting me in conference participations.

Finally I would like to thank my family and friends for continuous support. Special thanks go to Martin Axelsen for proof reading and being encouraging for all three years of my PhD.

# Contributions

## Included Peer-Reviewed Contributions

[A] S.T. Hansen, C. Stahlhut, L.K. Hansen (2013). Sparse Source EEG Imaging with the Variational Garrote. *Proceedings of the 3rd International Workshop on Pattern Recognition in NeuroImaging (PRNI), IEEE*, pp. 106-109.

[B] S.T. Hansen, C. Stahlhut, L.K. Hansen (2013). Expansion of the Variational Garrote to a Multiple Measurement Vectors Model. *Twelfth Scandinavian Conference on Artificial Intelligence (SCAI), ed./M. Jaeger. IOS Press*, pp. 105-114.

[C] S.T. Hansen, L.K. Hansen (2015). EEG Source Reconstruction Performance as a Function of Skull Conductance Contrast. *Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 827-831.

[D] S.T. Hansen, L.K. Hansen (2014). EEG Source Reconstruction using Sparse Basis Function Representations. *Proceedings of 4th International Workshop on Pattern Recognition in Neuroimaging (PRNI), IEEE.*

[E] S.T. Hansen, L.K. Hansen (2013). EEG Sequence Imaging: A Markov Prior for the Variational Garrote. Planned in *the proceedings of the 3rd NIPS 2013 Workshop on Machine Learning and Interpretation in NeuroImaging (MLINI).*

[F] S.T. Hansen, I. Winkler, L.K. Hansen, K.-R. Müller, S. Dähne (2015). Fusing Simultaneous EEG and fMRI Using Functional and Anatomical

Information. *Proceedings of the 2015 International Workshop on Pattern Recognition in NeuroImaging (PRNI), IEEE*, pp. 33-36.

[G] S.T. Hansen, L.K. Hansen. Spatio-Temporal Reconstruction of Brain Dynamics from EEG with a Markov Prior (2017). *NeuroImage*, vol. 148, pp. 274–283.

[H] S.T. Hansen, S. Hauberg, L.K. Hansen (2016). Data Driven Forward Model Inference for EEG Brain Imaging. *NeuroImage*, vol 139, pp. 249-258.

## Additional Contributions

- S.T. Hansen, L.K. Hansen (2015). Space-time Smooth EEG Source Reconstruction. Abstract at the *21st Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, and *International Conference on Basic and Clinical Multimodal Imaging (BACI)*
- S.T. Hansen, L.K. Hansen. Sparse Source Reconstruction with Fixed Temporal Support and Basis Functions. Poster at the 17th International Conference on AISTATS/MLSS.
- M.R. Andersen, S.T. Hansen, L.K. Hansen (2013). Learning the solution sparsity of an ill-posed linear inverse problem with the Variational Garrote, *International Workshop on Machine Learning for Signal Processing (MLSP), IEEE*, pp. 1-6.
- S.T. Hansen, S. Hauberg, L.K. Hansen (2016). Forward Models can be Inferred from EEG Data. Abstract accepted for the *22nd Annual Meeting of the Organization for Human Brain Mapping*. Available at `https://github.com/STherese/Preprints/blob/master/OHBM2016_preprint.pdf`.

## Software

- Implementations of the Variational Garrote and derived versions; teVG and MarkoVG. `https://github.com/STherese/VG_inverse_solvers`.

# Nomenclature

## Abbreviations and acronyms

| | |
|---|---|
| ARD | Automatic relevance determination |
| BCI | Brain computer interface |
| BEM | Boundary element method |
| BOLD | Blood oxygen level dependent signal |
| CCA | Canonical correlation analysis. |
| CT | Computed tomography |
| DTI | Diffusion tensor imaging |
| EEG | Electroencephalography |
| EIT | Electromagnetic impedance tomography |
| EP | Evoked potential |
| ERP | Event related potential |
| FDM | Finite difference method |
| FEM | Finite element method |
| FFA | Fusiform face area |
| fMRI | Functional magnetic resonance imaging |
| FOCUSS | Focal underdetermined system solver |
| GD | Gradient descent |
| Hrf | Hemodynamic response function |
| ICA | Independent component analysis |
| MarkoVG | VG with a Markov prior |
| MAP | Maximum a posteriori |
| MEG | Magnetoencephalography |
| M-FOCUSS | MMV extended focal underdetermined system solver |

| | |
|---|---|
| MMV | Multiple measurement vector |
| M-SBL | Multiple response sparse Bayesian learning |
| MSE | Mean squared error |
| MSP | Multiple source priors |
| mSPoC | Multimodal source power comodulation |
| OFA | Occipital face area |
| PCA | Principal component analysis |
| PMF | Paired mean field |
| SBL | Sparse Bayesian learning |
| S.d. | Standard deviation |
| S.e.m. | Standard error of the mean |
| SNR | Signal-to-noise ratio |
| SVD | Singular value decomposition |
| teVG | Time-expanded variational garrote |
| T-MSBL | Temporal correlation extension to M-SBL |
| VG | Variational garrote |

# Notation and Fixed Symbols

Matrices are defined in capital and bold, vectors lower case and bold, and finally scalars are defined in normal font.

| | |
|---|---|
| $B_{ij}$ | The matrix element of $\mathbf{B}$ in row $i$ and column $j$. |
| $\mathbf{B}^\top$ | The transpose of matrix $\mathbf{B}$. |
| $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. |
| $\nabla$ | The gradient. |
| $\sigma_f(b) = (1 + \exp(-b))^{-1}$ | The sigmoid function of $b$. |
| $F_1\text{-measure} = \dfrac{2 \cdot \mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{P}}$ | The source retrieval index, where TP, FP and P are true, false and actual positives. |
| $\mathbf{A}$ | The forward model. |
| $\mathbf{Y}$ | The data matrix. |
| $\mathbf{X}$ | The source matrix. |
| $\mathbf{E}$ | The noise with mean zero and variance $1/\beta$. |
| $K$ | The number of observations in $\mathbf{Y}$. |
| $N$ | The number of sources in $\mathbf{X}$. |

# Contents

Chapter 1

# Introduction and Motivation

Interest in understanding how the brain functions is the main motivation of this thesis. Although the general anatomy of the brain has been described in detail, it is partly unknown how the structure of brain networks translates into complicated human emotions, intelligence and perception. Identifying and alleviating the causes of neurological diseases such as schizophrenia and Parkinson's disease is also an important prospect of understanding the brain more comprehensively.

The neuronal activation has a temporal scale in the order of milliseconds and since electroencephalography (EEG) is usually recorded with a sampling frequency of around 1,000 Hz detailed temporal information can be acquired with this neuroimaging tool [Nunez and Srinivasan, 2006a]. EEG is consequently crucial in understanding the timing and temporal dynamics of brain processes. However, in order for EEG to provide also detailed spatial information, further higher order analysis must be performed. In this thesis detailed spatio-temporal information is obtained by localizing the EEG sources, i.e., the EEG signal measured at the scalp is mapped to the higher dimensional cortex space. This mapping requires an inverse solver with relevant source priors and a 'forward model' describing the reverse mapping, i.e. from cortex to scalp. We propose to infer the forward model of a subject using their EEG data combined with a prior over forward models, thus facilitating a highly accessible subject-specific mapping. We finally explore the combination of the temporally rich EEG and the spatially rich functional magnetic resonance imaging (fMRI).

Before describing the contribution of this thesis in greater detail Chapter 2 presents fundamental theory on the origins of EEG and fMRI. Chapter 2 also provides a description of EEG source localization including how EEG sources propagate to the EEG electrodes and how the inverse mapping can be approximated. Chapters 3 - 5 contain the contributions of this thesis and are followed by a conclusion containing perspectives of the presented work.

## 1.1 Contributions

The common goal of this thesis is to extract detailed information from EEG. This is approached from three angles:

**Chapter 3**: *Source localization.* The Bayesian sparsifying method, 'the Variational Garrote' is adapted to solve the EEG inverse problem using relevant modeling assumptions. Both spatial and temporal source priors are implemented to provide precise EEG based imaging. The spatial source distribution is modeled as being sparse with smooth and compact source components. The temporal dynamics of the EEG is modeled assuming coherence between neighboring time samples. The presented work is contained in contributions [A - E & G].

**Chapter 4**: *Forward models.* The conventional approach of mapping brain activity to EEG electrodes relies on detailed information about the head layers. This information is at best difficult to obtain and we therefore propose a data driven approach to estimate the forward model. A forward model for a new subject is inferred based on a parametrization of forward models and the subject's recorded EEG data. The inference procedure optimizes the 'model evidence' and provides the EEG sources in addition to a forward model. The presented work is contained in contribution [H].

**Chapter 5**: *Combining EEG with fMRI.* Through multimodal imaging multiple dimensions of brain activity can be considered and integrated. The multimodal Source Power Co-modulation method (mSPoC) actively searches for correlated components of the hemodynamic and electrophysical response to brain activation. The extraction of coupled components is based on functional correlation between, e.g., fMRI and power dynamics of EEG components. In this thesis we extend mSPoC to include anatomical information which enters the algorithm through the forward model. The presented work is contained in contribution [F].

Chapter 2

# Preliminaries

The following sections present fundamental theory of the thesis. The focus is therefore on the origin of EEG signals (section 2.1) and how the generators of EEG are localized (section 2.3). The hemodynamic response to brain activity is also briefly described to further clarify the benefits and drawbacks of EEG as well as to provide contextual background for the thesis study of correlating EEG and fMRI (section 2.2).

## 2.1 Electrophysics of the Brain

Electrical signals of the brain are in general easily collected and EEG was therefore one of the earliest techniques to acquire brain activity [Nisar and Yeap, 2014]. However, while human EEG has existed for a century we are still learning how to interpret the signal. A general consensus on what the EEG generators are has been established [Nunez and Srinivasan, 2006a] but we are still acquiring knowledge on how we can exploit the EEG to interpret how the brain functions in normal and pathological states.

The electrical activity that produces signal measurable at the scalp comes from the communication between neurons. To transmit information a neuron polar-

izes or depolarizes a receiving neuron through inhibitory or excitatory postsynaptic potentials (I/EPSPs). These synaptic inputs create current flows internally in the neuron as well as externally which in turn gives rise to an extracellular potential field. Synchronous activation of assemblies of neurons generates extracellular potentials that are summed into a signal detectable by scalp EEG electrodes. Summing of potentials is possible when the neurons' dendrites, where the PSPs take place, are aligned. As cortical pyramidal cells fulfill this condition they are believed to be EEG generators [Nunez and Srinivasan, 2006a].

The activity of neuronal assemblies is often described as being oscillatory; having a frequency, amplitude and phase [Nisar and Yeap, 2014]. These properties can for example be used to characterize EEG recordings of subjects during sleep [Borbély et al., 1981; Dang-Vu et al., 2008] or rest [Musso et al., 2010]. EEG occurring without stimuli is termed spontaneous EEG. In contrast, stimulus-locked EEG describes the neuronal response to an external stimulus, e.g., visual, auditory or sensory [Nunez and Srinivasan, 2006a]. The EEG response to a stimulus is often denoised by repeating the stimulus many times and performing stimulus-locked averaging [Baillet et al., 2001]. These responses are termed evoked potentials (EPs) or event-related potentials (ERPs). An ERP (or EP) has a stimulus dependent waveform which can be modulated by features of the subject, such as age [Polich, 1997] and disease history [McNeely et al., 2008], or by external features, such as preconditioning tasks [Schweinberger et al., 1995] and drug effects [Kähkönen et al., 2001].

A so-called 'differential ERP' can be used to contrast two different stimuli. It is computed as the averaged response to one set of stimuli minus the averaged response to another set of stimuli, e.g., a control task. The EEG response to seeing faces compared to scrambled faces is an example of a highly studied differential ERP signal. The ERP is created by showing the subject images of faces and scrambled faces [Henson et al., 2003]. Averaging of usually hundreds of trials in each condition is followed by a contrast subtraction of the two computed averages. The differential ERP is created in an effort to reveal the dynamics specific to a certain stimulus. In the case of the face-evoked response a larger negative peak occurs approximately 170 ms after stimulus onset [Bentin et al., 1996; Henson et al., 2003]. The peak is due to its timing called the N170 component. The face-evoked response is localized primarily to the occipital and fusiform face areas (O/FFAs) by EEG/MEG [Henson et al., 2009b] and fMRI [Kanwisher et al., 1997].

The many applications of EEG exploit the high temporal information it offers. Examples, besides the already mentioned, include brain computer interfaces (BCIs), e.g., for providing a communication pathway to locked-in patients [Höhne et al., 2011]. Neurofeedback is another example which facilitates (near) real time presentation of the current brain state [Mullen et al., 2015] and can

be used for entraining brain activity, e.g., with the purpose of decreasing symptoms of movement disorders [Thompson and Thompson, 2008]. EEG is also well suited for monitoring patients [Claassen et al., 2004] because of the high temporal resolution and its relatively portable acquisition system [Modarreszadeh and Schmidt, 1997]. EEG can therefore be measured without too much discomfort of the patient. Finally, outside the clinic EEG can facilitate brain scanning in more natural environments compared to, e.g., fMRI and MEG [Petersen et al., 2011; Stopczynski et al., 2014; Mullen et al., 2015].

One of the main challenges of EEG is its low spatial resolution. Section 2.3 describes how better spatial resolution can be obtained by source localizing the EEG generators. However, first a neuroimaging technique that has the inverse resolution properties of EEG is briefly discussed.

## 2.2 Hemodynamics of the Brain

The dynamics of cerebral blood flow (CBF) is another quantifying measure of brain activity [Ogawa et al., 1993]. The mechanism is as follows: When neurons become activated a higher demand of oxygen arises. The oxygen depletion signals the nearby blood vessels to dilate and thereby increase the blood flow to this area. The oxygen transporter, hemoglobin, has different magnetic properties depending on whether it is oxygenated or deoxygenated [Ogawa et al., 1993]. Since fMRI is sensitive to these magnetic properties it can be used to measure a blood oxygen level dependent (BOLD) signal [Stippich, 2007]. In response to increased neuronal activation the BOLD will increase due to a higher blood flow and peak after approximately 5 seconds [Flandin and Novak, 2013]. The BOLD will then decrease, often undershooting the baseline. In the beginning of the hemodynamic response there is a short initial dip in blood oxygenation, as shown in Fig. 2.1. In order to interpret the BOLD signal a baseline (or control) is needed to show whether a specific stimulus yields an increase or decrease in BOLD.

fMRI is termed an indirect measure of brain activity as it quantifies how the cerebral hemodynamics change in response to neuronal activity [Flandin and Novak, 2013]. Furthermore, while the spatial resolution is quite high in fMRI, down to millimeters [Baillet et al., 2001], the temporal resolution is limited by both the acquisition technique and the inherent resolution of the BOLD signal. The fast neuronal changes that happen on the millisecond-scale will therefore not be captured by hemodynamic measures which have a temporal resolution in the second-scale [Mather et al., 2013].

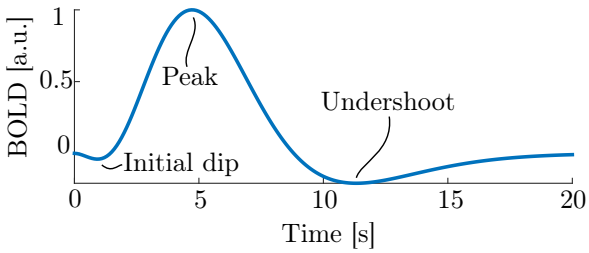



**Figure 2.1:** Example of a hemodynamic response function (hrf) with neuronal activation at time = 0 s.

The high spatial resolution of fMRI has been exploited in an asymmetrical integration with EEG where the fMRI activated areas act as location priors of the reconstructed EEG signal [Henson et al., 2011]. The exact relation between neuronal activity and hemodynamics, i.e., the neurovascular coupling, is however not fully understood. Several studies have, however, been able to correlate BOLD signal with EEG activity, see references in [Dähne et al., 2015]. There is thus growing evidence that fMRI and EEG can provide complementary information of the brain activity and that they combined can enhance our understanding of brain functioning and dysfunctioning.

## 2.3 EEG Source Localization

What EEG provides in temporal information it lacks in spatial resolution. The low resolution is partly caused by the low number of measurement points [Baillet et al., 2001], which usually ranges from 14 [Stopczynski et al., 2014] to 256 [Akalin Acar and Makeig, 2013]. But most importantly it is produced by spatial smearing of the signal occurring when the neuronal activity propagates from the brain to the electrodes on the scalp. Spatial smearing is primarily caused by the low conductivity of the skull as compared to the brain and scalp [Nunez and Srinivasan, 2006b]. According to Wolters et al. tissue anisotropy of the skull and white matter also greatly affect the forward fields [Wolters et al., 2006].

EEG source imaging has an unmixing effect of the EEG signal and thus reduces the blurring caused by volume conduction [Delorme et al., 2012]. In the framework of BCI Besserve et al. showed that decoding is enhanced when the EEG is source localized [Besserve et al., 2011]. Edelman et al. further showed that the separability of similar hand movements increases when based on source localized activity [Edelman et al., 2016]. Finally in [Ahn et al., 2012] it was shown that source localization can provide information not observable in scalp EEG.

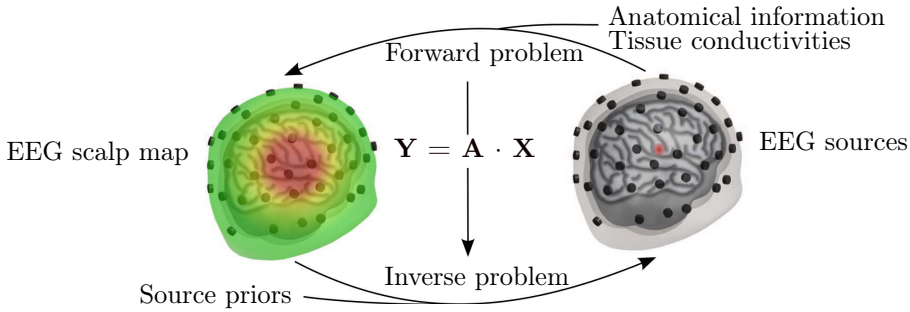



**Figure 2.2:** The relation between the forward and inverse problems. The forward problem requires prior knowledge of the head tissues, and the inverse problem is dependent on assumptions of the source distribution as well as on the accuracy of the constructed forward model.

The range of frequencies relevant for EEG means that the propagation of the electric fields is negligible [Hallez et al., 2007]. The quasi-static conditions are therefore met and the relation between the electrodes $\mathbf{Y}$ and the brain sources $\mathbf{X}$ is linear, illustrated in Fig. 2.2. The relation can thus be posed as a linear regression problem; $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A}$ is the forward model. The following sections explain the process of estimating the forward model and the inverse solution.

### 2.3.1 The Forward Problem

A forward model must be available in order to estimate the inverse solution and the forward problem must therefore be solved as the first step towards source localization [Hallez et al., 2007]. The forward model describes how a neuronal source in the brain maps to the electrodes on the scalp, see Fig. 2.3. It therefore contains the geometry and the electrical properties of the head layers lying between the sources and the electrodes.

The geometry of the head compartments, e.g., scalp, skull and brain, can be estimated by segmenting head scans recorded by MRI and/or CT. Several open source software libraries provide automatic segmentation such as SPM[1] [Ashburner and Friston, 2005] and FreeSurfer[2] [Dale et al., 1999]. However, for a detailed segmentation manual correction is often needed, e.g., for correcting the skull thickness [Perdue and Diamond, 2014].

[1] `http://www.fil.ion.ucl.ac.uk/spm/`
[2] `https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki`

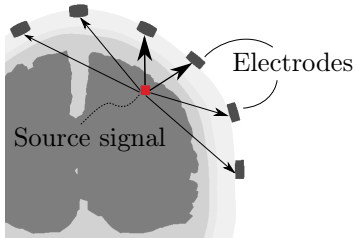



**Figure 2.3:** The forward problem. A forward model maps each source location in the brain to each electrode.

Compromises and simplifications are frequently made when specifying the forward model parameters. For example, the head is often modeled as containing only the scalp, skull and brain, and the soft tissues are assumed to have the same template conductivity [Wang and Ren, 2013]. When making these assumptions, only the skull:brain conductivity ratio remains to be estimated. This value is often set to 1:80 following the experimental findings reported in [Rush and Driscoll, 1968] and [Cohen and Cuffin, 1983]. However the skull:brain conductivity ratio is crucial for accurate source localization and is furthermore found to vary between persons [Awada et al., 1998]. In [Akalin Acar and Makeig, 2013] localization errors up to 30 mm were found when misspecifying the skull:brain conductivity ratio.

Electromagnetic impedance tomography (EIT) is an alternative to using template conductivity ratios. Currents are injected between sets of electrodes and the resulting potentials are measured [Baillet et al., 2001]. The conductivity of the head can then be approximated based on the applied currents and the measured potentials. However inferring the conductivities is an inverse problem as well as ill-posed [Baillet et al., 2001]. Diffusion tensor imaging (DTI) has also been suggested for the subject-specific estimation of conductivities [Tuch et al., 2001]. DTI measures the diffusion tensor of water which is related to the local conductivity, however the relation is not entirely understood [Plis et al., 2007]. Several other methods or experiments have been suggested for inferring the head conductivities but have provided different results, see overview in [Hallez et al., 2007]. In the case of missing or deficient conductivity measurements Plis et al. suggest to propagate the uncertainty of the skull conductivity to an uncertainty in the source localization estimates [Plis et al., 2007]. Lew et al. propose to estimate the skull conductivity from a defined set of values by simulated annealing where the objective function is the $l_2$-norm distance between the true and predicted EEG signals [Lew et al., 2007, 2009].

Realistic forward modeling is often divided into three types depending on how the head model is constructed, i.e., using boundary element method (BEM),

finite element method (FEM) or finite difference method (FDM). In the widely used BEM the surfaces of usually three to four head compartment are computed, with e.g. the brain tessellated into 8,196 vertices as a standard setting in SPM8. Each compartment is modeled as being isotropic and homogeneous [Hallez et al., 2007]. However as mentioned earlier the white matter of the brain and the skull have anisotropic conductivities and the skull is an inhomogeneous tissue [Baillet et al., 2001]. FEM and FDM are more flexible than BEM and can model these anisotropies. Instead of modeling only the boundaries as BEM does, FEM and FDM divide the head into many small *volumes* which within the same head compartments can have different conductivities.

The work presented in this thesis mainly takes advantage of the SPM8 toolbox where the more time consuming FEM and FDM models are not available. The BEM head model is therefore employed in this thesis.

### 2.3.2 The Inverse Problem

Establishing the anatomical origin of brain processes and functioning is an important application of EEG [Henderson et al., 1975]. While brain imaging based on EEG provides a direct measure of the electrical activity of the brain, the inverse problem must be solved for EEG to provide 'true' brain imaging. The inverse problem localizes the cortical sources that produces the electrical signal the electrodes measure, see Fig. 2.4.

Source localization methods are roughly divided into two categories; parametric and distributed (also called imaging) methods. Parametric methods fix the number of active dipoles and estimate their strengths and locations [Besserve et al., 2011], examples include equivalent dipole methods [Henderson et al., 1975; Scherg and Von Cramon, 1985]. Distributed methods fix the locations of a large number of dipoles and estimate their strengths, common examples

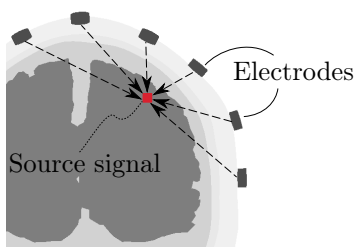



**Figure 2.4:** The inverse problem. All electrode potentials are used in approximating the EEG sources.

include methods employing $l_p$-norms [Hämäläinen and Ilmoniemi, 1994; Matsuura and Okabe, 1995; Gorodnitsky et al., 1995]. The parametric methods are reliant on a more or less subjective guess on the number of sources, while distributed methods avoid this they render the source localization problem underdetermined. Depending on the detail of the head model 1,000 to 100,000 sources must be estimated and the inverse problem is therefore very ill-posed and thus has no unique solution. The quality of a solution thus greatly depends on the type of regularization or source priors employed. Note that when the source orientations are free the number of variables to estimate increase three folds compared to employing fixed orientations [Henson et al., 2009a]. Modeling the sources as having fixed orientations is based on the assumption that the sources are oriented perpendicular to the cortex surface [Baillet et al., 2001].

Focal and sparse activation of the brain is often anticipated and assumed in EEG brain modeling [Gorodnitsky and Rao, 1997; Wipf and Rao, 2007; Friston et al., 2008; Zhang and Rao, 2011; Montoya-Martinez et al., 2012; Gramfort et al., 2013]. This assumption is partly based on the anatomy of the brain. EEG activity is for example believed to arise from assemblies of synchronously activated neurons. Additionally studies have shown that short ranging cortical connections are more numerous than long ranging cortical connections [Schüz and Braitenberg, 2002; Markov et al., 2014]. Further modeling assumptions include combining the neuron assemblies into bigger functional units thus reducing the ill-posedness of the problem by limiting the number of effective source locations to hundreds instead of thousands [Friston et al., 2008; Stahlhut et al., 2013]. This can for example be achieved by combining the surface elements of the tesselated cortex into source patches that are modeled as single source components in the inverse problem. In [Friston et al., 2008] the source patches have smooth and compact spatial distributions and in [Stahlhut et al., 2013] the functional units are defined as the Brodmann's areas of the brain .

Generally the interest of source localization is to examine the source distribution across time. Temporal properties of the EEG can therefore also be exploited to improve the inverse solution for example by assuming temporal smoothness. The generators of the scalp EEG are in ERP studies often believed to be active for some window of time. Temporal coherency can thus, e.g., be modeled by fixing the state of activation for each source across time samples, i.e., using multiple measurements vector (MMV) models [Wipf and Rao, 2007; Zhang and Rao, 2011; Friston et al., 2008]. These models assume that a source is either turned off or on for the duration of the investigated time window.

The multiple response sparse Bayesian learning (M-SBL) method [Wipf and Rao, 2007] is an MMV model providing row sparsity meaning that only a few sources (rows) are active across time. In M-SBL sparsity is obtained by automatic relevance determination (ARD) [Hansen and Rasmussen, 1994; MacKay,

1995] where the activity of each source is assigned a Gaussian prior with zero mean and a variance controlled by a hyperparameter. This hyperparameter is estimated by evidence maximization and controls whether the associated source is pruned or not. The T-MSBL [Zhang and Rao, 2011] is an extension to M-SBL that models temporal correlations within each source. ARD like inference is implemented in the multiple sparse priors model (MSP) [Friston et al., 2008] to recover the source components providing optimal free energy (or model evidence). Temporal smoothness is achieved by MSP using singular value decomposition (SVD) which estimates temporal basis functions, effectively ensuring row sparsity. M-FOCUSS [Cotter et al., 2005] is an MMV extension to the focal underdetermined system solver (FOCUSS) (Gorodnitsky and Rao, 1997). M-FOCUSS (and FOCUSS) employ a re-weighted norm minimization and obtain sparse solutions by regularization of the source estimate's $l_p$-norm such that $p \leq 1$. The assumption of a common sparsity structure across time as assumed by M-FOCUSS and the other MMV models might only be valid for a limited number of time samples. Alternative approaches to modeling temporal coherence are discussed in Chapter 3.

Chapter 3

# Spatio-Temporal Source Localization with the Variational Garrote

Unmixing of the EEG signal is in general ill-posed and therefore requires source priors based on known properties of the brain dynamics. As we are interested in reconstructing the time dependent source distribution, we construct source priors based both on spatial and temporal assumptions.

The variational garrote (VG) introduced in [Kappen, 2011; Kappen and Gómez, 2014] is the starting point of the developed inverse solvers. VG was chosen because of its high flexibility stemming from a spike and slab like separation of 'what' happens 'where'. This separation is facilitated by expressing the activation using two variables; one variable containing the source strengths and one binary variable controlling the activation states (active or inactive). The inference procedure of VG furthermore leads to intuitive updates rules of the source estimates, and finally, its low computational complexity encourages its use in real time EEG imaging.

In the following it is assumed that the source distribution of interest is sparse, e.g., because noise sources of biological origin have been removed by averaging and possibly minimized additionally by contrast subtraction of a control

condition. Further supporting the employment of sparse priors is when used in conjunction with independent component analysis (ICA). Delorme et al. showed that ICA can separate the EEG into scalp maps that are well described by dipolar sources [Delorme et al., 2012]. ICA can therefore be used to decompose the EEG into components having sparse source representations.

We have extended VG in two directions each with different temporal priors. The time-expanded VG (teVG) assumes a common sparsity profile across time while MarkoVG relaxes this strict smoothness requirement and allows for transient activity. Both VG extensions have been augmented to model smooth spatial source distributions by the incorporation of spatial basis functions.

Reconstruction by VG, teVG and MarkoVG of a synthetic dataset is seen in Fig. 3.1. The VG correctly located the two sources carrying activity (in blue and red) and the retrieved activity in these sources were limited to the relevant time samples (3.1B). However, false positives with low magnitude activity were also recovered (grey lines). This spurious activity was removed in Fig. 3.1C by teVG's prior assumption of temporal smoothness. However the true sources were now also given non-zero activity outside the relevant time samples. MarkoVG in Fig. 3.1D is seen to combine the best of VG and teVG by only recovering activity in the true locations *and* time samples.

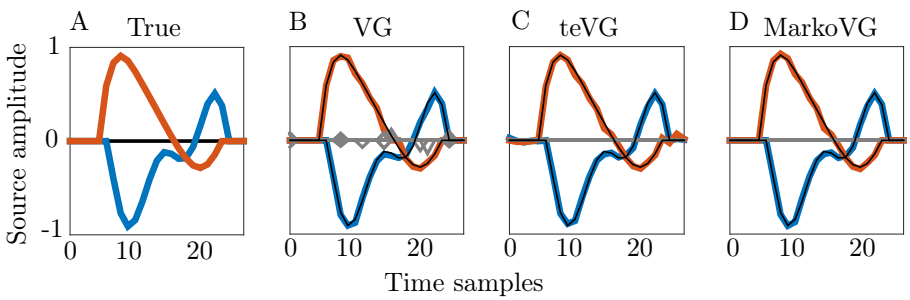



**Figure 3.1:** Reconstruction of a simulated dataset created using a randomly generated forward model of size $50 \times 500$. (A) Two out of the 500 sources were given the activity seen in red and blue, the remaining were set to zero. Noise was added to yield a signal-to-noise ratio (SNR) of 10 dB. The activity was reconstructed using (B) VG, (C) teVG and (D) MarkoVG. The true and estimated temporal dynamics of the two planted sources are seen in red and blue, while the grey indicate false positives. The black lines in B-D illustrate the true activation.

## 3.1 Validation Methods

The performance of the proposed algorithms is in the following judged based on simulations and real EEG data. Performance measures are in the simulations comprised of the localization error and the source retrieval index, i.e., the $F_1$-measure. The localization error is defined as the average Euclidean distance from all sources estimated to be active to their respective nearest located 'true' source. The $F_1$-measure is defined as [Rijsbergen, 1979; Makhoul et al., 1999]

$$F_1\text{-measure} = \frac{2\,\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{\text{TP} + \text{FP} + \text{P}}, \tag{3.1}$$

where TP, FP and P indicate the number of the true, false and actual positives. In a system where the solution is sparse the $F_1$-measure effectively emphasizes the influence of the number of correctly located sources compared to falsely located sources. In the same situation, the accuracy, which counts the number of true predictions, would in contrast be more influenced by the number of true negatives. Since any sparse estimation has a high number of true negatives the accuracy is suboptimal in judging the performance in sparse systems. The localization error is often used to estimate source localization performance [Friston et al., 2008; Akalin Acar and Makeig, 2013] and is a more 'soft' measure than the $F_1$-measure since the latter operates by 'all-or-nothing', i.e., a source is either correctly or incorrectly placed. The $F_1$-measure has the benefit compared to the localization error of also being measured in time samples were true activation is absent. Note that the $F_1$-measure could be augmented to count nearly correctly located sources as partly successful classifications.

The applied 'real' EEG data is from a subject included in the multimodal neuroimaging dataset [Wakeman and Henson, 2015]. The study paradigm investigated the previously mentioned face-evoked response, see also [Henson et al., 2003]. The subject thus viewed scrambled, famous and unfamiliar faces. In the following we use the first run of the data collection from subject 2. In this run 50 scrambled, 49 famous and 47 unfamiliar faces were presented to the subject. The famous and unfamiliar conditions were averaged and the averaged scrambled conditions subtracted, thus creating a differential ERP. Prior to averaging, preprocessing was performed following the MATLAB SPM8 scripts of R. Henson[1]. The applied paradigm has been highly investigated and there is general agreement that the occipital/fusiform face areas (O/FFAs) are active during face processing [Kanwisher et al., 1997; Henson et al., 2009b]. This observation is further supported by lesion studies [Eimer and McCarthy, 1999; Dalrymple et al., 2011] and the dataset therefore serves as a benchmark dataset to validate source localization methods.

[1] available at `ftp://ftp.mrc-cbu.cam.ac.uk/personal/rik.henson/wakemandg_hensonrn/SPMScripts/`

In the present chapter the forward model will be considered known and fixed. However, the proposed teVG also provides the starting point for inferring forward models as presented in Chapter 4. The forward model applied in this chapter was based on the structural scans of the previously mentioned subject. The head was segmented into three layers: scalp, skull and brain, and a canonical BEM head model was created using SPM8 [Phillips, 2000; Ashburner and Friston, 2005]. The cortex mesh consisted of 8,196 vertices. The head scan was coregistered to the 70 EEG electrodes using fiducials markers, head shape points and digitized electrode locations. We used the SPM8 template conductivity values for the head compartments to generate the forward model.

### 3.1.1 Implementation of Alternative Methods

The proposed VG algorithms are in the following compared to the multiple measurement vectors (MMV) models described in Section 2.3.2: T-MSBL, M-FOCUSS and MSP. MATLAB code[2]provided by Z. Zhang was used to implement the T-MSBL and M-FOCUSS algorithms. The T-MSBL was in this implementation given the noise level 'mild' or 'large' according to the signal-to-noise ratio (SNR) and M-FOCUSS was given the exact noise variance. The two models, especially M-FOCUSS, were thus given more information than MSP and the VG methods. Detailed noise information is not readily available in real data and M-FOCUSS was therefore not included in the real data examples. The SPM12 implemented version of the MSP algorithm [Friston et al., 2008] was applied in the 'MSP' setting providing inference of the sources by a combination of greedy search and automatic relevance determination (ARD).

## 3.2 The Variational Garrote

VG promotes sparsity analogously to a spike and slab prior on the source estimates. VG thus shares characteristics with the paired mean field model (PMF) [Titsias and Lázaro-Gredilla, 2011] and the Bayesian multi-task feature selection method [Hernández-Lobato et al., 2010], although all have different inference schemes. An alternative method to recover sparse solutions is by ARD [Hansen and Rasmussen, 1994; MacKay, 1995] as adopted in, e.g., sparse Bayesian learning methods (SBL) [Tipping, 2001; Wipf and Rao, 2007; Zhang and Rao, 2011]. A short introduction to Kappen's VG algorithm [Kappen, 2011; Kappen and Gómez, 2014] is provided to establish the foundation of the following sections' extensions.

[2] `http://dsp.ucsd.edu/~zhilin/Software.html`

As described earlier scalp EEG is a summation of the underlying EEG generators. The source strengths, $x_n$, in the $n = 1..N$ source locations are thus mapped to the electrode potentials, $y_k$, of electrode $k = 1..K$ through a design matrix/input matrix/forward model, $A_{kn}$. The generative model of EEG can thus be described as a linear regression problem

$$y_k = \sum_{n=1}^{N} A_{kn} x_n + e_k, \tag{3.2}$$

where the noise, $e_n \sim \mathcal{N}(0, 1/\beta)$, is i.i.d. with variance $1/\beta$.

The VG includes a binary variable, $s_n$, controlling the state of source $n$. If $s_n$ is 0 or 1, source $n$ is inactive or active, respectively. The generative model is now

$$y_k = \sum_{n=1}^{N} A_{kn} s_n x_n + e_k. \tag{3.3}$$

The solution is, as proposed by Kappen et al., found by computing the expectation of the binary variable, performing cross-validation to estimate the sparsity hyperparameter and using maximum a posteriori (MAP) to estimate the remaining variables. Algorithm 1 provides an overview of the VG inference procedure.

The free energy derived from Algorithm 1G equals the negative bound on the model evidence and is here presented in a dual formulation (which has reduced computational complexity compared to the original formulation)

$$\begin{aligned} F(\mathbf{m}, \mathbf{x}, \beta, \mathbf{z}, \lambda) = &- \frac{K}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{k=1}^{K} (z_k - y_k)^2 + \frac{K\beta}{2} \sum_{n=1}^{N} m_n (1 - m_n) x_n^2 \chi_{nn} \\ &- \gamma \sum_{n=1}^{N} m_n + \sum_{n=1}^{N} \left( m_n \log(m_n) + (1 - m_n) \log(1 - m_n) \right) \\ &+ N \log(1 + \exp(\gamma)) + \sum_{k=1}^{K} \lambda_k \left( z_k - \sum_{n=1}^{N} m_n x_n A_{kn} \right), \end{aligned} \tag{3.4}$$

where $z_k = \sum_{n=1}^{N} m_n x_n A_{kn}$ and $\lambda_k$ are Lagrange multipliers. The free energy is minimized to find MAP estimates of $\mathbf{m}$, $\mathbf{x}$, $\mathbf{z}$, $\lambda$, and $\beta$. Our VG extensions primarily involve augmentations of the activation state parameter, $s_n$, we therefore

**Algorithm 1:** (A) Posterior probability of VG, where $\mathbf{D}$ is the data: $\mathbf{A}$ and $\mathbf{y}$. (B) Prior on $\mathbf{s}$, where the hyperparameter $\gamma$ controls the degree of sparsity. (C) The likelihood of the data, where $\mathbf{b} = \mathbf{y}^T\mathbf{A}/K$, $\boldsymbol{\chi}$ is the covariance matrix of $\mathbf{A}$, and $\sigma_y^2$ is the variance of $\mathbf{y}$. (D) The prior $p(\mathbf{x}, \beta)$ is assumed flat and $p(\mathbf{D}|\gamma)$ is ignored in the optimization of $\mathbf{x}$ and $\beta$. (E) Posterior probability with the simplifications. (F) The binary variable is marginalized out. (G) The log of (F) with variational approximation (H) and lower bounded using Jensen's inequality.

(A) **PP:** $p(\mathbf{s}, \mathbf{x}, \beta|\mathbf{D}, \gamma) = \dfrac{p(\mathbf{x}, \beta)p(\mathbf{s}|\gamma)p(\mathbf{D}|\mathbf{s}, \mathbf{x}, \beta)}{p(\mathbf{D}|\gamma)}$;

(B) **where:** $p(\mathbf{s}|\gamma) = \prod_{n=1}^{N} p(s_n|\gamma)$, $p(s_n|\gamma) = \dfrac{\exp(\gamma s_n)}{1 + \exp(\gamma)}$;

(C) **and:** $p(D|\mathbf{s}, \mathbf{x}, \beta) =$
$\left(\dfrac{\beta}{2\pi}\right)^{K/2} \exp\left(-\dfrac{\beta K}{2}\left(\sum_{n,n'=1}^{N} s_n s_{n'} x_n x_{n'} \chi_{nn'} - 2\sum_{n=1}^{N} x_n s_n b_n + \sigma_y^2\right)\right)$;

(D) **Simplification:** $p(\mathbf{x}, \beta) \propto 1$, $p(\mathbf{D}|\gamma)$ is constant in optimizing $\mathbf{x}$, $\beta$.;

(E) **New PP:** $p(\mathbf{s}, \mathbf{x}, \beta|\mathbf{D}, \gamma) \propto p(\mathbf{s}|\gamma)p(\mathbf{D}|\mathbf{s}, \mathbf{x}, \beta)$;

(F) **Marginalization:** $p(\mathbf{x}, \beta|\mathbf{D}, \gamma) \propto \sum_{\mathbf{s}} p(\mathbf{s}|\gamma)p(\mathbf{D}|\mathbf{s}, \mathbf{x}, \beta)$;

(G) **Approximation:** $\log p(\mathbf{x}, \beta|\mathbf{D}, \gamma) \geq -\sum_{\mathbf{s}} q(\mathbf{s}) \log \dfrac{q(\mathbf{s})}{p(\mathbf{s}|\gamma)p(\mathbf{D}|\mathbf{s}, \mathbf{x}, \beta)}$;

(H) **where:** $q(\mathbf{s}) = \prod_{n=1}^{N} q_n(s_n)$, $q_n(s_n) = m_n s_n + (1 - m_n)(1 - s_n)$;

reprint its expectation

$$m_n = \sigma_f\left(\frac{\beta K}{2}\chi_{nn}x_n^2 + \gamma\right), \tag{3.5}$$

where $\sigma_f(a) = (1 + \exp(-a))^{-1}$. The variational mean, $m_n$, which describes the probability of source $n$ being active, is updated by interpolation of Eq. (3.5) and $m_n$ in the previous iteration, i.e.,

$$m_n^{it} = (1 - \eta_1)m_n^{it-1} + \eta_1\sigma_f\left(\frac{\beta^{it}p}{2}\chi_{nn}(x_n^{it})^2 + \gamma\right), \tag{3.6}$$

where the superscript $it$ indicates iteration number. The equation set with updates for $\mathbf{m}$, $\mathbf{x}$, $\mathbf{z}$, $\lambda$, and $\beta$ is in [Kappen, 2011; Kappen and Gómez, 2014] solved by fixed-point iteration until convergence. In the dual formulation of VG the equation set scales cubic in the number of observations, $K$, and linear in the number of sources, $N$.

The sparsity controlling hyperparameter, $\gamma$, was by Kappen et al. proposed found by cross-validation in an annealing/reheating process of increasing/decreasing the applied $\gamma$. The maximum value of $\gamma$ and the step size is heuristically set while the minimum $\gamma$ is defined as $\gamma_{\min} = \min_n \left( -\frac{p\chi_{nn} b_n}{2\sigma_y} \right) - \log \left( \frac{1-\epsilon}{\epsilon} \right)$, where $\epsilon$ is small. Instead of providing $\epsilon$ we propose to estimate $\gamma_{\min}$ empirically by increasing $\gamma$ from a very negative value until the VG solution starts to have non-zero values. It is noted that the solution converges quickly when all elements are zero and the proposed procedure does therefore not add significantly to the computation time. Furthermore we find that the annealing/reheating process traps the VG solution in local minima and we therefore suggest to apply each value of $\gamma$ independently by re-initializing $m_n = 0$ for each $\gamma$.

The related PMF method proposed by Titsias et al. also uses variational approximation [Titsias and Lázaro-Gredilla, 2011]. It deviates from VG in the inference scheme by estimating all variables using expectation-maximization. Furthermore the variational approximation is in [Titsias and Lázaro-Gredilla, 2011] defined as a joint distribution of $\mathbf{s}$ and $\mathbf{x}$. Kappen et al. showed similar performance of VG and PMF [Kappen and Gómez, 2014]. However when the inputs were highly correlated PMF was found to suffer more from local minima. The input matrix in the EEG inverse problem corresponds to the forward model which indeed has highly correlated columns since neighboring sources project very similarly to the electrodes. According to the findings in [Kappen and Gómez, 2014] PMF is therefore not suitable for solving the EEG inverse problem. Finally VG was shown empirically to be more computationally efficient compared to PMF.

We applied VG to EEG source localization in contribution [A], where we compared it to forward selection [Draper et al., 1966], the $l_1$-regularizer LASSO [Tibshirani, 1996] and to a sparse Bayesian learning method [Tipping, 2001]. VG was able to recover the highest number of true positives while also having no false positives. On the synthetic example in Fig. 3.1 VG was able to recover the true sources, but did however also have small spikes of activity in other locations.

## 3.3 The Time-Expanded Variational Garrote

The original VG considers one sample per observation as seen in Eq. (3.3). We propose to model multiple time samples and simultaneously exploit the inherent similarity between activation states in neighboring time samples. Temporal smoothness naturally occurs in many processes when recorded with sufficient

sampling rate thus including EEG signals [Cotter et al., 2005]. In teVG we model temporal coherency by constraining the activation state to be constant across time for each source location but allow the activity strength to vary. The problem is thus modeled as being an MMV problem and can be expressed as

$$Y_{kt} = \sum_{n=1}^{N} A_{kn} s_n X_{nt} + E_{kt}. \tag{3.7}$$

As seen, one variable for each source controls the activation state for all time samples, while the activation strength varies across time. The free energy in this framework is

$$\begin{aligned} F(\mathbf{m}, \mathbf{X}, \beta, \mathbf{Z}, \lambda) = &-\frac{TK}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k=1}^{T,K} (Z_{kt} - Y_{kt})^2 + \frac{K\beta}{2} \sum_{t,n=1}^{T,N} m_n (1 - m_n) X_{nt}^2 \chi_{nn} \\ &- \gamma \sum_{n=1}^{N} m_n + \sum_{n=1}^{N} (m_n \log(m_n) + (1 - m_n) \log(1 - m_n)) \\ &+ N \log(1 + \exp(\gamma)) + \sum_{t,k=1}^{T,K} \lambda_{kt} \left( Z_{kt} - \sum_{n=1}^{N} m_n X_{nt} A_{kn} \right). \end{aligned} \tag{3.8}$$

The inference procedure is only mildly adjusted compared to the original VG, as presented in contribution [B]. The variational mean updates are also very similar to the original formulation in Eq. (3.5)

$$m_n = \sigma_f \left( \frac{\beta p}{2} \chi_{nn} \sum_{t=1}^{T} X_{nt}^2 + \gamma \right). \tag{3.9}$$

The updates for the remaining variables can be seen in Appendix I. In contribution [C] we implemented gradient descent (GD) for $m_n$ such that the update rule in iteration $it$ is

$$m_n^{it} = m_n^{it-1} - \eta_2 \nabla F(m_n^{it-1}). \tag{3.10}$$

Where the gradient of the free energy can be seen in Appendix I and where $\eta_2 > 0$ is the learning rate which we set heuristically. We find that the VG solution is more robust to this parameter than the smoothing parameter $\eta_1$ in Eq. (3.6).

In most EEG imaging applications sources are predicted among thousands of possible locations. To decrease this number and at the same time decrease the computation time, neighboring sources can be combined into spatial basis functions. In contribution [D] we thus extended teVG to the framework of basis functions as introduced in the MSP model [Friston et al., 2008]. The basis

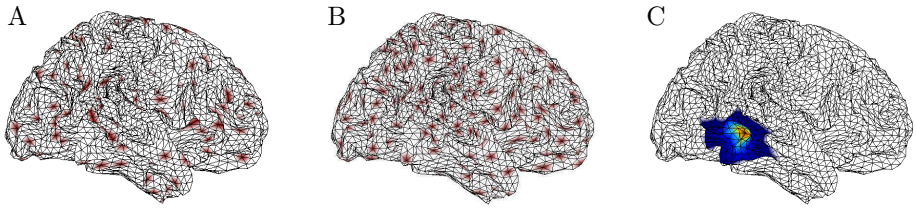



**Figure 3.2:** The basis functions in a mesh of 8,196 vertices. (A) The center vertices of MSP [Friston et al., 2008]; 512 distributed across both hemispheres (seen in red). (B) The proposed center vertices; 776 based on the connectivity of the mesh (seen in red). (C) The compact spatial extension of one source component. Each component covers 98-128 vertices. [From D and G].

function locations are in MSP evenly sampled from the source indices under the condition of left-right hemisphere symmetry. In the cortex mesh of 8,196 vertices 512 locations are chosen as basis function centers (Fig. 3.2A). Bilateral source components are created by combining the hemispheric symmetrically located components. In total 768 source components are thus formed. The spatial extension of each component is determined by the connectivity of the mesh. The spatial profile of a component thus extends from the center vertex to its neighbors of maximally eight degrees with smoothly decreasing strength (Fig. 3.2C).

We propose to place the centers of the source components by exploiting the connectivity of the mesh and thus promote even distribution in space as opposed to in source indices. In the proposed method we let the adjacency matrix control whether a randomly sampled vertex will serve as a center of a new source component. If a suggested vertex is at most a third degree neighbor to the already sampled centers it is accepted and included in the basis set. There are thus at least two vertices between the basis centers. By seeding the random number generator in MATLAB the chosen centers are reproducible. In our implementation 776 centers are defined in the cortex consisting of 8,196 vertices, see Fig. 3.2B. Each of the 8,196 source locations is a part of 8 to 16 basis functions. In the original MSP sampling technique the source locations are included in 3 to 17 basis functions.

The spatial basis functions can easily be incorporated into equation (3.7),

$$\text{teVG with basis functions:} \quad Y_{kt} = \sum_{c=1}^{C} A'_{kc} s'_c X'_{ct} + E_{kt}, \tag{3.11}$$

where $A'_{kc} = \sum_{n=1}^{N} A_{kn} B_{nc}$, $s'_c = \sum_{n=1}^{N} s_n B_{nc}$, $X'_{ct} = \sum_{n=1}^{N} B_{cn} X_{nt}$ and $c$ is

source component 1 through $C$.

In the following, teVG is validated through simulations and by the real EEG data described in Section 3.1.

### 3.3.1 Evaluation of teVG - Simulations

In Fig. 3.3 the computation time and $F_1$-measure were investigated on simulated data created by a forward model of realistic size with randomly generated elements. The five active sources were each temporally correlated using an autoregressive model (AR) of order 1 with AR coefficient 0.9. The teVG is compared to the MMV models, T-MSBL and M-FOCUSS, which are described in Section 3.1.1.

The computation time of teVG was for runs with less than 45 iterations shorter than that of M-FOCUSS, as seen in Fig. 3.3A. T-MSBL was for all investigated number of iterations slower than both methods. Computation time of M-FOCUSS stabilized already after 5 iterations which is attributed to the optimized implementation scheme for this algorithm[3]. Convergence of the M-FOCUSS solution was however not achieved after this low number of iterations as seen from Fig. 3.3B. The teVG on the other hand attained (near) optimal source retrieval after only a few iterations.

The T-MSBL did not perform well on the number of applied iterations presented in Fig. 3.3B. The number of iterations was too low causing TMSBL to recover many false sources, thus driving the $F_1$-measure down. However, T-MSBL did improve upon running more iterations and eventually obtained similar performance to teVG and M-FOCUSS (see contribution [B]).

In contribution [C] we showed that the cross-validation scheme for estimating $\gamma$ was efficient in providing a suitable sparsity level. The cross-validation procedure thus generally showed similar performance to an oracle setting of teVG where the sparsity level was chosen based on optimal source retrieval.

[3]In the current implementation M-FOCUSS prunes at each iteration the sources found irrelevant. As the computation time depends on the number of sources the pruning decreases computation time drastically.

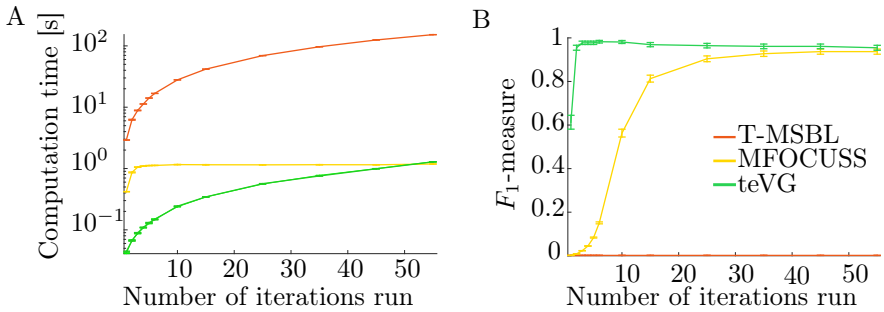



**Figure 3.3:** Comparison of teVG, T-MSBL and M-FOCUSS on a simulated example with a random design matrix of size $128 \times 8{,}196$. Five temporally correlated (AR(1) with AR coefficient 0.9) sources in three time samples were planted and noise was added to give an SNR of 3 dB. (A) Computation time and (B) $F_1$-measure as a function of the applied number of iterations. [From B].

### 3.3.2 Evaluation of teVG - Real EEG Data

In Fig. 3.4 the sources of the face-evoked EEG response are seen reconstructed by teVG with GD and the spatial basis functions of [Friston et al., 2008]. The basis functions including bilateral components were chosen as we in this dataset expect bilateral activity. The estimated spatial source distribution in Fig. 3.4A does indeed show high left-right symmetry. The temporal dynamics of the estimated sources with highest magnitude activity are seen in Fig. 3.4B. Sources in the vicinity of the FFAs and OFAs were also recovered by teVG and their temporal dynamics are shown in Fig. 3.4C. Especially the bilateral activity of the FFAs (in green) showed the well-known N170 peak.

In contribution [D] we compared the robustness of the teVG and MSP reconstructed sources of an EEG dataset with the same paradigm as the above. We applied a so-called split-half resampling setup where we divided both the faces and scrambled faces epochs into two datasets and calculated the differential ERP in each split. Reconstruction was then performed on each split with teVG and MSP. This procedure was replicated 100 times with different splits of data. The results showed that teVG was more robust than MSP. The teVG estimated temporal dynamics thus had a smaller variance across datasets and the strongest sources were more consistently located.

