bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states

David Sabbagh[a,b,c,*], Pierre Ablin[a], Gaël Varoquaux[a], Alexandre Gramfort[a], Denis A. Engemann[a,*]

[a] *Université Paris-Saclay, Inria, CEA, Palaiseau, France*
[b] *Inserm, UMRS-942, Paris Diderot University, Paris, France*
[c] *Department of Anaesthesiology and Critical Care, Lariboisière Hospital, Assistance Publique Hôpitaux de Paris, Paris, France*

**Abstract**

Predicting biomedical outcomes from Magnetoencephalography and Electroencephalography (M/EEG) is central to applications like decoding, brain-computer-interfaces (BCI) or biomarker development and is facilitated by supervised machine learning. Yet most of the literature is concerned with within-subject classification. Here, we focus on predicting continuous outcomes from M/EEG signal power across subjects. Considering different generative mechanisms for M/EEG signals and the biomedical outcome, we propose statistically-consistent predictive models that avoid source-reconstruction based on the covariance as representation. Our mathematical analysis and ground truth simulations demonstrated that consistent parameter estimation can be obtained with Source Power Comodulation (SPoC) supervised spatial filtering or by embedding with Riemannian geometry. Additional simulations revealed that Riemannian methods were more robust to model violations, in particular geometric distortions induced by individual anatomy. To estimate the relative contribution of brain dynamics and anatomy to prediction performance, we propose a novel model inspection procedure based on biophysical forward modeling. Applied to cross-subject prediction, the analysis revealed that the Riemannian model better exploited anatomical information while sensitivity to brain dynamics was similar across methods. We then probed the robustness of the models across different data cleaning options. Environmental denoising was globally important but Riemannian models were strikingly robust and continued performing even without preprocessing. Our results suggest each method has its niche: SPoC is practical for within-subject prediction while the Riemannian model may enable simple end-to-end learning.

*Keywords:* MEG/EEG, Neuronal oscillations, Machine Learning, Covariance, Spatial Filters, Riemannian Geometry

## 1. Introduction

Magnetoencephalography and Electroencephalography (M/EEG) access population-level neuronal dynamics across multiple temporal scales from seconds to milliseconds (Buzsáki and Draguhn, 2004; Hämäläinen et al., 1993). Its wide coverage of brain rhythms supports modeling cognition and brain health at different levels of organization from states to traits (Baillet, 2017; Buzsáki and Watson, 2012; da Silva, 2013). In the past decades, this has led to predictive modeling approaches in which cognitive or clinical targets are inferred from the electrophysiological signals (Woo et al., 2017; Besserve et al., 2007). In a common scenario, single-trial stimulus details are predicted from chunks of event-related signal, *e.g.* , visual orientation or auditory novelty (Cichy et al., 2015; King et al., 2013). With brain-computer-interfaces (BCI), one aims to read out cognitive states and translate them into control signals, *e.g.* , to capture movement-intentions (Wolpaw et al., 1991; Lotte et al., 2007; Tangermann et al., 2008). For biomarkers applications, the focus is on predicting medical diagnosis and other clinical endpoints (Engemann et al., 2018; Sami et al., 2018; Mazaheri et al., 2018).

What is the physiological source of M/EEG-based prediction? Similar to an analog radio, M/EEG receives signals containing multiplexed streams of information in different frequency 'channels' (van Wassenhove, 2016; Akam and

*Corresponding authors: 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau
*Email addresses:* dav.sabbagh@gmail.com (David Sabbagh), denis-alexander.engemann@inria.fr (Denis A. Engemann)

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kullmann, 2014; Panzeri et al., 2010). The signal comprises periodic and arrhythmic components which give rise to the characteristic 1/f power law regime (Dehghani et al., 2010; Linkenkaer-Hansen et al., 2001; He et al., 2010). M/EEG brain-dynamics originate from transient large-scale synchrony of distinct brain-networks where the anatomical regions involved communicate in different frequency bands (Hipp et al., 2012; Siegel et al., 2012). Typically, the frequency depends on the spatial scale of the network: as the scale becomes more local the spectral frequency increases (Buzsáki and Draguhn, 2004; Honey et al., 2007).

This motivates modeling approaches sensitive to both the temporal scale and the topography of the signal. Unfortunately, the neural sources of M/EEG cannot be observed and have to be inferred with uncertainty from their distorted representation on extra-cranial sensors. This argues in favor of statistical-learning techniques that can readily exploit high-density sensor arrays beyond sensor-wise statistical testing. So far this has been approached by explicit biophysical source modeling (Hämäläinen and Ilmoniemi, 1994; Khan et al., 2018; Westner et al., 2018), statistical approximations of biophysical generators through Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000; Makeig et al., 1995; Stewart et al., 2014; Subasi and Gursoy, 2010), spatial filtering approaches inspired by BCI (Nikulin et al., 2011; Dähne et al., 2014, 2013; Haufe et al., 2014; Dähne et al., 2014) or direct application of general purpose machine learning on the sensor time series (King et al., 2013).

Strikingly, the bulk of the literature on predictive modeling from M/EEG focuses on classification problems, evoked responses and within-subject targets. This is understandable, as evoked response analysis draws on rich resources from a long-standing history in experimental psychology (Coles and Rugg, 1995; Näätänen, 1975; Polich and Kok, 1995) and lend themselves to categorical problems as defined by experimental conditions. Besides, working on classification rather than regression may be more rewarding, as learning the boundary between classes is easier than estimating a full regression function (Hastie et al., 2005, chapter 7.3.1). Nevertheless, high-interest clinical targets other than diagnosis are often continuous and often involve cross-subject prediction (*e.g.* , prediction of risk scores, optimal drug-dosage, time of hospitalization or survival). Moreover, as EEG-recordings are combined across medical sites where different EEG-protocols are used, additional strain is put on spontaneous brain rhythms that can be accessed even if no particular task is used (Engemann et al., 2018). Yet, it is currently unclear how learning approaches based on brain rhythms compare as the data generating mechanism changes (within-subject vs cross-subject) or when the underlying probability model (*e.g.* log-linear vs linear relationship to power) is not a priori known.

In this paper we, therefore, focus on linking neural power spectra with their measure in M/EEG using appropriate models that facilitate prediction with high-dimensional regression. We aim to answer the following questions: 1) How can regression on M/EEG power-spectra be related to statistical models of the outcome and the neural signal? 2) What are the mathematical guarantees that a type of regression captures a given brain-behavior link? 3) How do ensuing candidate models perform in the light of model violations, uncertainty about the true data-generating process, variable noise, and different preprocessing options? The article is organized as follows: First we detail commonly used approaches for M/EEG-based predictive modeling. Subsequently, we develop a coherent mathematical framework for relating M/EEG-based regression to models of the neural signal, and, as a result, propose to conceptualize regression as predicting from linear combinations of uncorrelated statistical sources. Then we present numerical simulations which confront different regression models with commonly encountered model violations. Subsequently, we conduct detailed model comparisons on MEG data for within-subject and between-subject problems. Finally, we investigate practical issues related to availability of source-modeling and preprocessing options.

## 2. Methods

### *2.1. State-of-the art approaches to predict from M/EEG observations*

One important family of approaches for predictive modeling with M/EEG is relying on explicit biophysical source modeling. Here, anatomically constrained inverse methods are used to infer the most likely electromagnetic source configuration given the observations (Hämäläinen et al., 1993). Common techniques rely on fitting electrical-current dipoles (Mosher et al., 1992) or involve penalized linear inverse models to estimate the current distribution over a pre-specified dipole grid (Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006; Van Veen and Buckley, 1988; Hauk and Stenroos, 2014). Anatomical prior knowledge is injected through the well-defined forward model: Maxwell equations enable computing leadfields from the geometry and composition of the head, which predict propagation from a known

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

source to the sensors (Hämäläinen et al., 1993; Mosher et al., 1999). From a signal-processing standpoint, when these steps lead to a linear estimation of the sources, they can be thought of as biophysical spatial filtering. Prediction is then based on the estimated source-signals, see for example (Westner et al., 2018; Kietzmann et al., 2019; Khan et al., 2018).

A second family is motivated by unsupervised decomposition techniques such as Independent Component Analysis (Hyvärinen and Oja, 2000; Makeig et al., 1996), which also yield spatial filters and estimates of maximally independent sources that can be used for prediction (Stewart et al., 2014; Wang and Makeig, 2009; Subasi and Gursoy, 2010). Such methods model the data as an independent set of statistical sources that are entangled by a so-called mixing matrix, often interpreted as the leadfields. Here, the sources are purely statistical objects and no anatomical notion applies directly. In practice, unsupervised spatial filters are often combined with source modeling and capture a wide array of situations ranging from single dipole-sources to entire brain-networks (Hild II and Nagarajan, 2009; Brookes et al., 2011; Delorme et al., 2012).

Finally, a third family directly applies general-purpose machine learning on sensor space signals without explicitly considering the data generating mechanism. Following a common trend in other areas of neuroimaging research (Dadi et al., 2019; Schulz et al., 2019; He et al., 2019), linear prediction methods have turned out extraordinarily well-suited for this task, *i.e.* , logistic regression (Andersen et al., 2015), linear discriminant analysis (Wardle et al., 2016), linear support vector machines (King et al., 2013).

The success of linear models deserves separate attention as these methods enable remarkable predictive performance with simplified fast computation. While interpretation and incorporation of prior knowledge remain challenging, significant advances have been made in the past years. This has led to novel methods for specifying and interpreting linear models (Haufe et al., 2014; van Vliet and Salmelin, 2019). Recent work has even suggested that for the case of learning from evoked responses, linear methods are compatible with the statistical models implied by source localization and unsupervised spatial filtering (King et al., 2018; King and Dehaene, 2014; Stokes et al., 2015). Indeed, if the target is linear in the source signal, *i.e.* , due to the linear superposition principle, the mixing amounts to a linear transform that can be captured by a linear model with sufficient data. Additional source localization or spatial filtering should therefore be unnecessary in this case.

On the other hand, the situation is more complex when predicting targets from brain rhythms, *e.g.* , induced responses (Tallon-Baudry and Bertrand, 1999) or spontaneous oscillations. As brain-rhythms are not strictly time-locked to external events, they cannot be accessed by averaging. Instead, they are commonly represented by the signal power in shorter or longer time windows and often give rise to log-linear models (Buzsáki and Mizuseki, 2014; Roberts et al., 2015). A consequence of such non-linearities is that it cannot be readily captured by a linear model. Moreover, simple tricks such as log-transforming the power estimates only address the issue when applied at the source-level, not the sensor-level where the leadfields have already spatially smeared the signal.

This leads back to spatial filtering approaches. Beyond source localization and unsupervised filtering, supervised spatial filtering methods have recently become more popular beyond the context of BCIs. These methods solve generalized eigenvalue problems to estimate coordinate systems constructed with regard to criteria relevant for prediction. For example, spatio-spectral-decomposition (SSD) is an unsupervised technique that enhances SNR with regard to power in surrounding frequencies (Nikulin et al., 2011). On the other hand, common spatial patterns (CSP) and Source Power Comodulation (SPoC) focus on correlation with the target (Dähne et al., 2014; Koles, 1991; Dähne et al., 2013), whereas Dmochowski et al. (2012) have proposed variants of Canonical Correlation Analysis (CCA) (Hotelling, 1992; Dähne et al., 2014) without orthogonality constraint to focus on shared directions of variation between related datasets or by proposing shared envelope correlations as optimization target (Dähne et al., 2014). This yields a two-step procedure: 1) spatial filters model the correlation induced by the leadfields and provide unmixed time-series 2) some non-linear transforms such as logarithms are applied to these time-series as the validity of linear equations is now secured.

A more recent single-step approach consists in learning directly from spatially correlated power-spectra with linear models and Riemannian geometry (Barachant et al., 2011, 2013; Yger et al., 2017). This mathematical framework provides principles to correct for the geometric distortions arising from linear mixing of non-linear sources. This is achieved by using a Riemannian metric, immune to linear transformations, which allows transforming the covariance matrices used for representing the M/EEG signal to Euclidean objects for which linear models apply. This approach has turned out to be promising for enhancing classification of event-related data and has been the core ingredient of several winning solutions in recent data analysis competitions, *e.g.* , the seizure prediction challenge organized by the University of Melbourne in 2016.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Importantly, the majority of approaches has been mainly used for within-subject prediction problems. Here we will explicitly extend the reasoning in bridging within-subject and cross-subject prediction, both, theoretically and at the level of data analysis.

*Notation.* Scalars $s \in \mathbb{R}$ are written with lowercase, vectors $\boldsymbol{s} = [s_1, \ldots, s_N] \in \mathbb{R}^N$ with bold lowercase, and matrices $\boldsymbol{M} \in \mathbb{R}^{N \times M}$ with bold uppercase. $\boldsymbol{I}_N$ is the identity matrix of size $N$. $[\cdot]^\top$ represents vector or matrix transposition. $\mathrm{Tr}(\cdot)$ and $\mathrm{rank}(\cdot)$ are respectively the trace and the rank operators. The $\ell_2$ norm of a vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|_2^2 = \sum_i x_i^2$. We denote by $\mathcal{M}_P$ the space of $P \times P$ square real-valued matrices, $\mathcal{S}_P = \{\boldsymbol{M} \in \mathcal{M}_P, \boldsymbol{M}^\top = \boldsymbol{M}\}$ the subspace of symmetric matrices, $\mathcal{S}_P^{++} = \{\boldsymbol{S} \in \mathcal{S}_P, \boldsymbol{x}^\top S \boldsymbol{x} > 0, \forall \boldsymbol{x} \in \mathbb{R}^P\}$ the subspace of $P \times P$ symmetric positive definite matrices, $\mathcal{S}_P^{+} = \{\boldsymbol{S} \in \mathcal{S}_P, \boldsymbol{x}^\top S \boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \mathbb{R}^P\}$ the subspace of $P \times P$ symmetric semi-definite positive (SPD) matrices, $\mathcal{S}_{P,R}^{+} = \{\boldsymbol{S} \in \mathcal{S}_P^{+}, \mathrm{rank}(\boldsymbol{S}) = R\}$ the subspace of SPD matrices of fixed rank R. All matrices $\boldsymbol{S} \in \mathcal{S}_P^{++}$ are full rank, invertible (with $\boldsymbol{S}^{-1} \in \mathcal{S}_P^{++}$) and diagonalizable with real strictly positive eigenvalues: $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ with $\boldsymbol{U}$ an orthogonal matrix of eigenvectors of $\boldsymbol{S}$ ($\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{I}_P$) and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the diagonal matrix of its eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n > 0$. For a matrix $\boldsymbol{M}$, $\mathrm{diag}(\boldsymbol{M}) \in \mathbb{R}^P$ is its diagonal. We also define the exponential and logarithm of a matrix: $\forall \boldsymbol{S} \in \mathcal{S}_P^{++}, \log(\boldsymbol{S}) = \boldsymbol{U}\ \mathrm{diag}(\log(\lambda_1), \ldots, \log(\lambda_n))\ \boldsymbol{U}^\top \in \mathcal{S}_P$, and $\forall \boldsymbol{M} \in \mathcal{S}_P, \exp(\boldsymbol{M}) = \boldsymbol{U}\ \mathrm{diag}(\exp(\lambda_1), \ldots, \exp(\lambda_n))\ \boldsymbol{U}^\top \in \mathcal{S}_P^{++}$. $\mathcal{N}(\mu, \sigma^2)$ denotes the normal (Gaussian) distribution of mean $\mu$ and variance $\sigma^2$. Finally, $\mathbb{E}_s[\boldsymbol{x}]$ represents the expectation of any random variable $\boldsymbol{x}$ *w.r.t.* its subscript $s$ when needed.

