`invertmeeg`: A Benchmark and Unified Python Library for EEG Inverse Solvers

Lukas Hecker
lukas.hecker.job@gmail.com

Abstract

Electroencephalography (EEG) source imaging remains difficult to compare systematically because inverse solvers are distributed across different software packages, programming languages, and evaluation protocols. We present a frozen four-scenario EEG benchmark of 106 solvers evaluated on a shared BioSemi-32 / `ico3` setup, together with `invertmeeg`, an open-source Python package that currently exposes 118 inverse solvers through a consistent two-step interface built on the `MNE-Python` ecosystem. The benchmark spans focal, multi-source, spatially extended, and low-SNR source configurations and uses earth mover's distance (EMD) as the primary metric, with average precision (AP), mean localization error (MLE), and correlation used for complementary ranking. Across this benchmark, no single solver dominates every regime: flexible subspace and hybrid methods perform best overall, while Bayesian methods remain particularly competitive under extended-source and low-SNR conditions. The package is available via `pip install invertmeeg` and imported in Python as `invert`.

1 Introduction

The reconstruction of neural source activity from EEG sensor measurements is a fundamental problem in functional neuroimaging [1, 2]. Given a forward model that describes how source currents produce sensor-level signals, the inverse problem seeks to recover the underlying source distribution. Because the number of potential source locations (typically 10^3–10^4) far exceeds the number of sensors (typically 10^1–10^2), the problem is severely underdetermined and requires regularization or prior assumptions to obtain a unique solution [3].

Over the past four decades, a rich landscape of inverse methods has emerged. Minimum norm approaches [4] minimize the L_2 norm of the source distribution; the LORETA family [5, 6] incorporates spatial smoothness constraints; beamformers [7] construct spatial filters that maximize signal power at a target location while suppressing interference; Bayesian methods [8, 9] cast the problem in a probabilistic framework with structured priors; sparse recovery algorithms [10] exploit the assumption that only a small number of sources are simultaneously active; and subspace scanning methods such as MUSIC [11] exploit the orthogonality between signal and noise subspaces. More recently, deep learning approaches have been proposed that learn the inverse mapping from data [12, 13].

Despite this wealth of methods, practitioners face a fragmentation problem. Implementations are distributed across different toolboxes—`MNE-Python` [14], Brainstorm [15], FieldTrip [16], SPM [17]—each supporting only a subset of methods, often with incompatible data formats and interfaces. Many published algorithms lack publicly available implementations entirely, or exist only as standalone scripts tied to specific datasets. This fragmentation impedes systematic comparison of methods under controlled conditions and makes it difficult for researchers to evaluate which method is most appropriate for their specific experimental paradigm.

The central contribution of this paper is a frozen EEG benchmark of 106 solvers across 4 evaluation scenarios that span focal dipoles, multiple focal sources, spatially extended patches, and low-SNR conditions. The benchmark uses a shared BioSemi-32 forward model, `ico3` source space, simulation pipeline, and evaluation protocol, allowing direct comparison across solver families. To make this benchmark reproducible and extensible, we package the underlying infrastructure in `invertmeeg`, an open-source Python library that currently exposes 118 inverse solvers under a single, consistent application programming interface (API). All solvers inherit from a common `BaseSolver` class and follow a two-step workflow: (1) compute an inverse operator from a forward model, and (2) apply the operator to EEG data. The package builds on `MNE-Python` for data structures, forward modeling, and visualization, ensuring compatibility with existing analysis pipelines. Source code, documentation, and tutorials are available at `https://github.com/LukeTheHecker/invertmeeg` and `https://lukethehecker.github.io/invertmeeg/`. To provide an intuition for what these benchmark differences look like in practice, Figure 1 contrasts twelve solver families on the same extended-source EEG sample before the paper moves to aggregated rankings and scenario-level summaries.

Beyond the benchmark, `invertmeeg` provides:

- A **simulation framework** with diffusion-basis and contiguous Gaussian patch sources plus realistic structured sensor noise for generating diverse synthetic EEG data with known ground truth.
- **Automatic regularization** via generalized cross-validation (GCV), L-curve, and product methods, with a shared implementation across all linear solvers.
- A **standardized evaluation suite** with metrics including earth mover's distance (EMD), average precision (AP), mean localization error (MLE), correlation, and spatial dispersion.

The remainder of this paper is organized as follows. Section 2 describes the software architecture and the solver interface. Section 3 provides an overview of the implemented solver categories. Section 4 details the shared regularization framework. Section 5 presents the simulation methodology. Section 6 describes the evaluation metrics, benchmark scenarios, and results. Section 7 provides usage examples. Section 8 discusses solver selection guidelines, related work, and limitations. Section 9 concludes the paper.

2 Software Architecture

A meaningful benchmark requires that all solvers are compared under identical conditions with a shared forward model, simulation pipeline, and evaluation protocol. The design of `invertmeeg` is guided by two principles: (1) every solver, regardless of its algorithmic family, should expose the same interface, and (2) adding a new solver should require minimal boilerplate. This section describes the main architectural components. Figure 2 provides an overview of the architecture and the two-step workflow.

