Imaging Neuroscience
an open access journal

Research Article

Distorting anatomy to test MEG models and metrics

José David López[a,b], Yael Balbastre[c,d], John Ashburner[c], Alberto Mariola[c], James J. Bonaiuto[e,f], Gareth Barnes[c]

[a]Engineering Faculty, Universidad de Antioquia UDEA, Medellín, Colombia
[b]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland
[c]Department of Imaging Neuroscience, UCL Institute of Neurology, University College London, London, United Kingdom
[d]Department of Experimental Psychology, University College London, London, United Kingdom
[e]Institut de Sciences Cognitives Marc Jeannerod, CNRS, UMR5229, Lyon, France
[f]Université de Lyon, Université Claude Bernard Lyon 1, Lyon, France

Corresponding Author: Gareth Barnes (g.barnes@ucl.ac.uk)

ABSTRACT

The magnetoencephalographic and electroencephalographic (M/EEG) source reconstruction problem is an ill-posed model inversion, so it must be constrained by imposing biologically and physically plausible assumptions. Different M/EEG source reconstruction methods entail different assumptions about the underlying current distribution, yet all produce subjectively plausible current estimates. This work aims to develop an objective method that can be used to test any M/EEG analysis pathway. We make use of advances in diffeomorphic brain shape modeling to construct a set of parametrically deformable cortical surfaces that are representative of the population. These deformed (surrogate) brains provide a quantifiable parametric distortion from the ground-truth anatomy. If the current flow giving rise to the MEG data were generated on the true cortical manifold, MEG current estimates should be selective of the true anatomy. We show in simulation and with empirical data how the correct reconstruction assumptions depend closely on the true anatomy. We present a method to quantify the performance of MEG source reconstruction algorithms (and metrics of fit) in terms of millimeters of distortion.

Keywords: MEG/EEG brain imaging, diffeomorphic modeling, brain anatomy

1. INTRODUCTION

Magnetoencephalographic (MEG) and electroencephalographic (EEG) recordings enable reconstruction of spatio-temporal images of human brain function. Estimating these images from data measured outside the head requires additional assumptions, which are continually being refined, tested, and updated. We have proposed that anatomy (available from non-invasive Magnetic Resonance Imaging—MRI) provides a valuable ground truth against which to decide between assumption sets (Barnes et al., 2006; Little et al., 2018; López et al., 2017; Stevenson et al., 2014). We know that the primary current flow (which gives rise to the external magnetic fields) is bound (in the cell bodies of pyramidal neurons) to anatomy. We hypothesize that more accurate current flow models will be more likely given the correct, rather than incorrect, assumptions about anatomy.

The main issue addressed by this paper is the generation of realistic (brain-like) surrogate anatomy. Barnes et al. (2006) used rotated versions of the true anatomy within a spherical volume. Stevenson et al. (2014) and Little et al. (2018) progressively removed higher-order components from spherical harmonic approximations to the cortical manifold. While López et al. (2017) produced surrogate brains from linear mixtures of these harmonic components. All these cases, while providing proof of

Received: 1 April 2025 Revision: 4 March 2026 Accepted: 5 March 2026 Available Online: 11 March 2026

The MIT Press
© 2026 The Authors. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

principle, set a relatively low bar on algorithm performance as the surrogate brains were not anatomically realistic.

In contrast, Troebinger et al. (2014) found higher evidence for generative models (of the MEG data) based on an individual's true cortical anatomy, rather than another person's brain warped into the same individual's head. Here, we pursue this direction of using realistic surrogates but take advantage of recent advances in diffeomorphic brain shape modelling (Ashburner et al., 2019). This method was originally developed to enhance data mining of large MRI datasets, by parameterizing each structural MRI into a relatively small number of latent variables. Importantly, these latent variables, which describe a population based on large structural databases, are Gaussian. This means that it is possible to describe an individual's brain shape as a vector of values and construct perturbations of this vector that distort this brain within realistic limits. The deformations here are subtler than one might expect when exchanging brains (in which patterns of gyrification may change), as the individual topology is preserved.

The methods can be summarized as follows (Fig. 1). We create surrogate cortices that deviate increasingly from the true anatomy. We then reconstruct MEG data onto these cortices using different algorithms (or models of brain function) and fit metrics. We describe a method to summarize any processing pipeline in terms of its sensitivity to the ground truth anatomy in terms of distortion (in mm). Algorithms that favor large distortions, or do not care onto which anatomy current flow is reconstructed, are unlikely to be useful.

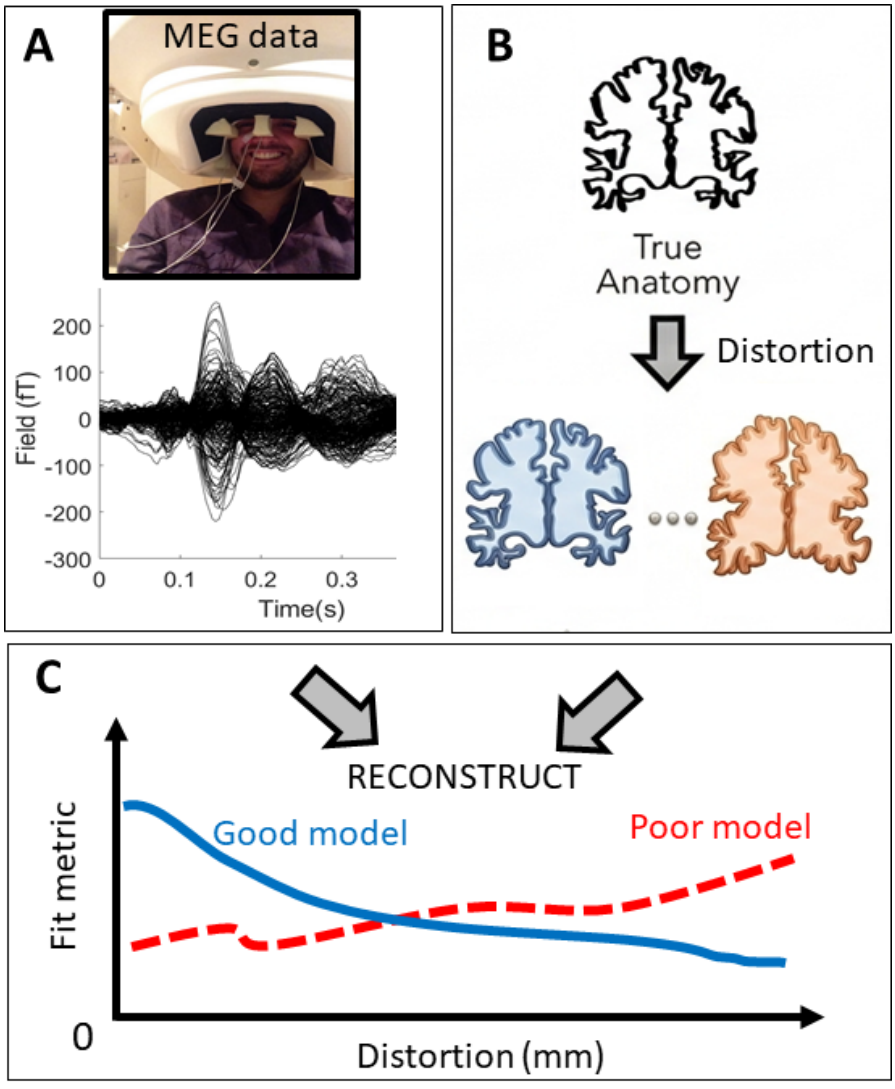

Fig. 1. Overview of paper. (A) We take evoked response data from an MEG head-cast experiment. (B) These data are then reconstructed (under different modelling assumptions) onto the true, then progressively more distorted, versions of the subject's cortical surface. (C) We plot a fit metric against distortion for each model. Models (and metrics) which reflect underlying brain function will be sensitive to the true anatomy; whereas poor models will not.

2. METHODS

We first describe the process of diffeomorphic modeling (or creating the surrogate brains). We then describe the parameter space we have chosen to work in, which involves stepping away from the true brain shape in different (pseudo-random) directions (in the high dimensional space).

2.1. Diffeomorphic modeling

The diffeomorphic algorithms used in this paper have been described extensively elsewhere (Balbastre et al., 2018). In summary, the shape of an individual's brain, described by its tissue probability map, can be parameterized into 100 components each with a Gaussian distribution. Average brain shape is located at zero in this high-dimensional space and any individual's brain shape can be described by a 100-component vector of z-scores. In this work, we extract the cortical mesh of an individual and then distort this mesh by perturbing it progressively through this space. A more comprehensive explanation follows.

