Learning Lead Field Corrections for EEG–fMRI Functional Connectivity Alignment

Anonymous authors
Paper under double-blind review

Abstract

Joint analysis of EEG and BOLD-fMRI signals is a route to studying how brain structure relates to brain function. This paper presents a model that learns a low-rank correction to the T1-derived lead field per subject. The correction is fitted by matching the EEG-side parcel covariance and the fMRI-derived parcel covariance, each estimated within-subject from the corresponding modality's own resting data. We call this construction *Lead-Field Low-Rank Correction (LF-LoRA)*. The covariance matching has the form $\Sigma_z \leftrightarrow \hat{\Sigma}_z = B\Sigma_y B^\top$, where $\Sigma_{\{y,z\}}$ represents the EEG/BOLD-parcel signal covariance and B is a linear EEG-to-parcel operator built from the corrected lead field $A = A_0 + UV^\top$. The simple model structure is consistent with the linear forward-modeling assumption from which the lead field is defined, and its solution may help explain subject-specific structure in the EEG forward model (Hämäläinen et al., 1993). The linear model admits a decoupled temporal operator which aligns Σ_y and Σ_z at lag $\tau > 0$. The temporal operator provides a means to investigate the dynamics of neural activity from EEG. We discuss the solution space of LF-LoRA. We evaluate the construction on the LEMON cohort (Babayan et al., 2019). Experiments validate the subject specificity of the learned model and the utility of reading $UV^\top$ on cortex.

1 Introduction

Functional connectivity (FC) is a matrix of pairwise temporal correlations between neural signals. FC is a reliable network biomarker that characterizes neural activities in individual subjects (Finn et al., 2015). Resting-state functional MRI (fMRI) is the de facto source of parcel-level FC (Biswal et al., 1995). However, the cost and motion sensitivity of MRI (Power et al., 2012) limits longitudinal and large-scale use. On the other hand, electroencephalography (EEG) is inexpensive and portable, and can be acquired even in patients who cannot undergo fMRI (e.g. during rehabilitation). But EEG FC is often considered less reliable and less comparable to fMRI FC (Hipp et al., 2012; Wirsich et al., 2020). An fMRI-quality FC recovered from a subject's own EEG would open a low-cost, clinically feasible channel for network imaging.

Both modalities probe the underlying neural activity. fMRI measures the blood oxygenation level-dependent (BOLD) signal, which correlates spatially with cortex but is distinct from neural electrical activity. Scalp EEG measures electrical potentials directly, through a linear mixture of cortical source amplitudes via the lead field (Hämäläinen et al., 1993). A natural strategy is therefore to recover source amplitudes from EEG, aggregate to parcel-level second-order statistics, and match these to the subject's fMRI parcel FC. In particular, let *lead field* $A \in \mathbb{R}^{C\times N}$ be a linear map from N cortical source amplitudes to C scalp potentials (Hämäläinen et al., 1993). If A is known, an estimate of the cortical activity can be obtained from EEG using the pseudoinverse $A^\dagger$. The second-order statistics of the EEG-derived time series and those of the BOLD can be compared. Conversely, the relation provides a means to learn about A: the fMRI FC and EEG-derived FC make constraint on A, and A is learned by enforcing this match.

This work presents an approach of this strategy by using EEG-fMRI FC matching to inform A. Specifically, an anatomical baseline A_0 is derived from a T1-MRI boundary-element model (BEM) (Mosher et al., 1999), which carries 10–30% systematic error from uncertain tissue conductivities and imperfect head-compartment geometry (Vorwerk et al., 2014). For each subject, fMRI provides a parcel-level FC matrix, which we adopt as the information source. We use it to constrain the fol-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

lowing model of the EEG parcel FC. The anatomical baseline receives an additive low-rank ($r < C$) correction, $A = A_0 + UV^\top$. A known atlas-based parcellation (Fischl et al., 2004) composed with the corrected pseudoinverse defines a linear EEG-to-parcel operator B; the EEG-derived parcel FC is then $B\Sigma_y B^\top$. The pair (U, V) is estimated subject-wise by minimizing the discrepancy between $B\Sigma_y B^\top$ and the fMRI-derived FC. We name this construction *Lead-Field Low-Rank Correction (LF-LoRA)*.

Aligning the EEG-side and fMRI-side parcel FCs has been studied at the cohort level (Hipp et al., 2012; Wirsich et al., 2020; Nentwich et al., 2020; Wirsich et al., 2021). We use the same alignment as a per-subject training objective. More discussions, as well as quantiative results are reported in §4.2. The structure learned by LF-LoRA is subject-specific and remains stable across sessions. We validate the subject specificity of the learned model in the fingerprinting protocol of Finn et al. (2015), recently adapted to EEG (Yang et al., 2022; Kampel et al., 2024; Donoso, 2025). Beyond the alignment metric, the correction $UV^\top$ can be read on cortex: its dominant components are localized on cortex and expose subject-specific structure that the anatomical baseline does not capture.

A resting-state EEG/fMRI session yields long time series, on the order of 10^5 EEG samples and 10^3 fMRI samples per subject; the instantaneous covariance matching used in the loss above ignores how each signal correlates with itself at later times. At delay $\tau > 0$, the EEG signal has a second-order statistic that pairs its value at t with its value at $t + \tau$; computing the same statistic on the fMRI parcel time series gives a matching reference, and we use these pairs as additional supervision beyond the zero-lag loss. In the framework of LF-LoRA, a lag-τ can be naturally implemented by a linear operator: $y_{t+\tau} = \Phi A x_t$, $\Phi \in \mathbb{R}^{C\times C}$. The loss is augmented to enforce parcel-FC agreement at lags $\tau > 0$. Similar to the correction parameters (U, V), the time-lag operator Φ admits a per-subject reading: ΦA exposes lag-τ coupling between EEG channels and cortical regions distant from each channel's lead-field support.

2 Related Work

Sensor-to-source reconstruction is a classical problem. MNE (Gramfort et al., 2014) and eLORETA (Pascual-Marqui et al., 1994) are regularized linear inverses built entirely on the anatomical baseline $A_{0,s}$ with no dependence on subject data, and serve as the classical reference point in §4.2. LCMV beamforming (Van Veen et al., 1997) likewise leaves $A_{0,s}$ untouched but constructs an adaptive spatial filter from the sensor covariance Σ_y; its data dependence sits at the sensor side, in interference and noise suppression. The ConvDip family (Hecker et al., 2021; Jiao et al., 2024; Morik et al., 2024; Yokoyama et al., 2024) is a learned-network parameterization for the same sensor-to-source problem. None of these is trained against an FC-alignment objective.

A separate line of work skips source reconstruction and aligns second-order summaries: an FC built from EEG is compared with an FC built from fMRI. Hipp et al. (2012) and Wirsich et al. (2020) take the EEG-side FC to be Hilbert-envelope correlations of band-limited EEG; under this construction, reported EEG–fMRI parcel-FC agreement falls in the range $r \approx 0.1$–0.2. The estimator is phase-insensitive and narrow-band: it discards the relative phase between regions and operates on a single frequency window per estimate. The lead field A does not enter the model, hence no per-subject correction can be fit against an FC reference within this class. Donoso (2025) also skips source reconstruction but in a different way: a foundation model predicts fMRI directly from EEG, without constructing an FC on either side.

The EEG-to-parcel map in our model is a linear matrix $B \in \mathbb{R}^{P\times C}$ assembled from the corrected lead field A and a parcel projection. The predicted parcel FC then has the form $B\Sigma_y B^\top$, where Σ_y is the subject's own EEG sensor covariance. For every subject, the predicted FC therefore contains Σ_y as an explicit factor and cannot be produced from the parameters of B alone. This dependence is imposed by the explicit form of the prediction. The dependency structure is weak in parameter-rich nonlinear models.

When placed in this setting (§4.2), the loss admits weights under which the predicted FC is almost entirely determined by the weights, with the model's dependence on the subject's own EEG nearly vanishing. Under the FC-alignment loss, the linear EEG-to-parcel parameterization additionally excludes this degenerate class from our hypothesis space.

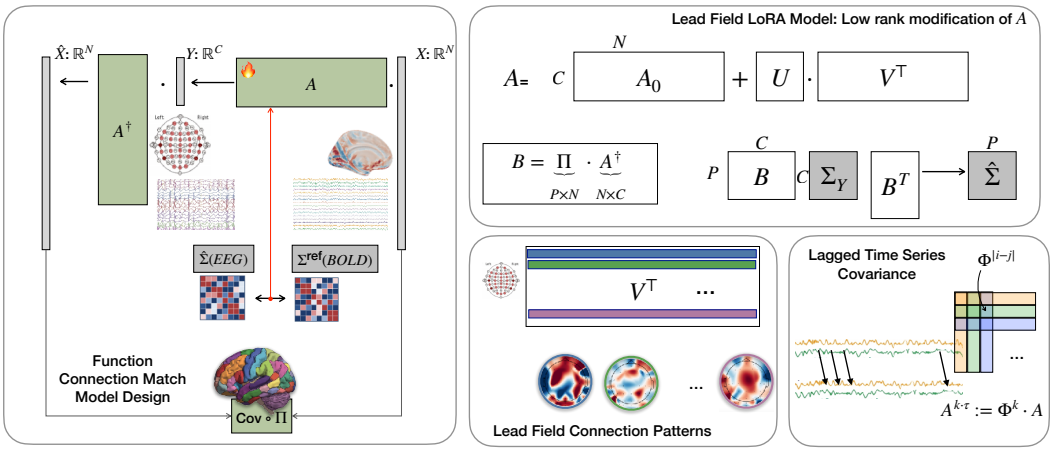


Figure 1: Method illustration; four panels. **(a) Left:** schematic of the computational structure. Red arrows: supervision information flow; Arrows: computation flow; green nodes: operators. **(b) Top-right:** LF-LoRA's parameterization of the EEG-to-parcel covariance $\hat{\Sigma}$ (instantaneous case). **(c) Bottom-left:** the learned LF correction $V^\top$ visualized on cortex. **(d) Bottom-right:** the lag extension of LF-LoRA: matching between lagged covariances of EEG and fMRI.