2.1 The Solver Factory

The primary user-facing entry point is the `Solver()` factory function:

```
from invert import Solver

solver = Solver("eloreta")     # eLORETA
```

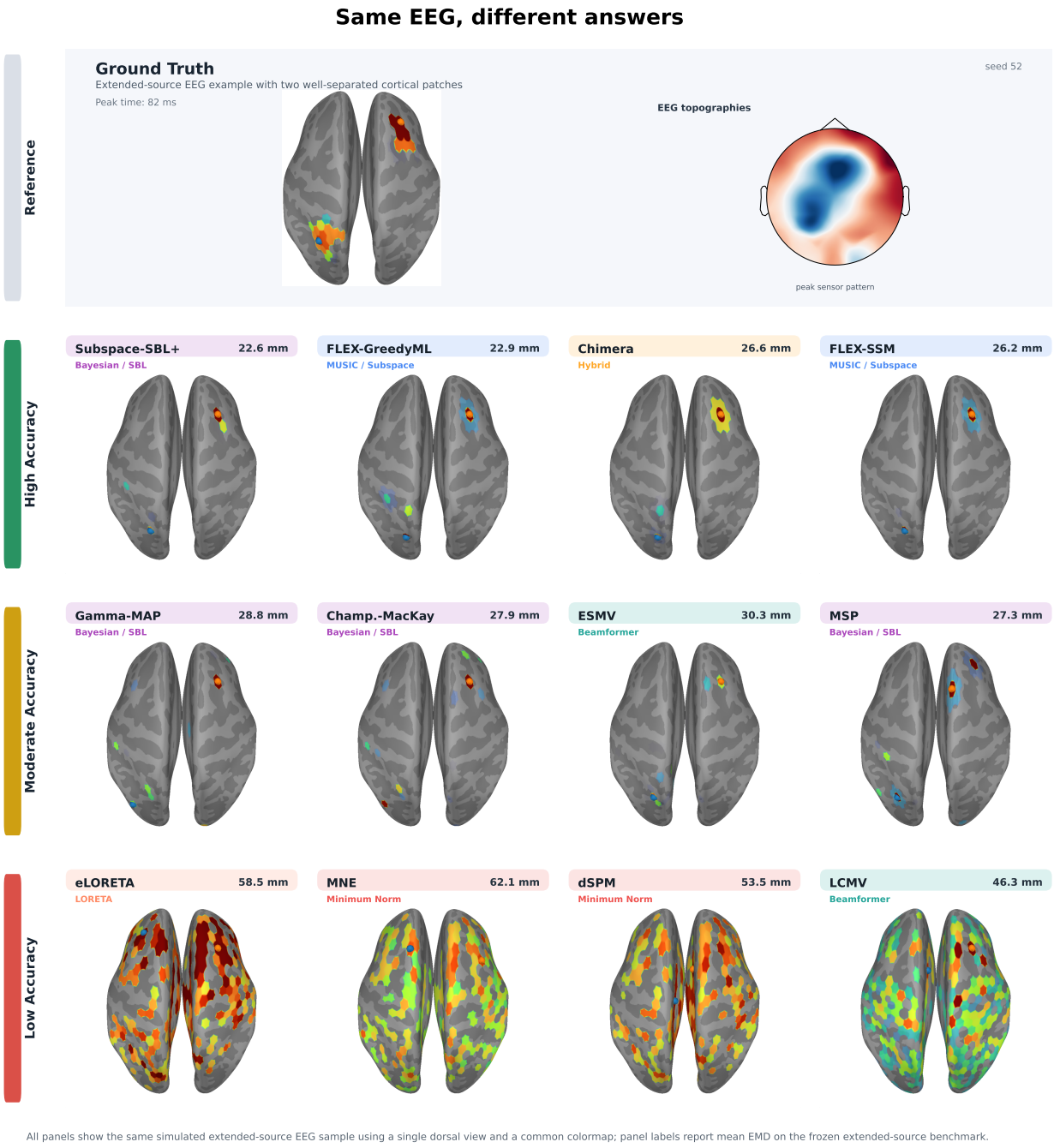

Figure 1: Qualitative solver spectrum for one representative extended-source EEG sample from the frozen benchmark. The ground-truth row shows two well-separated contiguous Gaussian patches together with the sensor topography at the selected peak time. The remaining panels apply twelve benchmarked solvers to the same simulated EEG data and annotate each reconstruction with its solver family and mean EMD on the frozen extended-source scenario. Row labels summarize relative performance on this specific scenario, not a universal ordering across all EEG source priors. In particular, the example favors methods that preserve compact patch structure; broader priors can be more appropriate when the underlying activation is genuinely diffuse.

