A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals

A Preprint

Alejandro Ojeda[12]*, Marius Klug[3], Kenneth Kreutz-Delgado[2], Klaus Gramann[345], Jyoti Mishra[1]
[1] Neural Engineering and Translation Labs, Department of Psychiatry, University of California San Diego, USA
[2] Department of Electrical and Computer Engineering, University of California San Diego, USA
[3] Institute of Psychology and Ergonomics, Technische Universität Berlin, Germany
[4] Center for Advanced Neurological Engineering, University of California San Diego, USA
[5] School of Software, University of Technology Sydney, Australia

February 23, 2019

Abstract

Electroencephalographic (EEG) source imaging depends upon sophisticated signal processing algorithms for data cleaning, source separation, and localization. Typically, these problems are addressed separately using a variety of heuristics, making it difficult to systematize a methodology for extracting robust EEG source estimates on a wide range of experimental paradigms. In this paper, we propose a unifying Bayesian framework in which these apparently dissimilar problems can be understood and solved in a principled manner using a single algorithm. We explicitly model the effect of non-brain sources by augmenting the lead field matrix with a dictionary of stereotypical artifact scalp projections. We propose to populate the artifact dictionary with non-brain scalp projections obtained by running Independent Component Analysis (ICA) on an EEG database. Within a parametric empirical Bayes (PEB) framework, we use an anatomical brain atlas to parameterize a source prior distribution that encourages sparsity in the number of cortical regions. We show that, in our inversion algorithm, PEB+ (PEB with the addition of artifact modeling), the sparsity prior has the property of inducing the segregation of the cortical activity into a few maximally independent components with known anatomical support. Artifacts produced by electrooculographic and electromyographic activity as well as single-channel spikes are also segregated into their respective components. Of theoretical relevance, we use our framework to point out the connections between Infomax ICA and distributed source imaging. We use real data to demonstrate that PEB+ outperforms Infomax for source separation on short segments of data and, unlike the popular Artifact Subspace Removal algorithm, it can reduce artifacts without significantly distorting clean epochs. Finally, we analyze mobile brain/body imaging data to characterize the brain dynamics supporting heading computation during full-body rotations. In this example, we run PEB+ followed by the spectral analysis of the activity in the retrosplenial cortex, largely replicating the findings of previous experimental literature.

***K*eywords** Electroencephalographic source imaging, Parametric Empirical Bayes, block-sparse learning, Infomax, Independent Component Analysis, ICA, EEG, subspace, artifact, rejection, mobile brain/body imaging, MoBI

*For correspondence: `alojeda@ucsd.edu` (AO)

Introduction

The electroencephalogram (EEG) is a noninvasive functional brain imaging modality that allows the study of brain electrical activity with excellent temporal resolution. Compared to other noninvasive imaging modalities such as fMRI, PET, SPECT, and MEG, EEG acquisition can be mobile and more affordable [Mcdowell et al., 2013, Mehta and Parasuraman, 2013], allowing the widespread study of human cognition and behavior under more ecologically valid experimental conditions [Makeig et al., 2009]. Imaging cognitive processes while participants engage naturally with their environment (natural cognition in action [Gramann et al., 2014]) has potential for developing a new generation of applications in brain-computer interfaces (BCI), mental health, rehabilitation, and neuroergonomics [Mishra and Gazzaley, 2014, Jungnickel and Gramann, 2016, Wagner et al., 2016]. However, despite impressive methodological advances in the estimation of the electrical activity of the cortex from EEG voltages recorded on the scalp, a number of practical and theoretical issues remain unsolved.

Imaging EEG source activity (also known as electromagnetic source imaging or ESI) is challenging for several reasons. First, since many configurations of currents in the brain can elicit the same EEG scalp topography [Michel and Murray, 2012], it entails solving an ill-posed inverse problem [Lopes da Silva, 2013]. Second, the EEG signal is often contaminated by artifacts of non-brain origin such as electrooculographic (EOG) and electromyographic (EMG) activity that need to be identified and removed. Third, there is evidence that large-scale brain responses measured by EEG are generated by underlying cortical dynamics that evolve over time and can exhibit nonlinear features [Breakspear, 2017, Khambhati et al., 2018], thereby rendering the simplifying assumptions of linearity and stationarity used by most inverse methods hard to justify. These problems are usually addressed separately using a variety of heuristics, making it difficult to systematize a methodology for obtaining biologically plausible EEG source estimates in the presence of artifacts and nonlinear and nonstationary dynamics. The objective of this paper is to develop a unifying Bayesian framework in which these, apparently dissimilar, problems can be understood and solved in a principled manner using a single algorithm.

To cope with the ill-posed nature of the inverse problem and ensure functional images with biological relevance, several inverse algorithms have been proposed that seek to estimate EEG sources subject to neurophysiologically reasonable spatial [Haufe et al., 2011, Friston et al., 2008, Trujillo-Barreto et al., 2004, Pascual-Marqui et al., 2002, Baillet et al., 2001], spatiotemporal [Martínez-Vargas et al., 2015, Valdés-Sosa et al., 2009, Trujillo-Barreto et al., 2008], and frequency-domain [Gramfort et al., 2013] constraints, just to mention a few examples. These approaches can work relatively well when the EEG samples are corrupted by Gaussian noise and the signal to noise ratio (SNR) is high. In practice, however, raw EEG data are affected by many other types of noise such as interference from the 50/60 Hz AC line, pseudo-random muscle activity, and mechanically induced artifacts, among others. Thus, before source estimation, non-Gaussian artifacts need to be removed from the data.

There is a plethora of methods for dealing with artifacts corrupting the EEG signal [Mannan et al., 2018, Islam et al., 2016]. Popular approaches used in real-time BCI applications are based on adaptive noise cancellation [Kilicarslan et al., 2016] or Artifact Subspace Removal (ASR) [Mullen et al., 2015] algorithms. The former has the inconvenience that an additional channel recording purely artifactual activity (i.e., EOG or EMG activity not admixed with EEG) needs to be provided, while the latter rests on the assumption that the statistics of data and artifacts stay the same after an initial calibration phase. In studies where the data can be analyzed offline, artifactual components can be largely removed using Independent Component Analysis (ICA) [Jung et al., 2000]. ICA-based cleaning, however, has the drawback that non-brain components need to be identified for removal, which is usually done manually based on the practitioner's experience.

ICA is a special case of blind source separation (BSS) method [Cichocki and Amari, 2002] that can be used to linearly decompose EEG data into components that are maximally statistically independent. ICA has been used to analyze event-related potentials (ERP) under the assumptions that during the task 1) the decomposition is stationary and 2) that brain components can be modeled as a predefined number of dipolar point processes with fixed spatial location and orientation [Makeig and Onton, 2011]. The stationarity assumption can be relaxed using a mixture of ICA models [Palmer et al., 2011] while the selection of brain scalp projections is typically done either manually or automatically based on the residual variance afforded by a dipole fitting algorithm. The practical use of ICA has been limited by its computational cost and the need for user intervention. Only recently, a real-time recursive ICA algorithm has been proposed [Hsu et al., 2016], as well as a number of automatic methods for minimizing the subjectivity of manual component selection [Tamburro et al., 2018, Pion-Tonachini et al., 2017, Radüntz et al., 2017]. Despite these advances, turning ICA into a brain imaging modality requires that after source separation, we solve the inverse problem of localizing the set of identified brain components into the cortical space.

One way of estimating EEG sources subject to multiple assumptions (constraints) in a principled manner

is to use the framework of parametric empirical Bayes (PEB) [Morris, 1983, Casella, 1985]. In this framework, constraints are used to furnish prior probability density functions (pdfs). Empirical Bayes methods use data to infer the parameters controlling the priors (hyperparameters), such that those assumptions that are not supported by the data can be automatically discarded without user intervention. Here we use priors to "encourage" source images to belong to a functional space with biological relevance, but the exact form of those priors is determined by the data (empirically). In the context of sparsity-inducing priors, PEB is sometimes referred to as Sparse Bayesian Learning (SBL) [Tipping, 2001]. The PEB framework has been applied extensively to brain imaging for inverting hierarchical models of fMRI and PET [Friston et al., 2002] as well as EEG responses [Henson et al., 2011, Friston et al., 2008].

In this paper, within the PEB framework, we propose a probabilistic generative model (PGM) of how the raw EEG time series arises from the superposition of the brain and non-brain (artifact) source activity. We use an anatomical brain atlas to parameterize the source prior such that we induce sparsity in the number of cortical regions active at any given time (this is known in the SBL literature as block-sparsity [Zhang and Rao, 2013]). Then, our model can be inverted using an efficient block SBL algorithm recently proposed by Ojeda et al. [2018]. Henceforth, we refer to this new approach as PEB+ (PEB with the addition of artifact modeling). Our main contribution is that, by explicitly modeling non-brain sources within the PEB framework, we can unify three of the most common problems in EEG analysis: data cleaning, source separation, and source imaging. In addition, we show that the PEB+ algorithm has online adaptation, thereby allowing it to capture transient brain dynamics. On the theoretical side, we point out the connections between distributed source imaging and ICA, two popular approaches that are often perceived to be at odds with one another.

The rest of the paper is organized into two main sections concerned with methods and results respectively. **Methods:** In Section 2.1 we propose the augmented PGM of the EEG taking into account the contribution of artifact sources and motivate all our modeling assumptions. In Section 2.2 we outline the model inversion algorithm. In Sections 2.3-2.3.1 we point out the connections between PEB+ and ICA. **Results:** In Section 3.1 we explain how to construct an empirical dictionary of artifact scalp projections. In Sections 3.2, 3.3, and 3.4 we investigate the respective source imaging, separation, and data cleaning capabilities of the PEB+ algorithm.

Throughout this paper we use lowercase and uppercase bold characters and symbols to denote column-vectors and matrices respectively, $\hat{\mathbf{a}}$ is an estimate of the parameter vector $\mathbf{a}$, and $\mathbf{I}_N$ is a $N \times N$ identity matrix.

Methods

It has been shown that popular source estimation algorithms used in ESI such as weighted minimum l_2-norm [Baillet et al., 2001], FOCUSS [Cotter et al., 2005, Gorodnitsky and Rao, 1997], minimum current estimation [Huang et al., 2006], sLORETA [Pascual-Marqui et al., 2002], beamforming [Van Veen et al., 1997], variational Bayes [Friston et al., 2008], and others can be expressed in a unifying Bayesian framework [Wipf and Nagarajan, 2009]. We extend this framework by explicitly modeling non-brain artifact sources.

2.1 Augmented probabilistic generative model

In source imaging, the neural activity is often referred to as the primary current density (PCD) [Baillet et al., 2001] and it is defined on a grid of known cortical locations (the source space). Typically, a vector of N_y EEG measurements at sample k, $\mathbf{y}_k \in \mathbb{R}^{N_y}$, relates to N_g PCD values, $\mathbf{g}_k \in \mathbb{R}^{N_g}$, through the following linear equation [Dale and Sereno, 1993],

$$\mathbf{y}_k = \mathbf{L}\mathbf{g}_k + \mathbf{e}_k, \quad k = 1, \ldots, N \tag{1}$$

where $\mathbf{e}_k \in \mathbb{R}^{N_y}$ represents the measurement noise vector. The PCD is projected to the sensor space through the lead field matrix $\mathbf{L} = [\mathbf{l}_1, \ldots, \mathbf{l}_{N_g}] \in \mathbb{R}^{N_y \times N_g}$ ($N_y \ll N_g$) where each column $\mathbf{l}_i$ denotes the scalp projection of the ith unitary current dipole with fixed orientation within the source space. When dipole orientations are considered, then $\mathbf{L} \in \mathbb{R}^{N_y \times 3N_g}$ and we determine a source vector $\mathbf{g}_k \in \mathbb{R}^{3N_g}$. The lead field matrix is usually precomputed for a given electrical model of the head derived from a subject-specific MRI [Hallez et al., 2007]. Alternatively, if an individual MRI is not available, an approximated lead field matrix obtained from a high-resolution template can be used [Huang et al., 2016]. Then, the inverse problem of the EEG can be stated as the estimation of a source configuration $\hat{\mathbf{g}}_k$ that is likely to produce the scalp topography $\mathbf{y}_k$.