Diffeomorphic shape modeling assumes that any individual anatomy can be described by a diffeomorphic (i.e., smooth and invertible) transform applied to a common template (Miller et al., 1993). Conversely, given a collection of N observed anatomies, an empirical template encoding the average anatomy of a population can be obtained by minimising its distance to all elements of the collection in a procedure known as geodesic shooting (Ashburner & Friston, 2011; Miller et al., 2006). In this framework, the diffeomorphism for subject n is encoded by an "initial velocity" $\mathbf{v}_n \in \mathbb{R}^{3M}$ (where M is the number of voxels) and the distance between anatomies is penalized on the squared gradients (or some other differential operator $\mathbf{D} \in \mathbb{R}^{F\times 3\,M}$ that describes smoothness) of this velocity field. After template construction, the posterior distribution of velocity fields can be efficiently estimated by performing a Bayesian residual component analysis (Bishop, 1999; Kalaitzis & Lawrence, 2011), a variant of principal component analysis that allows 'residual' variance (here, known smoothness) to be factored out of the projected variance when determining principal modes of variation. This procedure assumes that velocities can be factorized as $\mathbf{v}_n = \mathbf{W}\mathbf{z}_n + \boldsymbol{\varepsilon}_n$, where $\mathbf{W} \in \mathbb{R}^{3M\times K}$ contains $K \ll 3M$ principal components shared across individuals, $\mathbf{z}_n \in \mathbb{R}^K$ is the latent projection of the individual onto the principal basis, and $\boldsymbol{\varepsilon}_n$ is the residual vector with a covariance structure derived from the differential operator $\mathbf{D}$. The model further assumes that the latent coordinates stem from a standard Gaussian, the residuals have the same covariance structure as the differential regularizer used during Geodesic shooting, and each principal component stems from a Gaussian with variance α_k, which can be summarized as:

$$\mathbf{z}_n \sim \mathcal{N}(\mathbf{0},\mathbf{I}), \mathbf{w}_k \sim \mathcal{N}\left(\mathbf{0},\alpha_k\mathbf{I}\right), \mathbf{v}_n \sim \mathcal{N}\left(\mathbf{W}\mathbf{z}_n, \sigma^2\left(\mathbf{D}^{\mathsf{T}}\mathbf{D}\right)^{-1}\right) \quad (1)$$

The model is fitted using variational Bayesian inference under the approximation that the posterior factorises across $\mathbf{W}$ and $\mathbf{z}_n$, while the variances σ^2 and α_k are estimated by maximum-likelihood. This analysis allows each velocity field (and, therefore, each anatomy) to be described by a small number of latent projections (Ashburner et al., 2019; Balbastre et al., 2018; Zhang & Thomas Fletcher, 2015). New anatomies can then be encoded in this low-dimensional latent space by registering them with the template and projecting the resulting velocities onto the principal basis. Furthermore, random anatomies can be sampled from the learnt model by sampling latent codes from the standard Normal distribution.

Here, 662 structural MRI scans (with a 1:1 male/female ratio) were randomly sampled from four large datasets ("ABIDE", n.d; "COBRE", n.d; "IXI Dataset", n.d; "OASIS", n.d). They were segmented (Ashburner & Friston, 2005), resliced to 1.5 mm, and the grey and white matter tissues were diffeomorphically aligned with their common mean (Ashburner & Friston, 2011). A 100-component Bayesian residual model was fitted to the resulting velocity fields. Brains were rigidly aligned, meaning that brain size was encoded within at least one of these modes.

2.2. Parameter space

The shape model encodes variability using 100 eigenvectors to account for normal brain shape variation. Each individual therefore has a vector ($\mathbf{z}_n$) of 100 values that describe their brain shape and a corresponding cortical surface mesh. We then distorted this surface mesh by manipulating this individual vector. For clarity we now refer to the individual's true $\mathbf{z}_n$ prior to distortion as $\mathbf{z}_0$; distorted versions of this original vector are $\boldsymbol{z}_\alpha$. To avoid any confounds due to the volume of the mesh changing (i.e., sources becoming nearer to the sensors), we first modulated different ranges of components (e.g., 1–100, 4–100 etc) and selected a range (8–100) that maintained an approximately constant brain volume (Supplementary Fig. S1). We sampled each of these 93 components across a range of fixed points, which traversed a range of z-scores which modulated the distortion from the true brain in millimetres. A distortion trajectory is the defined by the random choices of -1 or +1 (denoted by s_i) at each of the components to be distorted (Eq. 2). For the lowest 7 components s_i is set to zero to avoid distorting the

brain size. The distortion at each component i then moves systematically from either $\{-Z \text{ to } +Z\}$ or $\{+Z \text{ to } -Z\}$ accordingly. The magnitude of this distortion (or point on the range between $+Z$ and $-Z$) is indexed by j which varies from 1 to N (Eq. 4). In this case, $N = 17$, $Z = 3$.

The random selection of direction per component (determined by the random seed that initializes the trajectory) means that it is possible to create several close-to-orthogonal trajectories for each subject (Supplementary Fig. S2). In other words, for each subject it is possible to create surrogate cortices that are distant to one another yet equally displaced (determined by j) from the true cortex (Fig. 2B).

For higher-order components, we define a random seed which gives a deterministic vector of random signs $\boldsymbol{s}$, where:

$$s_i = 2r - 1,\ 8 \le i \le 100 \tag{2}$$

where r is drawn from a random binary distribution $P(r=1) = 0.5$, $P(r=0) = 0.5$, and s_i is therefore (pseudo-randomly) either +1 or −1.

For the lower-order components (controlling brain size), s_i is set to zero.

$$s_i = 0,\ 0 < i \le 7 \tag{3}$$

$$\delta_{z_{i,j}} = \frac{2Z\left(j - \frac{N+1}{2}\right)s_i}{N},\ 1 < i \le 100,\ j \in \{1,2..N\}\ in\ \mathbb{Z} \tag{4}$$

For clarity of notation, let $\boldsymbol{\alpha}$ comprise both the choice of a specific random seed (determining the vector $\boldsymbol{s}$) and any point in the trajectory (indexed by j).

Then any distorted brain can be represented as

$$\boldsymbol{z}_{\alpha} = \boldsymbol{z}_0 + \boldsymbol{\delta z}_{\infty} \tag{5}$$

where $\boldsymbol{z}_{\alpha}$ contains the eigenvalues describing a brain distorted by $\boldsymbol{\delta z}_{\infty}$ from the original ($\boldsymbol{z}_0$).

2.3. Implementation

The diffeomorphic modeling was used to parameterize each subject's MRI. The same MRI was used to extract the initial (ground-truth) individual white-matter cortical surface, constructed using FreeSurfer (Fischl, 2012). This mesh consisted of ~ 30,000 vertices.

This initial (FreeSurfer) surface was then deformed using warps above to generate an additional 17 surfaces (Eq. 4). This parameterized distortion had a direction (in the high-dimensional space) determined **s** (dependent on the choice of random seed). Each new surface gave rise to a new source model defining a set of current dipoles at each mesh vertex. The current dipole orientation was set to be normal to the cortical surface. The extra-cranial field due to this primary current flow, in turn, depends on the intervening media. Here, we used a single-shell (or corrected-sphere) forward model (Nolte, 2003) to provide a 10th-order multiple spherical harmonic approximation

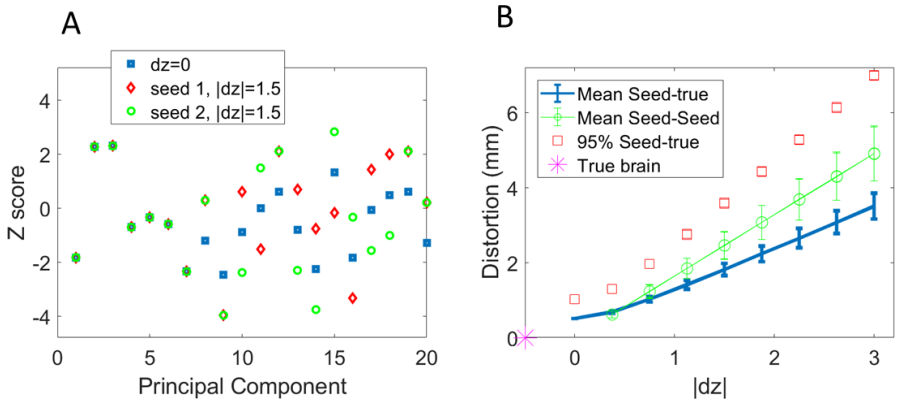

Fig. 2. (A) Example of the first 20 (of 100) components defining the true cortical surface **z** ($\delta z = 0$, blue squares) for an individual. In addition, two trajectories (based on different random seeds) are shown for $|\delta z| = 1.5$. Note that only components 8 and beyond are distorted. Each component is moved by $|\delta z|$ but the direction of movement is determined by the pseudo-random vector **s** (Eqs. 2 and 3). (B) Mean and standard deviation vertex-vertex distances between cortices on eight trajectories (eight pseudo-random vectors **s**) and the true surface (blue solid). The mean 95th percentiles of these distances are shown as red squares. Mean and standard deviation trajectory-trajectory (or seed-seed) vertex-vertex distances are shown as green circles. Note that at $|\delta z| = 0$ there is a non-zero distance to the true cortical surface. This is because the parameterization of the surfaces does not account for all spatial detail. An additional point (asterisk), at distance zero, has been added to represent the true (undistorted) cortical surface. Note that each trajectory consists of pairs of cortices at + $|\delta z|$ and - $|\delta z|$. These cortices are distinct yet distorted by the same distance in mm from $|\delta z| = 0$. This means that 8 ((N-1)/2) pairs of distorted cortices overlap on this plot.

to the inner-skull boundary (see also Supplementary Figure S5 for single sphere model). That is, the exploration of parameter space is an exploration of different source models (each defined by a different surface mesh).