Subject fingerprinting. Finn et al. (2015) introduced a subject-level identification protocol on FC matrices: each subject's FC from one session is matched against the FCs of all subjects from a held-out session, with a correct identification when the same subject's other-session FC is the closest match. The protocol has since been transported to EEG (Yang et al., 2022; Kampel et al., 2024). We adopt it in §4.4 as an external probe, asking whether adding the per-subject correction $U_s V_s^\top$ to $A_{0,s}$ improves the identification of subjects.

3 Method

Computational Model. Fig. 1 illustrates the overall design of the method. The signal model is the standard linear forward map as in Fig. 1(a): $y_t \approx A\, x_t$,where $x_t \in \mathbb{R}^N$ represents the cortical source amplitudes at time t, $y_t \in \mathbb{R}^C$ the scalp potentials, $A \in \mathbb{R}^{C\times N}$ the lead field. In our computational model, the pseudoinverse $A^\dagger$ represents an estimation of the source amplitudes from EEG as $\hat{x} = A^\dagger y$ (leftmost bar in the top-left panel in Fig. 1). A baseline A_0 can be derived using a boundary-element method (BEM) (Mosher et al., 1999). Due to uncertain tissue conductivities and imperfect head-compartment geometry, the BEM-derived A_0 has a systematic error of roughly 10–30% (Vorwerk et al., 2014), which we plan to calibrate against paired fMRI data.

The reference parcel functional connectivity (FC) $\Sigma^{\text{ref}} \in \mathbb{R}^{P\times P}$ is the Pearson correlation matrix of P parcel time series from the BOLD recording of a subject, where the parcel-aggregation matrix $\Pi \in \mathbb{R}^{P\times N}$ groups cortical sources into parcels (bottom right of the panel). We use the Destrieux atlas, so $P = 148$ (Destrieux et al., 2010). The corresponding EEG-side parcel FC takes the form $\hat{\Sigma} = B\,\Sigma_y\,B^\top \in \mathbb{R}^{P\times P}$,with $\Sigma_y \in \mathbb{R}^{C\times C}$ the empirical sensor covariance and $B \in \mathbb{R}^{P\times C}$ a prediction operator built from A (defined below). The calibration objective is $\hat{\Sigma} \approx \Sigma^{\text{ref}}$, which is performed per subject.

For subject s, we parameterize the lead field as

$$A_s = A_0 + U_s V_s^\top, \qquad U_s \in \mathbb{R}^{C\times r}, \quad V_s \in \mathbb{R}^{N\times r}, \tag{1}$$

with r a small integer (typically $r = 10$; ablation in Appendix C). The form is additive because the BEM-derived baseline A_0 is order-of-unity correct on the geometry and conductivity priors. The form is low rank because the per-subject data budget (a few minutes of resting EEG) is far smaller than the $CN \approx 3.5 \times 10^5$ entries of a full $C \times N$ operator, but suffices to identify a few dominant correction modes. The correction is per-subject because the systematic error of A_0 varies across subjects (skull conductivity, electrode placement, head geometry). From A_s we obtain the prediction

operator

$$B_s \;=\; \Pi A_s \big(A_s A_s^\top + \lambda I_C\big)^{-1} \;\in\; \mathbb{R}^{P\times C}, \tag{2}$$

where Π is the parcel-aggregation matrix and the remaining factor is a regularized pseudoinverse. The calibration minimizes

$$\mathcal{L}_s \;=\; \big\| \hat{\Sigma}_s - \Sigma^{\text{ref}_s} \big\|_F^2 \;=\; \big\| B_s\, \Sigma_{y,s}\, B_s^\top - \Sigma^{\text{ref}_s} \big\|_F^2. \tag{3}$$

Identifiability. In our learning framework, the dependency chain from (U, V) (omit subject index for clarity) to the calibration objective $\mathcal{L}$ is as follows:

$$(U,V) \xrightarrow[\text{(i) bilinear}]{A_0+UV^T} A \xrightarrow[\text{(ii) inversion}]{A^\dagger} A^\dagger \xrightarrow[\text{(iii) linear}]{\Pi\,\cdot} B \xrightarrow[\text{(iv) quad}]{\cdot\,\Sigma_y\,\cdot^T} \widehat{\Sigma} \xrightarrow[\text{(v) quad}]{\|\cdot-\Sigma^{\text{ref}}\|_F^2} \mathcal{L}. \tag{4}$$

There are two inherent ambiguities that we will resolve by fixed design choices. In step (i), the parameterization (1) admits a free rank-r transform that leaves A unchanged (gauge freedom): $(U, V) \mapsto (UM^{-1},\ VM^\top)$ has no effect on A and any downstream objects. We fix this freedom by orthonormalizing the columns of U_s after each gradient step (Appendix G). On the other hand, step (iv) is a quadratic operation that admits multiple equivalence classes of solutions in the space of B. Concurrent work (Anonymous, 2026) gives a closed-form lower bound on the inter-class Frobenius distance, $\|B_0 - B_1\|_F^2 \;\geq\; \frac{4\,\lambda_{\min(P,C)}(\Sigma_s^{\text{ref}})}{\sigma_{\max}(\Sigma_y)}$, where $\lambda_{\min(P,C)}(\Sigma_s^{\text{ref}})$ is the smallest among the top $\min(P, C)$ eigenvalues of the reference parcel covariance and $\sigma_{\max}(\Sigma_y)$ is the largest eigenvalue of the sensor covariance. Both are computable before training, so the right-hand side serves as an a-priori per-subject diagnostic of how cleanly the matching equation discriminates among classes.

The separation of classes in B-space implies a corresponding separation in A-space, by tracing back through the pseudo-inverse $A \mapsto \Pi A^\dagger$. This map is locally Lipschitz, with constant controlled by $\sigma_{\min}(A)$, which can in turn be controlled by $\sigma_{\min}(A_0)$ and the correction term $UV^\top$. This implies that if $\|UV^\top\|_{\text{op}}$ is kept inside an admissible range (a controlled fraction of $\sigma_{\min}(A_0)$ of Anonymous (2026)), the gradient flow cannot cross into a neighbouring class. Hence gradient descent initialized at A_0 stays in the equivalence class of A_0 as long as $\|UV^\top\|_{\text{op}}$ remains in the admissible range.

3.1 Extension to lagged second-order supervision

The covariance matrices $\hat{\Sigma}$ and Σ^{ref} used in the matching are computed from two signals coincident times. Synchronized EEG and fMRI recordings are used to obtain these statistics. Since the recording time is much longer than the timescale of FC dynamics. We can also obtain from the same data a family of lagged covariance matrices. Specifically, let the existing covariance of a Z (a generic multivariate signal) $\Sigma_Z = \mathbb{E}[Z_t Z_t^\top]$ be the *zero-lag* reference, and denote it as $\Sigma_{Z,0}$. Then for a short lag τ, the *lag-τ* covariance is

$$\Sigma_{Z,\tau} = \mathbb{E}[Z_t Z_{t+\tau}^\top]. \tag{5}$$

In terms of neural sources, the lag-τ covariance tells how parcel amplitudes at time t co-vary with parcel amplitudes at time $t + \tau$. The lag-τ covariance on the EEG side and that of the fMRI side are then two different views of the same underlying source process. We employ the same principle of matching these two views to extend the calibration objective to a family of lag-τ losses, one for each τ.

Taking advantage of the linear structure of the operators of the model, we can decouple the lag-τ dynamics using a small operator $\Phi_s \in \mathbb{R}^{C\times C}$ (per subject s):

$$A_{s,j\cdot\tau} = (\Phi_s)^j A_s, \tag{6}$$

where the new operator represents a unit-τ temporal evolution operator. For a delay of $t = j \cdot \tau$, the lagged lead field is obtained as $\Phi_s^j A_s$. Given K lag steps, we can then setup a multi-lag loss

$$L_s^{\text{lag}}(\Phi_s, U_s, V_s) = \sum_{j=0,j'\geq j}^{K-1} \big\| B_{s,j}\, \hat{\Sigma}_{y,(j,j')}\, B_{s,j'}^\top - \Sigma^{\text{ref}}_{s,(j,j')} \big\|_F^2, \tag{7}$$

where all time units τ are omitted for brevity. In addition to the zero-lag covariance matching, the loss (7) provides $(K-1)(K-2)/2$ more off-diagonal supervision signals ($j \neq j'$) that constrain the temporal evolution. We will see in empirical study (§4.3) that learning Φ_s primarily improves these off-diagonal blocks ($j \neq j'$), which do not enter the zero-lag loss.