## 3.4 MarkoVG

The claim of stationary activity made by teVG is inflexible in settings where the EEG signal is transient and where the interest is in the timing of the brain source activation. For example in Fig. 3.1C it is seen how activity outside the relevant time samples are predicted by teVG. Thus instead of modeling the activation state as one variable for each source location we in contribution [E] proposed to have separate but connected state activation variables across time. MarkoVG controls the activation state by a Markov prior such that the activation state in one time sample depends on the activation state in the previous.

In MarkoVG the activation state in source location $n$ for time sample $t$ (i.e. $S_{nt} = j$) thus depends on the activation state of the same source in the previous time sample (i.e. $S_{n,t-1} = i$) and on the state transition probabilities. If $i, j = 0, 1$ indicate the possible states, the transition probability can be formulated as $\Gamma_{ji} = \mathrm{P}(S_{nt} = j | S_{n,t-1} = i)$. The full $2 \times 2$ transition matrix,

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix} = \begin{bmatrix} 1 - \Gamma_{10} & \Gamma_{01} \\ \Gamma_{10} & 1 - \Gamma_{01} \end{bmatrix}, \tag{3.12}$$

can be parametrized by two parameters, where the relations $\Gamma_{00} + \Gamma_{10} = 1$ and $\Gamma_{01} + \Gamma_{11} = 1$ are used.

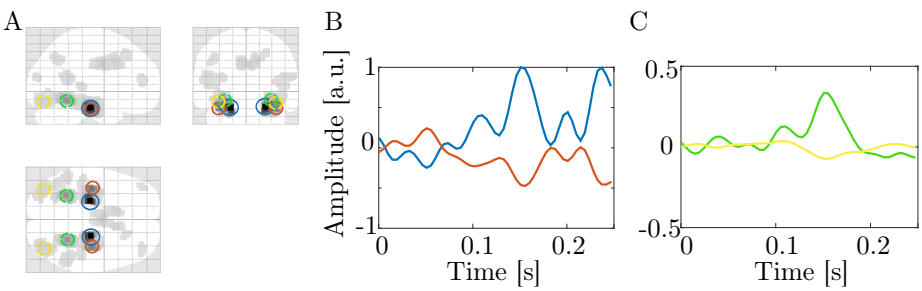



**Figure 3.4:** Source distribution estimated by teVG with GD and spatial bilateral basis functions on the face-evoked EEG response. (A) The spatial source distribution 151 ms after stimuli onset. The blue and red circles indicate the strongest and second strongest source components, which were both bilateral. The green circles indicate the locations of source components in the proximity of the FFAs and the yellow in the proximity of the OFAs. (B) The temporal dynamics of the strongest and second strongest sources. (C) The temporal dynamics of the FFA and OFA components.

The dual free energy with the Markov prior is

$$
\begin{aligned}
F = &- \frac{KT}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k} (Y_{kt} - Z_{kt})^2 + \frac{K\beta}{2} \sum_{t,n} M_{nt}(1 - M_{nt}) X_{nt}^2 \chi_{nn} \\
&- \sum_{n,t} \left[ M_{nt} \log \frac{\Gamma_{10}}{\Gamma_{00}} + M_{n,t-1} \log \frac{\Gamma_{01}}{\Gamma_{00}} + (M_{nt} M_{n,t-1}) \log \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \right] \\
&+ NT \log \frac{1}{\Gamma_{00}} + \sum_{n,t} \left[ M_{nt} \log(M_{nt}) + (1 - M_{nt}) \log(1 - M_{nt}) \right] \\
&+ \sum_{t,k} \lambda_{kt} \left( Z_{kt} - \sum_{n} A_{kn} M_{nt} X_{nt} \right). \qquad (3.13)
\end{aligned}
$$

The MAP updates of $Z_{kt}$, $X_{kt}$, $\lambda_{kt}$ and $\beta$ can be seen in contribution [E] and Appendix [J]. The update rule for the variational mean is reprinted here

$$
M_{nt} = \sigma_f \left( \frac{K\beta}{2} \chi_{nn} X_{nt}^2 + \gamma_1 + \gamma_2 \left( M_{n,t-1} + M_{n,t+1} \right) \right). \qquad (3.14)
$$

The hyperparameters are now, $\gamma_1 = \log \left( \frac{\Gamma_{10}\Gamma_{01}}{\Gamma_{00}^2} \right)$ which controls the sparsity, and $\gamma_2 = \log \left( \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \right)$ which controls the temporal smoothness. Having $\Gamma_{01} + \Gamma_{10} = 1$ leads to $\gamma_2 = 0$ and hence a restoration of the original variational mean as seen in Eq. (3.5). In contribution [E] we found that the optimal values of the sparsity and smoothness were highly connected. The source retrieval index suggested that a linear relation between the parameters was appropriate for optimal performance. More precisely we found that the magnitude of the smoothness should be slightly smaller than the magnitude of the sparsity. We therefore in the following implement the relation $\gamma_2 = -0.9\gamma_1$ and thus avoid cross-validating over two parameters. By propagating the defined relation to the transition probability some intuition on its effects can be gained. The relation $\gamma_2 = -\gamma_1$ would imply that the probability of switching from an inactive to an active state is the same as changing from an active to an inactive state, i.e. $\Gamma_{10} = \Gamma_{01}$. In this situation smoothness is promoted when $\Gamma_{10} = \Gamma_{01} < 0.5$, however sparsity is not encouraged and we thus include a factor in the smoothness-sparsity relation to ensure sparse solutions. The algorithmic setup is outlined in Algorithm 2.

As seen from Algorithm 2 we now use the free energy calculated on validation sets to estimate the optimum parameter setting. We find that this measure is less prone to overfitting than the originally proposed technique in [Kappen, 2011] where the parameter setting was based on the mean squared validation error of the predicted EEG signal.

**Algorithm 2:** The MarkoVG framework including GD updates for $\mathbf{M}$. The free energy is presented in Eq. (3.13) and the four fold cross-validation apply 20 combinations of $\gamma_1$ and $\gamma_2$. Convergence is based on $\beta$ and a maximum of 500 iterations are run.

**Input:** EEG ($\mathbf{Y}$) and forward model ($\mathbf{A}$).
**Output:** The reconstructed source strengths ($\mathbf{X}$) and the expectation of the activation states ($\mathbf{M}$).

```
for each cross fold do
    for each i_γ do
        γ_1 = γ_all(i_γ);
        γ_2 = −0.9γ_1;
        M = 0_{N×T};
        iter = 1;
        while not converged AND iter ≤ iter_max do
            [M, X, β]=MarkoVG(Y_train, A_train, γ_1, γ_2);
            iter = iter + 1;
        end
        FEval(i_γ)=FreeEnergy(Y_validation, A_validation, γ_1, γ_2, M, X, β);
    end
    IDX(i_fold) = arg min_{i_γ}(FEval);
end
γ_{1,opt} = γ_all(median(IDX));
γ_{2,opt} = −0.9γ_{1,opt};
[M, X, β]=MarkoVG(Y, A, γ_{1,opt}, γ_{2,opt});
```

Spatial smoothness can easily be incorporated into the MarkoVG following the same procedure as done for the teVG in Eq. (3.11). GD updates are also similarly introduced for MarkoVG. In the following MarkoVG is evaluated using simulations and real data.

### 3.4.1 Evaluation of MarkoVG - Simulations

We demonstrate the influence of the sparsity and smoothness parameters on the MarkoVG solution in Fig. 3.5. In this example two sources were active and had the temporal dynamics seen in Fig. 3.5A and the activation states seen in Fig. 3.5B (black/white corresponds to $S_{nt} = 0/1$). The remaining of the in total 500 sources were inactive for all 25 time samples. The source distribution was projected to 50 observations through a random forward model and corrupted with noise to yield an SNR of 10 dB.

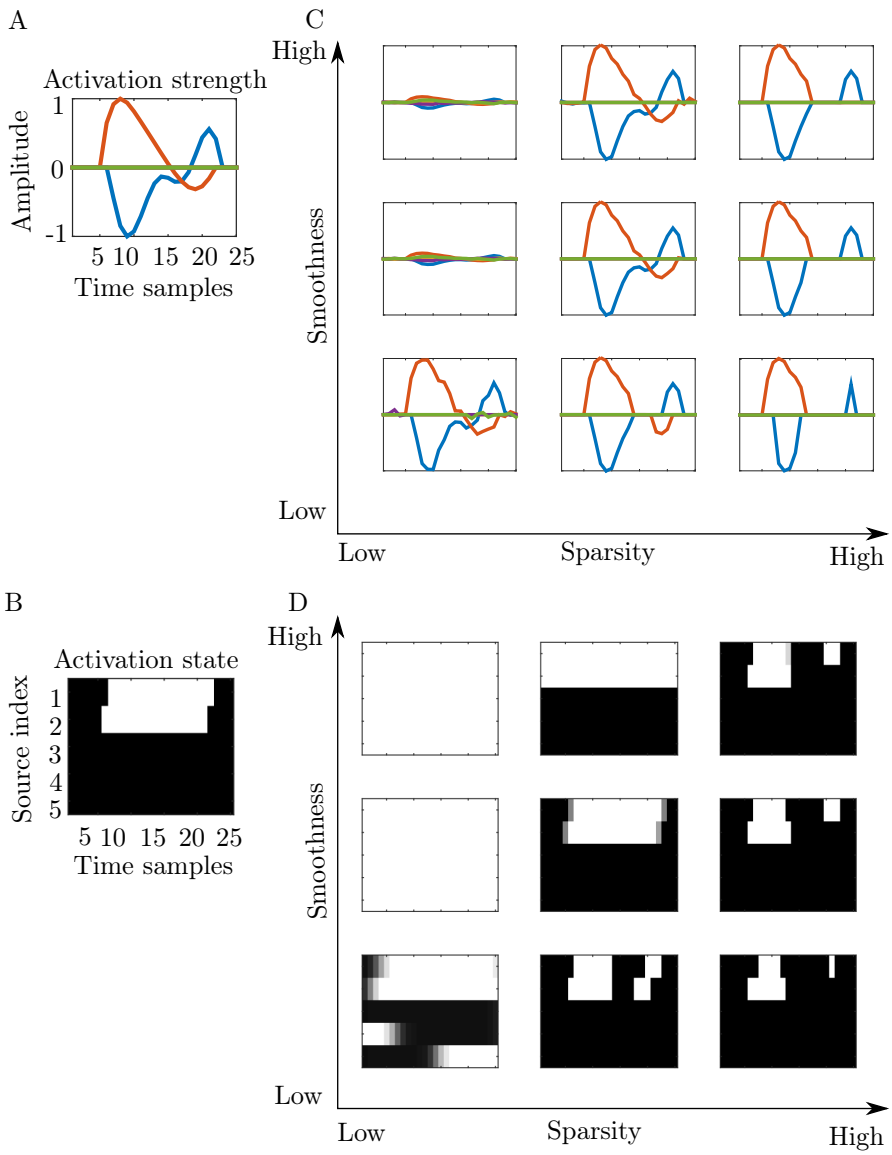



**Figure 3.5:** Smoothness/sparsity properties of MarkoVG illustrated using simulated data. The forward model was a randomly generated matrix of size $50 \times 500$. Noise was added to yield an SNR of 10 dB. (A) The temporal dynamics of the two active sources (red and blue). (B) The activation states of the source distribution, where white/black indicate active/inactive. The active sources had source indices 1 and 2. For visualization, only the first five source indices are shown. (C) Reconstruction of the sources with varying degrees of sparsity and smoothness. (D) The activation state of these reconstructions. The parameter setting for the solution in the center plots of C and D was estimated by cross-validation and corresponds to Fig. 3.1D. [From G].

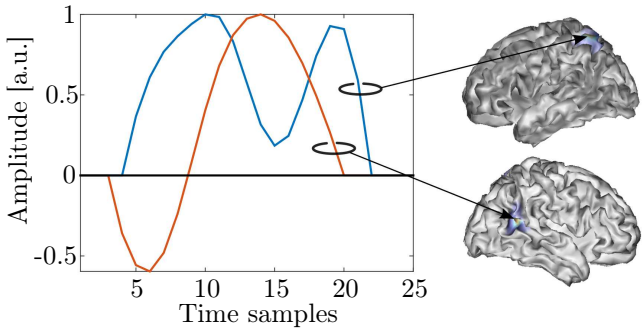



**Figure 3.6:** Example of simulated data constructed using a true forward model. Here one source is placed in each hemisphere. [From G].

The MarkoVG reconstruction of the signal is seen in Fig. 3.5C-D for different combinations of low/high smoothness and sparsity. Low sparsity (first column) yielded activity in other than the true locations and high sparsity (last column) pruned true activity in time samples of low magnitude activity. Low smoothness (bottom row) had a similar effect as high sparsity, while high smoothness (top row) favored more stationary solutions. The center plots of Fig. 3.5C-D, where the parameters were estimated by cross-validation, show a solution containing the correct level of sparsity and smoothness.

Next, datasets mimicking EEG activity were investigated. The real forward model described in Section 3.1 was therefore used in generating datasets from source distributions similar to that of Fig. 3.6. One to four source components, each with a spatial profile as seen in Fig. 3.2C, were randomly planted in any of the 8,196 possible locations. The source components had temporally non-stationary activity in 25 time samples.

Fig. 3.7 illustrates the behavior of the MarkoVG solution as a function of smoothness and sparsity on the example shown in Fig. 3.6. The simulated EEG signal was corrupted with noise to give an SNR of 8 dB. In these simulations the MarkoVG used the proposed basis function set of 776 basis functions to reconstruct from. A perfect reconstruction was thus not expected as the planted source components could be centered in any of the 8,196 vertices. In Fig. 3.7A it is seen that the averaged free energy of the validation sets was minimum just above the dashed line which indicate where $\gamma_1 = -\gamma_2$. As minimum localization error and maximum $F_1$-measure were also found in this region, the free energy was seen capable of choosing the optimal parameter setting, and validated our choice of defining $\gamma_2 = -0.9\gamma_1$

The performance of MarkoVG is further analyzed in Fig. 3.8 where it is also

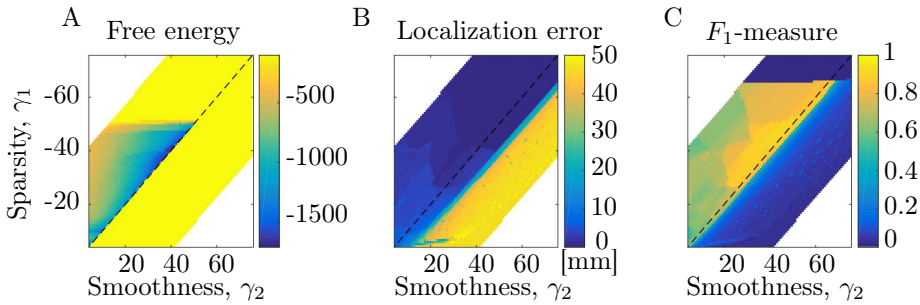



**Figure 3.7:** Performance of MarkoVG on the example in Fig. 3.6 as a function of sparsity and smoothness. (A) The averaged free energy of the four validation sets used in the cross-validation. (B) The localization error. (C) The $F_1$-measure. The dashed line indicates where $\gamma_1 = -\gamma_2$ and the white space where combinations of $\gamma_1$ and $\gamma_2$ are not meaningful. [From G].

compared to MSP, T-MSBL and M-FOCUSS. To obtain a fair comparison we also implemented these methods with the 776 basis functions. We furthermore included a version of T-MSBL, 'T-MSBL cross', for which we performed four-fold cross-validation to estimate the regularization parameter, as similarly done for MarkoVG. For each investigated SNR level we generated 100 repetitions of data with source distributions similar to that of Fig. 3.6. As seen in Fig. 3.8 MarkoVG obtained lowest localization error and T-MSBL and MarkoVG had similar $F_1$-measure and outperformed MSP and M-FOCUSS. It is furthermore noted that the cross-validated version of T-MSBL did not perform as well as the version of T-MSBL which was given the noise level.

### 3.4.2 Evaluation of MarkoVG - Real EEG Data

The earlier described face-evoked EEG response is in Fig. 3.9 reconstructed by MarkoVG, MSP and T-MSBL. As similarly done for teVG and this dataset we applied the spatial basis functions modeling bilateral activity. The source components estimated to have highest and second highest magnitude activity were for all three methods, as teVG, bilateral, see Fig. 3.9. The temporal dynamics of these source components are seen in the top panel and their locations are indicated in the bottom panel. The time samples for which the spatial distributions are shown corresponds to the timing of peak activity as estimated by each algorithm, corresponding to 151-155 ms after stimulus onset. The strongest sources exhibited for all methods the N170 peak. MarkoVG displayed a temporally more focal activity around this ERP component compared to the other

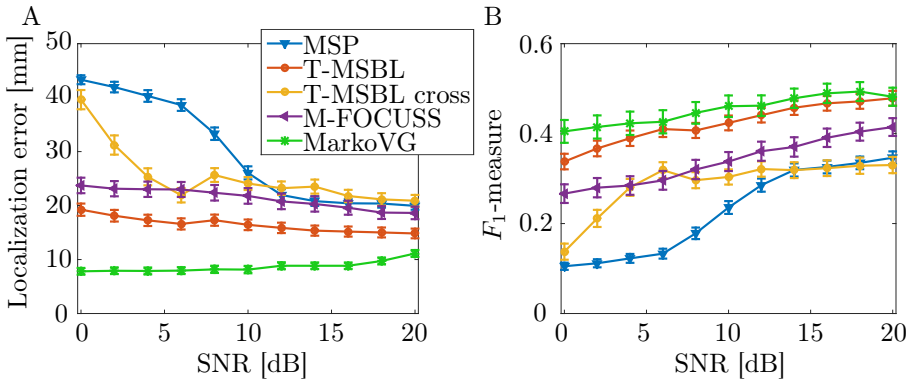



**Figure 3.8:** Performance of MarkoVG, T-MSBL, M-FOCUSS and MSP on 100 repetitions of data corresponding to Fig. 3.6. In 'T-MSBL cross' we performed cross-validation to find the regularization parameter. The errorbars indicate the standard error of the mean (s.e.m). (A) The localization error. (B) The $F_1$-measure. [From G].

algorithms.

The methods furthermore all recovered activity in vicinity of the OFAs, in fact the second strongest bilateral source component of the MarkoVG solution was found in the OFAs. Furthermore both T-MSBL and MarkoVG localized the strongest components as being in the FFAs. Finally MarkoVG provided a spatially more sparse distribution of sources compared to T-MSBL and MSP.

## 3.5 Relation to Other Work

Understanding the temporal dynamics of brain activity and materializing the obtained knowledge into source priors is important in EEG imaging and therefore also highly researched. Suitable temporal priors have the potential to not only retrieve accurate temporal information but also strengthen the spatial specificity. Invoking MMV models has, e.g., in convex relaxation methods shown to decrease the false recovery rate in ill-posed problems [Eldar and Rauhut, 2010]. Furthermore, [Cotter et al., 2005] showed a performance gain of the MMV extension, M-FOCUSS, relative to the original single measurement vector model, FOCUSS. However, the assumption of a fixed sparsity profile is not valid when the source signal is of transient nature. Activity outside the actually activated time samples might therefore mistakenly be retrieved by MMV models thus

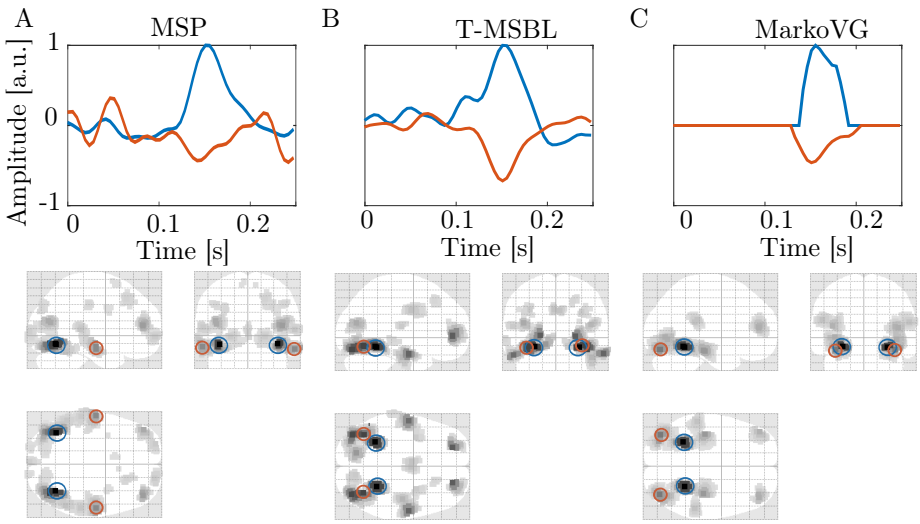



**Figure 3.9:** Source distribution estimated using spatial bilateral basis functions on the face-evoked EEG response by (A) MSP, (B) T-MSBL and (C) MarkoVG. The lower panel shows the spatial source distribution 151-155 ms after stimuli onset. The blue and red circles indicate the strongest and second strongest source components, which are both bilateral. The top panel shows the temporal dynamics of these sources. [Modified in G].

leading to wrong estimation of the source signal timing.

The temporal prior was in [Montoya-Martinez et al., 2012; Gramfort et al., 2013] made more flexible by assuming temporal coherence but not enforcing strict common support. Montoya-Martinez et al. introduced the term structured sparsity in the temporal domain where some activation patterns are preferred over others. Sources that are active/inactive for a number of time samples is, e.g., favored over sources with highly oscillating activation states. This type of structured sparsity was in [Montoya-Martinez et al., 2012] achieved by the sparse group lasso [Friedman et al., 2010].

While [Montoya-Martinez et al., 2012] worked directly in the spatio-temporal domain [Gramfort et al., 2013] used a combined time-frequency domain. Gramfort et al. applied a time-frequency dictionary to achieve structured sparsity. The temporal dynamics of each source are thus represented as a linear combination of a subset of the time-frequency dictionary atoms. In [Lina et al., 2014] source localization of oscillatory brain activity was also achieved by time-frequency analysis. Lina et al. used wavelet representation to model temporal

dynamics and spatial regions were controlled by hidden state variables. The activation state was thus like in VG controlled by a separate variable which in [Lina et al., 2014] determined the activation states of areas of approximately 6 $cm^2$.

The proposed framework in [Stahlhut et al., 2013] combined spatio-temporal basis functions with a so-called *source correction* term. The latter is beneficial when the basis functions are incapable of modeling all of the source dynamics. The method, called Aquavit, was adapted to employ spatial basis functions, either defined by Brodmann's areas or by the MSP source components. Aquavit is a hierarchical Bayesian model in which the hyperparameters are estimated by MAP and the remaining parameters by ARD variational approximation.

The VG sparsity hyperparameter was in [Andersen et al., 2013] proposed inferred, as opposed to estimated by cross-validation. This was achieved by simplifying the prior on the state activation variable ($S_{nt}$) to a binomial distribution. The new hyperparameter was assigned a Beta distribution and the hyperparameter was then approximated by MAP after fixing the hyper-hyperparameters. According to [Kappen and Gómez, 2014] performing inference of the regularization parameter leads to suboptimal performance when the columns of the input matrix are correlated. This was shown for the PMF model where local minima problems became more severe and deteriorated the PMF performance [Kappen and Gómez, 2014]. In sparse Bayesian learning identifiability issues arise when inferring both the regularization parameters and the noise variance because of a tight coupling between the two [Wipf and Rao, 2007]. A similar phenomenon is hypothesized to be the cause of the inferior performance of PMF. Kappen et al. showed that augmenting the PMF algorithm to the same annealing framework as proposed for VG improved the PMF solution.

In summary we have in this chapter described the implementation of physiological relevant priors for the variational garrote and thereby adapted the algorithm to the inverse problem of EEG. The two extensions, teVG and MarkoVG, differ in their modeling of the temporal dynamics. The teVG assumes stationarity of the source locations while MarkoVG allows transient sources. MarkoVG is thus preferred as it lets the data control the (non-)stationarity. However, it is hypothesized that teVG by being a less flexible model is more suitable when the signal-to-noise ratio is low, as well as in scenarios where the signal is truly stationary. Spatial basis functions were incorporated in both VG versions and using these we were able to recover both the OFAs and the FFAs.

Chapter 4

# Forward Model Inference

High spatial specificity of EEG imaging is contingent on an appropriate forward model which in turn requires correct modeling of the head compartments and their electrical properties. As described earlier this normally demands information that is either hard or at best difficult to attain. In the following we attempt to describe the variability among forward models and to facilitate more accessible construction of person-specific forward models.

The proposed method does not require structural scans or conductivity measurements to estimate the head properties but instead uses the recorded EEG to optimize over a parametrized space of forward models. This data driven approach recovers the forward model and EEG sources that combined best explain the recorded EEG. In the following sections the method is described and validated on simulated and real EEG data.

## 4.1 The Method

The estimation of a forward model relies in our approach on a low-dimensional representation of head anatomy and physiology build using a corpus of forward models. The parametrization of the corpus of forward models is outlined in Fig.

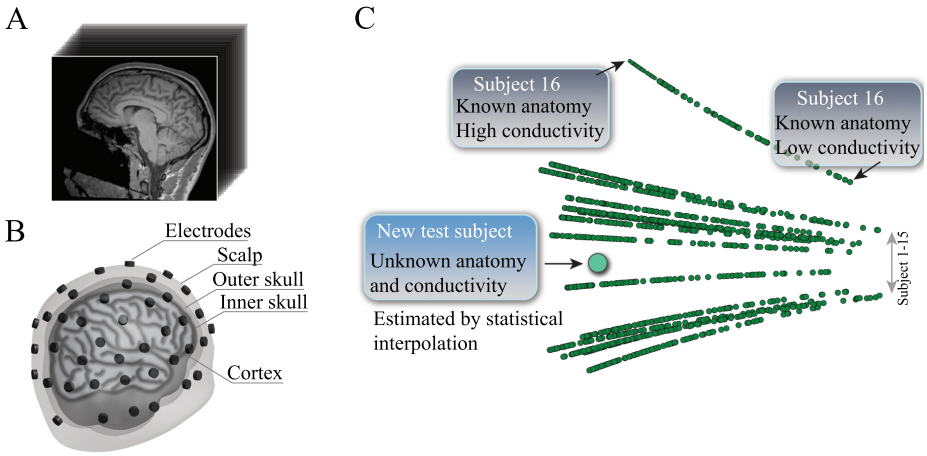



**Figure 4.1:** Forward model representation. (A) Structural scans of the corpus subjects. (B) The segmentations of brain, skull and scalp are used to construct a three-layered BEM head model. (C) For each subject 100 forward models are created each with skull:brain conductivity ratios between 1:250 to 1:15. The forward model corpus is represented using PCA with the locations of the forward models illustrated as green dots. Each semi-horizontal line corresponds to one subject. [From H].

4.1. The structural scans of each subject (Fig. 4.1A) are segmented into relevant head layers (Fig. 4.1B) from which forward models are constructed. Several replicates of forward models are generated for each subject, each incorporating different conductivity ratios of the head compartments. As previously noted the skull has a much lower conductivity than the other head layers and is therefore expected to greatly affect the accuracy of the forward model [Lew et al., 2009]. Research has consequently been widely focused on this tissue and we also suggest to in particular examine the effect of the skull conductivity.

A low-dimensional representation of the forward model corpus is obtained by principal component analysis (PCA) [Jolliffe, 2002]. The constructed forward models are thus collectively decomposed using PCA and projected onto a subset of the new basis. This step is exemplified in Fig. 4.1C where the two principal components explaining most of the forward model variance form the basis. Here two principal components were chosen for visualization purposes, the optimal number of components can, e.g., be inferred based on the generalization error [Hansen et al., 1999].

As PCA is a generative model new forward models can be simulated essentially

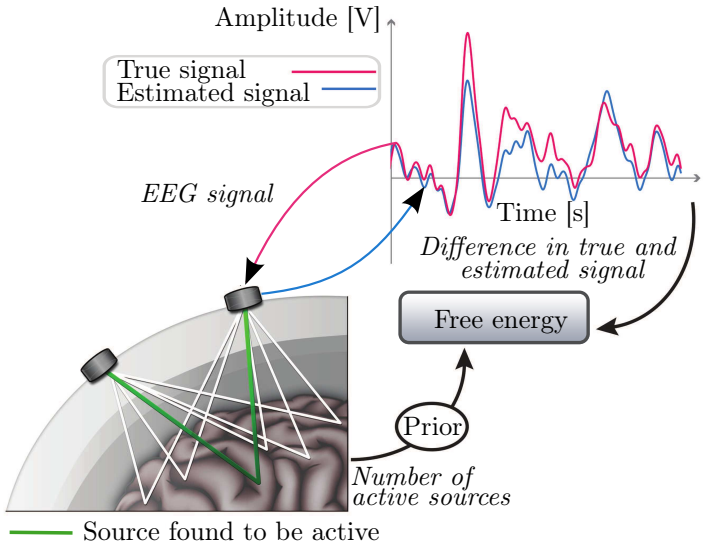



**Figure 4.2:** The free energy; a bound on the model evidence. Minimizing the free energy provides a tradeoff between the goodness of fit and the number of non-zero sources. [From H].

by interpolation of the forward models contained in the corpus. We use the free energy of teVG in Eq. (3.8) to infer the optimal forward model for a new subject. Minimizing the free energy corresponds to maximizing the lower bound on the model evidence and should thus indicate the source configuration and best suited forward model for the EEG provided. The free energy, as seen in Fig. 4.2, consists of 1) the goodness of fit between the EEG signal and the EEG signal as predicted by the model priors and the forward model. And 2) a penalization term proportional to the degree of prior sparsity and effective number of non-zero sources, thus promoting sparse relative to dense solutions.

In summary a forward model is inferred for a new subject by optimizing the free energy in the low-dimensional PCA representation. The EEG data of the new subject and the forward model parametrization thus provides a data driven approach to infer both a person-specific forward model and the EEG generators.

## 4.2 Evaluation

The following PCA representation was used in validating the proposed method on both simulated and real EEG data. For visualization purposes we created a two-dimensional parametrization.

The corpus of forward models was generated as in Section 3.1 using the sMRI scans of the 16 subjects in the multimodal neuroimaging dataset. The T1-weighted 1 mm isotropic structural scans seen in Fig. 4.1A were acquired using a Siemens 3T Tim Trio MRI system [Wakeman and Henson, 2015]. The 70 electrodes and 50-100 head shape points were mapped by a 3D digitizer relative to three fiducials: the left and right pre-auricular points and the nasion. For further acquisition details we refer to [Wakeman and Henson, 2015].

The structural scans were in SPM8 segmented into brain, skull and scalp from which BEM head models [Phillips, 2000] were build, see Fig. 4.1B. The brain mesh was created by affine transformation of a template mesh to fit the sMRI of each subject [Ashburner and Friston, 2005]. The forward models were generated with SPM8 template conductivity values for the brain and scalp, both corresponding to 0.33 S/m, while the skull:soft tissue conductivity ratio was drawn from a uniform distribution of 1:250 to 1:15. This interval of ratios was constructed from the literature values in [Stahlhut et al., 2011; Oostendorp et al., 2000]. The head model of each subject was combined with 100 samples of skull:brain (i.e. skull:soft tissue) conductivity ratios from which 100 forward models were build. In total 1,600 forward models were constructed of which 49 were judged as being outliers based on their high $l_2$-norm[1], these were excluded from further analysis.

The vectorized forward models were placed as rows of a $1{,}551 \times 70 \cdot 8{,}196$ sized matrix from which the average forward model was computed and subtracted. PCA using eigenvalue decomposition of this matrix provided the principal components, which were sorted according to the variance they explained. The two first principal components contained 73 % of the total forward model variability and were chosen to form the linear basis of the low-dimensional PCA representation. The curve describing the cumulative sum over explained variance is in Fig. 4.3 seen to be quite steep with 25 components explaining 99.9% of the variance.

The described forward model inference procedure was tested in a leave-one-out setup, where one subject was excluded from the forward model corpus. The PCA representation was thus build on 15 of the 16 subjects and a forward

[1] The $l_2$-norm was computed on the vectorized forward models.

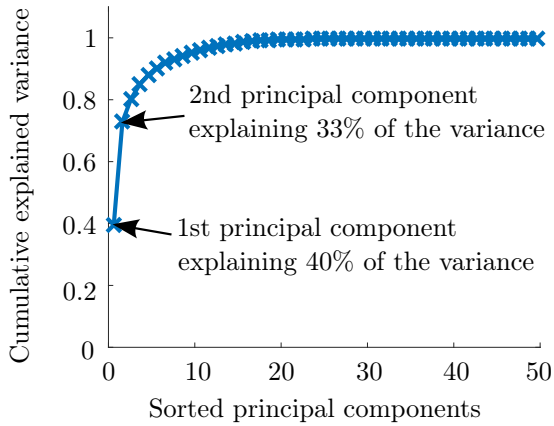



**Figure 4.3:** The cumulative sum of the explained variance of the principal components. The principal components are sorted in decreasing order relative to how much variance they explain. Only the first 47 principal components are shown, combined these explain $(100-2^{-52})\%$ of the variance.

model of the held-out subject was inferred.

The projection of the forward models to the constructed PCA space is seen in Fig. 4.1C. The first principal component was dominated by the skull:brain conductivity ratio; each semi-horizontal line in Fig. 4.1C thus consists of forward models belonging to one subject with decreasing conductivity from left to right. Fig. 4.1C thus confirms that the degree of variability among forward models is greatly influenced by the skull:brain conductivity ratio.

Subject-specific features of the head geometry also influenced the decomposition. It was tested whether the brain, skull or head size explained the vertical position of the subjects in Fig. 4.1C. However only small correlations were seen. Subject 16 is an outlier in Fig. 4.1 and analysis of the subjects' brain sizes revealed that subject 16 indeed had the biggest brain by far (Fig. 4.4). However we did not find a linear correspondence between the brain sizes and the vertical locations of the subjects in the PCA space. More complex factors of the head anatomy must therefore be contributing to the variability.

The matrix properties of real and PCA predicted forward models are presented in Table 4.1. The two types of forward models shared similar matrix coherence [Donoho et al., 2006] and condition number [Belsley et al., 2005]. Both had high coherence and condition number indicating that the forward fields in general are highly correlated and that the forward models are ill-conditioned. Particularly

| | Real | PCA predicted |
|---|---|---|
| $1-\text{coh}$ [$\times 10^{-4}$] | 2.7 (0.95 - 5.5) | 2.6 (1.0 - 4.1) |
| $\kappa$ | 97.3 (59.2 - 795.6) | 109.4 (61.5 -281.1) |

**Table 4.1:** Median (and interval) of the matrix properties of the forward models, where coh is the coherence and $\kappa$ is the condition number. As the coherence of all forward models approached 1, we show 1 minus the coherence. [From H].

high condition numbers were found for the forward models of subject 16 (see Fig. 4.1C). Without this subject the maximal condition number of the real forward models was 150.9.

Skull:brain conductivity ratio impacted the matrix coherence of the forward models as seen in Fig. 4.5. The coherence approached 1 as the ratio numerically decreased, meaning that decreasing the skull conductivity relative to the brain conductivity increases the ill-posedness of the problem. In contribution [C] we furthermore showed that the source localization error in general increases with decreasing skull conductivity.

### 4.2.1 Simulations

The simulation study was divided into two parts. In the first study we investigated whether the free energy could recover the relevant forward model from the corpus of real forward models. In the second study the PCA representation was optimized with respect to the free energy to predict new forward models. Note that in the following the localization error was calculated as the Euclidean distance between each true source and the estimated source which had maximal activation in the hemisphere of the true source. The reported error is a sum across the planted sources.

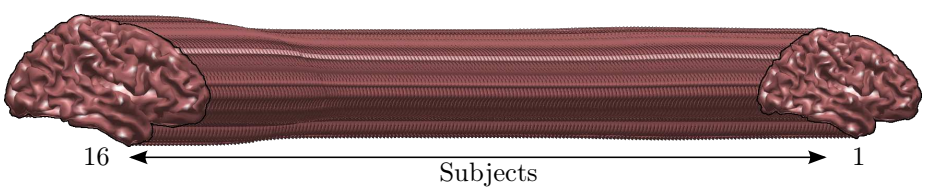



**Figure 4.4:** The brain meshes of the subjects smoothed/interpolated relative to their vertical PCA locations (from top to bottom) in Fig. 4.1C. [From H].

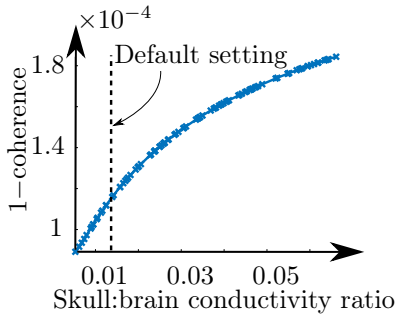



**Figure 4.5:** The coherence of the BEM forward models as a function of skull:brain conductivity ratio. All forward models approached a coherence of 1, for visualization purposes we therefore show 1 minus the coherence. The default setting of the ratio in SPM8 is indicated by a dashed line. [From C].

In the first study we simulated two posteriorly located sources (Fig. 4.6A) having synchronous sinusoidal temporal dynamics (Fig. 4.6B). The constructed source distribution was projected to sensor space using a forward model from the test subject (black cross in Fig. 4.7B and 4.8) and corrupted with noise to an SNR of 5 dB.

The sparsity levels were estimated using four fold cross-validation where the electrodes were divided into folds such that the electrodes in each fold were evenly spread across the scalp (Fig. 4.7A). The cross-validation mean squared error (MSE) determined the optimal setting of the sparsity for each of the forward

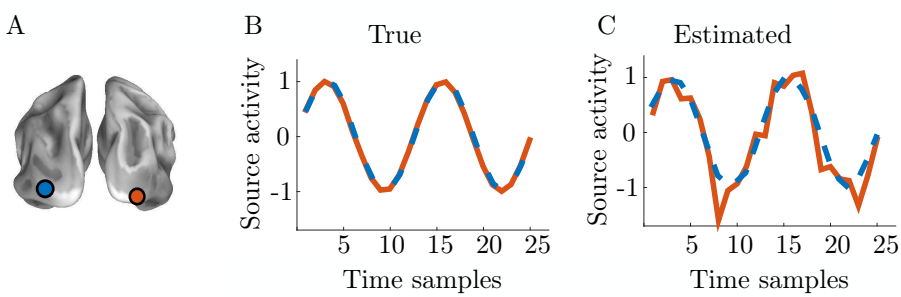



**Figure 4.6:** The planted and estimated activity. (A) The inflated brain showing the locations of the two planted sources (posterior view). The estimated sources completely overlapped with these. (B) The real and (C) estimated time courses of the two sources. [From H].

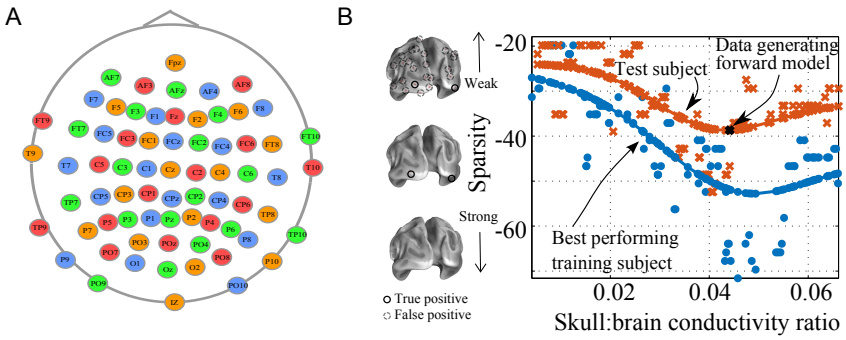



**Figure 4.7:** Sparsity estimation. (A) The partitioning of the 70 electrodes into four folds, each designated one color. (B) Four fold cross-validation was used to estimate the sparsity levels for the forward models belonging to the test subject (red crosses), the best $F_1$-measure performing non-test subject (blue dots), and the remaining subjects (not shown). The data generating forward model is indicated by a black cross. The estimated sparsity was smoothed relative to their conductivity ratios (full lines). [From H]

models, shown in Fig. 4.7B for the test subject (red) and the training subject having best $F_1$-measure (blue). To reduce the noise introduced by the coarse 10-step cross-validation procedure the obtained sparsity levels were smoothed as a function of the skull:brain conductivity ratio within each subject. The forward models combined with their smoothed sparsity values were applied to reconstruct the simulated source distribution in Fig. 4.8.

As seen in Fig. 4.8A the forward models belonging to the test subject with skull:brain conductivity ratios close to the true value obtained lowest free energy among all forward models and additionally had low cross-validation error (Fig. 4.8B). The location of the minimum free energy also coincided with optimal localization error (Fig. 4.8C) and $F_1$-measure (Fig. 4.8D). The reconstructed source locations corresponded exactly to the planted and had similar temporal dynamics as seen in Fig. 4.6C. We thus conclude that the free energy is a viable measure to predict performance of the forward models. We furthermore note that when the geometry is known but the skull:brain conductivity ratio is misspecified the summed localization error of the two sources is up to 30 mm, i.e. on average up to 15 mm (Fig. 4.8C).

In the second simulation study the inferred PCA forward models were validated. We simulated the EEG for this study by projecting the source distribution seen in Fig. 4.9 to sensor space using a forward model from the test subject. Again

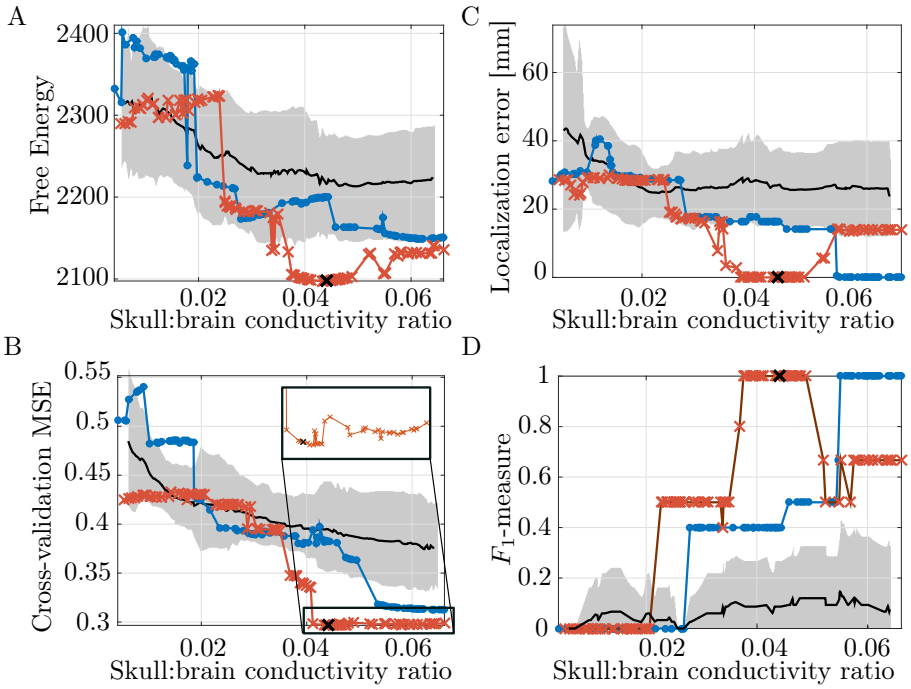



**Figure 4.8:** Reconstruction of simulated data using 'real' forward models. A forward model of the subject (black cross) was used to project the simulated sources in Fig. 4.6A-B to sensor space and noise was added to yield SNR = 5 dB. The teVG was applied to the smoothed sparsity levels (Fig. 4.7B) in combination with the corpus forward models. The performance of the test subject (red), the training subject with highest $F_1$-measure (blue) and the average across all training subjects (black with s.d. in grey) is shown. (A) The free energy. (B) The cross-validation MSE with zoom-in plot. (C) The summed localization error. (D) The $F_1$-measure. [From H].

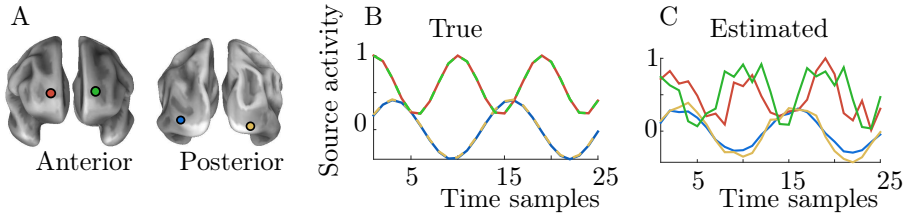



**Figure 4.9:** The planted and estimated activity in the study of inferred PCA forward models. (A) The locations of the four planted sources. The estimated sources completely overlapped with these. (B) The real and (C) estimated time courses of the four sources. [From H].

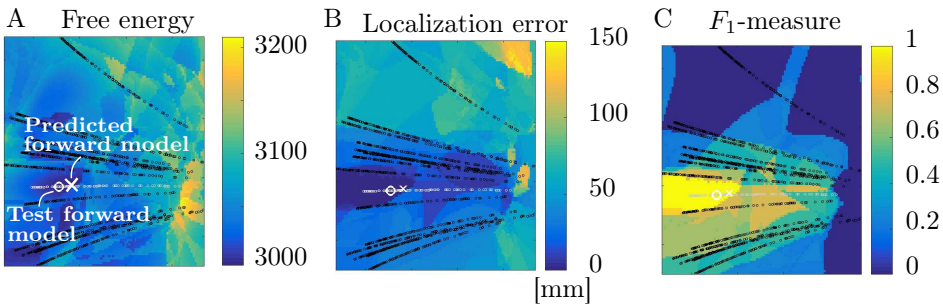



**Figure 4.10:** Inference of a forward model based on the simulated data of Fig. 4.9 with SNR = 5 dB. (A) The free energy, (B) the localization error and (C) the $F_1$-measure, calculated in the 2D PCA space spanned by the training forward models. The forward models of the test and training subjects are shown in white and black, respectively. [From H].

noise was added to yield an SNR of 5 dB. The previously described electrode cross-validation setup was used to estimate the sparsity levels for 250 training forward models. The sparsity levels were subsequently interpolated to cover the entire PCA space using kernel regression [Nadaraya, 1964] with a Gaussian kernel of width 1. The 2D PCA representation was generated as described earlier and importantly with all of the forward models of the test subject excluded from the decomposition (Fig. 4.10).

The PCA forward model with minimum free energy, white cross in Fig. 4.10A, was located close to the data generating forward model. Again optimal free energy was co-located with optimal localization error and $F_1$-measure (Fig. 4.10B-C). The estimated sources were thus located in the true source locations and had comparable temporal dynamics to these (Fig. 4.9C). The performance of

**Table 4.2:** Performance of forward models on the simulated data in Fig. 4.9. The investigated forward models comprise the inferred (white cross in Fig. 4.10A) and template/subject-specific forward models. The fourth forward model corresponds to the data generating forward model and $\sigma$ =skull:brain conductivity ratio. [From H].

| Forward models | Inferred from free energy | Template MRI, template $\sigma$ | Subject MRI, template $\sigma$ | Subject MRI, true $\sigma$ |
|---|---|---|---|---|
| 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 |

the optimal PCA forward model is in Table 4.2 compared to that of template and subject-specific forward models. The performance of the forward model with true geometry and skull:brain conductivity ratio was comparable to that of the free energy inferred forward model. Lowest performance was obtained by the template forward model with which non of the true sources were recovered.

In contribution [H] we repeated the leave-one out setup for five more subjects. These were in separate turns removed from the PCA representation and used as test subjects. A reasonable forward model could in general be inferred when the data generating forward model was contained in the span of training forward models. For three subjects the performance of the inferred forward models was comparable to using the subject-specific forward model. For the fourth subject one of the sources were not recovered and the fifth subject was the outlier in Fig. 4.1, i.e. subject 16. Indeed only one of the planted sources was recovered for this 'outlier subject', indicating the need for a larger forward model corpus.

### 4.2.2 Real EEG Data

Finally, the forward model inference pipeline was tested on the differential face ERP [Wakeman and Henson, 2015] described in section 3.1. We maintained the same test subject as used in the above simulation studies and in Section 3. The inference procedure followed the steps of the second simulation study.

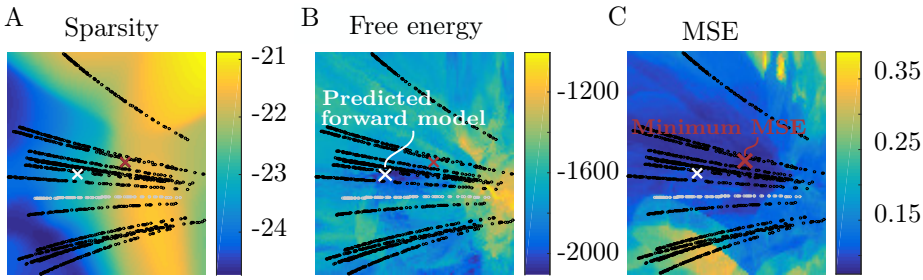



**Figure 4.11:** Inference of a forward model based on the face-evoked EEG response. (A) The averaged sparsity profile. B) The averaged free energy and (C) the averaged cross-validation MSE based on the 12 sparsity profiles. The PCA locations of the forward models belonging to the test and training subjects are shown in white and black, respectively. The 2D PCA space was based on the training forward models. [From H].

However, instead of fixing the smoothing parameter for the sparsity interpolation, the bandwidth of the smoothing kernel was assigned a uniform prior and marginalized numerically in 12 steps from 1/4 to 3. The free energy was thus calculated for each of the 12 sparsity profiles and evaluated on the average of these (Fig. 4.11).

The minimum of the averaged free energy (Fig. 4.11B) was found in a location with low cross-validation MSE (Fig. 4.11C). The inferred forward model combined with teVG gave the source distribution seen in Fig. 4.12. The response peaked 160 ms after stimulus onset, corresponding to the face-related N170 EEG component, in sources located near the O/FFAs. We recovered similar locations and time courses when using the forward model derived from the subject's MRI scan and template conductivity ratios (see supplementary material of contribution [H]).

## 4.3 Relation to Other Work

Studies have shown that uncertainties of the forward model is an important limiting factor in achieving EEG imaging with high spatial specificity [Oostenveld and Oostendorp, 2002; Plis et al., 2007; Lew et al., 2007, 2009; Stahlhut et al., 2012; Akalin Acar and Makeig, 2013; Windhoff et al., 2013; Aydin et al., 2014]. Our results add quantitatively to this view, especially regarding the skull:brain conductivity ratio where we saw the implications of its misspecification.

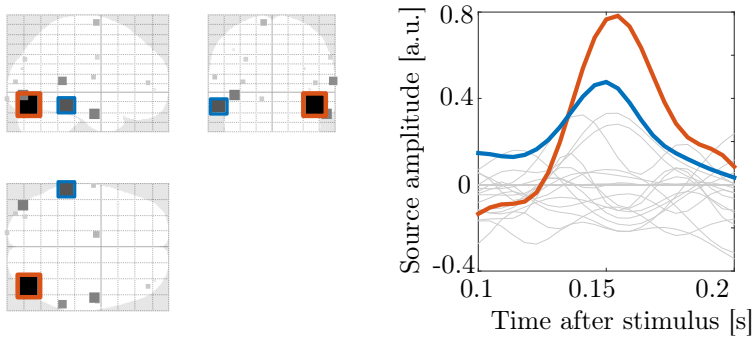



**Figure 4.12:** The recovered sources of the face-evoked ERP using the free energy predicted forward model. (A) The locations of the estimated activity. Source magnitudes are illustrated by the darkness and size of the squares; with large magnitude sources indicated by dark and big squares. (B) The time courses of the strongest (red) and second strongest magnitude source (blue). The remaining recovered sources are shown in grey. [From H].

Whether the skull conductivity can be estimated from the EEG data has been discussed by Plis et al. who concluded it unfeasible to estimate both the skull conductivity and the source locations [Plis et al., 2007]. This finding was partly based on the inability of the goodness of fit between true and predicted EEG signals to recover a meaningful skull conductivity (figure 8 in [Plis et al., 2007]). Although not clearly stated in [Plis et al., 2007] the performance measure was not based on unbiased test estimates perhaps explaining why source locations and conductivity could not be estimated simultaneously. Lew et al. in contrast showed that the goodness of fit can be used to estimate both source locations and the skull:brain conductivity ratio as well as the brain conductivity [Lew et al., 2009]. This was shown in low noise settings, however reconstruction deteriorated for larger noise levels, probably caused by overfitting as in [Plis et al., 2007]. An important dissimilarity of the two studies lies in the inference framework, while Lew et al. only used the EEG data to find the source strengths, Plis et al. also used it to estimate the source locations. These were in [Lew et al., 2009] found by simulated annealing. Compared to [Plis et al., 2007] the source estimates in [Lew et al., 2009] were thus 'more independent' of the EEG data which was also used to infer the conductivity configuration.