In the generative model presented above, the noise term $\mathbf{e}_k$ is assumed to be Gaussian and spatially uncorrelated with variance λ. This simplification is acceptable as long as EEG topographies are not affected by non-Gaussian pseudo-random artifacts generated by eye blinks, lateral eye movements, facial and neck muscle activity, body movement, among others. Therefore, before source estimation, EEG data are usually heavily preprocessed and cleaned [Bigdely-Shamlo et al., 2015]. Since artifacts contribute linearly to the sensors, ideally, one would like to characterize their scalp projections to describe more accurately

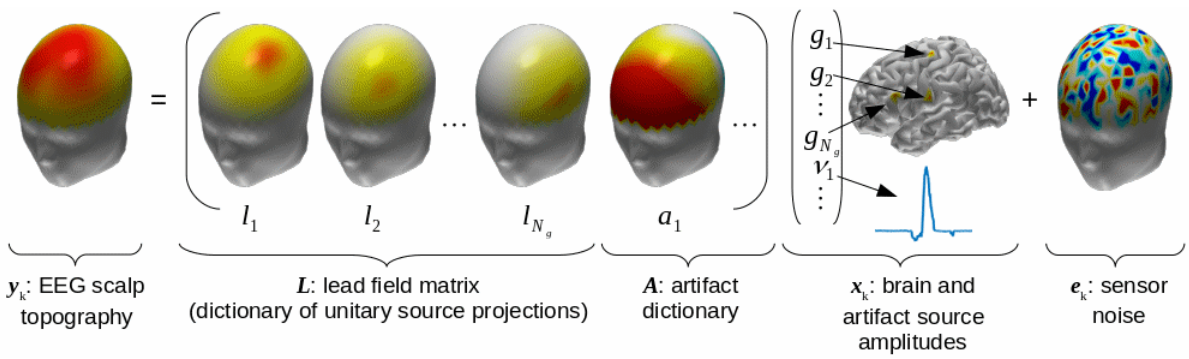

Figure 1: Proposed augmented generative model of the EEG. The model postulates that the EEG scalp topography $\mathbf{y_k}$ arises from the linear superposition of brain $g_{i,k}$ and artifact $\nu_{j,k}$ components weighted by their respective scalp projections $\mathbf{l_i}$ and $\mathbf{a_j}$, corrupted by spatially uncorrelated Gaussian noise $\mathbf{e_k}$.

the signal acquisition. To this end, we propose the following generalization of Eq (1),

$$\mathbf{y}_k = \mathbf{L}\mathbf{g}_k + \mathbf{A}\boldsymbol{\nu}_k + \mathbf{e}_k \tag{2}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{N_\nu}$ is a vector of N_ν artifact sources and $\mathbf{A} = [\mathbf{a_1}, \ldots, \mathbf{a_{N_\nu}}] \in \mathbb{R}^{\mathbf{N_y} \times \mathbf{N}_\nu}$ is a dictionary of artifact scalp projections (see Fig 1).

Although the entries of $\mathbf{A}$ that correspond to muscle activity may be obtained based on a detailed electromechanical model of the body [Böl et al., 2011], in most studies this approach may not be feasible due to computational and budgetary constraints. Janani et al. [2017] modelled $\mathbf{A}$ by expanding the lead field matrix to account for the contribution of putative scalp sources, which were assumed to be the generators of EMG activity. They used sLORETA to estimate brain and scalp sources simultaneously. Although this approach was shown to be as effective as ICA-based artifact removal, it was suggested by the authors that the use of the non-sparse solver sLORETA may lead to unrealistic configurations of brain and non-brain sources. Similarly, Fujiwara et al. [2009] augmented the magnetic lead field matrix to model the scalp contribution of two current dipoles located behind the eyes and used a Bayesian approach, that has similarities with ours, to estimate brain and eye source activity from MEG data. Although successful for removing EOG activity, in their formulation, Fujiwara et al. [2009] ignored other types of artifacts that are harder to model such as those produced by muscular activity.

In this paper, we take an empirical view inspired by the success of ICA-based artifact removal approaches. We propose constructing the dictionary $\mathbf{A}$ using a set of stereotypical artifact scalp projections such as those obtained from applying ICA to a database of EEG recordings [Bigdely-Shamlo et al., 2013a]. We then rewrite Eq (2) in a compact manner as follows,

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{e}_k \tag{3}$$

where the gain matrix is now $\mathbf{H} \triangleq [\mathbf{L}, \mathbf{A}]$ and $\mathbf{x}_k \triangleq [\mathbf{g}_k^T, \boldsymbol{\nu}_k^T]^T$ is the augmented vector of hidden (latent) brain and artifact sources (see Fig 1).

Note that, structurally, the standard generative model in Eq (1) and the augmented one in Eq (3) are identical. They differ however in that in Eq (3) we are explicitly modeling the instantaneous spatial contribution of non-brain sources to the scalp topography $\mathbf{y}_k$. Therefore, we may be able to dispense with computationally expensive preprocessing data cleaning procedures. The assumption of Gaussian measurement noise yields the following likelihood function,

$$p(\mathbf{y}_k|\mathbf{x}_k, \lambda) = N(\mathbf{y}_k|\mathbf{H}\mathbf{x}_k, \lambda\mathbf{I}_{N_y}) \tag{4}$$

Since Eq (3) does not have a unique solution, to obtain approximated source maps with biological interpretation we introduce constraints. One way of incorporating constraints in a principled manner is to express them in the form of the prior pdf of the sources $p(\mathbf{x}_k)$. Since the neural generators of the EEG are assumed to be the electrical currents produced by distributed neural masses that become locally synchronized in space and time [Nunez and Srinivasan, 2006], here we chose a parameterized prior $p(\mathbf{x}_k)$ that induces source maps to be globally sparse (seeking to explain the observed scalp topography by a few spots of cortical activity) and locally correlated (so that we obtain spatially smooth maps as opposed to maps formed by scattered isolated sources). Artifactual sources, on the other hand, can be assumed to be uncorrelated from one another. We use a Gaussian prior to express these modeling assumptions as follows,

$$p(\mathbf{x}_k|\boldsymbol{\gamma}) = N(\mathbf{x}_k|0, \boldsymbol{\Sigma}_x) \tag{5}$$

where the covariance matrix $\boldsymbol{\Sigma}_x$ has a block diagonal structure [Zhang and Rao, 2013] defined as

$$\boldsymbol{\Sigma}_x = \begin{bmatrix} \boldsymbol{\Sigma}_g & \\ & \boldsymbol{\Sigma}_\nu \end{bmatrix} \tag{6}$$

In Eq (6), the brain source prior covariance is defined as

$$\mathbf{\Sigma}_g = \begin{bmatrix} \gamma_1 C_1 & & \\ & \ddots & \\ & & \gamma_{N_{ROI}} C_{N_{ROI}} \end{bmatrix} \tag{7}$$

and $\mathbf{\Sigma}_\nu = \text{diag}(\gamma_{N_{ROI}+1}, \ldots, \gamma_{N_{ROI}+N_\nu})$ is the covariance of artifact sources. The matrices $\mathbf{C}_i \in \mathbb{R}^{N_i \times N_i}$ encode the intra-group brain source covariances and are precomputed based on source distance respecting the local folding of the cortex as described in [Ojeda et al., 2018]. $\boldsymbol{\gamma} \in \mathbb{R}^{N_{ROI}+N_\nu}$ denotes a nonnegative scale vector that encodes the sparsity profile of the group of sources. Here we define $N_{ROI} = 148$ groups based on anatomical regions of interest (ROI) obtained from the Destrieux cortical atlas [Destrieux et al., 2010]. We note that although other atlases could be used within the framework outlined in this section (e.g. the popular Desikan-Killiany 68-ROI cortical atlas [Desikan et al., 2006]), the discussion of issues pertaining to the selection of an optimal cortical parcellation is beyond the scope of this paper.

Next, let model $\mathcal{M} = \{\lambda, \boldsymbol{\gamma}\}$ be the set of hyperparameters that encode the generative model proposed above. We use the Bayes theorem to express the posterior pdf of the sources given the data and model as,

$$p(\mathbf{x}_k|\mathbf{y}_k, \mathcal{M}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M})p(\mathbf{x}_k|\mathcal{M})}{p(\mathbf{y}_k|\mathcal{M})} \tag{8}$$

Note that $p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M}) \triangleq p(\mathbf{y}_k|\mathbf{x}_k, \lambda)$ and $p(\mathbf{x}_k|\mathcal{M}) \triangleq p(\mathbf{x}_k|\gamma)$ because the likelihood and priors are independent of $\boldsymbol{\gamma}$ and λ respectively. The density function $p(\mathbf{y}_k|\mathcal{M})$ is known as the model evidence [MacKay, 2008a] and, as we will show latter, its optimization allows us to reshape our modeling assumptions in a data-driven manner.

The model evidence is the normalization constant of the posterior in Eq (8), therefore it can be ignored while searching for the maximum a posteriori (MAP) source estimate $\hat{\mathbf{x}}_k = \mathbf{x}_{MAP}$, which we find as the mode of the numerator of Eq (8) conditioned on $\mathcal{M}$,

$$\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M})p(\mathbf{x}_k|\mathcal{M}) \tag{9}$$

We readily determine the functional form of the posterior as $p(\mathbf{x}_k|\mathbf{y}_k, \mathcal{M}) = N(\mathbf{x}_k|\hat{\mathbf{x}}_k, \mathbf{\Sigma}_{x|y})$, with the following conditional mean and covariance [Wipf and Nagarajan, 2009],

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{\Sigma}_x \mathbf{H}^T \mathbf{\Sigma}_y^{-1} \mathbf{y}_k \\ \mathbf{\Sigma}_{x|y} &= \mathbf{\Sigma}_x - \mathbf{\Sigma}_x \mathbf{H}^T \mathbf{\Sigma}_y^{-1} \mathbf{H} \mathbf{\Sigma}_x \end{aligned} \tag{10}$$

where the model data covariance is given by the following expression,

$$\mathbf{\Sigma}_y = \lambda \mathbf{I}_{Ny} + \mathbf{H}\mathbf{\Sigma}_x \mathbf{H}^T \tag{11}$$

In case that we need to obtain the cleaned EEG signal $\bar{\mathbf{y}}_k$, e.g. for visualization or scalp ERP analysis, we subtract the artifact signal from the data as follows

$$\bar{\mathbf{y}}_k = \mathbf{y}_k - \mathbf{A}\hat{\boldsymbol{\nu}}_k = \mathbf{F}\mathbf{y}_k \tag{12}$$

where $\mathbf{F} = \mathbf{I} - \mathbf{A}\mathbf{\Sigma}_\nu \mathbf{A}^T \mathbf{\Sigma}_y^{-1}$ is a spatial filtering operator and $\hat{\boldsymbol{\nu}}_k$ is obtained (if needed) by selecting the last N_ν elements of the vector $\hat{\mathbf{x}}_k$. Likewise, the estimated PCD vector $\hat{\mathbf{g}}_k$ can be obtained by selecting the first N_g elements of $\hat{\mathbf{x}}_k$ or using the formula

$$\hat{\mathbf{g}}_k = \mathbf{\Sigma}_g \mathbf{L}^T \mathbf{\Sigma}_y^{-1} \mathbf{y}_k \tag{13}$$

and the source activity specific to the the ith ROI can be obtained using the formula

$$\hat{\mathbf{g}}_{i,k} = \gamma_i \mathbf{C}_i \mathbf{L}^T \mathbf{\Sigma}_y^{-1} \mathbf{y}_k \tag{14}$$

Further analysis of the source time series (e.g. ERP and connectivity analysis) can be done by averaging the source activity obtained in Eq (14) within ROIs,

$$\bar{g}_{i,k} = \frac{1}{N_i} \sum_{j \in \text{ROI}_i} \hat{\mathbf{g}}_{j,k} \tag{15}$$

where $\text{ROI}_i \subset \{1, \ldots, N_g\}$ is the subset of indices that belong to the ith ROI.

2.2 Model learning

The source estimates and cleaned data can be obtained analytically by evaluating the formulas given in Eq (10)-(13). These formulas, however, are model dependent because they depend on the specific values of the hyperparameters λ and $\boldsymbol{\gamma}$. In this section we outline the algorithm for learning those.

To evaluate Eq (10) conditioned on the optimal model estimate we maximize the posterior density,

$$\hat{\mathcal{M}} = \arg\max_{\lambda, \boldsymbol{\gamma}} p(\mathcal{M}|\mathbf{y}_k) \tag{16}$$

where $p(\mathcal{M}|\mathbf{y}_k) \propto p(\mathbf{y}_k|\mathcal{M})p(\mathcal{M})$ and $p(\mathcal{M})$ is a hyperprior. To determine the evidence described in Eq (8) we need to marginalize out the sources,

$$p(\mathbf{y}_k|\mathcal{M}) = \int p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M})p(\mathbf{x}_k|\mathcal{M})d\mathbf{x}_k \tag{17}$$

which for a linear Gaussian model like ours, is readily expressed as [Barber, 2012],

$$p(\mathbf{y}_k|\mathcal{M}) \propto \frac{\exp\big(-\frac{1}{2}\mathbf{y}_k^T \mathbf{\Sigma}_y^{-1} \mathbf{y}_k\big)}{|\mathbf{\Sigma}_y|^{-1/2}} \tag{18}$$

Next, we need to specify the hyperprior $p(\mathcal{M})$. Assuming that λ and γ_i are independent yields the factorization $p(\mathcal{M}) = p(\lambda)\prod_i p(\gamma_i)$. Since $\mathcal{M}$ contains only scale hyperparameters, a popular choice is to assume Gamma hyperpriors on a log-scale of λ^{-1} and γ_i^{-1} [Tipping, 2001]. When the scale and shape parameters of the Gamma tend to zero the hyperpriors become flat (noninformative), in which case the optimization of the model posterior density depends only

on the model evidence. This choice of hyperprior has the effect of assigning a high probability to low values of γ_i, which tends to shrink the irrelevant components of $\mathbf{x}_k$ to zero, leading to a sparsifying behavior know as Automatic Relevance Determination (ARD) [Neal, 1996, MacKay, 1992].