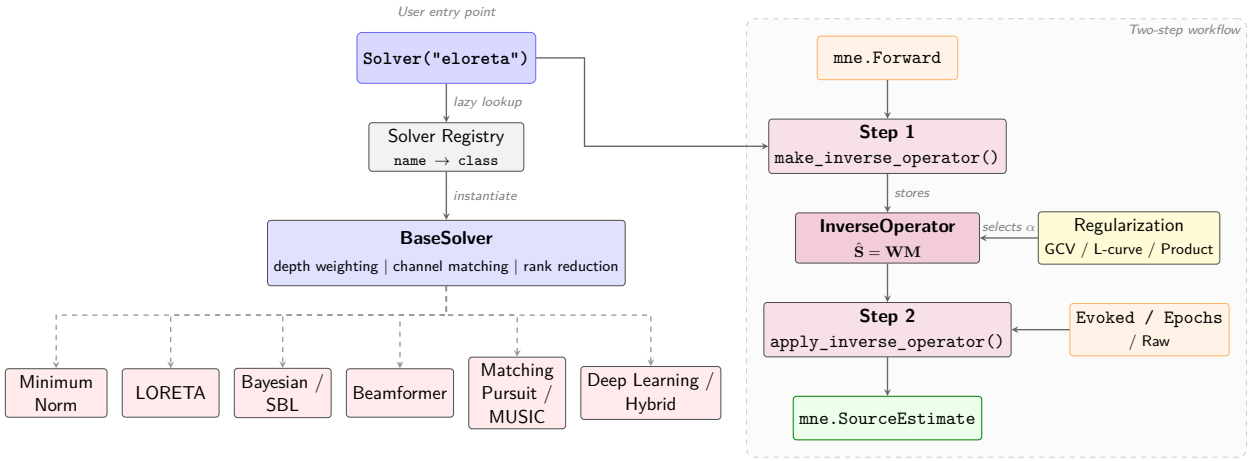

Figure 2: Architecture of `invertmeeg`. The `Solver()` factory lazily resolves a name to a concrete solver class inheriting from `BaseSolver`. The two-step workflow separates inverse operator computation (Step 1, depending only on the forward model) from application to data (Step 2). Regularization parameter selection happens at apply time.

```
solver = Solver("champagne")   # Champagne (default update rule)
solver = Solver("em-champagne") # Champagne with EM update rule
```

The package is installed via `pip install invertmeeg` but imported as `invert`; the code snippets below follow that Python namespace.

The factory maintains a registry that maps solver names and aliases to constructor functions. Many solvers are registered under multiple aliases for convenience (e.g., `"l1"`, `"fista"`, and `"mce"` all resolve to the same minimum current estimate solver). The registry is populated lazily: solver modules are imported only when the factory is first invoked, which keeps the initial import of `invertmeeg` lightweight. For solvers that represent parametric variants of the same algorithm (e.g., the seven Champagne update rules), the registry stores `functools.partial` objects that pre-fill the relevant keyword arguments.

Calling `list_solvers()` returns a sorted list of all registered aliases.

2.2 BaseSolver and the Two-Step Interface

Every solver class inherits from `BaseSolver`, which defines the contract:

1. `make_inverse_operator(forward, alpha="auto", ...)` accepts an `MNE-Python Forward` object, computes the inverse operator, and stores it internally.

2. `apply_inverse_operator(mne_obj)` accepts an `MNE-Python` data object (`Evoked`, `Epochs`, or `Raw`), applies the precomputed inverse operator, and returns an `mne.SourceEstimate`.

This separation has a practical motivation: the inverse operator depends only on the forward model and regularization parameters, not on the data. Once computed, it can be applied to multiple recordings without recomputation, which is particularly relevant for beamformers and minimum norm methods where the operator is a matrix multiplication.

The base class handles several concerns shared across all solvers:

- **Forward model preparation**: conversion of free-orientation forward models to fixed orientation, optional common average referencing of the leadfield, and depth weighting of leadfield columns.
- **Channel matching**: automatic intersection of channels present in the forward model and the data object.
- **Regularization parameter grid**: construction of a logarithmically spaced grid of regularization parameters, scaled by the largest eigenvalue of the leadfield Gram matrix $\mathbf{L}\mathbf{L}^\top$ (see Section 4).
- **Rank reduction**: optional signal subspace selection via SVD-based rank estimation.
- **Output conversion**: mapping of the source matrix back to an `mne.SourceEstimate` with correct vertex indices, timing, and subject metadata.

2.3 InverseOperator

For solvers that produce a linear inverse operator (a matrix $\mathbf{W} \in \mathbb{R}^{n_{\text{sources}} \times n_{\text{channels}}}$), the precomputed operator is wrapped in an `InverseOperator` object. Applying the operator to a data matrix $\mathbf{M} \in \mathbb{R}^{n_{\text{channels}} \times n_{\text{times}}}$ is then a single matrix multiplication:

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{M}, \tag{1}$$

where $\hat{\mathbf{S}} \in \mathbb{R}^{n_{\text{sources}} \times n_{\text{times}}}$ is the estimated source activity. When automatic regularization is enabled (`alpha="auto"`), the solver stores one `InverseOperator` per candidate regularization parameter; the optimal one is selected at apply time based on the chosen regularization criterion (Section 4).

2.4 Solver Metadata

Each solver class carries a `SolverMeta` dataclass with fields for the full solver name, acronym, category, a brief description, and literature references. This metadata serves two purposes: it enables programmatic generation of solver catalogs and documentation, and it allows users to inspect a solver's provenance at runtime.

2.5 Depth Weighting

Sources deeper in the brain produce weaker signals at the sensor level, biasing unweighted solutions toward superficial cortex. `invertmeeg` applies column-wise depth weighting to the leadfield matrix $\mathbf{L}$ before computing the inverse operator:

$$\tilde{\mathbf{L}}_{\cdot,j} = \frac{\mathbf{L}_{\cdot,j}}{\|\mathbf{L}_{\cdot,j}\|^d}, \tag{2}$$

where $d \in [0, 1]$ controls the weighting strength (default $d = 0.5$). After source estimation, the weighting is removed to recover physical amplitude units.

Table 1: Benchmark depth-weighting policy. Category defaults apply unless a solver-specific override is listed. Solvers not shown use their package defaults without an explicit benchmark override.