Other techniques for inferring the conductivity ratios have more recently been suggested. Akalin Acar et al., for example, specifically optimized for the skull:brain conductivity ratio while also estimating the EEG generators in their 'simultaneous tissue conductivity and source location estimation' (SCALE) [Akalin Acar et al., 2016]. In the optimization scheme SCALE uses a local linearization of

the forward model with respect to the skull:brain conductivity ratio [Gençer and Acar, 2004]. Performance is based on the compactness of the reconstructed independent component (IC) scalp maps and the goodness of fit between the predicted and actual IC scalp maps. The performance of the forward models is contingent on having high quality structural scans from which the head compartments can be derived.

Stahlhut et al. proposed a method, called SOFOMORE, that also approximates both a forward model and a source distribution. SOFOMORE models the uncertainty of the forward model by a multivariate Gaussian distribution [Stahlhut et al., 2011]. The mean of this distribution is defined as the conventionally build forward model with a precision parameter for each forward field estimated based on the EEG data. SOFOMORE thus relies as SCALE on person-specific anatomical detail for initialization. In practice the new forward model is only mildly adapted from the mean, primarily in regions of the estimated activity. Furthermore, the forward fields are modeled as being statistically independent and does therefore not optimize the forward model as one entity as SCALE or our proposed method.

An analog to the free energy we have presented was used in [Henson et al., 2009b] to compare a small set of forward models. Henson et al. showed that basing the forward model on anatomical information is far superior to a template forward model. The method does however not provide the possibility of interpolating between forward models and thus relies on the optimal model being in the set of proposed forward models and consequently requires anatomical information.

Our method is, in contrast to [Akalin Acar et al., 2016; Stahlhut et al., 2011; Henson et al., 2009b], data driven and does not require any structural information. The proposed forward model inference pipeline can therefore be applied when MRI or CT scans are unavailable. Research or clinical facilities where access to scanners is limited or non-existing can therefore benefit from such an inference scheme. Furthermore, the suggested framework enables subject-specific forward models for children, or people in general with difficulty of being in a scanner, e.g., because of pain or claustrophobia. If a structural scan is available the free energy can be optimized with respect to the compartment conductivities such that the forward model parametrization depends on conductivities alone and assume the head geometry known.

A further advantage of potentially obsoleting structural scans is the construction of an adaptive forward model. Rice et al. showed that the EEG signal is dependent on subject posture because of brain shifting allowed by the CSF layer [Rice et al., 2013]. A forward model derived from head scans of a subject lying down in an MRI scanner might therefore be ill-fitted for reconstructing sources of EEG recorded while the subject was sitting or standing.

Chapter 5

# Multimodal Integration of EEG and fMRI

When the electrophysical and hemodynamic brain signals are both modulated by a stimulus a synergistic effect can arise from fusing EEG and fMRI and thereby advance our understanding of the brain [Laufs, 2012]. Decomposition of electrophysical and hemodynamic signals into correlated components is an important step towards providing highly detailed information on the temporal and spatial scale. The multimodal source power comodulation (mSPoC) [Dähne et al., 2013] provides decomposition by extracting fMRI components which are maximally correlated with the power dynamics of the extracted EEG components. We suggest to augment the coupling recovered by mSPoC to exploit both the functional *and* anatomical information of the modalities. We extend the method with EEG and fMRI in mind and assume that there is a correspondence between the locations of the correlated EEG and fMRI components. An early multimodal study [Singh et al., 1998] motivates a spatial link between EEG and fMRI by showing the EEG sources of alpha activity to be located in areas of fMRI activity.

The following presented method builds on the work performed by Dähne et al. [Dähne et al., 2014, 2013] and was accomplished through collaboration with these authors. The framework was first formulated in the precursor to mSPoC, i.e., the Source Power Comodulation (SPoC) method [Dähne et al., 2014]. SPoC extracts components from EEG signals having power dynamics which are mod-

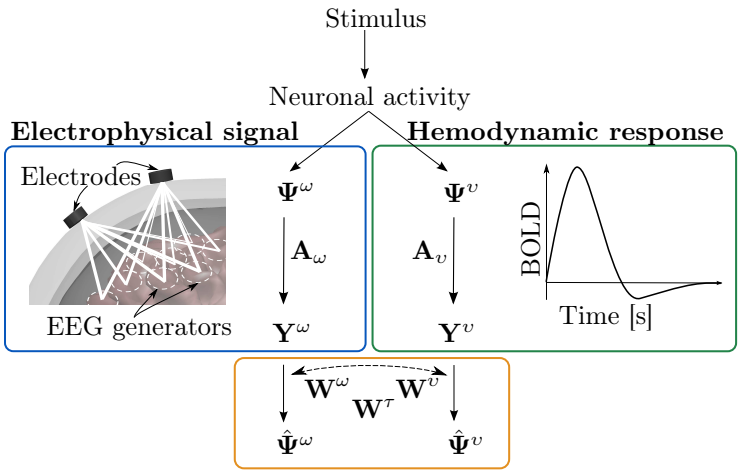



**Figure 5.1:** The generative model and the mSPoC derived components. A stimulus generates a brain response which in turn elicits both electrophysical ($\mathbf{\Psi}^{\omega}$) and hemodynamic components ($\mathbf{\Psi}^{\upsilon}$). These signals are mapped to their respective sensor spaces ($\mathbf{Y}^{\omega}/\mathbf{Y}^{\upsilon}$) through a forward model ($\mathbf{A}^{\omega/\upsilon}$). Maximally correlated EEG and fMRI source components ($\hat{\mathbf{\Psi}}^{\omega/\upsilon}$) are extracted by mSPoC using spatial activation filters ($\mathbf{W}^{\omega/\upsilon}$) and a temporal filter ($\mathbf{W}^{\tau}$) modeling the time delay between EEG and fMRI.

ulated by a target variable, e.g., changes in auditory intensity [Dähne et al., 2014]. SPoC thus facilitates a way to isolate the neuronal response to a certain stimulus. The target stimuli are in mSPoC substituted by another imaging modality such that the extracted EEG component has power dynamics which comodulate with, e.g., a component of a hemodynamic signal [Dähne et al., 2013].

## 5.1 The Method

Before describing the proposed augmentation a short introduction to mSPoC is provided for completeness. The mSPoC adopts the model framework illustrated in Fig. 5.1. We denote the EEG and fMRI related variables by superscript $\omega$ and $\upsilon$, respectively. In general superscripts are used to indicate variable relations, while we as previously reserve subscripts for referring to elements or structures/dimensions of the variables.

### 5.1.1 The Generative Model

The assumed generative models of the EEG and fMRI signals are defined as

$$\mathbf{Y}^{\omega} = \mathbf{A}^{\omega}\mathbf{\Psi}^{\omega} + \mathbf{E}^{\omega}, \tag{5.1}$$

$$\mathbf{Y}^{\upsilon} = \mathbf{A}^{\upsilon}\mathbf{\Psi}^{\upsilon} + \mathbf{E}^{\upsilon}, \tag{5.2}$$

where $\mathbf{Y}^{\omega} \in \mathbb{R}^{K^{\omega}\times T^{\omega}}/\mathbf{Y}^{\upsilon} \in \mathbb{R}^{K^{\upsilon}\times T^{\upsilon}}$ are the EEG/fMRI signals as mapped by the temporal dynamics of the $C$ source components, $\mathbf{\Psi}^{\omega/\upsilon} \in \mathbb{R}^{C^{\omega/\upsilon}\times T^{\omega/\upsilon}}$, and the spatial activation patterns of these components, $\mathbf{A}^{\omega/\upsilon} \in \mathbb{R}^{K^{\omega/\upsilon}\times C^{\omega/\upsilon}}$. The generative models are thus very similar to that of Eq. (3.2) but with the important difference that the component space is now smaller than the sensor space, i.e. $C^{\omega/\upsilon} < K^{\omega/\upsilon}$. $\mathbf{E}$ contains the activity not explained by the source components.

### 5.1.2 The mSPoC Framework

When the EEG and fMRI are the result of the same stimulus pattern mSPoC assumes that certain temporal dynamics are shared between the two modalities. The mSPoC thus extracts an EEG component ($\hat{\boldsymbol{\psi}}^{\omega}$) which have bandpower dynamics ($\hat{\boldsymbol{\phi}}$) that are correlated with an fMRI component ($\hat{\boldsymbol{\psi}}^{\upsilon}$). Note that a 'hat' on a variable indicates that this is an estimate of the true variable. In the following it is described how to extract the component pair with highest correlation. However, the procedure of finding the $C \leq \min(C^{\omega}, C^{\upsilon})$ coupled components is the same when the EEG and fMRI signals are deflated with respect to the previous pairs of components [Householder, 2013], i.e., recovered component pairs are projected out of the data before a new pair is extracted.

In order to integrate the two modalities the EEG must be aligned with the more slowly varying fMRI, the EEG is thus divided into epochs each having the duration of the fMRI sampling interval (TR). The EEG in epoch $l = 1..L$ is in the following denoted $\mathbf{Y}_l^{\omega}$ and has the covariance $\mathbf{\Sigma}_l^{\omega}$, while the covariance of the entire EEG signal is denoted $\mathbf{\Sigma}^{\omega}$.

Next, the extraction of EEG and fMRI source components is defined

$$\hat{\boldsymbol{\psi}}^{\omega} = (\mathbf{w}^{\omega})^{\top}\mathbf{Y}^{\omega}, \tag{5.3}$$

$$\hat{\boldsymbol{\psi}}^{\upsilon} = (\mathbf{w}^{\upsilon})^{\top}\mathbf{Y}^{\upsilon}, \tag{5.4}$$

where $\mathbf{w}^{\omega/\upsilon} \in \mathbb{R}^{K^{\omega/\upsilon}}$ are so-called spatial filters. The power of the EEG component in epoch $l$ is approximated by the epoch-wise component variance, i.e.,

$\hat{\phi}_l = (\mathbf{w}^\omega)^\top \mathbf{\Sigma}_l^\omega \mathbf{w}^\omega$. Finally, a temporal filter $\mathbf{w}^\tau \in \mathbb{R}^{K^\tau}$ is employed to model possible time delays between $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\psi}}^\upsilon$. The $K^\tau$th order finite impulse response filter is applied to the EEG component's bandpower dynamics in epoch $l$ by

$$h(\hat{\phi}_l) = \sum_{k=0}^{K^\tau-1} w_{k+1}^\tau \cdot \hat{\phi}_{l-k} \tag{5.5}$$

$$= (\mathbf{w}^\omega)^\top \left( \sum_{k=0}^{K^\tau-1} w_{k+1}^\tau \cdot \mathbf{\Sigma}_{l-k}^\omega \right) \mathbf{w}^\omega = (\mathbf{w}^\omega)^\top \mathbf{\Sigma}_l^h \mathbf{w}^\omega, \tag{5.6}$$

where $\mathbf{\Sigma}_l^h = \sum_{k=0}^{K^\tau-1} w_{k+1}^\tau \cdot \mathbf{\Sigma}_{l-k}^\omega$. The temporally filtered power time series $(h(\hat{\boldsymbol{\phi}}))$ can be re-expressed such that optimization with respect to $\mathbf{w}^\tau$ is simplified. Defining the rows of the matrix $\hat{\Phi} \in \mathbb{R}^{K^\tau \times L}$ to contain $\hat{\boldsymbol{\phi}}$ temporally delayed by 1 through $K^\tau$ epochs we can now define $h(\hat{\boldsymbol{\phi}}) = \hat{\Phi}^\top \mathbf{w}^\tau$.

Using the above definitions the mSPoC objective is to extract an EEG and fMRI component pair such that the temporally filtered bandpower dynamics of the EEG component are maximally correlated with the dynamics of the fMRI component, i.e.,

$$\max_{\mathbf{w}^\omega, \mathbf{w}^\tau, \mathbf{w}^\upsilon} \operatorname{Cov}\left(h(\hat{\boldsymbol{\phi}}), \hat{\boldsymbol{\psi}}^\upsilon\right) \text{ s.t. } \operatorname{Var}\left(\hat{\boldsymbol{\psi}}^\omega\right) = \operatorname{Var}\left(h(\hat{\boldsymbol{\phi}})\right) = \operatorname{Var}\left(\hat{\boldsymbol{\psi}}^\upsilon\right) = 1. \tag{5.7}$$

Optimization proceeds by alternating between optimizing for $\mathbf{w}^\tau$ and $\mathbf{w}^\upsilon$ solving a canonical correlation analysis (CCA) problem, and optimizing for $\mathbf{w}^\omega$ solving a generalized eigenvalue problem as illustrated in Algorithm 3.

**Algorithm 3:** The framework of the mSPoC algorithm.

**Input:** EEG ($\mathbf{Y}^\omega$) and fMRI ($\mathbf{Y}^\upsilon$) signal.
**Output:** $\mathbf{w}^\upsilon, \mathbf{w}^\tau$ and $\mathbf{w}^\omega$ yielding maximally correlated components.
**for** *each repetition* **do**
  Initialize $\mathbf{w}^\omega$ randomly;
  **while** *not converged* **do**
    Calculate $\Phi$;
    Solve $[\mathbf{w}^\upsilon, \mathbf{w}^\tau] = \text{CCA}\,(\mathbf{Y}^\upsilon, \Phi)$;
    Calculate $\mathbf{\Sigma}^{h\upsilon} = \sum_l \hat{\psi}_l^\upsilon \cdot \mathbf{\Sigma}_l^h = \sum_l \hat{\psi}_l^\upsilon \cdot \sum_k w_{k+1}^\tau \cdot \mathbf{\Sigma}_{l-k}^\omega$;
    Solve for $\mathbf{w}^\omega$ in the generalized eigenvalue problem $\mathbf{\Sigma}^{h\upsilon}\mathbf{w}^\omega = \lambda \mathbf{\Sigma}^\omega \mathbf{w}^\omega$;
    Calculate $h(\boldsymbol{\phi})$ and $\hat{\boldsymbol{\psi}}^\upsilon = (\mathbf{w}^\upsilon)^\top \mathbf{Y}^\upsilon$;
    Calculate correlation of $h(\boldsymbol{\phi})$ and $\hat{\boldsymbol{\psi}}^\upsilon$ and check for convergence;
  **end**
  Store correlation along with $\mathbf{w}^\upsilon, \mathbf{w}^\tau$ and $\mathbf{w}^\omega$;
**end**

### 5.1.3 Spatial Augmentation

The mSPoC coupling is as described based on a functional relation between the bandpower dynamics of the EEG component and the dynamics of the fMRI component. We propose to extend this relation to include a prior belief of similar anatomical locations of the correlated components. This is achieved by correlating the spatial patterns of the EEG and fMRI components. The spatial patterns can be approximated based on the spatial filters, i.e. $\hat{\mathbf{a}}^{\omega} \propto \mathbf{\Sigma}^{\omega}\mathbf{w}^{\omega}$ and $\hat{\mathbf{a}}^{\upsilon} \propto \mathbf{\Sigma}^{\upsilon}\mathbf{w}^{\upsilon}$ [Haufe et al., 2014]. Now, in order to correlate the spatial patterns of the fMRI and EEG they must be brought to the same domain. We choose to work in the EEG sensor domain and thus project the spatial pattern of the fMRI to this space using a forward model, $\mathbf{A} \in \mathbb{R}^{K^{\omega} \times K^{\upsilon}}$, as found by solving the EEG forward problem. Thus $\hat{\mathbf{a}}^{\upsilon}$ in the EEG sensor domain becomes $\tilde{\mathbf{a}} = \mathbf{A}\hat{\mathbf{a}}^{\upsilon}$.

The spatial pattern of the EEG is now set to approach the projected fMRI pattern, i.e., $\hat{\mathbf{a}}^{\omega} \rightarrow \tilde{\mathbf{a}}$, corresponding to $\mathbf{\Sigma}^{\omega}\mathbf{w}^{\omega} \rightarrow \tilde{\mathbf{a}}$. This can be achieved by following the technique described by the linearly constrained minimum variance (LCMV) beamformer [Van Veen et al., 1997]. As the name implies the variance is minimized, i.e., $\min \mathrm{Var}(\hat{\boldsymbol{\psi}}^{\omega}) = \min \left((\mathbf{w}^{\omega})^{\top}\mathbf{\Sigma}^{\omega}\mathbf{w}^{\omega}\right)$, under the linear constraint $(\mathbf{w}^{\omega})^{\top}\tilde{\mathbf{a}} = 1$. We modify this expression to facilitate incorporation into the mSPoC method such that we instead maximize $((\mathbf{w}^{\omega})^{\top}\tilde{\mathbf{a}})^2$ and linearly constrain the variance,

$$\max_{\mathbf{w}^{\omega}} (\mathbf{w}^{\omega})^{\top}\mathbf{\Sigma}^{a}\mathbf{w}^{\omega} \quad \text{s.t.} \quad \mathrm{Var}\left(\hat{\boldsymbol{\psi}}^{\omega}\right) = 1, \tag{5.8}$$

where $\mathbf{\Sigma}^{a} = \tilde{\mathbf{a}}\tilde{\mathbf{a}}^{\top}$ is a rank-one matrix. Eq. (5.8) can be formulated as the generalized eigenvalue problem $\mathbf{\Sigma}^{a}\mathbf{w}^{\omega} = \lambda\mathbf{\Sigma}^{\omega}\mathbf{w}^{\omega}$ and incorporated into the mSPoC generalized eigenvalue problem seen in Algorithm (3),

$$((1-\gamma)\mathbf{\Sigma}^{h\upsilon} + \gamma\mathbf{\Sigma}^{a})\mathbf{w}^{\omega} = \lambda\mathbf{\Sigma}^{\omega}\mathbf{w}^{\omega}. \tag{5.9}$$

The parameter $\gamma \in [0, 1]$ now determines the weight given the functional information relative to the anatomical, where if $\gamma = 0$ the original mSPoC is recovered while if $\gamma = 1$ only anatomical information is used. We call the augmented version mSPoC$_{\gamma}$ and propose to find $\gamma$ using cross-validation.

We develop an extension to our augmentation which seek to avoid projecting noise in the fMRI pattern to the EEG space. The fMRI pattern that we project therefore only contains the single highest magnitude element of the fMRI pattern, i.e., $\hat{\mathbf{a}}^{\upsilon}$ has only one non-zero element. We term this version mSPoC$_{\bar{\gamma}}$.

When the applied forward model contains dipoles of free orientation, i.e., defining $\mathbf{A} \in \mathbb{R}^{K^{\omega} \times K^{\upsilon} \times 3}$, the projection of the fMRI pattern involves estimating the

projection orientation. The projection of the fMRI pattern's element $i$ to EEG space is computed as a linear combination of the three directions, i.e.,

$$\tilde{\mathbf{a}} = \alpha_1 \mathbf{A}_{(:,i,1)} \hat{a}_i^v + \alpha_2 \mathbf{A}_{(:,i,2)} \hat{a}_i^v + \alpha_3 \mathbf{A}_{(:,i,3)} \hat{a}_i^v = \tilde{\mathbf{A}} \boldsymbol{\alpha}, \tag{5.10}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{3\times 1}$, $\tilde{\mathbf{A}} \in \mathbb{R}^{K^\omega \times 3}$, and $\mathbf{A}_{(:,i,d)} \in \mathbb{R}^{K^\omega \times 1}$ is the forward field of $i$ in orientation $d = 1, 2, 3$. Note that if the fMRI pattern is not thresholded its projection, $\tilde{\mathbf{a}}$, will be a sum of each non-zero element's projection, c.f. Eq. (5.10); each with separately estimated $\boldsymbol{\alpha}$. For now we assume that the thresholding is performed.

We suggest to recover $\boldsymbol{\alpha}$ by minimizing the squared distance between $\tilde{\mathbf{a}}$ and $\hat{\mathbf{a}}^\omega$,

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}} ||\hat{\mathbf{a}}^\omega - \tilde{\mathbf{A}} \boldsymbol{\alpha}||_2^2. \tag{5.11}$$

The posed least squares problem in Eq. (5.11) has the closed-form solution

$$\boldsymbol{\alpha} = \left(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}}\right)^{-1} \tilde{\mathbf{A}}^\top \hat{\mathbf{a}}^\omega. \tag{5.12}$$

## 5.2 Evaluation

In validating our proposed methods we simulated two datasets which were 1) based on the forward model introduced in section 3.1 and 2) based on the forward model described below containing free orientation of the dipoles. The two datasets were created based on the generative models in Eq. (5.1)-(5.2). The simulated EEG had a sampling frequency of 200 Hz and the frequency range of interest corresponded to the alpha range, here defined as 8-12 Hz. The fMRI had a sampling frequency of 1 Hz, thus each EEG epoch contained 200 samples. The functional coupling between the fMRI component and the envelope of the EEG component consisted of a shared oscillation of maximally 1/4 Hz.

In addition to the paired component the EEG and fMRI had 14 and 499 background sources, respectively. The number of fMRI background sources were increased to 4999 in dataset 2 as the simulated fMRI had a much higher spatial resolution in this dataset. The EEG and fMRI were constructed as a weighted sum of the normalized source and background activity. The ratio of the weights from this linear combination of activity was used to estimate the SNR. The degree to which the anatomical information entered the inference scheme of mSPoC$_\gamma$ and mSPoC$_{\bar{\gamma}}$ was estimated by five-fold cross-validation.

The mSPoC versions were finally validated on a combined EEG and fMRI dataset generously provided by collaborator S. Dähne. In the study an auditory

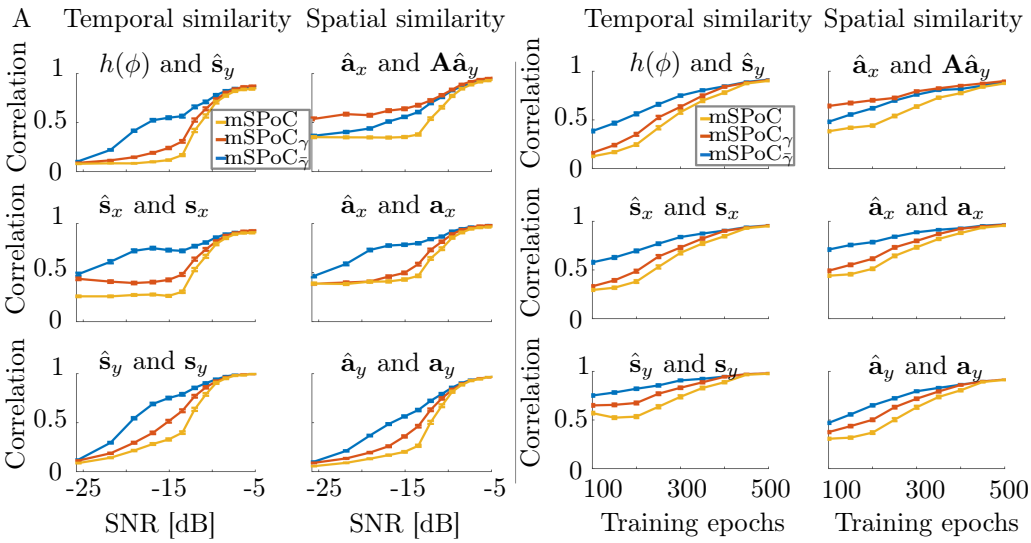



**Figure 5.2:** The performance of the original mSPoC (yellow), mSPoC$_\gamma$ (red) and mSPoC$_{\tilde{\gamma}}$ (blue) on dataset 1 as a function of (A) SNR, with 250 training epochs, and (B) number of trainings epochs, with SNR $= -12$ dB. The average across 500 repetitions is shown with the errorbars showing s.e.m. [From F].

stimulus signaled a subject to squeeze a ball with the right hand five times with a frequency of 1-2 Hz. In total 48 stimuli were administered. Simultaneous fMRI and EEG were recorded using a 3T MRI system (voxel size $= 3 \times 3 \times 4$ mm, TR $= 2$ s) and a 31 electrode EEG set. The data was preprocessed using SPM8 and MATLAB scripts provided by S. Dähne. The preprocessing reduced the number of EEG electrodes to 25 and the number of fMRI gray matter voxels to 54,590. For further details please refer to [Dähne, 2015]. A four-layered BEM forward model was constructed from an SPM8 segmentation of the subject's structural scans and by lead field computation done in OpenMEEG [Gramfort et al., 2010] using FieldTrip [Oostenveld et al., 2011]. The forward model mapped the 54,590 brain grid points to the 25 electrode and was generously provided by D. Miklody.

### 5.2.1 Simulated Dataset 1

We investigated the original and augmented versions of mSPoC in Fig. 5.2 with respect to the SNR and the number of epochs used for training.

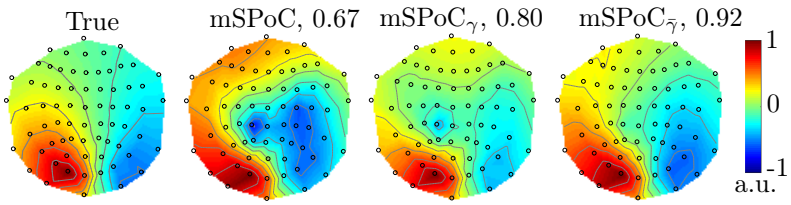



**Figure 5.3:** An example of the true and recovered spatial EEG patterns in dataset 1. Training was performed on 250 epochs, corresponding to 4 min, with SNR $= -12$ dB. The correlation between the true and estimated spatial patterns are indicated above the scalp maps. [From F].

The temporal similarity was calculated on a withheld test dataset of 600 epochs. The first row shows the correlation between the component pairs, the second and third rows show how the recovered EEG and fMRI components correlated with the true components.

For SNRs lower than $-10$ dB or training epochs less than 400 the augmented mSPoCs, especially mSPoC$_{\bar{\gamma}}$, outperformed the original mSPoC on all performance levels. The augmentation thus not only improved the correlation between the component pair but also their similarity with the true components. The mSPoC$_\gamma$ provided highest correlation of the EEG and fMRI projected spatial patterns. However, since these correlation values were not accompanied by high similarity between the estimated and true components, mSPoC$_\gamma$ might be overfitting the EEG pattern ($\hat{\mathbf{a}}^\omega$) to the noise components in the projected fMRI pattern ($\hat{\mathbf{a}}^\upsilon$).

Fig. 5.3 illustrates an example of the true and recovered spatial patterns of the EEG component as estimated by the mSPoC versions. The mSPoC$_{\bar{\gamma}}$ reconstructed the simulated scalp map with highest accuracy and since mSPoC$_{\bar{\gamma}}$ in general yielded better performance we investigated this version more closely. In Fig. 5.4 we thus evaluated for mSPoC$_{\bar{\gamma}}$ what effect the anatomical information has on the temporal similarity of the extracted component pair. There was a general tendency that lower SNRs required higher anatomical information. Furthermore, in regimes of good SNR too much anatomical information deteriorated the solution probably caused by the thresholding of $\hat{\mathbf{a}}^\upsilon$, as the same effect was not seen with mSPoC$_\gamma$ (not shown).

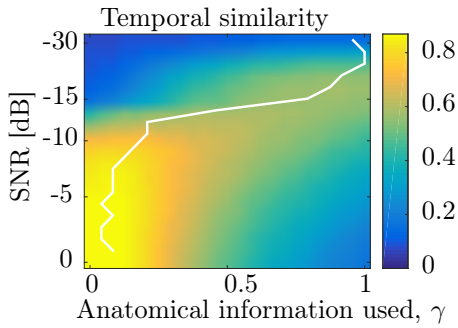



**Figure 5.4:** The temporal similarity between $h(\hat{\boldsymbol{\phi}})$ and $\hat{\boldsymbol{\psi}}^{v}$ as a function of SNR and the anatomical strength in mSPoC$_{\bar{\gamma}}$ on dataset 1. The correlation is an average over 500 runs, each trained on 250 epochs and reported on a test set of 600 epochs. The white line shows the $\gamma$-value with highest correlation across SNRs. [From F].

### 5.2.2 Simulated Dataset 2

The forward model with free orientation was in the second dataset used to simulate data. We therefore now generated random direction vectors for the planted source components to describe their 3D orientation. The temporal similarity was estimated on a test set of 600 epochs.

As seen in Fig. 5.5, mSPoC$_{\bar{\gamma}}$ extracted component pairs having higher correlations than the original mSPoC. Note that in order to calculate the cross-modal similarity of the spatial patterns, the fMRI was projected to EEG space using Eq. (5.10). For the original mSPoC the direction vector was defined as $\boldsymbol{\alpha} = [1\ 1\ 1]^{\top}$. For mSPoC$_{\bar{\gamma}}$ the direction vector was optimized to provide minimum least squares distance between the EEG pattern and the projected fMRI pattern following Eq. (5.12). The mSPoC$_{\bar{\gamma}}$ should therefore at the very least provide as high similarity of the patterns as the original mSPoC. The cross-modality spatial similarity will thus naturally be bigger for mSPoC$_{\bar{\gamma}}$, however importantly this was accompanied by higher performance on all similarity scores.

### 5.2.3 Real Data

The application of mSPoC and mSPoC$_{\bar{\gamma}}$ is exemplified on the combined EEG-fMRI study described in Section 5.2. We applied mSPoC and mSPoC$_{\bar{\gamma}}$ to the

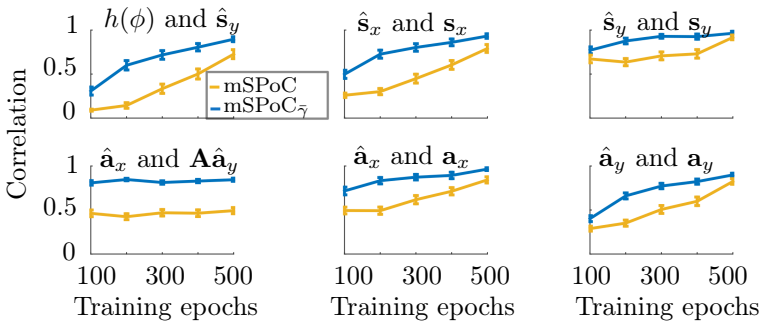



**Figure 5.5:** The performance of mSPoC (yellow) and mSPoC$_{\bar{\gamma}}$ (blue) on dataset 2 as a function of the number of trainings epochs and with SNR $= -12$ dB. The average across 50 repetitions is shown with s.e.m. indicated by errorbars.

fMRI and the band passed EEG data (16-22 Hz; corresponding to the sensorimotor rhythm). The anatomical augmentation of mSPoC$_{\bar{\gamma}}$ was through cross-validation estimated to be $\gamma = 0.1$. The extracted EEG-fMRI component pair had a functional correlation of 0.94 for mSPoC$_{\bar{\gamma}}$ and 0.82 for mSPoC. The recovered EEG and fMRI spatial patterns of mSPoC and mSPoC$_{\bar{\gamma}}$ can be seen in Fig. 5.6A and B, respectively. The EEG components were source localized and overlayed on the fMRI spatial pattern in Fig. 5.6B. As the spatial pattern is a vector we used the variational garrote (VG) for source reconstruction. The source reconstructed EEG components were in the visualization thresholded to retain the same number of voxels as found in the mSPoC extracted fMRI component, corresponding to 54 voxels. The average distance between the 54 EEG sources and their nearest fMRI component activation was 20 mm for mSPoC$_{\bar{\gamma}}$ and 33 mm for the original mSPoC. As seen in Fig. 5.6B the mSPoC EEG component had some spurious frontal activation explaining the larger localization difference. Overall mSPoC$_{\bar{\gamma}}$ improved the functional correlation between the modalities and provided more focal activation of the EEG sources.

## 5.3 Relation to Other Work

As mentioned in Chapter 2 EEG and fMRI measure different features of the brain activity [Laufs, 2012]. This fact complicates a direct integration of the two modalities, however it is also what makes a successful integration so attractive. Multimodal fusion methods are often divided into early and late fusion

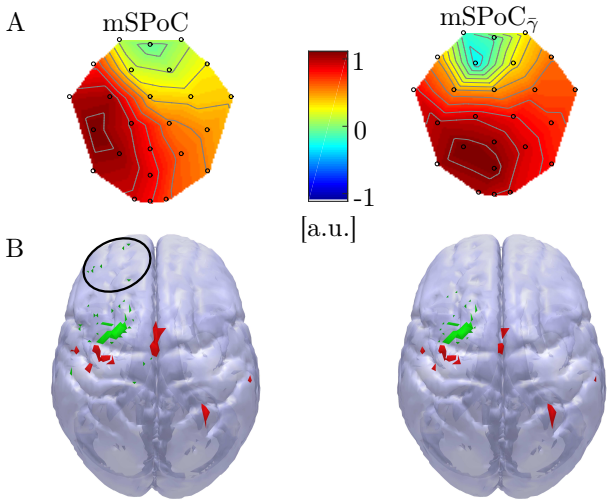



**Figure 5.6:** Spatial patterns recovered on motor data by mSPoC (left) and mSPoC$_{\bar{\gamma}}$ (right). (A) The spatial pattern of the EEG component. (B) The spatial pattern of the fMRI component (red) and the EEG component reconstructed using VG (green). The black circle indicates the frontal activity of the source localized mSPoC EEG component. Surface rendering was performed with 'MatCro'[1].

approaches. An example of an early fusion approach is, e.g., CCA where filters are applied to the modalities to extract components that are maximally correlated [Correa et al., 2010]. When the coupling between modalities is non-linear as for EEG and fMRI CCA is suboptimal. Substituting the EEG signal in the CCA problem with, e.g., its bandpower leads to problems of interpretability of the extracted filters as the non-linear transformation of the sensor level EEG signal does not respect the generative model and because noise will distort the decomposition. The extracted spatial filters can therefore not be transformed into spatial patterns and thereby reveal to what degree a component is expressed. CCA was compared to mSPoC in [Dähne et al., 2013] where it was found that the components extracted by mSPoC had higher mutual correlation.

ICA can be used to perform both early and late fusion of modalities [Lei et al., 2012]. The so-called joint ICA is an example of an ICA method providing early fusion [Calhoun et al., 2006]. The objective of joint ICA is to recover spatial maps from the fMRI that are coupled with time courses of the EEG across subjects. Joint ICA thus relies on having multiple subjects to estimate correlated components. In late fusion independent components are extracted

[1] http://www.mccauslandcenter.sc.edu/CRNL/tools/surface-rendering-with-matlab

from each modality separately and the components with high cross-modality correlations are chosen as the coupled component pair [Eichele et al., 2008].

Late fusion has the drawback of blindly extracting components from one modality without considering the other modality. The mSPoC has the unique feature of being an early fusion method, and thereby directly searches for correlated components, *while* still respecting the generative models of the modalities. In [Dähne, 2015] it was shown that mSPoC outperforms the late fusion ICA (which respects the generative models) and the early fusion CCA (which does not respect the generative models). We have shown that the performance of mSPoC can be even further improved by the inclusion of anatomical information into the extraction of correlated components.

Chapter 6

# Conclusion

The main focus of this thesis has been on developing methods for acquiring detailed and accurate spatio-temporal information about brain dynamics. No existing neuroimaging modality is capable of providing brain scans of both high spatial and high temporal detail. However high temporal resolution is inherent to EEG and by using signal processing techniques its spatial resolution can be improved. We furthermore chose EEG as the starting point of our investigations because of its accessibility and portability which effectively facilitate naturalistic brain studies as well as widespread use. We tackled the problem of increasing the spatial resolution of EEG from three angles.

First, we developed methods for locating the sources of EEG activity using physiologically relevant a priori knowledge. The ill-posed inverse problem was solved by a Bayesian procedure having separate parametrization of source locations and strengths. Effectively this allowed for temporal smoothness to be imposed on the locations while maintaining a more flexible representation of the strength. The enforced temporal coherency improved the localization of sources compared to using a single measurement vector model and compared to the multiple measurement vectors models MSP, M-FOCUSS and T-MSBL. The implementation of smooth and compact spatial basis functions provided further improvements. This was seen on the reconstructed sources of the face evoked response where the source components corresponding to the fusiform face area were located deeper and closer to the locations identified by fMRI [Kanwisher

et al., 1997; Iaria et al., 2008], MEG [Henson et al., 2009b] and multimodal studies [Henson et al., 2011]. The spatial basis functions were thus seen to increase sensitivity to deep EEG sources as also found in [Friston et al., 2008; Trujillo-Barreto et al., 2004].

Secondly, we developed a novel method to generate subject-specific forward models with the aim of increasing source localization accuracy when the head geometry is unknown. Even the best inverse solver is reliant on a forward model which adequately describe the geometry and biophysics of the head [Oostenveld and Oostendorp, 2002; Plis et al., 2007; Lew et al., 2007, 2009; Stahlhut et al., 2012; Akalin Acar and Makeig, 2013; Windhoff et al., 2013]. The novelty of our proposed inference procedure lies in the fact that no structural scans are required. Instead the EEG of a new subject is combined with a low-dimensional representation of forward models for the recovery of a suitable forward model and the EEG sources. The proposed forward model inference scheme can be employed beyond EEG, for example supporting other tools within neuroscience, e.g., transcranial magnetic stimulation, direct current stimulation and EIT, and outside the field of neuroscience, e.g., in heart modeling [Rudy, 2015] and seismology [Tommasi, 1998; Klimchuk et al., 2004].

Thirdly, we looked for fMRI components which were comodulated with the dynamics of the EEG components' power using the framework in the multimodal source power comodulation (mSPoC). The extracted spatial distribution of the fMRI components can thus be used to describe the spatial distribution of the comodulated EEG component or be used as a prior in source localizing the EEG component. The original mSPoC only considered temporal/functional coupling between the two modalities but we showed that by adding spatial information through the forward model better performance could be achieved.

## 6.1 Future Directions

### 6.1.1 Real Time Imaging and Noise Modeling

The inference scheme of VG is of low computational complexity and thus encourages use in real time EEG imaging and neurofeedback. In order to reduce computation time further different time optimization techniques can be employed. The cross-validation scheme for estimating the sparsity of VG is the main bottleneck for obtaining very low computation time and should thus be optimized. Instead of estimating the regularization parameter for each time sample, cross-validation could, e.g., be performed on EEG recorded prior to the

actual brain imaging. The estimated parameters could then be held fixed for a number of time samples during which the regularization parameter could be optimized again. Another approach is to decrease the computational complexity by introducing a so-called active set as in the applied implementation of M-FOCUSS [Zhang and Rao, 2011]. The sources found to be inactive are in this method removed from the set of potential relevant sources, thus iteratively reducing the source dimension. A final suggestion is to employ ICA as a preprocessing step. Delorme et al showed the ability of ICA to extract components of near dipolar distribution [Delorme et al., 2012]. ICA can therefore be used to provide sparse focal distribution thus simplifying the source localization task as, e.g., proposed in [Pion-tonachini et al., 2015]. It should be mentioned that ICA in general also provide removal of non-brain electrical activity [Frølich et al., 2015].

The proposed inverse solvers assumed the data to contain additive uncorrelated Gaussian noise which is only representative of the truth when certain preprocessing steps are employed. Whitening of the data [Pedersen et al., 2008] is one commonly used technique [Bolstad et al., 2009; Gramfort et al., 2013]. It requires the covariance of the data which is often approximated from baseline data. We tested our solvers on the averaged differential EEG response of two conditions, the EEG data should thus have a relatively high signal-to-noise ratio. In future extensions we would like to generalize the VG algorithms to be more invariant to the noise distribution. This would be a requirement for applying the inverse solvers to real time imaging where (contrast) averaging is not possible.

### 6.1.2 More Detailed Forward Models

Forward models are being built with increasing detail, e.g., containing many tissue types which are meticulously segmented and meshed [Windhoff et al., 2013; Stenroos and Nummenmaa, 2016]. However these methods demand a high level of expertise of the user and structural scans of high quality. Whether our simplified BEM head models are detailed enough to capture the relevant head geometry could be tested by generating test data based on a detailed forward model.

A recent study by [Stenroos and Nummenmaa, 2016] showed that including the CSF layer in a BEM head model improves the solution but also that the lack of a CSF layer can be partly compensated by decreasing the skull conductivity. In our study the skull:brain conductivity ratio was adapted to the EEG data thus allowing for some degree of compensation for the omission of a CSF layer. However in order to build more accurate forward models the CSF should be

included. Creating more complex forward models would in general require a higher dimensional forward model representation thus in effect demanding a bigger database of corpus subjects. Databases collected under the Biomedical Informatics Research Network (BIRN) contain structural scans of many subjects and could for example be used.

The curse of dimensionality complicates a grid search for locating the optimal forward model when the forward model representation is expanded. Alternatively a forward model can be recovered using optimization tools based on, e.g., Bayesian optimization schemes, such as BayesOpt [Martinez-Cantin, 2014]. Such optimization tools could also be used for MarkoVG to efficiently estimate its two parameters (i.e. sparsity and smoothness) separately.

The generated forward model representation relies on a specific configuration of the EEG electrodes, which however varies across studies. Thus to further generalize the utility of the proposed forward model inference procedure, the corpus forward models should contain forward field mappings to many points on the scalp, i.e., corresponding to high-density EEG systems. In this way the forward model representation could be applied to new subjects with any electrode configuration.

The proposed forward model inference pipeline can potentially provide dynamic forward models and thus offer a framework that respect the head position of the subject. The importance of considering the head position was shown in [Rice et al., 2013] where the occipital signal power increased on averaged by 80 % in supine compared to prone body positions. If the goal of the forward model inference is to provide a stationary forward model the forward model should be optimized based on multiple EEG paradigms with EEG sources located in different brain areas.

### 6.1.3 Integrating Multiple Modalities

The functional and anatomical relationship between the EEG and fMRI signals is far from understood. Even though evidence exists that the two modalities are related during certain brain processes [Moosmann et al., 2003; Singh et al., 1998] a direct mapping from one to the other is not described [Daunizeau et al., 2007; Laufs, 2012]. It is thus not guaranteed that correlations between EEG and fMRI can be found using the EEG source power as assumed in mSPoC or by the anatomical constraint implemented in this thesis. Focusing on the latter, tests could be performed to verify that, e.g., the anatomical constraint automatically adapts to the actual spatial overlap between the anatomical origins of the EEG and fMRI. Along the same line, useful studies include examining how accurate

the forward model, which maps fMRI activity to the EEG space, needs to be in order to improve the correlation. If a template forward model is sufficient, the augmented mSPoC would hold promise for very widespread use.

In future studies simultaneously measured multimodal datasets should be investigated closer. It would also be interesting to explore whether the same correlation strengths are recovered when the electrophysical and hemodynamic signals are not collected simultaneously. The multimodal neuroimaging dataset studying face perception would be a prime candidate for this study. The dataset could additionally help investigate the effect of including MEG as well, which was recorded simultaneously with EEG.

Appendix A

# Sparse Source EEG Imaging with the Variational Garrote

Included in the proceedings of the 3rd International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2013.

[Hansen, S.T., Stahlhut, C., & Hansen, L.K. (2013) "Sparse Source EEG Imaging with the Variational Garrote." *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on. IEEE.* doi: 10.1109/PRNI.2013.36.]

2013 3rd International Workshop on Pattern Recognition in Neuroimaging

# Sparse Source EEG Imaging with the Variational Garrote

Sofie Therese Hansen, Carsten Stahlhut, Lars Kai Hansen
*Department of Applied Mathematics and Computer Science*
*Technical University of Denmark*
*Kgs. Lyngby, Denmark*
*sofha@dtu.dk, csta@dtu.dk, lkai@dtu.dk*

***Abstract*—EEG imaging, the estimation of the cortical source distribution from scalp electrode measurements, poses an extremely ill-posed inverse problem. Recent work by Delorme et al. (2012) supports the hypothesis that distributed source solutions are sparse. We show that direct search for sparse solutions as implemented by the Variational Garrote (Kappen, 2011) provides excellent estimates compared with other widely used schemes, is computationally attractive, and by its separation of 'where' and 'what' degrees of freedom paves the road for the introduction of genuine prior information.**

***Keywords*-EEG; Imaging; Variational Garrote; LASSO; Sparse Bayesian Modeling; Sparsity**

## I. Introduction

We are interested in real-time imaging of human brain function by electroencephalography (EEG). The EEG imaging problem is of significant theoretical interest and *real-time* EEG imaging has many potential applications including quality control, in-line experimental design, brain state decoding, and neuro-feedback. In mobile applications these possibilities are attractive as elements in systems for personal state monitoring and well-being, and indeed in clinical settings where proper care requires imaging under quasi-natural conditions [1]. The first real-time mobile systems are based on reconstruction methods using basic Tikhonov regularization [1]. However, the computational challenges induced by the highly ill-posed nature of the EEG imaging problem escalate in mobile real-time systems and new algorithms may be necessary [2].

In recent work by Delorme et al. [3] it is argued that independent components of EEG signals are dipolar in nature. In particular it was shown that a direct dipolar fit can explain much of the spatially distributed signal measured in scalp electrodes. This is in line with a large literature, see e.g. [4] and references herein, suggesting sparse localized sources and motivates reconstruction algorithms that emphasize sparsity, thus contrasting the distributed spatial source patterns promoted in classical alternatives [5].

Unfortunately, the quest for sparse solutions to the EEG imaging problem is combinatorial and an exact solution will not be feasible in realistic real-time systems. Many promising approximation schemes have been proposed for the general problem and many have been applied to the EEG imaging problem. We here investigate a recent alternative for sparse recovery proposed by Kappen [6]. The soft active set construction is of particular interest to real-time EEG as it enables separation of the location and magnitude estimation aspects of the reconstruction task, and furthermore leads to a relative low-complexity set of non-linear equations that are iterated towards the solution.

## II. The EEG Inverse Problem

In the quasi static approximation the relation between dipolar sources placed at the cortical surface $w_i$ and the measured potentials at multiple scalp locations $y_\mu$ is instantaneous and linear $y_\mu = \sum_{i=1}^{n} w_i X_{i\mu} + \xi_\mu$. We have denoted the forward model by $X_{i\mu}$ and allowed for measurement noise $\xi_\mu$, which is further assumed to be independent of the source signal. In a typical laboratory setting the number of measured scalp signals $p$ can be $32 - 256$, while the source distribution can be represented by $n = 1000 - 10,000$ locations. Thus we face a severely underdetermined problem and regularization is necessary to ensure a well-defined solution, see e.g., [7] for an early review. As we have noted key processes appear to be rather dipolar, thus searching for sparse localized solutions seems well-motivated.

## III. The Variational Garrote

The so-called Variational Garrote (VG) introduces sparseness into the regression problem by adding the binary 'location' variable $s_i \in \{0, 1\}$ for absent/present parameters [6]. Thus, the modified linear problem reads

$$y_\mu = \sum_{i=1}^{n} w_i s_i X_{i\mu} + \xi_\mu. \tag{1}$$

The location variable is a latent binary variable with a prior $p(\mathbf{s}|\gamma) = \prod_{i=1}^{n} p(s_i|\gamma)$ where $p(s_i|\gamma) = \dfrac{\exp(\gamma s_i)}{1 + \exp(\gamma)}$. Parameter $\gamma$ will in general be assumed negative $\gamma < 0$, reflecting a bias towards sparsity.

The optimal solution to (1) can be obtained with a variational approximation proposed in [6]. First the posterior probability of the model given the data is established based on a Gaussian noise assumption, $\xi \sim N(0, \beta^{-1})$, $p(\mathbf{s}, \mathbf{w}, \beta|\mathbf{D}, \gamma) \propto p(\mathbf{w}, \beta)p(\mathbf{s}|\gamma)p(\mathbf{D}|\mathbf{s}, \mathbf{w}, \beta)$, with $\mathbf{D}$ being the full data set, while the prior over sources and noise variance is assumed to be uniform $p(\mathbf{w}, \beta) \propto 1$. The

978-0-7695-5061-9/13 $26.00 © 2013 IEEE
DOI 10.1109/PRNI.2013.36

discrete variable $\mathbf{s}$ is marginalized out, giving rise to the marginal posterior, $p(\mathbf{w}, \beta|\mathbf{D}, \gamma)$. The resulting expression to maximize is now

$$\log p(\mathbf{w}, \beta|\mathbf{D}, \gamma) \propto \log \sum_{\mathbf{s}} p(\mathbf{s}|\gamma) p(\mathbf{D}|\mathbf{s}, \mathbf{w}, \beta). \quad (2)$$

Invoking Jensen's inequality and a variational posterior over source locations, $q(\mathbf{s})$, we bound the log-likelihood in (2) from below by

$$-\sum_{\mathbf{s}} q(\mathbf{s}) \log \frac{q(\mathbf{s})}{p(\mathbf{s}|\gamma) p(\mathbf{D}|\mathbf{s}, \mathbf{w}, \beta)} \equiv -F(q, \mathbf{w}, \beta). \quad (3)$$

The variational free energy $F(q, \mathbf{w}, \beta)$ is minimized, corresponding to maximizing the log-likelihood (2). We assume $q(\mathbf{s}) = \prod_{i=1}^{n} q_i(s_i)$ and factors $q_i(s_i) = m_i s_i + (1-m_i)(1-s_i)$, where $m_i$ is the probability that $s_i = 1$ [6].

As noted, the EEG problem is severely underdetermined, therefore we can simplify the model using a dual formulation with update rules for $p$ Lagrange multipliers $\lambda_\mu, \hat{y}_\nu$ [6]

$$\begin{aligned} A_{\mu\nu} =& \delta_{\mu\nu} + p^{-1} \sum_{i=1}^{n} m_i X_{i\mu} X_{i\nu} / ((1-m_i)\chi_{ii}), \quad (4) \\ y_\mu =& \sum_{\nu=1}^{p} A_{\mu\nu} \hat{y}_\nu, \quad \beta^{-1} = p^{-1} \sum_{\mu=1}^{p} \hat{y}_\mu y_\mu, \\ \lambda_\mu =& \beta \hat{y}_\mu, \quad w_i = \frac{1}{\beta p \chi_{ii}(1-m_i)} \sum_{\mu=1}^{p} \lambda_\mu X_{i\mu}, \\ m_i^{-1} =& 1 + \exp\left(-\tfrac{\beta p}{2} w_i^2 \chi_{ii} - \gamma\right), \quad v_i = m_i w_i. \end{aligned}$$

The computational complexity is dominated by a term $\propto np^2$ (4) which is much lower than the $n^3$ complexity of a direct implementation.

## IV. EVALUATION

We investigate the Variational Garrote in a series of simulation experiments, and in a benchmark EEG data set. First we investigate a simple setup based on a random forward model, while the remaining simulations and the benchmark data are based on a high-dimensional EEG Boundary Element Method (BEM) forward model [8]. For the first set of simulations we form $p = 50$ measurements and $n = 100$ unknown sources and apply the noise precision $\beta^{-1} = 1$. Here we let a single source element in the 'true' generating model be set to unity, while the rest are set to 0. VG is run on this data set with $\gamma = -10$ (found in pilot experiment) and $\mathbf{m}$ initialized with three different strategies. The swift convergence of the estimated probabilities of the location indicators is illustrated in Fig. 1.

Next, VG and three currently used approximate solvers; least absolute shrinkage and selection operator (LASSO) [9], Forward selection [10], and Sparse Bayesian Modeling (SBM) [11], are tested in a more realistic EEG setting using synthetic sources. The latter consists of 10 sources set to the value 1, and the rest 0. However, now using a normalized forward model as $\mathbf{X}$ created using OpenMEEG [8] mapping $n = 8196$ sources to $p = 128$ electrodes.

The four methods all have a single hyper-parameter to tune, in the VG we follow [6] and tune the sparsity parameter $\gamma$. The data set is first split into a training and test set with $p_{\text{test}} = 10$ and $p_{\text{train}} = 118$. Within the training set we further perform $K$-fold cross-validation to tune the four methods' hyper-parameters. We use $K = 2, ..., 15$, i.e., the training set is subdivided to consist of a training set $((K-1)/K)$ and a validation set $(1/K)$. For each $K$, performances are reported in terms of the normalized mean

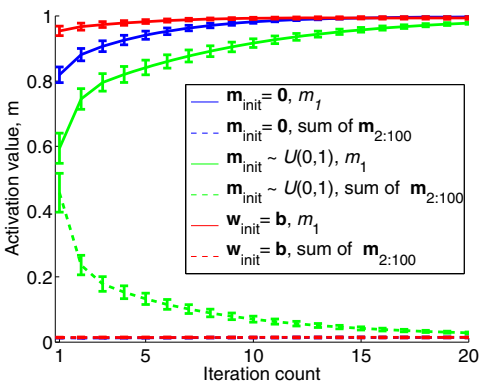



Figure 1: Simulated data ($p = 50, n = 100$, single active source, SNR= $-1.4$dB) applied to VG. Activation of the true source (full), sum of non-sources (dotted). When using $\mathbf{m}_{\text{init}} = \mathbf{0}$ the planted source is recovered in 88.3% of the 1000 repetitions, for $\mathbf{m}_{\text{init}} \sim U(0,1)$ 88.7% and using $\mathbf{w}_{\text{init}} = \mathbf{b}$ 89.6%. The found sources with $m_i > 0.5$ are considered recovered.