To obtain robust hyperparameter estimates in an online fashion, we use a block of consecutive measurements $\mathbf{Y} = [\mathbf{y}_k, \ldots, \mathbf{y}_{k+N}]$ as opposed to a single data point $\mathbf{y}_k$. Here we assume that the parameters characterizing the generative model vary at a slower rate compared to the brain and artifact dynamics producing the measurements. In other words, sources producing a given data block $\mathbf{Y}$ are not expected to go silent from one sample to the next. This assumption can be further motivated by the fact that the EEG can be temporally segmented into a sequence of discrete quasi-stable microstates, each of which consisting of a scalp configuration lasting for approximately 80 to 120 milliseconds before transitioning to a different microstate [Khanna et al., 2015, Van De Ville et al., 2010, Koenig et al., 2002]. With this in mind, we ignore short-term correlations within a $\mathbf{Y}$ block (iid assumption) and approximate the evidence of the ensemble as,

$$p(\mathbf{Y}|\mathcal{M}) \approx \prod_{k=1}^{N} p(\mathbf{y}_k|\mathcal{M}) \tag{19}$$

where N is set so that we deal with consecutive blocks of approximately 40 milliseconds in a way that we don't miss microstates.

The maximization of the model block evidence is equivalent to minimizing the so-called type-II Maximum Likelihood (ML-II) cost function [Barber, 2012], which is obtained by applying $-2\log(\cdot)$ to Eq (19),

$$\mathcal{L}(\mathcal{M}) = \underbrace{\log|\mathbf{\Sigma}_y|}_{\text{Complexity}} + \underbrace{\text{trace}\big(\mathbf{C}_y\mathbf{\Sigma}_y^{-1}\big)}_{\text{Accuracy}} \tag{20}$$

where $\mathbf{C}_y = N^{-1}\mathbf{Y}\mathbf{Y}^T$ is the empirical data covariance. Note that the data should be zero-mean, which can be achieved with a high-pass filtering stage during preprocessing. Eq (20) embodies a tradeoff between model complexity and accuracy. Geometrically, the complexity term represents the volume of an ellipsoid defined by Σ_y. In particular, as the axes of the ellipsoid shrink due to the pruning of irrelevant sources, the volume is reduced. The second term measures model accuracy; i.e., how similar are the empirical and analytic covariances $\mathbf{C}_y$ and $\mathbf{\Sigma}_y$. We update the model on every block by solving the following optimization problem

$$\hat{\lambda}, \hat{\boldsymbol{\gamma}} = \arg\max_{\lambda,\boldsymbol{\gamma}} \mathcal{L}(\mathcal{M}) \tag{21}$$

Eq (21) can be optimized very efficiently following a two-stage approach proposed by Ojeda et al. [2018]. In the first stage we learn a coarse-grained non-sparse model by solving the constrained optimization problem:

$$\hat{\lambda}, \hat{\gamma}_F = \arg\max_{\lambda,\gamma_F} \mathcal{L}(\mathcal{M}) \quad \text{subject to } \gamma_i = \gamma_F > 0, \quad \forall i = 1, \ldots, N_\gamma \tag{22}$$

In the second stage we fix λ to the value $\hat{\lambda}$ and, starting from the value $\gamma_i = \hat{\gamma}_F$, we learn the sparse model by solving the optimization problem:

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \mathcal{L}(\hat{\lambda}, \boldsymbol{\gamma}), \quad \text{subject to } \boldsymbol{\gamma} \succeq 0 \tag{23}$$

An intuitive explanation for why the optimization of the ML-II cost function in Eq (20) yields sparse sources can be found in [Tipping and Faul, 2003]. We point the reader interested in the details of the two-stage algorithm outlined in this section to [Ojeda et al., 2018] while the MATLAB code and examples can be freely downloaded from the *Distributed Source Imaging (DSI)* toolbox repository[2].

2.3 Connections between PEB+ and ICA

In the analysis presented above, the matrix $\mathbf{H}$ is prespecified. In this section, we analyze the generative model of Eq. (3) from the ICA viewpoint. ICA is a blind source separation method that seeks to estimate the source time series (often called activations in the ICA literature) $\mathbf{x}_k$ from the data time series $\mathbf{y}_k$ *without* knowing the gain (mixing) matrix $\mathbf{H}$. In ICA, we assume that the latent sources are instantaneously independent, which yields the following prior distribution

$$p(\mathbf{x}_k) = \prod_{i=1}^{N_x} p_i(x_{i,k}) \tag{24}$$

To simplify the exposition, we assume the same number of sensors and sources, $N_y = N_x$, and the interested reader can find the case $N_y < N_x$ in [Le et al., 2011, Lewicki and Sejnowski, 1998]. From these premises, the objective of the algorithm is to learn the unmixing matrix $\mathbf{H}^{-1}$ such that we can estimate the sources with $\hat{\mathbf{x}}_k = \hat{\mathbf{H}}^{-1}\mathbf{y}_k$. The unmixing matrix $\hat{\mathbf{H}}^{-1}$ can be learned up to a permutation and rescaling factor, which has the inconvenience that the order of the learned components can change depending on the starting point of the algorithm and data quality. We can use a data block $\mathbf{Y}$ to write the likelihood function

$$p(\mathbf{Y}|\mathbf{H}, \lambda) = \prod_{k=1} p(\mathbf{y}_k|\mathbf{H}, \lambda) \tag{25}$$

under the assumption of independent data collection. However, we should point out that in ICA, $\mathbf{Y}$ is usually a data block longer than the one considered in Section 2.2, thus the iid data assumption is harder to

[2] https://github.com/aojeda/dsi

justify. To alleviate this situation the data are usually whitened during preprocessing. We can obtain each factor in Eq (25) by integrating out the sources as follows

$$p(\mathbf{y}_k|\mathbf{H},\lambda) = \int p(\mathbf{y}_k|\mathbf{x}_k,\mathbf{H},\lambda)p(\mathbf{x}_k)d\mathbf{x}_k \tag{26}$$

As noted by MacKay [2008b], assuming that the data are collected in the noiseless limit, $\lambda \to 0$, transforms the Gaussian likelihood $p(\mathbf{y}_k|\mathbf{x}_k,\mathbf{H},\lambda)$ into a Dirac delta function, in which case Eq (26) leads to the Infomax algorithm of Bell and Sejnowski [1995]. The learning algorithm essentially consists in finding the gradient of the log likelihood, $\log p(\mathbf{Y}|\mathbf{H},\lambda)$, with respect to $\mathbf{H}$ and updating $\mathbf{H}$ on every iteration such that the probability of the data increases. As pointed out by Comon [1994], the ICA model is uniquely identifiable only if at most one component of $\mathbf{x_k}$ is Gaussian. Therefore, the prior densities $p_i(x_{i,k})$ are usually assumed to exhibit heavier tails than the Gaussian and, in particular, the prior $p_i(x_{i,k}) \propto \cosh^{-1} x_{i,k}$ yields the popular ICA contrast function $\tanh(\mathbf{H}^{-1}\mathbf{y}_k)$. Note that this prior is not motivated by a biological consideration but by a mathematical necessity.

It is remarkable that ICA can learn columns of $\hat{\mathbf{H}}$ that are consistent with bipolar (single or bilaterally symmetric) cortical current source scalp projections without using any anatomical or biophysical constraint whatsoever [Makeig et al., 1997]. Onton et al. [2006] have shown that other columns may correspond to different stereotypical artifact scalp projections as well as a set of residual scalp maps that are difficult to explain from a biological standpoint. Delorme et al. [2012] have shown that the best ICA algorithms can identify approximately 30% of dipolar brain components (approximately 21 brain components out of 71 possible in a 71-channel montage). Although ICA has proven to be a useful technique for the study of brain dynamics [Makeig and Onton, 2011], we must wonder if its performance can be improved, perhaps by making BSS of EEG data less "blind". In other words, if we know a priori what kind of source activity we are looking for (dipolar cortical activity, EOG and EMG artifacts and so on), why limit ourselves to a purely blind decomposition?

In this paper, we advocate the use of as much information as we can to help solve the ill-posed inverse problem. In that sense, the use of a prespecified lead field matrix in the generative model of the EEG forces inverse algorithms to explain the data in terms of dipolar sources, because the lead field is precisely an overcomplete dictionary of dipolar projections of every possible source there is in a discretized model of the cortex. It has been shown that source estimation can greatly benefit from the use of geometrically realistic subject-specific [Cuspineda et al., 2009] or, alternatively, population-based approximated lead fields matrices [Valdés-Hernández et al., 2009]. Furthermore, augmenting the lead field dictionary with a set of stereotypical artifact projections, as proposed in Section 2.1, furnishes a more realistic generative model of the EEG in a way that renders blind decomposition unnecessary or at least suboptimal for brain imaging.

2.3.1 Independent components through PEB+

Since the source activity measured in the EEG is mixed by the volume conduction effect, ideally, we would like the PEB+ framework to exhibit the ICA property of yielding maximally independent (demixed) source time series. In this section we show that this is indeed the case. We start by rewriting the biologically motivated source prior of Eq (5) as

$$p(\mathbf{x}_k|\gamma) = \prod_{i=1}^{N_\gamma} p_i(\mathbf{x}_{i,k}|\gamma_i) \tag{27}$$

where each factor is a Gaussian pdf and i indexes a group of sources or an artifact component. To write Eq (27) as the ICA prior in Eq (24) we need to integrate out the hyperparameter γ_i from each factor as follows:

$$p(\mathbf{x}_k) = \prod_{i=1}^{N_\gamma} \underbrace{\int p_i(\mathbf{x}_{i,k}|\gamma_i)p(\gamma_i)d\gamma_i}_{p_i(\mathbf{x}_{i,k}) \text{ is a Student t-distribution}} \tag{28}$$

which, given our choice of hyperprior on γ_i, renders each marginalized prior $p_i(\mathbf{x}_{i,k})$ a heavy-tailed Student t-distribution [Tipping, 2001]. We note that in our development we take the route of optimizing the γ_i hyperparameters rather than integrating them out because the former approach yields a simpler algorithm and tends to produce more accurate results in ill-posed inverse problems [MacKay, 1996]. Moreover, the optimization of γ_i allows for automatic removal of irrelevant brain and artifact components that are not supported by the data, thereby eliminating the subjectivity implicit in manual component selection. Assuming the prior in Eq (27), the ICA data likelihood of Eq (25) becomes exactly the evidence of Eq (19), with the difference that in the PEB+ algorithm the $\mathbf{H}$ matrix is known and the evidence is optimized on small blocks of data, which gives our algorithm the ability to run in an online manner and to capture transient brain dynamics.

We summarize the advantages of using the PEB+ framework over ICA for source separation and imaging of EEG data as follows:

- It deals gracefully with the overcomplete case ($N_y \ll N_x$) by finding the MAP source estimator, which always exist even in the presence of rank-deficient data, e.g. after removing the common average reference.

- It deals with the redundancy in brain responses by inducing independence over groups of sources.
- The use of the ARD prior allows for the automatic selection of components in a data-driven manner, thereby eliminating the subjectivity of selecting components based on practitioner's experience.
- It can adapt to non-stationary dynamics by updating the model on smaller blocks of data.
- It can be used in online applications by leveraging fast evidence optimization algorithms.
- Artifact removal, source separation, and imaging can be obtained simultaneously as a consequence of optimizing the evidence of a biologically informed generative model.
- It facilitates subject-level analysis because we estimate the same number of cortical source activations per subject, each of which has known anatomical support. This eliminates the complications of clustering ICs and dealing with missing components [Bigdely-Shamlo et al., 2013b] while allowing the use of more straightforward and widespread statistical parametric mapping techniques [Penny et al., 2007].

Results

The PEB framework has been validated extensively on simulated and real data elsewhere [Ojeda et al., 2018, Zhang and Rao, 2013, Henson et al., 2011, Friston et al., 2008]. In this section, we study the effects of modeling artifactual sources in the quality of the model inversion. We also discuss source separation through PEB+, as well as its use for data cleaning. Then in Section 3.5 we show an application of PEB+ to the study of heading computation during full-body rotations in the context of a mobile brain/body imaging (MoBI) experiment.

3.1 Construction of artifact projection dictionary A

To characterize artifactual ICs we used data from two different studies made public under the umbrella of the BNCI Horizon 2020 project[3] [Brunner et al., 2015]. Since in the next two sections we investigate different features of the PEB+ algorithm rather than the biological interpretation of its results, we don't dwell into the details of the experimental paradigms used in each study, and we direct the interested reader to the respective publications referenced below.

[3] http://bnci-horizon-2020.eu/database/data-sets

3.1.1 Data set 1: Error related potentials

The first study, *013-2015*, provided EEG data from 6 subjects (2 independent sessions per subject and 10 blocks per session) collected by Chavarriaga et al. [2010] using an experimental protocol designed to study error potentials during a BCI task. EEG samples were acquired at a rate of 512 Hz using a Biosemi ActiveTwo system and a 64-channels montage placed according to the extended 10/20 system.

3.1.2 Data set 2: Covert shifts of attention

The second data set, *005-2015*, provided EEG and EOG data from 8 subjects collected by Treder et al. [2011] using an experimental protocol designed to study the EEG correlates of shifts in attention. The EEG was recorded using a Brain Products actiCAP system, digitized at a sampling rate of 1000 Hz. The montage employed had 64 channels placed according to the 10/10 system referenced to the nose. In addition, an EOG channel (labeled as EOGvu) was placed below the right eye. To measure vertical and horizontal eye movements, from the total of 64 EEG channels, two were converted into bipolar EOG channels by referencing Fp2 against EOGvu, and F10 against F9, thereby yielding a final montage of 62 EEG channels.