2.4. Framework overview

All the inversion is carried out within the statistical parametric mapping (SPM) Bayesian framework.

From the starting point that

$$\mathbf{Y} = \mathbf{L}_\alpha \mathbf{J} + \boldsymbol{\varepsilon} \tag{6}$$

Where $\mathbf{Y} \epsilon \mathbb{R}^{N_c \times N_t}$ are the measured MEG data with N_c channels and N_t time points. $\boldsymbol{L}_\alpha \epsilon \mathbb{R}^{N_c \times N_v}$ is the gain (or lead-field) matrix linking the N_v source elements to all MEG channels for brain mesh α. The true underlying current density at each source over the N_t time points is $\mathbf{J} \epsilon \mathbb{R}^{N_v \times N_t}$. The unmodeled noise is given by $\boldsymbol{\varepsilon}$.

The most likely posterior estimate $\hat{\mathbf{J}}_\alpha$ for the α-th surrogate brain is given by

$$\hat{\mathbf{J}}_\alpha = \mathbf{Q}_\alpha \mathbf{L}_\alpha^T \left(\boldsymbol{L}_\alpha \mathbf{Q}_\alpha \mathbf{L}_\alpha^T + \mathbf{Q}_\varepsilon \right)^{-1} \mathbf{Y} \tag{7}$$

Where $\hat{\mathbf{J}}_\alpha$ is the estimated current density, $\mathbf{Q}_\alpha$ is the estimated source level covariance, and $\mathbf{Q}_\varepsilon$ is the estimated sensor level noise covariance.

To provide a null in which the physical (lead-field) information was distorted, we switched $\boldsymbol{L}_\alpha$ to $\boldsymbol{L}_{\alpha,null} \epsilon \mathbb{R}^{N_c \times N_v}$ in which the N_c rows are randomly interchanged. In this way, the data $\boldsymbol{Y}$ remains unchanged but any biophysical link to the true anatomy (determined by the lead fields) is broken:

$$\hat{\mathbf{J}}_{\alpha,\text{null}} = \boldsymbol{Q}_\alpha \boldsymbol{L}_{\alpha,\text{null}}^T \left(\boldsymbol{L}_{\alpha,\text{null}} \boldsymbol{Q}_\alpha \boldsymbol{L}_{\alpha,\text{null}}^T + \boldsymbol{Q}_\varepsilon \right)^{-1} \boldsymbol{Y} \tag{8}$$

The underlying source distribution is estimated by maximizing the negative variational Free energy (F_α) (Friston et al., 2007) (also known as the evidence lower bound or ELBO, https://en.wikipedia.org/wiki/Evidence_lower_bound):

$$F_\alpha = -\frac{N_t}{2}\log\left|\boldsymbol{\Sigma}_\alpha\right| - \frac{N_t}{2}\text{tr}\left(\boldsymbol{\Sigma}_Y \boldsymbol{\Sigma}_\alpha^{-1}\right) - \frac{N_t N_c \log(2\pi)}{2} - \frac{\left(\hat{\boldsymbol{\lambda}} - \mathbf{v}\right)^T \Pi \left(\hat{\boldsymbol{\lambda}} - \mathbf{v}\right)}{2} - \frac{\log\left|\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} \Pi\right|}{2} \tag{9}$$

Where $|\cdot|$ is the matrix determinant operator, "tr" $(\cdot)$ is the trace operator, $\boldsymbol{\Sigma}_Y = \frac{\mathbf{Y}\mathbf{Y}^T}{N_t}$ is the data covariance, and $\boldsymbol{\Sigma}_\alpha = \left(\boldsymbol{L}_\alpha \boldsymbol{Q}_\alpha \boldsymbol{L}_\alpha^T + \boldsymbol{Q}_\varepsilon\right)$ is the posterior model covariance at sensor level. The hyperparameters $\boldsymbol{\lambda}$ are assumed to have Gaussian prior and posterior distributions: $q_0(\boldsymbol{\lambda}) = N(\boldsymbol{\lambda}; \mathbf{v}, \Pi^{-1})$ and $q(\boldsymbol{\lambda}) = N(\boldsymbol{\lambda}; \hat{\boldsymbol{\lambda}}, \boldsymbol{\Sigma}_\lambda)$ respectively (Friston et al., 2008).

The key insight to this equation is in the first (sparsifying) term which penalizes the solution based on the determinant of the modeled data covariance. Namely, if the model covariance has a large volume (a high rank, or more constituent sources), it is penalized (Wipf & Nagarajan, 2009).

2.5. Inversion schemes

We use common inversion algorithms to explain the measured MEG data as current flow constrained to each surface on the trajectory in turn. Each algorithm can be defined by the modeling assumptions that determine the covariance of neuronal current flow across the cortex (López et al., 2014; Mosher et al., 2003). The models tested were all implemented in the open-source software package SPM (Litvak et al., 2011; Tierney et al., 2025), and were:

- **IID**: Minimum Norm algorithm (Hämäläinen et al., 1993). In SPM, this is implemented as single diagonal source variance component and a diagonal sensor level noise variance component which are then weighted (based on the free energy optimization above) in order to regularize the solution (Friston et al., 2008; López et al., 2014).
- **GS**: Greedy Search Algorithm—enforcing sparseness (Friston et al., 2008; López et al., 2014). In SPM, this is implemented in terms of multiple pre-defined patches of cortex (or priors) which are optimally mixed to explain the measured data outside the skull.
- **EBB**: Empirical Bayes Beamformer. In this case, a single prior is defined using beamformer assumptions (unit gain, no zero-lag covariant sources etc.) (Van Veen et al., 1997). This prior is then mixed with a diagonal estimate of sensor noise (as in the IID case) (Belardinelli et al., 2012).

In this work, we look at three additional metrics through which to quantify an inversion scheme.

- **Variance explained** (i.e., the squared correlation coefficient between measured and predicted data):

$$R^2 = \frac{\text{SST} - \text{SSR}}{\text{SST}} \tag{10}$$

where

$$\text{SSR} = \left\|\boldsymbol{Y} - \boldsymbol{L}\boldsymbol{J}\right\|_F^2 \text{ and } \text{SST} = \left\|\boldsymbol{Y}\right\|_F^2 \tag{11}$$

Where $\|\cdot\|_F$ denotes the Frobenius norm.

- **Cross-validation error** (Bonaiuto, Rossiter, et al., 2018; Browne, 2000; Little et al., 2018). Here, we compute cross-validation error as the average root mean square (RMS) difference in femto-Tesla (over the 10 cross-validation runs) between the predicted and measured values in the 10% of MEG channels (N_c' below) excluded from the source reconstruction (and randomly selected in each cross-validation run). For example:

$$CV = \frac{1}{N_c'}\sum_{1}^{N_c'}\sqrt{\frac{1}{N_t}\sum_{1}^{N_t}\left(y_{i,j}-\hat{y}_{i,j}\right)^2} \tag{12}$$

- **Maximum normalized projected source power**. This metric is based on the observation that power-estimates (normalized by noise) are maximized for correct source models (Vrba, 2002). This works for classical LCMV beamformer estimates in which there is no explicit optimization and had been used in several studies of this nature (Barnes et al., 2006; Hillebrand & Barnes, 2003, 2011). For each source, the normalized projected power is given by the weights (which map the sensor data to source location θ), the data covariance $\mathbf{\Sigma}_Y$ and normalized by the power projected by the sensor noise $\mathbf{\Sigma}_\varepsilon$.

$$\max\left|\Gamma_\theta = \frac{\boldsymbol{W}_\theta^T \mathbf{\Sigma}_Y \boldsymbol{W}_\theta}{\boldsymbol{W}_\theta^T \mathbf{\Sigma}_\varepsilon \boldsymbol{W}_\theta}\right| \tag{13}$$

  where θ covers the range of source locations and orientation. This metric does not appear in the main manuscript but is used in Supplementary Figure S7.