### 2.2. *The biophysical data-generating mechanism*

MEG and EEG signals are produced by electrical sources in the brain that emerge from the synchronous activity of neurons. These neural current generators form *physiological sources* in the brain. We will assume the existence of $M$ such sources, $\boldsymbol{z}(t) \in \mathbb{R}^M$, where $t$ represents time. These sources can be thought as localized current sources, such as a patch of cortex with synchronously firing neurons, or a large set of patches forming a network. In this work, we are interested in predicting a target variable $y \in \mathbb{R}$ from multivariate MEG/EEG signals $\boldsymbol{x}(t) \in \mathbb{R}^P$, where $P$ corresponds to the number of sensors. The underlying assumption is that the physiological sources are at the origin of the signals $\boldsymbol{x}$, and that they are statistically related to $y$. Often they are even the actual generators of $y$, *e.g.* , the neurons producing the finger movement of a person. Here, we embrace the statistical machine learning paradigm where one aims to learn a predictive model from a set of $N$ labeled training samples, $(\boldsymbol{x}_i(t), y_i), i = 1, \ldots, N$. The physics of the problem and the linearity of Maxwell's equations show that MEG/EEG acquisition is linear: the signals measured are obtained by linear combination of the underlying physiological sources. This leads to:

$$\boldsymbol{x}_i(t) = \boldsymbol{G}_i\, \boldsymbol{z}_i(t) \ , \tag{1}$$

where $\boldsymbol{G}_i \in \mathbb{R}^{P \times M}$ is the leadfield, also commonly referred to as gain matrix. Note that here the $j$-th column of $G_i$ is not necessarily constrained to be the forward model of a focal electrical current dipole in the brain. It can also correspond to large distributed sources. The area outside the cloud in Fig. 1 illustrates what was just described.

A natural approach to estimate a regression model in this setting consists in estimating the locations, amplitudes and extents of the sources from the MEG/EEG data. This estimation known as in the inverse problem (Baillet, 2017) can be done for example using the Minimum Norm Estimator (MNE) approach (Hämäläinen and Ilmoniemi, 1984). From the estimated sources, one can than learn to predict $y$. While approaching the problem from this perspective has some benefits, such as the ability to exploit the geometry and the physical properties of the head tissues of each subject, there are certain drawbacks. First, the inverse problem is ill-posed and notoriously hard to solve. Second, computing $\boldsymbol{G}_i$ requires costly MRI acquisitions and time-consuming manual labor by experienced MEG practitioners: anatomical coregistration and tedious data-cleaning to mitigate electromagnetic artefacts caused by environmental or physiological sources of non-interest outside of the brain. The purpose of this article is to show how to learn a regression model without biophysical source modeling, hence avoiding those drawbacks, which can be critical for clinical practice. Note that, in this article, we use the term *generative model* in its statistical sense, which, here, amounts to assuming a specific probabilistic model to account for the M/EEG observations and the prediction targets.

### 2.3. *Statistical approximation of the target and the M/EEG signals through generative models*

Independent Component Analysis (Hyvärinen and Oja, 2000) is another popular approach to model M/EEG signals (Makeig et al., 1997). We consider $Q \leq P$ *statistical sources* $\boldsymbol{s}(t) \in \mathbb{R}^Q$, that correspond to unknown latent

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

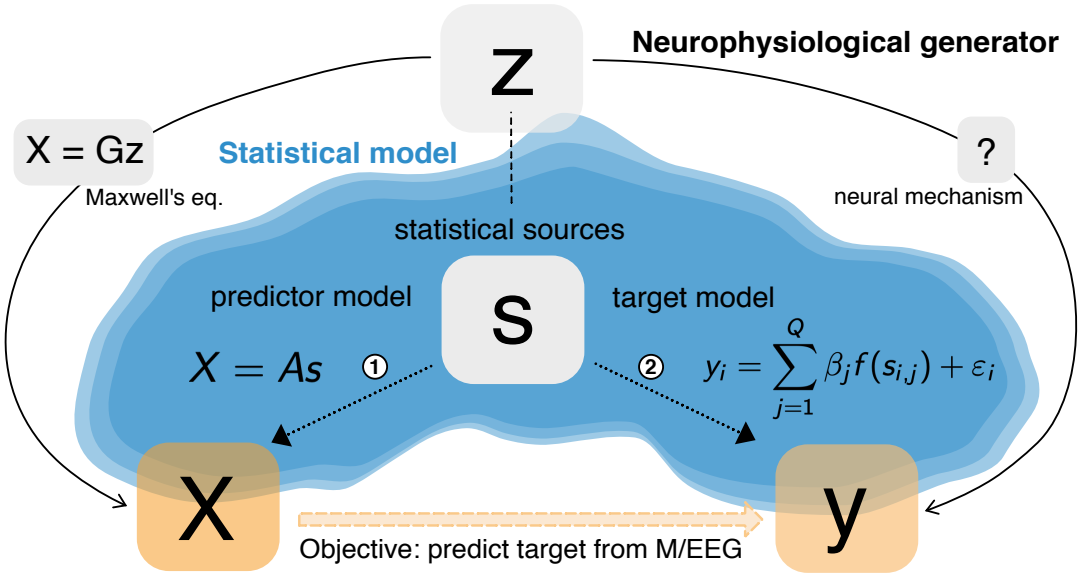


Figure 1: **Generative model for regression with M/EEG.** Unobservable neuronal activity $z$ gives rise to observed M/EEG data $\boldsymbol{X}$ and an observed neurobehavioral target $y$. The M/EEG data $\boldsymbol{X}$ is obtained by linear mixing of $z$ through the leadfield $\boldsymbol{G}$. The target $y$ is derived from $z$ through often unknown neural mechanisms. The statistical model (blue cloud) approximates the neurophysiological data-generating mechanisms with two sub-models, one for $\boldsymbol{X}$ (path 1), one for $y$ (path 2). Both models are based on a vector $s$ of uncorrelated statistical sources that, may refer to localized cortical activity or synchronous brain networks. The ensuing model generates $y$ from a linear combination of the statistical sources $s$. The generative model of $\boldsymbol{X}$ follows the ICA model (Hyvärinen and Oja, 2000) and assumes linear mixing of the source signals by $\boldsymbol{A}$, interpreted as a linear combination of the columns of the leadfield $\boldsymbol{G}$. The generative model of $y$ assumes a linear model in the parameters $\beta$ but allows for non-linear functions in the data, such as the power or the log-power. The mechanisms governing path 1 implies that the sources $s$ appear geometrically distorted in $\boldsymbol{X}$. This makes it impossible for a linear model to capture this distortion if $y$, in path 2, is generated by a non-linear function of $s$. This article focuses on how to mitigate this distortion when performing regression on M/EEG source power.

variables. These variables are assumed to be statistically related to the target variable $y$ and to be linearly related to measured signal $\boldsymbol{x}(t)$. The area inside the cloud depicted in Fig. 1 illustrates the generative models.

### *2.3.1. Generative model of the M/EEG signals*

Denoting by $\boldsymbol{a}_j \in \mathbb{R}^P$, $j = 1 \dots Q$ the *source patterns* (Haufe et al., 2014), which correspond to topographies on the sensor array, the generative model of M/EEG observations reads:

$$\boldsymbol{x}_i(t) = \sum_{j=1}^{Q} \boldsymbol{a}_j \, s_{i,j}(t) + \boldsymbol{n}_i(t) \ ,$$

where $\boldsymbol{s}_{i,j}(t) \in \mathbb{R}^Q$ is the $j$-th source time-series of sample $i$ and $\boldsymbol{n}_i(t) \in \mathbb{R}^P$ is a contamination due to noise. This model is conveniently written in matrix form:

$$\boldsymbol{x}_i(t) = \boldsymbol{A} \, \boldsymbol{s}_i(t) + \boldsymbol{n}_i(t) \ , \tag{2}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_Q] \in \mathbb{R}^{P \times Q}$ is the *mixing matrix*. Note that the mixing matrix $\boldsymbol{A}$ is assumed to be fixed across samples and across time: it does not depend on $i$ nor $t$. We also assume that the components $\{s_{i,j}, j = 1 \dots Q\}$ of the sources are zero-mean, uncorrelated, and independent from the noise. Each quantity in the right-hand side of Eq. (2), $A$, $\boldsymbol{s}_i(t)$ and $\boldsymbol{n}_i(t)$, is unknown and should be inferred from $\boldsymbol{x}_i(t)$. This setting encompasses both within-subject studies, where the samples $\boldsymbol{x}_i(t)$ are epochs of signal from a unique subject, and cross-subject studies where the samples $i$ represent the full signal of multiple subjects.

Note that this statistical generative model is a simplification of the biophysical generative mechanism: the real sources $z_i$ may not be independent, the gain $\boldsymbol{G}_i$ is sample-dependent in a cross-subject study, and the number of true sources may exceed the number of sensors, $M \gg P$.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

### 2.3.2. *Generative models of the target*

The proposed framework models $y_i$ as linear combination of the statistical sources $\boldsymbol{s}_i$. We considered three variants of generative regression models, allowing for modeling non-linearities:

1. Linearity in the sources at a certain time $\tau$:

$$y_i = \sum_{j=1}^{Q} \beta_j s_{i,j}(\tau) + \varepsilon_i \tag{3}$$

2. Linearity in the sources powers:

$$y_i = \sum_{j=1}^{Q} \beta_j p_{i,j} + \varepsilon_i \tag{4}$$

3. Linearity in the sources log-powers:

$$y_i = \sum_{j=1}^{Q} \beta_j \log(p_{i,j}) + \varepsilon_i \tag{5}$$

where $p_{i,j} = \mathbb{E}[s_{i,j}^2(t)]$ is the power of the $j$-th source of sample $i$ and $\varepsilon_i$ is an additive random perturbation. In practice, one prefers frequency specific models, where the previous relationships are obtained after $s_{i,j}(t)$ has been bandpass filtered in a specific frequency range. The broadband covariance (computed on the raw signal without temporal filtering) largely reflects low-frequency power as consequence of the predominant 1/f power spectrum, hence, is rarely of interest for predicting. In frequency-specific models, the powers are replaced by band-powers: power of the source in the chosen frequency band. The noise in the target variable depends on the context: it can represent intrinsic measurement uncertainty of $y_i$, for example sampling rate and latency jitter in behavioral recordings, inter-rater variability for a psychometric score, or simply a model mismatch. The true model may not be linear for example.

The first model ( 'linear-in-signal' , Eq. 3) is commonly used for evoked-responses studies, where the response is time-locked to the event of interest onset with a constant delay $\tau$. In this setting, assuming no acquisition noise ($\boldsymbol{n}_i(t) = 0$), both Eq. (2) and (3) are linear. Therefore, the relationship between $y_i$ and $\boldsymbol{x}_i(\tau)$ is linear. We can then perform regression directly by fitting a linear model to $\boldsymbol{x}_i(\tau)$. In the limit of no noise $\varepsilon_i$ and $\boldsymbol{n}_i(t)$, the prediction of $y_i$ is even perfect: we say that the linear model expressed in $\boldsymbol{x}$ is statistically *consistent*. Interestingly, as its performance does not depend on $\boldsymbol{A}$, the model is blind to the mixing matrix. Contrary to the first model (Eq. 3), the second model ('linear-in-powers', Eq. 4) and the third model ('log-linear-in-powers', Eq. 5) are not linear with respect to the sources, calling for more sophisticated approaches. These models are commonly used in the literature and support numerous statistical learning models on M/EEG (Blankertz et al., 2008; Dähne et al., 2014; Grosse-Wentrup* and Buss, 2008). In particular Buzsáki and Mizuseki (2014) discusses a wide body of evidence arguing in favor of log-linear relationships between brain dynamics and cognition.

In this article, we shall focus on linear regression on source power (second and third model) with methods that turn out to be blind to the mixing matrix and enable statistically consistent regression. For this purpose, we will summarize the M/EEG signal by its between-sensors covariance and rely on recent developments in research on spatial filters and theoretical results from Riemannian geometry.