Policy	**Depth weight** d
Minimum Norm default	0.5
LORETA default	0.0
Bayesian / SBL default	0.0
MUSIC / Subspace default	0.5
Backus-Gilbert override	0.0
GBF override	1.0
GFT-L1 override	0.0
MNE override	1.0
S-MAP override	0.0
TV override	0.0

3 Solver Catalog

`invertmeeg` currently registers 118 solver implementations. The frozen EEG benchmark reported here evaluates 106 of them under a common four-scenario setup. Rather than presenting the full mathematical derivation of each solver, we summarize the main families and refer the reader to the original publications. A complete solver reference with canonical identifiers and aliases is provided in the Appendix and as a machine-readable supplementary CSV.

3.1 Minimum Norm Methods

Minimum norm methods seek the source distribution $\hat{\mathbf{s}}$ that minimizes a weighted norm of the source vector subject to the data fit:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{L}\mathbf{s} - \mathbf{m}\|_2^2 + \alpha \|\mathbf{R}\mathbf{s}\|_p^p, \tag{3}$$

where $\mathbf{L}$ is the leadfield matrix, $\mathbf{m}$ is the measurement vector, α is the regularization parameter, $\mathbf{R}$ is a weighting matrix, and $p \in \{1, 2\}$ selects the norm.

For $p = 2$ and $\mathbf{R} = \mathbf{I}$, Equation (3) yields the classic minimum norm estimate (MNE) [4] with the closed-form solution:

$$\hat{\mathbf{s}} = \mathbf{L}^\top (\mathbf{L}\mathbf{L}^\top + \alpha \mathbf{I})^{-1} \mathbf{m}. \tag{4}$$

Variants include weighted MNE (wMNE), which uses noise-weighted columns, and dynamic statistical parametric mapping (dSPM) [18], which normalizes source power by noise variance.

For $p = 1$, Equation (3) promotes sparse solutions; `invertmeeg` implements this via the FISTA algorithm [19] as well as combined L_1/L_2 penalties.

3.2 LORETA Family

Low Resolution Electromagnetic Tomography (LORETA) [5] adds a spatial smoothness constraint by incorporating the cortical graph Laplacian $\mathbf{\Delta}$ into the regularization term. The family includes LORETA, standardized LORETA (sLORETA) [6], and exact LORETA (eLORETA) [20], which differ in their normalization strategies and have distinct localization properties.

Table 2: Solver categories in `invertmeeg`. Package totals are derived from the live registry; benchmark counts come from the frozen four-scenario EEG snapshot.

Category	**Package**	**Bench.**	**Representative methods**
Minimum Norm	16	16	MNE, wMNE, dSPM, MCE, MxNE-BCD, LAURA, TV
LORETA	5	5	LORETA, sLORETA, eLORETA, SSLOFO
Bayesian / SBL	31	29	Champagne variants, Gamma-MAP, MSP, Subspace-SBL
Beamformers	25	24	LCMV, DICS, SAM, ESMV family, MCMV, EBB
MUSIC / Subspace	14	13	MUSIC, RAP-MUSIC, FLEX-MUSIC, SSM, AP, GreedyML
Matching Pursuit	10	10	OMP, CoSaMP, ReMBo, FLEX-OMP, SubSMP
Dipole Fitting	2	2	ECD, SESAME
Deep Learning	6	3	FC, CovCNN, CovCNN-KL, LSTM, CNN, DeepSIF
Hybrid / Baseline	9	4	Chimera, Hydra, APSE, graph-Wiener variants, Random Noise
Total	**118**	**106**	

3.3 Bayesian Methods

Bayesian approaches model the source distribution as a random variable with a structured prior and optimize the model evidence (type-II maximum likelihood). The Champagne algorithm [8] and its variants form the largest subgroup, with seven update rules (MacKay, Convexity/MM, EM, AR-EM, TEM, LowSNR, Adaptive) and optional noise learning. Other Bayesian solvers include Gamma-MAP [21], Multiple Sparse Priors (MSP) [9], coherent maximum entropy on the mean (cMEM) [22], and Sparse Bayesian Learning (SBL) variants.

3.4 Beamformers

Beamformers construct spatial filters that pass signals from a target location while minimizing contributions from elsewhere. The linearly constrained minimum variance (LCMV) beamformer [7] solves:

$$\mathbf{w}_j = \frac{\mathbf{C}^{-1}\mathbf{l}_j}{\mathbf{l}_j^\top \mathbf{C}^{-1}\mathbf{l}_j}, \tag{5}$$

where $\mathbf{C}$ is the data covariance matrix and $\mathbf{l}_j$ is the leadfield column for source j. `invertmeeg` implements 25 beamformer variants, including DICS for frequency-domain analysis [23], synthetic aperture magnetometry (SAM) [24], eigenspace-projected variants (ESMV family), multiple-constraint approaches (MCMV), and the empirical Bayesian beamformer (EBB).

3.5 Matching Pursuit and Sparse Recovery

These methods greedily or iteratively select active sources from the leadfield dictionary. Orthogonal matching pursuit (OMP) [25] iteratively selects the leadfield column most correlated with the residual; CoSaMP [26] and subspace pursuit [27] extend this with two-stage selection and pruning. `invertmeeg` includes 10 matching pursuit variants, including FLEX variants that combine greedy selection with flexible subspace scanning, ReMBo (random embedding), and hierarchical subspace methods (SubSMP, ISubSMP).