2.6. MEG simulations

All simulations were run within the current (2026) SPM release (Tierney et al., 2025). The shared code (*spm_sim_example_distort_mesh.m*), based on simulating a single source in the motor cortex of the canonical subject, can be used to create Figure 4. The source consisted of 336 trials of a single dipole (with no spatial extent) simulated at MNI coordinates [x = -40, y = -18, z = 48] and a single-trial sensor-level SNR ≈ -1.77 dB and -17.43 dB at peak and median sensors respectively.

See also Supplementary Figure S7 which shows the generalizability of the method using common algorithms (sLORETA, eLORETA, LCMV) with/without orientation constraints (*spm_sim_example_distort_mesh_daiss.m*) and the maximal normalized projected source power metric (Barnes et al., 2006; Hillebrand & Barnes, 2003, 2011; Vrba, 2002) as a cost function.

2.7. MEG empirical data

We used previously published data from MEG recordings of eight subjects using head-casts (Bonaiuto, Meyer, et al., 2018; Little et al., 2018) (see data-availability section for an updated link). The head-cast minimizes subject movement during the experiment and gives a precise estimate of the relative positions of the sensor array and head (Meyer et al., 2017; Troebinger et al., 2014). In brief, recordings were made using a 275-channel CTF Omega MEG system. The data were sampled at 1200 Hz, filtered (5th-order Butterworth band-pass filter: 2–100 Hz, Notch filter: 50 Hz), and down-sampled to 250 Hz. Eye-blink artefacts were removed using multiple source eye correction (Berg & Scherg, 1994). Participants completed a visually cued action decision-making task in which they responded to a Random Dot Kinematogram (RDK) (displayed for 2 s with coherent motion, either to the left or right). Following a 500 ms delay, an instruction cue appeared, pointing either to the left or the right, and participants were instructed to press the corresponding button (left or right) as quickly and as accurately as possible.

The processing of the data was either focused on trials locked to the onset of the instruction cue (0 to 500 ms), the random dots (-2500 to -2000 ms) or to the button press (-200 to 300 ms). Time zero and epoching were as defined in Bonaiuto, Meyer, et al. (2018). These trials were first averaged and filtered to a 5–90 Hz bandwidth.

3. RESULTS

We now move through the space of possible cortical surfaces and examine the impact on MEG reconstruction algorithms and fit metrics. We first show how the cortex is progressively deformed from the anatomical ground truth. We show the effect of this distortion on the reconstruction of simulated data. We go on to test whether this deformation has any influence on the metric or method of source reconstruction for real data (recorded using a head-cast). The logic is that distortions from the true anatomy will be of little consequence for sub-optimal metrics or reconstruction assumptions.

3.1. Deformations

Figure 2A shows two potential distortion trajectories (Eqs. 2 to 5). The shape of any brain in the population can be defined by a 100-element vector $\boldsymbol{z}_0$ (blue squares). We then define a pseudo-random deformation trajectory (Eqs. 2 and 3) to move each person's cortical mesh about

$\boldsymbol{z}_0$ in 17 successive steps (Eqs. 4 and 5). Cortical surfaces 1 and 17 are at the extremes of the trajectory, with surface 9 at the origin, where $\delta z = 0$ (i.e., closest to the original cortical surface). Figure 2B shows the average distance of each vertex to the corresponding vertex on the original cortical surface (distributions shown in Supplementary Fig. S3). The true cortical surface (asterisk) on the x-axis is added for reference (as the trajectories do not capture all of the anatomical variance). Note that surfaces 1 and 17 and all pairs (2,16; 3,15 etc.), which have equal (but opposite) distortion of $\boldsymbol{z}_0$, overlap on this plot. Also shown in Figure 2B are the distances between surfaces lying on different trajectories (determined by the pseudo-random vector $\boldsymbol{s}$) and one another. That is, even though the trajectories are all a similar distance from the true brain, they are also distant from one another. Supplementary Figure S2 shows the correlation (mean $r^2 = 0.046$) between the vectors $\delta \boldsymbol{z}$ that define each trajectory.

Figure 3 shows three points on one trajectory of brain deformations, alongside similarly deformed scalp surfaces. As it is difficult to spot these cortical deformations by eye, the subject's scalp surface is shown below to illustrate the deformation as $|\delta \boldsymbol{z}|$ increases in magnitude. Note the gradual distortion of the face moving from left (truth) to right (distortion of lips and face). Coronal slices through the original (top left) and deformed (top right) surfaces are shown in Figure 4A.

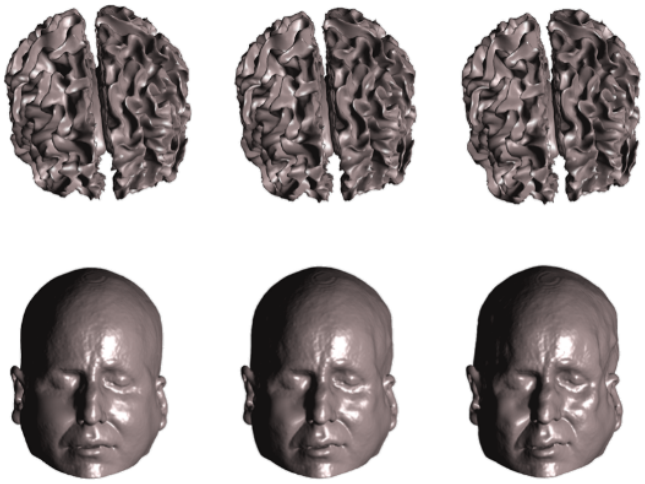

Fig. 3. Top panel. Cortical meshes from one example trajectory at points $|\delta z| = 0$, $|\delta z| = 1.5$ and $|\delta z| = 3$ (left-right respectively), which show increasing amounts of distortion. Figure 4A shows a superposition of coronal slices in which the difference between surfaces is more apparent. For illustration only, the lower panel shows the (70% reduced in size) scalp surface of this subject distorted in the same way. Note the gradual distortion of the face moving from left (truth) to right (distorted).

3.2. Simulated data

We first show an example of this methodology with simulated data. A single source was placed in the motor cortex based on the true brain shape (i.e., true orientation and position of this source). The simulated data were reconstructed onto brains moving across eight separate distortion trajectories. Reconstructions were based either on IID (distributed) or EBB (sparse) assumptions. We show the goodness of fit using Free-Energy in this case.

Figure 4 provides an example in simulation (*spm_sim_example_distort_mesh.m* creates this). A single source is simulated on the cortical surface (Fig. 4A). It is reconstructed with matching and ill-matching assumption sets (in this case EBB and IID algorithms respectively) onto progressively more distorted surfaces. A coronal slice through the true cortex and a distorted cortex at $|\delta \boldsymbol{z}| = 3$ are shown in Figure 4A. The site of the original current element is moved (in orientation and position) such that its impact on the measurement channels has changed (Fig. 4B). Figure 4C shows these data reconstructed under EBB (top) and minimum norm assumptions (lower) onto the true (left) and distorted cortices (right). The distortion, however, has no noticeable impact on the estimated field at the sensor level between algorithms or between distortions (Fig. 4D). In Figure 4E the Free energy cost function is used to quantify the solutions as distortion is increased for eight different distortion trajectories. The peak of the cost function (or best anatomical model) in each case is marked (diamonds for EBB, circles for IID). For the IID (incorrect assumptions) note that the cost-function increases (model becomes more likely) as the underlying anatomy becomes more distorted. For EBB (correct assumptions) however, the best models tend to peak closer to zero distortion. Figure 4F shows that the estimated distortion, in millimeters, can be used to judge between inversion schemes.

3.3. Real data

We now apply the same approach to real data (Bonaiuto, Meyer, et al., 2018). These data were recorded using subject-specific head-casts so as to minimize co-registration error. In an ideal scenario, the model fit metric should improve for brain structures closer to reality. We split this section into a comparison between algorithms, and then between metrics of fit.