## 2.4. *Statistical estimation procedures*

*M/EEG signal representation.* The second and third model (Eq. 4 and 5) involve powers *i.e.* the squared amplitude of the signal. This hints at the second-order moment of the signal: its covariance matrix. The covariance matrix of the $i$-th signal $\boldsymbol{x}_i(t)$ reads:

$$\boldsymbol{\Sigma}_i = \mathbb{E}_t \left[ \boldsymbol{x}_i(t)\, \boldsymbol{x}_i(t)^\top \right] \in \mathbb{R}^{P \times P} \quad .$$

The diagonal of this matrix represents the variance of each sensor, while the off-diagonal terms contain the unnormalized correlations between each pair of signals with squared amplitude units. Negative values in the off-diagonal express negative correlation. It is noteworthy that the number of parameters to estimate in a covariance grows quadratically with dimension. As a consequence, many more samples are required than there are sensors to accurately estimate such matrices (Engemann and Gramfort, 2015). This motivates the use of dimensionality reduction techniques, based

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on so-called *spatial filters*, which can be seen as low-rank regularization of the covariance matrix (Engemann and Gramfort, 2015; Coelho Rodrigues et al., 2017). A second motivation for spatial filter is to avoid working with rank deficient matrices that commonly result from popular techniques for cleaning data.

*Spatial filtering.* Spatial filtering consists in computing linear combinations of the signals to produce so-called *virtual sensors*. The weights of the combination form a spatial filter. Considering $R \leq P$ filters, it corresponds to the columns of a matrix $\boldsymbol{W} \in \mathbb{R}^{P \times R}$. If $R < P$, then spatial filtering reduces the dimension of the data. Note that, the covariance matrix of 'spatially filtered' signals $\boldsymbol{W}^\top \boldsymbol{x}_i$ is readily obtained from $\boldsymbol{\Sigma}_i$: $\boldsymbol{C}_i = \boldsymbol{W}^\top \boldsymbol{\Sigma}_i \boldsymbol{W}$. The question is now of how to estimate $\boldsymbol{W}$.

In a statistical-learning scenario, two approaches exist to estimate $\boldsymbol{W}$, depending if the target $y$ is used or not. A first strategy is to project the data into a subspace that captures most of the variance. This is achieved by Principal Component Analysis (PCA). As this strategy is not exploiting $y$, it is *unsupervised*. We denote the filters in this case by $\boldsymbol{W}_{\text{UNSUP}} = \boldsymbol{U}$, where $\boldsymbol{U}$ contains the eigenvectors corresponding to the top $R$ eigenvalues of the average covariance matrix $\overline{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Sigma}_i$. A second strategy we investigated is to use a supervised spatial filtering, namely the SPoC algorithm (Dähne et al., 2014). This approach was originally developed for multiple within-subject predictions, *e.g.* in BCI, and we adapt it here to a general problem that can also accommodate across-subjects predictions, where one observation corresponds to one subject instead of one trial. The main idea is to use the information contained in the target variable to guide the decomposition, giving preferences to sources whose power correlates with $y$. More formally the filters $\boldsymbol{W}$ are chosen to maximize the covariance between the power of the filtered signals and $y$. Denoting by $\boldsymbol{\Sigma}_y = \frac{1}{N} \sum_{i=1}^{N} y_i \boldsymbol{\Sigma}_i$ the weighted average covariance matrix, the first filter $\boldsymbol{w}_{\text{SUP}}$ is given by: $\boldsymbol{w}_{\text{SUP}} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^\top \boldsymbol{\Sigma}_y \boldsymbol{w}}{\boldsymbol{w}^\top \overline{\boldsymbol{\Sigma}} \boldsymbol{w}}$ . In practice, all the other filters in $\boldsymbol{W}_{\text{SUP}}$ are obtained by solving a generalized eigenvalue decomposition problem (Dähne et al., 2014). At this stage, the regression task now amounts to predicting $y_i$ from the spatially-filtered covariance matrices $\boldsymbol{C}_i$. As typical regression algorithms require vectors of features as input, we now need to transform each $\boldsymbol{C}_i$ into a vector $\boldsymbol{v}_i$.

*Vectorization.* A simple idea is to use the diagonal of $\boldsymbol{C}_i$ *i.e.* the variance of the spatially filtered signals. This approach is consistent if the (virtual-)sensors do correspond to the statistical sources of Eq. (2), which can be achieved using the SPoC algorithm with $R = P$. Hence, using a full-rank supervised spatial filtering step followed by a 'diag' vectorization $\boldsymbol{v}_i = \text{diag}(\boldsymbol{C}_i)$ achieves consistency for model (4). Similarly the 'log-diag' vectorization $\boldsymbol{v}_i = \log(\text{diag}(\boldsymbol{C}_i))$ achieves consistency for model (5). While this approach is appealing by its simplicity, it requires to first estimate the statistical sources by spatial filtering. To achieve consistency in one step, without assuming that the statistical sources have been recovered by spatial filtering, a first natural idea is to stack the coefficients of the upper triangular part of $\boldsymbol{C}_i$: $\boldsymbol{v}_i = \text{Upper}(\boldsymbol{C}_i) \in \mathbb{R}^{\frac{P(P+1)}{2}}$. While this approach can seem naive one can show that it is statistically consistent for model (4) (Sabbagh et al., 2019). To achieve consistency for model (5) one needs to resort to Riemannian geometry.

*Riemannian vectorization.* Covariance matrices are mathematical objects with a specific structure: they are positive definite. The space of positive definite matrices is a manifold (Förstner and Moonen, 2003): it is a high-dimensional surface and not an vector space, as it is not stable by subtraction of two covariance matrices. In particular, in order to properly vectorize the covariance matrices we should first map them to an Euclidean space, to be able to rely on standard vector arithmetics. A concise introduction to the basic concepts of Riemannian matrix manifolds can be found in (Absil et al., 2009) and is summarized below qualitatively. A Riemannian manifold $\mathcal{M}$ is a curved space that, at each point, locally resembles an Euclidean flat space called the tangent space at this point. The Riemannian manifold is endowed with a inner product which defines a scalar product on the tangent space, as well as a *geodesic distance* $d$ between two matrices on the manifold: the length of the shortest curve linking the two, given by summing the scalar product along the curve. In turn, the Riemannian mean of a set of matrices $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N$ can be defined as the matrix $\bar{\boldsymbol{C}}$ minimizing the sum of squared distances $\sum_{i=1}^{N} d(\boldsymbol{C}_i, \boldsymbol{C})^2$. Importantly, there is a mapping between the manifold and the tangent space at $\bar{\boldsymbol{C}}$, that approximately conserves distances: it represents a matrix $\boldsymbol{C}_i$ in a vector space where Euclidean distances approximate the original geodesic distances. This allows to define the *vectorization operator* at $\bar{\boldsymbol{C}}$, $\mathcal{P}_{\bar{\boldsymbol{C}}} : \mathcal{M} \to \mathbb{R}^K$ which explicitly captures the local Euclidean properties of the Riemannian manifold:

$$d(\boldsymbol{C}_i, \boldsymbol{C}_j) \simeq \|\mathcal{P}_{\overline{\boldsymbol{C}}}(\boldsymbol{C}_i) - \mathcal{P}_{\overline{\boldsymbol{C}}}(\boldsymbol{C}_j)\|_2 \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

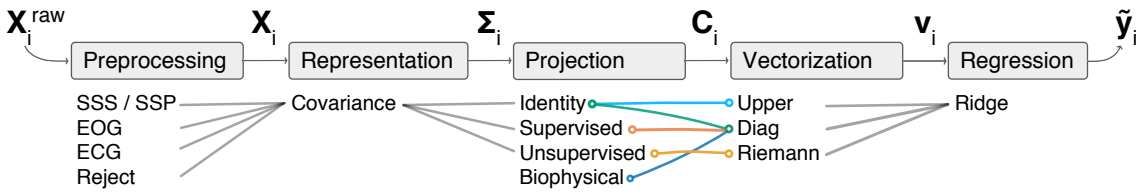


Figure 2: Proposed regression pipeline. Signals are first preprocessed, then represented by the between-sensors covariance, projection is optionally applied (identity means no spatial filtering $\boldsymbol{W} = \boldsymbol{I}$). The matrix is vectorized and, finally, handed over to standard Ridge regression with linear kernel. The choices for each step are detailed below each box. The preprocessing steps include removal of environmental artefacts with a) Signal-Space Separation (SSS) or Signal-Space Projection (SSP), b) removal of ocular (EOG) or cardiac artifacts (ECG) and, c) rejection of bad data segments. The five model types investigated in this work are depicted by the colors used for visualizing subsequent analyses: 'diag' in green, 'upper' in sky blue, 'SPoC' in red, 'Riemann' in orange and the biophysical 'MNE' model in dark blue. Note that only the most relevant combinations of projection and vectorization were investigated. The preprocessing procedure and the MNE model are introduced in the section *Cross-subject regression: age prediction*.

This vectorization operator is key for machine learning applications: it projects points in $\mathcal{M}$ on to $\mathbb{R}^K$, and the geodesic distance $d$ on $\mathcal{M}$ is approximated by the Euclidean $\ell_2$ distance on $\mathbb{R}^K$. Once vectorized, the matrices can be used as input for any regression technique that makes the assumption that data points live in a Euclidean space. Importantly, using different metrics leads to different vectorization operators which may have fundamentally different properties. A popular metric on the manifold of positive definite matrices is the *geometric* one which defines the geometric distance $d_G$ (Förstner and Moonen, 2003). This distance has the *affine invariance* property:

$$\text{For } \boldsymbol{W} \text{ invertible, } d_G(\boldsymbol{W}^\top \boldsymbol{C}\boldsymbol{W}, \boldsymbol{W}^\top \boldsymbol{C}'\boldsymbol{W}) = d_G(\boldsymbol{C}.\boldsymbol{C}') \ , \tag{7}$$

This distance is invariant by linear transform. Since the signals $\boldsymbol{x}$ are obtained as a linear transform of the sources $\boldsymbol{s}$, this distance potentially allows accessing the sources without explicitly inverting the linear mixing matrix.

*Regression models.* In this paper we consider four different regression models defined by the different options used for the spatial filtering and the vectorization steps. The '*upper*' model uses no spatial filtering followed by an upper vectorization. The '*diag*' model uses no spatial filtering followed by a log-diag vectorization. The '*SPoC*' model uses supervised spatial filtering based on the generalized eigenvalue problem as described by Dähne et al. (2014), followed by a log-diag vectorization. The '*Riemannian*' model uses the unsupervised spatial filtering proposed by Sabbagh et al. (2019), followed by a Riemannian vectorization. This last model is not influenced by the choice of spatial filters due to the affine-invariance property of the geometric distance $d_G$. Combining it with different spatial filtering options is therefore not of theoretical interest. The proposed pipeline is summarized in Fig. 2.

Under the linear power generative model (4), the Upper vectorization is a consistent vectorization: the relationship between $y_i$ and $\boldsymbol{v}_i = \text{Upper}(\boldsymbol{C}_i)$ is linear. However, under the log-linear power generative model (5), the Riemannian vectorization *w.r.t.* the geometric distance is consistent: the relationship between $y_i$ and $\boldsymbol{v}_i = \mathcal{P}_{\overline{\boldsymbol{C}}}(\boldsymbol{C}_i)$ is linear (Sabbagh et al., 2019). Specifically, the results from Sabbagh et al. (2019) make three assumptions: a) that there is no noise in the target variable ($\varepsilon_i = 0$), b) that the noise subspace is shared between subjects, c) that the sources are uncorrelated and independent from the noise. The proof relies on the affine invariance property of the geometric distance. Consequently, in this setting, linear regression methods applied on the $\boldsymbol{v}_i$ have no approximation error. The same can be shown with the 'Upper' vectorization approach for model (2) and (4).

### 2.5. Model violations

The current theoretical analysis implies that the mixing matrix $\boldsymbol{A}$ must be common to all subjects and the covariance matrices must be full rank. If these conditions are not satisfied, the consistency guarantees are lost, rendering model performance an empirical question.

*Individual mixing matrix.* A model where the mixing matrix $\boldsymbol{A}$ is subject-dependent reads: $\boldsymbol{x}_i(t) = \boldsymbol{A}_i \boldsymbol{s}_i(t) + \boldsymbol{n}_i(t)$ . Such situations typically arise when performing cross-subject regression due to individual head geometry, individual head positions in MEG and individual locations of EEG electrodes.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Rank-deficient signals.* M/EEG data is often rank-reduced for cleaning the data, leading to rank-deficient covariance matrices. Assuming the rank $R < P$ is the same across subjects, the corresponding covariance matrices do not belong to the $\mathcal{S}_P^{++}$ manifold anymore but to the $\mathcal{S}_{P,R}^{+}$ manifold of positive semi-definite matrices of fixed rank $R$. Unfortunately, the geometric distance cannot be used in this case and one can prove that no affine invariant distance exists on this manifold (Bonnabel and Sepulchre, 2009). One way to handle the rank-deficiency is to project covariance matrices on $\mathcal{S}_R^{++}$ with spatial filtering to make them full rank, and then use the geometric distance (Sabbagh et al., 2019). Since the Riemannian geometric distance on this manifold is affine invariant, supervised and unsupervised filtering will lead to the same prediction performance.

### 2.6. Model-inspection by error-decomposition

The link between the data-generating mechanism and the proposed regression models allows us to derive an informal analysis of variance (Gelman et al., 2005) for estimating the importance of the data generating factors head geometry, uniform global power and topographic detail. Given the known physics from Eq. (1), the data covariance can be written $\boldsymbol{\Sigma}_i = G_i \boldsymbol{\Sigma}_i^z G_i^\top$, where $\boldsymbol{\Sigma}_i^z$ is the covariance matrix of the physiological sources. The input to the regression model is therefore affected by both the head geometry expressed in $G_i$, and the covariance of the sources. We can hence compute degraded versions of the full covariance in which only specific components of the signal are retained based on computation of leadfields. Subsequent model comparisons against the full models then allow isolating the relative merit of each component. Following common practice, we considered electrical dipolar sources $z_i(t) \in \mathbb{R}^M$, with $M \approx 8000$, and we computed the leadfield matrix $\boldsymbol{G}_i$ with a boundary element model (BEM) (Gramfort et al., 2014). We then defined two alternative models which are only based on the anatomical information or, additionally, on the global signal power in a given frequency band without topographic structure.

*Model using anatomy.* Assuming $z_i^F(t) \sim \mathcal{N}(0, \boldsymbol{I}_M)$, we have that $\boldsymbol{\Sigma}_i^F = \boldsymbol{G}_i \boldsymbol{G}_i^\top$. This corresponds to only taking into account anatomy with the leadfields in the regression model.