3.1.3 Data preprocessing and IC scalp maps clustering

After transforming each data file to the *.set* format, both studies were processed using the same pipeline written in MATLAB (R2017b The MathWorks, Inc., USA) using the EEGLAB toolbox [Delorme et al., 2011]. The pipeline consisted of a 0.5 Hz high-pass forward-backward FIR filter and re-referencing to the common average, followed by the Infomax ICA decomposition of the continuous data. We pooled all the preprocessed data sets and randomly assigned them to one of two groups: 80 % to the training set and 20 % to the test set. The training set was used to construct the artifact dictionary and the test set was used to evaluate the performance of the PEB+ algorithm.

To construct an artifact dictionary from a heterogeneous EEG database, we need to represent each independent scalp map into a common (co-registered) channel space. To that end, we used the coordinates of the common channels between the montages used in the two studies described above to estimate a linear transformation from the 62-channel space to the 64-channel one. After co-registration, we pooled both studies in the training set resulting in a matrix of 64 channels by 6774 independent scalp maps (101 sessions and blocks yielding 64 ICs each plus 5 sessions yielding 62 ICs each). It is worth emphasizing that we only warped IC scalp maps and not the actual data or IC activations.

Next, we used the matrix of co-registered independent scalp maps to estimate clusters using the k-means algorithm. Clusters were labeled as Brain, EOG, EMG, or Unknown (scalp maps of unknown origin) by an expert. Unknown clusters were not used further in this paper. Fig 2 shows a visualization of the IC scalp maps using the t-distributed stochastic neighbor embedding (t-sne) algorithm [Van Der Maaten and Hinton, 2008]. The t-sne algorithm allows us to represent each 64-dimensional IC scalp map as a dot in a 2D space in a way that similar and dissimilar scalp maps are modeled by nearby and distant points respectively with high probability. We ran the k-means algorithm for several numbers of clusters, and we stopped at 13 after noticing that many small islands scattered at the periphery of Fig 2 started to be either mislabeled as Brain or labeled consistently as EOG, EMG or Unknown. The grey points in the figure denote most of the scalp maps labeled as non-brain.

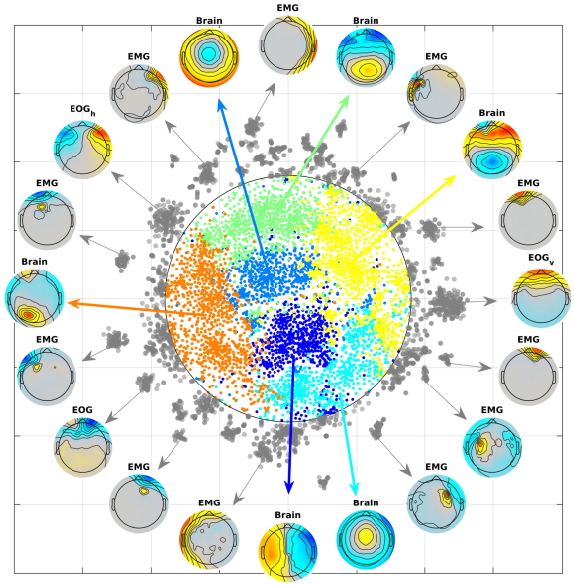

Figure 2: t-sne visualization of IC scalp map clusters. We used the t-sne algorithm to represent each 64-dimensional scalp map as a dot in a 2D space in a way that similar and dissimilar scalp maps are modeled by nearby and distant points respectively with high probability. The clusters were estimated using the k-means algorithm. The grey points indicate mostly non-brain or mislabeled scalp projections.

Using the insights from Fig. 2, we completed the augmented PGM of Eq (3) by building the $\mathbf{A}$ dictionary as follows:

$$\mathbf{A} = \left[\mathbf{a}_{EOG_v}, \mathbf{a}_{EOG_h}, \mathbf{a}_{EMG_1}, \ldots, \mathbf{a}_{EMG_{11}}, \mathbf{I}_{N_y}\right] \tag{29}$$

where $\mathbf{a}_{EOG_v}$ and $\mathbf{a}_{EOG_h}$ are the centroids of the vertical and horizontal EOG clusters respectively, $\mathbf{a}_{EMG_i}$ are the centroids of EMG clusters and we modeled spike artifacts affecting each individual channel with the columns of the identity matrix $\mathbf{I}_{N_y}$, with $N_y = 64$.

3.1.4 Calculation of subject-specific H matrices

Depending on the montage of each subject, we nonlinearly warped their 62 or 64-channel montage to the scalp surface of a four-layer (scalp, outer skull, inner skull, and cortex) "Collin27" template using the DSI toolbox. Then we computed the orientation-fixed lead field matrices $\mathbf{L}$ using the boundary element method solver in the OpenMEEG toolbox [Gramfort et al., 2010]. Next, we calculated individual $\mathbf{A}$ matrices by linearly warping its columns from the 64-channel space to the space defined by the head surface of the template. Finally, we divided each column of the augmented dictionary $\mathbf{H}$ by its norm so that their relative contribution to the scalp EEG could be determined by the amplitude of the source activation vector $\mathbf{x}_k$.

3.2 PEB+ artifact model validation

In this section, we investigate whether the approach of explicitly modeling artifact scalp projections used in PEB+ yields significantly better source estimates than the traditional PEB. To quantify the support in the data for a given model we used the Bayes factor. In Bayesian model selection/comparison, the Bayes factor is used as an alternative to classical hypothesis testing, replacing the p-value as a measure of evidential strength while avoiding the abuse to which the latter is often subjected to these days [Stern, 2016]. The Bayes factor between two generative models $\mathcal{M}_i$ and $\mathcal{M}_j$ is denoted as $B_{i,j}$ and is defined as the ratio between the evidence of each model:

$$B_{i,j} = \frac{p(\mathbf{Y}|\mathcal{M}_i)}{p(\mathbf{Y}|\mathcal{M}_j)} \tag{30}$$

Interpreting $B_{i,j}$ is straightforward, for instance $B_{i,j} > 1$ indicates that there is more evidence in favor of generative model $\mathcal{M}_i$ over $\mathcal{M}_j$. Likewise, $B_{i,j} \approx 1$ means that there is no conclusive evidence in favor of any of the models considered. Usually, a Bayes factor in $2\log_e$ units higher than 2, 6, and 10 is respectively considered positive, strong and very strong evidence in favor of $\mathcal{M}_i$ [Kass and Raftery, 1995]. Conversely, note that Eq (30) can be interpreted as evidence in favor of $\mathcal{M}_j$ simply by flipping the ratio.

As we explained earlier, in this paper we characterize artifacts empirically rather than in a mechanistic principled way. Therefore, next we used Bayes factors to assess the performance of the PEB+ algorithm under different variants of the artifact dictionary $\mathbf{A}$. In our analysis we considered the following models:

- $\mathcal{M}_0$: ignoring artifacts, $\mathbf{A} \in \emptyset$ (classic PEB)

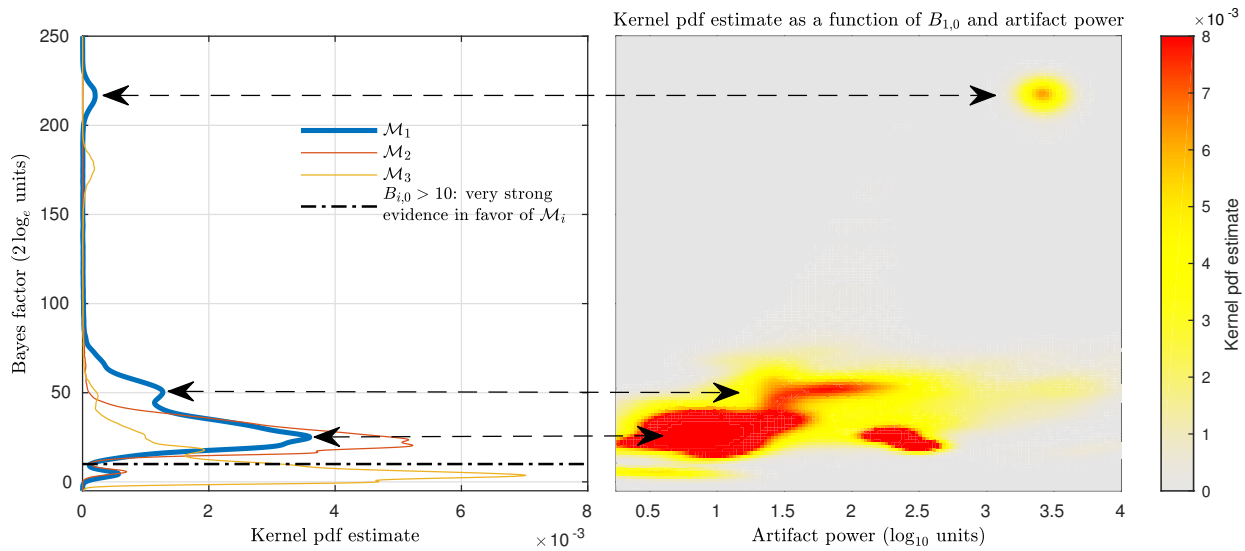

Figure 3: Comparison of generative models with different variants of the artifact dictionary with respect to $\mathcal{M}_0$. **Left:** Bayes factor kernel pdf estimate for each artifact model. **Right:** $B_{1,0}$ kernel pdf estimate as a function of artifact power.

- $\mathcal{M}_1$: modeling artifacts with Eq (29),
- $\mathcal{M}_2$: modeling EOG and EMG components only, $\mathbf{A} = [\mathbf{a}_{EOG_v}, \mathbf{a}_{EOG_h} \mathbf{a}_{EMG_1}, \ldots, \mathbf{a}_{EMG_{11}}]$,
- $\mathcal{M}_3$: modeling single channel spikes only, $\mathbf{A} = \mathbf{I}_{N_y}$.

To that end, we estimated the brain and artifact source time series of all subjects in the testing set under each model. We collected the log evidence afforded by each model in blocks of 40 msec of data and computed Bayes factors with respect to $\mathcal{M}_0$.

The left panel of Fig 3 shows the kernel pdf estimate of each Bayes factor: $B_{1,0}$, $B_{2,0}$, $B_{3,0}$ in blue, orange, and yellow traces respectively. The area above the dot-dashed black trace represents the probability space in which modeling artifacts yielded a better generative model for the EEG signal. We computed the probability of having very strong evidence in favor of model $\mathcal{M}_i$ by integrating over Bayes factors higher than 10, $P_i(10 \leq B_{i,0}) = \int_{B_{i,0} \geq 10} f_i(B_{i,0})$, where $f_i(B_{i,0})$ denotes a pdf as a function of model $\mathcal{M}_i$. As shown in Table 1, models $\mathcal{M}_1$ and $\mathcal{M}_2$ yielded the highest probability. Since $\mathcal{M}_1$ includes $\mathcal{M}_2$ and $\mathcal{M}_3$, henceforth we use the artifact dictionary given by Eq (29).

$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
0.9704	0.9705	0.4202

Table 1: Artifact model comparison. Probability that there is very strong support in the data in favor of generative model $\mathcal{M}_i$ over $\mathcal{M}_0$, $P_i(10 \leq B_{i,0})$.

To further illustrate the importance of modeling artifact components, the right panel of Fig 3 shows the kernel pdf estimate of $B_{1,0}$ as a function of the artifact power. The artifact power was calculated as the maximum RMS power over the artifact sources for each 40 msec window. We note that the shape of its pdf depicted on the left panel seems to be determined by the higher performance of the PEB+ algorithm under different clusters of artifactual activity expanding several orders of magnitude.

Fig 4 shows an example of applying PEB+ to an epoch of data of 4 seconds centered around an eye blink event. Panel **A** shows a 32-channel subset of the raw and reconstructed (cleaned) EEG traces in black and red respectively. In addition to an eye blink, we have a lateral eye movement event around the -1 sec latency. Panel **B** shows the estimated EOG_v and EOG_h artifact source activity in blue and orange respectively. We note that these artifact sources are active only at the latencies where the EEG is affected and mostly zero elsewhere. In panel **C**, the first and last two columns represent the raw and cleaned EEG topographies at the maximum of the lateral eye movement and eye blink events respectively. Panel **D** shows different views of the estimated cortical source maps underlying the raw topographies in **C**. The cleaned topographies in **C** are obtained after the estimated artifact sources are projected out of the data. Panel **E** shows the log evidence for generative models $\mathcal{M}_1$ and $\mathcal{M}_0$ in blue and orange respectively. Note that both traces differ mostly only when artifacts occur and higher log evidence in favor of model $\mathcal{M}_1$ indicates that source estimation benefits from modeling artifacts.