3.6 MUSIC and Subspace Methods

Multiple signal classification (MUSIC) [11] localizes sources by scanning for leadfield columns that lie in the signal subspace. RAP-MUSIC [28] extends this with recursive projection to resolve multiple sources. `invertmeeg` provides 13 subspace methods: MUSIC, RAP-MUSIC, TRAP-MUSIC, and their FLEX variants, as well as signal subspace matching (SSM), alternating projections (AP), ExSo-MUSIC, and maximum-likelihood scanners (ExhaustiveML, GreedyML) that search for the source configuration maximizing the data likelihood.

3.7 Dipole Fitting

Equivalent current dipole (ECD) fitting models the source as one or more point dipoles whose positions and moments are optimized to fit the data. SESAME [29] uses a sequential Monte Carlo approach for multi-dipole estimation.

3.8 Deep Learning

`invertmeeg` includes six neural network architectures that learn the inverse mapping from simulated training data. These solvers require PyTorch as an optional dependency and use the simulation framework (Section 5) for on-the-fly training data generation—no pre-recorded data is needed. Three of these solvers (FC, CovCNN, CovCNN-KL) are included in the benchmark.

FC is a simple fully connected EEG inverse-model baseline in the ConvDip / ESINet family of early deep-learning approaches [12, 13]. Each time point is processed independently through a dense layer ($n_{\text{channels}} \to n_{\text{units}}$, tanh activation), a second dense layer ($n_{\text{units}} \to n_{\text{units}}$), and a linear output layer ($n_{\text{units}} \to n_{\text{dipoles}}$). With 600 hidden units and 1,284 dipoles, this yields $\sim 1.2 \times 10^6$ trainable parameters.

CovCNN [13] operates on the sensor covariance matrix $\mathbf{C} = \mathbf{M}\mathbf{M}^\top \in \mathbb{R}^{n_{\text{ch}} \times n_{\text{ch}}}$ rather than the raw time series, providing implicit temporal averaging and sign invariance. The covariance is passed through a frozen convolutional layer whose weights are derived from the leadfield matrix, projecting the $n_{\text{ch}} \times n_{\text{ch}}$ input to n_{dipoles} feature maps. The resulting features are flattened and processed by a multi-layer perceptron (MLP) with tanh activations, producing per-dipole source estimates. CovCNN is trained with cosine similarity loss.

CovCNN-KL shares the CovCNN architecture but is trained with Kullback-Leibler divergence loss after applying a softmax to the output logits, treating source localization as a distribution-matching problem. This reformulation improves localization accuracy compared to the regression-based CovCNN. In the benchmark configuration (24 hidden units), both CovCNN variants have $\sim 1.0 \times 10^6$ trainable parameters; FC has $\sim 1.2 \times 10^6$. All three are trained with the Adam optimizer for 200 epochs with early stopping (patience 60) and a batch size of 2,048 simulated samples.

The remaining (non-benchmarked) architectures include CovCNN-KL-FlexOMP, which augments CovCNN-KL with iterative sparse post-processing; a bidirectional LSTM for temporal modeling; and a CNN with LSTM encoder for spatiotemporal processing.

3.9 Flexible Extent Estimation (FLEX)

A common limitation of classical inverse methods is the assumption that each source is a point dipole. In practice, cortical generators often extend over patches of varying size. The FLEX (Flexible Extent Estimation) framework [30] addresses this by augmenting any base solver with a multi-order leadfield dictionary that allows simultaneous estimation of source location and spatial extent.

The core idea is to construct smoothed versions of the leadfield matrix $\mathbf{L}$ via iterative diffusion on the cortical adjacency graph. Given the graph Laplacian $\mathbf{\Delta}$ of the source space, a smoothing operator is defined as:

$$\mathbf{S} = \mathbf{I} - \alpha_{\text{diff}} \cdot \mathbf{\Delta}, \tag{6}$$

where α_{diff} controls the diffusion rate. Applying $\mathbf{S}$ to the leadfield produces leadfields for increasingly extended sources:

$$\mathbf{L}^{(k)} = \mathbf{L} \cdot \mathbf{S}^k, \quad k = 0, 1, 2, \ldots, K, \tag{7}$$

where $k = 0$ corresponds to point dipoles and increasing k produces progressively smoother spatial profiles. For each order k, a gradient matrix $\mathbf{B}^{(k)}$ tracks how smoothed activations map back to individual dipoles in the original source space.

The extended leadfield dictionary $\mathbf{L}_{\text{ext}} = [\mathbf{L}^{(0)} \mid \mathbf{L}^{(1)} \mid \cdots \mid \mathbf{L}^{(K)}]$ contains $K \cdot n$ candidate columns (one per dipole per order). The base solver then operates on this extended dictionary and selects the best (location, order) pair for each source using its native selection criterion—subspace projection for MUSIC variants, residual correlation for matching pursuit, likelihood maximization for ML scanners, evidence maximization for Bayesian methods, or beamformer output power for spatial filters. A mutual exclusivity constraint ensures that at most one order is active per dipole location. The selected sources are mapped back to the original source space via their corresponding gradient matrices $\mathbf{B}^{(k)}$.

This mechanism is general: `invertmeeg` applies it to subspace methods (FLEX-MUSIC, FLEX-SSM, FLEX-GreedyML), matching pursuit algorithms (FLEX-OMP, FLEX-CoSaMP, FLEX-SP, FLEX-ReMBo), beamformers (the FLEX-ESMV family), and Bayesian solvers (Flex-Champagne, Flex-NL-Champagne). In all cases, the data-driven order selection allows each source to adopt the spatial extent that best explains the measurements, without requiring the user to specify source size a priori. Across the frozen EEG benchmark, several FLEX variants rank near the top overall, particularly in the focal and multi-source regimes.