*Model using anatomy and power.* Assuming $z_i^F(t) \sim \mathcal{N}(0, \sigma_i^2 \boldsymbol{I}_M)$, where $\sigma_i$ is a scalar, we have that $\boldsymbol{\Sigma}_i^F = \sigma_i^2 \boldsymbol{G}_i \boldsymbol{G}_i^\top$. This allows to take into account a source strength that is subject specific. Specifically, we chose $\sigma_i^2 = \mathrm{Tr}(\boldsymbol{\Sigma}_i)/\mathrm{Tr}(\boldsymbol{G}_i \boldsymbol{G}_i^\top)$, such that $\mathrm{Tr}(\boldsymbol{\Sigma}_i^F) = \mathrm{Tr}(\boldsymbol{\Sigma}_i)$: the sum of powers of the signals is the same. This corresponds to taking into account the total power of the sources in a given frequency band and anatomy in the ensuing regression model. Note that we omitted frequency-specific notation for simplicity.

### 2.7. Simulations

We considered simulations to investigate theoretical performance as model violations are gradually introduced. We focused on the 'linear-in-powers' model (Eq. 4) and the 'log-linear-in-powers' model (Eq. 5). Independent identically distributed covariance matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N \in S_P^{++}$ and variables $y_1, \ldots, y_N$ were generated for each generative model. The mixing matrix $\boldsymbol{A}$ was defined as $\exp(\mu \boldsymbol{B})$ with the random matrix $\boldsymbol{B} \in \mathbb{R}^{P \times P}$ and the scalar $\mu \in \mathbb{R}$ to control the distance of $A$ from identity ($\mu = 0$ yields $\boldsymbol{A} = \boldsymbol{I}_P$). The target variable was linked to the source powers (*i.e.* the variance) without and with a log function and is corrupted by Gaussian noise: $y_i = \sum_j \alpha_j f(p_{ij}) + \varepsilon_i$, with $f(x) = x$ or $\log(x)$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is a small additive random perturbation. We chose $P = 5$, $N = 100$ and $Q = 2$. The affine invariance property of the geometric Riemannian distance should make the model blind to the mixing matrix $\boldsymbol{A}$ and enable to perfect out-of-sample prediction whatever its value when the target variable is not corrupted by noise ($\sigma = 0$). Then we corrupted the clean ground truth data in two ways: by increasing noise in the target variable and with individual mixing matrices deviating from a reference: $\boldsymbol{A}_i = \boldsymbol{A} + \boldsymbol{E}_i$, where entries of $\boldsymbol{E}_i$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. The reference $\boldsymbol{A}$ can then be thought of as representing the head of a mean-subject.

### 2.8. MEG data

When analyzing MEG data, we do not have access to the actual sources and do not know a priori which generative model, hence, which regression model performs best. Likewise, we cannot expect perfect out-sample prediction: the target variable may be noisy (leading irreducible error), data samples are finite (leading to estimation error) and numerous model violations may apply (mixing matrices may be different for each sample, data may be rank-deficient due to pre-processing, etc.). However, by performing model comparisons based on cross-validation errors, we can potentially infer which model provides the better approximation.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

#### 2.8.1. *Within-subject regression: cortico-muscular coherence*

We first focused on within-subject regression of continuous electromyogram (EMG) from MEG beta activity.

*Data acquisition.* We analyzed one anonymous subject from the data presented in (Schoffelen et al., 2011) and provided by the FieldTrip website to study cortico-muscular coherence (Oostenveld et al., 2011). The MEG recording was acquired with 151 axial gradiometers and the Omega 2000 CTF whole-head system. EMG of forceful contraction of the wrist muscles (bilateral musculus extensor carpi radialis longus) was concomitantly recorded with two silver chloride electrodes. MEG and EMG data was acquired at 1200Hz sampling-rate and online-filtered at 300Hz. For additional details please consider the original study (Schoffelen et al., 2011).

*Data processing and feature engineering.* The analysis closely followed the continuous target decoding example from the MNE-Python website (Gramfort et al., 2014). We considered 200 seconds of joint MEG-EMG data. First, we filtered the EMG above 20 Hz using a time-domain firwin filter design, a Hamming window with 0.02 passband ripple, 53 dB stop band attenuation and transition bandwidth of 5Hz (-6dB at 17.5 Hz) with a filter-length of 661 ms. Then we filtered the MEG between 15 and 30 Hz using an identical filter design, however with 3.75 Hz transition bandwidth for the high-pass filter (-6 dB at 13.1 Hz) and 7.5 Hz for the low-pass filter (-6 dB at 33.75 Hz). The filter-length was about 880ms. Note that the transition bandwidth and filter-length was adaptively chosen by the default procedure implemented in the filter function of MNE-Python. We did not apply any artifact rejection as the raw data was of high quality. The analysis then ignored the actual trial structure of the experiment and instead considered a sliding window-approach with 1.5 s windows spaced by 250 ms. Allowing for overlap between windows allowed to increase sample size.

We then computed the covariance matrix in each time window and applied Oracle Approximation Shrinkage (OAS) (Chen et al., 2010) to improve conditioning of the covariance estimate. The target was defined as the variance of the EMG in each window.

*Model evaluation.* For the within-subject analysis with overlapping windows, we applied 10-fold cross-validation without shuffling such that folds correspond to blocks of neighboring time windows preventing data-leakage between training and testing splits. The initialization of the random number generator used for cross-validation was fixed, ensuring identical train-test splits across models. Note that a Monte Carlo approach with a large number of splits would lead to significant leakage, hence, optimistic bias (Varoquaux et al., 2017). This, unfortunately, limits the resolution of the uncertainty estimates and precludes formalized inference. As we did not have any a priori interest in the units of the target, we used the $R^2$ metric, a.k.a. coefficient of determination, for evaluation.

#### 2.8.2. *Cross-subject regression: age prediction*

In a second MEG data example, we considered a cross-person regression problem in which we focused on age prediction using the Cam-CAN dataset (Taylor et al., 2017; Shafto et al., 2014).

*Data acquisition.* Cam-CAN dataset contains 643 subjects with resting state MEG data, between 18 and 89 years of age. MEG was acquired using a 306 VectorView system (Elekta Neuromag, Helsinki). This system is equipped with 102 magnetometers and 204 orthogonal planar gradiometers inside a light magnetically shielded room. During acquisition, an online filter was applied between around 0.03 Hz and 1000Hz. To support offline artifact correction, vertical and horizontal electrooculogram (VEOG, HEOG) as well as electrocardiogram (ECG) signal was concomitantly recorded. Four Head-Position Indicator (HPI) coils were used to track head motion. For subsequent source-localization the head shape was digitized. The recording lasted about eight minutes. For additional details on MEG acquisition, please consider the reference publications on the Cam-CAN dataset (Taylor et al., 2017; Shafto et al., 2014).

*MNE model for regression with source localization.* To compare the data-driven statistical models against a biophysics-informed method, for this dataset, we included a regression pipeline based on anatomically constrained minimum norm estimates (MNE) informed by the individual anatomy. Following common practice using the MNE software we used $Q = 8196$ candidate dipoles positioned on the cortical surface, and set the regularization parameter to $1/9$ (Gramfort et al., 2014). Concretely, we used the MNE inverse operator as any other spatial filter by multiplying the covariance with it from both sides. We then retained the diagonal elements which provides estimates of the source power. To obtain spatial smoothing and reduce dimensionality, we averaged the MNE solution using a cortical parcellation encompassing

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Definition of frequency bands

| name | low | $\delta$ | $\theta$ | $\alpha$ | $\beta_{low}$ | $\beta_{high}$ | $\gamma_{low}$ | $\gamma_{mid}$ | $\gamma_{high}$ |
|---|---|---|---|---|---|---|---|---|---|
| range (Hz) | 0.1 – 1.5 | 1.5 – 4 | 4 – 8 | 8 – 15 | 15 – 26 | 26 – 35 | 35 – 50 | 50 – 74 | 76 – 120 |

448 regions of interest from Khan et al. (2018). For preprocessing of structural MRI data we used the FreeSurfer software (Fischl 2012, `http://surfer.nmr.mgh.harvard.edu/`).

*Data processing and feature engineering.* This large dataset required more extensive data processing. We composed the preprocessing pipeline following current good practice recommendations (Gross et al., 2013; Jas et al., 2018). The steps are depicted in Fig. 2 in the broader context of the predictive modeling pipeline. The full procedure comprised the following steps: suppression of environmental artifacts, suppression of physiological artifacts (EOG/ECG) and rejection of remaining contaminated data segments.

To mitigate contamination by high-amplitude environmental magnetic fields, we applied the signal space separation method (SSS) (Taulu and Kajola, 2005). SSS decomposes the MEG signal into extracranial and intracranial sources and renders the data rank-deficient. Once applied, magnetometers and gradiometers become linear combinations of approximately 70 common SSS components, hence, become interchangeable (Garcés et al., 2017). For simplicity, we conducted all analyses on magnetometers. We kept the default settings of eight and three components for harmonic decomposition of internal and external sources, respectively, in concert with a 10-second sliding windows (temporal SSS). To discard segments in which inner and outer signal components were poorly separated, we applied a correlation-threshold of 98%. Since no continuous head monitoring data were available at the time of our study, we performed no movement compensation. The origin of internal and external multipolar moment space was estimated based on the head-digitization. To mitigate ocular and cardiac artifacts, we applied the signal space projection method (SSP) (Uusitalo and Ilmoniemi, 1997). This method learns principal components on data-segments contaminated by artifacts and then projects the signal into to the subspace orthogonal to the artifact. To reliably estimate the signal space dominated by the cardiac and ocular artifacts, we excluded data segments dominated by high-amplitude signals using the 'global' option from autoreject (Jas et al., 2017). To preserve the signal as much as possible, we only considered the first SSP vector based on the first principal component. As a final preprocessing step, we used the 'global' option from autoreject for adaptive computation of rejection thresholds to remove remaining high-amplitudes from the data at the epoching stage.

As the most important source of variance is not a priori known for the problem of age prediction, we considered a wide range of frequencies. We then bandpass filtered the data into nine conventional frequency bands (*cf.* Tab. 1) adapted from the Human-Connectome Project (Larson-Prior et al., 2013), and computed the band-limited covariance matrices with the OAS estimator (Chen et al., 2010). All pipelines were separately run across frequencies and features were concatenated after the vectorization step (see Fig. 2).

*Model evaluation.* We performed Monte Carlo (shuffle split) cross-validation with 100 splits and 10% testing data. The initialization of the random number generator used for cross-validation was fixed, ensuring identical train-test splits across models. This choice also allowed us to obtain more fine-grained uncertainty estimates than was possible with the time-series data used for within-subject regression. As absolute changes of the unit of the target is meaningful, we used the mean absolute error (MAE) as evaluation metric.

### *2.8.3. Statistical modeling*

We used ridge regression (Hoerl and Kennard, 1970) to predict from the vectorized covariance matrices and tuned its regularization parameter by generalized cross-validation (Golub et al., 1979) on a logarithmic grid of 100 values in $[10^{-5}, 10^{3}]$ on each training fold of a 10-fold cross-validation loop. For each model described in previous sections ('diag', 'upper', SPoC, Riemann, MNE), we standardized the features enforcing 0 mean and unit variance. This preprocessing step is standard for penalized linear models. To compare models against chance, we estimated the chance-level empirically through a dummy-regressor predicting the mean of the training data. Uncertainty estimation was obtained from the cross-validation distribution. Note that formal hypotheses testing for model comparison was not available for any of the datasets analyzed as this would have required several datasets, such that each average cross-validation score would have made one observation.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To save computation time, across all analyses, additional shrinkage for spatial filtering with SPoC and unsupervised was fixed at the mean of the value ranges tested in Sabbagh et al. (2019) *i.e.* 0.5 and $10^{-5}$, respectively.

### *2.8.4. Software*

All analyses were performed in Python 3.7 using the Scikit-Learn software (Pedregosa et al., 2011), the MNE software for processing M/EEG data (Gramfort et al., 2014) and the PyRiemann package (Congedo et al., 2013). We used the R-programming language and its ecosystem for visualizing the results (R Core Team, 2019; Allaire et al., 2019; Wickham, 2016; Clarke and Sherrill-Mix, 2017). Code used for data analysis can be found on GitHub[1].