4 EXPERIMENTS

4.1 DATASET AND PREPROCESSING

We use the LEMON cohort (Babayan et al., 2019): of the 227 participants we retain $N = 203$ with usable resting EEG, paired resting fMRI, and a T1-derived BEM lead field. To share a single sensor dimensionality across the cohort, we keep only the EEG channels valid in every retained subject, giving a $C = 43$-channel common montage. The experiments below need this shared channel set so that the forward operator has a uniform shape across the cohort. The fMRI reference Σ_s^{ref} is the Pearson correlation matrix of $P = 148$ Destrieux (Destrieux et al., 2010) parcel time series. Acquisition specs, QC criteria, the EEG / fMRI preprocessing pipelines, and the common-montage construction are in Appendix B.

4.2 WITHIN-SUBJECT CROSS-STATE ALIGNMENT OF EEG-DERIVED PARCEL FC

We use this experiment to ask whether the per-subject correction $A_s = A_{0,s} + U_s V_s^\top$ carries information about the *subject* rather than about the cortical state in which it was fit. Each LEMON subject contributes two resting recordings, eyes-closed (EC) and eyes-open (EO), at the same scanner visit but in distinct cortical states; fitting A_s on one state and evaluating on the other holds the subject fixed while changing the cortical state. Agreement on the held-out state is evidence that A_s has captured a stable per-subject feature of the EEG-to-source map rather than a state-specific overfit.

Specifically, A_s is fit on one resting EEG session (EC/EO), and on the held-out session the EEG-side parcel FC $\hat{\Sigma}_s = B_s \, \hat{\Sigma}_y \, B_s^\top$ is scored against the fMRI-derived reference Σ_s^{ref} via the upper-triangular Pearson r. Both directions (EC→EO and EO→EC) are run, and the cohort score is their average.

We run the same protocol on two existing methods. (i) With the classical inverses MNE (Gramfort et al., 2014) and eLORETA (Pascual-Marqui et al., 1994) on the anatomical baseline $A_{0,s}$, no parameter is trained against the FC-alignment loss; their alignment is that of the BEM template $A_{0,s}$, and the difference from ours is the contribution of $U_s V_s^\top$. (ii) With the amplitude-envelope FC of Hipp et al. (2012) and Wirsich et al. (2020) as the predictor, the alignment is computed from inter-channel envelope correlations in fixed bands (α, β). Cohort-level mean Pearson r values with 95% confidence intervals are collected in Table 1.

Table 1: EEG-fMRI parcel FC alignment on LEMON (203 subjects). Mean Pearson r between predicted and reference upper-triangular FC entries, cross-session protocol, direction-averaged.

Method	Mean $r \pm$ std	95% CI
MNE (frozen $A_{0,s}$)	0.053 ± 0.020	$[0.018,\ 0.095]$
eLORETA (frozen $A_{0,s}$)	0.049 ± 0.019	$[0.016,\ 0.089]$
α-envelope (8–13 Hz)	0.118 ± 0.056	$[0.022,\ 0.232]$
β-envelope (13–30 Hz)	0.150 ± 0.069	$[0.014,\ 0.277]$
Ours (EC → EO)	$\mathbf{0.628 \pm 0.131}$	—
Ours (EO → EC)	$\mathbf{0.630 \pm 0.133}$	—

MNE and eLORETA invert $A_{0,s}$ and use no fMRI information; the difference between their rows and ours in Table 1 therefore measures the effect of the fMRI-supervised correction $U_s V_s^\top$, not a difference between inverse methods.

A parameter-rich nonlinear model trained per-subject against the same FC-alignment loss can saturate the metric without learning any EEG-to-source structure. We probed this with a ConvDip backbone (Hecker et al., 2021), originally a cross-subject source localizer, retrofitted to per-subject training against the FC-alignment loss: on a single-subject smoke test ($N = 1$, sub-032301) the loss saturates at $r \approx 0.97$, yet the predicted parcel FC is unchanged when the EEG input is time-shuffled, replaced by a different subject's recording, or replaced by i.i.d. Gaussian noise. The model has memorized this subject's FC and the input plays no role. Whenever the model class is rich enough to encode an input-independent template, the FC-alignment loss can in principle be driven to zero by such a template.

Our linear low-rank correction $A_s = A_{0,s} + U_s V_s^\top$ sits between these two regimes: per-subject parameters are admitted, but the rank budget is small enough that the predicted FC must still flow through the EEG-to-source map, since Σ_y enters $B_s \Sigma_y B_s^\top$ as an explicit factor.

4.3 Multi-lag alignment of EEG-derived parcel FC

Section 3.1 introduced the multi-lag loss and the evolution operator Φ_s. We test the time-lagged extension of the FC alignment protocol in this experiment. The (7) loss is computed with $d = 5$ lag bins per subject. Different lag steps τ are scanned over $\{350, 700, 1050\}$ EEG samples, which correspond to $1400ms, 2800ms, 4200ms$ in real time, and equivalently to $\{1, 2, 3\}$ multiples of the fMRI TR. The detailed experiment setup is in Appendix D.

Table 2 reports the cohort-averaged Pearson correlation ρ between predicted and reference parcel covariances at initialization and after fitting, separately for the diagonal block ($j = 0$) and the off-diagonal blocks ($j \geq 1$). The diagonal-block correlation moves from 0.649 at initialization to 0.705 at $\tau = 700$, a small but consistent improvement. The off-diagonal blocks move much more: from $+0.042$ at initialization to $+0.216$ at $\tau = 700$, and to $+0.188$ at $\tau = 1050$. The zero-lag loss does not involve the off-diagonal blocks; these are the blocks on which fitting Φ_s changes ρ most ($\Delta_{\text{off}} = +0.174$ against $\Delta_{\text{diag}} = +0.055$ at $\tau = 700$).

At $\tau = 350$ (one TR) the off-block score ends at -0.109; at $\tau = 700$ and $\tau = 1050$ (two and three TR) it ends at $+0.216$ and $+0.188$. Within the scanned range, any τ of at least two TR gives comparable results.

Table 2: Temporal-extension alignment. Mean over 203 subjects of ρ_{diag} (diagonal-block correlation) and ρ_{off} (off-block correlation) at init vs final. $\Phi_s \in \mathbb{R}^{43\times 43}$.

Variant	$\rho_{\text{diag}}^{\text{init}}$	$\rho_{\text{diag}}^{\text{final}}$	Δ_{diag}	$\rho_{\text{off}}^{\text{init}}$	$\rho_{\text{off}}^{\text{final}}$	Δ_{off}
$\Phi_s, \tau = 350$	0.649	0.690	+0.041	−0.214	−0.109	+0.105
$\Phi_s, \tau = 700$	**0.649**	**0.705**	**+0.055**	**+0.042**	**+0.216**	**+0.174**
$\Phi_s, \tau = 1050$	0.649	0.705	+0.055	+0.001	+0.188	+0.187

4.4 Subject fingerprinting

Protocol. For each pair of subjects (i, j), subject i's EO-segment EEG and subject j's EC-segment EEG are passed through the same family of inverse operators $B(\alpha)$ and projected onto 148 parcels, yielding two parcel-FC vectors $\widehat{F}_i^{\text{EO}}(\alpha)$ and $\widehat{F}_j^{\text{EC}}(\alpha)$. We take the Pearson correlation of their flattened upper triangles, giving the 203×203 cross-subject cross-session similarity matrix

$$S_{ij}^{\alpha} = \text{Pearson}\Big(\text{vec}_{\triangle}\widehat{F}_i^{\text{EO}}(\alpha),\ \text{vec}_{\triangle}\widehat{F}_j^{\text{EC}}(\alpha)\Big), \quad i, j = 1, \ldots, 203.$$

At $\alpha = 0$ the operator $B(\alpha)$ reduces to the anatomy-only inverse ($A_{0,s}$, MNE style); at $\alpha = 1$ it uses the version with our learned correction, where the correction term is fitted during training against *the subject's own fMRI parcel-FC* as the reference target, so that the EEG-projected parcel-FC is driven toward the same subject's fMRI parcel-FC. Intermediate $\alpha \in (0, 1)$ is the overall scaling between these two endpoints. All other regularization (the ridge λ, etc.) is held at the MNE setting and does not vary with α, so that $\alpha = 0$ strictly recovers the MNE numerics.

We report two families of metrics. **Identification (1:N):** whether $\arg\max_j S_{ij}^{\alpha}$ hits the diagonal (top-1) for each row i. **Verification (1:1):** treating the 203 diagonal entries as positives and the $203 \times 202 = 41006$ off-diagonal entries as negatives, sweeping a threshold to obtain a pooled ROC, and reporting AUC. The 95% CI is from a cluster bootstrap (with-replacement on the 203 subjects, 1000 reps). We additionally report differential power $\text{DP} = \overline{S}_{\text{diag}} - \overline{S}_{\text{off}}$ as a complementary continuous summary of the same matrix. We sweep $\alpha \in \{0, 0.1, \ldots, 1.0\}$, 11 values in total. The baselines are the MNE and eLORETA inverse operators (each using the subject's own $A_{0,s}$, no learned correction), independent of α.