3.10 Hybrid Methods

`invertmeeg` includes three hybrid solvers that combine multiple algorithmic strategies to adapt to the characteristics of the data at hand.

Chimera adaptively selects between a beamformer (FLEX-ESMV) and a subspace scanner (SSM) based on properties estimated from the data. It first estimates the number of active sources $\hat{k}$ using the minimum description length (MDL) criterion [31] applied to the eigenvalues of the data covariance. If exactly one source is detected ($\hat{k} = 1$), Chimera computes an order-score ratio that compares the subspace fit of a point-dipole leadfield against smoothed (extended) leadfields: if the ratio indicates a focal source, the beamformer branch is used; otherwise, SSM handles multi-source

or extended configurations. This two-stage decision rule allows Chimera to route each data sample to the solver best suited to its source geometry.

Hydra combines a flexible maximum-likelihood scanner (FLEX-GreedyML) with an adaptive beamformer (Adapt-FLEX-ESMV), switching between them based on a variance-based heuristic. The mean sensor variance of the input data is compared against a calibrated threshold: high-variance data (characteristic of distributed or ongoing activity) is routed to the beamformer, while lower-variance data (characteristic of sparse focal sources) is processed by FLEX-GreedyML. In the frozen EEG benchmark snapshot, Hydra ties for the best global rank, illustrating how a simple switching heuristic can leverage the complementary strengths of sparse maximum-likelihood scanning and adaptive beamforming.

APSE (Alternating Projections with Source Enumeration) integrates subspace scanning with automatic model-order selection via information-theoretic criteria.

3.11 Novel Solvers Developed with AI Coding Agents

Beyond re-implementing published algorithms, `invertmeeg` includes more than 30 currently unpublished solver variants. These methods were developed by the first author in collaboration with AI coding agents (large language models used as interactive programming assistants) and represent novel combinations and extensions of existing algorithmic components. Families of novel solvers include FLEX-ESMV beamformer variants, Flex-Champagne Bayesian solvers, hybrid switching methods (Chimera, Hydra, APSE), subspace matching pursuit variants, and all deep learning solvers (CovCNN, CovCNN-KL, CNN, LSTM). Each novel solver is documented with its component references in the solver source code and in the appendix tables.

As an illustrative example, we describe **Subspace-SBL**—one of the strongest Bayesian solvers and a top-five method in the frozen benchmark—in detail. Subspace-SBL decomposes the inverse problem into two stages: a combinatorial source detection stage followed by a continuous amplitude estimation stage on the detected support.

Stage 1: Subspace detection. Given sensor data $\mathbf{Y} \in \mathbb{R}^{m\times T}$, we first form a regularized signal-subspace projector via signal subspace matching (SSM):

$$\mathbf{P}_Y = \mathbf{Y}\left(\mathbf{Y}^\top\mathbf{Y} + \lambda_1 \operatorname{tr}(\mathbf{Y}^\top\mathbf{Y})\,\mathbf{I}_T\right)^{-1}\mathbf{Y}^\top, \qquad \mathbf{C} = \mathbf{P}_Y^\top\mathbf{P}_Y, \tag{8}$$

where λ_1 is a small regularization constant. The matrix $\mathbf{C}$ captures the signal subspace robustly even under model errors. Sources are then detected greedily from the multi-order FLEX dictionary $\{\mathbf{L}^{(0)}, \ldots, \mathbf{L}^{(K)}\}$. At each step q, we select the (location, order) pair that maximizes the projected subspace score:

$$(d^*, k^*) = \arg\max_{d,k} \frac{\mathbf{a}_{d,k}^\top\,\mathbf{C}\,\mathbf{a}_{d,k}}{\mathbf{a}_{d,k}^\top\mathbf{a}_{d,k}}, \qquad \text{where } \mathbf{a}_{d,k} = (\mathbf{I} - \mathbf{P}_A)\,\ell_d^{(k)}, \tag{9}$$

and $\mathbf{P}_A$ is the (regularized) projector onto the subspace spanned by previously selected leadfield columns. After the greedy forward pass, a cyclic refinement procedure revisits each selected source and re-optimizes its (location, order) assignment while holding the others fixed, iterating until convergence.

Stage 2: Noise-learning SBL refinement. Let $\mathbf{L}_{\text{sel}} \in \mathbb{R}^{m\times\hat{k}}$ denote the leadfield columns corresponding to the $\hat{k}$ detected sources. We model the data as $\mathbf{Y} = \mathbf{L}_{\text{sel}}\mathbf{S} + \mathbf{N}$, with a Gaussian

prior $\mathbf{s}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\gamma))$ and heteroscedastic sensor noise $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\lambda))$. The model covariance is $\mathbf{\Sigma}_y = \mathbf{L}_{\mathrm{sel}}\,\mathrm{diag}(\gamma)\mathbf{L}_{\mathrm{sel}}^\top + \mathrm{diag}(\lambda)$. Source variance and noise parameters are estimated by iterating Type-II ML fixed-point updates:

$$\gamma_i \leftarrow \sqrt{\frac{\sum_t \bar{s}_{i,t}^2}{[\mathbf{L}_{\mathrm{sel}}^\top \mathbf{\Sigma}_y^{-1} \mathbf{L}_{\mathrm{sel}}]_{ii}}}, \qquad \lambda_j \leftarrow \sqrt{\frac{\sum_t (y_{j,t} - \hat{y}_{j,t})^2}{[\mathbf{\Sigma}_y^{-1}]_{jj}}}, \tag{10}$$