3.3.1. Comparing algorithms

Figure 5A shows the negative variational Free Energy for EBB, IID, and GS algorithms for one trajectory and a single

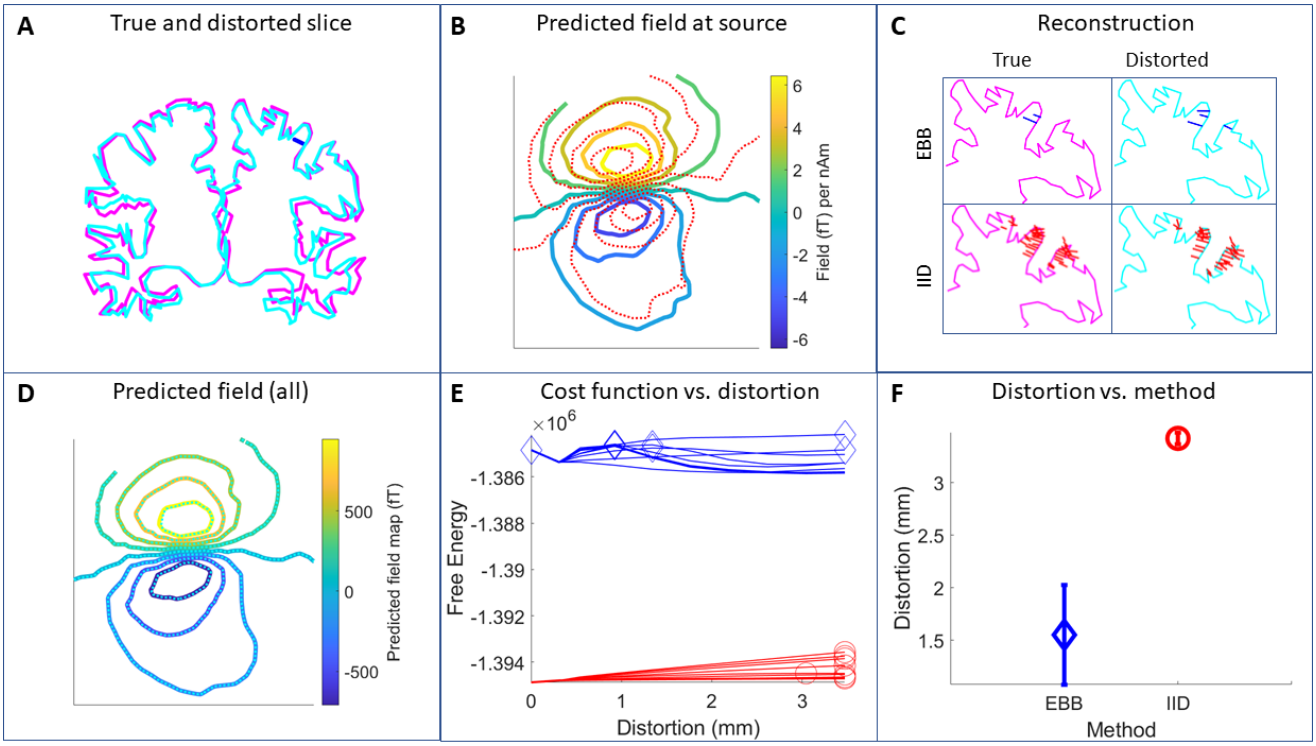

Fig. 4. (A) Simulated source (dark blue bar) on a coronal slice through the true cortical mesh (magenta). Also shown is a cortical mesh (thin light blue lines) at the extreme of the trajectory ($|\delta z| = 3$) with mean vertex-vertex distortion from original of 3.5 mm. (B) The magnetic field profile at sensor level due to the original source (colors) and alongside (dotted) the magnetic field pattern due to the same current element on the distorted surface. (C) The source reconstruction under EBB (top) and IID (lower) assumptions onto the undistorted (left) and maximally distorted cortices (right). (D) The field-maps predicted by all current estimates in C give rise to sensor-level field maps that explain 99% of the variance of one another. Shown are the field-maps due to IID reconstructions (lower panel C) of the same data onto distorted (dotted) and undistorted cortices. The field-maps explain 99% of the variance of one another. The same is the case for EBB (99% of IID field map explained) (E) The free energy associated with reconstructions onto the different cortices as distortion is increased for eight different trajectories and two different algorithms (EBB, IID). Diamonds and circles mark the peak metric (signifying most likely underlying cortex) for EBB and IID respectively. (F) Each peak (in E) gives a distortion. Taking the mean distortions across the eight trajectories gives a single distance measure for each algorithm.

dataset. The peaks are marked. Note that the best model (as quantified purely by Free Energy) is the GS algorithm and a brain distorted by 3 mm. Note also that IID and EBB algorithms favor brains closer to the ground truth (0 distortion). The original averaged data, source level images, Free energy values, and time-series for the EBB case (distorted and undistorted) are shown in Supplementary Figure S4. This process is now repeated across 8 trajectories and the peak fit metric (in mm of distortion) noted (Fig. 5B). For each algorithm the mean peak location (in mm) (large symbols) is taken to the next level of analysis. Figure 5C shows the empirical data for eight subjects and three datasets per subject. The distortions (per dataset) are ordered according to how close they are to zero in order to aid comparison. In addition, the null model (Eq. 8), using the EBB algorithm, is shown in black crosses. Figure 5D shows the mean and standard errors of the distances between the true brain and the algorithm-preferred brains. All algorithms come closer to the true brain than would be expected by chance ($p < 0.05$) (dotted, and set by the null models). In this case, the only significant difference (between algorithms, one-tailed t-test) is that EBB improves over GS ($t(23) = -1.89$, $p < 0.036$).

3.3.2. Comparing metrics

It is possible to run the same analysis but using different combinations of fit algorithm and metric. Figure 6A shows the data from Figure 5 plotted as a function of the number of datasets showing less than a certain distortion using the metric of Free Energy. Larger values to the left of the curve (or larger area under curve) indicate better performance. Figure 6B scores the same algorithms using cross-validation. In this case, note that in 40% of the datasets the true brain (zero distortion) is found. Figure 6C shows the algorithms scored according to variance explained. Note the relatively poor performance of IID and EBB algorithms under this metric. Summarizing in