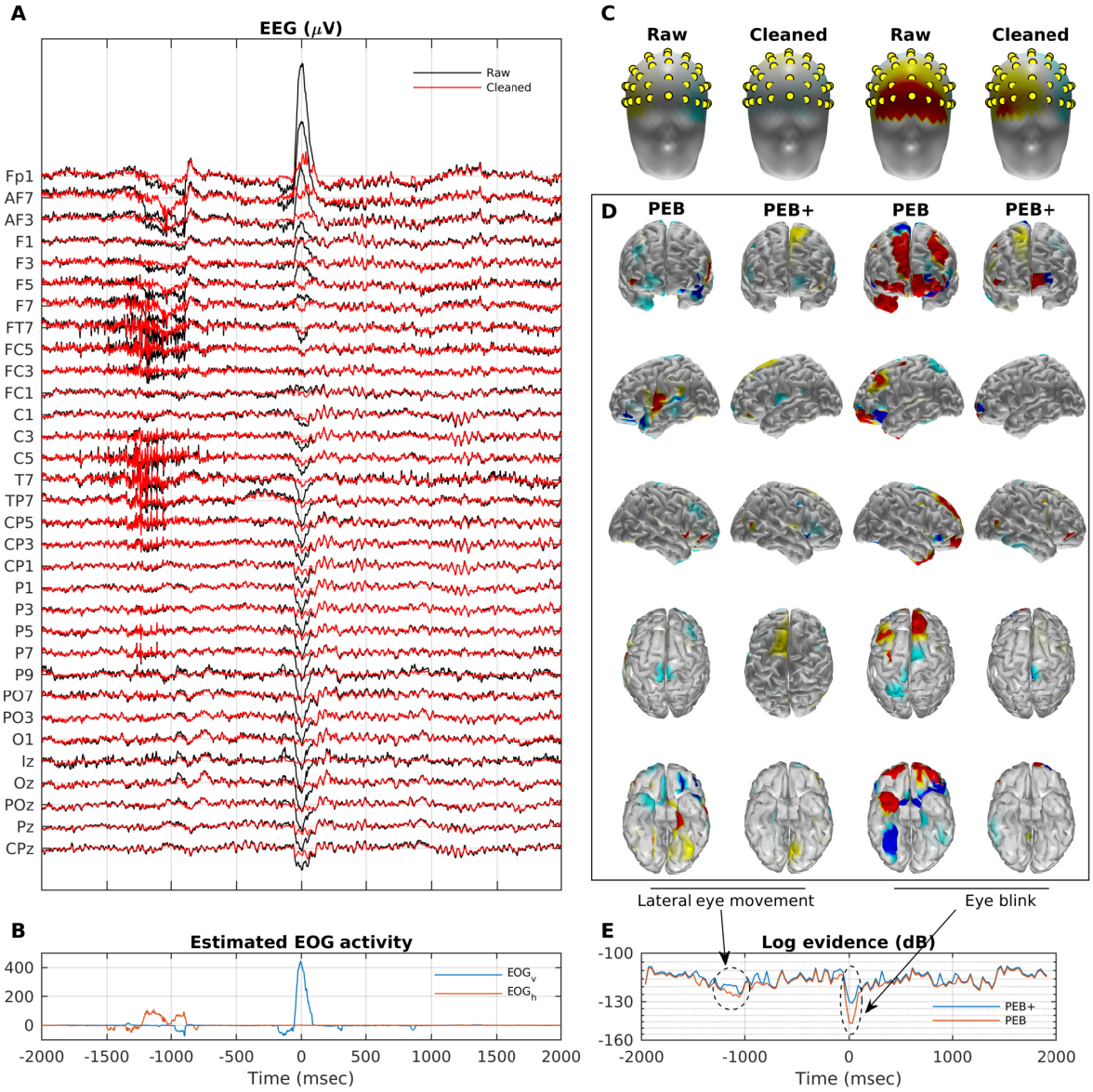

Figure 4: Example of applying PEB+ to an epoch of EEG with lateral eye movement and eye blink artifacts. **A**: 32-channel subset of raw and cleaned EEG traces. **B**: Estimated EOG_v and EOG_h artifact source activity. **C**: Columns 1 and 3 and 2 and 4 represent the raw and cleaned EEG topographies at the maximum of the lateral eye movement and eye blink events respectively. **D**: Different views of the estimated cortical source maps underlying the raw topographies in **C**. **E**: Log evidence yielded by PEB+ and PEB algorithms on consecutive 40 msec blocks of data along this epoch.

It is worth noting that, in the last column of panel **D**, some residual eye blink artifact seems to be mistakenly represented as a small activation in the frontal pole. We point out that, in practice, it may be extremely hard to totally remove artifactual activity because: 1) the use of a lead field matrix derived from a template head model may misfit the anatomy of the subject introducing errors in the **L** dictionary, 2) errors in the sensor locations can cause the EEG topography to shift with respect to the expected brain and artifact source projections, 3) EMG scalp projections are difficult to characterize due to their variability, as opposed to EOG projections that are more stereotyped, and 4) unmodeled muscle projections, such as those towards the back of the head that were largely ignored in this study. Despite all these issues, Figures 3 and 4 demonstrate that PEB+ can yield reasonably robust source estimates in the presence of artifacts. Furthermore,

panel **E** of Fig 4 suggests that we could use dips in the log evidence to inform subsequent processing stages of artifactual events that were not successfully dealt with.

3.3 Source separation performance

In this section, we investigate the source separation performance of the PEB+ algorithm. To that end, we used the test set to compare PEB+ and Infomax ICA regarding 1) volume conduction unmixing performance and 2) data size requirements for good source separation. We assessed the unmixing performance by calculating the mutual information reduction (MIR) achieved by each algorithm on data blocks of different sizes.

The MIR is an information theoretic metric that measures the total reduction in information shared between the components of two sets of multivariate time series. The mutual information (MI) between two given time series $x_{i,k}$ and $x_{j,k}$, $I(x_i, x_j)$, can be defined as the Kullback-Leibler (KL) divergence between their joint and marginal distributions:

$$I(x_i, x_j) = D_{KL}[p(x_i, x_j) \parallel p(x_i), p(x_j)] \tag{31}$$

where $I(x_i, x_j) > 0$ indicates that processes x_i and x_j share information while $I(x_i, x_j) = 0$ indicates that they are statistically independent such that

$$p(x_i, x_j) = p(x_i)p(x_j)\cancelto{0}{p(x_i|x_j)}$$

We define the MIR of source separation algorithm A with respect to B, as the difference in normalized total pairwise MI (PMI) achieved by each decomposition:

$$\begin{aligned} MIR_{A,B} = & \frac{2}{N_A(N_A-1)} \sum_{i=2}^{N_A} \sum_{j=1}^{i-1} I(x_i^A, x_j^A) - \\ & \frac{2}{N_B(N_B-1)} \sum_{i=2}^{N_B} \sum_{j=1}^{i-1} I(x_i^B, x_j^B) \end{aligned} \tag{32}$$

where x_i^A and x_i^B are the set of components yielded by each method and N_A and N_B are the number of components afforded by each decomposition. We note that to obtain a PMI that is not biased by the number of components, we normalize each summation by the number of unique (i, j) pairs. Here we calculated the MI using the non-parametric kernel pdf estimates of the quantities in Eq (31) for the the multichannel EEG data, the ROI-collapsed sources estimated by PEB+, and the ICs obtained by Infomax.

In Fig 5, the left panel shows a box plot of the MIR of PEB+ and Infomax calculated with respect to the MI of channel data. As indicated by the x-axis, we ran the experiment multiple times varying the data sizes from 0.5 to 500 seconds ($\sim$ 8 minutes). As expected, both algorithms reduce source MI, thereby reversing to some extent the mixing effect of the volume conduction. We see also that, on average, when the MIR is calculated in short blocks of data, PEB+ exhibits higher unmixing performance while Infomax seems to do better on longer blocks. This effect is more clearly represented in the panel on the right, which shows the box plot of the MIR of PEB+ with respect to Infomax. In that panel, distributions with entire

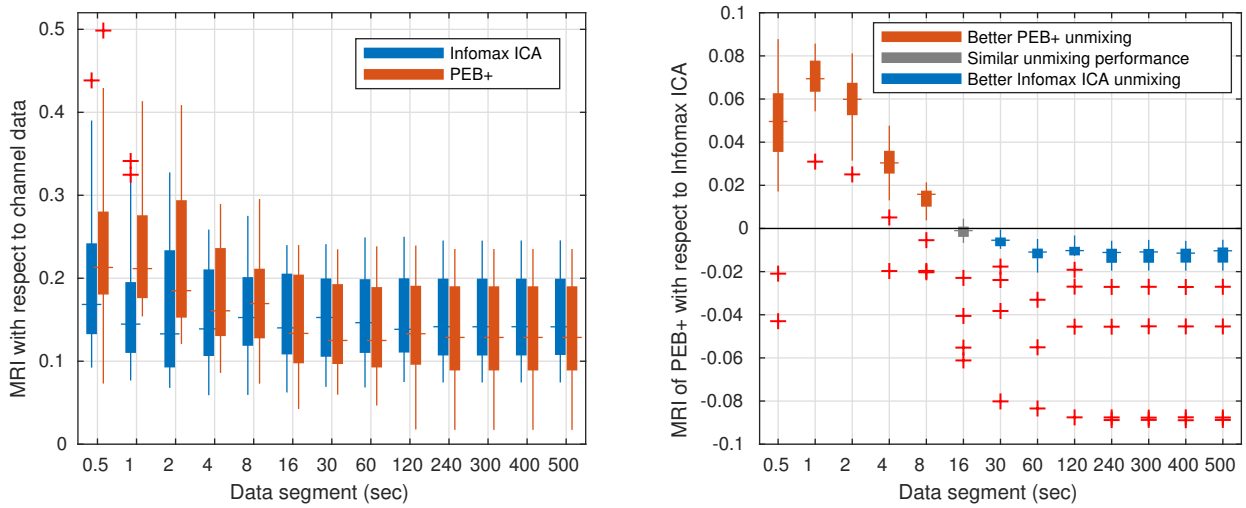

Figure 5: Source separation performance. **Left:** Box plot of MIR with respect to channel data computed on blocks of various sizes. **Right:** Box plot of MIR of PEB+ with respect to Infomax ICA for the same data blocks shown on the left. On each box, the central mark indicates the median, and the bottom and top edges indicate the 25th and 75th percentiles respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the + symbol. On the right, the distributions with entire positive (orange) or negative (blue) values indicate a significant source crosstalk reduction in favor of the PEB+ or Infomax algorithms respectively.

positive (orange) or negative (blue) values indicate a significant source crosstalk reduction performance in favor of the PEB+ or Infomax algorithms respectively. We note that PEB+ better captures transient dynamics for short 0.5-8 sec data blocks.

It is worth noting that with PEB+, it is possible to update the unmixing matrix (given by the term $\mathbf{\Sigma_x H^T \Sigma_y^{-1}}$ in Eq (10)) on a time-scale of tens of milliseconds because of the regularization induced by the multiple constraints. This allows for adaptation to non-stationary brain and artifact source dynamics. Infomax (and most ICA algorithms) on the other hand, requires larger data blocks to learn a global factorization of mixing matrix and source activations of reasonable quality. Moreover, Fig 5 suggests that the estimation of a global ICA model is suitable for identifying components that remain stationary over the whole experiment, but otherwise, it is suboptimal for capturing transient dynamics.

3.4 Data cleaning performance

In this section, we benchmark the data cleaning performance of the PEB+ algorithm against ASR. The ASR algorithm has gained popularity in recent years for its ability to remove a variety of high amplitude artifacts in an unsupervised manner, thereby enabling automatic artifact rejection for offline as well as real-time EEG-based BCI applications. Since in real data we do not have a ground truth for artifactual activity, we benchmark the methods according to the correlation between raw and cleaned data samples in blocks with negligible or no artifactual activity, where low correlation values indicate needless distortion of the brain activity.

We ran both algorithms for each subject in the test set and collected the following quantities on subsequent blocks of 40 msec: 1) the correlation between raw and cleaned data (computed as the correlation between the correspondent data blocks vectorized across channels and time points) and 2) the maximum RMS artifact power yielded by PEB+ as described in Section 3.2. ASR's performance depends on multiple parameters, but it has been indicated that the most critical one is the cutoff [Chang et al., 2018]. In the first experiment we used a cutoff equal to 5, which was the default value of EEGLAB's ASR plugin at the time of this publication.

In Fig 6, the left and right panels show the empirical kernel pdf estimation of the correlation as a function of the artifact's power for the ASR and PEB+ algorithms respectively. We see that in both methods, the correlation decreases as artifact power increases. This effect is expected and desired because cleaning algorithms are supposed to modify contaminated raw data. Towards low power artifact regions, however, ASR exhibits a significant amount of probability mass that spreads down to low correlation values while PEB+ seems to have most of its probability mass bounded from below at around 0.8. This result indicates that, at a cutoff of 5, ASR cleaning is overly aggressive to the point of significantly modifying the data in the absence of artifacts. These findings are in agreement to what was recently reported by Chang et al. [2018].

In Section 3.2 we showed that PEB+ can reduce the effects of eye-related artifacts. In Fig 7 we show an example of its performance removing EMG. The data correspond to an excerpt of EEG extracted from a subject selected at random from the test set while he/she was performing the respective cognitive task. In the left panel, the gray and colored traces represent contaminated and cleaned EEG signals respectively. The traces shown correspond to channels located on each side of the cap. Channels in these areas are often contaminated by EMG activity due to their proximity to the temporalis muscle [Fu et al., 2006]. As we can see, the higher amplitude decorrelated EMG activity is largely reduced. In the panel on the right, the gray and colored traces represent the power spectral density estimates of their respective channels on the left. We see that a significant amount of broad band power related to the EMG activity was removed, especially towards frequencies higher than 18 Hz. We note that a strong 50 Hz AC line noise remains in the cleaned data, this is expected because in our approach we do not model this type of artifacts and they can be relatively easily removed using a notch filter.