## 3. Results

### *3.1. Simulated data*

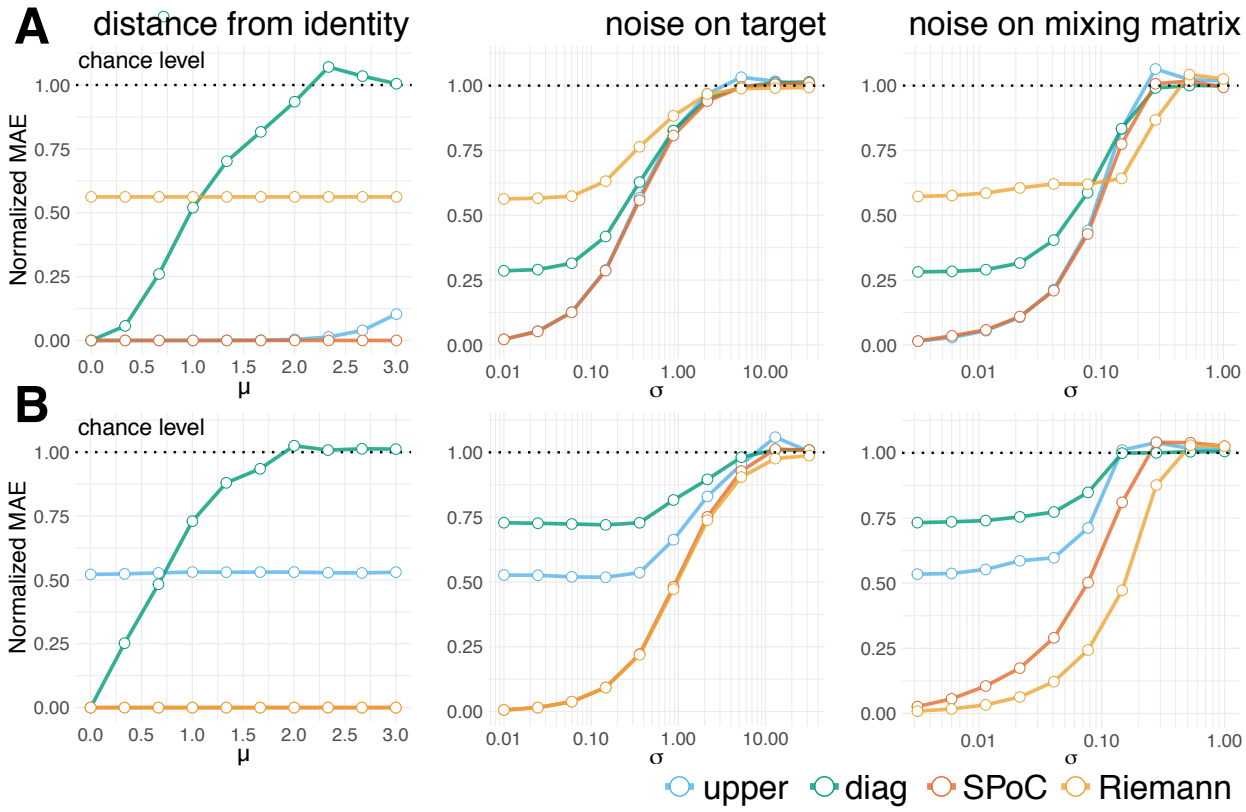


Figure 3: **Simulation-based model comparison across generative models** We focused on four regression models (indicated by color) each of which learnt from the covariance in distinct ways. The simulations performance across three types of model violations: the distance $\boldsymbol{\mu}$ between the mixing matrix $\boldsymbol{A}$ and the identity matrix $\boldsymbol{I}$ (left), noise on the target $y$ (middle) and individual noise on $\boldsymbol{A}$ (right). (**A**) Results for generative model in which y depends linearly on source variance. All but the Riemannian model achieve consistent regression when no mixing occurs (left). SPoC remained consistent throughout the simulated range. The 'upper' and SPoC models performed best as noise on the target (center) and noise on $\boldsymbol{A}$ (right) increased. (**B**) Results for generative model in which y depends log-linearly on source powers. The SPoC and Riemannian models achieve consistent regression across all simulated values (left). Likewise, both methods are more robust to noise on $y$ (center). Finally, the Riemannian model is most resilient to individual noise on the mixing matrix $\boldsymbol{A}$ (right).

We simulated data according to the linear and log-linear generative models and compared the performance of the proposed approaches. Fig. 3**A** displays the results for the linear generative model (4). The left figure shows the scores

[1] `https://github.com/DavidSabbagh/meeg_power_regression.git`

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of each method when the parameters $\mu$ controlling the distance from $A$ to $I_p$ is changed, and when the target variable $y$ is not corrupted by noise. We see that the Riemannian method is not affected by $\mu$ due to the affine invariance property of the geometric distance. At the same time, it is not the correct method for this generative model as is revealed by its considerable prediction error. Unsurprisingly, the 'diag' method is sensitive to changes in $A$, showing that is it not the right feature extraction method. On the contrary, both 'upper' and SPoC methods lead to perfect out-of-sample prediction (MAE = 0) even when the mixing matrix $\boldsymbol{A}$ is different from identity. This demonstrates consistency of these methods for the linear generative model. They both transparently access the statistical sources, either by being blind to the mixing matrix $\boldsymbol{A}$ ('upper') or by explicitly inverting it (SPoC). Hence they allow to perform regression on M/EEG brain signals without source localization. When we add noise in the target variable $y$ (middle) or individual noise in the mixing matrix (right) we have no theoretical guarantee of optimality for those methods. Yet, we see that both 'upper' and SPoC are equally sensitive to these model violations. The Riemannian method seems to be more robust than any other methods to individual noise in $\boldsymbol{A}$.

Fig. 3**B** displays the results for the log-linear generative model 5. In this case Riemann and SPoC performed best (left), the latter due to the linearity in the tangent space for the geometric distance. Both were equally sensitive to noise in target variable (middle) but, again, the Riemann method was more robust than others when we add individual noise in the mixing matrix (right). The simulations show that, under these idealized circumstances, 'upper', and SPoC are equivalent when the target $y$ depends linearly on source powers. When $y$ depends linearly on the log-powers, SPoC and Riemann are equivalent. However, when mixing matrices are individual, Riemann may be the best default choice, irrespective of the generative model of $y$.

### *3.2. Within-subject regression with MEG: Predicting continuous muscle contraction*

We then probed the regression models capitalizing on MEG data where the true data-generating mechanism is not a priori clear and multiple model violations may occur simultaneously. In a first step, we considered a problem where the unit of observation was individual behavior of one single subject with some unknown amount of noise affecting the measurement of the target. We analyzed an MEG dataset which provides concomitant EMG recordings that we chose as target (See section *Within-subject regression: cortico-muscular coherence* in Methods for details). At the time of the analysis, individual anatomical data was not available, hence we constrained the analysis to the sensor-space. The results are depicted in Fig. 4. The analysis revealed that only models including the cross-terms of the covariance predicted visibly better than chance (Fig. 4**A**). Importantly, extensive search for model order in the methods with projection step revealed important low-rank optima for SPoC and Riemann (Fig. 4**B**) with performance around 50% variance explained on unseen data. This is not surprising when considering the difficulty of accurate covariance estimation from limited data. Indeed, low-rank projection is one important method in biased estimation of covariance (Engemann and Gramfort, 2015). Interestingly, SPoC showed stronger performance with fewer components than Riemann, which is well explained by SPoC involving supervised projection which concentrate target-relevant information on fewer dimensions. However, it remains equivocal which statistical model best matches this regression problem. The best performing models all implied the log-linear model. Yet, compared to the linear-in-power 'upper' model, the low-rank SPoC and Riemann models also implied massive shrinkage on the covariances, leaving unclear if the type of model or biased covariance estimation explains their superior performance.

### *3.3. Cross-subject regression with MEG: Predicting age*

We then turned our attention to a regression problem that imposes the important model violation of varying source geometry due to individual anatomy while providing a clean target with virtually no measurement noise: predicting age from MEG. We analyzed resting-state MEG from about 600 subjects of the Cam-CAN dataset (Shafto et al., 2014; Taylor et al., 2017) focusing on the power spectral topography and between-sensor covariance of nine frequency bands as features (for details see Table 1). In this problem, each sample consists of MEG signals recorded from different persons, hence different brains. On theoretical grounds, one may therefore expect individual cortical folding, size and proportional composition of the head and its tissues to induce important distortions to the signal that may pose problems to data-driven approaches. Put technically, here, each sample can be said to have its own mixing matrix that contributes individual distortions to each observation. To investigate this point explicitly, here we conducted source localization to obtain topographic power estimates that corrected for individual head geometry based on biophysical prior knowledge. On the other hand, 8 minutes of MEG support accurate covariance estimation, hence, rendering

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

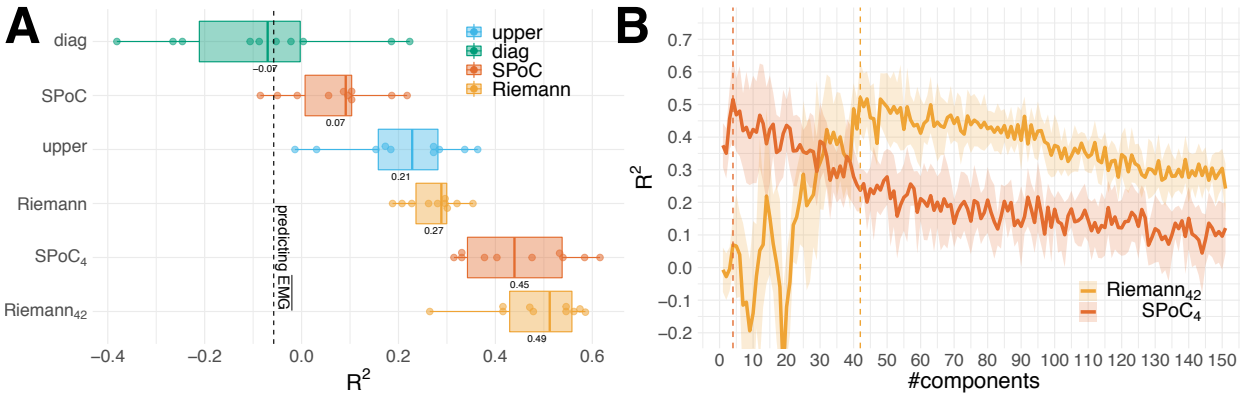


Figure 4: **Predicting continuous muscular activity on single-subject MEG. (A)** model comparison using cross-validation with 10 consecutive groups of approximately overlapping 80 epochs from one single subject. Models are depicted along the y-axis, expected out-of-sample performance ($R^2$) on the x-axis. The distribution is summarized by standard boxplots. Split-wise prediction scores are represented by dots. The model type is indicated by color. As covariance estimation is necessarily inaccurate with the short 1.5 second epochs, models may perform better when fit on a reduced subspace of the covariance. For SPoC and Riemann we, hence, reported alternative low-rank models (model order indicated by subscripts). **(B)** Exhaustive search for model order in pipelines with projection step. All values from 1 to the total number of 151 gradiometer channels were considered. One can spot well defined low-rank regimes in both models. However, SPoC supports a lower model order than Riemann. Only models explicitly considering the between-sensor correlation were successful. The best performance was achieved when projecting into a lower dimensional space with optima for SPoC and Riemann of order 4 and 42, respectively.

model order search less important for shrinkage. Fig. 5 displays the results for different regression models. The analysis revealed that all models performed clearly better than chance. The biophysically informed MNE regression model yielded the best performance (7.5y MAE), followed by the purely data-driven Riemannian method (8.0y MAE) and SPoC (8.8y MAE) (5**A**). The diagonal and upper-triangle models performed worse. Model order search did not reveal striking low-rank optima. Models above rank 40 seem approximately equivalent, especially when considering the estimation uncertainty of standard deviation above 1 year of MAE. For both SPoC and Riemann, the best low-rank model was close to the model at the theoretically derived rank of 65 (due to preprocessing with SSS, see Taulu and Kajola 2005). For subsequent analyses, we, nevertheless, retained the best models.

Strikingly, the only model not implementing a log transform, the 'upper' model, performed clearly worse than any other model. This suggests that the log-linear model is more appropriate in this regression problem. Yet, important difference in performance remain to be explained among the log-linear models. When considering how these methods treat the cross-terms, a more complex picture emerges. The 'diag' model ignores the cross-terms and performed least well among the log-linear models. The SPoC and Riemann models performed better than 'diag' and both analyzed the cross-terms, SPoC implicitly through the spatial filters. MNE performed best, but did not access the cross-terms of the covariance. Indeed, its spatial filter is constructed from the individual anatomy and some spatial whitening derived from the empty room recording, and not just the data covariance. This suggests that the cross-term models have learnt to some extent what MNE receives explicitly from individual anatomy, which raises the question of how sensitive the log-linear models are to anatomical information.

*Error decomposition.* To disentangle the factors explaining model performance, we devised a novel error-decomposition method derived from the proposed statistical framework (Fig. 1). Using a simulation-based approach, we computed degraded observations, *i.e.* , individual covariance matrices, that were either exclusively influenced by the individual anatomy in terms of the leadfields or also by additive uniform power (see section *Model-inspection by error-decomposition* in Methods for details). This has allowed us to estimate to which extent the log-linear models have learnt from anatomical information, global signal power of the MEG and topographic details. Fig. 6 compares three log-linear models based on the partially simulated observations: the 'diag' model and the best low-rank models previously found for SPoC and Riemann methods. One can see that all three error components improved overall prediction in similar ways, each improving performance between 2 and 4 years on average (Fig. 6**A**). Moreover, choice of algorithm mattered across all levels of the data generating scenario with Riemann leading and the 'diag' model trailing (Fig. 6**A**). Finally, the results suggest the presence of an interaction effect where both the leadfields and the uniform power components

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

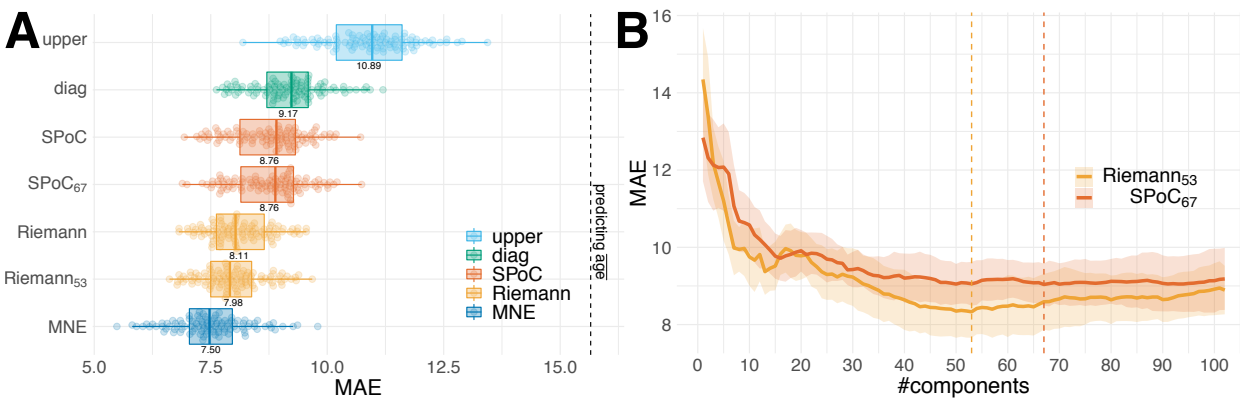


Figure 5: **Predicting age from cross-subject MEG. (A)** model comparison using Monte Carlo cross-validation with 100 splits sampled from 596 subjects. Models are depicted along the y-axis, expected out-of-sample performance (mean absolute error) on the x-axis. The distribution is summarized by standard boxplots. Split-wise prediction scores are depicted by dots. The model type is indicated by color. Here, covariance estimation was based on 8 minutes of MEG, hence, the impact of shrinkage should be small. For comparison with the single-subject data (Fig. 4), we nevertheless reported the alternative low-rank models (model order indicated by subscripts). **(B)** Exhaustive search for model order in pipelines with projection step. All values from 1 to the total number of 102 magnetometer channels were considered. One can see that performance starts to saturate around 40 to 50. No striking advantage of model order search was evident compared to deriving the order from prior knowledge on rank deficiency at a value of about 65. All models performed better than chance, however, models consistent with log-linear model and using correlation terms performed better. The Riemannian models performed closest to MNE that exploits biophysical prior knowledge.

were not equally important across models (Fig. 6**A,B**). For the Riemannian model, when only learning from anatomy, performance got as close as three years to the final performance of the full model (Fig. 6**B**). The 'diag' model, instead, only arrived at 5 years of distance from the equivalent model with full observations (Fig. 6**B**). On the other hand, the Riemannian model learnt rather little from the uniform power and only made its next leap forward when accessing topographic detail present in the full data but not in the degraded observations. Please note that these analyses are based on cross-validation. The resulting resampling splits do not count as independent samples. This precludes formal analysis of variance with an ANOVA model.