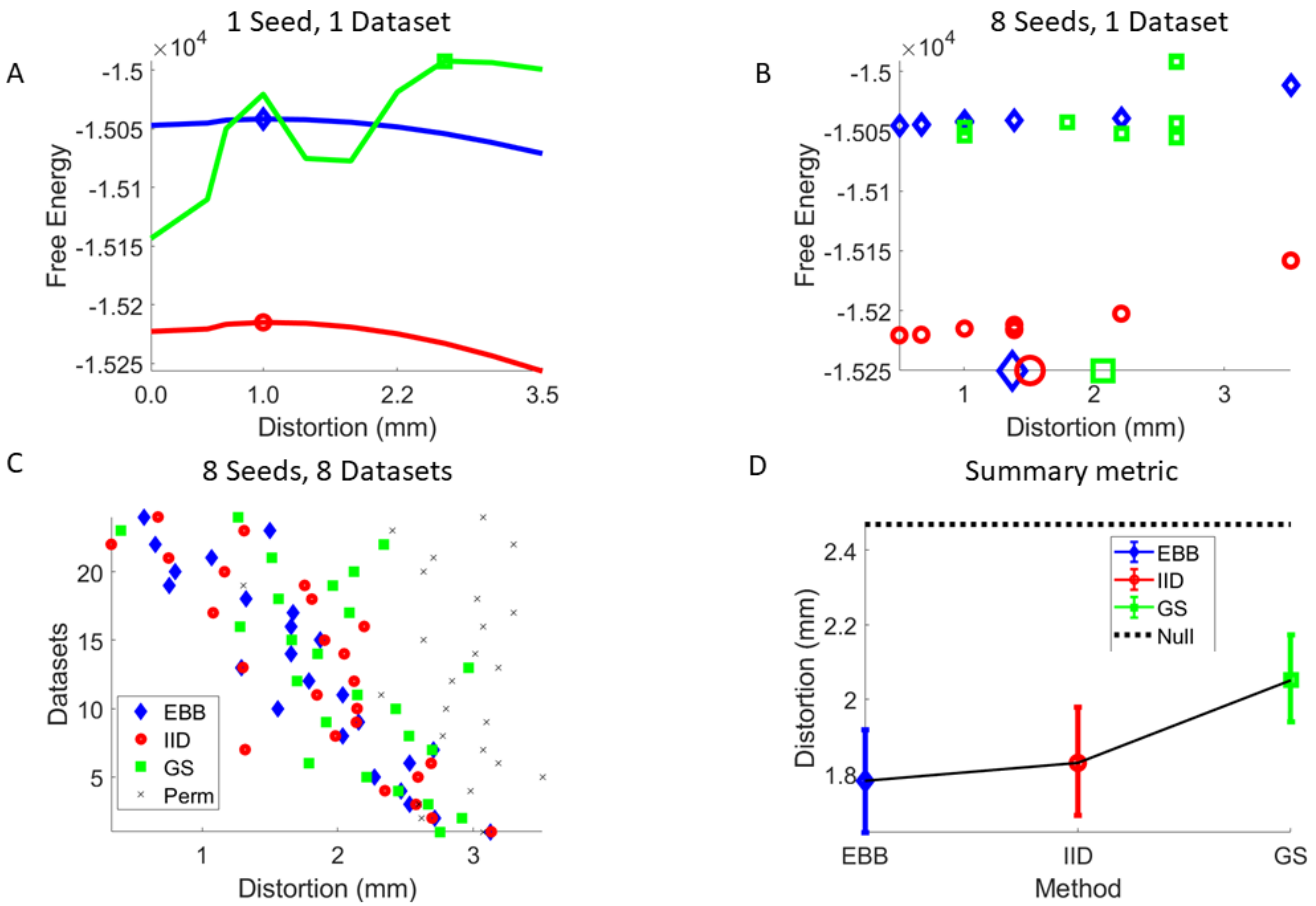

Fig. 5. Empirical head-cast data. Inverted using three different algorithms EBB (blue diamonds), IID (red circles), and GS (green squares). (A) Free Energy (one trajectory, 1 dataset) as the cortex is distorted from 0 to 3.5 mm (mean vertex-to-vertex distance). The peaks (marked) indicate which cortices are most likely to support the measured data. Note that, by the Free-Energy metric alone, the most-likely anatomy is found by the GS algorithm at a mean distortion of ~3 mm. (B) Peak locations for 8 trajectories for the same dataset for three different algorithms EBB (blue diamonds), IID (red circles), and GS (green squares). The mean distortions estimated for each algorithm in this dataset (large markers on x-axis) are taken to the next level of inference. (C). Estimated distortions for the three algorithms over eight subjects and three datasets per subject. Distortions are ordered (top to bottom) in terms of increasing mean distortion (for visualization purposes). The crosses show the distortions estimated using the null models for the EBB algorithm (D) One can then quantify algorithm performance in terms of the mean distortions (in **C**) in millimetres. All algorithms perform better than chance ($p < 0.05$, black-dotted, set by the null models). EBB significantly ($p < 0.05$) outperforms GS.

terms of mean distortion, it is clear that (for these data) cross-validation (dashed line) arrives at closest to the ground truth for EBB and IID algorithms. In this case, one-tailed t-tests show significant difference for EBB vs. GS ($t(23) = -5.49$, $p < 0.0001$) and IID vs. GS ($t(23) = -5.50$, $p < 0.0001$).

These comparisons are not limited between algorithms or metrics. Supplementary Figure S5 shows that alternative forward models (in this case the single sphere) can also be compared.

4. DISCUSSION

We leverage the knowledge of true cortical anatomy to compare among MEG analysis pathways using a metric of millimetres. We have done this using realistic and quantifiable deformations to distort an individual's cortical surface. We first show this in simulation (Fig. 4). We then show that the same applies for real-MEG data. All algorithms tested significantly ($p < 0.05$) outperformed chance behavior as defined by the null case (Fig. 5D). We found that inversion schemes based on the assumption sets behind IID and EBB were most selective of the true anatomy; and that this selectivity was improved when using a cross-validation metric (Fig. 6D).

This approach could be used to empirically determine best practice at any stage of the MEG (or EEG) pathway from the choice of forward models to sensor type. Importantly the metric of performance is in terms of millimetre distortion from true anatomy.

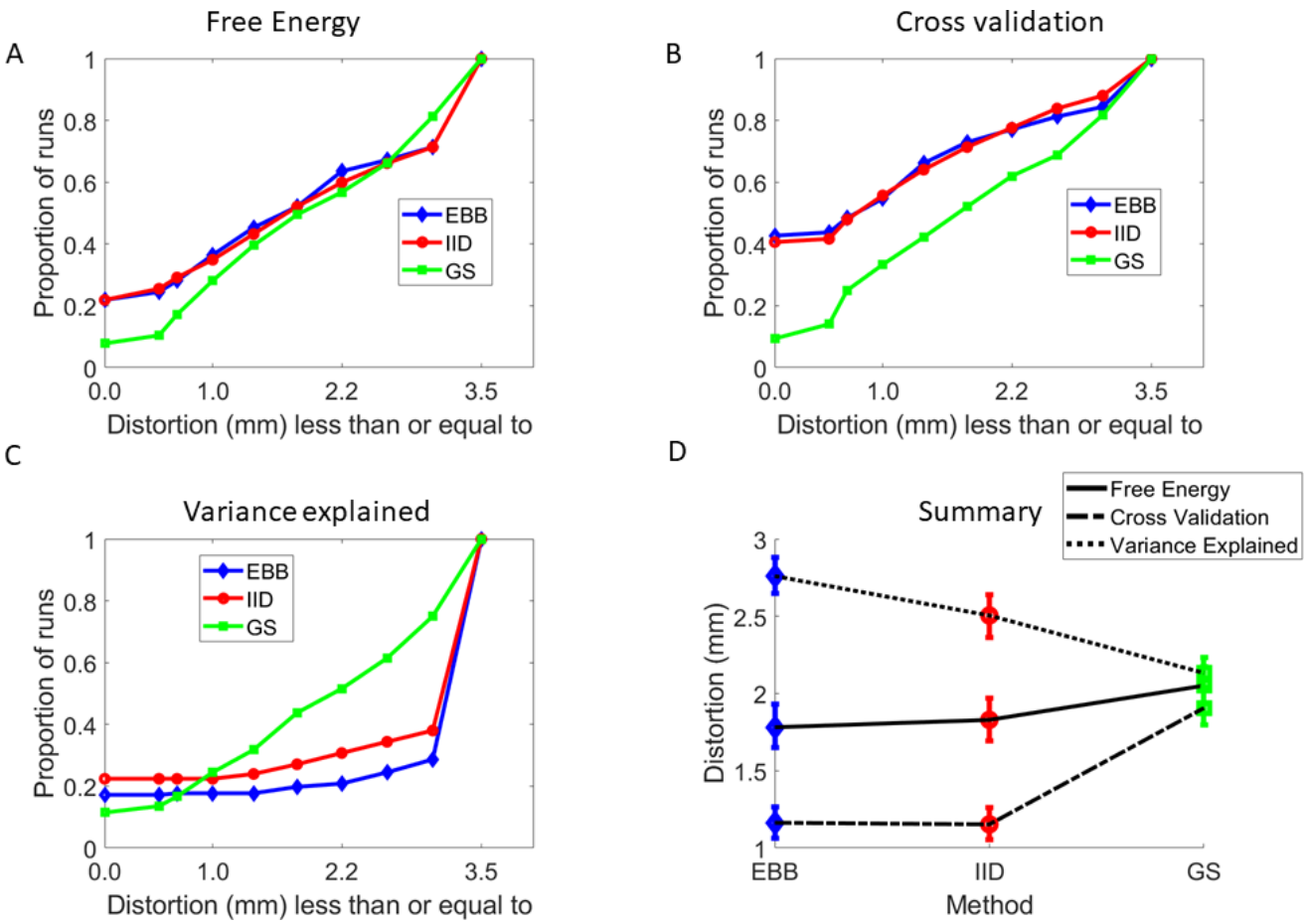

Fig. 6. (A–C) Comparing metrics and algorithms EBB (blue triangles), IID (red circles), GS (green squares). (A) The Free Energy metric and the three different algorithms; (B) The cross-validation metric; (C) the variance explained. (D) Summary figure showing mean distortions for the three algorithms scored using the three different metrics: Free-energy (solid line), Cross-validation (dashed line), and variance explained (dotted line).

We should note that these findings are not restricted to a Bayesian framework. One can select other metrics such as peak normalized projected power (Barnes et al., 2006; Hillebrand & Barnes, 2003, 2011). Several commonly used algorithms (including LCMV (Van Veen et al., 1997), eLORETA (Pascual-Marqui et al., 2011), sLORETA (Pascual-Marqui, 2002)) are made use of in the demo-code (spm_sim_example_distort_daiss.m). Supplementary Figure S7 shows the sensitivity of these algorithms to cortical distortion with and without orientation constraints. As one might expect, the use of the cortical orientation makes all algorithms more sensitive to changes in the underlying anatomy.

It is interesting to note the contrast between the simulated (Fig. 4) and real data (Figs. 5 and 6). In simulation, there was a very clear difference between IID and EBB models. In the real data, however, there was little difference. There could be a number of reasons for this: i) The activity was primarily cortical and superficial and so it could be more difficult to distinguish between IID (typically superficially biased) and EBB estimates. ii) The EBB is not a classic beamformer algorithm (Van Veen et al., 1997) but rather uses an empirical beamformer prior as part of a generative model (Belardinelli et al., 2012). In cases in which a conventional beamformer might fail (for example zero lag correlated and distinct sources) there is no useful information in the prior and it becomes flat—defaulting to minimum norm behavior. In other words, EBB and IID will tend to the same algorithm in certain cases. iii) Finally, and perhaps most-likely, neither the assumptions between EBB or IID generative models match the true neuronal generative model and/or algorithm performance is constrained by some other factor (for example, sub-optimal volume conductor model, tissue conductivity, or segmentation errors).