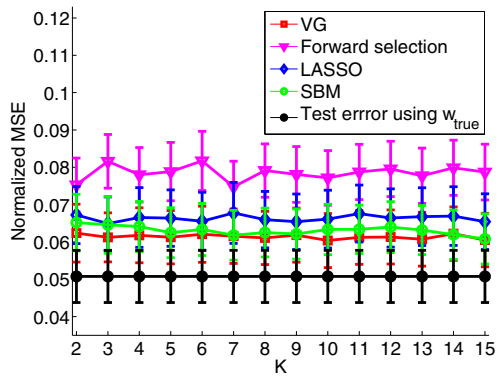
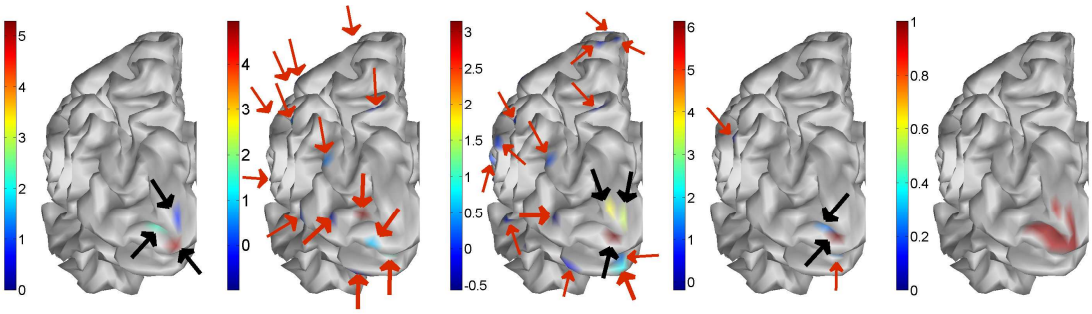



Figure 2: Normalized mean squared test error after performing 50 two-level $K$-fold cross-validations, $K = 2, ..., 15$. The algorithms are optimized wrt one parameter; for VG the sparsity level $\gamma$, for Forward selection the size of the active set, for LASSO the regularization parameter $\lambda$ and for SBM the precision of the noise $\beta$. The solution of VG is in the form of $\mathbf{v}$, while the remaining use the weight distributions directly. Ten sources out of 8196 are defined to be active in the 'true' weight distribution.




(a) VG (b) Forward selection (c) LASSO (d) SBM (e) True

Figure 3: Sources estimated through ten-fold cross-validation in the context of a 3D cortex structure are compared with the 'true' distribution. The solution for VG corresponds to $\mathbf{v}$ including a threshold on the activation so that, $P(s_i|\mathbf{D}) = m_i > 0.5$ (maximum marginal posterior). The solutions presented for the three other algorithms are the weight distributions with a treshold of $10^{-10}$. Heavy or light arrows indicate sources with magnitudes larger or less than 0.5, respectively. Black arrows indicate true sources and red false sources. View is from the back of the left hemisphere. No sources are found in the right hemisphere for VG, only low-strength sources for LASSO, one low strength for SBM while Forward selection returns many distributed sources. Note individual color maps are used.

squared error (nMSE). The above procedure is repeated 50 times. Fig. 2 compares the performances of the four methods. To set a scale we also include the nMSE for the 'true' source distribution. We find that VG outperforms the alternatives for all $K$. It is noted (not shown) that the convergence speed of VG in the applied 'EEG setting' is similar to that of the simple setup applied in Fig. 1. Interestingly, the performances of the methods with their respective optimized hyper-parameters are stable with respect to fold size $K$. Further, inspection of the estimates reveals that, e.g., LASSO - also referred to as minimum current estimate in the present context [12] - is less sparse than VG, and in fact has many small 'false' sources. Fig. 3 visualizes the spatial structure of the found sources in the context of a 3D 'cortex'.

For the simulation we also check how well VG with optimization of sparsity using the electrode cross-validation procedure is able to identify the actual source locations. For this experiment we plant 10 sources and estimate source distributions for a range of sparsity parameters ($\gamma$). In Fig. 4 we show that the cross-validation error as function of the sparsity control parameter indeed is minimized in the same range as the source retrieval index $F_1 = \frac{2\cdot\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$ [13] is maximized. For comparison we show the similar plot for SBM (where the cross-validated hyper-parameter is the noise precision). Here the test error-optimal solution has a somewhat lower source retrieval index than obtained by VG.

For a final test of the performance of VG in the context of 'real' EEG we turn to the SPM face recognition benchmark data [14].We focus on reconstructing the sources and their activation in a time window $100\text{ms} < t < 200\text{ms}$, at sampling rate fs $= 200$Hz. We make a simple extension of the model to allow for a time constant $s_i$ with time varying activation strengths $w_i$, effectively decoupling the 'where' ($s$) and 'what' ($w$) degrees of freedom which leads to only minor modifications to the inference scheme [15]. In Fig. 5 we show the resulting activation time courses and in the cortex inset arrows indicate the locations of the corresponding sources. Both time courses (N170 components) and locations are consistent with the general findings of [14] and [16].

## V. DISCUSSION AND CONCLUSION

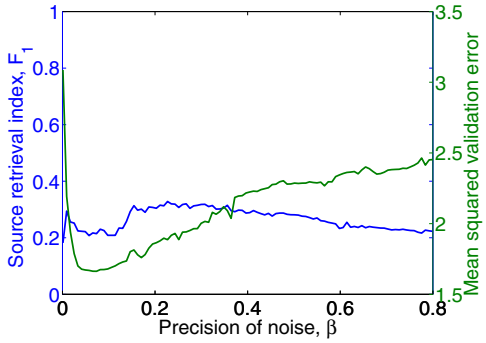
EEG imaging is a hard, underdetermined inverse problem. We hypothesize that solutions of interest are sparse and note that sparsity constraints can regularize the problem. We have shown that direct search for sparse solutions as implemented by Kappen's Variational Garrote [6] can outperform solutions based on convex relaxations (LASSO, minimum current estimate), forward feature selection, sparse Bayesian learning, both in terms of cross-validation error on test data, and in terms of quality of the solutions. In a quasi-realistic setting with an EEG forward model we found that the VG solution provides an excellent reconstruction of the planted sources. Finally, we noted that the VG model allows separation of where and what degrees of freedom, and used this to analyze a benchmark face recognition data set assuming that the locations were constant, while activations change in time. The resulting time courses and locations for a single trial were found to be consistent with the solutions proposed earlier based on averaging over multiple epochs.

## ACKNOWLEDGMENT

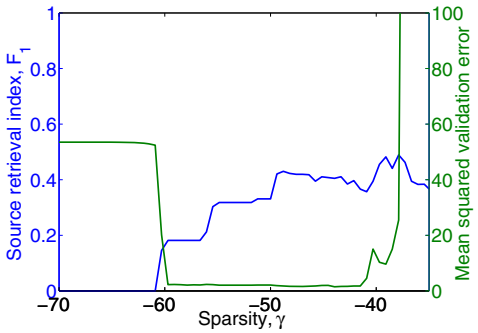
This work is supported by the Danish Lundbeck Foundation via the Center for Integrated Molecular Brain Imaging.

## REFERENCES

[1] A. Stopczynski, J. Larsen, C. Stahlhut, M. Petersen, and L. Hansen, "A smartphone interface for a wireless eeg headset with real-time 3d reconstruction," *Affective Computing and Intelligent Interaction*, pp. 317–318, 2011.




(a) Optimization comparison for SBM.




(b) Optimization comparison for VG.

Figure 4: Comparison of ability to recover sources and obtained validation error. A synthetic source distribution with ten planted sources is applied together with an EEG forward model as input, mapping $n = 8196$ sources to $p = 128$ electrodes. The means of ten repetitions are shown.

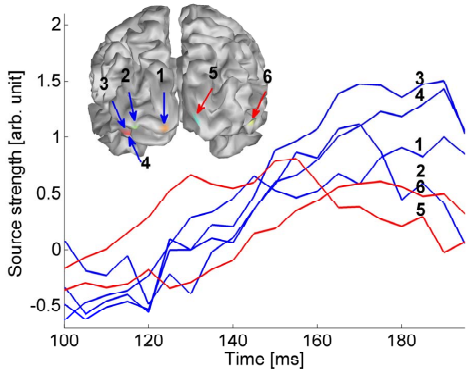



Figure 5: The sources found in a single face-evoked epoch in the time window 100-200ms by the temporally extended VG. Inset shows posterior view of the cerebrum cortex with projected sources found at 180ms. Additionally two anterior frontal sources are found in the left hemisphere. The temporal source distribution is found using five-fold cross-validation.


[2] M. Petersen, C. Stahlhut, A. Stopczynski, J. Larsen, and L. Hansen, "Smartphones get emotional: mind reading images and reconstructing the neural sources," *Affective Computing and Intelligent Interaction*, pp. 578–587, 2011.

[3] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent eeg sources are dipolar," *PloS one*, vol. 7, no. 2, p. e30135, 2012.

[4] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the m/eeg inverse problem using accelerated gradient methods." *Physics in Medicine and Biology*, vol. 57, no. 7, pp. 1937–1961, 2012.

[5] R. Pascual-Marqui, M. Esslen, K. Kochi, D. Lehmann *et al.*, "Functional imaging with low-resolution brain electromagnetic tomography (loreta): a review," *Methods and findings in experimental and clinical pharmacology*, vol. 24, no. suppl C, pp. 91–95, 2002.

[6] H. J. Kappen, "The Variational Garrote," *arXiv preprint arXiv:1109.0486*, 2011. [Online]. Available: http://arxiv.org/abs/1109.0486

[7] D. Titterington, "Common structure of smoothing techniques in statistics," *International Statistical Review/Revue Internationale de Statistique*, pp. 141–170, 1985.

[8] A. Gramfort, T. Papadopoulo, E. Olivi, M. Clerc *et al.*, "Openmeeg: opensource software for quasistatic bioelectromagnetics," *Biomed. Eng. Online*, vol. 9, no. 1, p. 45, 2010.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[10] N. Draper and H. Smith, "Applied regression analysis. Series in probability and mathematical statistics," Wiley, 1981.

[11] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[12] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *Biomedical Engineering, IEEE Transactions on*, vol. 42, no. 6, pp. 608–615, 1995.

[13] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in *Proceedings of DARPA Broadcast News Workshop*, 1999, pp. 249–252.

[14] R. N. Henson, Y. Goshen-Gottstein, T. Ganel, L. J. Otten *et al.*, "Electrophysiological and haemodynamic correlates of face perception, recognition and priming," *Cerebral cortex*, vol. 13, no. 7, pp. 793–805, 2003.

[15] S. T. Hansen, "Sparse EEG imaging," Master thesis, *Technical University of Denmark*, 2013.

[16] K. J. Friston, L. Harrison, S. J. Kiebel, C. Phillips *et al.*, "Multiple sparse priors for the m/eeg inverse problem." *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.

Appendix B

# Expansion of the Variational Garrote to a Multiple Measurement Vectors Model

Included in the conference proceedings of the Twelfth Scandinavian Conference on Artificial Intelligence (SCAI), 2013.

[Hansen, S.T., Stahlhut, C., & Hansen, L.K. (2013). "Expansion of the Variational Garrote to a Multiple Measurement Vectors Model." *In M. Jaeger (Ed.), Twelfth Scandinavian Conference on Artificial Intelligence. (pp. 105-114). IOS Press. (Frontiers in Artificial Intelligence and Applications*, Vol. 257). doi:10.3233/978-1-61499-330-8-105.]

*Twelfth Scandinavian Conference on Artificial Intelligence*
*M. Jaeger et al. (Eds.)*
*IOS Press, 2013*

*doi:10.3233/978-1-61499-330-8-105*


# Expansion of the Variational Garrote to a Multiple Measurement Vectors Model

Sofie Therese HANSEN [a,1], Carsten STAHLHUT [a] and Lars Kai HANSEN [a]

[a] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark*

**Abstract.** The recovery of sparse signals in underdetermined systems is the focus of this paper. We propose an expanded version of the Variational Garrote, originally presented by Kappen (2011), which can use multiple measurement vectors (MMVs) to further improve source retrieval performance. We show its superiority compared to the original formulation and demonstrate its ability to correctly estimate both the sources' location and their magnitude. Finally evidence is given of the high performance of the proposed algorithm compared to other MMV models.

**Keywords.** Variational Garrote, sparsity, Bayesian inference, temporally correlated sources, multiple measurements vector (MMV)

## Introduction

Estimation of sparse solutions to underdetermined linear regression problems is a very active research field and new algorithms are continuously emerging [1][2][3][4]. Important modeling tasks where the number of potential explaining features is much higher than the number of independent responses available for estimation are numerous. A key example includes locating the magneto-/electroencephalography (M/EEG) generators responsible for a measured brain wave signal. Here the potential sources are counted in the thousands while the number of electrodes used to record the signal is typically less, only a few hundreds or even less [5]. Additional examples are given in [1].

The contribution of the present paper is to expand upon a promising new algorithm, the so-called Variational Garrote (VG) proposed in [6] which was recently applied to EEG brain imaging in [5]. Our goal and contribution is to expand the algorithm to work for a stream of regression problems, also known as the multiple measurement vector (MMV) problem. Our main result is a sparsity-promoting solver that allows separation of the *where* and *what* degrees of freedom which is highly relevant to the brain imaging example. We present numerical evidence that the algorithm is competitive with existing algorithms - both with respect to the source retrieval capability and the computation time.

[1]Corresponding Author: Sofie Therese Hansen, Technical University of Denmark, DTU Compute, Matematiktorvet, 2800 Kgs. Lyngby, Denmark; E-mail: sofha@dtu.dk.

## 1. The Linear Regression Problem

The starting point of many real world modeling tasks is a linear model describing the relationship between the input variables and their resulting responses [7]. The linear regression problem is defined as

$$y_\mu = \sum_{i=1}^{n} w_i X_{i\mu} + \xi_\mu, \tag{1}$$

where $y_\mu$ is one of $p$ responses which is related through the weights $w_i$ ($i = 1..n$) to the input variables contained in the so-called design matrix $\mathbf{X}$. Finally an additive noise term $\xi_\mu$ is included. The latter is often assumed to be IID and here we assume that it is zero mean, and normally distributed with precision (inverse variance) $\beta$. Eq. (1) can for example be solved using ordinary least squares or by invoking sparsity promoting methods such as Basis Pursuit, or Least Absolute Shrinkage and Selection Operator (LASSO), which exploits the $L_1$-norm to shrink weights towards zero and set some exactly to zero [7][8]. Yet other approaches are based on Bayesian inference, such as Sparse Bayesian Learning (SBL) [1][9].

## 2. The Variational Garrote

The Variational Garrote (VG) is introduced by Kappen [6] and combines a sparsity promoting representation with a new mean field solution technique to provide a computationally efficient algorithm with excellent performance [6][5]. It expands the linear regression problem by adding a binary indicator variable $\mathbf{s}$ encoding the presence of a given feature

$$y_\mu = \sum_{i=1}^{n} w_i s_i X_{i\mu} + \xi_\mu. \tag{2}$$

The result of introducing this variable ($s_i = 0, 1$) is effectively implementing the $L_0$-norm, i.e., a direct search for a solution containing a minimal number of features that is able to describe the relation between the input and the response. It is formally an extension to Breiman's non-negative Garrote [10], however, where the restriction on $s_i$ is less rigorous as $s_i$ can take all non-negative values (up to a predefined bound). The solution scheme for VG also departs from the one applied to the non-negative Garrote, and we here give a detailed description for completeness.

The indicator variable $\mathbf{s}$ is assigned a prior $p(\mathbf{s}|\gamma) = \prod_{i=1}^{n} p(s_i|\gamma)$ where $p(s_i|\gamma) = \frac{\exp(\gamma s_i)}{1+\exp(\gamma)}$ [6]. When sparsity is sought the hyperparameter $\gamma$ will be set to a negative value, as more negative values promote more sparse solutions. Approximate Bayesian inference is used to find a solution to Eq. (2) [6]. First the posterior probability of the model given the data $\mathbf{D}$ is written

$$p(\mathbf{s}, \mathbf{w}, \beta|\mathbf{D}, \gamma) = \frac{p(\mathbf{w}, \beta) p(\mathbf{s}|\gamma) p(\mathbf{D}|\mathbf{s}, \mathbf{w}, \beta)}{p(\mathbf{D}|\gamma)}, \tag{3}$$

and this expression is to be maximized with respect to $\beta$ and $\mathbf{w}$ (a vector containing the elements $w_i$, $i = 1..n$). A uniform prior is assigned to these two parameters which simplifies the above expression. Further simplifications are made by marginalizing $\mathbf{s}$ out, giving rise to the following marginal log posterior

$$\log p(\mathbf{w},\beta|\mathbf{D},\gamma) = \log \sum_{\mathbf{s}} p(\mathbf{s}|\gamma)p(D|\mathbf{s},\mathbf{w},\beta) + constant, \tag{4}$$

where the constant represents the log of the denominator in Eq. (3). As this constant will not influence the further maximization it is disregarded in the following. We can further rewrite Eq. (4) by introducing the variational approximation $q(\mathbf{s})$, which is a fully factorized distribution and satisfies $q(\mathbf{s}) = \prod_{i=1}^{n} q_i(s_i)$, where $q_i(s_i) = m_i s_i + (1-m_i)(1-s_i)$ [6]. Additionally, Jensen's inequality [11] is exploited

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

The expression $-F(q,\mathbf{w},\beta)$ should now be maximized, hopefully giving a high lower bound on the log-likelihood. This corresponds to minimizing $F(q,\mathbf{w},\beta)$, which by a physics analogy is termed the variational free energy. First $F(q,\mathbf{w},\beta)$ is rewritten

$$\begin{aligned} F(\mathbf{m},\mathbf{w},\beta) = &-\frac{p}{2}\log\frac{\beta}{2\pi} + \frac{p\beta}{2}\sigma_y^2 \\ &+ \frac{p\beta}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_i m_j w_i w_j \chi_{ij} + \sum_{i=1}^{n} m_i(1-m_i)w_i^2\chi_{ii} - 2\sum_{i=1}^{n} m_i w_i b_i\right) \\ &- \gamma\sum_{i=1}^{n} m_i + n\log(1+\exp(\gamma)) \\ &+ \sum_{i=1}^{n}\left(m_i\log(m_i) + (1-m_i)\log(1-m_i)\right). \end{aligned} \tag{6}$$

Before continuing with this expression for the variational free energy it is reformulated following Kappen to reveal the proper low rank of the involved linear system, so to speak "provided by the underdetermined nature of the problem". Here this gives rise to the computational efficient dual formulation

$$\begin{aligned} F(\mathbf{m},\mathbf{w},\beta,\mathbf{z},\lambda) = &-\frac{p}{2}\log\frac{\beta}{2\pi} + \frac{\beta}{2}\sum_{\mu=1}^{p}(z_\mu - y_\mu)^2 + \frac{p\beta}{2}\sum_{i=1}^{n} m_i(1-m_i)w_i^2\chi_{ii} \\ &- \gamma\sum_{i=1}^{n} m_i + n\log(1+\exp(\gamma)) \\ &+ \sum_{i=1}^{n}(m_i\log(m_i) + (1-m_i)\log(1-m_i)) \\ &+ \sum_{\mu=1}^{p}\lambda_\mu\left(z_\mu - \sum_{i=1}^{n} m_i w_i X_{i\mu}\right), \end{aligned} \tag{7}$$

where the variables $z_\mu = \sum_{i=1}^{n} m_i w_i X_{i\mu}$ are included along with the Lagrange multipliers $\lambda$.

Solving equation set (7) by setting the partial derivatives with respect to the parameters equal to zero gives the following final equation set

$$A_{\mu\nu} = \delta_{\mu\nu} + \frac{1}{p}\sum_{i=1}^{n} \frac{m_i X_{i\mu} X_{i\nu}}{(1-m_i)\chi_{ii}}, \tag{8}$$

$$y_\mu = \sum_{\nu=1}^{p} A_{\mu\nu}\hat{y}_\nu, \tag{9}$$

$$\frac{1}{\beta} = \frac{1}{p}\sum_{\mu=1}^{p} \hat{y}_\mu y_\mu, \tag{10}$$

$$\lambda_\mu = \beta\hat{y}_\mu, \tag{11}$$

$$w_i = \frac{1}{\beta p \chi_{ii}(1-m_i)} \sum_{\mu=1}^{p} \lambda_\mu X_{i\mu}, \tag{12}$$

$$m_i = \left(1 + \exp\left(-\frac{\beta p}{2} w_i^2 \chi_{ii} - \gamma\right)\right)^{-1}, \tag{13}$$

which is to be iterated to convergence [6]. The implementation of this formulation scales close to linearly in the number of variables ($n$), whereas the original implementation scales cubic in $n$. Eq. (8) has the highest computational complexity in the above equation set, by scaling in the order of $p^2 n$, here assuming that $p$ is much smaller than $n$. When the number of features $n$ is several thousands and the number of observations $p$ is at most a couple of hundreds, as in the EEG source reconstruction problem, then computation time is greatly reduced using the dual formulation.

### *2.1. Multiple Measurement Vectors - the Time-Expanded VG*

We suggest expanding the dual formulation of VG to be applicable to predefined time windows, forming an MMV problem. Several studies have shown that invoking MMV models may significantly improve the ability to correctly identify relevant variables [1][14]. For VG the extension, which we call the time-expanded VG (teVG for short), gives rise to fairly simple revisions to the original formulation presented above. We propose to keep the location indicator (*where*), that is the binary variable $\mathbf{s}$, constant within the time window but allow for the weights $\mathbf{w}$ (*what*) to vary freely. In the context of EEG this is a natural assumption as a network of sources is active for 10-50ms, representing 10-100 samples depending on sampling frequency, however, the activity of the network may vary at frequencies up to 100Hz, i.e., fluctuate from sample to sample [12].

Inspecting Eq. (5), the above assumption implies that only minor revisions to the likelihood $p(D|\mathbf{s},\mathbf{w},\beta)$ are necessary. The variational free energy becomes

$$
\begin{aligned}
F(\mathbf{m},\mathbf{w},\beta,\mathbf{z},\lambda) = &-\frac{Tp}{2}\log\frac{\beta}{2\pi} + \frac{\beta}{2}\sum_{t=1}^{T}\sum_{\mu=1}^{p}(z_{\mu t}-y_{\mu t})^2 + \frac{p\beta}{2}\sum_{t=1}^{T}\sum_{i=1}^{n} m_i(1-m_i)w_{it}^2\chi_{ii} \\
&-\gamma\sum_{i=1}^{n} m_i + n\log(1+\exp(\gamma)) \\
&+\sum_{i=1}^{n}(m_i\log(m_i)+(1-m_i)\log(1-m_i)) \\
&+\sum_{t=1}^{T}\sum_{\mu=1}^{p}\lambda_{\mu t}\left(z_{\mu t}-\sum_{i=1}^{n} m_i w_{it} X_{i\mu}\right). \qquad (14)
\end{aligned}
$$

As apparent from Eq. (14) the parameters $\mathbf{y}$, $\mathbf{w}$, $\mathbf{z}$ and $\lambda$ are now functions of time samples. The resulting equation set is

$$
w_{it} = \frac{1}{p\beta(1-m_i)\chi_{ii}}\sum_{\mu=1}^{p}\lambda_{\mu t}X_{i\mu}, \qquad (15)
$$

$$
A_{\mu\nu} = \delta_{\mu\nu} + \frac{1}{p}\sum_{i=1}^{n}\frac{m_i}{(1-m_i)\chi_{ii}}X_{i\mu}X_{i\nu}, \qquad (16)
$$

$$
y_{\mu t} = \sum_{\nu=1}^{p} A_{\mu\nu}\hat{y}_{\nu t}, \qquad (17)
$$

$$
\frac{1}{\beta} = \frac{1}{Tp}\sum_{t=1}^{T}\sum_{\mu=1}^{p}\hat{y}_{\mu t}y_{\mu t}, \qquad (18)
$$

$$
\lambda_{\nu t} = \beta\hat{y}_{\nu t}, \qquad (19)
$$

$$
m_i = \left(1+\exp\left(-\frac{\beta p}{2}\chi_{ii}\sum_{t=1}^{T}w_{it}^2-\gamma\right)\right)^{-1}. \qquad (20)
$$

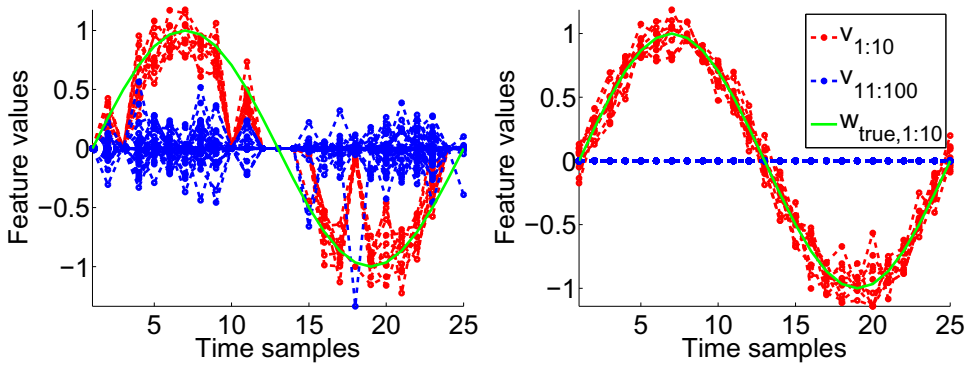
In the implementation of the above equations, it is again the calculation of $\mathbf{A}$, Eq. (16), that dominates, and since this is independent of the added dimension of time, the computational complexity is still governed by $p^2n$.

## 3. Experimental Evaluation

The time-expanded VG (teVG) is investigated in two types of experiments. First a comparison is made with the single time-shot dual formulation originally presented by Kappen and then comparisons are made with SBL varieties.

### *3.1. Comparison of the Original and the Time-Expanded Variational Garrote*

To verify the claimed improvement gained from expanding VG a simulation experiment is performed. A random design matrix is created, where $n = 100$ and $p = 50$. The response is formed by the design matrix and a weight distribution which contains 10 sine waves across 25 time samples placed in the weights $w_i$, $i = 1...10$, additionally noise is




(a) Single time-shot VG. (b) Time-expanded VG.

**Figure 1.** The feature values, true and estimated, as function of time samples. The actual appearance of the defined ten equal strength active sources is shown in green. The same ten sources estimated are seen in red, and the estimated remaining irrelevant variables are seen in blue. The estimated feature values are presented as **v** which corresponds to the element-wise multiplication of **w** and **m**. For each level of sparsity 100 iterations are applied. Five-fold cross-validation is used to find an optimum level of sparsity. SNR=10.

added. One sine wave can be seen as the green trace in Figure 1. The single time-shot VG is run separately on each of the time samples while the entire data is fed to teVG. For both algorithms an optimum level of sparsity $\gamma$ is found through five-fold cross-validation. The results can be seen in Figure 1.

Further comparisons of the two VG formulations can be seen in Figure 2, where they are compared on the basis of the mean squared error (MSE) on a test set. The algorithms are compared across 25 levels of signal to noise ratio (SNR) and are presented along with the MSE obtained when using the true weight distribution. The MSE is defined as $\mathrm{MSE} = \mathrm{E}[(\mathbf{X}^T\mathbf{v} - \mathbf{y})^2]$, where **v** is the element-wise multiplication of **w** and **m**.

### *3.2. Comparisons with Sparsity Enforcing MMV Algorithms*

The teVG is now compared with algorithms which also exploit multiple time samples to obtain improved solutions. A toolbox created by Zhang[2], is used to investigate the algorithms T-SBL, T-MSBL, MSBL and MFOCUSS. The two former algorithms, T-SBL and T-MSBL, were originally presented by Zhang and Rao in [1] and apply a so-called block sparse Bayesian learning framework (bSBL). It is thus assumed, as with VG, that a weight's mode of activity (on/off) is constant within a time window. T-MSBL is a computationally faster edition of T-SBL. The MSBL is an algorithm similar to T-MSBL and was introduced by Wipf and Rao in [2]. Finally MFOCUSS is an expanded version of FOCUSS, which finds a reweighted minimum norm solution [3]. In the current implementation a noise level (high, medium or low) must be known for the algorithms T-SBL, T-MSBL and MSBL. MFOCUSS and an additional implementation of MSBL (called MSBL2 in the following figures) use the actual noise variance. In the following results teVG is not given any prior knowledge of the noise level. The sparsity level for the method is found using the following relation, which is inspired by Kappen's suggestion of a definition for the minimum sparsity level

[2]Can be downloaded at `http://dsp.ucsd.edu/~zhilin/TMSBL.html`

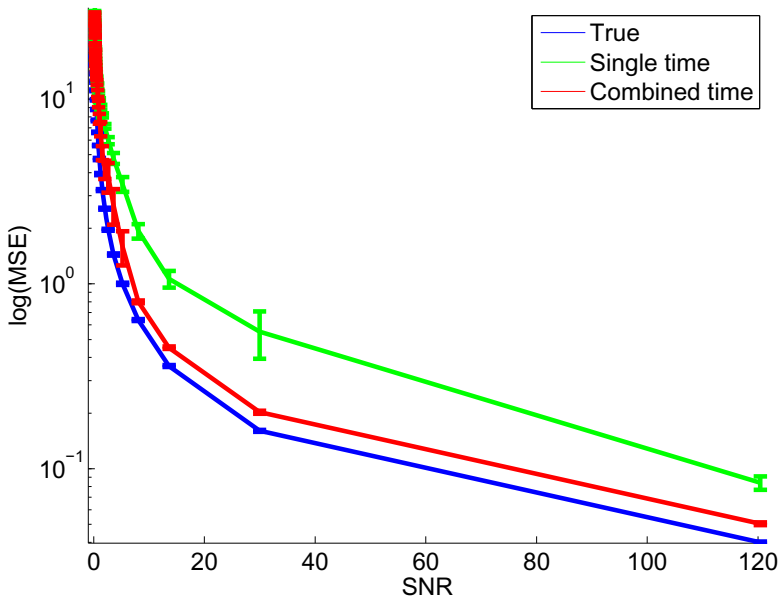



**Figure 2.** Test error as function of 25 levels of SNR. The sources' temporal development are estimated one time sample at a time (single time-shot VG) and combined (teVG), respectively. The simulated data is the same as used in Figure 1. The algorithms are compared with applying the 'true' source distribution. For each level of sparsity 100 iterations are applied in ten five-fold cross-validations.

$$\gamma = -\max_{i=1:n}\left(\left|\frac{pb_i^2}{2\sigma_y^2\chi_{ii}}\right|\right), \tag{21}$$

where $\sigma_y^2$ is the variance of the response and $\mathbf{b}$ is the input-output covariance vector. This approach of estimating $\gamma$ is used rather than cross-validation to speed up computation.

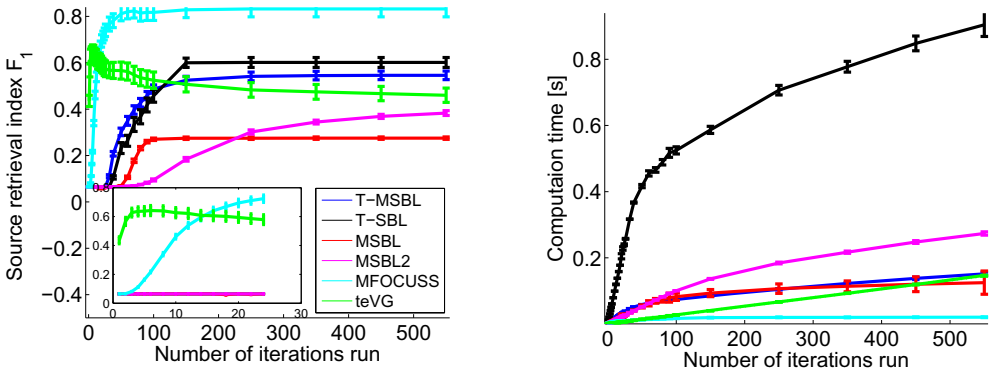
The algorithms are compared on the basis of their computation time and their ability to recover sources through the source retrieval index [13]

$$F_1 = \frac{2\cdot\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}. \tag{22}$$

A weight is deemed relevant in the model if it has a weight of considerable magnitude, the cut-off level is here set to $10^{-5}$.

First a simulation is made where a random design matrix of size $150\times 30$ is applied along with a weight distribution where five sources are active and with each source being an AR(1) process with AR coefficients 0.9. Three time samples are constructed. Noise is added to the response giving an SNR of 5dB. The results of the experiment can be seen in Figure 3.

A further simulation is made to mimic an actual application of the time-expanded algorithms with the application being EEG source reconstruction. Here the design matrix $\mathbf{X}$ would be equal to a forward model of, in this example, size $8196\times 128$, corresponding to mapping 128 electrodes to 8196 sources. Additionally, the SNR level is reduced to 3dB, as the noise is assumed to be high in EEG settings. The results can be seen in Figure 4. Notice that T-SBL has been left out of these comparisons due to too high computation




(a) Source retrieval, based on sources with magnitude $\geq 1e-5$, includes zoom inset

(b) Computation time

**Figure 3.** Comparison of MMV algorithms on simulated data with SNR=5dB and five sources temporally correlated across three time samples as AR(1) processes with AR coefficients of 0.9. A random design matrix of size 150x30 is applied. T-MSBL, T-SBL and MSBL need a noise level predefined, while MSBL2 and MFOCUSS is here given the exact noise variance. The teVG is not given information of the noise in the data.

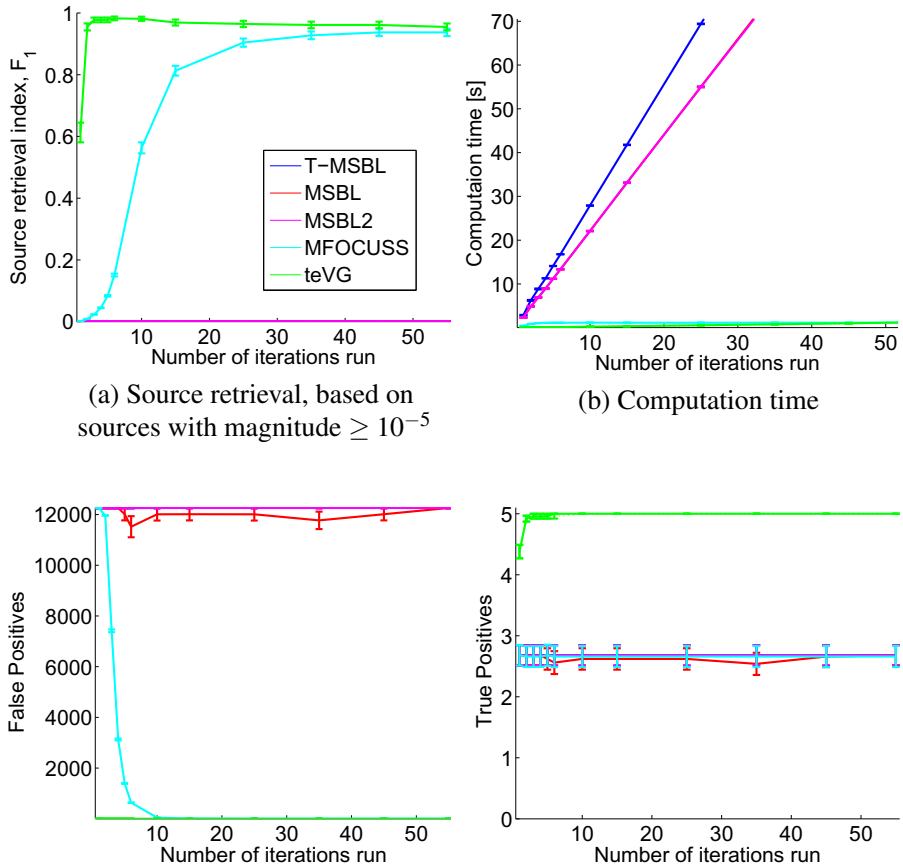
time. In addition to an evaluation of the source retrieval performance and computation time, Figures 4(a) and (b), the number of false and true positives are also depicted, see Figures 4(c) and (d).

## 4. Discussion

The evaluations of teVG confirmed the expected improvement over the original single time-shot algorithm and furthermore we see evidence of high source retrieval capability and low computation time when comparing it to other MMV algorithms.

The first statement follows from Figure 1 where teVG finds activity in the correct weights for all time samples, while the single time-shot VG fails to recover the activity in some of the time samples and additionally is seen to be more sensitive to noise and produces spurious sources. Also when comparing the test MSE achieved by the two formulations, the time-expanded version is clearly superior. The results thus support the hypothesis that using multiple samples increases the ability to correctly identify variables as being relevant or irrelevant, as also reported in other studies [14]. The number of consecutive time samples to apply is important to consider carefully in relation to the nature of the signal. That is, if a transient signal is expected, the size of the measurement vectors should be kept as low as possible, while still retaining recovery capability. Zhang et al. argued that for an EEG signal, as few as 5 time samples should be used [1].

In the comparison with T-MSBL,T-SBL, MSBL and MFOCUSS the teVG is indeed competitive. In the simulation where the smaller design matrix was used, teVG achieves the highest source retrieval score when comparing the algorithms run on few iterations (less than 10), see Figure 3 (a). However beyond this point VG is outperformed by especially MFOCUSS. It should be kept in mind that in this comparison we favor MFOCUSS since it is supplied with the actual noise variance, which is not known in most applications. Turning to the evaluations on the larger design matrix with the dimensions corre-




(a) Source retrieval, based on sources with magnitude $\geq 10^{-5}$

(b) Computation time

(c) Number of false positives

(d) Number of true positives

**Figure 4.** Comparison of MMV algorithms on simulated data with SNR = 3dB and five sources temporally correlated across three time samples as AR(1) processes with AR coefficients of 0.9. A random design matrix/forward model of size $8196 \times 128$ is applied. T-MSBL, T-SBL and MSBL needs a noise level predefined, while MSBL2 and MFOCUSS is here given the exact noise variance. The teVG is not given information of the noise in the data.

sponding to an EEG forward model, teVG is superior throughout all investigated iterations, see Figure 4(a). Even more importantly it reaches close to perfect source retrieval after only a few iterations (ca. 5!). From Figure 4(c) it can be seen that the SBL models are not capable of setting the irrelevant variables to zero but instead retain them in the model. At the same time they only find about half of the relevant variables as apparent from figure 4(d).

With respect to computation time, teVG and MFOCUSS are fastest up until applying 100 iterations to reach a solution, see Figure 3(b). Thereafter MFOCUSS is less time consuming than also teVG. We expect that VG can be made even more effective by using more specific optimization schemes, e.g., such as to stop updating weights which have not been changing for a certain number of iterations.

## 5. Conclusion

We have presented a time-expanded version (teVG) of the Variational Garrote, originally proposed by Kappen (2011). The algorithm produces both location indicators as well as weight magnitude estimates.

The teVG showed excellent performance compared to the original single time-shot version when data has smooth sparse support. In one of the presented simulations we used temporal variable weight vectors corresponding to single sine waves across 25 time samples. In an EEG setting with a sampling frequency of 250Hz these simulated sources would correspond to alpha brainwaves of 10Hz. Its use in EEG is therefore motivated. Additional motivation derives from its source retrieval properties and low computational complexity and the high speed of convergence making it a promising candidate for real-time EEG source reconstruction.

## Acknowledgement

This work is supported by the Danish Lundbech Foundation via the Center for Integrated Molecular Brain Imaging.

## References

[1] Zhang, Z., Rao, B.D., *Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning*, IEEE Journal of Selected Topics in Signal Processing, Special Issue on Adaptive Sparse Representation of Data and Applications in Signal and Image Processing, vol. 5, no. 5, pp. 912-926, (2011).

[2] Wipf, D. P., Rao, B. D., *An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem*. IEEE Transactions on Signal Processing, 55(7),pp. 3704-3716 (2007).

[3] Gorodnitsky, I. F., Rao, B. D., *Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm*. IEEE Transactions on Signal Processing, 45(3), pp. 600-616 (1997).

[4] Montoya-Martinez, J., Artes-Rodriguez, A., Hansen, L.K., Pontil, M. *Structured sparsity regularization approach to the EEG inverse problem.* In Proc. 3rd International Workshop on Cognitive Information Processing (CIP 2012), IEEE (2012).

[5] Hansen, S.T., Stahlhut, C., Hansen, L.K., *Sparse Source EEG Imaging with the Variational Garrote*. Proceedings of the 3rd International Workshop on Pattern Recognition in NeuroImaging, (2013).

[6] Kappen, H. *The Variational Garrote*. arXiv preprint arXiv:1109.0486. Published at http://arxiv.org/abs/1109.0486, (2011).

[7] Bishop, C. M., *Pattern recognition and Machine Learning*, Springer, New York 1st ed. (2006).

[8] Tibshirani, R., *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society. Series B, 58(1), pp. 267-288 (1996).

[9] Tipping, M. E., *Sparse Bayesian learning and the relevance vector machine*. The Journal of Machine Learning Research, 1(1), pp. 211-244 (2001).

[10] Breiman, L., *Better subset regression using the nonnegative garrote*. Technometrics, 37(4), pp. 373-384 (1995).

[11] Jensen, J. L. W. V., *"Sur les fonctions convexes et les ingalits entre les valeurs moyennes*. Acta Mathematica 30 (1): 175193 (1906).

[12] Nunez, P., Srinivasan, R., *Electric Fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2nd edition (2006).

[13] Makhoul, J., Kubala, F., Schwartz, R., Weischedel, R., *Performance measures for information extraction*. Proceedings of DARPA Broadcast News Workshop, pp. 249-252 (1999).

[14] Eldar, Y. C., Rauhut, H., *Average case analysis of multichannel sparse recovery using convex relaxation*, IEEE Trans. on Information Theory, vol. 56, no. 1, pp. 505-519 (2010).

Appendix C

# EEG Source Reconstruction Performance as a Function of Skull Conductance Contrast

Included in the conference proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[Hansen, S.T., & Hansen, L.K. (2015). "EEG Source Reconstruction Performance as a Function of Skull Conductance Contrast." *In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 827-831). doi:10.1109/ICASSP.2015.7178085.]

# EEG SOURCE RECONSTRUCTION PERFORMANCE AS A FUNCTION OF SKULL CONDUCTANCE CONTRAST

*Sofie Therese Hansen, Lars Kai Hansen*

Technical University of Denmark; Department of Applied Mathematics and Computer Science; DK-2800 Kgs. Lyngby

## ABSTRACT

Through simulated EEG we investigate the effect of the forward model's applied skull:scalp conductivity ratio on the source reconstruction performance. We show that having a higher conductivity ratio generally leads to improvement of the solution. Additionally we see a clear connection between higher conductivity ratios and lower coherence, thus a reduction of the ill-posedness of the EEG inverse problem. Finally we show on real EEG data the stability of the strongest source recovered across conductivity ratios.

***Index Terms*—** EEG source reconstruction, Inverse problem, Forward models, Sparsity, Variational approximation.

## 1. INTRODUCTION

The promise of electroencephalography (EEG) based brain imaging is high in settings where real-time and quasi-natural conditions are required [1]. These situations call for a neuroimaging technique with high temporal resolution and equipment which is portable and lightweight [2]. The span of applications for EEG source reconstruction includes clinical deployment where EEG is used in, e.g., diagnosis support [3, 4], home use for self-monitoring [2], and cognitive neuroscience brain imaging for exploring neural dynamics [1, 5, 6].

The EEG inverse problem is straightforward to formulate based on the linear relation between EEG sensors and generators provided by the quasi-static approximation of Maxwell's equations [1]. However, the problem is highly underdetermined; electrodes are counted in hundreds but potential sources in thousands. The solution is therefore non-trivial and research in the area is extensive [7, 8, 9, 10, 11]. Several studies have shown that invoking multiple measurement vector (MMV) models by assuming common support across time improve source recovery [8, 9, 10]. Also beneficial is complexity control obtained by enforcing sparsity; warranted by the assumption that the activity of interest is focal and by the dipolar nature of independent source components [10, 12].

The forward models (relating the EEG sources to the electrodes) are inherently very coherent and as explained in [13] this adds to the ill-posedness of the inverse problem. Often the forward model is assumed fixed and known. The boundary element method (BEM) is an efficient way of obtaining a forward model where the geometry of the layers between the electrodes and sources can be included [1]. For instance, the three-layered BEM models the scalp, skull and cortex, with information from either structural head scans or through template models [1]. The used conductivity ratios of these layers have been shown to affect the ability to correctly localize the EEG sources [14]. Especially the conductance of the skull is important and due to its complex structure the conductivity differs not only between subjects but also within a single subject [15]. Plis et al. suggest to account for the uncertainty of the skull conductance by including uncertainty in the source location estimate [15]. In [16] Lew et al. propose to include the skull conductivity as a parameter to model.

In [17] Wang et al. compare the source localization error obtainable with skull:scalp conductivity ratios 1/15, 1/20, 1/25 and 1/80 using both the data generating ratio and the three incorrect ratios. With a signal to noise ratio (SNR) of 10 dB and one planted source, an error of 9 mm is reported when using the true ratio. Single equivalent current dipole was used for source localization and the solution is thus biased by the knowledge of the number of planted sources.

The basic question of this paper is *how does the skull conductance contrast affect source reconstruction performance?*. We will answer this by systematically investigating the performance of the source reconstruction as a function of the skull conductivity with several state of the art inverse problem solvers. Rather than considering the error introduced when using a wrong conductance ratio as in [15, 16], we use the same ratio for generating the simulations as we use in the reconstruction. We thus demonstrate across 100 ratios the inherent variability of the reconstruction performance. For one of the solvers we also show that a simple cross-validation based optimization of hyperparameters leads to close to oracle performance for a large range of conductivities; we thus expand on the method used in [17]. This insight we use to investigate, in real EEG data, the stability of the recovered sources with respect to the applied conductivity ratio.

## 2. MATERIALS AND METHODS

### 2.1. Neuroimaging data

Structural magnetic resonance imaging (sMRI) scans from a subject participating in a multimodal study exploring the neural mechanisms of face perception were acquired

This work is supported by the Stibo-Foundation and the Danish Lundbeck Foundation via the Center for Integrated Molecular Brain Imaging (CIMBI).

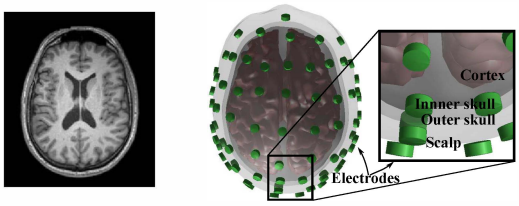



**Fig. 1**. MRI scan of a subject (Left) and the segmentation into scalp, skull and cortex (Right). The 70 electrodes are placed in the standard 10-20 system.

through the courtesy of Henson and Wakeman [18]; available at `//ftp.mrc-cbu.cam.ac.uk/personal/rik.henson/wakemandg_hensonrn/`. The T1 weighted images were recorded on a Siemens 3T Trio. We use the EEG data, recorded with 70 10-20 arranged Ag-AgCl electrodes, to create the average difference between seeing faces and scrambled faces. See stimulus design of the study in [5].

Preprocessing is done using MATLAB (MathWorks Inc.) in part by scripts provided by Wakeman and Henson. Segmentation of the subject's sMRI is done using SPM8 [19] and co-registration to the EEG electrodes is obtained through fiducials placed on the nasion and the left and right pre-auricular, together with headshape points. The cortex mesh is set to consist of 8196 vertices. The BEM, in the "bemcp" implementation (by Phillips [20]), is used to create 100 forward models having the conductivities $[1,\, c,\, 1] \cdot 0.33$ of brain:skull:scalp, where 100 samples of $c$ are drawn from a uniform distribution between 1/250 and 1/15 (in SPM8 default setting is 1/80). The constructed layers can be seen in Fig. 1 together with the electrodes.

### 2.2. Synthetic EEG data

Five synthetic data sets are constructed by planting two active sources; for all configurations one source is placed in each hemisphere (see Fig. 4). The sources are given the temporal dynamics of two synchronous sinusoids across 25 time samples. The synthetic source distributions are projected to the 70 electrodes through each of the 100 forward models. Additionally white noise is added to yield SNR$= 5\,$dB in ten repetitions (in total $5 \times 100 \times 10 = 5000$ data sets).

### 2.3. Source reconstruction methods

We will apply two source reconstruction methods to investigate the generated forward models; MFOCUSS [8] and our adapted version of the variational garrote (VG) [21]. The methods are similar in three ways. 1) They both have a regularization parameter that controls the density of the solution. 2) They assume common support across time samples. 3) Their relation to the $\ell_0$-norm solution.

We use the implementation of MFOCUSS as provided by Zhang; `http://dsp.ucsd.edu/~zhilin/TMSBL.html`. MFOCUSS by Cotter et al. [8] is an MMV version of FOCUSS (FOCal Underdetermined System Solution), which finds a reweighted minimum norm (MN) solution [7, 22], and thus approximates the $\ell_0$ (numerousity) regularizer [10]. FOCUSS is initialized with the MN solution and then iteratively increases large solution values and decreases small [7].

The regularization parameter of MFOCUSS is the noise level of the data. This parameter is in the experiments varied from $10^{-11}$ to 1 in 100 steps on the log10 scale. We use MFOCUSS in an oracle like setup where we report the solution with lowest localization error and highest F-measure across investigated noise levels. This is of course not applicable to real data where we would not have ground truth available. However on the synthetic data we want to use this approach to find the best possible solution obtainable using MFOCUSS and a forward model with a specific skull:scalp conductivity ratio. The localization error is the average Euclidean distance between the two planted and two estimated strongest sources. And F-measure $= 2 \cdot \mathrm{TP}/(\mathrm{P} + \mathrm{TP} + \mathrm{FP})$ [23]; TP, FP and P are the number of true, false and all actual positives, respectively. Perfect localization yields an F-measure of 1.

The second source reconstruction method we test is our MMV modified version of VG [21]. The $\ell_0$ regularization is achieved in VG for the single measurement by including a binary variable $s_n \in \{0, 1\}$ (for electrode $n = 1..N$) in the linear relation between electrodes and sources, modelling whether a source is active or not. In our modification, detailed in [24], we extend VG to MMV by simply fixing this binary variable across time; we call it the time-expanded VG (teVG). The binary variable has the prior $p(\mathbf{s}|\gamma) = \prod_{n=1}^{N} p(s_n|\gamma)$ where $p(s_n|\gamma) = \dfrac{\exp(\gamma s_n)}{1 + \exp(\gamma)}$ [21]. The hyperparameter $\gamma$ is thus sparsity controlling.

The solution scheme proposed by Kappen et al. is based on Bayesian inference by maximizing the posterior probability. As this is non-trivial, variational approximation is employed. First a marginalization over $\mathbf{s}$ is performed and $q(\mathbf{s}) = \prod_{n=1}^{N} q_n(s_n)$ is introduced, where $q_n(s_n) = m_n s_n + (1 - m_n)(1 - s_n)$ [21]. The parameter $m_n$ is the variational mean and can be interpreted as the probability of $s_n$ being active, thus it has values between 0 and 1. Next, Jensen's inequality is applied giving us the lower bound; i.e. the free energy $F$. We pose the free energy in a dual formulation following [21], and minimize it by setting the partial derivatives equal to zero, except for $\gamma$ which we estimate through cross-validation. In Fig. 2 we show how we split the 70 electrodes into four folds. The partitioning is done with the aim of maximizing the spread of the 17-18 electrodes contained in each fold.

Kappen et al. suggest to obtain the solution through fixed point iteration, which has a computational complexity that scales quadratic in the number of electrodes and linearly in

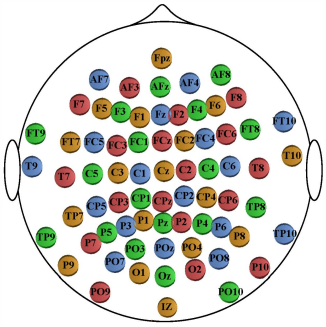



**Fig. 2**. Partioning of the 70 10-20 arranged EEG electrodes into four folds represented by the coloring.

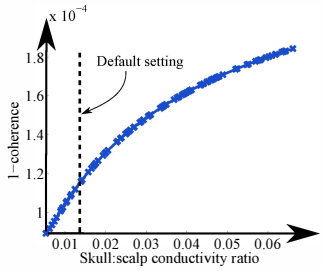



**Fig. 3**. '1− the coherence' of the 100 forward models.

the number of sources, thus keeping computation time low (further reduced using common support in teVG). As the VG solves a non-convex problem a smoothing parameter is needed to control the variational mean updates. How to set this parameter is non-trivial, therefore we suggest to control the updates of $\mathbf{m}$ through gradient descent. The update to $\mathbf{m}$ will now be $\mathbf{m} - \eta \frac{\partial F}{\partial \mathbf{m}}$. The parameter $\eta$ now needs to be set, however the solution is much less sensitive to this parameter, than the before mentioned smoothing parameter. A starting value of $\eta_0 = 10^{-3}$ is heuristically set. At each iteration the decrease or increase in the free energy will increase the $\eta$-value by $10\,\%$ or reduce it by $50\,\%$, respectively. Additional convergence speed is gained by raising $\eta$ to $\eta_0$ for a number of iterations. We call this faster VG version GDteVG (Gradient Descent time-expanded Variational Garrote).

We apply GDteVG in two settings. On the synthetic data we 1) run GDteVG in the same oracle setup as MFOCUSS, and 2) using cross-validation to find the regularization parameter, $\gamma$. On the real EEG data we can of course only do the latter. In the oracle setup we explore solutions with sparsity levels from $-250$ to $-1$ in 100 steps, and we cross-validate 25 steps of sparsity between $-150$ and $-10$.