where $\bar{\mathbf{S}} = \mathrm{diag}(\gamma)\mathbf{L}_{\mathrm{sel}}^\top \mathbf{\Sigma}_y^{-1}\mathbf{Y}$ is the posterior mean and $\hat{\mathbf{Y}} = \mathbf{L}_{\mathrm{sel}}\bar{\mathbf{S}}$ the predicted data. Sources with $\gamma_i < 10^{-3}\max(\gamma)$ are pruned, and the algorithm converges when the relative change in the Type-II ML objective falls below 10^{-8}. The final inverse operator is $\mathbf{K} = \mathbf{B}^\top \mathrm{diag}(\gamma)\mathbf{L}_{\mathrm{sel}}^\top \mathbf{\Sigma}_y^{-1}$, where $\mathbf{B}$ contains the FLEX gradient vectors that map detected atoms back to the full source space.

Subspace-SBL+ extends this by expanding the SSM-detected candidate set with graph-neighborhood atoms (one hop in the cortical adjacency graph) before running the SBL refinement. This local expansion addresses SSM's sensitivity to vertex-level misalignment: the Bayesian refinement automatically prunes spurious candidates while retaining those that genuinely explain data variance.

The design philosophy behind these variants follows a common pattern: combine fast geometric methods (subspace scanning, beamforming) for discrete decisions with statistically regularized refinement (SBL or maximum likelihood) for continuous estimation. A dedicated paper on the agentic development process is in preparation.

4 Regularization

Many inverse solvers require a regularization parameter α that balances data fit against solution complexity. Choosing α appropriately is critical: too little regularization amplifies noise, while too much regularization oversmooths the solution. `invertmeeg` provides a shared regularization framework that is used by all solvers that produce linear inverse operators.

4.1 Regularization Parameter Grid

When `alpha="auto"` (the default), the solver constructs a grid of candidate values. A set of dimensionless knobs r_i is logarithmically spaced in $[10^{-10}, 10^{1}]$, and each is scaled by the largest eigenvalue $\lambda_{\max}$ of a reference matrix (by default $\mathbf{L}\mathbf{L}^\top$):

$$\alpha_i = r_i \cdot \lambda_{\max}(\mathbf{L}\mathbf{L}^\top). \tag{11}$$

This scaling ensures that the regularization grid is adapted to the magnitude of the forward model, making the same r-values meaningful across different sensor configurations and source spaces. For each α_i, a separate inverse operator $\mathbf{W}_i$ is precomputed and stored.

4.2 Generalized Cross-Validation (GCV)

GCV [32] selects the α that minimizes a leave-one-out prediction error estimate without explicit knowledge of the noise level. For a linear inverse operator that produces the hat matrix $\mathbf{H}_\alpha = \mathbf{L}\mathbf{W}_\alpha$, the GCV criterion is:

$$\mathrm{GCV}(\alpha) = \frac{\|\mathbf{m} - \mathbf{H}_\alpha \mathbf{m}\|^2}{(n - \gamma \cdot \mathrm{tr}(\mathbf{H}_\alpha))^2}, \tag{12}$$

where n is the number of channels and $\gamma \geq 1$ is a correction factor. Setting $\gamma = 1$ gives standard GCV; values slightly above 1 (modified GCV, MGCV) bias the selection toward stronger regularization, which can be helpful when the forward model is rank-deficient.

For iterative Bayesian solvers such as MacKay-Champagne and Convexity-Champagne, standard GCV selected unstable regularization values in our benchmark setup and led to degraded localization quality. To address this, we use the strong robust GCV (R1GCV) variant [33]:

$$\mathrm{R1GCV}(\alpha) = \big[\gamma + (1-\gamma)\cdot \mu_{12}(\alpha)\big] \cdot \mathrm{GCV}(\alpha), \quad \mu_{12}(\alpha) = \frac{\mathrm{tr}(\mathbf{H}_\alpha) - \mathrm{tr}(\mathbf{H}_\alpha^2)}{n \cdot \alpha}, \tag{13}$$

where μ_{12} is the negative derivative of the mean leverage $\mathrm{tr}(\mathbf{H}_\alpha)/n$ with respect to α. The weighting factor penalizes regions of the regularization grid where solution variance amplifies rapidly as α decreases, biasing the selection toward more stable operating points. We therefore use R1GCV as an empirical stability choice for the Champagne-family benchmark runs and do not claim a more general mechanistic explanation here.

4.3 L-Curve Method

The L-curve method [34] plots the solution norm $\|\hat{\mathbf{s}}\|_2$ against the residual norm $\|\mathbf{L}\hat{\mathbf{s}} - \mathbf{m}\|_2$ for each candidate α. The optimal α is located at the "corner" of the resulting L-shaped curve, which represents the best trade-off between data fit and solution regularity. `invertmeeg` detects the corner using a triangle-area heuristic: the candidate that maximizes the area of the triangle formed with the two endpoints of the curve is selected.

4.4 Product Method

The product method [3] selects the α that minimizes the product of the solution norm and the residual norm:

$$\alpha^* = \arg\min_\alpha \|\hat{\mathbf{s}}_\alpha\|_2 \cdot \|\mathbf{L}\hat{\mathbf{s}}_\alpha - \mathbf{m}\|_2. \tag{14}$$

This criterion does not require computing the hat matrix and is therefore computationally cheaper than GCV.