Overall, error decomposition suggests that all methods learn from anatomy. As predicted, models considering cross-terms were more sensitive. Strikingly, Riemannian model learnt more than any other model from the anatomical information, which may explain why it gets closest to MNE.

*Robustness to preprocessing choices.* This leads to a final consideration about error components. Commonly used preprocessing in M/EEG analysis is based on the idea to enhance signal-to-noise ratio by removing signals of non-interest, often using dedicated signal-space decomposition techniques (Uusitalo and Ilmoniemi, 1997; Taulu and Kajola, 2005; Hyvärinen et al., 2004). However, it is perfectly imaginable that such preprocessing removes information actually useful for predicting. At the same time, predictive models may learn the signal subspace implicitly, which could render preprocessing unnecessary. To investigate this issue for the current cross-subject regression problem, we sequentially repeated the analysis after activating the essential preprocessing steps one by one, and compared them to the baseline of extracting the features from the raw data. For this purpose, we considered an alternative preprocessing pipeline in which we kept all step unchanged but the SSS (Taulu and Kajola, 2005) for removal of environmental artifacts. We used instead a data-driven PCA-based SSP (Uusitalo and Ilmoniemi, 1997) computed on empty room recordings. The preprocessing pipeline is detailed in section *Cross-subject regression: age prediction*. Results are depicted in Fig. 7. The analysis revealed that the Riemannian model performed reasonably well when no preprocessing was done at all (Fig. 7**A**). It also turned out to be relatively robust to particular preprocessing choices. On the other hand, whether preprocessing was done or not turned out decisive for the 'diag' model and to some extent for the SPoC model (Fig. 7**B,C**). A few common tendencies became apparent. Across all models, while improving above baseline, SSP consistently led to worse performance than SSS. Second, performance was also slightly degraded by removing ocular and cardiac artifacts, suggesting that both shared variance with age. Removing EOG seemed to consistently degrade performance. On the other hand, removing ECG had virtually no impact for SPoC and the 'diag' model. For Riemann, both removing ECG and EOG additively deteriorated performance. Final bad epochs rejection had a negligible and

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

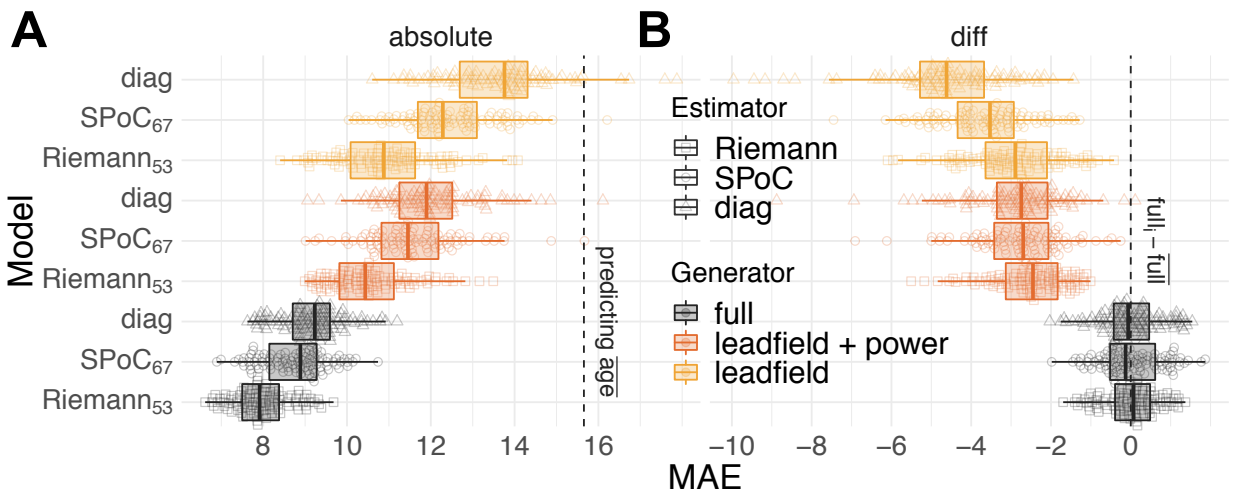


Figure 6: **Simulation-based error decomposition**. Model performance for observed data ('full') and partially simulated data for which spatio-spectral information was progressively removed: 'leadfield + power' muted topographic information keeping only spatially uniform power and information from the individual leadfields, 'leadfield' muted all electrophysiological variance. Data generating scenarios are represented by color. Shape indicates the regression model. (**A**) depicts absolute performance, (**B**), differences with the full observation for each model. One can see that all models learnt to predict age from all three components: anatomical variation across subjects, electrophysiological signal power and topographic information. However, the relative importance of each error component was clearly different across models. Riemannian model was most responsive to the leadfield component and least responsive to the uniform power.

inconsistent effect. Overall, the results suggest that the importance of preprocessing depended on the model, while minimal denoising with SSP or SSS always helped improve performance.

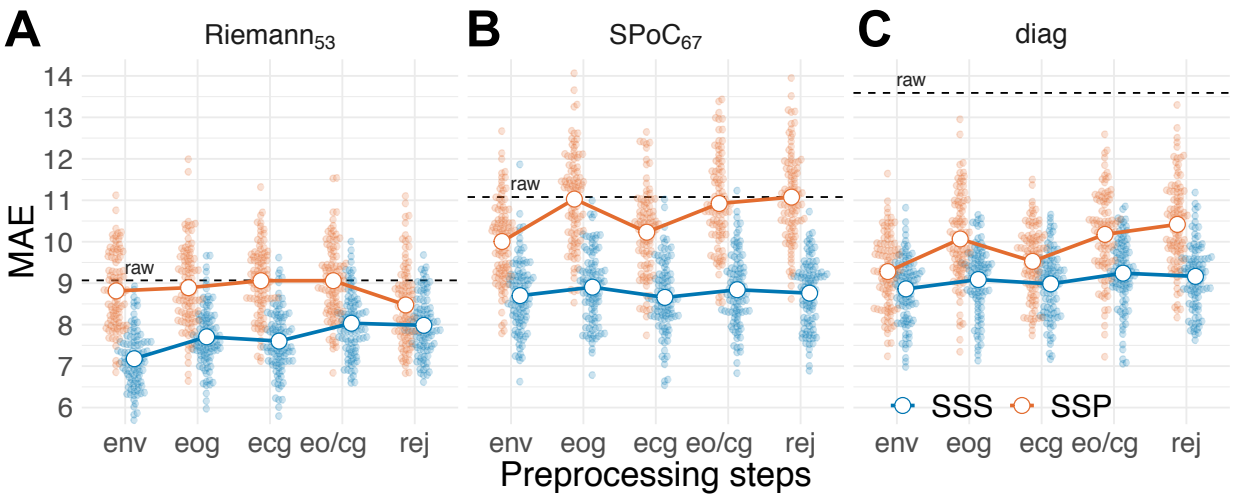


Figure 7: **Impact of preprocessing.** Model comparison across cumulative artifact removal steps: environmental artifacts (env), environmental + occular (eog), environmental + cardiac (ecg), environmental + occular + cardiac (eo/cg), environmental + occular + cardiac + bad segments (rej). Results are compared to the baseline of extracting features from raw data with no preprocessing (depicted by vertical dashed lines). The method for removal of environmental artifacts is indicated by color, *i.e.* , blue and red for SSS and SSP respectively. Note that the endpoint `rej` is identical to the full preprocessing conducted in previous analyses. Panels depict performance for the best Riemannian model (**A**), the best SPoC model (**B**), and the 'diag' model (**C**). One can see that the Riemann model, but not the 'log-diag' model, is relatively robust to preprocessing and its details.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## 4. Discussion

In this work, we have proposed a biophysics-aware framework for predictive modeling with M/EEG. We specifically considered regression tasks based on source power in which source localization is not practical. To the best of our knowledge, this is the first study systematically comparing alternative approaches for predicting continuous targets from M/EEG in a coherent theoretical framework from the level of mathematical analysis over simulations down to analysis of real data. Here, we focused on the band-limited between-channels covariance as fundamental representation and investigated distinct regression models. Mathematical analysis identified different models supporting perfect prediction under ideal circumstances when the target is either linear or log-linear in the source power. We adapted techniques originating from within-person prediction *i.e.* the SPoC spatial filtering approach (Dähne et al., 2014) and projection with Riemannian geometry (Congedo et al., 2017) for cross-person prediction typically encountered in biomarker development. Our simulation-based findings were consistent with the mathematical analysis and suggested that the regression models based on adaptive spatial filtering or Riemannian geometry were more robust across data generating scenarios and model violations. Subsequent analysis of MEG focused on a) differences between models in within-subject and cross-subject prediction, b) the contribution of anatomical and electrophysiological factors, for which we proposed a novel error decomposition method, and c) the impact of preprocessing, investigated through large-scale analysis of M/EEG data. Our findings suggest that, consistent with simulations, Riemannian methods are generally a good bet across a wide range of settings with considerable robustness to different choices of preprocessing.

### *4.1. What distinguishes within-subject from cross-subject prediction in the light of model violations?*

Unsurprisingly, no model performed perfectly when applied to MEG data in which the actual data generating mechanism is by definition unobservable, multiple model violations may occur and information is only partially available. One important source of differences in model violation is related to whether targets are defined at the single-subject level or across-subjects. When predicting targets from ongoing segments of neural time-series within a subject, covariance estimation becomes non-trivial as the event-related time windows are too short for accurate estimation. Even if regularized covariance estimates provide an effective remedy, there is not one shrinkage recipe that works in every situation (Engemann and Gramfort, 2015). In this study, we have relied on the oracle approximating shrinkage (OAS) (Chen et al., 2010) as a default method in all analyses. Yet, we found that additional low-rank shrinkage (Engemann and Gramfort, 2015; Woolrich et al., 2011; Tipping and Bishop, 1999; Rodrigues et al., 2018), as implied by the SPoC method (Dähne et al., 2014), or the unsupervised projection for the Riemannian model (Sabbagh et al., 2019), improved performance considerably for event-related prediction. A spatial-filter method like SPoC (Dähne et al., 2014) can be particularly convenient in this context. By design, it concentrates the variance most important for prediction on a few dimensions, which can be easily searched for, ascending from the bottom of the rank spectrum. Riemannian methods can also be operated in low-rank settings (Sabbagh et al., 2019). However, model order search may be more complicated as the best model may be anywhere in the spectrum. This can lead to increased computation times, which may be prohibitive in realtime settings such as BCI (Lotte et al., 2007; Tangermann et al., 2008).

Issues with the numerical rank of the covariance matrix also appear when predicting across subjects. The reason for this is fundamentally different and rather unrelated to the quality of covariance estimation. Many modern M/EEG preprocessing techniques focus on estimating and projecting out the noise-subspace, which leads to rank-deficient data. In our analysis of the Cam-CAN dataset (Shafto et al., 2014; Taylor et al., 2017), we applied the SSS method (Taulu and Kajola, 2005) by default, which is the recommended way when no strong magnetic shielding is available, as is the case for the Cambridge MEG-system on which the data was acquired (see also discussion in Jas et al. 2018). However, SSS reduces the rank down to 20% of the original input dimensionality, which may demand special attention when calibrating covariance estimation. Our results suggest that projection can indeed lead to slightly improved average prediction once a certain rank value is reached. Yet, thoughtful search of optimal model order may not be worth the effort in practice when a reasonably good guess of model order can be derived from the understanding of the preprocessing steps applied. Our findings, moreover, suggest, that a Riemann-based model is, in general, a reasonably good starting point, even when no model order search is applied. What seems to be a much more important issue in cross-subject prediction from M/EEG are the model violations incurred by individual anatomy its implied variation of source geometry. Our mathematical analysis and simulations demonstrated that not even the Riemannian is immune to those.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

### *4.2. What explains the performance in cross-subject prediction?*

Our results suggested that for the current regression problems the log-linear model was more appropriate than the linear-in-powers ones. This is well in line with practical experience and theoretical results highlighting the importance of log-normal brain dynamics (Buzsáki and Mizuseki, 2014). On the other hand, we observed substantive differences in performance within the log-normal models highlighting a non-trivial link between the cross-terms of the covariance and cross-subject variation. Indeed, the 'diag' model and the MNE model ignored the cross-terms of the covariance, yet MNE performed about 1.5 years better on average. This is rather unsurprising when recapitulating the fact that cross-subject regression on M/EEG implies individual anatomy. Indeed, our mathematical analysis and simulations identified this factor as important model violation. MNE source localization, by design, uses the head and brain geometry to correct for such violations. This suggests that the cross-term models that were more successful than the 'diag' model may be due to learning from the MEG data what MNE gets from the anatomical MRI.