## 3. RESULTS

We investigate the matrix properties of the forward models through their coherence and condition number. The coherences of the forward models are shown in Fig. 3. As the forward models are very coherent we show '1− the coherence'. Another matrix characteristic, the condition number, is high for these models (between $1.4 \cdot 10^{15}$ and $2.9 \cdot 10^{15}$) but is not found to be linked to the conductivity ratio.

The ability of the forward models to recover the sources they themselves have projected to the electrodes is now tested. We compare the forward models through localization error and F-measure in Fig. 4. MFOCUSS and GDteVG are run in the oracle setup described earlier and GDteVGcross is GDteVG with cross-validation on the regularization parameter. There is a general trend that forward models with lower contrast, i.e., lower conductivity ratios, perform worse (source configuration 1, 2, 3 and in part 4). In the the fifth source configuration we see relatively limited effects of the contrast. Generally the GDteVG solver shows the lowest localization error and highest F-measure, and there is a good correspondence between the oracle and the cross-validation forms of the method, which allows us to make inferences for the real EEG data where ground truth is unavailable.

Finally we investigate the real EEG data described in section 2.1. As we now do not know the ground truth and can not report localization error, we instead track the strongest source found as a function of conductivity ratio, see Fig. 5. Across the 100 conductivity ratios five different sources dominate the source reconstruction solutions. Generally it is agreed that the activity happens posteriorly and 87 of the tested conductivity ratios place the strongest source in the right posterior inferior temporal lobe (red and green in Fig. 5). The solution is rather robust with respect to the time of maximum activity; having a mean of $154.4$ ms with the standard deviation $1.5$ ms.

## 4. CONCLUSION

This contribution relies on advances in both EEG forward modeling [14, 15, 16, 17, 20] and inverse inference techniques [7, 8, 9, 10, 11, 12, 21, 22] to expand on the dependence of skull conductivity on the source reconstruction performance.

For establishing an accurate electrical forward model for EEG, the conductivity distribution is needed. Here we have investigated the sensitivity of the solution to one important aspect herein, namely the skull conductivity contrast. We found that increasing the skull conductivity contrast reduces the forward model coherence, hence, in this aspect reduces the ill-posedness of the linear inverse problem. This relation was confirmed in experiments using two different sparse inverse problem solvers MFOCUSS and the variational garrote (VG). Both showed a higher localization error for low contrasts. We found that a simple cross-validation scheme could reliably be used to infer the sparsity level in VG and using the cross-validation scheme we showed that the most active dipoles found in a real EEG data set was relatively stable to the skull conductivity contrast for a wide window of values.

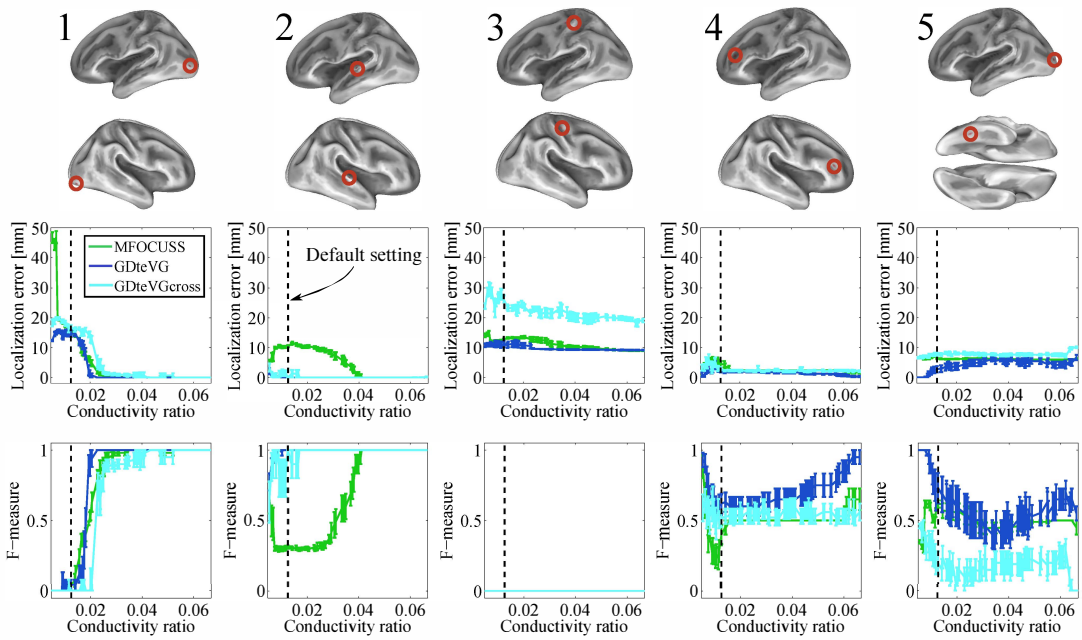



**Fig. 4**. Performance of MFOCUSS, GDteVG and GDteVGcross with forward models of varying skull:scalp conductivity ratios on synthetic data. The two former are favored by here reporting the lowest localization error (middle panel) and highest F-measure (lower panel) across regularization. At each conductivity ratio the investigated forward model is used as the data generating design matrix. The SPM8 default setting of the ratio is indicated by the dashed lines. Results are from five source configurations, each having two active sources; the locations of these can be seen in the top panel. Noise is added to give an SNR of 5 dB, this is done in ten repetitions. In the implementations with MFOCUSS the columns of the forward models are scaled to unit norm while they are merely centered for VG.

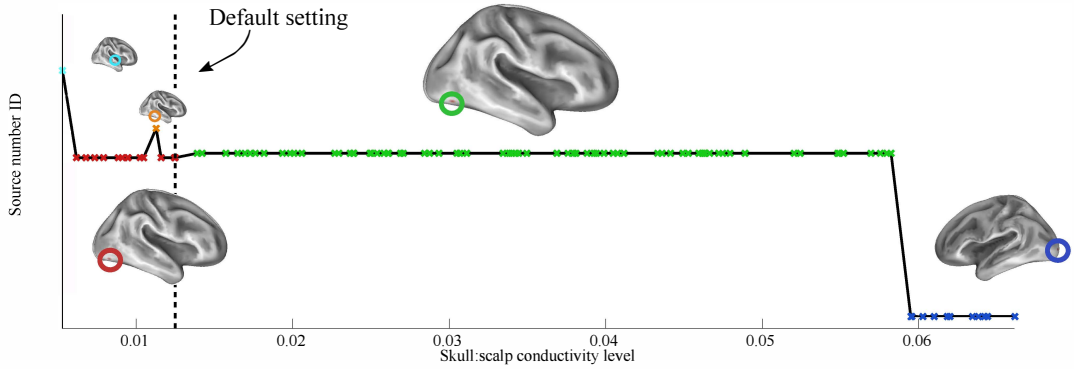



**Fig. 5**. Stability of the solution when using real EEG data across conductivity ratios (the SPM8 default setting is indicated). Each color represents a different dominating source, in total five unique strongest sources are found. Source reconstruction is performed using GDteVG with four-fold cross-validation. The EEG data is the differential signal of the conditions faces and scrambled faces. The signal is 23 time samples long, corresponding to 100 ms, and begins 100 ms after stimuli.

## 5. REFERENCES


[1] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *Signal Processing Magazine, IEEE*, vol. 18, no. 6, pp. 14–30, 2001.

[2] A. Stopczynski, C. Stahlhut, J. E. Larsen, M. K. Petersen, and L. K. Hansen, "The smartphone brain scanner: A portable real-time neuroimaging system," *PloS one*, vol. 9, no. 2, pp. e86733, 2014.

[3] M. Teplan, "Fundamentals of EEG measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[4] S. Hulbert and H. Adeli, "EEG/MEG-and imaging-based diagnosis of Alzheimers disease," *Reviews in the neurosciences*, vol. 24, no. 6, pp. 563–576, 2013.

[5] R. N. Henson, Y. Goshen-Gottstein, T. Ganel, L. J. Otten, A. Quayle, and M. D. Rugg, "Electrophysiological and haemodynamic correlates of face perception, recognition and priming," *Cerebral cortex*, vol. 13, no. 7, pp. 793–805, 2003.

[6] S. Kouider, C. Stahlhut, S. V. Gelskov, L. S. Barbosa, M. Dutat, V. De Gardelle, A. Christophe, S. Dehaene, and G. Dehaene-Lambertz, "A neural marker of perceptual consciousness in infants," *Science*, vol. 340, no. 6130, pp. 376–380, 2013.

[7] I. Gorodnitsky, J. George, and B. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroencephalography and clinical Neurophysiology*, vol. 95, pp. 231–251, 1995.

[8] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-delgado, "Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[9] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *Selected Topics in Signal Processing, IEEE*, vol. 5, no. 5, pp. 912–926, 2011.

[10] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods," *Physics in medicine and biology*, vol. 57, no. 7, pp. 1937–61, 2012.

[11] C. Stahlhut, H. T. Attias, K. Sekihara, and D. Wipf, "A hier-archical Bayesian M/EEG imaging method correcting for incomplete spatio-temporal priors," *10th IEEE International Symposium on Biomedical Imaging (ISBI 2013*, 2013.

[12] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PloS one*, vol. 7, no. 2, pp. 1–14, 2012.

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Transactions on Information Theory, IEEE*, vol. 52, no. 1, pp. 6–18, 2006.

[14] Z. Akalin Acar and S. Makeig, "Effects of forward model errors on EEG source localization," *Brain topography*, vol. 26, no. 3, pp. 378–396, 2013.

[15] S. M. Plis, J. S. George, S. C. Jun, D. M. Ranken, P. L. Volegov, and D. M. Schmidt, "Probabilistic forward model for electroencephalography source analysis., volume = 52, year = 2007," *Physics in medicine and biology*, , no. 17, pp. 5309–5327.

[16] S. Lew, C. Wolters, a. Anwander, S. Makeig, and R. MacLeod, "Low resolution conductivity estimation to improve source localization," *International Congress Series*, vol. 1300, pp. 149–152, 2007.

[17] G. Wang and D. Ren, "Effect of brain-to-skull conductivity ratio on EEG source localization accuracy," *BioMed research international*, vol. 2013.

[18] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston, "A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration," *Frontiers in human neuroscience*, vol. 5, pp. 76:1–16, 2011.

[19] J. Ashburner, C.-C. Chen, R. Moran, R. N. Henson, V. Glauche, and C. Phillips, "SPM8 manual," Tech. Rep., The FIL Methods Group, 2012.

[20] C. Phillips, *Source estimation in EEG: Combining anatomical and functional constraints*, Ph.D. thesis, 2000.

[21] H. J. Kappen and V. Gómez, "The Variational Garrote," *Machine Learning*, pp. 1–26, 2013.

[22] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[23] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in *Proceedings of DARPA Broadcast News Workshop*, pp. 249–252. Morgan Kaufmann Pub, 1999.

[24] S. T. Hansen, C. Stahlhut, and L. K. Hansen, "Expansion of the Variational Garrote to a Multiple Measurement Vectors Model," in *Twelfth Scandinavian Conference on Artificial Intelligence*, 2013, pp. 105–114.

Appendix D

# EEG Source Reconstruction using Sparse Basis Function Representations

Included in the proceedings of the 4th International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2014.

[Hansen, S.T. & Hansen, L.K. (2014) "EEG Source Reconstruction using Sparse Basis Function Representations", *Pattern Recognition in Neuroimaging (PRNI), 2014 International Workshop on. IEEE.* doi: 10.1109/PRNI.2014.6858521.]

# EEG Source Reconstruction using Sparse Basis Function Representations

Sofie Therese Hansen and Lars Kai Hansen
Technical University of Denmark
Department of Applied Mathematics and Computer Science
Kgs. Lyngby, Denmark, (sofha,lkai@dtu.dk)

***Abstract*—State of the art performance of 3D EEG imaging is based on reconstruction using spatial basis function representations. In this work we augment the Variational Garrote (VG) approach for sparse approximation to incorporate spatial basis functions. As VG handles the bias variance trade-off with cross-validation this approach is more automated than competing approaches such as Multiple Sparse Priors (Friston et al., 2008) or Champagne (Wipf et al., 2010) that require manual selection of noise level and auxiliary signal free data, respectively. Finally, we propose an unbiased estimator of the reproducibility of the reconstructed activation time course based on a split-half resampling protocol.**

## I. Introduction

We are interested in fast and accurate reconstruction of the sources of macroscopic EEG signals measured by surface electrodes. The aim of the reconstruction is to infer the spatio-temporal structure of the source signals supporting our long term goal: To combine EEG and behavioral data with neuroscience databases for real-time annotated and interpreted brain monitoring. The lack of spatial precision is one of the major challenges to such a program. The state of the art in high accuracy spatio-temporal reconstruction is represented by the multiple sparse priors (MSP) approach [1], [2], based on spatial basis functions and an approximate inference scheme using so-called automatic relevance determination (ARD) [3], [4], [5], and the Champagne framework [6], [7]. These methods are based on a combination of approximate Bayesian inference and simple heuristics, such as tuning temporal complexity by the Kaiser criterion and weakly informative hyper priors [1] or use noise estimates from other data [6]. The implicit regularization in ARD, leading to sparse solutions, is implemented by the divergence of certain precision parameters forming a highly complex optimization problem [8], [9]. Here we explore an alternative approach which combines approximate Bayesian inference and resampling methods for a more automated approach to source reconstruction in relevant spatial bases - without sacrificing accuracy or speed. The main contributions in this work are the following 1) We show that the MSP basis function approach can easily be adapted to the Variational Garrote (VG). 2) In the VG an unbiased cross-validation step is used to optimize and validate sparsity. 3) We introduce an unbiased measure of the *reproducibility* of the activation time course based on a split-half resampling protocol.

## II. The EEG Inverse Problem

Reconstruction of the EEG source distribution is based on the well-established linear relation between the measured scalp EEG potentials and the cortex level EEG generators [10], [11], [12], [13], [2]. Noting that the inverse mapping from electrodes to cortical sources is highly ill-posed, regularization is needed. We follow a Bayesian approach and control complexity by assigning priors. Sparsity of the solution is a well-known means for complexity control and at the same time motivated for EEG by recent work on the dipolar nature of independent signal components [14]. The Variational Garrote (VG), originally proposed by Kappen et al. [15] has been adapted to the EEG reconstruction problem [16]. The approach promotes sparse solutions by introducing a binary variable for each potential dipole location, implementing a flexible and adaptive description of the dipole support. A spatio-temporal multiple measurements vector approach is obtained by the additional assumption that the support is constant within a given time window (time expanded VG, teVG) [17].

The simple VG/teVG modifications to the linear regression problem, relating the EEG potentials $\mathbf{Y} \in \mathbb{R}^{K\times T}$ to the sources $\mathbf{X} \in \mathbb{R}^{N\times T}$ through lead fields $\mathbf{A} \in \mathbb{R}^{K\times N}$, are

$$\text{Linear reg.} \qquad Y_{kt} = \sum_{n=1}^{N} A_{kn} X_{nt} + \text{noise} \tag{1}$$

$$\text{VG} \qquad Y_{kt} = \sum_{n=1}^{N} A_{kn} S_{nt} X_{nt} + \text{noise} \tag{2}$$

$$\text{teVG} \qquad Y_{kt} = \sum_{n=1}^{N} A_{kn} s_{n} X_{nt} + \text{noise} \tag{3}$$

where $S_{nt}$ and $s_n$ are binary variables $\in \{0, 1\}$ and the additive noise is assumed to be Gaussian with zero mean. Using variational approximation, as described by [15] the solutions to VG and teVG can be found. Note that computation time is markedly reduced in teVG compared to individual VG reconstructions for each time point [17]. Furthermore source recovery is also greatly improved [17]. In the following, teVG is augmented to incorporate spatial basis functions. In order to make teVG more comparable with MSP, we assume in the following constant support across the whole time frame. It could be argued that dividing the data into smaller time windows would be more appropriate as brain activity can happen on short time scales [18].

### A. Implementation of Basis Functions

Basis functions are included to model the synchronous activity seen between neighboring neurons [20]. Following the framework of constructing spatial source components in MSP [1], we sample the basis functions from a coherence matrix based on the Green's function encompassing the connectivity of the cortical mesh [1]. The outcome is a set of source basis

978-1-4799-4149-0/14/$31.00 

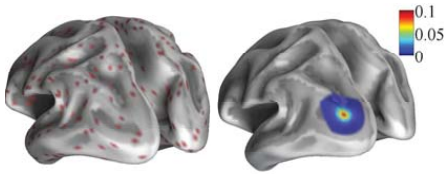



Fig. 1. Projection of the basis functions to the cortex. Left: center vertices (in red) of the $3 \times 256$ source components, and right: the spatial extent of one component. Note that the 256 bilateral components completely overlap with the 256 left and 256 right hemisphere components.

functions with compact support, see example in Fig. 1, which also shows the center vertices of 768 components.

We sample $C/3$ basis functions per hemisphere and additionally create $C/3$ bilateral functions from the unilateral functions. The $C$ basis functions can therefore model both unilateral and bilateral activity. The set of functions, make up the basis $\mathbf{B} \in \mathbb{R}^{N \times C}$. Transforming the regression problem, it now relates the observations across time $\mathbf{Y} \in \mathbb{R}^{K \times T}$ to the source functions $\mathbf{X}_C \in \mathbb{R}^{C \times T}$ through the lead fields $\mathbf{A} \in \mathbb{R}^{K \times N}$ and the basis $\mathbf{B}$

$$\mathbf{Y} = \mathbf{ABX}_C + \mathbf{noise} \Longleftrightarrow \mathbf{Y} = \mathbf{DX}_C + \mathbf{noise}, \quad (4)$$

where $\mathbf{D} \in \mathbb{R}^{K \times C}$. Following equations (1)-(3) teVG is easily applied to the constructed basis functions, where we now have $C$ binary variables. Note, the source functions are projected back to the original source space by $\mathbf{X} = \mathbf{BX}_C$. In the following analyses $C = 768$ basis functions are applied.

## III. Simulations

Through simulations emulating real EEG the localization accuracy is investigated. Two areas of activation are simulated having temporal dynamics as shown in Fig. 2, upper panel. Their locations correspond to the left and right fusiform face areas (FFAs) as found by Henson et al. (2003) [19] in an fMRI study. The peak activity in the left and right FFAs were reported in MNI space at $(-39, -51, -24)$ mm and $(42, -45, -27)$ mm, respectively. Each FFA is simulated to cover seven sources whose locations can be seen in Fig. 3(a). The forward field matrix employed is described in the next section where actual EEG source reconstruction is performed [21], [22]. The dimensions of the current simulation therefore correspond to the upcoming EEG analysis, where the cortex is divided into $N = 8196$ vertices and the EEG is measured through $K = 128$ electrodes.

The teVG is compared to MSP across 1000 simulations at signal to noise ratios (SNRs), 1 and 10[1]. Both methods apply

[1]SNR $= \langle \mathrm{Var(Signal)}/\mathrm{Var(Noise)} \rangle$ where $\langle ... \rangle$ is the time average.

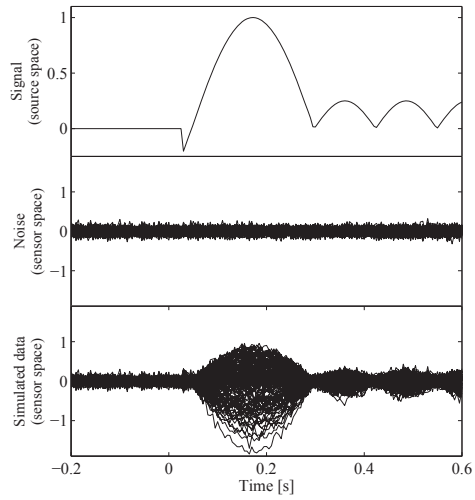



Fig. 2. Simulated data at source level together with noise and simulated data at sensor level. Seven sources having activity equal to top panel are planted in the left and right FFA. The location of these two patches can be seen in Fig. 3(a). Signal is projected to sensor space using a forward model and white noise (middle panel) is added to give the signal in the bottom panel, here SNR=10.

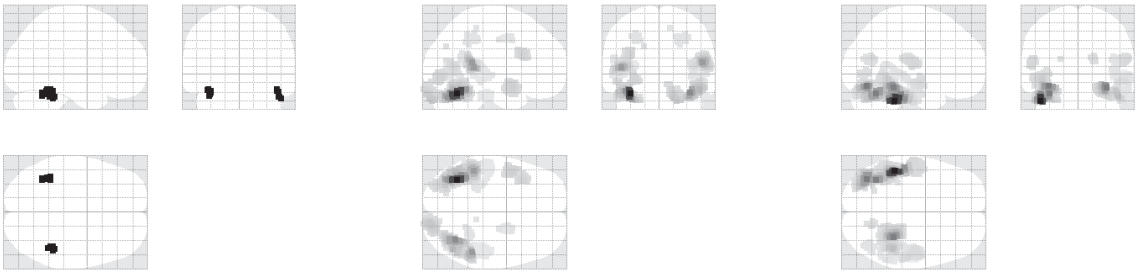
the described spatial basis functions. The temporal complexity is controlled in MSP by dimensionality reduction using principal component analysis, with a heuristic variance threshold applied to the time series before applying the ARD inference scheme. The sparsity level in teVG is selected in four-fold cross-validation (with resampling of electrodes), while the noise variance is estimated by the Bayesian scheme.

The mean localization errors are reported in Table I and the mean solutions for SNR $= 10$ are visualized in figures 3(b) and (c) using MSP and teVG respectively. The overall impression from Fig. 3 is that the teVG reconstruction is more focal, while MSP produces more scattered activity by reconstructing dipoles further away from the true activity, i.e. in the mid frontal/parietal and anterior temporal regions. Comparing in Table I MSP and teVG we find that MSP is on average $0.3$ mm more accurate, hence, the more automatic teVG approach leads to localization results that are close to state of the art methods. For both methods the time complexity of reconstruction depends on a number of choices (number of iterations etc.), here we have tuned these parameters to result in approximately the same time complexity (a total of $\Delta =$ 10-15 s including cross-validation in teVG).

In a related experiment the FFA patches reported in [19] are moved to the location of their respective nearest source

TABLE I. Localization errors for teVG and MSP on simulated left (L) and right (R) FFA activity. The error is measured as the Euclidean distance between the center of the planted activity (in the left or right hemisphere) to the source estimated to have largest magnitude (in the left or right hemisphere). The means of 1000 repetitions are listed. Due to the high number of repeats, standard deviations of the means are all below 0.4 mm rendering all positive and negative differences between MSP and teVG localization significant ($p < 0.05$). The overall mean localization errors are very similar (MSP= 18.3 mm and teVG= 18.6 mm).

| | FFAs from Henson et al. (2003) [19] | | FFAs from Henson et al. (2003) [19] projected to nearest basis function center | |
|---|---|---|---|---|
| | SNR = 1 | SNR = 10 | SNR = 1 | SNR = 10 |
| MSP | L: 13.8 mm; R: 14.5 mm | L:10.8 mm; R: 31.9 mm | L:13.4 ; R: 18.7 mm | L: 15.1 mm; R:27.9 mm |
| teVG | L: 19.1 mm; R: 22.8 mm | L: 18.0 mm; R: 21.6 mm | L: 17.7 mm; R: 21.1 mm | L: 16.3 mm; R: 12.2 mm |



(a) True (b) MSP (c) teVG

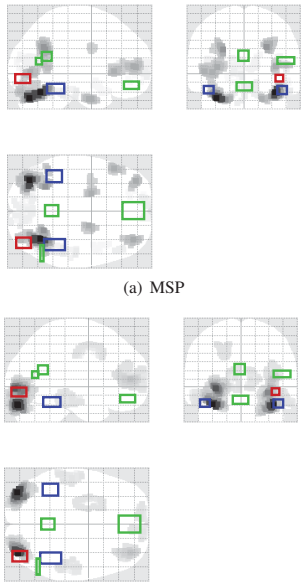
Fig. 3. Glass brain view of the simulated FFA source distribution, and the estimates by MSP and teVG, respectively. The two latter are averages across 1000 repetitions of reconstructions of the true response added with noise yielding an SNR= 10. Shown are 512 dipoles.

function centers. The aim here is to investigate whether the localization error decreases when the actual sources are close to a basis function center. The results are seen in Table I. The relative insignificant changes indicate that selection of basis functions is not critical to the result.

## IV. EEG Analysis

We now test teVG with basis functions on a benchmark 128-channel EEG data set from a subject viewing faces and scrambled faces. The paradigm of the face perception study is described in [19] and the data can be accessed via http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces/. The forward model used here, is a symmetric BEM head model produced in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) with the open source plug-in OpenMEEG [22]. As earlier, the cortex is divided into $N = 8196$ vertices which are projected to $C = 768$ basis functions. About 150 epochs are recorded of each of the two conditions. The source reconstructed difference of their means can be seen in Fig. 4(a) and Fig. 4(b) using MSP and teVG, respectively. The temporal dynamics of the two strongest sources at 170 ms are shown in Fig. 5. These correspond for both methods, to one source in the left and one in the right hemisphere. Fig. 5 shows in three panels the time courses of these two sources with their: a) 90% posterior distribution confidence intervals (only shown for MSP), and the resampling confidence intervals for the MSP (b) and teVG (c). Confidence intervals in figures 5(b) and (c) are estimated through 100 repetitions of split-half resampling of condition epochs, i.e. they arise from a total of $2 \cdot 100 = 200$ source reconstructions, see e.g., [23]. The differential responses found are close to zero in the pre-stimulus time window ($[-200, 0]$ ms) and peak at around 170 ms post-stimulus; the critical time scale for face perception [19]. The unbiased resampling based confidence intervals in the range 150-300 ms, are seen to be larger for MSP (5b) compared to teVG (5c). More specifically at $t = 170$ ms MSP has standard deviations of $0.42$ (left dipole) and $0.47$ (right dipole), where teVG has respectively $0.29$ and $0.23$. We note that the posterior and the resampling based uncertainties show some similarity although quite different in 'statistical meaning' - the former quantifies the uncertainty within the model given the specific data set, while the latter represents the expected variability under repeated experiments.

The consistency of the locations of the recovered dipoles (the strongest left and right) found when using the whole data set and in the split-halves is examined for the two methods. Compared to applying MSP to the mean difference across all epochs, the $2 \times 100$ splits recover on average the same left dipole $(29 + 43)/2 = 36$ times and the right dipole $(12+15)/2 = 13.5$ times. For teVG the locations of the splits' strongest dipoles comply $(58+57)/2 = 57.5$ times for the left and $(69 + 65)/2 = 67$ times for the right. Thus, in general teVG is more consistent with respect to the recovery of the strongest dipoles.

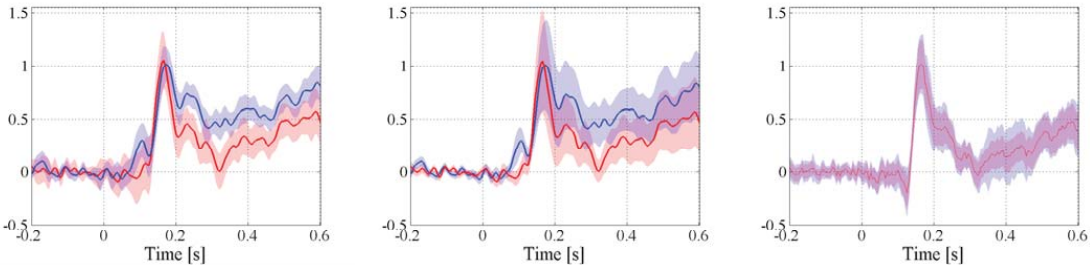


(a) MSP

(b) teVG

Fig. 4. Glass brain view of source reconstruction of the differential response of faces and scrambled faces at 170 ms post-stimulus. 512 dipoles are shown for both methods. The activation found with fMRI of the subject's face specific response are illustrated with colored squares, see [21]. The blue are in the FFA, the red in the OFA, the remaining clusters discovered are shown in green.

Inspecting in Fig. 4 the source locations for MSP and teVG, MSP seems to have a better overlap with the FFA (blue square in Fig. 4 as found by fMRI in [21]), whereas teVG seems to better locate the right occipital face area (red square).




(a) MSP (b) MSP (c) teVG

Fig. 5. Source reconstruction of the face specific response. The two sources with strongest activity are shown, these are for both methods placed in the left (blue curve) and the right (red curve) hemisphere, respectively. The confidence interval seen in (a) are estimated from the posterior distribution, while they are in (b) and (c) calculated through 100 split-half resampling of the epochs. The magnitude of the sources have been scaled by the maximum value found in the source reconstruction on all epochs. Note that the activity of the two sources are very similar, especially for teVG where the two dipoles are exactly bilateral.

## V. Conclusion

In conclusion we have shown that the MSP spatial basis function approach can be adapted easily to the Variational Garrote (VG). The VG applies cross-validation for estimation of its regularization strength, i.e, the effective sparsity level while the noise level is inferred, hence it does not require manual tuning of temporal complexity as in the MSP approach nor auxiliary signals for noise estimation as in Champagne. Finally, we used split-half resampling to provide an unbiased measure of the reproducibility of the activation time courses of the reconstructed spatial sources. For both MSP and teVG we found that the differential responses peaked around 170 ms post-stimulus, a timing that is well established for the face specific response [19]. The unbiased resampling based confidence intervals were found to be larger for MSP compared to teVG in the vicinity of this peak activity.

## Acknowledgment

This work is supported by the Danish Lundbeck Foundation via the Center for Integrated Molecular Brain Imaging.

## References

[1] K. J. Friston, L. Harrison, J. Daunizeau, S. J. Kiebel, C. Phillips, N. Trujillo-Barreto, R. N. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem." *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.

[2] G. Strobbe, P. van Mierlo, M. De Vos, B. Mijović, H. Hallez, S. Van Huffel, J. D. López, and S. Vandenberghe, "Bayesian model selection of template forward models for EEG source reconstruction," *NeuroImage*, 2014.

[3] D. J. MacKay *et al.*, "Bayesian nonlinear modeling for the prediction competition," *ASHRAE transactions*, vol. 100, no. 2, pp. 1053–1062, 1994.

[4] L. K. Hansen and C. E. Rasmussen, "Pruning from adaptive regularization," *Neural Computation*, vol. 6, no. 6, pp. 1223–1232, 1994.

[5] R. M. Neal, "Bayesian learning for neural networks," Ph.D. dissertation, University of Toronto, 1995.

[6] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, no. 1, pp. 641–655, 2010.

[7] J. P. Owen, D. P. Wipf, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data," *Neuroimage*, vol. 60, no. 1, pp. 305–323, 2012.

[8] M. Seeger and D. P. Wipf, "Variational Bayesian inference techniques," *Signal Processing Magazine, IEEE*, vol. 27, no. 6, pp. 81–91, 2010.

[9] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination." in *NIPS Advances in Neural Information Processing Systems 20, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. New York: Curan Associates*, 2007, pp. 1625–1632.

[10] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *International Journal of psychophysiology*, vol. 18, no. 1, pp. 49–65, 1994.

[11] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *Signal Processing Magazine, IEEE*, vol. 18, no. 6, pp. 14–30, 2001.

[12] H. Hallez, Vanrumste *et al.*, "Review on solving the forward problem in EEG source analysis," *Journal of neuroengineering and rehabilitation*, vol. 4, no. 46, pp. 1–29, 2007.

[13] C. Stahlhut, M. Mørup, O. Winther, and L. K. Hansen, "Simultaneous EEG source and forward model reconstruction (sofomore) using a hierarchical bayesian approach," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 431–444, 2011.

[14] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PloS one*, vol. 7, no. 2, p. e30135, 2012.

[15] H. J. Kappen and V. Gómez, "The Variational Garrote," *Machine Learning*, pp. 1–26, 2013.

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

[17] ——, "Expansion of the Variational Garrote to a Multiple Measurement Vectors Model," in *Twelfth Scandinavian Conference on Artificial Intelligence*. IOS, 2013, pp. 105–111.

[18] P. Nunez and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*, 2nd ed. Oxford University Press, USA, 2006.

[19] R. N. Henson, Y. Goshen-Gottstein, T. Ganel, L. J. Otten, A. Quayle, and M. D. Rugg, "Electrophysiological and haemodynamic correlates of face perception, recognition and priming," *Cerebral cortex*, vol. 13, no. 7, pp. 793–805, 2003.

[20] P. A. Salin and J. Bullier, "Corticocortical connections in the visual system: structure and function." *Physiological reviews*, vol. 75, no. 1, pp. 107–154, 1995.

[21] J. Ashburner, C.-C. Chen, R. Moran, R. N. Henson, V. Glauche, and C. Phillips, "SPM8 manual," The FIL Methods Group, Tech. Rep., 2012.

[22] A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc, "Forward field computation with OpenMEEG," *Computational intelligence and neuroscience*, no. 1, pp. 1–13, 2011.

[23] S. Strother, S. La Conte, L. Kai Hansen, J. Anderson, J. Zhang, S. Pulapura, and D. Rottenberg, "Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis," *Neuroimage*, vol. 23, pp. S196–S207, 2004.

Appendix E

# EEG Sequence Imaging: A Markov Prior for the Variational Garrote

Included in the planned workshop proceedings of the 3rd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging (MLINI), 2013.

[Hansen, S.T., & Hansen, L.K. (2013). "EEG Sequence Imaging: A Markov Prior for the Variational Garrote". Planned *In Proceedings of the 3rd NIPS Workshop on Machine Learning and Interpretation in Neuroimaging 2013*.]

# EEG Sequence Imaging: A Markov Prior for the Variational Garrote

Sofie Therese Hansen and Lars Kai Hansen

Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Kgs. Lyngby, Denmark
sofha@dtu.dk and lkai@dtu.dk, home page: www.compute.dtu.dk

**Abstract.** We propose the following generalization of the Variational Garrote for sequential EEG imaging: A Markov prior to promote sparse, but temporally smooth source dynamics. We derive a set of modified Variational Garrote updates and analyze the role of the prior's hyperparameters. An experimental evaluation is given in simulated data and in a benchmark EEG data set.

**Keywords:** Source reconstruction, Bayesian inference, the Variational Garrote, EEG, sparse Bayesian learning.

## 1 Sparse Sequence Reconstruction

The dynamics of electroencephalographic (EEG) sources is an active research field, see. e.g., [1–6]. We are interested in the spatio-temporal source distribution under well-defined brain activation. The contribution of the present paper is to expand upon a promising new algorithm, the so-called Variational Garrote (VG), first proposed in [7] and recently applied to EEG brain imaging in [8] and expanded to a fixed sparsity temporal model in [9]. Our goal and main contribution in this presentation is to relax the fixed sparsity model by introducing a flexible Markov prior forming a new algorithm that we refer to as MarkoVG.

The forward relation between cortical sources and electrode potentials is linear and we will here assume the forward propagation model known, although attempts have been made of estimating it from data, see e.g. [5]. Using a 'spike and slab' like representation, the linear relation between observations across time, $\mathbf{Y} \in \mathbb{R}^{K\times T}$, the forward model $\mathbf{A} \in \mathbb{R}^{K\times N}$ and the source matrix $\mathbf{X} \in \mathbb{R}^{N\times T}$ is modified in the Variational Garrote (VG) [7] as

$$Y_{k,t} = \sum_{n=1}^{N} A_{k,n} X_{n,t} + E_{k,t} \overset{VG}{\Longrightarrow} Y_{k,t} = \sum_{n=1}^{N} A_{k,n} S_{n,t} X_{n,t} + E_{k,t} \ , \qquad (1)$$

where $S_{n,t}$ is a $0,1$ binary variable controlling the spatio-temporal support of brain activity $X_{n,t}$ (i.e., at the dipolar location $n$ at time $t$). The variable $E_{k,t}$ is assumed to be i.i.d. normal noise with zero mean and unknown variance $1/\beta$. Aiming for temporally smooth and spatially sparse configurations we assign a

simple Markov model prior for the binary variables of a specific dipole location $n$, represented by a transition matrix $\Gamma_{i,j} = \mathrm{Prob}(S_{n,t} = i | S_{n,t-1} = j)$ with two free parameters, e.g., of the form

$$\Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} \\ \Gamma_{1,0} & \Gamma_{1,1} \end{bmatrix} = \begin{bmatrix} 1-\Gamma_{1,0} & \Gamma_{0,1} \\ \Gamma_{1,0} & 1-\Gamma_{0,1} \end{bmatrix} \quad . \tag{2}$$

The stationary distribution of $\Gamma$ is given by $(\Gamma_{0,1}/(\Gamma_{1,0}+\Gamma_{0,1}), \Gamma_{1,0}/(\Gamma_{1,0}+\Gamma_{0,1}))$, thus the ratio $\frac{\Gamma_{0,1}}{\Gamma_{1,0}}$ controls the prior sparsity.

The VG approach is based on approximate variational inference. Here we derive the modified update rules for the variational approximation. With uniform priors on $\mathbf{X}$ and $\beta$ we obtain a variational free energy $F(q, \mathbf{X}, \beta)$ which is minimized to obtain the optimal variational distribution $q$, the source estimates $\mathbf{X}$, and the noise parameter $\beta$. To reduce computation we use the *dual formulation* [7] introducing $Z_{k,t} = \sum_n A_{k,n} M_{n,t} X_{n,t}$ and Lagrange multipliers $\lambda_{k,t}$

$$F = -\frac{KT}{2}\log\frac{\beta}{2\pi} + \frac{\beta}{2}\sum_{t,k}(Y_{k,t} - Z_{k,t})^2 + \frac{K\beta}{2}\sum_{t,n} M_{n,t}(1-M_{n,t})X_{n,t}^2\chi_{n,n} \tag{3}$$

$$-\sum_{n,t}\left[M_{n,t}\log\frac{\Gamma_{1,0}}{\Gamma_{0,0}} + M_{n,t-1}\log\frac{\Gamma_{0,1}}{\Gamma_{0,0}} + (M_{n,t}M_{n,t-1})\log\frac{\Gamma_{0,0}\Gamma_{1,1}}{\Gamma_{0,1}\Gamma_{1,0}}\right] \tag{4}$$

$$+ NT\log\frac{1}{\Gamma_{0,0}} + \sum_{n,t}\left[M_{n,t}\log(M_{n,t}) + (1-M_{n,t})\log(1-M_{n,t})\right] \tag{5}$$

$$+ \sum_{t,k}\lambda_{k,t}\left(Z_{k,t} - \sum_n A_{k,n}M_{n,t}X_{n,t}\right) \quad , \tag{6}$$

where $\boldsymbol{\chi} = \mathbf{A}^T\mathbf{A}/K$. The variational estimates satisfy the following equation set (with $\sigma(x) = (1+\exp(-x))^{-1}$)

$$X_{n,t} = \frac{1}{K\beta}\frac{1}{(1-M_{n,t})\chi_{n,n}}\sum_k \lambda_{k,t}A_{k,n} \quad , \quad Z_{k,t} = Y_{k,t} - \frac{1}{\beta}\lambda_{k,t} \quad , \tag{7}$$

$$\beta = \frac{1}{TK}\sum_{t,k,c}\lambda_{k,t}\lambda_{c,t}C_{k,c,t}, \; C_{k,c,t} \equiv \delta_{k,c} + \frac{1}{K}\sum_n \frac{M_{n,t}}{(1-M_{n,t})\chi_{n,n}}A_{k,n}A_{c,n} \tag{8}$$

$$\lambda_{c,t} = \beta\hat{Y}_{k,t} \text{ with } \sum_c C_{k,c,t}\hat{Y}_{k,t} = Y_{k,t} \quad , \tag{9}$$

$$M_{n,t} = \sigma\left(\frac{K\beta}{2}\chi_{n,n}X_{n,t}^2 + \gamma_1 + \gamma_2\left(M_{n,t-1} + M_{n,t+1}\right)\right) \quad , \tag{10}$$

solved by iteration. Here the combination of Markov parameters; $\gamma_1 = \log\left(\frac{\Gamma_{1,0}\Gamma_{0,1}}{\Gamma_{0,0}^2}\right)$, $\gamma_2 = \log\left(\frac{\Gamma_{0,0}\Gamma_{1,1}}{\Gamma_{0,1}\Gamma_{1,0}}\right)$ determine the sparsity and smoothness of the solution: The parameter $\gamma_2$ is seen to control the degree of temporal smoothness, while $\gamma_1$ corresponds to Kappen's sparsity control parameter (with negative values favoring sparse solutions). If $\gamma_2 = 0$ we recover the original VG algorithm.

## 2 Experimental Evaluation

In the following a simulation example will serve to illustrate the properties of MarkoVG. We simulate a data set of size $N = 150$, $K = 25$ and $T = 25$. The weight distribution is controlled as three active sources with sine wave like time courses active at different time windows. The data are corrupted by normal additive noise (SNR= 5 dB). An example of the generated sources can be seen in the left panel of Fig. 1, while the right panel shows the estimated sources using MarkoVG. Here the parameter $\Gamma_{0,1} = 0.02$ is fixed, while the ratio $\frac{\Gamma_{0,1}}{\Gamma_{1,0}}$ is estimated through three-fold cross-validation among possible values ranging from $10^{-5}$ to 10, in 50 steps. For each step 25 iterations are performed.

We find that the relevant weights are recovered while one irrelevant weight (at time $t = 11$) is mistakenly judged as being relevant. Swift convergence of $M_{n,t}$ is seen in Fig. 2 for both the 'true' locations (depicted in blue, green and red corresponding to Fig. 1) and for one 'false' location, chosen as the location with non-zero activity at time $t = 11$ (depicted in gray). The color indicates the value of $M_{n,t}$; brightest or darkest indicate $M_{n,t}$=0 or 1 (minimal or maximal marginal posterior probability of activation in location $n$). In Fig. 2, MarkoVG is further evaluated using the source retrieval score ($F_1$-measure) and the mean squared error (MSE) on the estimated sources. A total of 100 randomly generated data sets similar to that shown in Fig. 1 are drawn. The performance measures are seen to converge quickly towards their optimum. Note that the iterations minimize the Free energy whose optimum is not simply related to source retrieval ($F_1$) nor MSE. Varying both $\gamma_1$ and $\gamma_2$ in a grid to inspect the influence of the smoothness and sparsity parameters on the $F_1$-measure and MSE we obtain Fig. 3. Note 'hot' colors indicate high values (better performance in $F_1$-measure and worse for the MSE). A ($\gamma_1$,$\gamma_2$)-region exists with high source retrieval ability and low MSE.

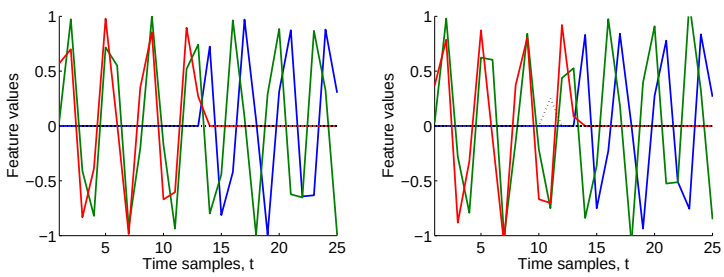



Fig. 1: The source time functions, true (left) and estimated (right). The active sources are in blue (n=1), green (n=2) and red (n=3) respectively, and the non-active sources (n=4:150) are all represented as black dotted lines. Three-fold cross-validation is used to find an optimum level of $\Gamma_{0,1}/\Gamma_{1,0}$. For each investigated level 25 iterations are applied. SNR= 5 dB.

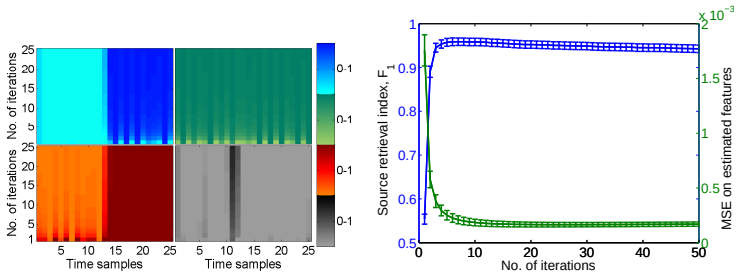



Fig. 2: Left: Convergence of $M_{n,t}$ during 25 iterations for the 3 planted active sources (blue, green and red) and the strongest false source (gray scale). Example corresponds to that shown in Fig. 1. Right: Evaluation of MarkoVG with 100 repetitions of data sets similar to that shown in Fig. 1, all with SNR around 5 dB. Performance shown as mean ± standard deviation of source retrieval $F_1$-measure and MSE on the weights.

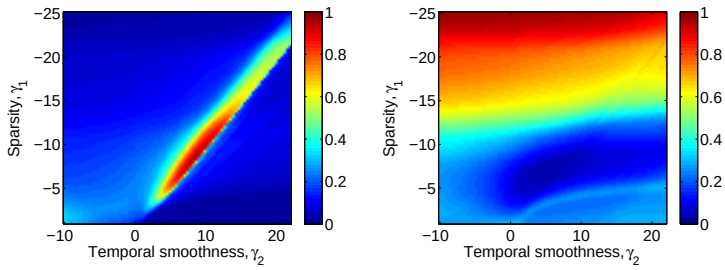



Fig. 3: Search for optimal smoothness and sparsity. Left: mean $F_1$, right: mean MSE, across 30 repetitions of data sets similar to Fig. 1(left). For each parameter combination 25 iterations are applied. SNR= 5 dB.

The existence of this region indicates that the MarkovVG representation indeed allows us to find sparse sources with limited bias on the source magnitudes.

The performance of MarkoVG is next examined on benchmark EEG data which is part of a multi-modal face response data set, available through the SPM website[1]. The data used here are collected from a single subject and used to demonstrate the modulation in brain activity when seeing faces vs. scrambled faces. The EEG signals were acquired with a 128 channel ActiveTwo system and downsampled from 2048 Hz to 200 Hz and averaged across trials, more specifications on the setup can be found in the SPM manual [10]. For reference we show the results using Friston et al.'s multiple sparse priors model (MSP) [3] as

[1] http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces/

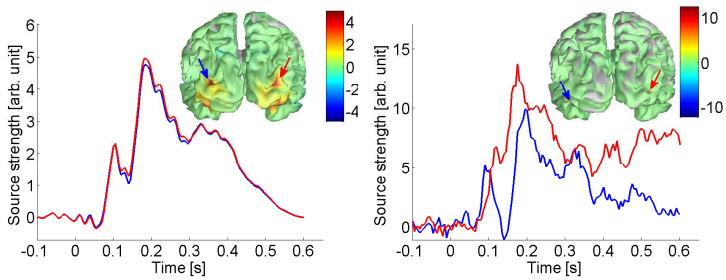



Fig. 4: The time evolution of the two strongest sources found in face response. In the inset of the cortex (posterior view) arrows point to the corresponding locations 180 ms post-stimuli. Left: Sources obtained using SPM's multiple sparse priors method. The color coding indicates that the solution obtained is rather dense. Right: Sources obtained using Zhang et al.'s T-MSBL method. As in MSP, estimate is rather dense.

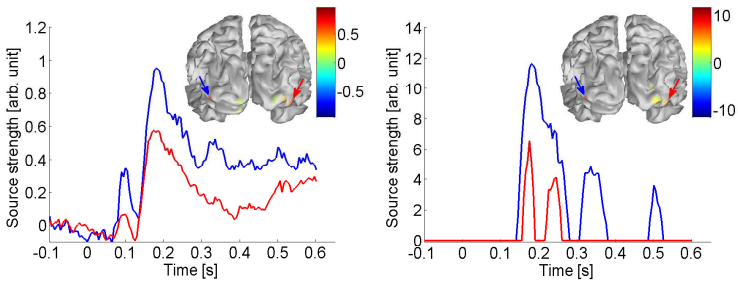



Fig. 5: The time evolution of the two strongest sources found in face response. In the inset of the cortex (posterior view) arrows point to the corresponding locations 180 ms post-stimuli. Left: Solution obtained using the temporally expanded VG (teVG) scheme. The color coding indicates more sparse solution than that obtained by MSP and T-MSBL. Right: Estimate by MarkoVG. As teVG, MarkoVG is also spatially very sparse. The time courses are, however, now sparse, thus the difference between face and scrambled face processing has been localized to focused shorter time intervals.

adopted in the SPM software and the result of Zhang et al.'s multiple measurements vector sparse Bayesian learning model, T-MSBL [4] in Fig. 4 left and right panels, respectively. These solutions should be compared to two versions of the VG: Time expanded VG (teVG) [9] and MarkoVG, both shown in Fig. 5. The MSP and T-MSBL estimates of the sources responsible for the *difference* between face and scrambled faces are very smooth in time, in fact resembling standard ERPs. This is also the case for the window-wise constant support model teVG, c.f., Fig. 5 (left), while the more flexible support recovery method MarkoVG finds a smaller number of active sites for the difference signal in Fig. 5 (right).

Also we note that teVG and MarkoVG in general find spatially sparser solutions, viz., the more extended gray areas in Fig. 4.

## 3 Conclusion

We have proposed MarkoVG assigning a Markov prior for promotion of temporally smooth sources in the Variational Garrote. We derived the modified variational update rules and identified the role of the Markov prior parameters. We showed that MarkoVG converges fast, as VG also does, and we found that sources are reliably estimated both in terms of location and source strength mean square error. In a benchmark EEG data set we showed that MarkoVG produced more focused activation than multiple sparse priors and temporal sparse Bayesian learning, both of which are more similar to our earlier VG generalization, teVG, which assumes constant temporal support in specified windows.

**Acknowledgement** We thank the Lundbeck Foundation for support through the Center for Integrated Molecular Brain Imaging (CIMBI).

## References

1. Gorodnitsky, I. F., Rao, B. D., *Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum norm algorithm.* IEEE Transactions on Signal Processing, 45(3), pp. 600-616 (1997).
2. Wipf, D. P., Rao, B. D., *An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem.* IEEE Transactions on Signal Processing, 55(7), pp. 3704-3716 (2007).
3. Friston, K. J., Harrison, L., Daunizeau, J., Kiebel, S. J., Phillips, C., Trujillo-Barreto, N., Henson, R. N., et al.: . *Multiple sparse priors for the M/EEG inverse problem*, NeuroImage 39(3), pp. 1104-1120 (2008).
4. Zhang, Z., Rao, B.D.: *Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning*, IEEE Journal of Selected Topics in Signal Processing, Special Issue on Adaptive Sparse Representation of Data and Applications in Signal and Image Processing 5(5), pp. 912–926 (2011).
5. 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(3), pp. 431-444 (2011).
6. Montoya-Martinez, J., Artes-Rodriguez, A., Hansen, L.K., Pontil, M. *Structured sparsity regularization approach to the EEG inverse problem.* In Proc. 3rd International Workshop on Cognitive Information Processing (CIP 2012), IEEE (2012).
7. Kappen, H.:*The Variational Garrote.* arXiv preprint arXiv:1109.0486. Published at http://arxiv.org/abs/1109.0486 (2011).
8. Hansen, S.T., Stahlhut, C., Hansen, L.K.: *Sparse Source EEG Imaging with the Variational Garrote*, 3rd International Workshop on Pattern Recognition in Neuroimaging (PRNI), pp. 106–109 (2013).
9. Hansen, S.T., Stahlhut, C., Hansen, L.K.: *Expansion of the Variational Garrote to a Multiple Measurement Vectors Model*, Scandinavian Conference on AI 2013 (2013).
10. Ashburner, J., Chen, C.-C., Moran, R., Henson, R. N., Glauche, V., Phillips, C.: *SPM8 Manual*, Available at http://www.fil.ion.ucl.ac.uk/spm/doc/ (2012).

Appendix F

# Fusing Simultaneous EEG and fMRI Using Functional and Anatomical Information

Included in the proceedings of the 5th International Workshop on Pattern Recognition in Neuroimaging (PRNI), 2015.

[Hansen, S.T., Winkler, I., Hansen, L.K., Müller, K.-R. & Dähne, S. (2015) "Fusing Simultaneous EEG and fMRI Using Functional and Anatomical Information." *Pattern Recognition in Neuroimaging (PRNI), 2015 International Workshop on. IEEE*, pp. 33-36. doi: 10.1109/PRNI.2015.22.]

# Fusing Simultaneous EEG and fMRI Using Functional and Anatomical Information

Sofie Therese Hansen*, Irene Winkler†, Lars Kai Hansen*, Klaus-Robert Müller†‡, Sven Dähne†
*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark.
†Department of Computer Science, Technical University of Berlin, Berlin, Germany.
‡Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea.