An alternative method of validation is to see how the distortion metric performs when one subject's data is used to estimate the best fitting anatomy from across the

group. Supplementary Figure S6 shows that (for EBB) the most likely anatomy matches the MEG data it corresponds to for 6 of 8 subjects ($p < 0.0001$).

We used the white matter (rather than pial) surface and made no attempt to optimise the orientation of current flow (Bonaiuto et al., 2020). We used exclusively cortical surfaces; but if the manifold on which sub-cortical sources sit is well defined (Attal & Schwartz, 2013; Meyer, Rossiter, et al., 2017) the same approach could be used. Likewise, rather than distort the cortical surface, the same code can be used to also distort any of the conductivity boundaries (such as the inner skull).

In the limit of infinite data, all estimates should converge toward the true anatomy. Using longer-recordings and novel data processing techniques (e.g., Hidden Markov Models—Woolrich et al., 2013) would allow one to make use of all data recorded (Little et al., 2018; Martinez-Vargas et al., 2016).

4.1. Limitations

We have made use of the Nolte Single Compartment (or Single Shell) model which has been shown to outperform single shell BEMs but not more complex models with more compartments (Stenroos et al., 2014). We also made use of the single sphere forward model (Supplementary Fig. S5). There were two motivations for this. Firstly, to test a model that does not depend on the inner skull-boundary (which may be crossed during distortion) but simply requires specification of the volume centre (Sarvas, 1987). Secondly, to show how the method could be generalized (to more complex forward models in future). On the positive side, the findings between the two models were similar—indicating that our results are not due to crossing the inner-skull boundary. Surprisingly however, the simpler sphere model outperformed the corrected-sphere model (Supplementary Fig. S5). Future work, using more complex forward models, could examine the role that changes in CSF and white-matter compartments have on the solution. The main computational drawback is that a new set of lead-fields must be computed for each distortion and seed.

It should be noted that the implementation of GS used here is most likely sub-optimal in that only a fixed number of discrete (pre-defined) cortical patches are allowed to be active. Previous work has combined the greedy-search with an automatic relevance determination (ARD) stage (Friston et al., 2008): the greedy search to find variance components (source distributions) that maximize the model evidence and ARD to prune away those that make least contribution to the evidence. Alternatively, more computationally intensive approaches have been suggested in which model evidence is assessed over multiple possible patch locations to overcome problems with sparse sampling; and this, in turn, is run over multiple iterations to avoid local maxima (Lopez et al., 2012).

The averaged evoked data shown are specific to one experiment derived from three time-windows and in a single frequency band over eight subjects. We cannot, therefore, generalize findings across other datasets.

The cortical deformations are based on eight (random) distortion trajectories for all subjects. In order to observe any change in how the measured MEG data can be expressed on the cortical surface, we rely on the distortion perturbing cortical sources that will influence the estimate of current flow. However, distortion trajectories must exist that minimally influence this estimate; in which case we would see no change in model fit with distortion.

4.2. Perspective

Note that these data are based on head-cast recordings. Here, subjects wore individualized foam helmets that fit the scalp internally and the MEG system externally. That is, the subjects' heads were immobilised in the scanner. Therefore, the relative location of the subject's anatomy and the sensors is well-defined and fixed. This scanning method is time-intensive, only appropriate for healthy compliant subjects, and carries a certain amount of risk (due to the potential of MEG system movement whilst the subject's head is fixed). We used head-cast data here as for this method the knowledge of true anatomy is essential. However, in practice, robustness to (inevitable) co-registration errors in an inversion algorithm is desirable. We suggest that this constitutes an initial stage to establish the relative performance of an algorithm (or metric). Once these algorithms (or assumptions about source covariance) have been established, the anatomical precision is no longer required.

Co-registration may also become less of an issue with a new generation of MEG sensors known as optically pumped magnetometers (OPMs) (Tierney et al., 2019), which can be worn for long periods, and their location relative to the underlying anatomy can be well defined. Although head-cast recording may be limited, OPMs provide an exciting and expanding field to test these methods further. This is important as OPMs allow subjects to perform tasks (like dancing (O'Neill et al., 2025)) which expand the literature yet cannot be replicated in many fixed scanning systems. In the limit of increasing measured data for an individual, these tests between competing anatomies should converge to a single solution (the true individual anatomy). The more data available per-subject, the more constrained (and well-posed) the problem becomes.

DATA AND CODE AVAILABILITY

The code is part of the current development version of SPM (https://github.com/spm/spm). The deformation templates will be automatically downloaded the first time the algorithm is used. The code to train the Bayesian Residual Component Analysis model is available at https://github.com/balbasty/BRCA. The MEG-evoked response data, individual MRIs, and cortical surfaces are available here: https://figshare.com/s/e3ff829753a76bcfda22. The additional code to reproduce the simulations and single-subject analyses can be found at https://github.com/barnesgr123/spm_distort/ and within the latest version of SPM. The complete dataset, originally from Bonaiuto, Meyer, et al. (2018) and Little et al. (2018), can be found at https://osf.io/eu6nx.

AUTHOR CONTRIBUTIONS

G.B., J.D.L., and Y.B.—conception. G.B., J.D.L., J.J.B., and A.M.—MEG-specific code. Y.B. and J.A.—Structure-specific code. All authors—writing.

DECLARATION OF COMPETING INTEREST

The authors have no competing interests.

ETHICS

We used previously published data from MEG recordings of eight subjects using head-casts (Bonaiuto, Meyer, et al., 2018; Little et al., 2018). The study protocol was in full accordance with the Declaration of Helsinki, and all participants gave written informed consent after being fully informed about the purpose of the study. The study protocol, participant information, and form of consent, were approved by the UCL Research Ethics Committee (reference number 5833/001).

ACKNOWLEDGMENTS

J.D.L. was funded by the Colombian MinCiencias project contract 1222-852-69927. Y.B. was funded in part by the MRC and Spinal Research Charity through the ERA-NET Neuron joint call (MR/R000050/1), and in part by a fellowship from the Royal Society (NIF\R1\232460). This research was supported by the Discovery Research Platform for Naturalistic Neuroimaging funded by the Wellcome Trust (226793/Z/22/Z), by core center funding from Wellcome Trust (203147/Z/16/Z), and by BBSRC grant BB/M009645/1. We would like to thank the UCL MEG group for their helpful and constructive comments.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/IMAG.a.1189#supplementary-data.

REFERENCES

ABIDE. (n.d.). Retrieved March 15, 2020, from http://fcon_1000.projects.nitrc.org/indi/abide/

Ashburner, J., Brudfors, M., Bronik, K., & Balbastre, Y. (2019). An algorithm for learning shape and appearance models without annotations. *Medical Image Analysis*, *55*, 197–215. https://doi.org/10.1016/j.media.2019.04.008

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*, *55*(3), 954–967. https://doi.org/10.1016/j.neuroimage.2010.12.049

Attal, Y., & Schwartz, D. (2013). Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: A MEG study. *PLoS One*, *8*(3), e59856. https://doi.org/10.1371/journal.pone.0059856

Balbastre, Y., Brudfors, M., Bronik, K., & Ashburner, J. (2018). Diffeomorphic brain shape modelling using Gauss-Newton optimisation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-00928-1_97

Barnes, G. R., Furlong, P. L., Singh, K. D., & Hillebrand, A. (2006). A verifiable solution to the MEG inverse problem. *NeuroImage*, *31*(2), 623–626. https://doi.org/10.1016/j.neuroimage.2005.12.036

Belardinelli, P., Ortiz, E., Barnes, G., Noppeney, U., & Preissl, H. (2012). Source reconstruction accuracy of MEG and EEG Bayesian inversion approaches. *PLoS One*, *7*(12), e51985. https://doi.org/10.1371/journal.pone.0051985

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241. https://doi.org/10.1016/0013-4694(94)90094-9

Bishop, C. M. (1999). Variational principal components. In *9th International Conference on Artificial Neural Networks: ICANN '99*, *1999* (pp. 509–514). IEEE. https://doi.org/10.1049/cp:19991160

Bonaiuto, J. J., Afdideh, F., Ferez, M., Wagstyl, K., Mattout, J., Bonnefond, M., Barnes, G. R., & Bestmann, S. (2020). Estimates of cortical column orientation improve MEG source inversion. *NeuroImage*, *216*, 116862. https://doi.org/10.1016/j.neuroimage.2020.116862