In this work we, therefore, derived a novel error-decomposition technique from the statistical framework presented in Fig. 1 to estimate the sensitivity of our M/EEG regression models to anatomy, spatially uniform power and topographic details. We applied this technique on the Cam-CAN dataset to investigate the cross-person prediction problem. Strikingly, all models learnt from anatomical information but the Riemannian models were the most sensitive to it. This also suggests that MEG captures age-related anatomical information from the individual leadfields and imminently raises the question which aspects of anatomy are concerned. Neuroscience of aging has suggested important alterations of the cortical tissues (Liem et al., 2017) that also generate M/EEG signals, such as cortical surface area, cortical thickness or cortical folding. Yet, more trivially, head size or posture are a common issue in MEG and could explain the present effect, which would be potentially less fascinating from a neuroscientific standpoint. We investigated this issue post-hoc by predicting age from the device-to-head transform describing the position of the head relative to the helmet and the coregistration transforms from head to MRI. Compared to the Riemannian model applied to the leadfields-only surrogate data, this resulted in three years lower performance of around 14 years error, which is close to the random guessing error and may at best explain the performance of the 'diag' model.

Interestingly, also the SPoC model was more sensitive to anatomy than the 'diag' model. This suggests that by learning adaptive spatial filters from the data to best predict age SPoC may implicitly also tune the model to the anatomical information conveyed by the leadfields. This seems even more plausible when considering that from a statistical standpoint, SPoC learns how to invert the mixing matrix $\boldsymbol{A}$ to get the statistical sources implied by the predictive model. This must necessarily yield a linear combination of the columns of $\boldsymbol{G}$. In other words, this means that SPoC does not learn to invert the leadfields $\boldsymbol{G}$ directly, otherwise the choice of the target would be irrelevant and any SPoC model would be equivalent. This also motivates the conjecture that differences between SPoC and Riemann should become smaller when the $\boldsymbol{G}_i$ are not correlated with the target (Riemann should still enjoy an advantage due to increased robustness to model violations) or even vanish when $\boldsymbol{G}$ is constant and no low-rank issues apply. The latter case is what we encountered in the event-related analysis where SPoC and Riemann where roughly on par, suggesting that both handled the distortions induced by $\boldsymbol{G}$.

The current analysis leaves it, unfortunately, equivocal by which mechanism the models learnt from the anatomy and why the Riemannian model was so much more proficient. As a speculation, one can imagine that changes in the leadfields translate into simple topographic displacements that the 'diag' model can easily capture, which would be in line with the performance of the 'diag' model on the leadfields-only surrogate data matching prediction from transformation matrices. With cross-terms included in the modeling, SPoC and, in particular, Riemann may better unravel the directions of variation with regard to the target by considering the entire geometry presented in the leadfields. Instead, for the case of the leadfields-only surrogates, SPoC is trying to find sources which literally do not exist, hence must offer a degraded view on $\boldsymbol{G}$.

Overall, our results suggest that Riemannian models may also be the right choice when the anatomy is correlated with the target and the primary goal is prediction. The enhanced sensitivity of the Riemannian model to source and head geometry may be precisely what brings them so close to MNE performance. As a conjecture, these properties may render Riemannian models particularly helpful in the case of EEG, where the leadfields should be less variable as the sensor cap is affixed to the head, which strongly limits variation due to head posture. Unfortunately, at the time of conducting the study, we did not have access to an EEG-dataset equivalent to Cam-CAN with age reported in high resolution, which leaves this point open for future work.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

### *4.3. How important is preprocessing for cross-subject prediction?*

It is up to now equivocal how important preprocessing is when performing predictive modeling. Some evidence suggests that preprocessing may be negligible when performing event-related decoding of evoked responses as a linear model may well learn to regress out the noise-subspace (Haufe et al., 2014). Our findings suggest a more complex situation when performing cross-subject regression from M/EEG signal power. Strikingly, performing no preprocessing was clearly reducing performance, for some models even dramatically, SPoC and in particular 'diag'. The Riemann model, on the other hand, was remarkably robust and performed even reasonably well without preprocessing. Among the preprocessing steps, the removal of environmental artifacts seemed to be most important and most of the time led to massive improvements in performance. Removing EOG and ECG artifacts mostly reduced performance suggesting that age-related information was present in EOG and ECG. For example, one can easily imagine that older subjects produced less blinks or showed different eye-movement patterns (Thavikulwat et al., 2015) and also cardiac activity may change across the lifespan (Attia et al., 2019).

Interestingly, our results suggest that the method used for preprocessing was highly important. In general, performance was clearly enhanced when SSS was used instead of SSP. Does this mean that SSP is a bad choice for removing environmental artifacts? Our results have to be interpreted carefully, as the situation is more complicated when considering how fundamentally different SSP and SSS are in terms of design. When performing SSS, one actually combines the information of independent gradiometer and magnetometer sensor arrays into one latent space of roughly 65 dimensions that is roughly 20% of the dimensionality of both sensor arrays (306 sensors in total). Even when analyzing the magnetometers only after SSS, one will also access the extra information from the gradiometers (Garcés et al., 2017). SSP on the other hand is less invasive and is applied separately to magnetometers and gradiometers. It commonly removes only few dimensions from the data, yielding a subspace greater than 280 in practice. Our results therefore conflate two effects: 1) learning from magnetometers and gradiometers versus learning from magnetometers only and 2) differences in strength of dimensionality reduction. To disentangle these factors, careful experimentation with more targeted comparisons is indicated. To be conclusive, such an effort may necessitate computations at the scale of weeks and should be investigated in a dedicated study. For what concerns the current results, the findings simply suggest that SSS is a convenient tool as it allows one to combine information from magnetometers and gradiometers into a subspace that is sufficiently compact to enable efficient parameter estimation. It is not clear though, if careful processing with SSP and learning on both sensors types would not lead to better results.

## 5. Conclusion

Our study has investigated learning continuous targets from M/EEG signal power from the perspective of generative models. Across datasets, the log-linear model turned out to be more appropriate. In the light of common empirical model violations and preprocessing options, models based on Riemannian geometry stood out in terms of performance and robustness. The overall performance level is remarkable when considering the simplicity of the model. Our results demonstrate that a Riemannian model can actually be used to perform end-to-end learning (Schirrmeister et al., 2017) involving nothing but signal filtering and covariance estimation and, importantly, without deep-learning (Roy et al., 2019). When using SSS, performance improves beyond the current benchmark set by the MNE model but probably not because of denoising but rather due to the addition of gradiometer information. This result has at least two important practical implications. First, this allows researchers and clinicians to quickly assess the limits of what they can hope to learn in an economical and eco-friendly fashion (Strubell et al., 2019). In this scenario, the Riemannian end-to-end model rapidly delivers an estimate of the overall performance that could be reached by extensive and long processing, hence, support practical decision making on whether a deeper analysis is worth the investment of time and resources. Second, this result suggests that if prediction is the priority, availability of MRI and precious MEG expertise for conducting source localization is not any longer the bottleneck. This could potentially facilitate data collection and shift the strategy towards betting on the law of large numbers: assembling an MEG dataset in the order of thousands is easier when collecting MRI is not a prerequisite.

It is worthwhile to consider important limitations of this study. Unfortunately, we have not had access to more datasets with other interesting continuous targets and additional modalities such as EEG. In particular the conclusions drawn from the comparison between within-subject regression and cross-subject regression may be expanded in the future when considering larger within-subject datasets and other targets for which the linear-in-powers model may be

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

more appropriate. Second, one has to critically acknowledge that the performance benefit for the Riemannian model may be partially explained by increased sensitivity to anatomical information, which might imply reduced specificity with regard to neuronal activity. In this context it is noteworthy that recent regression pipelines based on a variant of SPoC (Dähne et al., 2014) made use of additional spatial filtering for dimensionality reduction, *i.e.* , SSD (Nikulin et al., 2011) to isolate oscillatory components and discard arrhythmic (1/f) activity. This raises the question if the specificity of a Riemannian model could be enhanced in a similar way. Ultimately, what model to prefer, therefore, clearly depends on the strategic goal of the analysis (Bzdok et al., 2018; Bzdok and Ioannidis, 2019) and cannot be globally decided.

We hope that this study will provide the community with the theoretical framework and tools needed to deepen the study of regression on neural power spectra and safely navigate between regression models and geometric distortions governing M/EEG observations.

## Acknowledgements

This work was supported by a 2018 “médecine numérique” (for digital medicine) thesis grant issued by Inserm (French national institute of health and medical research) and Inria (French national research institute for the digital sciences). It was also partly supported by the European Research Council Starting Grant SLAB ERC-YStG-676943.

## References

G. Buzsáki, A. Draguhn, Neuronal oscillations in cortical networks, science 304 (2004) 1926–1929.

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

S. Baillet, Magnetoencephalography for brain electrophysiology and imaging, Nature Neuroscience 20 (2017) 327 EP –.

G. Buzsáki, B. O. Watson, Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease., Dialogues in clinical neuroscience 14 (2012) 345.

F. L. da Silva, Eeg and meg: relevance to neuroscience, Neuron 80 (2013) 1112–1128.

C.-W. Woo, L. J. Chang, M. A. Lindquist, T. D. Wager, Building better biomarkers: brain models in translational neuroimaging, Nature neuroscience 20 (2017) 365.

M. Besserve, K. Jerbi, F. Laurent, S. Baillet, J. Martinerie, L. Garnero, Classification methods for ongoing eeg and meg signals, Biological research 40 (2007) 415–437.

R. M. Cichy, F. M. Ramirez, D. Pantazis, Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans?, Neuroimage 121 (2015) 193–204.

J.-R. King, F. Faugeras, A. Gramfort, A. Schurger, I. El Karoui, J. Sitt, B. Rohaut, C. Wacongne, E. Labyt, T. Bekinschtein, et al., Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness, Neuroimage 83 (2013) 726–738.

J. R. Wolpaw, D. J. McFarland, G. W. Neat, C. A. Forneris, An eeg-based brain-computer interface for cursor control, Electroencephalography and clinical neurophysiology 78 (1991) 252–259.

F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, Journal of Neural Engineering 4 (2007) R1–R13.

M. W. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, C. Vidaurre, B. Blankertz, K.-R. Müller, Playing pinball with non-invasive bci, in: Proceedings of the 21st International Conference on Neural Information Processing Systems, Citeseer, pp. 1641–1648.

D. A. Engemann, F. Raimondo, J.-R. King, B. Rohaut, G. Louppe, F. Faugeras, J. Annen, H. Cassol, O. Gosseries, D. Fernandez-Slezak, S. Laureys, L. Naccache, S. Dehaene, J. D. Sitt, Robust EEG-based cross-site and cross-protocol classification of states of consciousness, Brain 141 (2018) 3179–3192.

S. Sami, N. Williams, L. E. Hughes, T. E. Cope, T. Rittman, I. T. Coyle-Gilchrist, R. N. Henson, J. B. Rowe, Neurophysiological signatures of alzheimer's disease and frontotemporal lobar degeneration: pathology versus phenotype, Brain 141 (2018) 2500–2510.

A. Mazaheri, K. Segaert, J. Olichney, J.-C. Yang, Y.-Q. Niu, K. Shapiro, H. Bowman, Eeg oscillations during word processing predict mci conversion to alzheimer's disease, NeuroImage: Clinical 17 (2018) 188–197.

V. van Wassenhove, Temporal cognition and neural oscillations, Current Opinion in Behavioral Sciences 8 (2016) 124–130.

T. Akam, D. M. Kullmann, Oscillatory multiplexing of population codes for selective communication in the mammalian brain, Nature Reviews Neuroscience 15 (2014) 111.

S. Panzeri, N. Brunel, N. K. Logothetis, C. Kayser, Sensory neural codes using multiplexed temporal scales, Trends in neurosciences 33 (2010) 111–120.

N. Dehghani, C. Bédard, S. S. Cash, E. Halgren, A. Destexhe, Comparative power spectral analysis of simultaneous elecroencephalographic and magnetoencephalographic recordings in humans suggests non-resistive extracellular media, Journal of computational neuroscience 29 (2010) 405–421.

K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, R. J. Ilmoniemi, Long-range temporal correlations and scaling behavior in human brain oscillations, Journal of Neuroscience 21 (2001) 1370–1377.

B. J. He, J. M. Zempel, A. Z. Snyder, M. E. Raichle, The temporal structures and functional significance of scale-free brain activity, Neuron 66 (2010) 353–369.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

J. F. Hipp, D. J. Hawellek, M. Corbetta, M. Siegel, A. K. Engel, Large-scale cortical correlation structure of spontaneous oscillatory activity, Nature neuroscience 15 (2012) 884.

M. Siegel, T. H. Donner, A. K. Engel, Spectral fingerprints of large-scale neuronal interactions, Nature Reviews Neuroscience 13 (2012) 121.

C. J. Honey, R. Kötter, M. Breakspear, O. Sporns, Network structure of cerebral cortex shapes functional connectivity on multiple time scales, Proceedings of the National Academy of Sciences 104 (2007) 10240–10245.

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

S. Khan, J. A. Hashmi, F. Mamashli, K. Michmizos, M. G. Kitzbichler, H. Bharadwaj, Y. Bekhti, S. Ganesan, K.-L. A. Garel, S. Whitfield-Gabrieli, et al., Maturation trajectories of cortical resting-state networks depend on the mediating frequency band, NeuroImage 174 (2018) 57–68.

B. U. Westner, S. S. Dalal, S. Hanslmayr, T. Staudigl, Across-subjects classification of stimulus modality from human meg high frequency activity, PLoS computational biology 14 (2018) e1005938.

A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural networks 13 (2000) 411–430.

S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, Independent component analysis of electroencephalographic data, in: Proceedings of the 8th International Conference on Neural Information Processing Systems, NIPS'95, MIT Press, Cambridge, MA, USA, 1995, pp. 145–151.

A. X. Stewart, A. Nuthmann, G. Sanguinetti, Single-trial classification of eeg in a visual object task using ica and machine learning, Journal of neuroscience methods 228 (2014) 1–14.

A. Subasi, M. I. Gursoy, Eeg signal classification using pca, ica, lda and support vector machines, Expert systems with applications 37 (2010) 8659–8666.

V. V. Nikulin, G. Nolte, G. Curio, A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition, NeuroImage 55 (2011) 1528–1535.