Result. Figure 2 shows the results. As α increases from 0 to 1, the ROC curves in panel A move monotonically toward the upper-left corner: within-subject and cross-subject similarities become

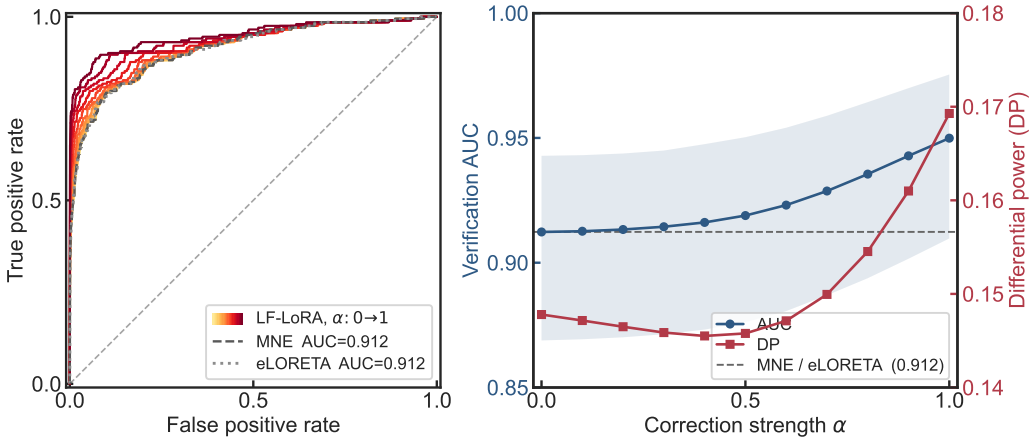


Figure 2: Subject fingerprinting on 203 subjects, swept over correction strength $\alpha \in \{0, 0.1, \ldots, 1.0\}$. **(A)** Verification ROC curves at the 11 values of α (curve colour: light yellow for $\alpha = 0$, deep red for $\alpha = 1$, see legend strip). MNE (grey dashed) and eLORETA (grey dotted) baselines computed without the learned correction; diagonal grey dashed line is chance. **(B)** Pooled verification AUC (steel blue, left axis) and differential power DP $= \overline{S}_{\text{diag}} - \overline{S}_{\text{off}}$ (cordovan red, right axis), as functions of α. Shaded blue band: 95% cluster-bootstrap CI on AUC (1000 reps over the 203 subjects). Horizontal grey dashed line: mean MNE / eLORETA AUC reference (0.912).

more separable as the correction is scaled up. The MNE and eLORETA reference curves nearly coincide with the $\alpha = 0$ ROC, confirming that the $\alpha = 0$ endpoint is matched exactly to the anatomy-only baseline.

Panel B plots the AUC of these curves (blue, left axis), which increases monotonically from 0.912 at $\alpha = 0$ (grey dashed line: MNE / eLORETA) to 0.950 at $\alpha = 1$ and stays below 1 over the whole range; we therefore use it as the main measure of how much the correction improves subject discrimination over $A_{0,s}$. DP (red, right axis) decreases slightly to 0.146 on $\alpha \in [0, 0.4]$ and then increases to 0.169 at $\alpha = 1$. The two curves measure different things: AUC depends only on the ranking of the within-subject score among the scores of each row, whereas DP is the difference of the two means; for small α the correction shifts both kinds of score by similar amounts, so the difference of means can decrease while the rankings already improve. Identification top-1 reaches 1.000 (203/203) for $\alpha \geq 0.5$, against $\sim$ 0.62 for the baselines; since it is constant on $[0.5, 1]$, it does not resolve the dependence on α, and we do not use it as the main metric.

4.5 Visualization

Representative correction modes. Figure 3 decomposes the correction $U_s V_s^\top$ for a high-r_s subject (sub-032331, $r_s = 0.917$) into r rank-1 terms $u_{s,k} v_{s,k}^\top$, of which the top three are shown in descending order of $\|v_{s,k}\|$ (since U_s is column-orthogonalized during training, this ordering coincides with descending $\|u_{s,k} v_{s,k}^\top\|_F$). In each row, column 1 plots $u_{s,k} \in \mathbb{R}^C$ as an EEG scalp topomap; columns 2–5 render $v_{s,k} \in \mathbb{R}^N$ on the fsaverage inflated cortex from four views, LH/RH $\times$ lateral/medial (the oct6→ico7 nearest-neighbour expansion is for display only). Each pair $(u_{s,k}, v_{s,k})$ is sign-flipped jointly so that the $|\cdot|$-largest component of $v_{s,k}$ is positive, fixing the overall sign.

Temporal-layer visualization. Figure 4 shows the per-channel parcel projection of the learned propagator Φ_s at the training lag τ. For two channels (Oz and AFz), and for each of two exemplar subjects, two parcel-aggregated head maps are plotted: "Aggregated Source" $= \Pi A_s[\text{ch}, :]$ is the lead-field projection without Φ_s; "Aggregated Source – Lag" $= \Pi \Phi_s A_s[\text{ch}, :]$ is the same projection composed with Φ_s at the training lag. Their difference equals $\Pi(\Phi_s - I)A_s$, i.e. the correction Φ_s applies to the lead field on that channel row, parcel-aggregated.

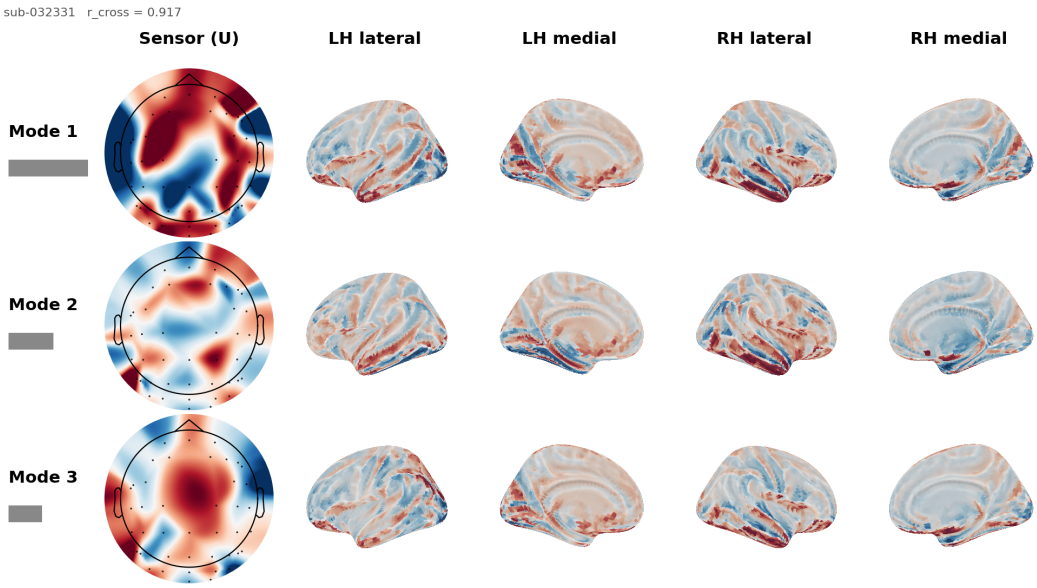


Figure 3: Representative correction modes for a high-r_s subject (sub-032331, $r_s = 0.917$). Top-3 columns of the trained (U_s, V_s) sorted by $\|v_{s,k}\|$ descending; the left grey bar in each row is proportional to $\|v_{s,k}\| / \max_j \|v_{s,j}\|$. Column 1 (`Sensor (U)`): $u_{s,k}$ on the EEG cap. Columns 2–5: $v_{s,k}$ on fsaverage inflated cortex, LH/RH $\times$ lateral/medial. Colormap `RdBu_r`; per-row symmetric limits at max $|u_{s,k}|$ (sensor) and at the 98th percentile of $|v_{s,k}|$ (cortex).

For each channel, the version with Φ_s shows non-trivial colour at parcels outside the channel's own lead-field support, structures absent from the baseline. The AFz baseline is concentrated frontally; with Φ_s applied, additional structured patches emerge on the lateral and posterior parcels. The Oz baseline is concentrated posteriorly; with Φ_s applied, non-zero signal appears in frontal and lateral-temporal regions. Since the difference between the two panels is $\Pi(\Phi_s - I)A_s[\text{ch}, :]$, these patches show that Φ_s relates the channel at time $t + \tau$ to parcels outside its lead-field support at time t.

5 Conclusion

We propose a subject-specific calibration of the anatomy-derived lead field A_0 via a low-rank additive correction $UV^\top$, trained to align EEG-projected parcel functional connectivity with the same subject's fMRI parcel-FC. The full training chain is summarized in (4). When A_0 serves as a warm start, the magnitude of the learned correction is subject to theoretical bounds that follow from the chain structure. We further extend the framework to the lagged-covariance regime: replacing instantaneous second-order supervision with lagged covariance matrices at multiple delays, the dynamics supervision signal is absorbed into a temporal operator Φ that encodes directed temporal relationships between parcels.

On the LEMON cohort (203 participants), the learned correction carries subject-specific structure correlated with the same subject's fMRI FC. Within-subject / cross-subject separability of EEG-derived parcel-FC increases monotonically with the correction magnitude α, reaching a pooled verification AUC of 0.950 at $\alpha = 1$, against 0.912 for the anatomy-only baselines MNE and eLORETA. This shows that paired EEG-fMRI data provides sufficient calibration signal to bring EEG-derived connectivity to a level where it supports reliable individual identification. Decomposing the learned correction and inspecting Φ yields interpretable correction modes and lagged inter-regional connectivity patterns, offering a qualitative view of the subject-specific structure encoded by the calibration.