To further illustrate the results of the analysis shown in Fig 6, in Fig 8 we compare the cleaning performance of PEB+ against ASR using two cutoff parameters (5 and 100) and ICA on a typical epoch of EEG containing lateral eye movement, eye blink, and muscle artifacts. As before, the EEG data were extracted from a subject selected at random from the test set. In panel **A**, the shaded area in blue indicates a segment of clean data, while green, red, and yellow areas indicate segments contaminated by lateral eye movement, muscle, and eye blink artifacts respectively. The correlation achieved by each method (computed by vectorizing all channels and samples in the blue segment) is displayed on the top left corner of each panel. All correlations were significant with p-values lower than 0.005. Panels **A**, **B**, **C**, and **D** show the raw and cleaned EEG traces produced by PEB+, ICA, ASR (100), and ASR (5) methods respectively. For the ICA approach we cleaned the data by removing the contribution of several stereotypical EOG and EMG components selected manually. PEB+ and ICA displayed similar performance in the sense that lateral eye movement and eye blink artifacts were largely removed, EMG was not totally removed by PEB+, while the clean data segment was minimally distorted, as indicated by correlations with the raw samples of 0.8922 and 0.8294 respectively. We note that the distortion introduced by ICA could be reduced by a more

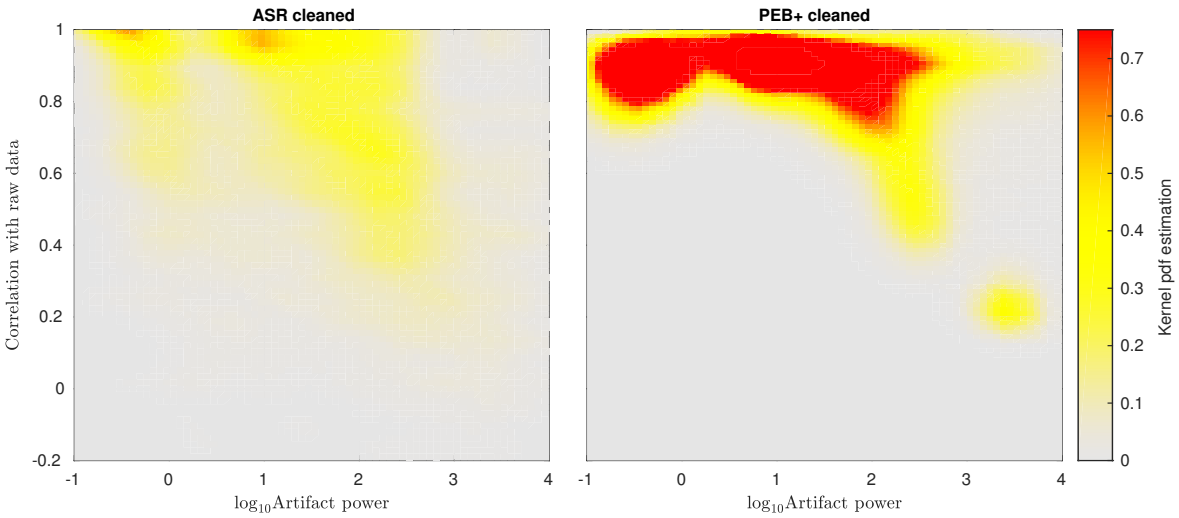

Figure 6: Data cleaning performance. Kernel pdf estimation of the correlation between raw and cleaned data as a function of artifact power. **Left:** Data cleaned by ASR using cutoff=5 (default). **Right:** Data cleaned by PEB+. Note that, as expected, in both algorithms the correlation drops as artifacts increase. Towards low amplitude artifacts, however, ASR significantly distorts the data while PEB+ does not.

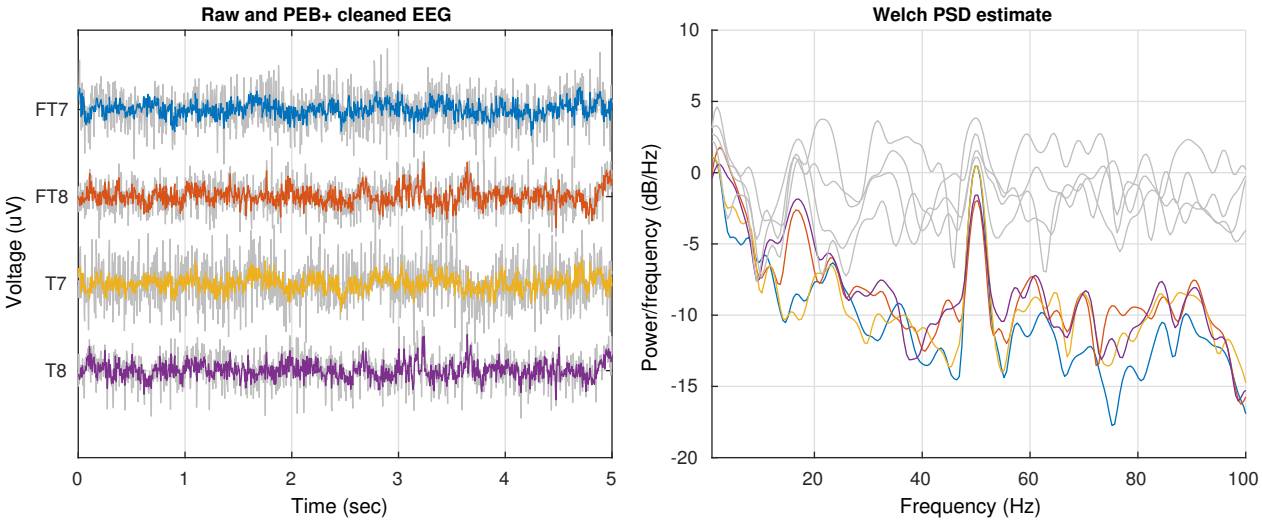

Figure 7: EMG artifact cleaning. EEG channels contaminated by EMG noise, the gray and colored traces represent raw and cleaned data respectively. **Left:** Excerpt of EEG signal. **Right:** Welch power spectral density estimation of the data shown on the left.

conservative selection of the artifactual ICs to remove. ASR (100) did not distort the clean data segment and removed the higher amplitude eye blink artifact but failed to remove lateral eye movement and muscle artifacts. ASR (5) removed all the artifactual activity, however it also significantly distorted the clean data segment as indicated by a correlation of 0.2468.

3.5 Heading computation during full-body rotations

We finalize the paper with an application of the PEB+ algorithm to MoBI data. MoBI experiments are notoriously difficult to analyze due to the amount of motion-induced artifacts as well as the presence of transient and stationary brain dynamics of variable duration across trials. Here, we try to replicate the main findings of a study that looked into the dynamics of the retrosplenial cortex (RSC) supporting heading computation during full-body rotations [Gramann et al., 2018].

Heading computation is key for successful spatial orientation of humans and other animals. The registration of ongoing changes in the environment, perceived through an egocentric first-person perspective has

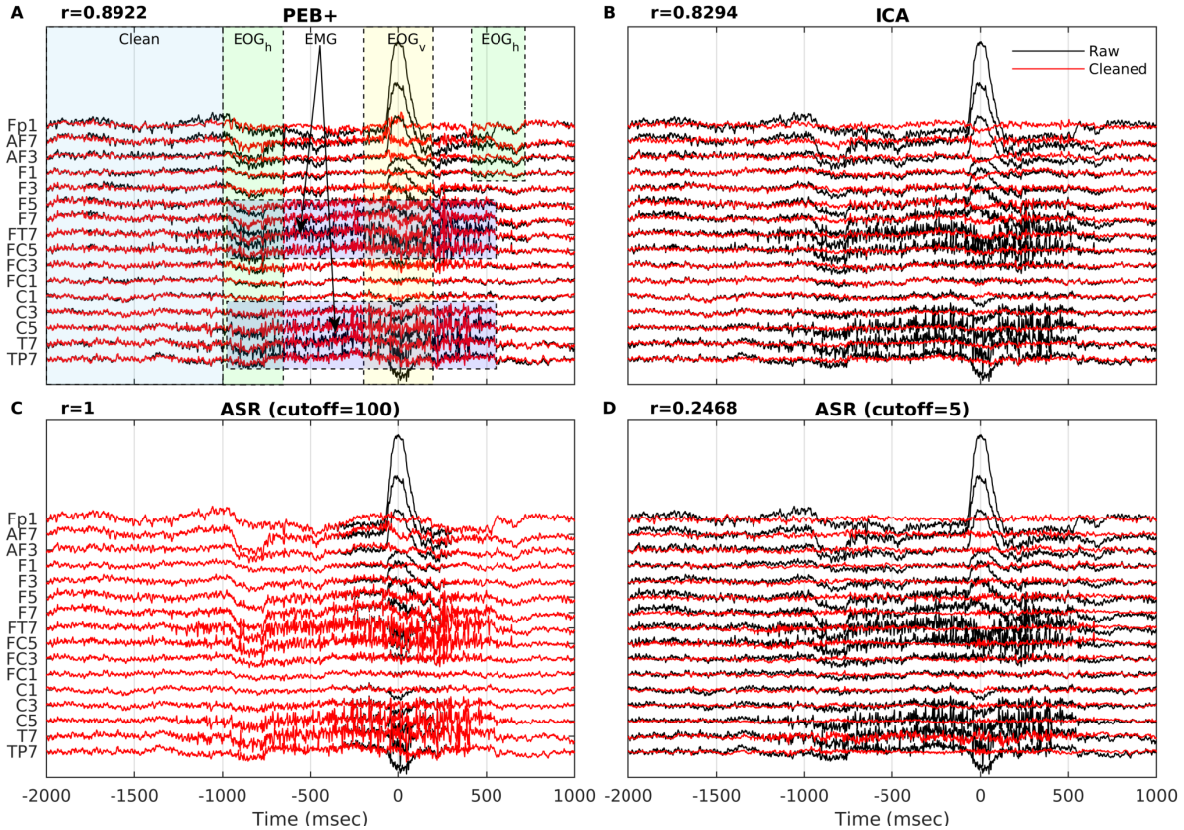

Figure 8: Example of the data cleaning performance of PEB+, ICA, and ASR on a noisy epoch. Only 16 channels are shown. **A**: PEB+. **B**: ICA. **C**: ASR with cutoff=100. **D**: ASR with cutoff=5 (current default value in EEGLAB's *clean_rawdata* plugin). In panel **A**, the segment of data used to compute the correlations between raw and cleaned EEG is indicated with the blue shaded area (-2000 msec to -1000 msec) while the areas shaded in green, yellow and red indicate lateral eye movement, eye blink, and muscle artifacts respectively.

to be integrated with allocentric, viewer-independent spatial information to allow complex navigation behaviors. The RSC provides the neural mechanisms to integrate egocentric and allocentric spatial information by providing an allocentric reference direction that contains the subject's current heading relative to the environment [Byrne et al., 2007]. Single-cell recordings in freely behaving animals have shown that the RSC is also implicated in heading computation [Sharp et al., 2001]. And although there is fMRI evidence that points to the same conclusion in humans that navigate in a virtual environment [Baumann and Mattingley, 2010], verifying this hypothesis in more naturalistic settings has remained elusive.

Recently, Gramann et al. [2018] used EEG synchronized to motion capture recordings combined with virtual reality (VR) to investigate the role of the RSC in heading computation of actively moving humans. Data were recorded from 19 participants using 157 active electrodes sampled at 1000 Hz and band-pass filtered from 0.016 Hz to 500 Hz using a BrainAmp Move System (Brain Products, Gilching, Germany). 129 electrodes were placed equidistant on the scalp and 28 were placed around the neck using a custom neckband. In that study, data from physically rotating participants were contrasted with rotations based on visual flow. In the physical rotation condition, participants wore a Vive HTC head-mounted display (HTC Vive; $2 \times 1080 \times 1200$ resolution, 90 Hz refresh rate, 110° field of view). They were placed in a sparse VR environment devoid of any landmark information facing an orienting beacon at the beginning of each trial. The beacon was then replaced by a sphere that started rotating around them to the left or the right at a fixed distance with two different, randomly selected, velocity profiles on each trial. Participants were instructed to rotate on the spot to follow the sphere and keep it in the center of their visual field. The sphere movement was completed at an eccentricity randomly selected between 30° and 150° relative to the initial heading. When the sphere stopped, they had to rotate back and press a controller button to indicate when they believed to have reached their initial heading

orientation. After the button press, the beacon would reappear and participants had to rotate to face the beacon and to start the next trial. In the joystick rotation condition, participants stood in front of a large TV screen (1.5 m viewing distance, HD resolution, 60 Hz refresh rate, 40″ diagonal size) controlling a gaming joystick to rotate in the same VR environment with an otherwise identical trial structure.

Using an ICA/dipole fitting approach, the data was analyzed with a focus on oscillatory activity of ICs located in or near the RSC. ICs were clustered using repetitive k-means clustering optimized to the RSC as the region of interest. Four subjects without an IC in the RSC were excluded from the analysis (21% of all participants). Subsequently, the wavelet (Morlet) time-frequency decomposition was computed for each IC in the RSC cluster for the rotation periods. The spectral baseline was defined as the 200 msec period before stimulus onset and subtracted from each time-frequency decomposition. To account for different trial durations, single trial time-frequency maps were linearly time-warped with respect to the presentation of the stimulus and rotation onset and offset to create time-warped event-related spectral perturbations (ERSPs). Using this approach, the data from the RSC cluster in the joystick rotation condition replicated previous studies using desktop navigation protocols and comparable data analysis approaches [Gramann et al., 2010, Chiu et al., 2012, Lin et al., 2015, 2018], exhibiting 1) a theta burst between stimulus onset and movement onset and 2) alpha and beta desynchronization during the rotation. The physical rotation, however, had drastically different properties: no clear theta burst was present before movement onset, and only minor desynchronization in higher beta bands, but synchronization in the alpha and low beta bands after movement onset and delta and theta bands during the rotation (see Fig 9 **A-B**).