S. Dähne, V. V. Nikulin, D. Ramírez, P. J. Schreier, K.-R. Müller, S. Haufe, Finding brain oscillations with power dependencies in neuroimaging data, NeuroImage 96 (2014) 334–348.

S. Dähne, F. Bießmann, F. C. Meinecke, J. Mehnert, S. Fazli, K.-R. Müller, Integration of multivariate data streams with bandpower signals, IEEE Transactions on Multimedia 15 (2013) 1001–1013.

S. Haufe, S. Dähne, V. V. Nikulin, Dimensionality reduction for the analysis of brain oscillations, NeuroImage 101 (2014) 583–597.

S. Dähne, F. C. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K.-R. Müller, V. V. Nikulin, Spoc: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters, NeuroImage 86 (2014) 111–122.

M. G. Coles, M. D. Rugg, Event-related brain potentials: An introduction., Oxford University Press, 1995.

R. Näätänen, Selective attention and evoked potentials inhumans—a critical review, Biological Psychology 2 (1975) 237–307.

J. Polich, A. Kok, Cognitive and biological determinants of p300: an integrative review, Biological psychology 41 (1995) 103–146.

T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, The elements of statistical learning: data mining, inference and prediction, The Mathematical Intelligencer 27 (2005) 83–85.

J. C. Mosher, P. S. Lewis, R. M. Leahy, Multiple dipole modeling and localization from spatio-temporal MEG data, IEEE Transactions on Biomedical Engineering 39 (1992) 541–557.

F.-H. Lin, J. W. Belliveau, A. M. Dale, M. S. Hämäläinen, Distributed current estimates using cortical orientation constraints, Human brain mapping 27 (2006) 1–13.

B. D. Van Veen, K. M. Buckley, Beamforming: A versatile approach to spatial filtering, IEEE assp magazine 5 (1988) 4–24.

O. Hauk, M. Stenroos, A framework for the design of flexible cross-talk functions for spatial filtering of eeg/meg data: Deflect, Human brain mapping 35 (2014) 1642–1653.

J. C. Mosher, R. M. Leahy, P. S. Lewis, EEG and MEG: forward solutions for inverse methods, IEEE Transactions on Biomedical Engineering 46 (1999) 245–259.

T. C. Kietzmann, C. J. Spoerer, L. K. A. Sörensen, R. M. Cichy, O. Hauk, N. Kriegeskorte, Recurrence is required to capture the representational dynamics of the human visual system, Proceedings of the National Academy of Sciences (2019).

S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, Independent component analysis of electroencephalographic data, in: Advances in neural information processing systems, pp. 145–151.

Y. Wang, S. Makeig, Predicting intended movement direction using eeg from human posterior parietal cortex, in: International Conference on Foundations of Augmented Cognition, Springer, pp. 437–446.

K. E. Hild II, S. S. Nagarajan, Source localization of eeg/meg data by correlating columns of ica and lead field matrices, IEEE Transactions on Biomedical Engineering 56 (2009) 2619–2626.

M. J. Brookes, M. Woolrich, H. Luckhoo, D. Price, J. R. Hale, M. C. Stephenson, G. R. Barnes, S. M. Smith, P. G. Morris, Investigating the electrophysiological basis of resting state networks using magnetoencephalography, Proceedings of the National Academy of Sciences 108 (2011) 16783–16788.

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

K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, Benchmarking functional connectome-based predictive models for resting-state fMRI, NeuroImage 192 (2019) 115–134.

M.-A. Schulz, T. Yeo, J. Vogelstein, J. Mourao-Miranada, J. Kather, K. Kording, B. A. Richards, D. Bzdok, Deep learning for brains?: Different linear and nonlinear scaling in uk biobank brain images vs. machine-learning datasets, bioRxiv (2019) 757054.

T. He, R. Kong, A. J. Holmes, M. Nguyen, M. R. Sabuncu, S. B. Eickhoff, D. Bzdok, J. Feng, B. T. Yeo, Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics, NeuroImage (2019) 116276.

L. M. Andersen, M. N. Pedersen, K. Sandberg, M. Overgaard, Occipital meg activity in the early time range (< 300 ms) predicts graded changes in perceptual consciousness, Cerebral Cortex 26 (2015) 2677–2688.

S. G. Wardle, N. Kriegeskorte, T. Grootswagers, S.-M. Khaligh-Razavi, T. A. Carlson, Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with meg, Neuroimage 132 (2016) 59–70.

S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, F. Bießmann, On the interpretation of weight vectors of linear models in multivariate neuroimaging, NeuroImage 87 (2014) 96 – 110.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

M. van Vliet, R. Salmelin, Post-hoc modification of linear models: Combining machine learning with domain information to make solid inferences from noisy data, NeuroImage (2019) 116221.

J.-R. King, L. Gwilliams, C. Holdgraf, J. Sassenhagen, A. Barachant, D. Engemann, E. Larson, A. Gramfort, Encoding and decoding neuronal dynamics: Methodological framework to uncover the algorithms of cognition, 2018.

J.-R. King, S. Dehaene, Characterizing the dynamics of mental representations: the temporal generalization method, Trends in cognitive sciences 18 (2014) 203–210.

M. G. Stokes, M. J. Wolff, E. Spaak, Decoding rich spatial information with high temporal resolution, Trends in cognitive sciences 19 (2015) 636–638.

C. Tallon-Baudry, O. Bertrand, Oscillatory gamma activity in humans and its role in object representation, Trends in cognitive sciences 3 (1999) 151–162.

G. Buzsáki, K. Mizuseki, The log-dynamic brain: how skewed distributions affect network operations, Nature Reviews Neuroscience 15 (2014) 264.

J. A. Roberts, T. W. Boonstra, M. Breakspear, The heavy tail of the human brain, Current opinion in neurobiology 31 (2015) 164–172.

Z. J. Koles, The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg, Electroencephalography and clinical Neurophysiology 79 (1991) 440–447.

J. P. Dmochowski, P. Sajda, J. Dias, L. C. Parra, Correlated components of ongoing eeg point to emotionally laden attention–a possible marker of engagement?, Frontiers in human neuroscience 6 (2012) 112.

H. Hotelling, Relations between two sets of variates, in: Breakthroughs in statistics, Springer, 1992, pp. 162–190.

A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by riemannian geometry, IEEE Transactions on Biomedical Engineering 59 (2011) 920–928.

A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a riemannian-based kernel for bci applications, Neurocomputing 112 (2013) 172–178.

F. Yger, M. Berar, F. Lotte, Riemannian approaches in brain-computer interfaces: A review, IEEE Transactions on Neural Systems and Rehabilitation Engineering 25 (2017) 1753–1762.

M. Hämäläinen, R. Ilmoniemi, Interpreting magnetic fields of the brain: minimum norm estimates, Technical Report TKK-F-A559, Helsinki University of Technology, 1984.

S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, T. J. Sejnowski, Blind separation of auditory event-related brain responses into independent components, Proceedings of the National Academy of Sciences 94 (1997) 10979–10984.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. Muller, Optimizing spatial filters for robust eeg single-trial analysis, IEEE Signal Processing Magazine 25 (2008) 41–56.

M. Grosse-Wentrup*, M. Buss, Multiclass common spatial patterns and information theoretic feature extraction, IEEE Transactions on Biomedical Engineering 55 (2008) 1991–2000.

D. A. Engemann, A. Gramfort, Automated model selection in covariance estimation and spatial whitening of meg and eeg signals, NeuroImage 108 (2015) 328–342.

P. L. Coelho Rodrigues, F. Bouchard, M. Congedo, C. Jutten, Dimensionality Reduction for BCI classification using Riemannian geometry, in: 7th Graz Brain-Computer Interface Conference (BCI 2017), Gernot R. Müller-Putz, Graz, Austria.

D. Sabbagh, P. Ablin, G. Varoquaux, A. Gramfort, D. Engemann, Manifold-regression to predict from meg/eeg brain signals without source modeling, 2019. ArXiv Preprint 1906.02687v2.

W. Förstner, B. Moonen, A metric for covariance matrices, in: Geodesy-The Challenge of the 3rd Millennium, Springer, 2003, pp. 299–309.

P.-A. Absil, R. Mahony, R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2009.

S. Bonnabel, R. Sepulchre, Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank, SIAM Journal on Matrix Analysis and Applications 31 (2009) 1055–1070.

A. Gelman, et al., Analysis of variance — why it is more important than ever, The annals of statistics 33 (2005) 1–53.

A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M. S. Hämäläinen, MNE software for processing MEG and EEG data, NeuroImage 86 (2014) 446–460.

J.-M. Schoffelen, J. Poort, R. Oostenveld, P. Fries, Selective movement preparation is subserved by selective increases in corticomuscular gamma-band coherence, Journal of Neuroscience 31 (2011) 6750–6758.

R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data, Computational intelligence and neuroscience 2011 (2011) 1.

Y. Chen, A. Wiesel, Y. C. Eldar, A. O. Hero, Shrinkage algorithms for mmse covariance estimation, IEEE Transactions on Signal Processing 58 (2010) 5016–5029.

G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, B. Thirion, Assessing and tuning brain decoders: cross-validation, caveats, and guidelines, NeuroImage 145 (2017) 166–179.

J. R. Taylor, N. Williams, R. Cusack, T. Auer, M. A. Shafto, M. Dixon, L. K. Tyler, R. N. Henson, et al., The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample, Neuroimage 144 (2017) 262–269.

M. A. Shafto, L. K. Tyler, M. Dixon, J. R. Taylor, J. B. Rowe, R. Cusack, A. J. Calder, W. D. Marslen-Wilson, J. Duncan, T. Dalgleish, et al., The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing, BMC neurology 14 (2014) 204.

B. Fischl, FreeSurfer, NeuroImage 62 (2012) 774–781.

J. Gross, S. Baillet, G. R. Barnes, R. N. Henson, A. Hillebrand, O. Jensen, K. Jerbi, V. Litvak, B. Maess, R. Oostenveld, et al., Good practice for conducting and reporting meg research, Neuroimage 65 (2013) 349–363.

M. Jas, E. Larson, D. A. Engemann, J. Leppäkangas, S. Taulu, M. Hämäläinen, A. Gramfort, A reproducible MEG/EEG group study with the MNE software: Recommendations, quality assessments, and good practices, Frontiers in Neuroscience 12 (2018) 1–18.

S. Taulu, M. Kajola, Presentation of electromagnetic multichannel data: the signal space separation method, Journal of Applied Physics 97 (2005) 124905.

bioRxiv preprint doi: https://doi.org/10.1101/845016; this version posted November 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

P. Garcés, D. López-Sanz, F. Maestú, E. Pereda, Choice of magnetometers and gradiometers after signal space separation, Sensors 17 (2017) 2926.
M. A. Uusitalo, R. J. Ilmoniemi, Signal-space projection method for separating MEG or EEG into components, Medical and Biological Engineering and Computing 35 (1997) 135–140.
M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, A. Gramfort, Autoreject: Automated artifact rejection for MEG and EEG data, NeuroImage 159 (2017) 417–429.
L. J. Larson-Prior, R. Oostenveld, S. Della Penna, G. Michalareas, F. Prior, A. Babajani-Feremi, J.-M. Schoffelen, L. Marzetti, F. de Pasquale, F. Di Pompeo, et al., Adding dynamics to the Human Connectome Project with MEG, Neuroimage 80 (2013) 190–201.
A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1970) 55–67.
G. H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–223.
F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.
M. Congedo, A. Barachant, A. Andreev, A new generation of brain-computer interface based on Riemannian geometry, arXiv e-prints (2013).
R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019.
J. Allaire, K. Ushey, Y. Tang, reticulate: Interface to 'Python', 2019. R package version 1.11.
H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2016.
E. Clarke, S. Sherrill-Mix, ggbeeswarm: Categorical Scatter (Violin Point) Plots, 2017. R package version 0.6.0.
A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, volume 46, John Wiley & Sons, 2004.
M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review, Brain-Computer Interfaces 4 (2017) 155–174.
M. Woolrich, L. Hunt, A. Groves, G. Barnes, Meg beamforming using bayesian pca for adaptive data covariance matrix regularization, Neuroimage 57 (2011) 1466–1479.
M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (1999) 611–622.
P. L. C. Rodrigues, M. Congedo, C. Jutten, Multivariate time-series analysis via manifold learning, in: 2018 IEEE Statistical Signal Processing Workshop (SSP), IEEE, pp. 573–577.
F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. K. Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, S. Riedel-Heller, T. Luck, M. Loeffler, M. L. Schroeter, A. V. Witte, A. Villringer, D. S. Margulies, Predicting brain-age from multimodal imaging data captures cognitive impairment, NeuroImage 148 (2017) 179 – 188.
A. T. Thavikulwat, P. Lopez, R. C. Caruso, B. G. Jeffrey, The effects of gender and age on the range of the normal human electro-oculogram, Documenta Ophthalmologica 131 (2015) 177–188.
Z. I. Attia, P. A. Friedman, P. A. Noseworthy, F. Lopez-Jimenez, D. J. Ladewig, G. Satam, P. A. Pellikka, T. M. Munger, S. J. Asirvatham, C. G. Scott, R. E. Carter, S. Kapa, Age and sex estimation using artificial intelligence from standard 12-lead ecgs, Circulation: Arrhythmia and Electrophysiology 12 (2019) e007284.
R. Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, T. Ball, Deep learning with convolutional neural networks for decoding and visualization of eeg pathology, in: 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), IEEE, pp. 1–7.
Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, J. Faubert, Deep learning-based electroencephalography analysis: a systematic review, Journal of Neural Engineering 16 (2019) 051001.
E. Strubell, A. Ganesh, A. McCallum, Energy and policy considerations for deep learning in nlp, arXiv preprint arXiv:1906.02243 (2019).
D. Bzdok, D. Engemann, O. Grisel, G. Varoquaux, B. Thirion, Prediction and inference diverge in biomedicine: Simulations and real-world data (2018).
D. Bzdok, J. P. Ioannidis, Exploration, inference, and prediction in neuroscience and biomedicine, Trends in neurosciences (2019).