The present study is a feasibility check conducted on healthy adult cross-session data. The corrected lead field has not been validated as a clinical biomarker; establishing clinical utility would require test-retest validation and clinical-label association on patient populations. The low-rank correction

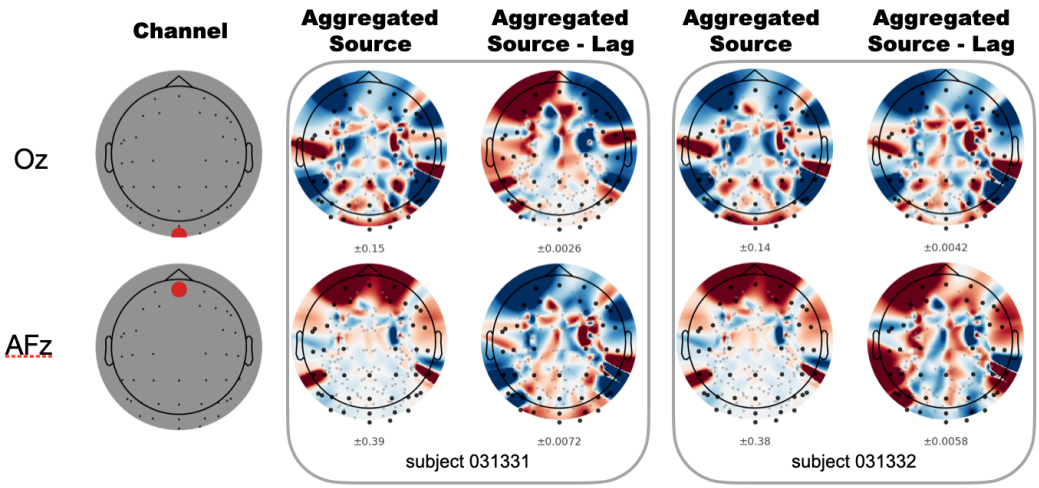


Figure 4: Propagator Φ_s at the training lag τ, shown as per-channel parcel maps for two exemplar subjects. Rows: channel (Oz, AFz; the locator column marks channel position on the 1020 montage). For each subject, two panels are shown: "Aggregated Source" $= \Pi A_s[\text{ch}, :]$ (parcel-aggregated lead field, no Φ_s); "Aggregated Source – Lag" $= \Pi\Phi_s A_s[\text{ch}, :]$ (lead field composed with Φ_s at the training lag). Colour map RdBu_r (red positive, blue negative); each panel auto-scaled to its own 98% percentile. Parcels outside the channel's own lead-field support that carry non-trivial colour in the "Lag" panel correspond to the spatial footprint of the temporal coupling captured by Φ_s at the training lag.

framework applies in principle to any linear inverse problem supervised by cross-modal second-order statistics, but its behavior outside the EEG-fMRI setting remains to be tested.

AI USE STATEMENT

In this work, we used generative AI tools (Claude Opus, Anthropic) for the following tasks with required disclosure. We used them to read and digest the literature. We used them to assist in designing the experiments, implementing the methods and the preprocessing pipeline of the LEMON data, analysing and interpreting the results, and drafting the experiment reports. We have not used generative AI tools for translation, or for qualitative or thematic data analysis. Additionally, we used generative AI tools to draw the pictorial elements in Figure 1; the composition of the figure and the arrangement of its elements were designed by the authors. We have reviewed all AI-assisted work; the code and data behind the figures were checked and cross-validated. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work is mostly methodological and develops mathematical models. Its experiments use a publicly released and widely used medical dataset (LEMON, Babayan et al., 2019). We are not aware of any ethical concern raised by this work.

REPRODUCIBILITY STATEMENT

The model, its loss and the lag extension are specified in Section 3; the hyperparameters and the training protocols are given in Appendices D and G, and the ablation over the number of lag bins d and the correction rank r in Appendix C. The experiments use the public LEMON dataset (Babayan et al., 2019); the subject selection and the EEG and fMRI processing steps are described in Section 4.1 and Appendix B. The supplementary material contains `lf_lora_demo.py`, a self-contained script, depending only on NumPy and PyTorch, that implements the correction $A_s = A_{0,s} + U_s V_s^\top$ with

the gauge fixing of U_s, the regularised inverse B_s, the block-Toeplitz multi-lag loss, the training loop and the evaluation of parcel FC. The script runs these steps on synthetic data of reduced dimension; it demonstrates the computation and does not reproduce the numbers reported on LEMON. A `README` in the supplementary material lists the software requirements and explains how to obtain LEMON.

References

Anonymous. Learning a forward operator from cross-modal second-order statistics. In *Submitted to The Fifteenth International Conference on Learning Representations*, 2026. under review.

Anahit Babayan, Miray Erbey, Deniz Kumral, Janis D. Reinelt, Andrea M. F. Reiter, Josefin Röbbig, H. Lina Schaare, Marie Uhlig, Alfred Anwander, Pierre-Louis Bazin, Annette Horstmann, Leonie Lampe, Vadim V. Nikulin, Hadas Okon-Singer, Sven Preusser, André Pampel, Christiane S. Rohr, Julia Sacher, Angelika Thöne-Otto, Sabrina Trapp, Till Nierhaus, Arno Villringer, and Daniel S. Margulies. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Scientific Data*, 6:180308, 2019. doi: 10.1038/sdata.2018.308.

Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995. doi: 10.1002/mrm.1910340409.

Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1):1–15, 2010. doi: 10.1016/j.neuroimage.2010.06.010.

Maël Donoso. Neural networks and foundation models: two strategies for EEG-to-fMRI prediction. *Frontiers in Systems Biology*, 5:1715692, 2025. doi: 10.3389/fsysb.2025.1715692.

Emily S. Finn, Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671, 2015. doi: 10.1038/nn.4135.

Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, Larry J. Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, and Anders M. Dale. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22, 2004. doi: 10.1093/cercor/bhg087.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S. Hämäläinen. MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–460, 2014. doi: 10.1016/j.neuroimage.2013.10.027.

Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993. doi: 10.1103/RevModPhys.65.413.

Lukas Hecker, Rebekka Rupprecht, Ludger Tebartz Van Elst, and Jürgen Kornmeier. ConvDip: A convolutional neural network for better EEG source imaging. *Frontiers in Neuroscience*, 15: 569918, 2021. doi: 10.3389/fnins.2021.569918.

Joerg F. Hipp, David J. Hawellek, Maurizio Corbetta, Markus Siegel, and Andreas K. Engel. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, 15 (6):884–890, 2012. doi: 10.1038/nn.3101.

Meng Jiao, Xiaochen Xian, Boyu Wang, Yu Zhang, Shihao Yang, Spencer Chen, Hai Sun, and Feng Liu. XDL-ESI: Electrophysiological sources imaging via explainable deep learning framework with validation on simultaneous EEG and iEEG. *NeuroImage*, 299:120802, 2024. doi: 10.1016/j.neuroimage.2024.120802.

Nikolas Kampel, Farah Abdellatif, N. Jon Shah, Irene Neuner, and Jürgen Dammers. Contrastive learning for neural fingerprinting from limited neuroimaging data. *Frontiers in Nuclear Medicine*, 4:1332747, 2024. doi: 10.3389/fnume.2024.1332747.

Marco Morik, Ali Hashemi, Klaus-Robert Müller, Stefan Haufe, and Shinichi Nakajima. Enhancing brain source reconstruction by initializing 3D neural networks with physical inverse solutions. arXiv preprint arXiv:2411.00143, 2024.

John C. Mosher, Richard M. Leahy, and Paul S. Lewis. EEG and MEG: forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46(3):245–259, 1999. doi: 10.1109/10.748978.

Maximilian Nentwich, Lei Ai, Jens Madsen, Qawi K. Telesford, Stefan Haufe, Michael P. Milham, and Lucas C. Parra. Functional connectivity of EEG is subject-specific, associated with phenotype, and different from fMRI. *NeuroImage*, 218:117001, 2020. doi: 10.1016/j.neuroimage.2020.117001.

Roberto D. Pascual-Marqui, Christoph M. Michel, and Dietrich Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18(1):49–65, 1994. doi: 10.1016/0167-8760(84)90014-X.

Jonathan D. Power, Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3):2142–2154, 2012. doi: 10.1016/j.neuroimage.2011.10.018.

Barry D. Van Veen, Wim van Drongelen, Moshe Yuchtman, and Akifumi Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880, 1997. doi: 10.1109/10.623056.

Johannes Vorwerk, Jae-Hyun Cho, Stefan Rampp, Hajo Hamer, Thomas R. Knösche, and Carsten H. Wolters. A guideline for head volume conductor modeling in EEG and MEG. *NeuroImage*, 100:590–607, 2014. doi: 10.1016/j.neuroimage.2014.06.040.

Jonathan Wirsich, Anne-Lise Giraud, and Sepideh Sadaghiani. Concurrent EEG- and fMRI-derived functional connectomes exhibit linked dynamics. *NeuroImage*, 219:116998, 2020. doi: 10.1016/j.neuroimage.2020.116998.

Jonathan Wirsich, João Jorge, Giannina Rita Iannotti, Elhum A. Shamshiri, Frédéric Grouiller, Rodolfo Abreu, François Lazeyras, Anne-Lise Giraud, Rolf Gruetter, Sepideh Sadaghiani, and Serge Vulliemóz. The relationship between EEG and fMRI connectomes is reproducible across simultaneous EEG-fMRI studies from 1.5T to 7T. *NeuroImage*, 231:117864, 2021. doi: 10.1016/j.neuroimage.2021.117864.

Yao-Yuan Yang, Angel Hsing-Chi Hwang, Chien-Te Wu, and Tsung-Ren Huang. Person-identifying brainprints are stably embedded in EEG mindprints. *Scientific Reports*, 12:17031, 2022. doi: 10.1038/s41598-022-21384-0.