Here, we used the PEB+ algorithm to re-analyze the data. To this end, we further down-sampled the data to 250 Hz, removed the neck channels, applied a 0.5 Hz high-pass forward and backward FIR filter, and subtracted the common average reference. We co-registered each subject-specific 129-channels montage with the head surface of the "Colin27" template, computed each lead field matrix, and linearly warped the **A** dictionary to the space of the individualized template as explained in Section 3.1.4. Then we ran the PEB+ algorithm for each condition and computed the ERSPs of the centroid source activity (see Eq (15)) within the RSC. The computation of ERSPs was identical to the previous one, only the IC activity of the RSC cluster was replaced by PEB+ RSC source activity of all subjects.

Fig 9 **C-D** shows the PEB+ group ERSP for the joystick and physical rotation conditions as well as their difference. The top panel shows in red the location of the RSC in our template brain. Despite the differences between the two methodologies, our results largely replicate those in [Gramann et al., 2018] displayed in panels **A-B**. A few differences between the two results are worth mentioning though. We point out that the differences in ERSP scales exhibited in panels **B** and **D** may be explained by different scales of the sources obtained by ICA and PEB+. Also, we note that the low-frequency power increase towards the end of the head rotation cycle in panel **D** Joystick condition can be explained by artifacts improperly removed near the end of a few trials. It should be emphasized that, unlike the approach used by Gramann et al. [2018], ours has the advantage of using data from all subjects without any cleaning in the time, channel, or trial domains, except for the inherent cleaning capabilities of the PEB+ algorithm. To increase the robustness to residual artifacts, Fig 4 **E** suggests that a future research direction could explore the use of the log evidence yield by PEB+ to automatically downplay the influence of artifactual trials into post hoc statistical summaries.

Conclusions

In this paper, we have extended the Parametric Empirical Bayes (PEB) framework previously proposed for electrophysiological source imaging [Henson et al., 2011, Wipf and Nagarajan, 2009] in two ways. First, we augmented the standard generative model of the EEG with a dictionary of artifact scalp projections obtained empirically. In our model, we captured EOG, EMG, and single-channel spike artifacts. Second, we used an anatomical atlas to parametrize a source prior that encourages sparsity in the number of active cortical regions, which has the desired property of inducing the segregation of the cortical electrical activity into a few maximally independent components with known anatomical support. We used these elements to develop the PEB+ inversion algorithm. Under the proposed framework, dissimilar problems such as data cleaning, source separation, and imaging can be understood and solved in a principled manner using a single algorithm. Furthermore, we used our framework to point out the connections between distributed source imaging and Independent Component Analysis (ICA), two of the most popular approaches for EEG analysis that are often perceived to be at odds with one another.

We used publicly available data from two independent studies to develop and test the proposed algorithm. In particular, we have shown that PEB+: 1) outperforms classic PEB for source imaging when artifacts are present in the data, while on clean data their performance is comparable, 2) outperforms Infomax ICA for source separation on short blocks of data, thereby showing potential for tracking non-stationary cortical dynamics, and 3) unlike the popular Artifact Subspace

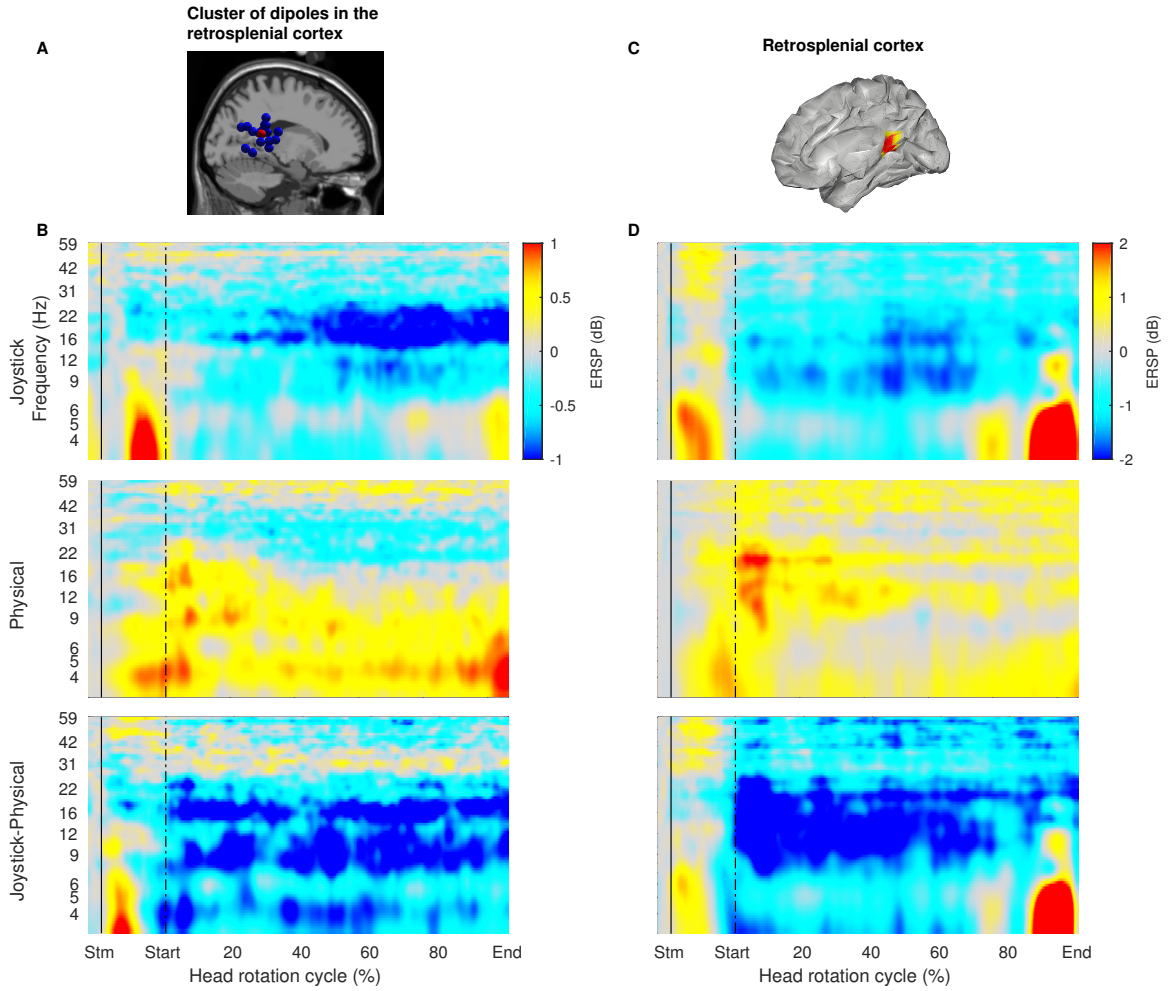

Figure 9: Event-related spectral perturbations (ERSPs) in the RSC. Panels **A** and **B** are adapted from Gramann et al. [2018]. **A**: Cluster of IC equivalent current dipoles in or near the RSC. **B**: ICA derived ERSPs of the joystick and physical rotation conditions and their difference. **C**: Location of the RSC in the cortical surface of our template. **D**: PEB+ derived ERSPs of the joystick and physical rotation conditions and their difference. The x-axes at the bottom of panels **B** and **D** are annotated with the stimulus onset (Stm), movement onset (Start), percentage of the head rotation cycle, and movement offset (End).

Removal algorithm, it can reduce artifacts without significantly distorting epochs of clean data. Furthermore, we were able to replicate the main finding of a study that looked into the dynamics of the retrosplenial cortex (RSC) supporting heading computation during full-body rotations.

The ability to estimate the time series of EEG sources that correspond to known anatomical locations accounting for the influence of artifacts without user intervention, as well as its online adaptation, makes the PEB+ algorithm appealing for established ERP paradigms as well as MoBI. We believe that the proposed algorithm can help to solve basic research questions employing EEG as the functional imaging modality, and at the same time constitute a biologically-grounded signal processing tool that can be useful to translational efforts.

Acknowledgements

This research was supported by NIMH Training Fellowship in Cognitive Neuroscience (AO), UC San Diego Chancellor's Research Excellence Scholarship (JM, AO), and UC San Diego School of Medicine start-up funds (JM). The PEB+ algorithm is copyrighted for commercial use (UC San Diego Copyright #SD2019-810) and free for research and educational purposes.

References

S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):

14–30, 2001. ISSN 10535888. doi: 10.1109/79.962275.

D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

O. Baumann and J. B. Mattingley. Medial Parietal Cortex Encodes Perceived Heading Direction in Humans. *Journal of Neuroscience*, 30(39):12897–901, 9 2010. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3077-10.2010.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995. ISSN 0899-7667. doi: 10.1162/neco.1995.7.6.1129.

N. Bigdely-Shamlo, K. Kreutz-Delgado, C. Kothe, and S. Makeig. EyeCatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5845–5848. IEEE, 7 2013a. ISBN 978-1-4577-0216-7. doi: 10.1109/EMBC.2013.6610881.

N. Bigdely-Shamlo, T. Mullen, K. Kreutz-Delgado, and S. Makeig. Measure projection analysis: A probabilistic approach to EEG source comparison and multi-subject inference. *NeuroImage*, 2013b. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.01.040.

N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 2015. ISSN 1662-5196. doi: 10.3389/fninf.2015.00016.

M. Böl, R. Weikert, and C. Weichert. A coupled electromechanical model for the excitation-dependent contraction of skeletal muscle. *Journal of the Mechanical Behavior of Biomedical Materials*, 2011. ISSN 17516161. doi: 10.1016/j.jmbbm.2011.04.017.

M. Breakspear. Dynamic models of large-scale brain activity. *Nature neuroscience*, 2017. ISSN 15461726. doi: 10.1038/nn.4497.

C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. d. R. Millán, F. Miralles, A. Nijholt, E. Opisso, N. Ramsey, P. Salomon, and G. R. Müller-Putz. BNCI Horizon 2020: towards a roadmap for the BCI community. *Brain-Computer Interfaces*, 2(1):1–10, 1 2015. ISSN 2326-263X. doi: 10.1080/2326263X.2015.1008956.

P. Byrne, S. Becker, and N. Burgess. Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, 2007. ISSN 0033295X. doi: 10.1037/0033-295X.114.2.340.

G. Casella. An introduction to empirical Bayes data analysis. *The American Statistician*, 39(2):83–87, 1985. ISSN 0003-1305. doi: 10.1080/00031305.1985.10479400.

C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung. Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1242–1245. IEEE, 7 2018. ISBN 978-1-5386-3646-6. doi: 10.1109/EMBC.2018.8512547.

R. Chavarriaga, J. del R. Millán, and J. del R. Millan. Learning From EEG Error-Related Potentials in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):381–388, 8 2010. ISSN 1534-4320. doi: 10.1109/TNSRE.2010.2053387.

T. C. Chiu, K. Gramann, L. W. Ko, J. R. Duann, T. P. Jung, and C. T. Lin. Alpha modulation in parietal and retrosplenial cortex correlates with navigation performance. *Psychophysiology*, 2012. ISSN 14698986. doi: 10.1111/j.1469-8986.2011.01270.x.

A. Cichocki and S.-i. Amari. *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, Ltd, Chichester, UK, 5 2002. ISBN 0471607916. doi: 10.1002/0470845899.

P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, 1994. ISSN 01651684. doi: 10.1016/0165-1684(94)90029-9.

S. Cotter, B. D. Rao, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7):2477–2488, 2005. ISSN 1053-587X. doi: 10.1109/TSP.2005.849172.

E. Cuspineda, C. Machado, T. Virues, E. Martínez-Montes, A. Ojeda, P. A. Valdés, J. Bosch, and L. Valdes. Source Analysis of Alpha Rhythm Reactivity Using LORETA Imaging with 64-Channel EEG and Individual MRI. *Clinical EEG and Neuroscience*, 40(3):150–156, 2009. ISSN 1550-0594. doi: 10.1177/155005940904000306.

A. M. Dale and M. I. Sereno. 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, 4 1993. ISSN 0898-929X. doi: 10.1162/jocn.1993.5.2.162.

A. Delorme, T. Mullen, C. Kothe, Z. Akalin Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig. EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing. *Computational Intelligence and Neuroscience*, 2011, 2011. ISSN 16875265. doi: 10.1155/2011/130714.

A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig. Independent EEG sources are dipolar. *PLoS ONE*, 2012. ISSN 19326203. doi: 10.1371/journal.pone.0030135.

R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.01.021.

C. Destrieux, B. Fischl, A. Dale, and E. Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.06.010.

K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3):1104–1120, 2 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.09.048.

K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, J. Ashburner, D. E. Glaser, R. N. a. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and Bayesian

Inference in Neuroimaging: Theory. *NeuroImage*, 16(2): 465–483, 6 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1090.

M. J. Fu, J. J. Daly, and M. C. Cavusoglu. A Detection Scheme for Frontalis and Temporalis Muscle EMG Contamination of EEG Data. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4514–4518. IEEE, 8 2006. ISBN 1-4244-0032-5. doi: 10.1109/IEMBS.2006.259511.