Bonaiuto, J. J., Meyer, S. S., Little, S., Rossiter, H., Callaghan, M. F., Dick, F., Barnes, G. R., & Bestmann, S. (2018). Lamina-specific cortical dynamics in human visual and sensorimotor cortices. *Elife*, *7*, e33977. https://doi.org/10.7554/eLife.33977

Bonaiuto, J. J., Rossiter, H. E., Meyer, S. S., Adams, N., Little, S., Callaghan, M. F., Dick, F., BEstmann, S., & Barnes, G. R. (2018). Non-invasive laminar inference with MEG: Comparison of methods and source inversion algorithms. *NeuroImage*, *167*, 372–383. https://doi.org/10.1016/j.neuroimage.2017.11.068

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108–132. https://doi.org/10.1006/JMPS.1999.1279

COBRE. (n.d.). Retrieved March 15, 2020, from http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html
Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021
Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., & Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, *39*(3), 1104–1120. https://doi.org/10.1016/j.neuroimage.2007.09.048
Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*, 220–234. https://doi.org/10.1016/j.neuroimage.2006.08.035
Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2), 413–497. https://doi.org/10.1103/RevModPhys.65.413
Hillebrand, A., & Barnes, G. R. (2003). The use of anatomical constraints with MEG beamformers. *NeuroImage*, *20*(4), 2302–2313. https://doi.org/10.1016/j.neuroimage.2003.07.031
Hillebrand, A., & Barnes, G. R. (2011). Practical constraints on estimation of source extent with MEG beamformers. *NeuroImage*, *54*(4), 2732–2740. https://doi.org/10.1016/j.neuroimage.2010.10.036
IXI Dataset. (n.d.). Retrieved March 14, 2020, from https://brain-development.org/ixi-dataset/
Kalaitzis, A. A., & Lawrence, N. D. (2011). A simple approach to ranking differentially expressed gene expression time courses through Gaussian Process Regression. *BMC Bioinformatics*, *12*(1), 180. https://doi.org/10.1186/1471-2105-12-180
Little, S., Bonaiuto, J., Meyer, S. S., Lopez, J., Bestmann, S., & Barnes, G. (2018). Quantifying the performance of MEG source reconstruction using resting state data. *NeuroImage*, *181*, 453–460. https://doi.org/10.1016/j.neuroimage.2018.07.030
Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., & Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, *2011*, 852961. https://doi.org/10.1155/2011/852961
Lopez, J. D., Espinosa, J. J., & Barnes, G. R. (2012). Random location of multiple sparse priors for solving the MEG/EEG inverse problem. In *Conference Proceedings: … Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2012* (pp. 1534–1537). IEEE. https://doi.org/10.1109/embc.2012.6346234
López, J. D., Litvak, V., Espinosa, J. J., Friston, K., & Barnes, G. R. (2014). Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. *NeuroImage*, *84*, 476–487. https://doi.org/10.1016/j.neuroimage.2013.09.002
López, J. D., Valencia, F., Flandin, G., Penny, W., & Barnes, G. R. (2017). Reconstructing anatomy from electro-physiological data. *NeuroImage*, *163*, 480–486. https://doi.org/10.1016/j.neuroimage.2017.06.049
Martinez-Vargas, J. D., Lopez, J. D., Baker, A., Castellanos-Dominguez, G., Woolrich, M. W., & Barnes, G. (2016). Non-linear parameter estimates from non-stationary MEG data. *Frontiers in Neuroscience*, *10*, 366. https://doi.org/10.3389/fnins.2016.00366
Meyer, S. S., Bonaiuto, J., Lim, M., Rossiter, H., Waters, S., Bradbury, D., Bestmann, S., Brookes, M., Callaghan, M. F., Weiskopf, N., & Barnes, G. R. (2017). Flexible head-casts for high spatial precision MEG. *Journal of Neuroscience Methods*, *276*, 38–45. https://doi.org/10.1016/j.jneumeth.2016.11.009
Meyer, S. S., Rossiter, H., Brookes, M. J., Woolrich, M. W., Bestmann, S., & Barnes, G. R. (2017). Using generative models to make probabilistic statements about hippocampal engagement in MEG. *NeuroImage*, *149*, 468–482. https://doi.org/10.1016/j.neuroimage.2017.01.029
Miller, M. I., Christensen, G. E., Amit, Y., & Grenander, U. (1993). Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(24), 11944–11948. https://doi.org/10.1073/pnas.90.24.11944
Miller, M. I., Trouvé, A., & Younes, L. (2006). Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, *24*(2), 209–228. https://doi.org/10.1007/s10851-005-3624-0
Mosher, J. C., Baillet, S., & Leahy, R. M. (2003). Equivalence of linear approaches in bioelectromagnetic inverse solutions. *IEEE Workshop on Statistical Signal Processing, 2003*, 294–297. https://doi.org/10.1109/SSP.2003.1289402
Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoenchephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, *48*(22), 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002
OASIS. (n.d.). Retrieved March 15, 2020, from https://www.oasis-brains.org/
O'Neill, G. C., Seymour, R. A., Mellor, S., Alexander, N. A., Tierney, T. M., Bernachot, L., Fahimi Hnazaee, M., Spedden, M. E., Timms, R. C., Bush, D., Bestmann, S., Brookes, M. J., & Barnes, G. R. (2025). Combining video telemetry and wearable MEG for naturalistic imaging. *Imaging Neuroscience*, *3*, imag_a_00495. https://doi.org/10.1162/IMAG_A_00495
Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, *103*(1), 25–26. https://doi.org/10.1016/S0013-4694(97)88020-4
Pascual-Marqui, R. D., Lehmann, D., Koukkou, M., Kochi, K., Anderer, P., Saletu, B., Tanaka, H., Hirata, K., John, E. R., Prichep, L., Biscay-Lirio, R., & Kinoshita, T. (2011). Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *369*(1952), 3768–3784. https://doi.org/10.1098/RSTA.2011.0081
Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology*, *32*(1), 11–22. https://doi.org/10.1088/0031-9155/32/1/004
Stenroos, M., Hunold, A., & Haueisen, J. (2014). Comparison of three-shell and simplified volume conductor models in magnetoencephalography. *NeuroImage*, *94*, 337–348. https://doi.org/10.1016/J.NEUROIMAGE.2014.01.006
Stevenson, C., Brookes, M., López, J. D., Troebinger, L., Mattout, J., Penny, W., Morris, P., Hillebrand, A., Henson, R., & Barnes, G. (2014). Does function fit structure? A ground truth for non-invasive neuroimaging. *NeuroImage*, *94*, 89–95. https://doi.org/10.1016/j.neuroimage.2014.02.033
Tierney, T. M., Alexander, N. A., Avila, N. L., Balbastre, Y., Barnes, G., Bezsudnova, Y., Brudfors, M., Eckstein, K.,

Flandin, G., Friston, K., Jafarian, A., Kowalczyk, O. S., Litvak, V., Medrano, J., Mellor, S., O'Neill, G., Parr, T., Razi, A., Timms, R., & Zeidman, P. (2025). SPM 25: Open source neuroimaging analysis software. *Journal of Open Source Software*, *10*(110), 8103. https://doi.org/10.21105/joss.08103

Tierney, T. M., Holmes, N., Mellor, S., López, J. D., Roberts, G., Hill, R. M., Boto, E., Leggett, J., Shah, V., Brookes, M. J., Bowtell, R., & Barnes, G. R. (2019). Optically pumped magnetometers: From quantum origins to multi-channel magnetoencephalography. *NeuroImage*, *199*, 598–608. https://doi.org/10.1016/j.neuroimage.2019.05.063

Troebinger, L., López, J. D., Lutti, A., Bradbury, D., Bestmann, S., & Barnes, G. (2014). High precision anatomy for MEG. *NeuroImage*, *86*, 583–591. https://doi.org/10.1016/j.neuroimage.2013.07.065

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056

Vrba, J. (2002). Magnetoencephalography: The art of finding a needle in a haystack. *Physica C: Superconductivity*, *368*(1–4), 1–9. https://doi.org/10.1016/S0921-4534(01)01131-5

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

Woolrich, M. W., Baker, A., Luckhoo, H., Mohseni, H., Barnes, G., Brookes, M., & Rezek, I. (2013). Dynamic state allocation for MEG source reconstruction. *NeuroImage*, *77*, 77–92. https://doi.org/10.1016/j.neuroimage.2013.03.036

Zhang, M., & Thomas Fletcher, P. (2015). Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability. *Medical Image Analysis*, *25*(1), 37–44. https://doi.org/10.1016/j.media.2015.04.009