***Abstract*—Simultaneously measuring electrophysical and hemodynamic signals has become more accessible in the last years and the need for modeling techniques that can fuse the modalities is growing. In this work we augment a specific fusion method, the multimodal Source Power Co-modulation (mSPoC), to not only use functional but also anatomical information. The goal is to extract correlated source components from electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). Anatomical information enters our proposed extension to mSPoC via the forward model, which relates the activity on cortex level to the EEG sensors. The augmented mSPoC is shown to outperform the original version in realistic simulations where the signal to noise ratio is low or where training epochs are scarce.**

## I. Introduction

Multimodal neuroimaging techniques hold promise for enhancing our knowledge of the normal and pathological brain [1], [2]. Integrating electrophysical and hemodynamic neural responses can facilitate information not obtainable with the individual techniques [3], [4]. The electrophysical signals, e.g. magneto/electro-encephalography (M/EEG), have a high temporal resolution but low spatial resolution, whereas the hemodynamic signals, e.g. functional magnetic resonance imaging (fMRI), have low temporal resolution but high spatial resolution [5]. A technique that exploits the specific advantages of the modalities is therefore highly sought, as no gold standard is established yet. For a more exhaustive overview of multimodal analysis techniques, the interested reader is referred to [2], [6], [7]. The fusion of electrophysical and hemodynamical signals must be approached very carefully as the two categories of modalities measure different attributes of the brain activity [2], [8]. Of the many attempts to solve this problem of modeling the *neurovascular coupling* and combining the modalities, we will mention only a few.

In EEG source localization spatial information from fMRI can be used to guide the reconstruction [3]. This can be done in the form of a strict or soft prior on where the neural activity should be located according to the fMRI signal. This is called asymmetrical integration as one modality is given preference [8]. An example of symmetrical integration is given in [9] where Valdes-Sosa et al. describe an approach to correlate the actual generators of fMRI and EEG, i.e. the vasomotor feedforward signal and the net primary current density, respectively. This technique thus rely on both inverse temporal (of the fMRI) and inverse spatial (of the EEG) problems. These inverse problems are however very sensitive to the applied assumptions [8].

As in [9] we are interested in fusing bandpower modulations contained in the EEG with the dynamics of the fMRI signal. A recently proposed method attempts to achieve this fusion by projecting both EEG and fMRI signals into a lower dimensional source-component space. The method is called multimodal Source Power Co-modulation (mSPoC)[10] and it is based on ideas presented in [11]. The mapping to component space is not to be confused with the previously mentioned *source localization*, i.e. the mapping of M/EEG recordings into a much higher dimensional cortex surface space. Specifically, mSPoC searches for components that exhibit maximal functional coupling between bandpower modulations (EEG component) and the hemodynamic response (fMRI component). Both modalities are given equal priority, thus making mSPoC a symmetrical fusion approach. Importantly, mSPoC uses only the temporal information of the modalities and is thus based on purely functional assumptions.

In this paper, we suggest to extend mSPoC, when working with EEG and fMRI, by adding anatomical information that makes the method asymmetrical in the spatial domain. We propose to achieve this by projecting the estimated spatial activation pattern of the fMRI component to the EEG sensor space and bias the spatial activation pattern of the corresponding EEG component towards this projection. We thereby give preference to the fMRI in the spatial domain, and thus exploit the high spatial resolution of the modality.

## II. Methods

### A. The Generative Model

Let $\mathbf{X} \in \mathbb{R}^{M_x \times T_x}$ denote the EEG dataset and $\mathbf{Y} \in \mathbb{R}^{M_y \times T_y}$ denote the fMRI dataset, where $M_{x/y}$ and $T_{x/y}$ denote the modality specific number of sensors and time samples, respectively. The central assumption we make is that both datasets are decomposable into what is called a set of *components* (or *factors*). The component $k$ is identified by a time course ($\mathbf{s}_k$) as well as a spatial activation pattern ($\mathbf{a}_k$) and may thus be regarded as a functional unit. The generative model for dataset $\mathbf{X}$ is then $\mathbf{X} = \mathbf{A}_x\mathbf{S}_x + \boldsymbol{\epsilon}_x$, where $\boldsymbol{\epsilon}_x$ is IID noise. $\mathbf{A}_x$ contains $K_x$ spatial patterns in the columns and $\mathbf{S}_x$ contains $K_x$ time courses in the rows. The same framework also holds for $\mathbf{Y}$.

978-1-4673-7145-2/15 $31.00 © 2015 IEEE
DOI 10.1109/PRNI.2015.22

### B. Multimodal Source Power Co-modulation (mSPoC)

For completeness and to facilitate the description of our expansion in section II-C we now describe mSPoC. For further information we refer to [10].

mSPoC seeks $K \leq \min(K_x, K_y)$ pairs of *functionally* coupled components in $\mathbf{X}$ and $\mathbf{Y}$. More specifically, the band-power dynamics of a component from the EEG are assumed to co-modulate with the time course of a component in the fMRI. In the following we assume the EEG to be band-pass filtered in a frequency range of interest and epoched to match the fMRI recordings. We use the notation $\mathbf{X}(e)$ to refer to the EEG signal in epoch $e$ and $\mathbf{C}_{xx}(e)$ to denote the covariance of $\mathbf{X}(e)$ while $\mathbf{C}_{xx}$ is the covariance of the full signal $\mathbf{X}$.

In mSPoC the extraction of components is parametrized by means of spatial filters. The time courses of the extracted components from $\mathbf{X}$ and $\mathbf{Y}$ are given as $\hat{\mathbf{s}}_x \stackrel{\text{def}}{=} \mathbf{w}_x^\top \mathbf{X}$ and $\hat{\mathbf{s}}_y \stackrel{\text{def}}{=} \mathbf{w}_y^\top \mathbf{Y}$ for $\mathbf{w}_x \in \mathbb{R}^{M_x}$ and $\mathbf{w}_y \in \mathbb{R}^{M_y}$. Moreover the filter-pattern relation is $\hat{\mathbf{a}}_x \propto \mathbf{C}_{xx}\mathbf{w}_x$ and $\hat{\mathbf{a}}_y \propto \mathbf{C}_{yy}\mathbf{w}_y$ [12]. The bandpower dynamics of an extracted component from the EEG are given by the variance computed over the epoch $e$ and is denoted by $\boldsymbol{\phi}(e)$, which is defined as $\boldsymbol{\phi}(e) \stackrel{\text{def}}{=} \operatorname{Var}\left[\mathbf{w}_x^\top \mathbf{X}(e)\right] = \mathbf{w}_x^\top \mathbf{C}_{xx}(e)\mathbf{w}_x$ .

In order to model temporal delays in the coupling between $\mathbf{X}$ and $\mathbf{Y}$, a finite-impulse-response (FIR) filter $\mathbf{w}_\tau \in \mathbb{R}^{N_\tau}$ is applied to the bandpower dynamics $\boldsymbol{\phi}$. Thus, the FIR-filtered bandpower dynamics are defined as

$$h(\boldsymbol{\phi})(e) \stackrel{\text{def}}{=} \sum_{i=0}^{N_\tau - 1} \mathbf{w}_\tau(i+1) \cdot \boldsymbol{\phi}(e-i) \tag{1}$$

$$= \mathbf{w}_x^\top \underbrace{\left( \sum_{i=0}^{N_\tau - 1} \mathbf{w}_\tau(i+1) \cdot \mathbf{C}_{xx}(e-i) \right)}_{\stackrel{\text{def}}{=}\mathbf{C}_h(e)} \mathbf{w}_x, \tag{2}$$

Note that $h(\boldsymbol{\phi})(e)$ depends on $\mathbf{w}_x$ as well as on $\mathbf{w}_\tau$.

With these definitions, the mSPoC objective function for functional coupling between $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$\max_{\mathbf{w}_x, \mathbf{w}_\tau, \mathbf{w}_y} \operatorname{Cov}(h(\boldsymbol{\phi}), \hat{\mathbf{s}}_y) \tag{3}$$
$$\text{s.t.} \quad \operatorname{Var}(\hat{\mathbf{s}}_x) = \operatorname{Var}(h(\boldsymbol{\phi})) = \operatorname{Var}(\hat{\mathbf{s}}_y) = 1.$$

Note that if $\mathbf{w}_x$ is provided, $\boldsymbol{\phi}$ can be computed and the optimization problem (now only for $\mathbf{w}_\tau$ and $\mathbf{w}_y$) reduces to the well known canonical correlation analysis (CCA) with temporal embedding of $\boldsymbol{\phi}$ [13]. If, on the other hand, $\mathbf{w}_\tau$ and $\mathbf{w}_y$ are provided, then $\mathbf{C}_h(e)$ and $\hat{\mathbf{s}}_y$ can be computed and the resulting optimization problem (now only for $\mathbf{w}_x$) leads to the following generalized eigenvalue problem

$$\mathbf{C}_z \mathbf{w}_x = \lambda \mathbf{C}_{xx} \mathbf{w}_x \ , \tag{4}$$

where $\mathbf{C}_z \stackrel{\text{def}}{=} \sum_e \mathbf{C}_h(e) \cdot \hat{\mathbf{s}}_y(e)$.

The mSPoC objective is optimized by iterating over:

0) Initialize $\mathbf{w}_x$ randomly.
1) Given $\mathbf{w}_x$, solve for $\mathbf{w}_\tau$ and $\mathbf{w}_y$ using CCA.
2) Given $\mathbf{w}_\tau$ and $\mathbf{w}_y$, solve for $\mathbf{w}_x$ using Eq. (4).
3) Repeat steps 1 and 2 until convergence.

### C. mSPoC With Anatomical Information: mSPoC$_\gamma$/ mSPoC$_{\tilde{\gamma}}$

We now include the assumption that the extracted components should not only be functionally linked but also anatomically related. Thus we assume the spatial activation pattern of the EEG component, denoted by $\hat{\mathbf{a}}_x$, to be similar to the activation pattern of the corresponding fMRI component, denoted by $\hat{\mathbf{a}}_y$. In order to relate $\hat{\mathbf{a}}_x$ to $\hat{\mathbf{a}}_y$, we project $\hat{\mathbf{a}}_y$ to the EEG sensor space using a lead field matrix $\mathbf{L} \in \mathbb{R}^{M_x \times M_y}$; i.e. $\overrightarrow{\mathbf{a}}_x^y \stackrel{\text{def}}{=} \mathbf{L}\hat{\mathbf{a}}_y$ is the projected fMRI pattern.

Finding the $\mathbf{w}_x$ that optimally extracts the temporal activity from the EEG given a spatial pattern, in our case $\overrightarrow{\mathbf{a}}_x^y$, can be achieved by the linearly constraint minimum variance (LCMV) beamformer [14]. In the LCMV approach, the optimal $\mathbf{w}_x$ is derived by minimizing $\mathbf{w}_x^\top \mathbf{C}_{xx}\mathbf{w}_x = \operatorname{Var}(\hat{\mathbf{s}}_x)$, subject to the constraint $\mathbf{w}_x^\top \overrightarrow{\mathbf{a}}_x^y = 1$. Here we find it convenient to modify [14] to instead maximize $(\mathbf{w}_x^\top \overrightarrow{\mathbf{a}}_x^y)^2$, corresponding to

$$\max_{w_x} \ \mathbf{w}_x^\top \mathbf{C}_{aa} \mathbf{w}_x \ \ \text{s.t.} \ \ \operatorname{Var}(\hat{\mathbf{s}}_x) = 1, \tag{5}$$

where $\mathbf{C}_{aa} \stackrel{\text{def}}{=} \overrightarrow{\mathbf{a}}_x^y \overrightarrow{\mathbf{a}}_x^{y\top}$ is a rank-one matrix. Note that the solution to this problem is (up to a scaling) identical to the one of the original LCMV, both satisfying $\mathbf{w}_x \propto \mathbf{C}_{xx}^{-1} \overrightarrow{\mathbf{a}}_x^y$.

These insights now allow us to integrate the anatomical information contained in the lead field matrix $\mathbf{L}$ into the mSPoC objective. Specifically, we replace Eq. (4) by

$$((1-\gamma)\mathbf{C}_z + \gamma \mathbf{C}_{aa})\mathbf{w}_x = \lambda \mathbf{C}_{xx}\mathbf{w}_x \ , \tag{6}$$

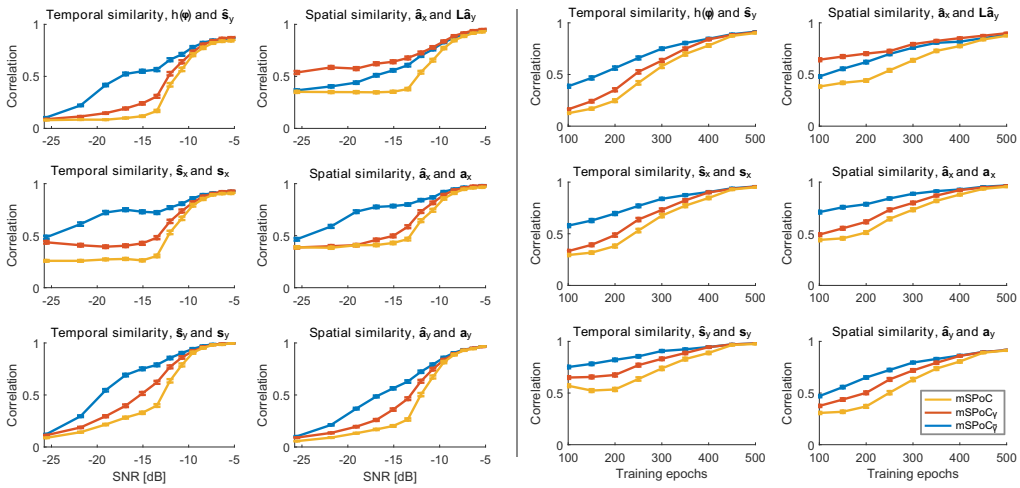
where $\gamma \in [0, 1]$ allows to smoothly interpolate between using only functional information ($\gamma = 0$, corresponding to the original mSPoC) and using only anatomical information ($\gamma = 1$). Thus $\gamma$ is the strength of the spatial constraint. We refer to this version of mSPoC as mSPoC$_\gamma$.

As a further variant, we propose to threshold the estimate of $\hat{\mathbf{a}}_y$ prior to the computation of $\overrightarrow{\mathbf{a}}_x^y$. Specifically, we propose to set all values in $\hat{\mathbf{a}}_y$ to zero, except for the maximum magnitude value. Thereby, we reduce the influence of noise while at the same time dealing with a known weakness of beamformers, namely a performance decrease when the spatial extent of the source increases [15]. This version of mSPoC will be referred to as mSPoC$_{\tilde{\gamma}}$.

### D. Simulated Data

The simulated EEG signal has 70 electrodes and the fMRI activity is assumed to be projected to the cortical surface consisting of 8196 vertices. The two modalities are set to have one source component in common. In addition to sharing temporal dynamics the component in the two modalities also share spatial patterns. The spatial pattern of the EEG component is thus the projection of the spatial pattern of the fMRI component through a lead field matrix. The lead field matrix was generated in SPM8 [16] using the boundary element method (BEM). Here a structural MRI scan of a subject was used; dataset was provided by Henson and Wakeman [3].

The generated EEG signal has frequencies between 8 and 12 Hz and has a sampling frequency of 200 Hz. The sampling




a) Influence of the SNR

b) Influence of the number of training epochs

Fig. 1. Correlations obtained using the original and augmented mSPoCs. a) Training was performed on 250 epochs ($\sim 4$ min) on data with a varying SNR. b) Training was performed on a varying number of epochs on data with SNR$= -12$ dB. The time course correlation is reported on a held out test set of 600 epochs (10 min data). Patterns are calculated on the training data and $\mathbf{L}$ is the lead field matrix. We show the mean of 500 repetitions together with the standard error on the mean. mSPoC$_\gamma$ and mSPoC$_{\bar{\gamma}}$ include crossvalidation to estimate $\gamma$.

frequency of fMRI is 1 Hz, meaning that the EEG is divided into epochs of 200 samples each. The fMRI component and the envelope of the paired EEG component have a frequency of maximum $1/4$ Hz, i.e the coupled activity oscillates with $1/4$ Hz or below. The fMRI signal is temporally delayed using a Gaussian shaped hemodynamic response function (HRF) with a length of 15 seconds ($N_\tau = 15$).

The EEG and fMRI additionally have 14 and 499, respectively, background source components. The spatial pattern of each source component is equivalent to a delimited Gaussian at the cortex level. The simulated EEG signal is a weighted sum over the Frobenius normalized source and background activity; i.e. $\mathbf{X} = \lambda \mathbf{X}_{source}/||\mathbf{X}_{source}||_F + (1-\lambda)\mathbf{X}_{background}/||\mathbf{X}_{background}||_F$. The SNR is here defined as $20 \log 10\,(\lambda/(1-\lambda))$. Same structure applies for the fMRI.

We vary both the SNR and the number of epochs available for training, and test the temporal correlation on 600 epochs. To estimate the level of anatomical information needed ($\gamma$) in mSPoC$_\gamma$ and mSPoC$_{\bar{\gamma}}$ we perform five-fold crossvalidation on the training data with $\gamma = (0, ..., 1)$ in 25 steps.

## III. Results

In Figure 1 we show the performance of the original mSPoC and the two augmented versions by their dependence on a) the SNR and b) the number of training epochs. The first row of Figure 1 shows the intermodal temporal and spatial similarity, i.e. the correlation between the estimated EEG and fMRI component. The middle row shows the temporal and spatial similarity of the true and estimated EEG component. The same is shown for fMRI in the bottom row. In real data applications we would only be able to show the top row.

As seen in Figure 1a) mSPoC$_\gamma$ and especially mSPoC$_{\bar{\gamma}}$ outperform the original mSPoC for most SNR levels. For higher SNRs ($> -10$ dB) the three methods converge to the same performance. It is evident from the middle and bottom rows of Figure 1a) that the augmentation improves the recovery of the true components both on the EEG side *and* on the fMRI side.

Superiority of mSPoC$_\gamma$ compared to mSPoC$_{\bar{\gamma}}$ is only found on the correlation of the estimated EEG pattern and the projected estimated fMRI pattern. This appears to be an overfitting of the EEG pattern to the noise components in the estimated fMRI pattern, as the spatial correlations of the estimated patterns of the two modalities to the true pattens do not show the same degree of improvement.

Analyzing Figure 1b) shows the same trend as Figure 1a), only now for the number of training epochs. The temporal correlation of the bandpowered EEG and the fMRI signal is highest using mSPoC$_{\bar{\gamma}}$ followed by mSPoC$_\gamma$. More specifically mSPoC$_{\bar{\gamma}}$ outperforms mSPoC when fewer than 400 epochs (here corresponding to 6.7 min) are available for training.

We investigate mSPoC$_{\bar{\gamma}}$ further in Figure 2 where we explore the influence on the temporal correlation of the SNR and the strength of the spatial constraint ($\gamma$). We see that when having low SNR mSPoC benefits from anatomical information and the biggest improvement can be found here compared to mSPoC (which corresponds to $\gamma = 0$). However, for higher levels of SNR relying too much on anatomical information

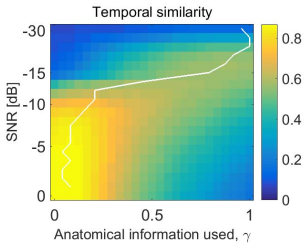



Fig. 2. For mSPoC$_{\bar{\gamma}}$ : The influence of $\gamma$ and the SNR on the correlation between the temporally convolved EEG component and the fMRI component, i.e. $h(\phi)$ and $\hat{s}_y$. A $\gamma = 0$ corresponds to the original mSPoC. The mean test correlation of 600 epochs across 500 data sets each trained on 250 epochs is shown. The white line indicates the optimum strength of the spatial constraint.

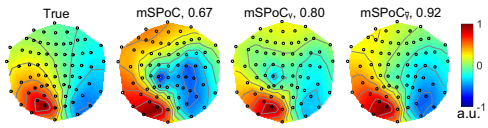



Fig. 3. An example of the estimated EEG patterns, where SNR$= -12$ dB and training is performed on 250 epochs ($\sim 4$ min). The correlations between the true and estimated patterns are shown.

worsen the solution and the spatial constraint should thus be small here.

An example of the estimated EEG patterns found by the three mSPoC versions is shown in Figure 3 together with the true pattern. The scalp map found with mSPoC$_{\bar{\gamma}}$ has highest correspondence with the true scalp map.

## IV. Discussion and Conclusion

The original multimodal Source Power Co-modulation (mSPoC) was in [10] compared to CCA where it was found to be superior. In this work we have introduced two expansions of mSPoC that exploit anatomical information. Both improve the extraction of correlated EEG and fMRI components. However, mSPoC$_{\bar{\gamma}}$ demonstrates higher potential for future use as it seems to avoid overfitting to noise components.

The next steps include investigating how mSPoC$_{\bar{\gamma}}$ reacts when certain assumptions are violated. We e.g. need to explore what happens when the paired components which are temporally correlated do not match spatially. The accuracy of the lead field matrix's projection of the fMRI to the EEG sensors could also be expected to influence the solution. The method should also be applied to real EEG and fMRI data to verify its applicability in multimodal neuroimaging. In the simulations we assumed that the fMRI activity was projected to the cortex. This projection can be performed with real data using a Voronoï interpolation scheme [17] or using SPM8 [16].

The preliminary results hold promise to improve the extraction of correlated pairs in situations where the SNR is low and/or where few training epochs are available. The method could thus potentially be used to investigate the correlation of the modalities as a function of time.

## Acknowledgment

LKH acknowledges support from the Danish Lundbeck Foundation via the Center for Integrated Molecular Brain Imaging (CIMBI). SD acknowledges funding by the German Research Foundation (DFG) grant no. MU 987/19-1 and support by the Bernstein Center for Computational Neuroscience, Berlin through the graduate program GRK 1589/1. KRM acknowledges support by the Brain Korea 21 Plus Program through the National Research Foundation of Korea funded by the Ministry of Education.

## References

[1] N. D. Schiff, "Multimodal neuroimaging approaches to disorders of consciousness," *The Journal of head trauma rehabilitation*, vol. 21, no. 5, pp. 388–397, 2006.

[2] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of neuroscience methods*, vol. 204, no. 1, pp. 68–81, 2012.

[3] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston, "A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration." *Frontiers in human neuroscience*, vol. 5, p. 76, 2011.

[4] S. M. Plis, V. D. Calhoun, M. P. Weisend, T. Eichele, and T. Lane, "MEG and fMRI Fusion for Non-Linear Estimation of Neural and BOLD Signal Changes." *Frontiers in neuroinformatics*, vol. 4, p. 114, 2010.

[5] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *Signal Processing Magazine, IEEE*, vol. 18, no. 6, pp. 14–30, 2001.

[6] F. Bießmann, S. M. Plis, F. C. Meinecke, T. Eichele, and K.-R. Müller, "Analysis of multimodal neuroimaging data," *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 26–58, 2011.

[7] S. Dähne, F. Bießman, W. Samek, S. Haufe, D. Goltz, C. Gundlach, A. Villringer, S. Fazli, and K.-R. Müller, "Multivariate machine learning methods for fusing functional multimodal neuroimaging data," *Proceedings of the IEEE*, 2015, accepted.

[8] X. Lei, P. A. Valdes-Sosa, and D. Yao, "EEG/fMRI fusion based on independent component analysis: integration of data-driven and model-driven methods." *Journal of integrative neuroscience*, vol. 11, no. 3, pp. 313–337, 2012.

[9] P. A. Valdes-Sosa, J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki, "Model driven EEG/fMRI fusion of brain oscillations." *Human brain mapping*, vol. 30, no. 9, pp. 2701–2721, 2009.

[10] S. Dähne, F. Bießmann, F. C. Meinecke, J. Mehnert, S. Fazli, and K.-R. Müller, "Integration of multivariate data streams with bandpower signals," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1001–1013, 2013.

[11] S. Dähne, F. C. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K.-R. Müller, and V. V. Nikulin, "SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters," *NeuroImage*, vol. 86, no. 0, pp. 111–122, 2014.

[12] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, 2014.

[13] F. Bießmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. K. Logothetis, and K.-R. Müller, "Temporal kernel cca and its application in multimodal neuronal data analysis," *Machine Learning*, vol. 79, no. 1-2, pp. 5–27, 2010.

[14] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24.

[15] M. J. Brookes, C. M. Stevenson, G. R. Barnes, A. Hillebrand, M. I. G. Simpson, S. T. Francis, and P. G. Morris, "Beamformer reconstruction of correlated sources using a modified source model," *NeuroImage*, vol. 34, no. 4, pp. 1454–1465.

[16] J. Ashburner, C.-C. Chen, R. Moran, R. N. Henson, V. Glauche, and C. Phillips, "SPM8 manual," The FIL Methods Group, Tech. Rep., 2012.

[17] C. Grova, S. Makni, G. Flandin, P. Ciuciu, J. Gotman, and J. B. Poline, "Anatomically informed interpolation of fMRI data on the cortical surface." *NeuroImage*, vol. 31, no. 4, pp. 1475–86, 2006.

Appendix G

# Spatio-Temporal Reconstruction of Brain Dynamics from EEG with a Markov Prior

Postprint of article published in NeuroImage.

Hansen, S.T., & Hansen, L.K. (2017). "Spatio-Temporal Reconstruction of Brain Dynamics from EEG with a Markov Prior". *NeuroImage*, vol 148, pp. 274-283.
doi: http://dx.doi.org/10.1016/j.neuroimage.2016.12.030

# Spatio-temporal reconstruction of brain dynamics from EEG with a Markov prior

Sofie Therese Hansen[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) can capture brain dynamics in high temporal resolution. By projecting the scalp EEG signal back to its origin in the brain also high spatial resolution can be achieved. Source localized EEG therefore has potential to be a very powerful tool for understanding the functional dynamics of the brain. Solving the inverse problem of EEG is however highly ill-posed as there are many more potential locations of the EEG generators than EEG measurement points. Several well-known properties of brain dynamics can be exploited to alleviate this problem. More short ranging connections exist in the brain than long ranging, arguing for spatially focal sources. Additionally, recent work (Delorme et al., 2012) argues that EEG can be decomposed into components having sparse source distributions. On the temporal side both short and long term stationarity of brain activation are seen. We summarize these insights in an inverse solver, the so-called "Variational Garrote" (Kappen and Gómez, 2013). Using a Markov prior we can incorporate flexible degrees of temporal stationarity. Through spatial basis functions spatially smooth distributions are obtained. Sparsity of these are inherent to the Variational Garrote solver. We name our method the MarkoVG and demonstrate its ability to adapt to the temporal smoothness and spatial sparsity in simulated EEG data. Finally a benchmark EEG dataset is used to demonstrate MarkoVG's ability to recover non-stationary brain dynamics.

*Keywords:* Source reconstruction, Bayesian inference, the Variational Garrote, EEG, Inverse problem, Temporal dynamics

## 1. Introduction

The large body of event-related potential (ERP) studies demonstrates that EEG is a productive tool for detailed and accurate understanding of brain dynamics. While ERP studies are typically based on the native scalp electrode measures, imaging of human brain dynamics is gaining interest. Imaging by source reconstruction solves one of the main issues with EEG scalp studies, namely the limited spatial specificity due to volume conduction (Nunez et al., 1997). EEG imaging is obtained by solving an inverse problem, where the measured EEG scalp data is used to reconstruct the location and strength

*Corresponding author
*Email addresses:* sofha@dtu.dk (Sofie Therese Hansen), lkai@dtu.dk (Lars Kai Hansen)

of the signal's cortical sources. However, the inverse problem is very ill-posed as the number of possible source locations exceeds the number of EEG electrodes by orders of magnitude (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994; Baillet et al., 2001; Hulbert and Adeli, 2013; De Ciantis and Lemieux, 2013). The inverse problem is based on a forward model which describes the mapping from a large set of hypothetical local sources to a smaller number of scalp electrodes. The forward model is constructed from electrophysiological first principles based on anatomical data and assumed values of conductivities of the various tissues; scull, scalp, etc. Attempts have been made at estimating the forward model from the EEG data, see e.g., (Stahlhut et al., 2011; Akalin Acar et al., 2016; Hansen et al., 2016). However, in the following we will assume the forward model known and focus on solving the inverse problem.

Although no gold standard EEG inverse solver has been established, the field is converging on methods that employ spatial sparsity (Gorodnitsky and Rao, 1997; Wipf and Rao, 2007; Vega-Hernández et al., 2008; Friston et al., 2008; Zhang and Rao, 2011; Stahlhut et al., 2011; Montoya-Martinez et al., 2012; Gramfort et al., 2013; Hansen et al., 2013c; Hansen and Hansen, 2014; Andersen et al., 2014). Evidence was presented, in recent work (Delorme et al., 2012) that the instantaneous independent components of EEG signals are dipolar and localized. In particular it was shown that the residual variance after a dipole fit to the component scalp maps is less than 5% for large fractions of the independent components. ICA can thus provide sparse source distributions supporting the search for sets of localized sources, and motivates reconstruction algorithms that emphasize sparsity in contrast to the distributed spatial source patterns promoted in classical alternatives (Pascual-Marqui et al., 1994, 2002). The connectivity of the brain tissue speaks in favor of focal and sparse solutions in general as there exists more short than long ranging cortical connections (Schüz and Braitenberg, 2002; Markov et al., 2011), enabling local coordination at typical EEG time scales. Sparsity can furthermore result from averaging repetitions of stimuli leaving only focal or a sparse network of activity, as in ERP studies.

Imaging strategies can in general be divided into two categories, each having their own limitations. Dipole fits assume the number of active dipoles to be fixed and estimate their locations (Scherg and Von Cramon, 1985). Meaningful solutions thus rely on a qualified guess at the number of active dipoles. In contrast, distributed imaging approaches estimate the source strength in a large number of source locations (Gorodnitsky et al., 1995; Friston et al., 2008). These methods thus avoid making subjective assumptions, but do render the EEG inverse problem underdetermined. Constraints are therefore needed to obtain unique solutions. These can, however, be formulated based on physiological assumptions (Haufe et al., 2008) and spatial priors obtained from other neuroimaging modalities can be incorporated (Henson et al., 2010).

As we are interested in brain dynamics, the goal is to reconstruct not only sources at a given moment in time, but rather the spatio-temporal source distribution from a sequence of scalp measurements. To stabilize the solution it is useful to impose some level of temporal smoothness. A basic scheme is to enforce that the locations of activity are fixed throughout an analysis window (Wipf and Rao, 2007; Friston et al., 2008; Ou et al., 2009; Zhang and Rao, 2011; Hansen et al., 2013c). While useful for short time windows, this may be less appropriate for more extended and non-stationary settings. Recently proposed methods enforce temporal coherency while also allowing for dynamic activation patterns (Montoya-Martinez et al., 2012; Gramfort et al., 2013). These methods model

the temporal dynamics more realistically by assuming brain areas to be sequentially or simultaneously activated during, e.g. a stimulus after which the activity returns to a baseline level (Gramfort et al., 2013). Both methods employ a mixed-norm scheme to recover what is hypothesized to be a structured sparsity pattern across time, see also (Haufe et al., 2008; Gramfort et al., 2012). These types of convex relaxation schemes are very interesting and are frequently applied to solve inverse problems in general (Vega-Hernández et al., 2008). We have started investigations in a recent alternative for sparse recovery proposed in (Kappen, 2011; Kappen and Gómez, 2013). The approach, called the Variational Garrote (VG), solves the sparse recovery problem directly without resorting to convex relaxation. In addition VG enables separation of the variables encoding the source locations and source magnitudes, which is relevant when a given dipole is active for an extended period (i.e., location is smooth in time) in which the activation magnitude involves high-frequency changes. Finally, a Bayesian inference scheme leads to a relatively low-complexity set of non-linear equations that are iterated towards the solution.

The contribution of the present paper is to advance our understanding of this new algorithm. VG has been applied to EEG brain imaging, and extended to the spatio-temporal domain by assuming a fixed sparsity profile in time windows (Hansen et al., 2013c,b). In this presentation our aim is to replace the fixed sparsity model with a more flexible Markov prior, which in a preliminary unpublished workshop note was named "MarkoVG" (Hansen and Hansen, 2013). Here we further develop this model by including spatial basis functions to obtain focal smooth sources, and improve on the optimization scheme. In the following sections we analyze our proposed inverse solver and show how the model's degree of spatial and temporal sparsity can be adapted to fit the data. Finally we demonstrate MarkoVG's application to the spatio-temporal reconstruction of the EEG response to a face perception task.

### *1.1. Notation*

In the following we have defined $\mathbf{X}^\top$ and $X_{ij}$ as the transpose and the scalar element in row $i$ and column $j$ of the matrix $\mathbf{X}$, respectively. Capital bold thus indicates a matrix, a vector is in lower case and in bold font, while a scalar is in normal font, either in lower or upper case. The $L_2$-norm of $\mathbf{x}$ is denoted by $||\mathbf{x}||_2$.

## 2. Methods

### *2.1. The Variational Garrote*

At the frequencies relevant for EEG acquisition the scalp EEG can be considered as a linear combination of the underlying brain activity (Baillet et al., 2001). The VG is therefore immediately applicable as it provides a framework to solve a linear inverse problem by imposing a "spike-and-slab" like representation (Ishwaran and Rao, 2005). For $T$ time samples the linear relation between $N$ possible brain sources, $\mathbf{X} \in \mathbb{R}^{N\times T}$, and $K$ EEG recordings, $\mathbf{Y} \in \mathbb{R}^{K\times T}$, is given by the forward model, $\mathbf{A} \in \mathbb{R}^{K\times N}$. This relation is modified in VG (Kappen, 2011) by introducing binary variables $\mathbf{S} \in \{0,1\}^{N\times T}$ that dictate the spatio-temporal activation states (inactive or active), i.e.

$$Y_{kt} = \sum_{n=1}^{N} A_{kn} S_{nt} X_{nt} + E_{kt}. \tag{1}$$

We assume the noise, $E_{kt}$, to be i.i.d. with zero mean and normally distributed with scalar unknown precision $\beta$. As seen in eq. (1) there will for each dipolar location $n$ and time sample $t$ be an estimate of its state $S_{nt}$ *and* its dipolar strength $X_{nt}$. The VG therefore supplies a framework which is highly flexible for including different priors into the solver.

### *2.2. Temporal coherence*

We now impose temporal coherence through the binary variable $S_{nt}$. In a previous study we suggested to enforce a strict prior on the temporal smoothness by keeping $S_{nt}$ fixed for each source for a given time window, while allowing for the activity strength, $X_{nt}$, to vary (Hansen et al., 2013c). Here we adapt the VG to provide a data-driven flexible degree of temporal smoothness of $S_{nt}$ by imposing a Markov prior on this variable (Hansen and Hansen, 2013; Hansen et al., 2013a). For dipole location $n$ the transition probabilities of $S_{nt}$ thus depend on the activation state at time sample $t-1$ and is given by $\Gamma_{ji} = \mathrm{P}(S_{nt} = j|S_{n,t-1} = i)$, where $i, j = 0, 1$. The full transition matrix can be described by two parameters, as $\Gamma_{00} + \Gamma_{10} = 1$ and $\Gamma_{01} + \Gamma_{11} = 1$, and is given by

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix} = \begin{bmatrix} 1-\Gamma_{10} & \Gamma_{01} \\ \Gamma_{10} & 1-\Gamma_{01} \end{bmatrix}. \tag{2}$$

Through different combinations of $\Gamma_{10}$ and $\Gamma_{01}$ the Markov prior thus enables flexibility in both temporal smoothness and spatial sparsity. Temporal smoothness is for example achieved by having large probabilities of staying in a state, i.e. large $\Gamma_{00}$ and $\Gamma_{11}$, while spatial sparsity is achieved by large probabilities of staying in or switching to an inactive state, i.e. large $\Gamma_{00}$ and $\Gamma_{01}$.

Lucka et al. suggested to use hierarchical Bayesian methods to solve the inverse problem of EEG using fully Bayesian inference methods (Lucka et al., 2012). Practically, Lucka et al. showed these methods' strengths in the single-measurements setup for especially deeply located sources. Here we follow Kappen et al., and instead solve the inverse problem by turning to approximate variational Bayesian inference (Kappen, 2011; Kappen and Gómez, 2013). First we define the posterior

$$p(\mathbf{S}, \mathbf{X}, \beta|\mathbf{D}, \mathbf{\Gamma}) = \frac{p(\mathbf{X}, \beta)p(\mathbf{S}|\mathbf{\Gamma})p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta)}{p(\mathbf{D}|\mathbf{\Gamma})}, \tag{3}$$

where $\mathbf{D}$ is the data. Since we intend to optimize the posterior with respect to the source dipole activations we can ignore the denominator of eq. (3). Next we follow Kappen et al. by 1) assuming a flat prior on $\mathbf{X}$ and $\beta$, 2) marginalizing over $S_{nt}$, and 3) introducing the variational approximation $q(\mathbf{S}) = \prod_{n=1}^{N} q_{nt}(S_{nt})$, where $q_{nt}(S_{nt}) = M_{nt}S_{nt} + (1 - M_{nt})(1 - S_{nt})$. $M_{nt} \in [0, 1]$ describes the posterior probability of source $n$ being active at time sample $t$, corresponding to the probability of $S_{nt}$ being 1. The marginal log-likelihood is by these definitions

$$\log \sum_{\mathbf{S}} p(\mathbf{S}|\Gamma)p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta) \geq -\sum_{\mathbf{S}} q(\mathbf{S}) \log \left( \frac{q(\mathbf{S})}{p(\mathbf{S}|\Gamma)p(\mathbf{D}|\mathbf{X}, \mathbf{S}, \beta)} \right) = -F(q, \mathbf{X}, \beta). \tag{4}$$

where $F$ is an estimate of model evidence and is the so-called "variational free energy". As the free energy describes an upper bound on the negative log-likelihood it is minimized

to find the optimal solution. As suggested in (Kappen, 2011; Kappen and Gómez, 2013) we apply a *dual formulation* to reduce the computational complexity by defining $Z_{kt} = \sum_n A_{kn} M_{nt} X_{nt}$ and Lagrange multipliers, $\lambda_{kt}$. The free energy with the Markovian prior can be derived from eq. (4) and is given by

$$\begin{aligned} F = &- \frac{KT}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k} (Y_{kt} - Z_{kt})^2 + \frac{K\beta}{2} \sum_{t,n} M_{nt}(1 - M_{nt}) X_{nt}^2 \chi_{nn} \\ &- \sum_{n,t} M_{nt} \log \frac{\Gamma_{10}}{\Gamma_{00}} + M_{n,t-1} \log \frac{\Gamma_{01}}{\Gamma_{00}} + (M_{nt} M_{n,t-1}) \log \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \\ &+ NT \log \frac{1}{\Gamma_{00}} + \sum_{n,t} \left[ M_{nt} \log(M_{nt}) + (1 - M_{nt}) \log(1 - M_{nt}) \right] \\ &+ \sum_{t,k} \lambda_{kt} \left( Z_{kt} - \sum_n A_{kn} M_{nt} X_{nt} \right). \end{aligned} \tag{5}$$

Here we define $\boldsymbol{\chi} \in \mathbb{R}^{N \times N}$ to be the covariance of the forward model $\mathbf{A}$.

Calculating the partial derivatives of the free energy and equating them to zero, yields the following equation set

$$X_{nt} = \frac{1}{K\beta} \frac{1}{(1 - M_{nt})\chi_{nn}} \sum_k \lambda_{kt} A_{kn}, \quad Z_{kt} = Y_{kt} - \frac{1}{\beta} \lambda_{kt}, \tag{6}$$

$$\beta = \frac{1}{TK} \sum_{t,k,c} \lambda_{kt} \lambda_{ct} C_{kct} \tag{7}$$

$$\text{where } \; C_{kct} = \delta_{kc} + \frac{1}{K} \sum_n \frac{M_{nt}}{(1 - M_{nt})\chi_{nn}} A_{kn} A_{cn}, \tag{8}$$

$$\lambda_{ct} = \beta \hat{Y}_{ct} \text{ def. } \sum_c C_{kct} \hat{Y}_{ct} = Y_{kt}, \tag{9}$$

$$M_{nt} = \sigma \left( \frac{K\beta}{2} \chi_{nn} X_{nt}^2 + \gamma_1 + \gamma_2 \left( M_{n,t-1} + M_{n,t+1} \right) \right), \tag{10}$$

where $\sigma(x) = (1 + \exp(-x))^{-1}$ and where the estimated source strength of source $n$ in time sample $t$ is given by $V_{nt} = M_{nt} X_{nt}$. While Kappen et al. solve the equation set by iteration, we implement gradient descent for the variational mean to ensure convergence. The complexity of the equations set is dominated by the computation of the tensor $\mathbf{C}$ and its inversion; which are of order $NK^2T$ and $K^3T$, respectively.

Inspecting the modified VG equations, it is clear that the combination of the Markov parameters $\gamma_1 = \log \left( \frac{\Gamma_{10}\Gamma_{01}}{\Gamma_{00}^2} \right)$ and $\gamma_2 = \log \left( \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \right)$ dictates how sparse and smooth the solution will be. The parameter $\gamma_2$ thus determines the degree of temporal smoothness, and $\gamma_1$ corresponds to a sparsity control parameter, where more negative values will yield more sparse solutions. We note that if $\Gamma_{01} + \Gamma_{10} = 1$ then $\gamma_2 = 0$ and the original VG formulation of the variational mean is obtained.

In Fig. 1 we show how applying different combinations of sparsity and smoothness affect the MarkoVG solution in a simulation. In the example we synthesized 25 time samples containing non-stationary support on the activation for two out of 500 sources

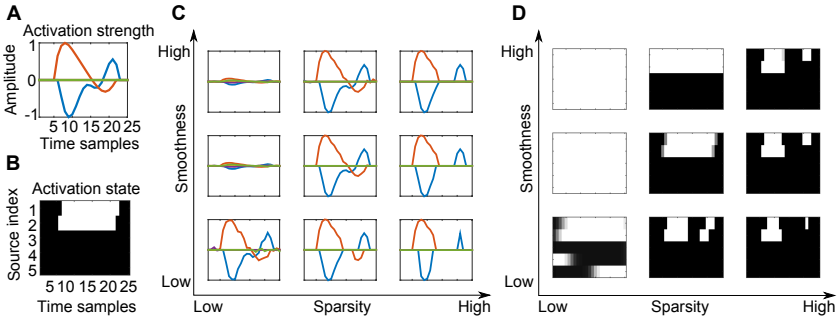



Figure 1: Example of sparsity and smoothness dependency. A) The simulated signal. Two out of 500 sources are active. These were projected through a random forward model with 50 observations and added with noise to give a signal-to-noise ratio (SNR) of 10 dB. B) True activation states for five of the sources. White indicate active state while black illustrates an inactive state. Source 1 and 2 were active in some time samples, and the remaining 498 were completely turned off. C) and D) The MarkoVG estimated signal and their activation state for different combinations of sparsity and smoothness degree. The parameter setting for the solution in the mid insets of C) and D) was found through four-fold cross-validation on the 50 observations.

(Fig. 1A and B). A random forward model of size $50 \times 500$ was used to project the signal to 50 observations. From Fig. 1C and D we see that if the sparsity is set too high some of the relevant time samples are turned off and that this can only be partly remedied by a high temporal smoothness. If on the other hand the sparsity is too low we obtain activity in other than the relevant source locations. It is furthermore evident from Fig. 1 that having too little temporal smoothness will cause the solver to miss activation in the time samples of low magnitude activity. However, the temporal smoothness must not be too large or activity outside the activated period will emerge. We can therefore conclude that only the right amount of temporal smoothness improves the solution. Importantly we also demonstrate that we are able to match the true signal's properties (Fig. 1A and B) using four fold cross-validation to find the optimal level of sparsity and smoothness (mid inset in Fig. 1C and D).

### *2.3. Spatial coherence*

EEG activity arises when regional active neurons are active in synchrony (Baillet et al., 2001) and therefore many EEG inverse solvers incorporate an assumption of spatial smoothness (Phillips et al., 2002; Pascual-Marqui et al., 2002; Friston et al., 2008; Haufe et al., 2008). To obtain a spatially smooth source distribution we introduce spatial basis functions following the implementation described in the multiple sparse priors model (MSP) (Friston et al., 2008). The basis functions are based on the adjacency matrix, which describes how the source space is connected. The connectivity contained in the adjacency matrix is propagated to neighbors' neighbors in eight steps and finally a thresholding is performed. This translates to basis functions extending from their center to maximally their eighth-order neighbors (Fig. 2A). The degree of smoothness is controlled by a parameter that is set as suggested in (Friston et al., 2008). With this setting each basis

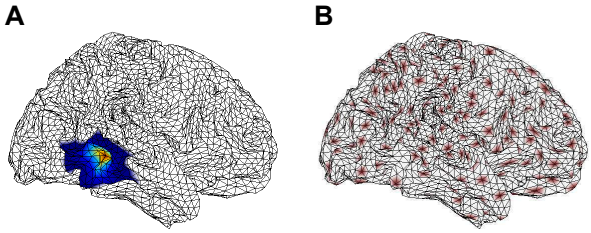



Figure 2: The spatial basis functions. A) Example of one basis function's spatial distribution. Red/blue indicate high/low numeric activity. B) Centers of the 776 sampled basis functions.

function covers between 98-128 dipoles when the cortex surface has been segmented into a mesh of 8196 nodes.

In order to reduce the complexity of the inverse solver not all dipoles (or nodes) in the source space will serve as centers of basis functions. Friston et al. placed the centers by first sampling 256 evenly spaced source indices (Friston et al., 2008), and then also included their symmetrically located sources on the other hemisphere. Finally the hemispherical symmetric centers were paired to create 256 additional basis functions. In total 768 basis function were created with 512 representing unilateral activity and 256 bilateral activity (Friston et al., 2008).

To ensure an even distribution in space, rather than in the source index, we propose to sample the basis function centers based on the connectivity of the mesh and thereby obtain better coverage of the cortex. We thus let the adjacency matrix determine whether a randomly sampled center should be included. The precise requirement is that there must be at least three vertices between all centers. By seeding the random generator the locations of the basis function centers are controlled. With our applied seeding, 776 centers are obtained, shown in Fig. 2B. In both sampling techniques all locations are part of more than 1 basis function. In the original method each source location is included in 3 to 17 basis functions while in our method each source location is a part of 8 to 16 basis functions.

### *2.4. MarkoVG*

The inference scheme for VG with a Markovian prior is explained in eqs. (1)-(10) and seen implemented in `https://github.com/STherese/VG_inverse_solvers`. The spatial basis functions are included simply by projecting the forward model $\mathbf{A}$ onto these. This produces the reduced forward model $\mathbf{A}_{\text{basis}} = \mathbf{AB}$, where $\mathbf{B}$ contains the basis functions in the columns. The optimum level of sparsity and temporal smoothness is determined through four-fold cross-validation on the electrodes. The free energy is used to estimate the optimum setting for each fold, and the median across these defines the parameter setting. Since we are working with EEG recorded from 70 electrodes 17-18 electrodes are in each fold, and we therefore believe that it is likely that the brain activity will be seen to some degree by all four folds. We compare our proposed method with the below three inverse solvers, which all produce temporally stationary source distributions.

### 2.5. Multiple sparse priors

MSP (Friston et al., 2008) specifies another Bayesian approach of finding sparse source distributions. As previously described, inference is based on a number of spatial basis functions with compact support, these are pruned or incorporated in the solution through a restricted maximum likelihood procedure. Smooth temporal source distributions are obtained by creating temporal projectors from the EEG signal. In effect this creates rather stationary temporal activation patterns. The implementation used in the following experiments is from the SPM12 software (Ashburner et al., 2014).

### 2.6. T-MSBL

T-MSBL (Zhang and Rao, 2011) is an extension of the multiple measurement vectors (MMV) sparse Bayesian learning (SBL) (Wipf and Rao, 2007) method which exploits temporal correlation to obtain smooth temporal dynamics. T-MSBL assumes a block-structure where temporal correlations are modeled in the blocks. Automatic relevance determination (ARD) (Hansen and Rasmussen, 1994; MacKay, 1995) is applied to identify the active sources (blocks) and prune the irrelevant. In the following experiments we employ the implementation provided in the toolbox by Zhilin Zhang `http://dsp.ucsd.edu/~zhilin/TMSBL.html`. We apply two versions of T-MSBL. In the first version we follow the recommendations of the toolbox and set the noise level to "large" if SNR$<$ 6 dB and to "mild" if SNR$\geq$6. These two intervals translate into two numerical noise levels. We refer to the first version as "T-MSBL" which is partly favored in the simulations as the true noise level is provided. In the second version we perform four-fold cross-validation to estimate the regularization parameter, we call this version "T-MSBL cross".

### 2.7. M-FOCUSS

The FOCal Underdetermined System Solver (FOCUSS) employs a reweighted norm minimization and finds sparse solutions by defining the regularization norm to be equal to or less than 1 (Gorodnitsky and Rao, 1997). M-FOCUSS is an MMV extension developed in (Cotter et al., 2005) and also here extended to be applicable to noisy data. The latter version is the so-called regularized M-FOCUSS which performs iterative weighting using the diagonal matrix $\mathbf{W} \in \mathbf{R}^{N\times N}$ to find the dipole estimates $\mathbf{X}$, i.e. in iteration $it$

$$\mathbf{W}_{n,n}^{(it)} = ||X_{n,:}^{(it-1)}||_2^{1-p/2}, \text{ with } p \in [0,2] \tag{11}$$

$$\mathbf{X}^{(it)} = \mathbf{W}^{(it)}\mathbf{W}^{(it)^\top}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^{(it)}\mathbf{W}^{(it)^\top}\mathbf{A}^\top + \lambda\mathbf{I})^{-1}\mathbf{Y}. \tag{12}$$

We use the implementation of the regularized M-FOCUSS provided in the same toolbox as the T-MSBL algorithm. The regularization parameter, $\lambda$, can be approximated by the noise variance of the data (Zhang and Rao, 2011) and in the simulations we therefore use the exact noise variance for this parameter. The M-FOCUSS is thus favored in the simulations. As suggested by Cotter et al. we set the norm to be $p = 0.8$, which should, according to the authors, provide a reasonable balance between being sparse and not having to many local minima.

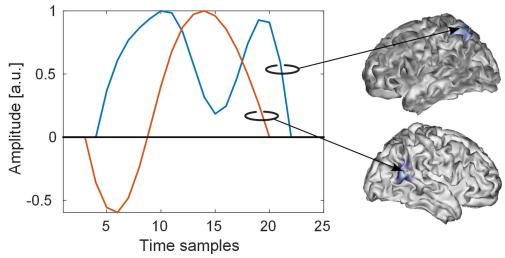



Figure 3: Example of two simulated sources. In each repetition one to four sources were planted.

*2.8. Simulations*

We first evaluate MarkoVG in a simulation study. In line with previous EEG simulation studies (Friston et al., 2008; Stahlhut et al., 2011; Montoya-Martinez et al., 2012; Gramfort et al., 2013) we generated synthetic EEG signals by randomly planting one to four sources and projecting their temporal dynamics through a forward model. The sources were modeled as having a spatial distribution given by the earlier described basis functions where the centers could be placed in any of the dipoles of the mesh (and not only in the reduced set used for reconstruction). The source signal was projected to scalp EEG electrodes through a forward model generated for subject "A" in the real data experiment described below. The forward model contains the projection of 8196 dipoles to 70 EEG electrodes.

The temporal dynamics of the sources were each generated from random white noise which was low pass filtered to yield frequency content up to 20 Hz. We created a signal of 25 time samples. To obtain varying degrees of non-stationarity we only kept the activity in the mid section of these time samples. An example of the temporal dynamics of a set of sources is shown in Fig. 3. Noise was added to yield SNRs of 0 to 14 dB.

Across the applied SNRs 100 data sets were used to compare MarkoVG to MSP, T-MSBL and M-FOCUSS. All inverse solvers used the 776 earlier described basis functions in their reconstruction. The performance was judged based on a source retrieval score called the $F_1$-*measure* (Rijsbergen, 1979; Makhoul et al., 1999), as well as the source localization error. The $F_1$-*measure* is defined as

$$F_1\text{-}measure = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{TP + FP + P}, \tag{13}$$

where $TP$, $FP$ and $P$ are the true, false and actual positives, respectively. It is noted that this is a rather strict measure as only a perfect correspondence between planted and estimated activity will yield perfect source retrieval, i.e. $F_1\text{-}measure = 1$. Since we only used a subset of the possible basis function centers (776 out of 8196) to reconstruct from, perfect reconstruction was only obtainable when the planted sources were basis functions from the subset. However, since the basis functions describe locally coherent activation some of the actual activity can be recovered even if the planted source component is not directly contained in the set used for reconstruction.

We defined the localization error as the Euclidean distance between each estimated source/dipole and the nearest planted source. The reported error is the average over all

estimated sources at all time samples containing simulated activity. We proceeded in this way as considering only the maximum magnitude source from the estimation would disregard any spurious activity located far from the true sources.

### *2.9. Benchmark EEG data*

To further investigate MarkoVG we applied it to EEG recorded during a well studied paradigm, namely the multi-subject multimodal dataset studying face recognition (Wakeman and Henson, 2015). Images of famous faces, unfamiliar faces and scrambled faces were presented to 19 subjects in six runs of 7.5 minutes. We investigated face perception from the 70-channel EEG data recorded in run 1 for three subjects, here termed "A", "B" and "C". In this run the subjects were presented with approximately 50 famous, 50 unfamiliar and 50 scrambled faces. As we are interested in finding the response to faces we averaged over the two face conditions and subtracted the average of the scrambled face condition (see EEG sensor data in Supplementary Fig. 1 and 2A). For further information on the experimental setup used in the data collection we refer to the documentation provided by Wakeman et al. (Wakeman and Henson, 2015). We built forward models in SPM8 using a three layered boundary element method head model (Phillips, 2000). The head model was the result of segmenting T1-weighted MRI scans of the three investigated subjects.