Hikaru Yokoyama, Naotsugu Kaneko, Noboru Usuda, Tatsuya Kato, Hui Ming Khoo, Ryohei Fukuma, Satoru Oshino, Naoki Tani, Haruhiko Kishima, Takufumi Yanagisawa, and Kimitaka Nakazawa. M/EEG source localization for both subcortical and cortical sources using a convolutional neural network with a realistic head conductivity model. *APL Bioengineering*, 8(4):046104, 2024. doi: 10.1063/5.0226457.

A Notation

Table 3 collects the symbols used in this paper, together with their shapes and typical scales.

Table 3: Notation and scales.

Symbol	Meaning	Scale / shape
s	subject index	$s = 1, \ldots, 203$
C	EEG channels	$C = 43$
N	cortical sources	$N = 8196$
P	parcels (Destrieux atlas)	$P = 148$
T_{EEG}	per-session EEG samples	$\sim 7.5 \times 10^4$ at 250 Hz
T_{fMRI}	per-session fMRI samples	~ 650 at TR $= 1.4$ s
TR	fMRI repetition time	$1.4\,\text{s} = 350$ EEG samples
$y_t \in \mathbb{R}^C$	sensor-level EEG at time t	—
$x_t \in \mathbb{R}^N$	source amplitude vector at time t	—
$\varepsilon_t \in \mathbb{R}^C$	sensor-side noise at time t	—
$A \in \mathbb{R}^{C \times N}$	forward operator ("lead field")	true, unknown
A_0 .	BEM-computed baseline	10–30% systematic error
U_s, V_s	low-rank correction factors	$C \times r$, $N \times r$, $r = 10$
Π	source-to-parcel averaging matrix	$P \times N$
Σ_s^{ref}	fMRI-derived reference parcel FC	$P \times P$, Pearson
Σ_s^{pred}	our predicted parcel FC	$P \times P$
τ	EEG-side lag (time-delay)	samples (ex. $\tau = 2$ TR)
d	number of lags in block covariance	small integer
Φ_s	temporal evolution operator (per subject)	$C \times C$ (fitted)

B Dataset and preprocessing details

LEMON (Babayan et al., 2019) is the publicly released subset of the MPI Leipzig Mind-Brain-Body database. It comprises $N = 227$ healthy adults stratified by design into a younger group ($N = 153$, age 20–35 years, mean 25.1 ± 3.1) and an older group ($N = 74$, age 59–77 years, mean 67.6 ± 4.7). For each participant the release provides high-density EEG, a resting-state fMRI run, and a high-resolution structural T1; Additional semantic information accompanying the recordings is also distributed but is not used here. EEG was acquired with a BrainAmp MR plus amplifier and an active ActiCAP electrode set, 61 scalp channels in a 10-10 layout plus one vertical EOG channel, sampled at 2500 Hz, referenced to FCz with the ground at the sternum. The resting protocol consists of 16 min arranged as 16 alternating 60 s blocks, 8 eyes-closed and 8 eyes-open with eyes-closed first. Resting fMRI was acquired on a 3T Siemens MAGNETOM Verio with a multi-band gradient-echo EPI sequence (multi-band factor 4, TR $= 1400$ ms, TE $= 30$ ms, isotropic voxel size 2.3 mm), giving a single run of 657 volumes (15 min 30 s) per subject. The structural T1 is an MP2RAGE volume at 1 mm isotropic resolution.

LEMON meets the modality requirements of our pipeline at the single-subject level: each participant contributes the structural T1 needed to construct the anatomical baseline $A^{(0)}$, the resting EEG used to fit the per-subject correction, and a paired resting-state fMRI run from which the parcel-FC reference Σ_s^{ref} is computed. The cohort size also supports reporting alignment scores at the cohort level rather than per subject.

Of the 227 LEMON participants we retain $N = 203$. The hard inclusion criterion is the joint availability of three streams: raw resting-state EEG, the paired resting-state fMRI BOLD run, and a usable T1-derived BEM lead field; the remaining exclusions come from a conservative quality-control pass on the EEG side (channel-level peak-to-peak and z-score outliers, ICA failing to isolate stable EOG components) and on the fMRI side (head-motion outliers and `fmriprep` warnings).

On the EEG side, the raw recording for each retained subject is resampled to 250 Hz, mapped to the standard_1005 montage with the position-less VEOG channel dropped, bandpass-filtered to 0.5–45 Hz with a zero-phase FIR, notched at 50 Hz, re-referenced to the common average, and cleaned by automated bad-channel interpolation followed by ICA-based EOG removal. The BEM lead field

for each subject covers 57–61 channels, with the channel set varying mildly across subjects. Cross-subject comparisons require a single sensor dimensionality shared by the whole cohort; to obtain it, we retain only the channels for which every included subject has a valid recording in both the eyes-closed and eyes-open conditions. The resulting common montage has $C = 43$ channels and is used in all downstream training and evaluation. This step is non-standard: most EEG analyses preserve the native per-subject channel set; we accept the per-subject information loss in exchange for a cohort-uniform shape of the forward operator and the sensor covariance.

On the fMRI side, the resting-state run is processed with the standard `fmriprep` pipeline (motion correction, slice-timing correction, co-registration, MNI 2 mm normalization). Parcel time series are extracted as the spatial mean within each of the $P = 148$ regions of the Destrieux atlas (Destrieux et al., 2010), and the parcel-FC reference Σ_s^{ref} is their Pearson correlation matrix.

C d / r ABLATION

Setup. A 30-subject subset of LEMON, single seed. Train on the EC session, evaluate parcel-FC Pearson r on the EO session of the same subject. Spatial-only fit: $\Phi_s = I$ throughout (temporal operator disabled), so the only free parameters are (U_s, V_s) per subject. Optimization uses Adam at learning rate 10^{-3} for 500 steps, with the gauge fix on U_s applied each step. The Tikhonov scale in $B_s = \Pi A_s (A_s A_s^\top + \lambda I_C)^{-1}$ is set per subject and recomputed at every step as $\lambda_s = c \cdot \mathrm{tr}(A_s A_s^\top)/C$ with $c = 10^{-3}$, so λ_s tracks the magnitude of $A_s A_s^\top$ and the conditioning of $A_s A_s^\top + \lambda_s I_C$ is invariant to the overall scale of A_s. The training loss is the block-Toeplitz multi-lag form of §3.1 with d lag-bins per side: bin k on the EEG side is the lag-k EEG covariance at a stride of one EEG sample (4 ms at 250 Hz), bin k on the fMRI side is the lag-k parcel covariance at a stride of one TR (1.4 s). Reported numbers are the cohort mean and across-subject standard deviation.

Table 4: One-dimensional sweeps with everything else fixed at the headline operating point. The $d = 5$, $r = 10$ entry is shared between the two sweeps.

d ($r = 10$)	Mean $r \pm$ std	r ($d = 5$)	Mean $r \pm$ std
1	0.525 ± 0.152	5	0.458 ± 0.141
3	0.527 ± 0.154	10	$\mathbf{0.667 \pm 0.104}$
5	$\mathbf{0.667 \pm 0.104}$	20	0.616 ± 0.145
10	0.525 ± 0.147		

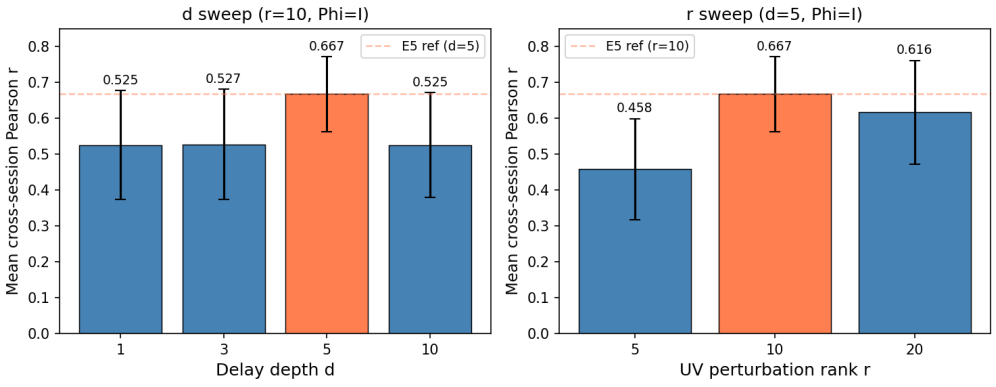


Figure 5: d sweep at $r = 10$ (left) and r sweep at $d = 5$ (right). Bars are the cohort mean; error bars are the across-subject standard deviation on the same 30 subjects. Coral marks the shared headline configuration; the dashed line is its mean.

Reading. On the d axis, $d = 1$, 3, 10 all sit near 0.525 and are within their across-subject std of one another; only $d = 5$ lifts past 0.6. The drop from $d = 5$ to $d = 10$ shows that r does not increase with d under this protocol: $d = 10$ returns to 0.525. On the r axis, $r = 5$ is too low to absorb the

per-subject correction; $r = 20$ is mildly worse than $r = 10$, consistent with the per-subject training budget supporting on the order of 10 identifiable correction modes.

Items not in this revision. Multi-seed error bars (across optimizer seeds at fixed data) and an alternative-atlas sanity check (Schaefer-100 in place of the Destrieux-148 atlas) were not run. The std columns above are across the 30 subjects, not across seeds.

D Multi-lag temporal extension: training protocol

Section 4.3 reports results on the full 203-subject LEMON cohort. Training follows a two-stage schedule: stage 1 fits the spatial correction (U_s, V_s) from zero-lag supervision only; stage 2 freezes (U_s, V_s) and fits the temporal operator Φ_s against the multi-lag loss L_s^{lag} (Eq. 7).