Y. Fujiwara, O. Yamashita, D. Kawawaki, K. Doya, M. Kawato, K. Toyama, and M.-a. Sato. A hierarchical Bayesian method to resolve an inverse problem of MEG contaminated with eye movement artifacts. *NeuroImage*, 45(2):393–409, 4 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.12.012.

I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616, 1997. ISSN 1053587X. doi: 10.1109/78.558475.

K. Gramann, J. Onton, D. Riccobon, H. J. Mueller, S. Bardins, and S. Makeig. Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation. *Journal of Cognitive Neuroscience*, 2010. ISSN 0898929X. doi: 10.1162/jocn.2009.21369.

K. Gramann, D. P. Ferris, J. Gwin, and S. Makeig. Imaging natural cognition in action, 2014. ISSN 01678760.

K. Gramann, F. U. Hohlefeld, L. Gehrke, and M. Klug. Heading computation in the human retrosplenial complex during full-body rotation. *bioRxiv*, 2018. doi: 10.1101/417972.

A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc. OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical engineering online*, 9:45, 1 2010. ISSN 1475-925X. doi: 10.1186/1475-925X-9-45.

a. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, and M. Kowalski. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 70:410–22, 4 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.12.051.

H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. Van Huffel, and I. Lemahieu. Review on solving the forward problem in EEG source analysis, 2007. ISSN 17430003.

S. Haufe, R. Tomioka, T. Dickhaus, C. Sannelli, B. Blankertz, G. Nolte, and K. R. Müller. Large-scale EEG/MEG source localization with spatial flexibility. *NeuroImage*, 54(2):851–859, 1 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.09.003.

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*, 5, 2011. ISSN 1662-5161. doi: 10.3389/fnhum.2011.00076.

S.-H. Hsu, T. R. Mullen, T.-P. Jung, and G. Cauwenberghs. Real-Time Adaptive EEG Source Separation Using Online Recursive Independent Component Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016. ISSN 1534-4320. doi: 10.1109/TNSRE.2015.2508759.

M. X. Huang, A. M. Dale, T. Song, E. Halgren, D. L. Harrington, I. Podgorny, J. M. Canive, S. Lewis, and R. R. Lee. Vector-based spatial-temporal minimum L1-norm solution for MEG. *NeuroImage*, 2006. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.01.029.

Y. Huang, L. C. Parra, and S. Haufe. The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage*, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.12.019.

M. K. Islam, A. Rastegarnia, and Z. Yang. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 2016. ISSN 09877053. doi: 10.1016/j.neucli.2016.07.002.

A. S. Janani, T. S. Grummett, T. W. Lewis, S. P. Fitzgibbon, E. M. Whitham, D. DelosAngeles, H. Bakhshayesh, J. O. Willoughby, and K. J. Pope. Evaluation of a minimum-norm based beamforming technique, sLORETA, for reducing tonic muscle contamination of EEG at sensor level. *Journal of Neuroscience Methods*, 288:17–28, 8 2017. ISSN 01650270. doi: 10.1016/j.jneumeth.2017.06.011.

T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 2000. ISSN 00485772. doi: 10.1017/S0048577200980259.

E. Jungnickel and K. Gramann. Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects. *Frontiers in Human Neuroscience*, 10, 6 2016. ISSN 1662-5161. doi: 10.3389/fnhum.2016.00306.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. ISSN 1537274X. doi: 10.1080/01621459.1995.10476572.

A. N. Khambhati, A. E. Sizemore, R. F. Betzel, and D. S. Bassett. Modeling and interpreting mesoscale network dynamics, 2018. ISSN 10959572.

A. Khanna, A. Pascual-Leone, C. M. Michel, and F. Farzan. Microstates in resting-state EEG: Current status and future directions, 2015. ISSN 18737528.

A. Kilicarslan, R. G. Grossman, and J. L. Contreras-Vidal. A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements. *Journal of Neural Engineering*, 13(2):026013, 4 2016. ISSN 1741-2560. doi: 10.1088/1741-2560/13/2/026013.

T. Koenig, L. Prichep, D. Lehmann, P. V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, and E. John. Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *NeuroImage*, 16(1):41–48, 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1070.

Q. V. Le, A. Karpenko, J. Ngiam, and A. Ng. ICA with reconstruction cost for efficient overcomplete feature learning. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in*

Neural Information Processing Systems 24, pages 1017–1025. Curran Associates, Inc., 2011.

M. S. Lewicki and T. J. Sejnowski. Learning nonlinear overcomplete representations for efficient coding. *Advances in Neural Information Processing Systems 10 (NIPS'97)*, pages 556–562, 1998.

C. T. Lin, T. C. Chiu, and K. Gramann. EEG correlates of spatial orientation in the human retrosplenial complex. *NeuroImage*, 2015. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.07.009.

C. T. Lin, T. C. Chiu, Y. K. Wang, C. H. Chuang, and K. Gramann. Granger causal connectivity dissociates navigation networks that subserve allocentric and egocentric path integration. *Brain Research*, 2018. ISSN 18726240. doi: 10.1016/j.brainres.2017.11.016.

F. Lopes da Silva. EEG and MEG: Relevance to neuroscience, 2013. ISSN 08966273.

D. J. C. MacKay. Bayesian Interpolation. *Neural Computation*, 4(3):415–447, 1992. ISSN 0899-7667. doi: 10.1162/neco.1992.4.3.415.

D. J. C. MacKay. Hyperparameters: Optimize, or integrate out? *Maximum entropy and bayesian methods*, pages 43–59, 1996. ISSN 9048144078. doi: 10.1007/978-94-015-8729-7.

D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2008a. ISBN 0521642981. doi: 10.2277/0521642981.

D. J. C. MacKay. Independent Component Analysis and Latent Variable Modelling. In *Information Theory, Inference, and Learning Algorithms*, pages 437–440. Cambridge University Press, 2008b. ISBN 0521642981. doi: 10.2277/0521642981.

S. Makeig and J. Onton. *ERP Features and EEG Dynamics*. Oxford University Press, 12 2011. ISBN 9780199940356. doi: 10.1093/oxfordhb/9780195374148.013.0035.

S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*, 1997. ISSN 0027-8424. doi: 10.1073/pnas.94.20.10979.

S. Makeig, K. Gramann, T.-P. P. Jung, T. J. Sejnowski, and H. Poizner. Linking brain, mind and behavior. *International Journal of Psychophysiology*, 73(2):95–100, 8 2009. ISSN 01678760. doi: 10.1016/j.ijpsycho.2008.11.008.

M. M. N. Mannan, M. A. Kamran, and M. Y. Jeong. Identification and Removal of Physiological Artifacts From Electroencephalogram Signals: A Review. *IEEE Access*, 6:30630–30652, 2018. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2842082.

J. D. Martínez-Vargas, F. M. Grisales-Franco, and G. Castellanos-Dominguez. Estimation of M/EEG Nonstationary Brain Activity Using Spatio-temporal Sparse Constraints. In *Artificial Computation in Biology and Medicine*, pages 429–438. Springer International Publishing, 2015. doi: 10.1007/978-3-319-18914-7{_}45.

K. Mcdowell, C. T. Lin, K. S. Oie, T. P. Jung, S. Gordon, K. W. Whitaker, S. Y. Li, S. W. Lu, and W. D. Hairston. Real-world neuroimaging technologies. *IEEE Access*, 2013. ISSN 21693536. doi: 10.1109/ACCESS.2013.2260791.

R. K. Mehta and R. Parasuraman. Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00889.

C. M. Michel and M. M. Murray. Towards the utilization of EEG as a brain imaging tool, 2012. ISSN 10538119.

J. Mishra and A. Gazzaley. Closed-Loop Rehabilitation of Age-Related Cognitive Disorders. *Seminars in Neurology*, 34(05):584–590, 12 2014. ISSN 0271-8235. doi: 10.1055/s-0034-1396011.

C. N. Morris. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, 1983. ISSN 1537274X. doi: 10.2307/2287098.

T. R. Mullen, C. A. E. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs. Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567, 11 2015. ISSN 0018-9294. doi: 10.1109/TBME.2015.2481482.

R. M. Neal. *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*. Springer New York, New York, NY, 1996. ISBN 978-0-387-94724-2. doi: 10.1007/978-1-4612-0745-0.

P. L. Nunez and R. Srinivasan. *Electric Fields of the Brain*. Oxford University Press, 1 2006. ISBN 9780195050387. doi: 10.1093/acprof:oso/9780195050387.001.0001.

A. Ojeda, K. Kreutz-Delgado, and T. Mullen. Fast and robust Block-Sparse Bayesian learning for EEG source imaging. *NeuroImage*, 174(8):449–462, 7 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.03.048.

J. Onton, M. Westerfield, J. Townsend, and S. Makeig. Imaging human EEG dynamics using independent component analysis. *Neuroscience & Biobehavioral Reviews*, 30(6):808–822, 1 2006. ISSN 01497634. doi: 10.1016/j.neubiorev.2006.06.007.

J. Palmer, K. Kreutz-Delgado, and S. Makeig. AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components. *San Diego, CA: Technical report, Swartz Center for Computational Neuroscience*, 2011.

R. D. Pascual-Marqui, M. Esslen, K. Kochi, and D. Lehmann. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods and findings in experimental and clinical pharmacology*, 24 Suppl C:91–5, 1 2002. ISSN 0379-0355. doi: 841[pii]. URL `http://www.ncbi.nlm.nih.gov/pubmed/12575492`.

W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols. *Statistical Parametric Mapping*. Elsevier, 2007. ISBN 9780123725608. doi: 10.1016/B978-0-12-372560-8.X5000-1.

L. Pion-Tonachini, S. Makeig, and K. Kreutz-Delgado. Crowd labeling latent Dirichlet allocation. *Knowledge*

and Information Systems, 53(3):749–765, 12 2017. ISSN 0219-1377. doi: 10.1007/s10115-017-1053-1.

T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert. Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features. *Journal of Neural Engineering*, 2017. ISSN 17412552. doi: 10.1088/1741-2552/aa69d1.

P. E. Sharp, H. T. Blair, and J. Cho. The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosciences*, 24(5):289–294, 5 2001. ISSN 01662236. doi: 10.1016/S0166-2236(00)01797-5.

H. S. Stern. A Test by Any Other Name: P Values, Bayes Factors, and Statistical Inference. *Multivariate Behavioral Research*, 2016. ISSN 00273171. doi: 10.1080/00273171.2015.1099032.

G. Tamburro, P. Fiedler, D. Stone, J. Haueisen, and S. Comani. A new ICA-based fingerprint method for the automatic removal of physiological artifacts from EEG recordings. *PeerJ*, 6:e4380, 2 2018. ISSN 2167-8359. doi: 10.7717/peerj.4380.

M. E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1:211–245, 2001. ISSN 15324435. doi: 10.1162/15324430152748236.

M. E. Tipping and A. C. Faul. Fast marginal likelihood maximisation for sparse Bayesian models. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, Florida, 1 2003. AISTATS. URL `http://www.miketipping.com/papers.htm`.

M. S. Treder, A. Bahramisharif, N. M. Schmidt, M. A. Van Gerven, and B. Blankertz. Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. *Journal of NeuroEngineering and Rehabilitation*, 2011. ISSN 17430003. doi: 10.1186/1743-0003-8-24.

N. J. Trujillo-Barreto, E. Aubert-Vázquez, and P. A. Valdés-Sosa. Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21(4):1300–1319, 4 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2003.11.008.

N. J. Trujillo-Barreto, E. Aubert-Vázquez, and W. D. Penny. Bayesian M/EEG source reconstruction with spatio-temporal priors. *NeuroImage*, 39(1):318–335, 1 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.07.062.

P. A. Valdés-Hernández, N. von Ellenrieder, A. Ojeda-Gonzalez, S. Kochen, Y. Alemán-Gómez, C. Muravchik, and P. A. Valdés-Sosa. Approximate average head models for EEG source imaging. *Journal of Neuroscience Methods*, 185(1):125–132, 12 2009. ISSN 01650270. doi: 10.1016/j.jneumeth.2009.09.005.

P. A. Valdés-Sosa, M. Vega-Hernández, J. M. Sánchez-Bornot, E. Martínez-Montes, and M. A. Bobes. EEG source imaging with spatio-temporal tomographic non-negative independent component analysis. *Human Brain Mapping*, 30(6):1898–1910, 6 2009. ISSN 10659471. doi: 10.1002/hbm.20784.

D. Van De Ville, J. Britz, and C. M. Michel. EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*, 2010. ISSN 0027-8424. doi: 10.1073/pnas.1007841107.

L. Van Der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 2008. ISSN 1532-4435. doi: 10.1007/s10479-011-0841-3.

B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 1997. ISSN 0018-9294. doi: 10.1109/10.623056.

J. Wagner, S. Makeig, M. Gola, C. Neuper, and G. Muller-Putz. Distinct Band Oscillatory Networks Subserving Motor and Cognitive Control during Gait Adaptation. *Journal of Neuroscience*, 36(7):2212–2226, 2 2016. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3543-15.2016.

D. Wipf and S. Nagarajan. A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*, 44(3):947–966, 2 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.02.059.

Z. Zhang and B. D. Rao. Extension of SBL Algorithms for the Recovery of Block Sparse Signals With Intra-Block Correlation. *IEEE Transactions on Signal Processing*, 61(8):2009–2015, 4 2013. ISSN 1053-587X. doi: 10.1109/TSP.2013.2241055.