It has been shown that face perception exhibits partially bilateral activation (Eimer and McCarthy, 1999; Henson et al., 2009). In the source reconstruction we therefore employed the basis functions set described by Friston et al. comprising both unilateral and bilateral basis functions. For comparison we also show the solutions obtained using MSP, T-MSBL and M-FOCUSS; also with the basis function set described by Friston et al. As the EEG signal was averaged over many repetitions we judged the noise level needed for T-MSBL to be "mild". For M-FOCUSS an estimate of the noise variance was calculated based on a 100 ms pre-stimulus window.

## 3. Results

### *3.1. Simulations*

In Fig. 4 we investigated the effect different combinations of sparsity and smoothness levels have on the MarkoVG solution. Note that these combinations have to respect the specification of the prior probabilities, i.e. the columns of the matrix in eq. (2) must sum to 1 and have elements with values between 0 and 1. The relevant combinations of sparsity and temporal smoothness were therefore contained in the band shown in Fig. 4.

Fig. 4A demonstrates that low free energy calculated on the validation folds in a cross-validation scheme coincides with low localization error (Fig. 4B) and high $F_1$-*measure* (Fig. 4C). This is evidence that the free energy can be used to optimize parameters for performance. The optimal solution was located in the vicinity of the dashed line where $\gamma_2 = -\gamma_1$ and more precisely just above the dashed line, particularly when considering the $F_1$-*measure*. In the following we assumed the relation $\gamma_2 = -0.9\gamma_1$, and thus reduced the search space of the optimal parameter setting from two to one parameter. Some intuition on the implications of the defined relation can be gained by propagating the relation to the transition probabilities, i.e., setting $\gamma_2 = -\gamma_1$ implies that the probability of staying in an inactive and active state are equal ($\Gamma_{00} = \Gamma_{11}$). Hence, implies that transitioning from an inactive to an active state is as likely as the reverse. In this case

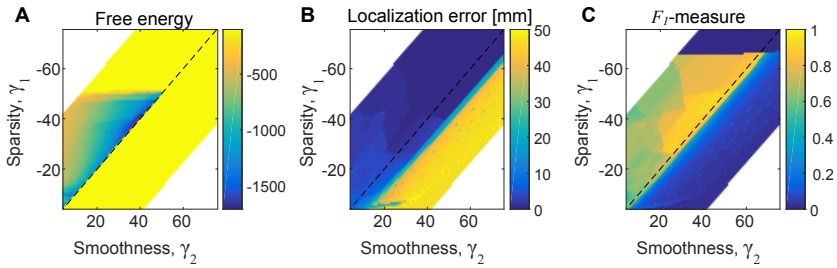



Figure 4: Example of the dependence of sparsity and smoothness on the MarkoVG solution illustrated on simulated data. The simulated source distribution for this example can be seen in Fig. 3 and was created using an EEG forward model. The black dashed line indicates $\gamma_1 = -\gamma_2$. White areas in the plots illustrate where combinations of $\gamma_1$ and $\gamma_2$ are not meaningful, see text. A) The free energy calculated on the validation sets in a four-fold cross-validation scheme. Shown is the mean across these folds. B) The localization error averaged across time and estimated sources, and C) $F_1$-*measure* of the MarkoVG solution; 0 indicates no correct sources are retrieved/many false sources are retrieved and 1 indicates all correct and no false sources are recovered.

there is no sparsity bias, only temporal smoothness is enforced if $\Gamma_{00} > 0.5$. To promote sparse solutions we heuristically applied the factor 0.9, based on complete scans of the parameter space as seen in Fig. 4.

In Fig. 5 we applied the above mentioned sparsity-smoothness relation and performed cross-validation on one parameter for MarkoVG. We compared MarkoVG to MSP, T-MSBL, T-MSBL with cross-validation, and M-FOCUSS, in 100 simulations across different levels of noise. It can be observed that MarkoVG achieved the best localization error and that MarkoVG and T-MSBL outperformed the other methods with respect to the $F_1$-*measure*. We further observe that cross-validation was not effective for T-MSBL.

### *3.2. Face perception EEG data*

Fig. 6 presents the source distributions of the face perception data as estimated by MSP, T-MSBL, M-FOCUSS and MarkoVG for three subjects. The temporal dynamics of the basis function/source component having highest activity in the time interval 130 to 200 ms after stimuli onset is shown in blue in the top panel with the locations marked in blue in the lower panel. The second largest valued basis function is similarly shown in red. These basis functions were bilateral in several of the shown examples.

Most solvers recovered contrast activity in or close to the expected areas, i.e. the left and right occipital face areas (OFAs) and fusiform face areas (FFAs). More specifically MarkoVG placed the strongest activation (blue circles) in or near the FFA for all subjects. Focusing on subject "A" the strongest basis function for T-MSBL, M-FOCUSS and MarkoVG were located in the FFAs, marked with blue circles in Fig. 6A. All four methods found activation in the OFAs; MSP had its strongest activated sources close to the OFA, the same for the second strongest activation for M-FOCUSS and finally MarkoVG had its second strongest activation in the OFA.

The presented inverse solvers generally showed the well-known temporal response to viewing faces, i.e. the N170 ERP component. The N170 component normally appears 130-200 ms after presentation of a face (Itier and Taylor, 2004). The solvers peaked

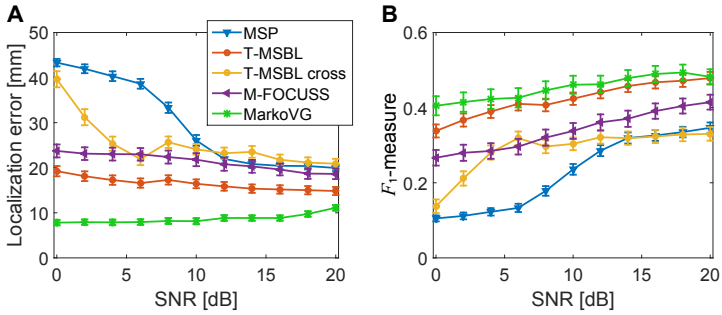



Figure 5: Performance on simulated data; created with a real EEG forward model. For each 11 SNR levels 100 simulations were run with simulated activity consisting of one to four sources (basis functions) randomly placed on the cortex, each having non-stationary temporal dynamics, see example in Fig. 3. Errorbars indicate standard error of the mean. A) Localization error averaged across time and sources. B) The source retrieval score, $F_1$-*measure*; 1 indicates optimal retrieval. Note, perfect performance is not expected as the locations of the planted sources were drawn from the entire mesh and sought reconstructed based on a subset.

between 150 ms and 170 ms after stimuli, however, the N170 peak was less defined for MSP, TMSBL and M-FOCUSS for subject "B". It is further noted that MarkoVG differed from the other methods by being temporally as well as spatially more sparse.

In Fig. 7 we show the averaged observed ERP as well as the ERP predicted by MarkoVG. These are quite similar, however with a slight bias towards zero of the MarkoVG predicted ERPs. This is similarly demonstrated for MSP, T-MSBL and M-FOCUSS in Supplementary Fig. 1. To avoid scaling issues in a comparison we show in supplementary Fig. 2B the temporal correlations across channels between the observed and predicted EEG signals for subject "A". These were again similar for all methods.

## 4. Discussion

Solving the ill-posed inverse problem of EEG and obtaining detailed spatio-temporal knowledge of cognitive processes require us to make relevant prior assumptions on the solution. Such assumptions should be based on prior knowledge of the brain, for example, from brain anatomy and physiology. Common assumptions include on the spatial side sparsity and smoothness, meaning that the source distribution of interest is believed to consist of relatively few source patches, each having temporally coherent source strength. Sparsity is a common assumption when solving ill-posed inverse problems in general as it mitigates the non-uniqueness of the problem (Donoho et al., 2006). In EEG imaging it is, as mentioned earlier, further motivated by the existence of more short ranging connections than long ranging. Sparsity has previously been obtained through regularization of the inverse problem, e.g., by imposing the $L_p$-norm, where $p \leq 1$ (Gorodnitsky and Rao, 1997; Matsuura and Okabe, 1995). When $p = 1$ the problem is still convex however the correct solution is only guaranteed under certain conditions that are usually not met because of the highly correlated columns of the EEG forward model (Donoho et al., 2006). Furthermore, studies have shown that the $L_1$-norm produces spurious sources (Liu et al.,

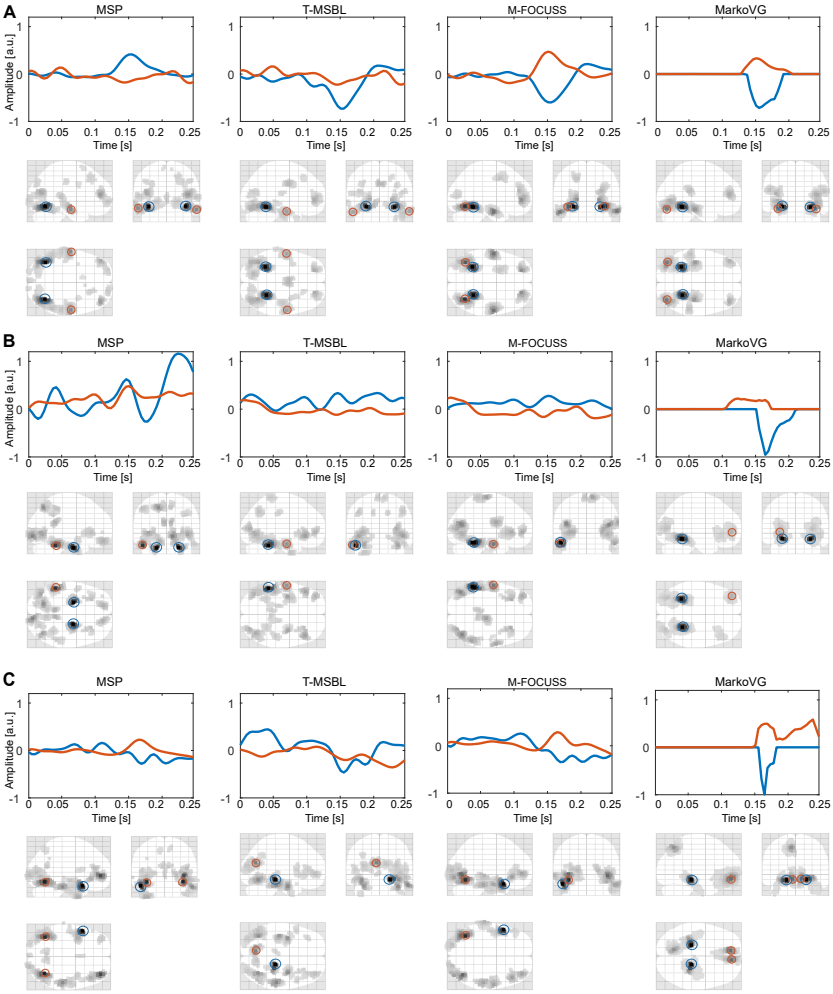



Figure 6: Reconstruction of EEG face perception data for subject "A", "B" and "C". The two strongest basis functions' temporal dynamics (top) and their locations (bottom) for MSP, T-MSBL, M-FOCUSS and MarkoVG. The highest magnitude source is shown in blue and second highest in red. The highest magnitude basis functions were for several of the examples bilateral. The glass brains show the activity for the 512 maximum magnitude dipoles at the time sample with highest magnitude source, here at 151-170 ms after stimuli onset. Source strengths are directly comparable between sources, subjects and solvers but are in arbitrary unit due to lack of units of the forward model Litvak (2016)

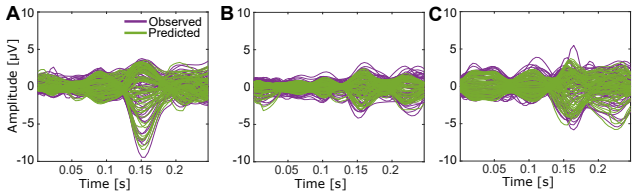



Figure 7: Averaged ERP as observed and as predicted by MarkoVG. Shown are all electrodes for the time interval 0 to 250 ms after stimuli onset for subjects "A", "B" and "C".

2004; Hansen et al., 2013c). Employing $L_p$-norms where $p < 1$ will produce more sparse solutions, it however also implies non-convexity.

Bayesian approximations such as SBL (Tipping, 2001) should produce fewer local minima (Zhang and Rao, 2011) and are therefore also very promising. ARD is in SBL used to prune away variables by assigning a hyperparameter to each variable dictating whether to keep or discard the variable. Extending SBL to the MMV framework one hyperparameter controls all the time samples for each variable (Wipf and Rao, 2007). As more samples are available in determining whether a variable is relevant in the MMV model an improved solution is obtained (Wipf and Rao, 2007; Zhang and Rao, 2011). However, the assumption of common sparsity profile across time does not always hold physiologically.

More flexible ways of handling and exploiting temporal coherency have been proposed (Montoya-Martinez et al., 2012; Gramfort et al., 2013), wherein signals are modeled as being non-stationary. Structured sparsity profiles are achieved by Gramfort et al. by time-frequency analysis and modeling each active source as a summation of Gabor atoms (Gramfort et al., 2013) . Montoya-Martines et al. avoid synthesizing a dictionary containing temporal patterns by using the sparse group LASSO regularizer (Montoya-Martinez et al., 2012). While these two methods both base their solution on regularization through the $L_1/L_2$-norm, we propose to obtain sparser solutions through the VG (Kappen, 2011) which approximates the $L_0$-norm regularizer. Furthermore the VG has the favorable trait of estimating both the *state of activation* (active/non-active) and the *activation strength* of the active sources. This allows for modeling the temporal dynamics in EEG as having smooth temporal support, while allowing for more rapid changes in the dipole strength, under the assumption that the location of activation varies slower than the activation strength. By applying a Markovian prior on the support, the level of smoothness in the temporal sparsity profile is adapted to the observed EEG data.

We extended MarkoVG by incorporating spatial basis functions inspired by earlier implementations (Friston et al., 2008). Spatial smooth compact source patches are motivated by knowledge of the EEG generators' spatial extension, which is estimated to be at least $5 \times 5$ mm$^2$ (Baillet et al., 2001) and often extending several centimeters (Michel, 2009). Incorporating spatial basis functions has the additional benefit of reducing the computational complexity when there are fewer basis functions than original sources. There is however a risk that the center of a "true" source is a source with low activity in the basis functions. This is a potential bias that is incurred to counter the uncertainty and ill-posedness of the EEG inverse problem. These assumptions are argued to impose limited bias as in (Friston et al., 2008). Another related possible concern is the

use of fixed orientations and simplified forward models in general. Several studies argue for improving source reconstruction by using as detailed and accurate forward models as possible (Akalin Acar and Makeig, 2013; Windhoff et al., 2013), and we note that it is indeed possible to combine MarkoVG with any type of forward model. Incorporation of both flexible dipole orientations and spatial coherency could, for example, be achieved by the so-called 'sparse basis fields' introduced by Haufe et al. (Haufe et al., 2011). Finally, increasing the number of spatial basis functions should also be investigated in future work.

In summary we applied the physiological meaningful assumptions that the underlying EEG generators are spatially smooth and sparse, and temporally variable smooth/sparse. The effectiveness of implemented assumptions can in general be validated through simulation studies. We therefore tested the performance of MarkoVG in a controlled setting where we also compared the proposed algorithm to three other solvers. On the synthetic data we found that MarkoVG was better at identifying the correct active sources and time samples. MarkoVG was thus more effective in recovering the sparsity level, both spatially and temporally. Importantly, we showed that even when favoring the T-MSBL and especially the M-FOCUSS algorithm superior performance to MarkoVG was not accomplished.

In real data hypothesized assumptions can be validated using other imaging techniques such as fMRI or through lesion studies which can provide information about the location of specific information processing in the brain. The estimated temporal dynamics can be validated through single cell recordings that can indicate when specific brain areas are involved in an EEG response. We tested our algorithm on the EEG response to seeing faces as compared to scrambled faces. In this paradigm we know from fMRI studies (Henson et al., 2003) and combined EEG/MEG studies (Henson et al., 2009) that the activated areas include the FFA and OFA. Studies of patients with lesions in the FFA and OFA further validate the importance of these areas in face perception (Eimer and McCarthy, 1999; Dalrymple et al., 2011). The face/scrambled face contrasted fMRI recordings included in the multimodal study we extracted EEG from, have been analyzed at the group level in Fig. 3b in (Wakeman and Henson, 2015). The O/FFAs were also here dominating the face response as compared to the scrambled face condition. Furthermore, the face-sensitive response was sparse and largely symmetric across hemispheres.

The MSP, T-MSBL, M-FOCUSS and MarkoVG confirmed the existence of face-sensitive activity in the FFA and OFA. Frontal activation was also recovered by all algorithms, partly agreeing with the before mentioned fMRI study which also contained frontal activation. The temporal dynamics recovered by the inverse solvers in the most strongly activated sources showed focused activity around the N170 component. This was especially true for MarkoVG. The strongest sources as estimated by MSP, T-MSBL and M-FOCUSS had activity in the entire time window including peaks around 50 ms. A study of the single cell recordings from the inferior temporal cortex of the macaque brain response to faces revealed predictive power in the response after approximately 100 ms (Kiani et al., 2005), thus indicating the relevant face response starts well after 50 ms. Our study thus indicates that this evidence can be transferred to humans. We also note that an ERP study on humans has shown that significant differences between faces and noise textures begins 130 ms after stimuli onset (Rousselet et al., 2008).

The reasoning behind promoting zero activation is a model of focal brain activation, i.e., specific brain areas become active as a response to given stimuli and then return to their baseline level. This is for example appropriate when reconstructing contrast EEG

responses. Furthermore the ill-posedness of the inverse problem and the poor signal-to-noise levels of EEG in general obstruct accurate recovery of dense source activations. By sparsity promoting priors we focus on activity in time samples having sufficient evidence.

For reference we provide the computation time required by each inverse solver per iteration. Computed on a laptop with 2.1-GHz 64-bit i7 processor the time spend per iteration is 225 ms for MSP, 3 ms for T-MSBL, 0.6 ms for M-FOCUSS and 100 ms for MarkoVG. We note that MSP in general requires least iterations per inverse problem, and since MarkoVG performs cross-validation to estimate the sparsity level it is slowest among the tested algorithms. However, active set based optimization as implemented in TMSBL and M-FOCUSS (in which inactive variables are pruned), could dramatically reduce the computation time. This is a current topic of research. Furthermore, we note that the goal of MarkoVG is to explore the implementation of meaningful physiologically priors in order to solve the severely ill-posed inverse EEG problem rather than being fast.

In conclusion we have introduced temporal smoothness in the support of the brain dynamics within the so-called MarkoVG framework, and demonstrated how it can adapt to the degree of temporal coherency and spatial sparsity underlying the recorded EEG signal. In simulations and in real data MarkoVG showed promise as a tool for EEG dynamic imaging. Further improvements on the algorithm involves increasing the flexibility of the model by optimizing two free parameters instead of working with a fixed relation between them as here. Such more complex optimization of parameters could be accomplished by Bayesian optimization methods such as proposed by (Snoek et al., 2012).

**Acknowledgements**

We thank D. Wakeman and R. Henson for making their multimodal dataset freely available. We also thank the anonymous reviewers for their constructive comments. The work was supported in part by the Novo Nordisk Foundation Interdisciplinary Synergy Program 2014 ["Biophysically adjusted state-informed cortex stimulation (BASICS)"] (STH) and the Danish Lundbeck Foundation via the Center for Integrated Molecular Brain Imaging and by the Innovation Fund Denmark project, "Neuro-technology for 24/7 mental state monitoring" (LKH).

**References**

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

Akalin Acar, Z., Makeig, S., 2013. Effects of forward model errors on EEG source localization. Brain topography 26, 378–396. doi:10.1007/s10548-012-0274-6.

Andersen, M.R., Winther, O., Hansen, L.K., 2014. Bayesian inference for structured spike and slab priors, in: Advances in Neural Information Processing Systems, pp. 1745–1753.

Ashburner, J., Barnes, G., Chen, C.C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., et al., 2014. SPM12 Manual .

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. Signal Processing Magazine, IEEE 18, 14–30.

Cotter, S.F., Rao, B.D., Engan, K., Kreutz-delgado, K., Member, S., 2005. Sparse Solutions to Linear Inverse Problems With Multiple Measurement Vectors. IEEE Transactions on Signal Processing 53, 2477–2488.

Dalrymple, K.A., Oruc, I., Duchaine, B., Pancaroglu, R., Fox, C.J., Iaria, G., Handy, T.C., Barton, J.J., 2011. The anatomic basis of the right face-selective N170 in acquired prosopagnosia: a combined ERP/fMRI study. Neuropsychologia 49, 2553–2563.

De Ciantis, A., Lemieux, L., 2013. Localisation of epileptic foci using novel imaging modalities. Current opinion in neurology 26, 368.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG sources are dipolar. PloS one 7, e30135.

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

Eimer, M., McCarthy, R.A., 1999. Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. Neuroreport 10, 255–259.

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

Gorodnitsky, I., George, J., Rao, B., 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalography and clinical Neurophysiology 95, 231–251. URL: http://www.sciencedirect.com/science/article/pii/001346949500107A.

Gorodnitsky, I.F., Rao, B.D., 1997. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. Signal Processing, IEEE Transactions on 45, 600–616.

Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. Physics in medicine and biology 57, 1937–1961. doi:10.1088/0031-9155/57/7/1937.

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M.S., Kowalski, M., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. NeuroImage 70, 410–422. doi:10.1016/j.neuroimage.2012.12.051.

Hämäläinen, M.S., Ilmoniemi, R., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Medical & biological engineering & computing 32, 35–42.

Hansen, L.K., Hansen, S.T., Stahlhut, C., 2013a. Mobile real-time EEG imaging Bayesian inference with sparse, temporally smooth source priors, in: 2013 International Winter Workshop on Brain-Computer Interface (BCI), IEEE. pp. 6–7. doi:10.1109/IWW-BCI.2013.6506608.

Hansen, L.K., Rasmussen, C.E., 1994. Pruning from adaptive regularization. Neural Computation 6, 1223–1232. URL: http://www.mitpressjournals.org/doi/abs/10.1162/neco.1994.6.6.1223.

Hansen, S.T., Hansen, L.K., 2013. EEG Sequence Imaging: A Markov Prior for the Variational Garrote, in: Unpublished note for the International Workshop, Machine Learning and Interpretation in Neuroimaging, available at http://orbit.dtu.dk/fedora/objects/orbit:127330/datastreams/file_61f34d92-2e60-4871-8a41-08f1d69f5c47/content.

Hansen, S.T., Hansen, L.K., 2014. EEG source reconstruction using sparse basis function representations, in: 2014 International Workshop on Pattern Recognition in Neuroimaging, IEEE. pp. 1–4. doi:10.1109/PRNI.2014.6858521.

Hansen, S.T., Hauberg, S., Hansen, L.K., 2016. Data-driven forward model inference for eeg brain imaging. NeuroImage 139, 249–258.

Hansen, S.T., Stahlhut, C., Hansen, L.K., 2013b. Expansion of the Variational Garrote to a Multiple Measurement Vectors Model, in: Scandinavian Conference on AI 2013 (2013), SCAI.

Hansen, S.T., Stahlhut, C., Hansen, L.K., 2013c. Sparse Source EEG Imaging with the Variational Garrote, in: Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on, IEEE. pp. 106–109.

Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.R., Nolte, G., 2008. Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage 42, 726–38. doi:10.1016/j.neuroimage.2008.04.246.

Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., Müller, K.R., 2011. Large-scale EEG/MEG source localization with spatial flexibility. NeuroImage 54, 851–859. doi:10.1016/j.neuroimage.2010.09.003.

Henson, R., Goshen-Gottstein, Y., Ganel, T., Otten, L., Quayle, A., Rugg, M., 2003. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cerebral cortex 13, 793–805.

Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Human brain mapping 31, 1512–31. doi:10.1002/hbm.20956.

Henson, R.N., Mouchlianitis, E., Friston, K.J., 2009. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. NeuroImage 47, 581–9. doi:10.1016/j.neuroimage.2009.04.063.

Hulbert, S., Adeli, H., 2013. EEG/MEG-and imaging-based diagnosis of Alzheimers disease. Reviews in the neurosciences 24, 563–576.

Ishwaran, H., Rao, J.S., 2005. Spike and slab variable selection: frequentist and Bayesian strategies. Annals of Statistics , 730–773.

Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cerebral cortex 14, 132–142.

Kappen, H., 2011. The Variational Garrote. arXiv preprint arXiv:1109.0486 arXiv:arXiv:1109.0486v2.

Kappen, H.J., Gómez, V., 2013. The Variational Garrote. Machine Learning , 1–26doi:10.1007/s10994-013-5427-7.

Kiani, R., Esteky, H., Tanaka, K., 2005. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. Journal of neurophysiology 94, 1587–1596. doi:10.1152/jn.00540.2004.

Litvak, V., 2016. JISCMail - SPM Archives. URL: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1611{&}L=SPM{&}P=R372{&}I=-3{&}X=A3C1321019C707A995{&}Y=sofha{%}40dtu.dk{&}d=No+Match{%}3BMatch{%}3BMatches.

Liu, H., Gao, X., Schimpf, P.H., Yang, F., Gao, S., 2004. A recursive algorithm for the three-dimensional imaging of brain electric activity: shrinking LORETA-FOCUSS. Biomedical Engineering, IEEE Transactions on 51, 1794–1802.

Lucka, F., Pursiainen, S., Burger, M., Wolters, C.H., 2012. Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: depth localization and source separation for focal primary currents. NeuroImage 61, 1364–82. doi:10.1016/j.neuroimage.2012.04.017.

MacKay, D.J.C., 1995. Probable networks and plausible predictions a review of practical Bayesian methods for supervised neural networks. Network: Computation in Neural Systems 6, 469–505. doi:10.1088/0954-898X/6/3/011.

Makhoul, J., Kubala, F., Schwartz, R., Weischedel, R., 1999. Performance measures for information extraction, in: Proceedings of DARPA Broadcast News Workshop. Morgan Kaufmann Pub, pp. 249–252.

Markov, N., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M., Giroud, P., Ercsey-Ravasz, M., Pilaz, L., et al., 2011. Weight consistency specifies regularities of macaque cortical networks. Cerebral Cortex 21, 1254–1272.

Matsuura, K., Okabe, Y., 1995. Selective minimum-norm solution of the biomagnetic inverse problem. IEEE transactions on bio-medical engineering 42, 608–15. doi:10.1109/10.387200.

Michel, C.M., 2009. Electrical neuroimaging. Cambridge University Press.

Montoya-Martinez, J., Artes-Rodriguez, A., Hansen, L.K., Pontil, M., 2012. Structured sparsity regularization approach to the EEG inverse problem, in: Cognitive Information Processing (CIP), 2012 3rd International Workshop on, IEEE. pp. 1–6.

Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein, R.B., Cadusch, P.J., 1997. EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalography and clinical neurophysiology 103, 499–515.

Ou, W., Hämäläinen, M.S., Golland, P., 2009. A distributed spatio-temporal EEG/MEG inverse solver. NeuroImage 44, 932–46. doi:10.1016/j.neuroimage.2008.05.063.

Pascual-Marqui, R., Esslen, M., Kochi, K., Lehmann, D., et al., 2002. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. Methods and findings in experimental and clinical pharmacology 24, 91–95.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of psychophysiology 18, 49–65.

Phillips, C., 2000. Source estimation in EEG. PhD thesis. URL: http://orbi.ulg.ac.be/bitstream/2268/156112/1/Phillips_localis.pdf.

Phillips, C., Rugg, M.D., Friston, K.J., 2002. Anatomically Informed Basis Functions for EEG Source Localization: Combining Functional and Anatomical Constraints. NeuroImage 16, 678–695. doi:10.1006/nimg.2002.1143.

Rijsbergen, C.J.V., 1979. Information Retrieval. 2nd ed., Butterworth-Heinemann. URL: http://dl.acm.org/citation.cfm?id=539927.

Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2008. Time course and robustness of ERP object and face differences. Journal of vision 8, 3.1–18. doi:10.1167/8.12.3.

Scherg, M., Von Cramon, D., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 62, 32–44. doi:10.1016/0168-5597(85)90033-4.

Schüz, A., Braitenberg, V., 2002. The human cortical white matter: Quantitative aspects of cortico-cortical long-range connectivity. Cortical Areas: Unity and Diversity , 377–384doi:10.1201/9780203299296.ch16.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, pp. 2951–2959.

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

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. The Journal of Machine Learning Research 1, 211–244.

Vega-Hernández, M., Martínez-Montes, E., Sanchez-Bornot, J.M., Lage-Castellanos, A., Valdés-Sosa, P.A., 2008. Penalized least squares methods for solving the EEG inverse problem. Statistica Sinica 18, 1535–1551.

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.

Windhoff, M., Opitz, A., Thielscher, A., 2013. 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. doi:10.1002/hbm.21479.

Wipf, D.P., Rao, B.D., 2007. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. Signal Processing, IEEE Transactions on, 55, 3704–3716.

Zhang, Z., Rao, B.D., 2011. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. Selected Topics in Signal Processing, IEEE Journal of 5, 912–926.

# Supplementary Information for: *Spatio-temporal reconstruction of brain dynamics from EEG with a Markov prior*

Sofie Therese Hansen[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*

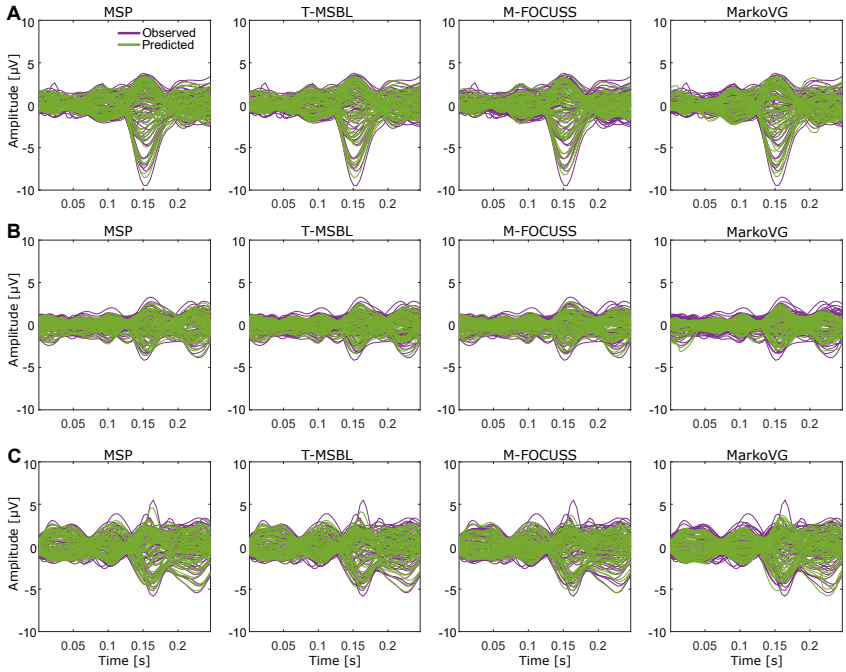



Figure 1: Observed and predicted ERP for each channel for subject "A", "B" and "C".

*Email addresses:* `sofha@dtu.dk` (Sofie Therese Hansen), `lkai@dtu.dk` (Lars Kai Hansen)

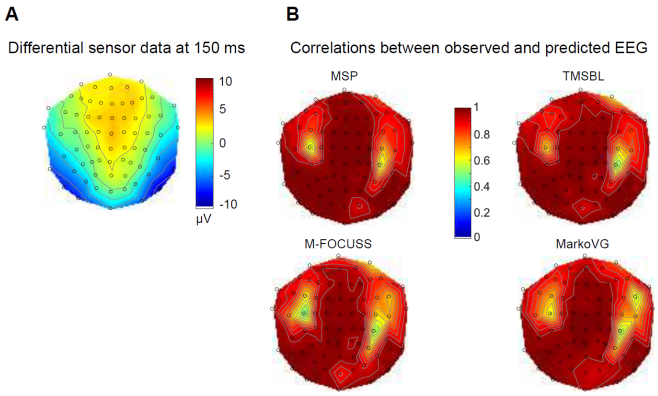



Figure 2: Scalp maps belonging to subject "A" (as included in the main paper). A) The face/scrambled face differential ERP 150 ms after stimuli onset. B) Correlations across time (0-250 ms) between the observed EEG and as predicted by the inverse solvers.

Appendix H

# Data Driven Forward Model Inference for EEG Brain Imaging

Postprint of article published in NeuroImage.

Hansen, S.T., Hauberg, S. & Hansen, L.K. (2016). "Data Driven Forward Model Inference for EEG Brain Imaging". *NeuroImage*, vol 139, pp. 249-258.
doi: http://dx.doi.org/10.1016/j.neuroimage.2016.06.017

# 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

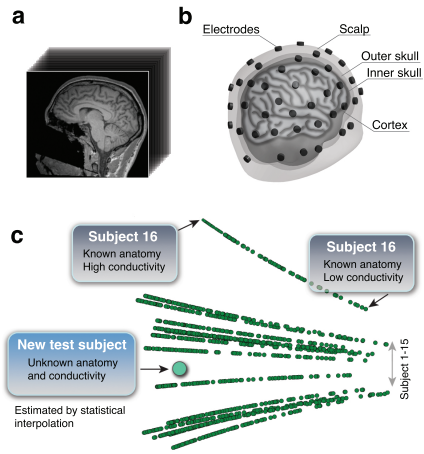



Figure 1: The process of creating forward models and their projection to PCA space. (**a**) For each of the 16 subjects, a T1-weighted image is used to construct a forward model. (**b**) The forward model is here constructed using a three-layered BEM head model (scalp-skull-brain). For each subject, 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.

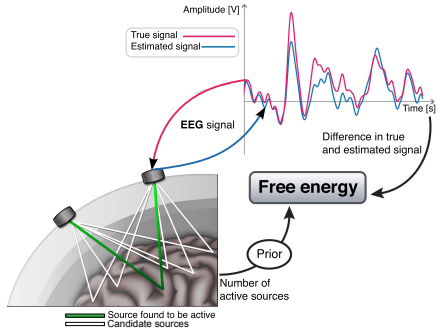



Figure 2: The free energy summarizing the ability of a model to describe the data and its complexity. We calculate the free energy based on the inference scheme proposed in the Variational Garrote (VG). Formally, the free energy in VG expresses the Bayesian combination of data fit (difference in true 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 $70 \times 8196$ 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

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 $\gamma$. 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 $\gamma$ changes. In order not to depend on a particular choice of the bandwidth, we consider a uniform prior on this parameter,

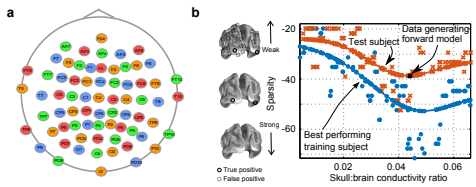



Figure 3: Estimation of optimal sparsity. (a) Partitioning of the 70 EEG electrodes into four folds. Each fold is represented by one color. (b) The sparsity levels obtained by cross-validation using the mean squared error (MSE) between the true and predicted signal. For visualization purposes, we only 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) |
| $\kappa$ | 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 ($\kappa$) 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.

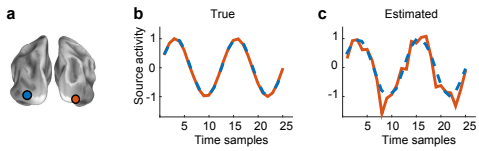



Figure 4: Source distributions of the planted activity and as estimated using the free energy-optimal forward model. (a) Posterior view of the inflated brain showing the locations of the two planted sources. The locations of the estimated sources were identical to these. (b) The real and (c) estimated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

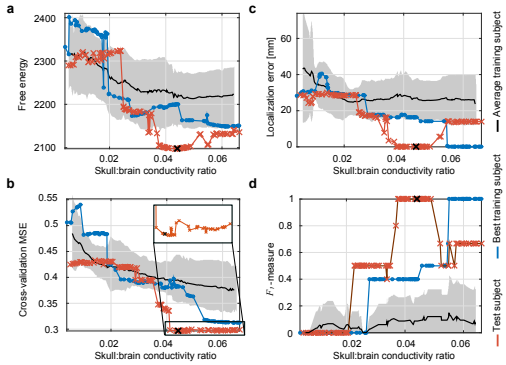



Figure 5: Forward model prediction among the database test and training subjects for simulated data. The source signal in Fig. 4a-b was projected to sensor space with the forward model indicated by the black cross. The smoothed sparsity levels in Fig. 3b were applied to the Variational Garrote in combination 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

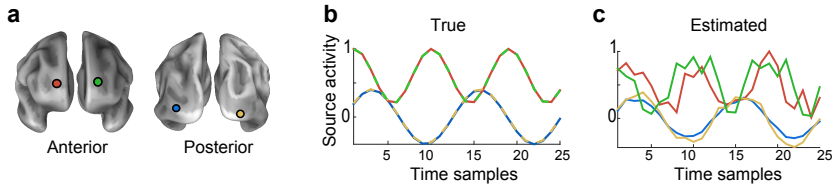



Figure 6: Source distributions of the planted activity and as estimated using the free energy-optimal forward model. (a) Posterior view of the inflated brain showing the locations of the four planted sources. The locations of the estimated sources were identical to these. (b) The real and (c) estimated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

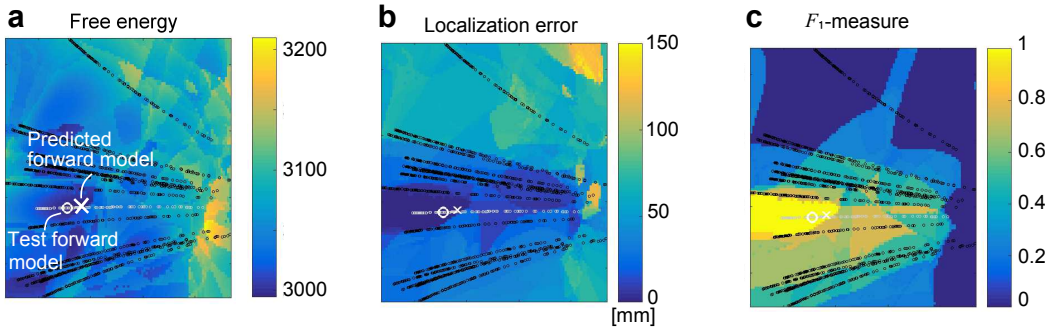



Figure 7: Search for the optimal forward model in the 2D PCA space created by the training subjects; simulated EEG data (Fig. 6). A test forward model (white circle) from the withheld test subject was used to generate the test EEG data. The 2D PCA space of forward models wherein (a) the free energy, (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

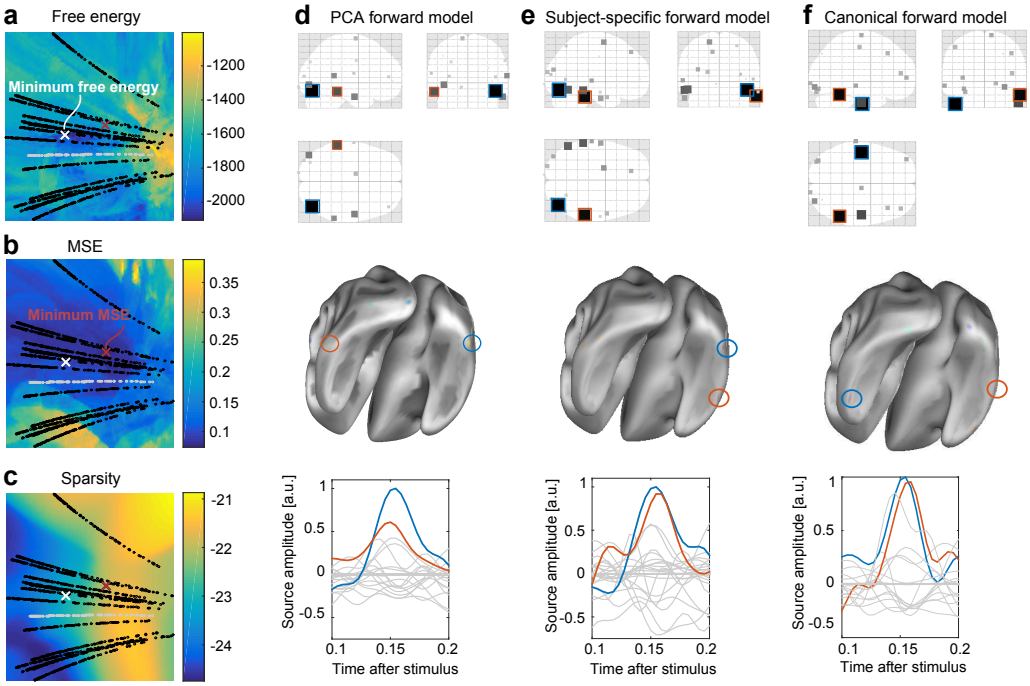



Figure 8: Search for the optimal forward model; real EEG data recorded from the test subject. (**a-c**) The 2D PCA space created by 15 training subjects, whose projections are seen in black. The test subject is seen in gray. The minimum free energy is indicated by a white cross and minimum MSE by a red 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 ($\sigma$). 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 $\sigma$ | Subject head, template $\sigma$ | Subject head, true $\sigma$ |
|---|---|---|---|---|
| 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-

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 $\epsilon$ 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 $\gamma$ 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

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

Here, $\chi$ 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 $\gamma$ 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`.

# Supplementary Material

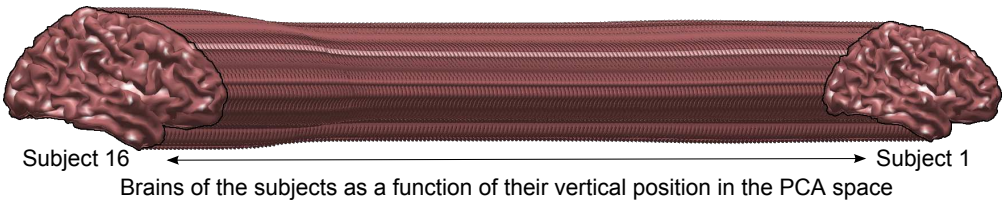
**Data-driven forward model inference for EEG brain imaging**

Sofie Therese Hansen, Søren Hauberg and Lars Kai Hansen

June 8, 2016

## Forward Model Analysis

### Locations of subjects in the PCA space




**Supplementary Figure 1:** The segmented brains of the 16 subjects, smoothed/interpolated as a function of their locations in the vertical dimension (from top to bottom) of the PCA space in Fig. 1c (main paper). The left-most brain thus belongs to subject 16.

### Performance of the PCA projection

We have in the main text used two principal components (PCs) for the projection of the forward models. We now investigate the influence the number of PCs has on the PCA forward models' similarity with the real forward models (Supplementary Fig. 2) as well as their ability to reconstruct sources (Supplementary Fig. 3).

We use the following expression to estimate a Scaled Difference (ScD)

$$\text{ScD} = \frac{||\mathbf{A}_{\text{PCA}} - \mathbf{A}_{\text{true}}||_2^2}{||\mathbf{A}_{\text{true}} - \mu||_2^2} \tag{1}$$

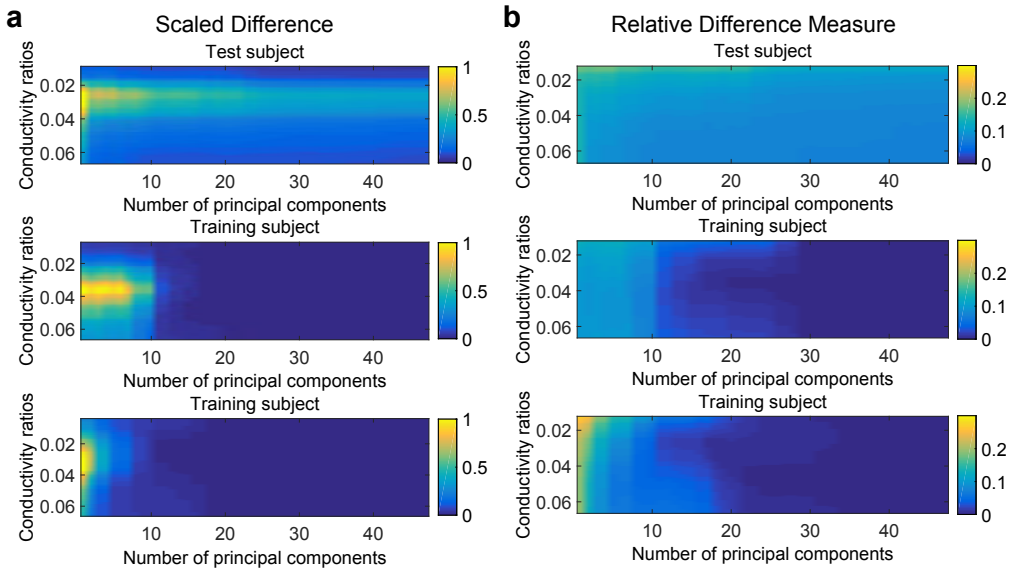
between a "true" forward model, $\mathbf{A}_{\text{true}}$, and its corresponding projection from the PCA space, $\mathbf{A}_{\text{PCA}}$. Here $\mu$ is the average of the training forward models. We also compute the Relative Difference Measure (RDM) [1] given by

$$\text{RDM} = \left|\left| \frac{\mathbf{A}_{\text{PCA}}}{||\mathbf{A}_{\text{PCA}}||_2} - \frac{\mathbf{A}_{\text{true}}}{||\mathbf{A}_{\text{true}}||_2} \right|\right|_2 . \tag{2}$$

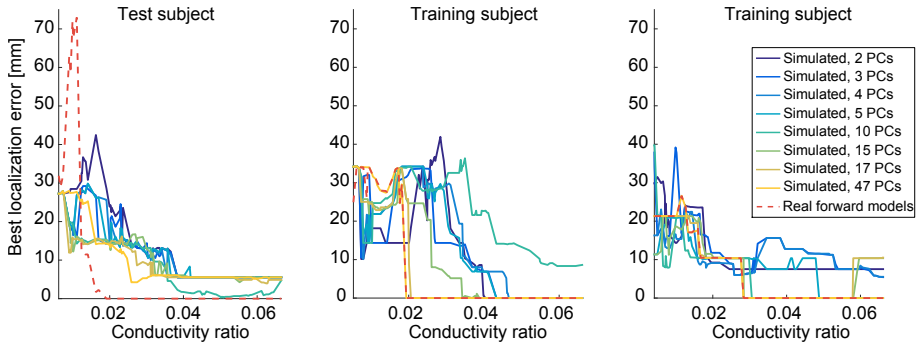
Supplementary Figure 2a and b show these two measures for a test subject and two training subjects as a function of the number of PCs used to create $\mathbf{A}_{\text{PCA}}$. The maximum number of PCs is here 47 which corresponds to the number of PCs needed to describe $(1-\epsilon)100\%$[1] of the variance in the training data. The ScD (Supplementary Fig. 2a) converges after applying about ten PCs for the training subjects while the ScD for the test subject decreases in the first 20 PCs but never reaches the same low level as the training subjects. Note that the high ScD for medium conductivities (especially for few PCs) is an artifact of the calculation, since when $\mathbf{A}_{\text{PCA}} \to \mu$ then $\frac{||\mathbf{A}_{\text{PCA}} - \mathbf{A}_{\text{true}}||_2^2}{||\mathbf{A}_{\text{true}} - \mu||_2^2} \to \frac{||\mu - \mathbf{A}_{\text{true}}||_2^2}{||\mathbf{A}_{\text{true}} - \mu||_2^2} = 1$. The RDM converges for the training subjects after 10-15 PCs while again the test subject does not attain zero difference. However the RDM of the training subject in the lower panel of Supplementary Figure 2b is for the first few PCs higher than the RDM of the test subject, but this relation quickly reverses.

Supplementary Figure 3 shows the performance of the PCA forward models as reflected by the best obtainable localization error across applied regularizations. For this experiment we have planted two posteriorly located sources with sinusoidal activity across 25 time samples. The source distribution is projected to the electrodes through a real forward model with a specific skull:brain conductivity ratio and finally noise is added. A forward model is extracted from the PCA space in the

[1] $\epsilon = 2.2204 \cdot 10^{-16}$ as defined in MATLAB, MathWorks Inc.




**Supplementary Figure 2:** Difference between true and simulated forward models as function of principal components used for the projection. (**a**) Scaled Difference and (**b**) Relative Difference Measure. Fifteen training subjects were used to create the PCA projection, here results from two are shown along with the test subject.




**Supplementary Figure 3:** Performance of source reconstruction using the PCA generated forward models of a test subject and two training subjects. In total fifteen subjects were used to build the PCA forward model space. Different numbers of applied principal components for the projection are investigated. The first 17 and 47 components contain respectively 99% and $(1-\epsilon)100\%$ of the variance in the training data. For comparison the reconstruction performance of the real forward model (used to generate the simulated data) is also shown. Here two posterior sources were planted and SNR= 5 dB. The localization error is a sum of the distances between the two planted and estimated sources.

location corresponding to where the aforementioned real forward model would project to. Using the PCA forward model source reconstruction is performed. This procedure is repeated for all 100 levels of conductivity ratios. In Supplementary Figure 3 we thus compare the performance of the PCA reconstructed forward models for subjects included in creating the PCA space and for a withheld test subject as a function of the conductivity ratio. We additionally investigate the influence the numbers of PCs used in generating the simulated forward models has on the performance.

We conclude from these initial experiments that a few PCs give a reasonable representation of a test forward model. The performance of the PCA forward models for the shown test subject do not show a clear tendency that applying more PCs gives better reconstruction, e.g. 10 PCs are better than 15, 17 and 47 PCs for a wide range of skull:brain conductivity ratios (Supplementary Fig. 3). However if more than two sources were planted more PCs might be need.

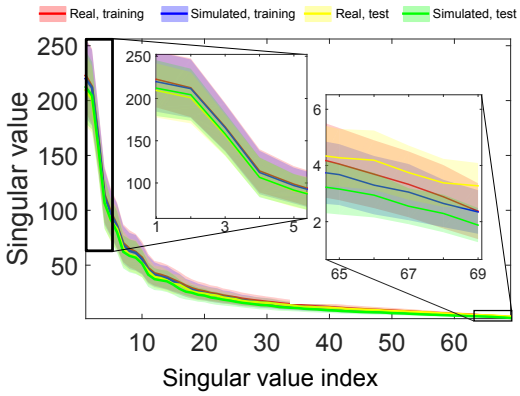
In Supplementary Fig. 3 it is seen that forward models with very low conductivity ratios has poor performance even when using the correct forward model. The investigated span of ratios was defined by literature values in [2]. The lower ratio of 1:250 is very low compared to experimental findings. We have seen in other studies that the matrix coherence of the forward models is highest for lower conductivity ratios [3] thus increasing the ill-posedness of the inverse problem. Studies have also

shown worse localization performance when the conductivity ratio is low [3, 4]

Note that the PCA forward model from the direct projection of the true data generating forward model does not necessarily correspond to the optimal model due to loss of information in the projection. Thus other PCA forward models in a relevant search space could be able to yield a better source reconstruction.

### Singular Value Decomposition

The simulated forward models from the 2D PCA projection are compared to the real forward models through their singular values. Singular value decomposition is performed on each of the forward models in the four groups: 1) real, used for training, 2) simulated, used for training, 3) real, used for test and 4) simulated, used for test. Other than showing the difference between the real and simulated forward models the analysis will also reveal any differences in the test and training sets. The mean and standard deviation of the mean of the four groups can be seen in Supplementary Figure 4. The singular

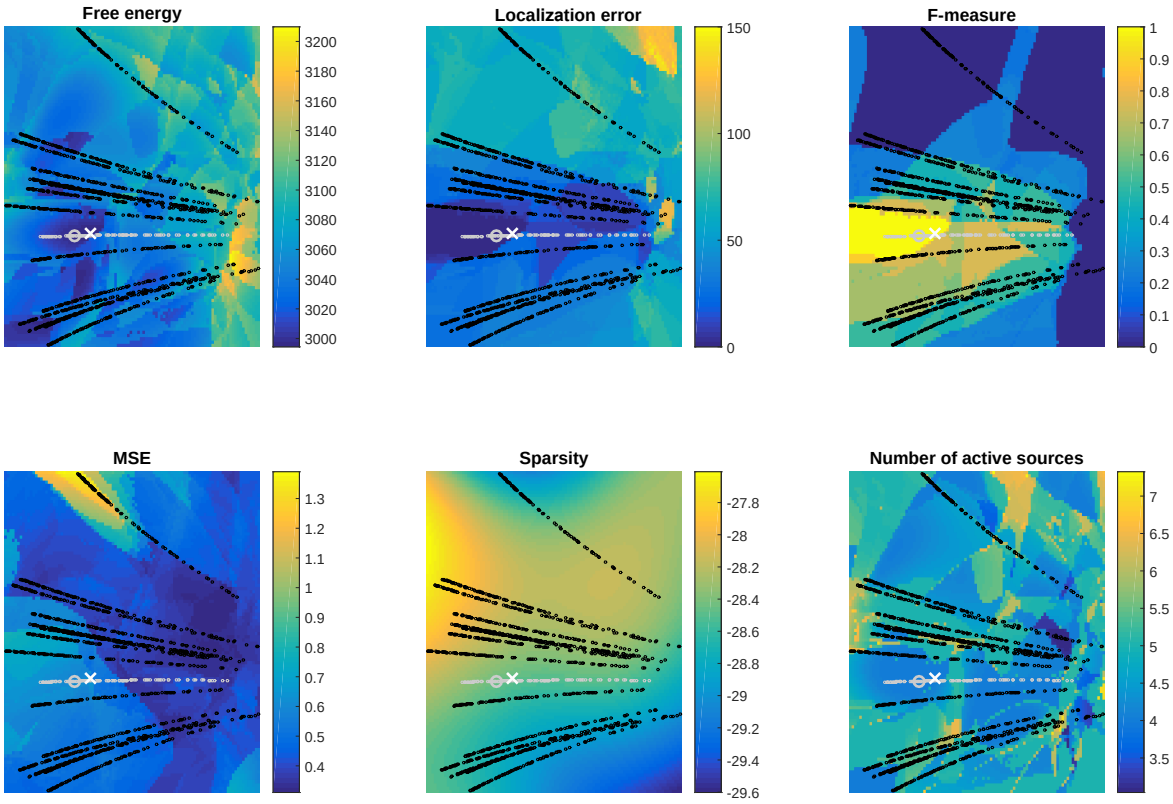



**Supplementary Figure 4:** Singular values of simulated (PCA) and real forward models. Comparisons is done after conversion to single precision. The ranks of the forward models are 69.

values are very similar both between training and test and between real and simulated forward models.