Stage 1: spatial correction. Each subject receives a rank-r=10 correction $A_s = A_0 + U_s V_s^\top$, with $U_s \in \mathbb{R}^{C\times 10}$, $V_s \in \mathbb{R}^{N\times 10}$ ($C = 43$, $N \approx 8196$ BEM source vertices). The zero-lag term $\|B_s \hat{\Sigma}_y^{(0,0)} B_s^\top - \Sigma_s^{\text{ref}}\|_F^2$ is summed over all 203 subjects and minimized with Adam at learning rate 10^{-3} for 300 steps. After each step, U_s is replaced by the Q-factor of its QR decomposition, keeping it column-orthonormal (gauge fix from the main text). Every 20 steps the cohort-mean cross-session r_s is recorded; the snapshot with the highest value is taken as the stage-1 output and both U_s^* and V_s^* are frozen before stage 2 begins.

Stage 2: temporal operator. The multi-lag loss uses $d = 5$ lag steps. Three variants are run independently, scanning lag stride $\tau \in \{350, 700, 1050\}$ EEG samples (1400, 2800, 4200 ms, i.e. $1, 2, 3$ multiples of the LEMON fMRI TR). Each variant is optimized with Adam at learning rate 10^{-4} for 200 steps, with an ℓ_2 regularizer on the free parameters weighted by $\lambda_\Phi = 10^{-3}$. $\Phi_s = I_C + \Delta_s$, where $\Delta_s \in \mathbb{R}^{C\times C}$ is initialized to zero. The lagged lead field at step j is $A_{s,j\tau} = (I_C + \Delta_s)^j A_s^*$, computed by iterating $A \mapsto (I_C + \Delta_s)A$. The regularizer penalizes $\|\Delta_s\|_F^2$.

Block-Toeplitz covariance construction. Given lag-step count d and stride τ, the EEG-side block input is a $Cd \times Cd$ matrix whose (i, j)-th block ($C \times C$) is the empirical lag-$|i - j|\tau$ covariance of the eyes-closed EEG,

$$\hat{\Sigma}_y^{(i,j)} = \frac{1}{T} \sum_{t=|i-j|\tau+1}^{T} y_{t-|i-j|\tau}\, y_t^\top.$$

The covariance is normalized by $\text{tr}(\hat{\Sigma}_y^{(0,0)})$ so that the trace of the zero-lag block equals one across subjects. The fMRI-side target is constructed analogously from the z-scored parcel time series: the (i, j) block of Σ_s^{ref} is the lag-$|i - j|$-TR cross-covariance of the parcel amplitudes. The full $Pd \times Pd$ reference is then rank-truncated to its top Cd eigenvectors, making its effective rank match that of the Cd-dimensional predicted block matrix.

Scale normalization. The EEG-derived prediction $B_{s,i\tau}\, \hat{\Sigma}_y^{(i,j)}\, B_{s,j\tau}^\top$ carries an arbitrary amplitude scale, whereas the z-scored fMRI reference has unit parcel variances in its diagonal blocks. Before computing the Frobenius loss we apply a block-wise correlation normalization: for each diagonal block (i, i) of the predicted matrix, let $d_i^{(s)} = \sqrt{\text{diag}(B_{s,i\tau}\, \hat{\Sigma}_y^{(i,i)}\, B_{s,i\tau}^\top)} \in \mathbb{R}^P$; the (i, j) block is then divided entry-wise by $d_i^{(s)} (d_j^{(s)})^\top$. The same normalization is applied to the reference matrix. The stage-2 loss is the squared Frobenius norm between the two normalized block matrices, plus the regularization term.

Evaluation metrics. ρ_{diag} is the Pearson correlation between predicted and reference computed on the strict upper triangle of each diagonal block (i, i), pooled over all d diagonal blocks; it captures the zero-lag parcel-FC alignment that stage 1 already targets. ρ_{off} pools over all entries of the off-diagonal blocks ($i \neq j$), capturing the cross-lag parcel-FC alignment that Φ_s is designed to improve. Both quantities are reported at initialization (stage-2 step 0, where $\Phi_s = I$) and at the final step of stage 2.

E Representative modes on fsaverage

The main-text Fig. 3 shows the rank-1 decomposition $U_s V_s^\top = \sum_k u_{s,k} v_{s,k}^\top$ for a high-r_s subject (sub-032331, $r_s = 0.917$). Here we place a second high-r_s subject (Fig. 6) alongside the cohort minimum (Fig. 7) to compare the learned modes of a high-r_s and a low-r_s subject. Layout, mode ordering, sign convention and color limits are the same as in Fig. 3.

The two subjects differ most in the sensor (topomap) column. Sub-032315 ($r_s = 0.837$) produces smooth, low-spatial-frequency $u_{s,k}$ patterns in all three modes: dipolar or gently graded arrangements consistent with a volume-conducted source. Sub-032354 ($r_s = 0.199$), by contrast, carries mode 2 and mode 3 topomaps with extrema concentrated at one or two electrodes, with little surrounding support. A volume-conducted cortical source does not produce a scalp pattern of that shape; the pattern is consistent with U_s having fitted a channel-specific artefact or per-electrode noise. The mode-magnitude bars are also less differentiated in sub-032354 (modes 2–3 approach mode 1 in $\|v_{s,k}\|$), whereas sub-032315 follows the same concentrated hierarchy seen in the main-text exemplar. These observations concern two subjects. The cortex panels show that $v_{s,k}$ is not degenerate; we draw no conclusion about anatomical localisation from them.

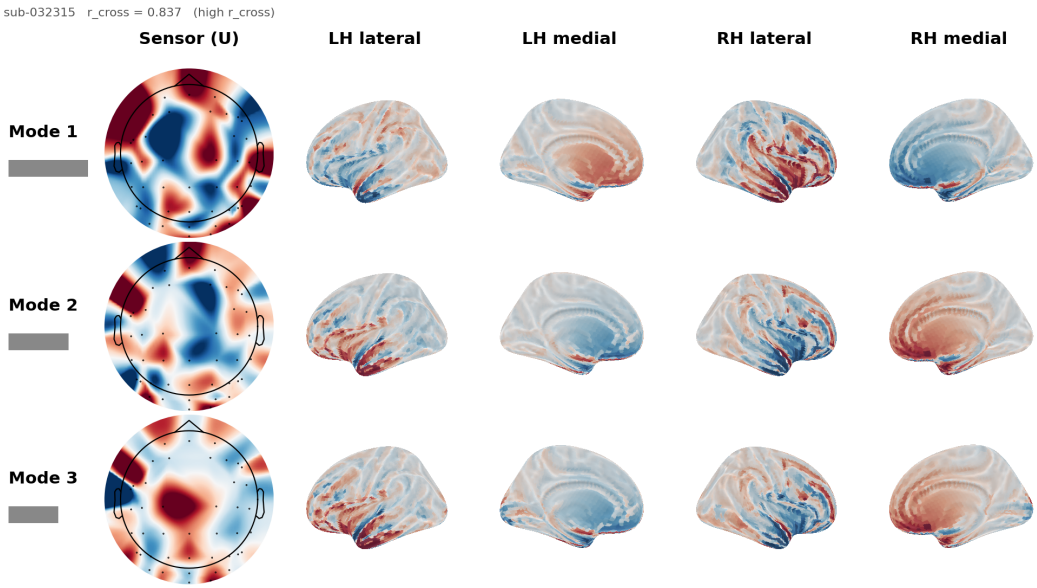


Figure 6: Top-3 correction modes for sub-032315 ($r_s = 0.837$, second-highest in the cohort). Layout identical to Fig. 3.

F Temporal operator: parameter budget

The lag extension in §3.1 fits a per-subject operator $\Phi_s \in \mathbb{R}^{C\times C}$. At the source level, the corresponding object is an operator $\Phi^N \in \mathbb{R}^{N\times N}$ with $\mathbb{E}[x_{t+\tau} \mid x_t] = \Phi^N x_t$ under a first-order Markov assumption on x_t. With $N = 8196$, Φ^N has $\sim 6.7\times10^7$ entries per subject, against $T_{\text{EEG}} \sim 7.5\times10^4$ samples per subject; it cannot be estimated from one subject's data, and a low-rank Φ^N would impose a structure that we have no per-subject data to support. Φ_s has $C^2 = 43^2 \approx 1.8 \times 10^3$ entries.

The two operators are related by

$$\Phi_s \;\approx\; A_s \, \Phi^N \, A_s^\dagger$$

(Moore–Penrose). Components of Φ^N in the null space of A_s do not affect sensor-level lagged covariances, so Φ_s retains the part of Φ^N that these covariances depend on.