## Performance of the 2D PCA Generated Forward Models

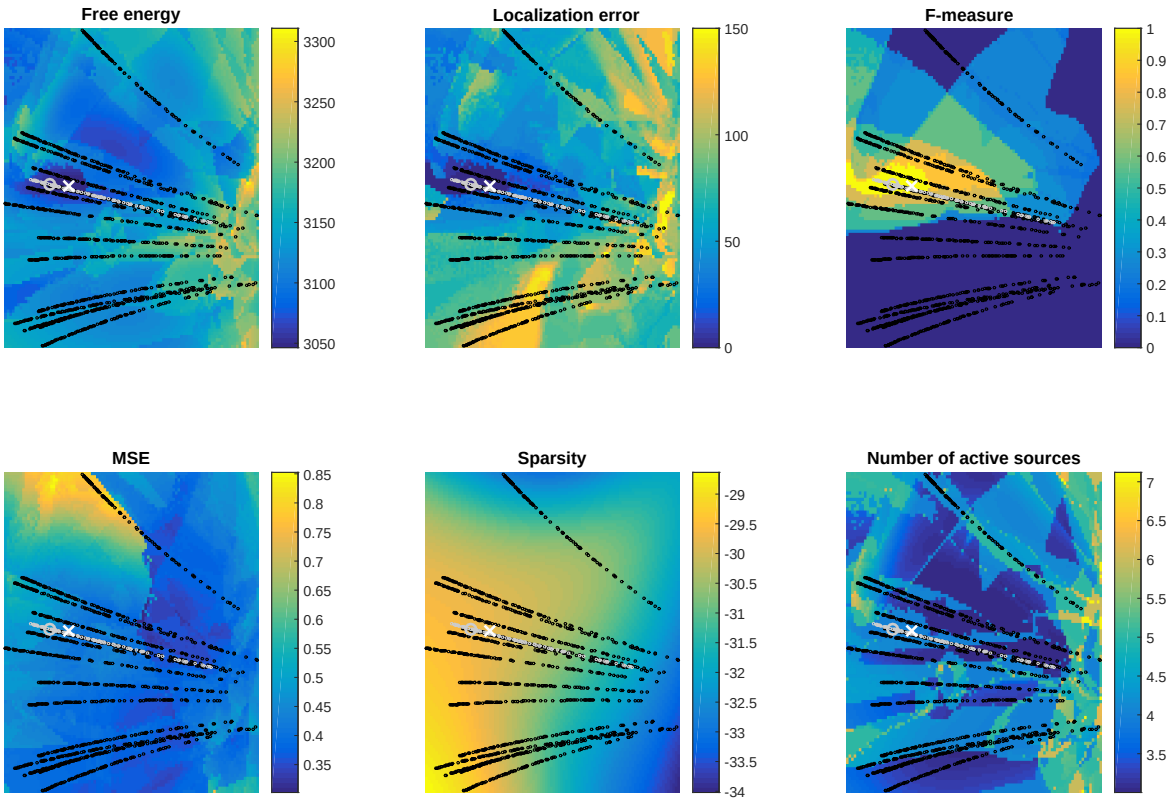
The simulation study in Figure 7 and 8 is here replicated for several subjects. As mentioned we use the free energy to recover the optimal forward model in a constructed 2D PCA space. Below we show this for six subjects on synthetic data (Supplementary Fig. 5-10). The test subject of Supplementary Figure 5 is the same as the one used in the paper. The procedure is as follows: 1) A 2D PCA space is constructed from the 15 training subjects. 2) A real forward model from the remaining test subject is used to generate the synthetic EEG data together with a source density of four active sources; two in each hemisphere; one posterior and one anterior. The activity of the sources are sinusoidal and bilaterally synchronized. 3) The sparsity level is estimated through four-fold cross-validation on the electrodes for 250 randomly chosen PCA training forward models. The sparsity levels are then intra/extrapolated and smoothed to cover the search space. 4) A $100 \times 100$ grid containing all training subjects' projected forward models is explored. 5) The PCA forward model with lowest free energy is reported and used for source reconstruction. In Supplementary Figure 11 real EEG is used instead of the synthetic (a description of preprocessing of the real EEG data can be found in the main paper).




**Supplementary Figure 5:** Test subject 2, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 2 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see Methods). Finally we show the estimated number of sources.

**Supplementary Table 1:** Test subject 2: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 5) and template/subject specific forward models. 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**.

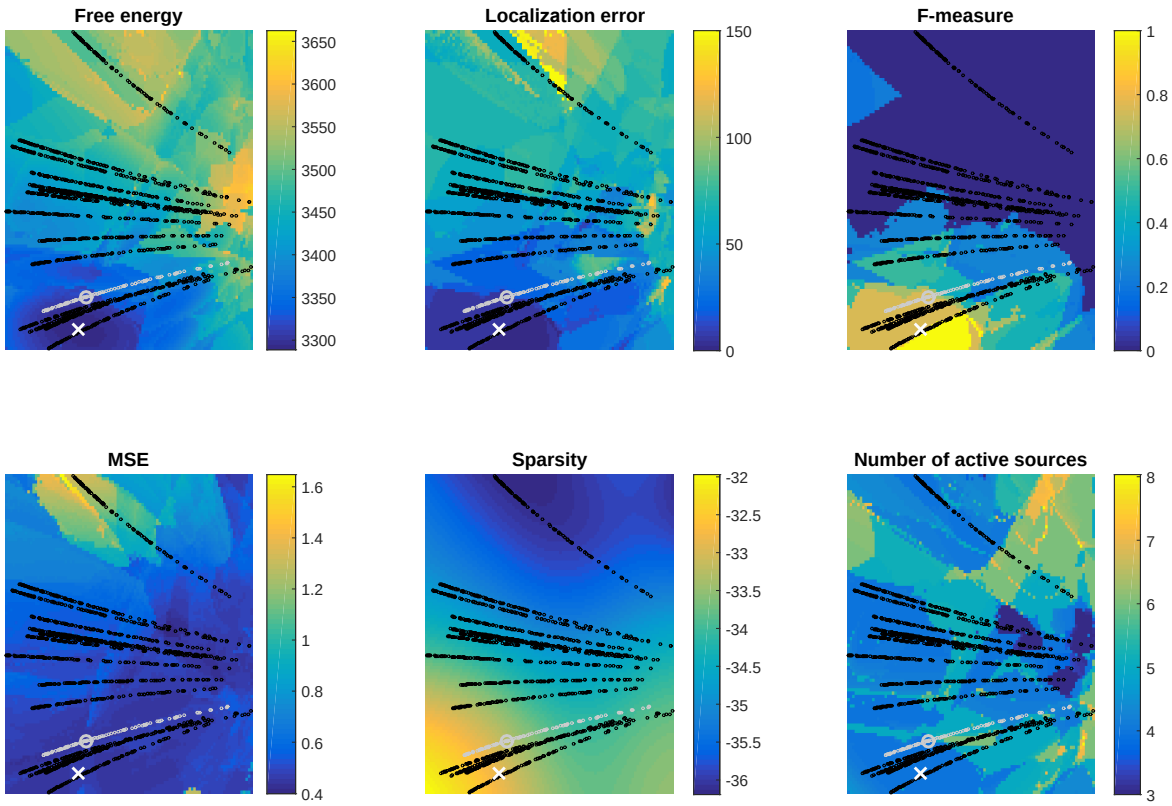
| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 2994 | 3192 | 3057 | 2956 |
| MSE | 0.63 | 0.55 | 0.93 | 0.61 |
| F-measure | 1 | 0 | 0.44 | 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 |




**Supplementary Figure 6:** Test subject 7, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 7 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see **Methods**). Finally we show the estimated number of sources.

**Supplementary Table 2:** Test subject 7: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 6) and template/subject specific forward models. Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior.

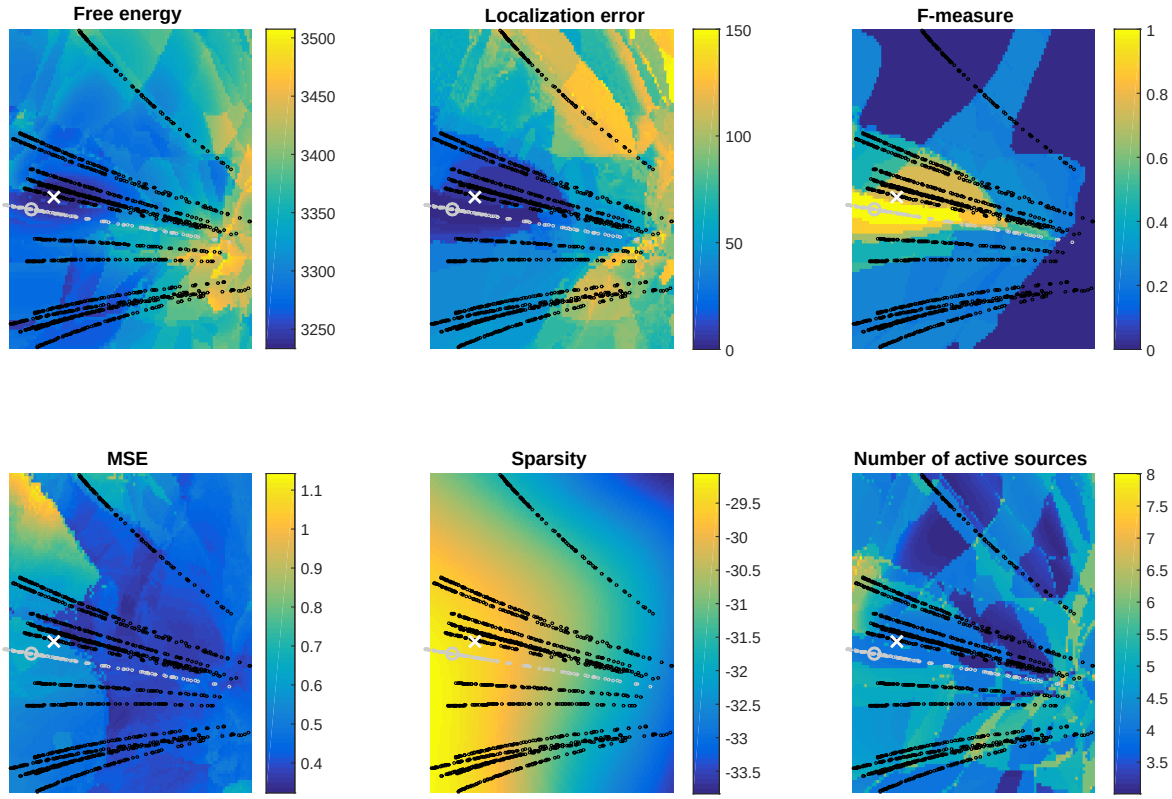
| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 3046 | 3192 | 3132 | 3027 |
| Mean cross-val error | 0.43 | 0.38 | 0.41 | 0.41 |
| F-measure | 1 | 0 | 0.25 | 1 |
| Localization error | | | | |
| Left posterior | 0 mm | 16.7 mm | 11.5 mm | 0 mm |
| Right posterior | 0 mm | 18.0 mm | 0 mm | 0 mm |
| Left anterior | 0 mm | 19.7 mm | 12.1 mm | 0 mm |
| Right anterior | 0 mm | 54.1 mm | 41.9 mm | 0 mm |
| Sum | 0 mm | 108.4 mm | 65.5 mm | 0 mm |
| Estimated number of active sources | 4.0 | 3.3 | 4.0 | 4.0 |




**Supplementary Figure 7:** Test subject 9, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 9 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see **Methods**). Finally we show the estimated number of sources.

**Supplementary Table 3:** Test subject 9: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 7) and template/subject specific forward models. Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior.

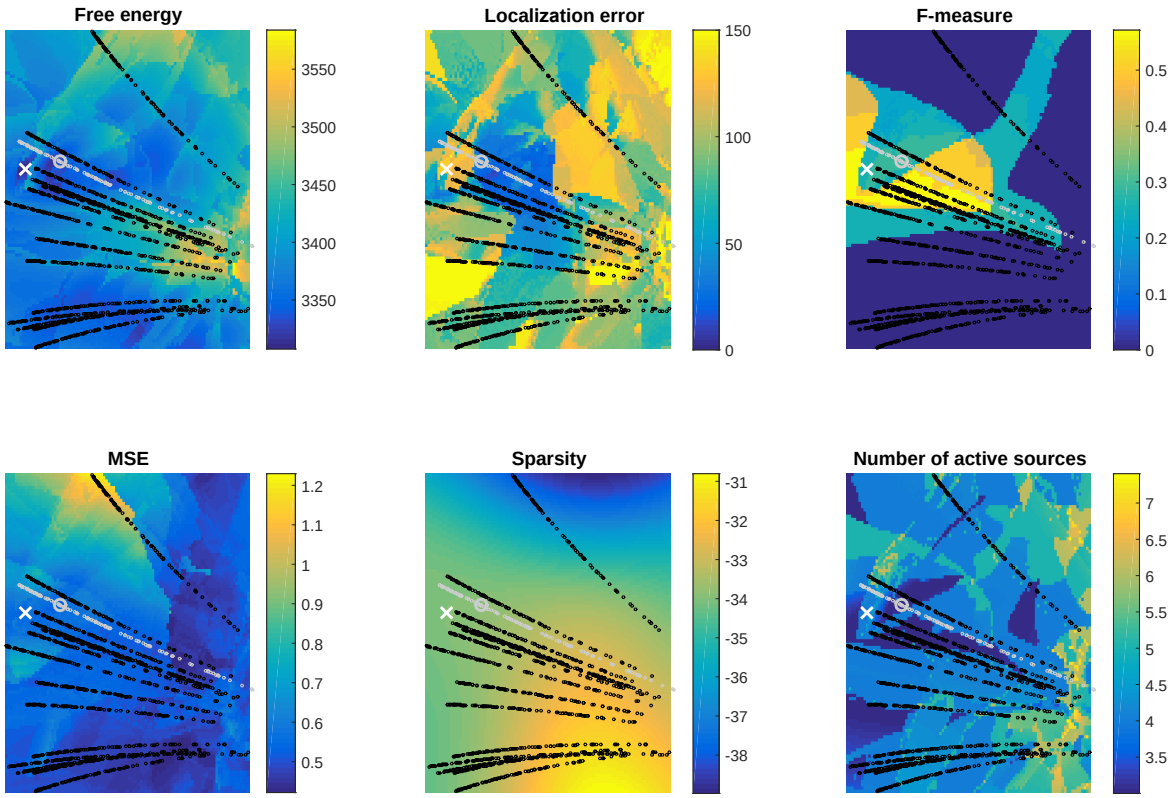
| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 3298 | 3661 | 3416 | 3273 |
| Mean cross-val error | 0.47 | 0.55 | 0.43 | 0.63 |
| F-measure | 0.75 | 0 | 0.57 | 0.75 |
| Localization error | | | | |
| Left posterior | 0 mm | 19.7 mm | 12.2 mm | 0 mm |
| Right posterior | 0 mm | 20.2 mm | 20.1 mm | 0 mm |
| Left anterior | 5.7 mm | 13.1 mm | 0 mm | 5.7 mm |
| Right anterior | 0 mm | 23.7 mm | 0 mm | 0 mm |
| Sum | 5.7 mm | 76.7 mm | 32.3 mm | 5.7 mm |
| Estimated number of active sources | 4.0 | 3.3 | 3.0 | 4.0 |




**Supplementary Figure 8:** Test subject 11, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 11 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see **Methods**). Finally we show the estimated number of sources.

**Supplementary Table 4:** Test subject 11: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 8) and template/subject specific forward models. Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior.

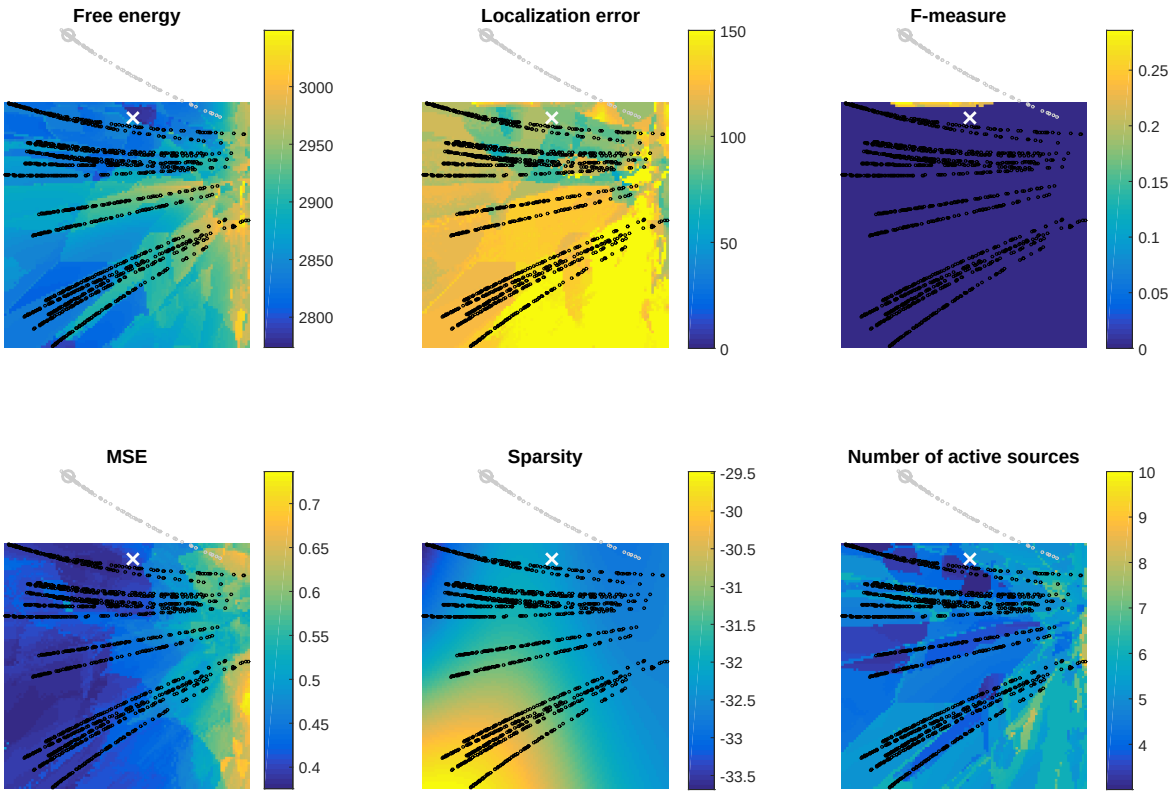
| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 3233 | 3390 | 3329 | 3181 |
| Mean cross-val error | 0.61 | 0.40 | 0.43 | 0.30 |
| F-measure | 0.75 | 0 | 0.57 | 1 |
| Localization error | | | | |
| Left posterior | 0 mm | 16.7 mm | 11.5 mm | 0 mm |
| Right posterior | 0 mm | 18.0 mm | 0 mm | 0 mm |
| Left anterior | 5.7 mm | 13.1 mm | 0 mm | 0 mm |
| Right anterior | 0 mm | 44.1 mm | 35.3 mm | 0 mm |
| Sum | 5.7 mm | 91.9 mm | 46.8 mm | 0 mm |
| Estimated number of active sources | 4.0 | 3.2 | 3.0 | 4.0 |




**Supplementary Figure 9:** Test subject 12, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 12 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see **Methods**). Finally we show the estimated number of sources.

**Supplementary Table 5:** Test subject 12: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 9) and template/subject specific forward models. Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior. * Estimated source closer to planted right posterior than right anterior.

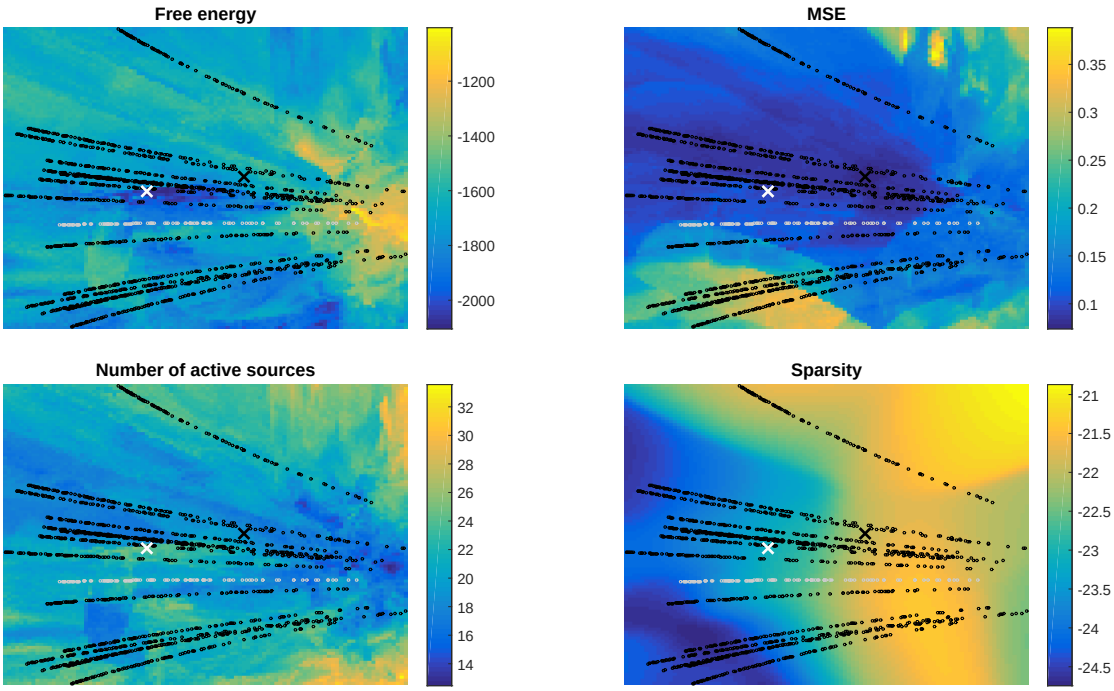
| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 3308 | 3433 | 3341 | 3260 |
| Mean cross-val error | 0.56 | 0.46 | 0.46 | 0.36 |
| F-measure | 0.25 | 0 | 0.25 | 1 |
| Localization error | | | | |
| Left posterior | 0 mm | 17.3 mm | 13.8 mm | 0 mm |
| Right posterior | 6.7 mm | 18.4 mm | 0 mm | 0 mm |
| Left anterior | 28.2mm | 4.8 mm | 15.1 mm | 0 mm |
| Right anterior | 96.7 mm* | 40.3 mm | 42.5 mm | 0 mm |
| Sum | 131.6 mm | 80.8 mm | 71.5 mm | 0 mm |
| Estimated number of active sources | 4.0 | 4.1 | 4.0 | 4.0 |




**Supplementary Figure 10:** Test subject 16, simulated EEG data: Search for optimum forward model (white cross) in the PCA space created by 15 training subjects (black dots). The projections of the forward models of test subject 16 are shown in gray, where the larger circle indicate the forward model used to generate the simulated EEG. The free energy is used to find the optimum forward model. The localization error depicted is the sum of errors in mm across the four planted sources. F-measure scale: 0 (zero sources are correctly identified) to 1 (perfect correspondence of planted and recovered locations of sources). MSE: normalized mean square four-fold cross-validation error. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest (the kernel bandwidth of the sparsity smoothing was set to 1, see **Methods**). Finally we show the estimated number of sources. Note that the training subjects does not cover the projection of subject 16 thus making the recovery of a good forward model difficult for this test subject.

**Supplementary Table 6:** Test subject 16: Performance of simulated forward model with lowest free energy (white cross in Supplementary Fig. 10) and template/subject specific forward models. Four sources were planted, one in each half hemisphere, i.e. left/right and posterior/anterior.

| | Optimized forward model | Template head, template $\sigma_{skull:brain}$ | Subject head, template $\sigma_{skull:brain}$ | Subject head, true $\sigma_{skull:brain}$ |
|---|---|---|---|---|
| Free energy | 2774 | 2850 | 2783 | 2730 |
| Mean cross-val error | 0.41 | 0.42 | 0.42 | 0.64 |
| F-measure | 0 | 0.29 | 0.75 | 0.75 |
| Localization error | | | | |
| Left posterior | 33.7 mm | 26.5 mm | 0 mm | 0 mm |
| Right posterior | 25.8 mm | 14.9 mm | 0 mm | 0 mm |
| Left anterior | 26.2 mm | 54.5 mm | 17.6 mm | 15.4 mm |
| Right anterior | 5.9 mm | 0 mm | 0 mm | 0 mm |
| Sum | 91.6 mm | 95.9 mm | 17.6 mm | 15.4 mm |
| Estimated number of active sources | 6.0 | 3.4 | 4.2 | 4.0 |




**Supplementary Figure 11:** Search for the forward model best reconstructing the EEG data from test subject 2 in the PCA space created by 15 training subjects (corresponding to the lines formed by the black dots). The projections of the forward models of test subject 2 are shown in gray. The free energy is used to find the optimum forward model (white cross). MSE: normalized mean square four-fold cross-validation error (minimum MSE is found at the black cross). The number of active sources at lowest free energy is estimated to 24.6. The applied sparsity level is found through four-fold cross-validation on 250 training forward models which were subsequently used to extrapolate the sparsity level of the space of interest. The kernel bandwidth of the sparsity smoothing was varied between 1/4 to 3 in steps of 1/4, the figures shown are thus averages over applying these 12 levels of smoothing.

# Computational Complexity

Although the forward model inference procedure has not been time optimized we provide some indication of computational complexity for reference. Computed on a laptop with 2.1-GHz 64-bit i7 processor the sparsity estimation for one forward model takes approx. 15 min. Computing the sparsity for 250 forward models as we did will thus require approx. 2.5 days. When having the sparsity it takes approx. 5 min to estimate the source density and free energy for one forward model, meaning that calculating forward models in a $100 \times 100$ grid would take approx. 35 days. However, many steps can be performed in parallel and computation can therefore be distributed to several C/GPUs. Assuming 100 workers the forward model space can thus effectively be optimized in one day.

# References

[1] Gramfort, A., Papadopoulo, T., Olivi, E. and Clerc, M. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical Engineering Online* **9** (2010).

[2] 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).

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

[4] Wang, G., Ren, D. . Effect of brain-to-skull conductivity ratio on EEG source localization accuracy. *BioMed Research International* **459346** (2013). doi:10.1155/2013/459346

Appendix I

# Supplement to teVG

The following provides supporting information for the teVG as presented in Section 3.3.

Dual representation of the free energy $F$ (same as in eq. (3.8))

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

The partial derivatives of the free energy

$$\frac{\partial F}{\partial X_{nt}} = \beta K m_n(1-m_n)\chi_{nn}X_{nt} - \sum_{k=1}^{K}\lambda_{\mu t}m_n A_{kn} \tag{I.2}$$

$$\frac{\partial F}{\partial Z_{kt}} = \beta(Z_{kt}-Y_{kt}) + \lambda_{kt} \tag{I.3}$$

$$\frac{\partial F}{\partial \beta} = -\frac{KT}{2\beta} + \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}(Z_{kt}-Y_{kt})^2 + \frac{K}{2}\sum_{t=1}^{T}\sum_{n=1}^{N}m_n(1-m_n)X_{nt}^2\chi_{nn} \tag{I.4}$$

$$\frac{\partial F}{\partial m_n} = \frac{\beta K}{2}\sum_{t=1}^{T}(1-2m_n)X_{nt}^2\chi_{nn} - \gamma + \log\left(\frac{m_n}{1-m_n}\right) - \sum_{t=1}^{T}\sum_{k=1}^{K}\lambda_{kt}X_{nt}A_{kn} \tag{I.5}$$

$$\frac{\partial F}{\partial \lambda_{kt}} = Z_{kt} - \sum_{n=1}^{N}m_n X_{nt}A_{kn}. \tag{I.6}$$

Setting the above equal to zero yields the equation set

$$X_{nt} = \frac{1}{K\beta(1-m_n)\chi_{nn}}\sum_{k=1}^{K}\lambda_{kt}A_{kn}, \tag{I.7}$$

$$C_{kk'} = \delta_{kk'} + \frac{1}{K}\sum_{n=1}^{N}\frac{m_n}{(1-m_n)\chi_{nn}}A_{kn}A_{k'n}, \tag{I.8}$$

$$\sum_{k'=1}^{K}C_{kk'}\hat{Y}_{k't} = Y_{kt}, \tag{I.9}$$

$$\frac{1}{\beta} = \frac{1}{TK}\sum_{t=1}^{T}\sum_{k=1}^{K}\hat{Y}_{kt}Y_{kt}, \tag{I.10}$$

$$\lambda_{kt} = \beta\hat{y}_{kt} \tag{I.11}$$

$$m_n = \sigma_f\left(\frac{\beta K}{2}\chi_{nn}\sum_{t=1}^{T}X_{nt}^2 + \gamma\right). \tag{I.12}$$

where $\sigma_f(a) = (1+\exp(-a))^{-1}$.

Appendix J

# Supplement to MarkoVG

The following provides supporting information for the MarkoVG as presented in Section 3.4.

Dual representation of the free energy $F$ (same as in eq. (3.13))

$$\begin{aligned} F = &- \frac{KT}{2} \log \frac{\beta}{2\pi} + \frac{\beta}{2} \sum_{t,k} (Y_{kt} - Z_{kt})^2 + \frac{K\beta}{2} \sum_{t,n} M_{nt}(1 - M_{nt}) X_{nt}^2 \chi_{nn} \\ &- \sum_{n,t} \left[ M_{nt} \log \frac{\Gamma_{10}}{\Gamma_{00}} + M_{n,t-1} \log \frac{\Gamma_{01}}{\Gamma_{00}} + (M_{nt} M_{n,t-1}) \log \frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}} \right] \\ &+ NT \log \frac{1}{\Gamma_{00}} + \sum_{n,t} \left[ M_{nt} \log(M_{nt}) + (1 - M_{nt}) \log(1 - M_{nt}) \right] \\ &+ \sum_{t,k} \lambda_{kt} \left( Z_{kt} - \sum_{n} A_{kn} M_{nt} X_{nt} \right), \end{aligned} \tag{J.1}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix}. \tag{J.2}$$

The partial derivatives of the free energy

$$\frac{\partial F}{\partial X_{nt}} = K\beta M_{nt}(1-M_{nt})X_{nt}\chi_{nn} - \sum_k \lambda_{kt}A_{kn}M_{nt}, \tag{J.3}$$

$$\frac{\partial F}{\partial Z_{kt}} = \beta(Z_{kt}-Y_{kt}) + \lambda_{kt}, \tag{J.4}$$

$$\frac{\partial F}{\partial \beta} = -\frac{KT}{2\beta} + \frac{1}{2}\sum_{t,k}(Y_{kt}-Z_{kt})^2 + \frac{K}{2}\sum_{t,n} M_{nt}(1-M_{nt})X_{nt}^2\chi_{nn}, \tag{J.5}$$

$$\frac{\partial F}{\partial M_{nt}} = \frac{K\beta}{2}(1-2M_{nt})X_{nt}^2\chi_{nn} + \log\left(\frac{M_{nt}}{1-M_{nt}}\right) + \log\left(\frac{\Gamma_{00}^2}{\Gamma_{10}\Gamma_{01}}\right), \tag{J.6}$$

$$+ (M_{n(t-1)} + M_{n(t+1)})\log\left(\frac{\Gamma_{01}\Gamma_{10}}{\Gamma_{00}\Gamma_{11}}\right) - \sum_k \lambda_{kt}A_{kn}X_{nt}, \tag{J.7}$$

$$\frac{\partial F}{\partial \lambda_{kt}} = Z_{kt} - \sum_n A_{kn}M_{nt}X_{nt}. \tag{J.8}$$

Setting the above equal to zero yields the equation set

$$X_{nt} = \frac{1}{K\beta}\frac{1}{(1-M_{nt})\chi_{nn}}\sum_k \lambda_{kt}A_{kn}, \tag{J.9}$$

$$Z_{kt} = Y_{kt} - \frac{1}{\beta}\lambda_{kt}, \tag{J.10}$$

$$\beta = \frac{1}{TK}\sum_{t,k,c}\lambda_{kt}\lambda_{ct}C_{kct}, \tag{J.11}$$

$$Y_{kt} = \frac{1}{\beta}\sum_c \lambda_{ct}C_{kct}, \tag{J.12}$$

$$\lambda_{ct} = \beta\hat{Y}_{ct}, \tag{J.13}$$

$$M_{nt} = \sigma_f\left(\frac{K\beta}{2}\chi_{nn}X_{nt}^2 + \gamma_1 + \gamma_2\left(M_{n(t-1)} + M_{n(t+1)}\right)\right), \tag{J.14}$$

where $C_{kct} = \delta_{kc} + \frac{1}{K}\sum_n \frac{M_{nt}}{(1-M_{nt})\chi_{nn}}A_{kn}A_{cn}$, $\sum_c C_{kct}\hat{Y}_{ct} = Y_{kt}$, $\sigma_f(a) = (1+\exp(-a))^{-1}$, $\gamma_1 = \log\left(\frac{\Gamma_{10}\Gamma_{01}}{\Gamma_{00}^2}\right)$ and $\gamma_2 = \log\left(\frac{\Gamma_{00}\Gamma_{11}}{\Gamma_{01}\Gamma_{10}}\right)$.

# Bibliography


Ahn, M., Hong, J. H., and Jun, S. C. (2012). Feasibility of approaches combining sensor and source features in brain-computer interface. *Journal of neuroscience methods*, 204(1):168–178.

Akalin Acar, Z., Acar, C. E., and Makeig, S. (2016). Simultaneous head tissue conductivity and EEG source location estimation. *NeuroImage*, 124:168–180.

Akalin Acar, Z. and Makeig, S. (2013). Effects of forward model errors on EEG source localization. *Brain topography*, 26(3):378–396.

Andersen, M. R., Hansen, S. T., and Hansen, L. K. (2013). Learning the solution sparsity of an ill-posed linear inverse problem with the Variational Garrote. *Machine Learning for Signal Processing (MLSP), IEEE International Workshop on*, pages 1–6.

Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3):839–851.

Awada, K. A., Jackson, D. R., Baumann, S. B., Williams, J. T., Wilton, D. R., Fink, P. W., and Prasky, B. R. (1998). Effect of conductivity uncertainties and modeling errors on EEG source localization using a 2-D model. *IEEE transactions on bio-medical engineering*, 45(9):1135–1145.

Aydin, Ü., Vorwerk, J., Küpper, P., Heers, M., Kugel, H., Galka, A., Hamid, L., Wellmer, J., Kellinghaus, C., Rampp, S., and Wolters, C. H. (2014). Combining EEG and MEG for the reconstruction of epileptic activity using a calibrated realistic volume conductor model. *PLoS ONE*, 9(3).

Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *Signal Processing Magazine, IEEE*, 18(6):14–30.

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

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of cognitive neuroscience*, 8(6):551–565.

Besserve, M., Martinerie, J., and Garnero, L. (2011). Improving quantification of functional networks with EEG inverse problem: evidence from a decoding point of view. *NeuroImage*, 55(4):1536–1547.

Bolstad, A., Van Veen, B., and Nowak, R. (2009). Space-time event sparse penalization for magneto-/electroencephalography. *NeuroImage*, 46(4):1066–81.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., and Lehmann, D. (1981). Sleep deprivation: Effect on sleep stages and EEG power density in man. *Electroencephalography and Clinical Neurophysiology*, 51(5):483–493.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Kiehl, K. A. (2006). Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. *NeuroImage*, 30(2):544–553.

Claassen, J., Mayer, S. A., Kowalski, R. G., Emerson, R. G., and Hirsch, L. J. (2004). Detection of electrographic seizures with continuous EEG monitoring in critically ill patients. *Neurology*, 62(10):1743–1748.

Cohen, D. and Cuffin, B. (1983). Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, 56(1):38–51.

Correa, N. M., Adali, T., Li, Y.-O., and Calhoun, V. D. (2010). Canonical correlation analysis for data fusion and group inferences: Examining applications of medical imaging data. *IEEE Sig Proc*, 27(4):39–50.

Cotter, S. F., Rao, B. D., Engan, K., and Kreutz-delgado, K. (2005). Sparse Solutions to Linear Inverse Problems With Multiple Measurement Vectors. *IEEE Transactions on Signal Processing*, 53(7):2477–2488.

Dähne, S. (2015). *Decomposition Methods for the Fusion of Multimodal Neuroimaging Data*. Phd thesis, Technische Universität Berlin.

Dähne, S., Bießmann, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2013). Integration of Multivariate Data Streams With Bandpower Signals. *IEEE Transactions on Multimedia*, 15(5):1001–1013.

Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate Machine Learning Methods for Fusing Multimodal Functional Neuroimaging Data. *Proceedings of the IEEE*, 103(9):1507–1530.

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., and Nikulin, V. V. (2014). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86:111–122.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.

Dalrymple, K. A., Oruç, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., Handy, T. C., and Barton, J. J. S. (2011). The anatomic basis of the right face-selective N170 in acquired prosopagnosia: a combined ERP/fMRI study. *Neuropsychologia*, 49(9):2553–2563.

Dang-Vu, T. T., Schabus, M., Desseilles, M., Albouy, G., Boly, M., Darsaud, A., Gais, S., Rauchs, G., Sterpenich, V., Vandewalle, G., Carrier, J., Moonen, G., Balteau, E., Degueldre, C., Luxen, A., Phillips, C., and Maquet, P. (2008). Spontaneous neural activity during human slow wave sleep. *PNAS*, 105(39):15160–15165.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pélégrini-Issac, M., Lina, J.-M., and Benali, H. (2007). Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. *NeuroImage*, 36(1):69–87.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PloS one*, 7(2):1–14.

Donoho, D. L., Elad, M., and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *Transactions on Information Theory, IEEE*, 52(1):6–18.

Draper, N., Smith, H., and Pownell, E. (1966). Applied regression analysis.

Edelman, B. J., Baxter, B., and He, B. (2016). EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE transactions on bio-medical engineering*, 63(1):4–14.

Eichele, T., Calhoun, V. D., Moosmann, M., Specht, K., Jongsma, M. L. A., Quiroga, R. Q., Nordby, H., and Hugdahl, K. (2008). Unmixing concurrent EEG-fMRI with parallel independent component analysis. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 67(3):222–34.

Eimer, M. and McCarthy, R. (1999). Prosopagnosia and structural encoding of faces: Evidence from event related potentials. *Neuroreport*, 10:255–259.

Eldar, Y. C. and Rauhut, H. (2010). Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation. *IEEE Transactions on Information Theory*, 56(1):505–519.

Flandin, G. and Novak, M. J. U. (2013). fMRI Data Analysis Using SPM. In S. Ulmer, O. J., editor, *fMRI, Basics and Clinical Application*, chapter 6, pages 64–66. Springer-Verlag Berlin Heidelberg, 2nd edition.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736*.

Friston, K. J., Harrison, L., Daunizeau, J., Kiebel, S. J., Phillips, C., Trujillo-Barreto, N., Henson, R. N., Flandin, G., and Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3):1104–1120.

Frølich, L., Andersen, T. S., and Mørup, M. (2015). Classification of independent components of EEG into multiple artifact classes. *Psychophysiology*, 52(1):32–45.

Gençer, N. G. and Acar, C. E. (2004). Sensitivity of EEG and MEG measurements to tissue conductivity. *Physics in Medicine and Biology*, 49(5):701–717.

Gorodnitsky, I., George, J., and Rao, B. (1995). Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical Neurophysiology*, 95:231–251.

Gorodnitsky, I. and Rao, B. (1997). Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616.

Gramfort, A., Papadopoulo, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical engineering online*, 9:45.

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M. S., and Kowalski, M. (2013). Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 70:410–422.

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel, S., and Lemahieu, I. (2007). Review on solving the forward problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, 4(46):1–29.

Hämäläinen, M. S. and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1):35–42.

Hansen, L. K., Larsen, J., Nielsen, F. A., Strother, S. C., Rostrup, E., Savoy, R., Lange, N., Sidtis, J., Svarer, C., and Paulson, O. B. (1999). Generalizable patterns in neuroimaging: how many principal components? *NeuroImage*, 9(5):534–544.

Hansen, L. K. and Rasmussen, C. E. (1994). Pruning from adaptive regularization. *Neural Computation*, 6(6):1223–1232.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., and Biessmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110.

Henderson, C., Butler, S., and Glass, A. (1975). The localization of equivalent dipoles of EEG sources by the application of electrical field theory. *Electroencephalography and Clinical Neurophysiology*, 39(2):117–130.

Henson, R. N., Goshen-Gottstein, Y., Ganel, T., Otten, L. J., Quayle, A., and Rugg, M. D. (2003). Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cerebral cortex*, 13(7):793–805.

Henson, R. N., Mattout, J., Phillips, C., and Friston, K. J. (2009a). Selecting forward models for MEG source-reconstruction using model-evidence. *Neuroimage*, 46(1):168–176.

Henson, R. N., Mouchlianitis, E., and Friston, K. J. (2009b). MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *NeuroImage*, 47(2):581–589.

Henson, R. N., Wakeman, D. G., Litvak, V., and Friston, K. J. (2011). 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:Article 76.

Hernández-Lobato, D., Hernández-Lobato, J. M., Helleputte, T., and Dupont, P. (2010). Expectation Propagation for Bayesian Multi-task Feature Selection. In Balcázar, J. L., Bonchi, F., Gionis, A., and Sebag, M., editors, *Machine Learning and Knowledge Discovery in Databases*, volume 6321 of *Lecture Notes in Computer Science*, pages 522–537. Springer Berlin Heidelberg.

Höhne, J., Schreuder, M., Blankertz, B., and Tangermann, M. (2011). A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. *Frontiers in neuroscience*, 5:Article 99.

Householder, A. S. (2013). *The Theory of Matrices in Numerical Analysis*. 2nd edition.

Iaria, G., Fox, C. J., Waite, C. T., Aharon, I., and Barton, J. J. S. (2008). The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: Neuropsychological and neuroimaging evidence. *Neuroscience*, 155(2):409–422.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Kähkönen, S., Ahveninen, J., Jääskeläinen, I. P., Kaakkola, S., Näätänen, R., Huttunen, J., and Pekkonen, E. (2001). Effects of haloperidol on selective attention: a combined whole-head MEG and high-resolution EEG study. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 25(4):498–504.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 17(11):4302–4311.

Kappen, H. (2011). The Variational Garrote. *arXiv preprint arXiv:1109.0486*.

Kappen, H. J. and Gómez, V. (2014). The Variational Garrote. *Machine Learning*, 96(3):269–294.

Klimchuk, J. A., Tanner, S. E. M., and De Moortel, I. (2004). Coronal Seismology and the Propagation of Acoustic Waves along Coronal Loops. *The Astrophysical Journal*, 616(2):1232–1241.

Laufs, H. (2012). A personalized history of EEG-fMRI integration. *NeuroImage*, 62(2):1056–1067.

Lei, X., Valdes-Sosa, P. a., and Yao, D. (2012). EEG/fMRI fusion based on independent component analysis: integration of data-driven and model-driven methods. *Journal of integrative neuroscience*, 11(3):313–337.

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

Lew, S., Wolters, C. H., Anwander, A., Makeig, S., and MacLeod, R. S. (2009). Improved EEG source analysis using low-resolution conductivity estimation in a four-compartment finite element head model. *Human brain mapping*, 30(9):2862–2878.

Lina, J. M., Chowdhury, R., Lemay, E., Kobayashi, E., and Grova, C. (2014). Wavelet-based localization of oscillatory sources from magnetoencephalography data. *IEEE Transactions on Biomedical Engineering*, 61(8):2350–2364.

MacKay, D. J. C. (1995). Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3):469–505.

Makhoul, J., Kubala, F., Schwartz, R., and Weischedel, R. (1999). Performance measures for information extraction. In *Proceedings of DARPA Broadcast News Workshop*, pages 249–252. Morgan Kaufmann Pub.

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C., and Kennedy, H. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral cortex*, 24(1):17–36.

Martinez-Cantin, R. (2014). BayesOpt: a Bayesian optimization library for nonlinear optimization, experimental design and bandits. *The Journal of Machine Learning Research*, 15(1):3735–3739.

Mather, M., Cacioppo, J. T., and Kanwisher, N. (2013). How fMRI Can Inform Cognitive Theories. *Perspectives on Psychological Science*, 8(1):108–113.

Matsuura, K. and Okabe, Y. (1995). Selective minimum-norm solution of the biomagnetic inverse problem. *IEEE Transactions on Biomedical Engineering*, 42(6):608–615.

McNeely, H. E., Lau, M. A., Christensen, B. K., and Alain, C. (2008). Neurophysiological evidence of cognitive inhibition anomalies in persons with major depressive disorder. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 119(7):1578–1589.

Modarreszadeh, M. and Schmidt, R. (1997). Wireless, 32-channel, EEG and epilepsy monitoring system. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering'*, volume 3, pages 1157–1160.

Montoya-Martinez, J., Hansen, L. K., and Massimiliano, P. (2012). Structured sparsity regularization approach to the EEG inverse problem. *3rd International Workshop on Cognitive Incromation Processing.*

Moosmann, M., Ritter, P., Krastel, I., and Brink, A. (2003). Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *Neuroimage*, 20(1):145–158.

Mullen, T., Kothe, C., Chi, M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., and Cauwenberghs, G. (2015). Real-time Neuroimaging and Cognitive

Monitoring Using Wearable Dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., and Winterer, G. (2010). Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. *NeuroImage*, 52(4):1149–1161.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Nisar, H. and Yeap, K. H. (2014). EEG/ERP Analysis Methods and Applications. In Nidal, K. and Malik, A. S., editors, *EEG/ERP Analysis - Methods and Applications*, chapter 1. CRC Press.

Nunez, P. and Srinivasan, R. (2006a). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2nd edition.

Nunez, P. L. and Srinivasan, R. (2006b). A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, 117(11):2424–2435.

Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M., and Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophysical journal*, 64(3):803–812.

Oostendorp, T. F., Delbeke, J., and Stegeman, D. F. (2000). The conductivity of the human skull: results of in vivo and in vitro measurements. *IEEE transactions on bio-medical engineering*, 47(11):1487–1492.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011.

Oostenveld, R. and Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human brain mapping*, 17(3):179–192.

Pedersen, M. S., Baxter, B., Templeton, B., Rishøj, C., Theobald, D. L., Hoegh-rasmussen, E., Casteel, G., Gao, J. B., Dedecius, K., Strim, K., Christiansen, L., Hansen, L. K., Wilkinson, L., He, L., Bar, M., Winther, O., Sakov, P., and Hattinger, S. (2008). The Matrix Cookbook. Technical report, Technical University of Denmark.

Perdue, K. L. and Diamond, S. G. (2014). T1 magnetic resonance imaging head segmentation for diffuse optical tomography and electroencephalography. *Journal of biomedical optics*, 19(2):026011.

Petersen, M., Stahlhut, C., and Stopczynski, A. (2011). Smartphones get emotional: mind reading images and reconstructing the neural sources. *Affective Computing Intelligent Interaction*, pages 578–587.

Phillips, C. (2000). *Source estimation in EEG*. Phd thesis, University de Liege, Belgium.

Pion-tonachini, L., Member, S., Hsu, S.-h., and Member, S. (2015). Real-time EEG Source-mapping Toolbox ( REST ): Online ICA and Source Localization. In *Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE*, pages 4114–4117.

Plis, S. M., George, J. S., Jun, S. C., Ranken, D. M., Volegov, P. L., and Schmidt, D. M. (2007). Probabilistic forward model for electroencephalography source analysis. *Physics in medicine and biology*, 52(17):5309–5327.

Polich, J. (1997). EEG and ERP assessment of normal aging. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 104(3):244–256.

Rice, J. K., Rorden, C., Little, J. S., and Parra, L. C. (2013). Subject position affects EEG magnitudes. *NeuroImage*, 64:476–484.

Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann, 2nd edition.

Rudy, Y. (2015). The forward problem of electrocardiography revisited. *Circulation. Arrhythmia and electrophysiology*, 8(3):526–528.

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

Scherg, M. and Von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(1):32–44.

Schüz, A. and Braitenberg, V. (2002). The Human Cortical White Matter: Quantitative Aspects of Cortico-Cortical Long-Range Connectivity. In Almut Schuez and Miller, R., editors, *Cortical Areas: Unity and Diversity*, chapter 16. CRC Press.

Schweinberger, S. R., Pfütze, E.-M., and Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3):722–736.

Singh, M., Patel, P., and Al-Dayeh, L. (1998). fMRI of brain activity during alpha rhythm. *Int. Soc. Mag. Res. Med.*

Stahlhut, C., Attias, H. T., Sekihara, K., and Wipf, D. (2013). A hierarchical Bayesian M/EEG imaging method correcting for incomplete spatio-temporal priors. *10th IEEE International Symposium on Biomedical Imaging (ISBI 2013)*, pages 560–563.

Stahlhut, C., Attias, H. T., and Stopczynski, A. (2012). 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*, pages 1538–1541.

Stahlhut, C., Mørup, M., Winther, O., and Hansen, L. K. (2011). Simultaneous EEG Source and Forward Model Reconstruction (SOFOMORE) Using a Hierarchical Bayesian Approach. *Journal of Signal Processing Systems*, 65(3):431–444.

Stenroos, M. and Nummenmaa, A. (2016). Incorporating and compensating cerebrospinal fluid in surface-based forward models of magneto- and electroencephalography. *bioRxiv preprint preprint*, (January).

Stippich, C. (2007). *Clinical functional MRI.* Springer-Verlag Berlin Heidelberg.

Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K., and Hansen, L. K. (2014). The smartphone brain scanner: a portable real-time neuroimaging system. *PloS one*, 9(2):e86733.

Thompson, M. and Thompson, L. (2008). Biofeedback for movement disorders (dystonia with Parkinson ' s disease ): Theory and preliminary results biofeedback for movement disorders. *Journal of Neurotherapy : Investigations in Neuromodulation , Neurofeedback and Applied Neuroscience*, 6(4):51–70.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 1(1):211–244.

Titsias, M. and Lázaro-Gredilla, M. (2011). Spike and slab variational inference for multi-task and multiple kernel learning. *Advances in neural information processing systems*, 24:2339–2347.

Tommasi, A. (1998). Forward modeling of the development of seismic anisotropy in the upper mantle. *Earth and Planetary Science Letters*, 160(1-2):1–13.

Trujillo-Barreto, N., Aubert-Vázquez, E., and Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21(4):1300–1319.

Tuch, D. S., Wedeen, V. J., Dale, A. M., George, J. S., and Belliveau, J. W. (2001). Conductivity tensor mapping of the human brain using diffusion tensor MRI. *Proceedings of the National Academy of Sciences*, 98(20):11697–11701.

Van Veen, B. D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE transactions on bio-medical engineering*, 44(9):867–880.

Wakeman, D. G. and Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*, 2:150001.

Wang, G. and Ren, D. (2013). Effect of brain-to-skull conductivity ratio on EEG source localization accuracy. *BioMed research international*, page 459346.

Windhoff, M., Opitz, A., and Thielscher, A. (2013). 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(4):923–935.

Wipf, D. P. and Rao, B. D. (2007). An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem. *IEEE Transactions on Signal Processing*, 55(7):3704–3716.

Wolters, C. H., Anwander, A., Tricoche, X., Weinstein, D., Koch, M. A., and MacLeod, R. S. (2006). Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. *NeuroImage*, 30(3):813–826.

Zhang, Z. and Rao, B. D. (2011). Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *Selected Topics in Signal Processing, IEEE*, 5(5):912–926.