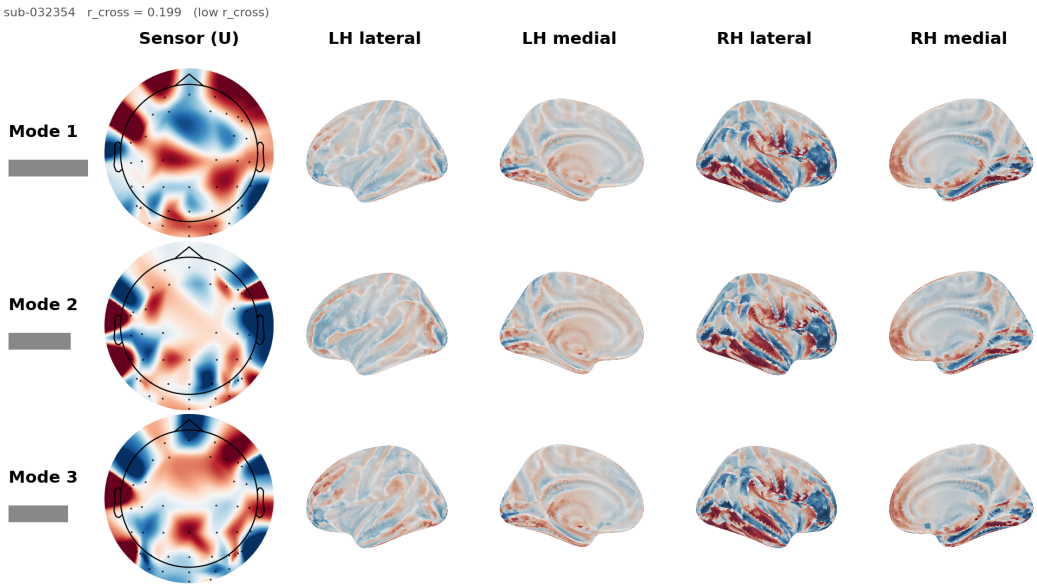


Figure 7: Top-3 correction modes for sub-032354 ($r_s = 0.199$, cohort minimum). Modes 2–3 show topomaps concentrated at one or two electrodes, in contrast to the smooth dipolar patterns in Fig. 3 and Fig. 6.

G Two-stage training protocol

Two stages. The fit is carried out in two stages. Stage 1 fits the spatial correction (U_s, V_s) against the zero-lag loss of §3 with $\Phi_s = I_C$. Stage 2 freezes (U_s, V_s) and fits the temporal perturbation Δ_s against the multi-lag loss L_s^{lag} of §3.1; since (U_s, V_s) is fixed, any decrease of L_s^{lag} in stage 2 is due to Δ_s.

Stage 1: gauge fixing. The parameterization $A_s = A_{0,s} + U_s V_s^\top$ admits a free $\mathrm{GL}(r)$ action $(U_s, V_s) \mapsto (U_s M^{-1}, V_s M^\top)$ that leaves A_s, B_s, and any downstream loss invariant (§3). We pick a representative within each orbit by retracting U_s onto the Stiefel manifold $\mathrm{St}(C, r) = \{U \in \mathbb{R}^{C\times r} : U^\top U = I_r\}$ after each gradient step: a thin QR factorization $U_s = QR$ followed by the assignment $U_s \leftarrow Q$ leaves $U_s V_s^\top$ unchanged (the factor R is absorbed into V_s on the next iteration) but renders (U_s, V_s) uniquely determined within its orbit. With orthonormal columns of U_s, $\|U_s V_s^\top\|_F = \|V_s\|_F$.

Stage 2: frozen (U_s, V_s), fit Δ_s. With (U_s, V_s) frozen at the Stage 1 representative, Δ_s is fitted against L_s^{lag} at $d = 5$ lag bins per subject and lag step $\tau \in \{350, 700, 1050\}$ EEG samples (equivalently $\{1, 2, 3\}$ multiples of the fMRI TR). The lag-bin weights w_j of §3.1 normalize the scale difference between the diagonal block ($j = 0$) and the off-diagonal blocks ($j \geq 1$) so that neither dominates the loss.

Hyperparameters. Full optimizer settings, learning rates, and step counts for both stages are specified in Appendix D.

H Per-electrode visualization of the trained Φ_s

This appendix extends the channel-pair view in §4.5 (Fig. 4) to a wider sample of 16 electrodes spread across the scalp, all from the same exemplar subject (sub-032331). We restate the construction of the main text to fix notation: each thin panel is a 3-row strip in which the top row is a locator (red dot marking the channel position on the 1020 montage), the middle row is the spatial-stage forward-operator row $(\Pi A_s)[\text{ch}, :]$ with $A_s = A_{0,s} + U_s V_s^\top$, and the bottom row is the Φ-only contribution

at $\Delta t = 100$ ms, $(\Pi({\Phi_s}^t - I)A_s)[\text{ch}, :]$ with $t = \Delta t/\tau$ and $\tau = 2800$ ms the training lag. Each sub-panel is auto-scaled to its own 98th percentile (red positive, blue negative); the absolute scales of the middle and bottom rows differ by roughly two orders of magnitude (middle $\sim 10^{-1}$, bottom $\sim 10^{-3}$), which is why they are not on a shared colour scale.

The middle row is the row of ΠA_s, where A_s includes the correction $U_s V_s^\top$ fitted in stage 1; it is not the row of $\Pi A_{0,s}$. Some channels still show a predominantly local lobe in A_s (e.g. AF3 / AFz / AF4 lighting up the front, F4 right-frontal); other channels, especially central and posterior ones, carry a multi-polar pattern in A_s already, because the spatial stage moved a non-trivial fraction of each channel's row away from a single-lobe shape during the fit against instantaneous covariance.

Both rows are spatially structured. We ask whether the Φ-only map at $\Delta t = 100$ ms, a lag not used in training, is spatially structured, and how it relates to the A_s row. Across the 16 channels we observe the following. First, every Φ-only map is patchy and signed rather than noise-like: it carries spatially coherent patches of consistent sign, typically a handful across the head, on the order of a parcel cluster in extent. Second, there is partial spatial overlap with the channel's A_s row but the patterns are not the same map rescaled. Some channels (e.g. C3, P3) have a Φ-only map whose patch geometry tracks the same regions that are active in A_s, just with a different sign distribution; others (e.g. AFz, FT7) have a Φ-only map that brings non-trivial mass to lateral or anterior-temporal parcels that the corresponding A_s row does not emphasise.

If Φ_s^t at $\Delta t = 100$ ms only rescaled the channel's row of A_s, the Φ-only map would be a multiple of the A_s row. The observed maps are not: for some channels they place weight on parcels that the A_s row does not emphasise. This is consistent with Φ_s encoding lagged couplings between the channel and parcels outside its lead-field support. Establishing which long-range couplings are recovered would require a source-resolved ground truth, which LEMON does not provide.

The 16 channels in Figure 8 sample the anterior half of the montage in the top row of panels and the central / posterior half in the bottom row.

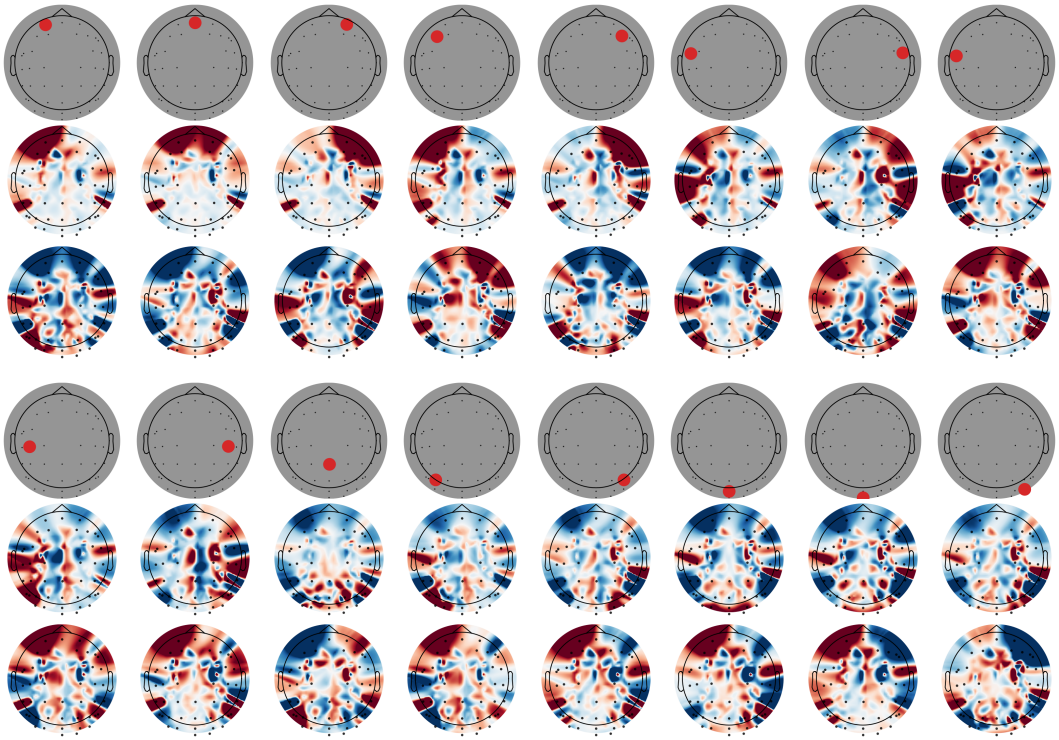


Figure 8: Per-electrode parcel maps for 16 channels of exemplar subject sub-032331. Each thin panel is a 3-row strip: top — channel locator (red dot on 1020 montage); middle — spatial-stage forward-operator row $(\Pi A_s)[\text{ch}, :]$ with $A_s = A_{0,s} + U_s V_s^\top$; bottom — Φ-only contribution $(\Pi(\Phi_s{}^t - I)A_s)[\text{ch}, :]$ at $\Delta t = 100$ ms, with $t = \Delta t/\tau$ and $\tau = 2800$ ms the training lag. Colour map RdBu_r (red positive, blue negative); each sub-panel auto-scaled to its own 98th percentile. Top row of panels: 8 channels in the anterior half of the montage; bottom row: 8 channels in the central / posterior half.