5 Simulation Methodology

A meaningful comparison of inverse solvers requires synthetic data with known ground truth that spans the diversity of source configurations encountered in practice. Existing simulation approaches often rely on simple Gaussian patches centered on fixed cortical locations, which produces unrealistically regular spatial patterns and may favor solvers whose assumptions match these idealized geometries. We designed a simulation framework that generates more diverse and challenging source configurations through two complementary spatial models, structured sensor noise, and flexible temporal dynamics.

5.1 Spatial Source Models

`invertmeeg` provides two spatial source models that produce qualitatively different source geometries.

Diffusion basis patches. The first model constructs source patterns using a graph-based spatial basis derived from the cortical mesh adjacency. Given the adjacency matrix $\mathbf{A}$ of the source space, the diffusion-based gradient operator is:

$$\mathbf{G} = \mathbf{I} - \alpha_{\text{diff}} \cdot \mathbf{\Delta}(\mathbf{A}), \tag{15}$$

where $\mathbf{\Delta}(\mathbf{A})$ is the graph Laplacian and α_{diff} controls the smoothing strength. Multi-order spatial patterns are generated by iteratively applying $\mathbf{G}$ to a seed dipole, where the spatial order parameter k controls source extent: $k = 0$ produces point dipoles, $k = 1$–2 produces cortical patches, and $k \geq 3$ produces broadly distributed activity.

Contiguous Gaussian patches. The second model generates spatially extended sources with irregular, anatomically plausible geometries. For each source, a contiguous region of a specified size is grown from a random seed vertex using randomized breadth-first search (BFS) on the cortical adjacency graph. At each step, a random vertex is selected from the frontier (the set of unvisited neighbors of the current region), producing irregular region shapes that conform to the cortical surface topology rather than forming symmetric discs.

Within each region, one or two spatial components are created with Gaussian-weighted profiles. For each component, a center vertex c is chosen and the graph hop distance d_v from c to every vertex v in the region is computed via BFS restricted to the region. The spatial weights follow a Gaussian kernel:

$$w_v = \exp\left(-\frac{d_v^2}{2\sigma^2}\right), \tag{16}$$

where σ controls the spatial smoothness (in graph-hop units). For rank-2 patches, a second component with a different center produces two overlapping smooth maps within the same region, yielding more complex spatial patterns.

This approach produces sources that are (1) contiguous and connected on the cortical surface, (2) irregularly shaped due to the randomized BFS growth, and (3) smoothly varying within the patch rather than sharply bounded. These properties make contiguous Gaussian patches substantially more diverse than standard Gaussian blobs and more representative of the spatially extended generators observed in electrophysiological recordings of oscillatory or evoked activity.

5.2 Temporal Dynamics

Source time courses are drawn from a pre-generated pool of colored noise signals with power spectral density $S(f) \propto 1/f^{\beta}$, where β is uniformly sampled from a configured range. This produces time courses ranging from white noise ($\beta = 0$) to smooth, slowly varying signals ($\beta = 3$). Inter-source correlations are introduced via Cholesky decomposition of a correlation matrix, allowing control over the degree of temporal dependence between simultaneously active sources.

5.3 Sensor Noise Model

Sensor-level noise is added at a specified signal-to-noise ratio (SNR in dB). For the benchmark evaluation, we use a "realistic" noise model that combines three components to approximate the structure of empirical EEG noise:

$$\mathbf{C}_{\text{noise}} = \mathbf{A}\mathbf{A}^{\top} + \alpha \cdot \mathbf{C}_{\text{banded}} + \sigma^2 \mathbf{I}, \tag{17}$$

where $\mathbf{A} \in \mathbb{R}^{m \times r}$ is a random basis generating r low-rank artifact-like components (simulating environmental interference or physiological artifacts), $\mathbf{C}_{\text{banded}}$ captures spatial correlations between

neighboring channels (with elements $[\mathbf{C}_{\text{banded}}]_{ij} = \rho^{|i-j|}$ for a coefficient ρ), and $\sigma^2\mathbf{I}$ adds a white noise floor. All components are trace-normalized before combination to keep their relative contributions interpretable. The low-rank dimension r is randomly drawn from a configured range (typically 2–8), introducing variability in the noise structure across simulations.

This structured noise model is more challenging than i.i.d. Gaussian noise because it introduces spatial correlations and low-rank interference patterns that can bias inverse solutions toward artifactual source configurations, particularly for methods that assume spatially white noise.

5.4 Simulation Modes

Source activity in the benchmark consists of the selected source patches, representing the common assumption that a small number of discrete cortical generators are active. The framework additionally supports a mixture mode that combines sparse patches with smooth spatiotemporal background activity, but this mode is not used in the current benchmark.

6 Evaluation and Benchmarking

`invertmeeg` provides a standardized evaluation suite for comparing solver performance on simulated data with known ground truth.

6.1 Metrics

The following metrics are computed between the true source estimate $\mathbf{y}_{\text{true}}(t)$ and the predicted estimate $\mathbf{y}_{\text{pred}}(t)$. For the localization-style metrics (EMD, AP, MLE, and SD), the implementation first collapses each spatiotemporal estimate to a single unsigned spatial map by averaging absolute activity across time,

$$\bar{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T} |\mathbf{y}(t)|. \tag{18}$$

Distances are computed from Euclidean distances between dipole positions and reported in millimeters in the tables and figures. Correlation is the only metric that operates on the full time series. We designate EMD as the primary metric and AP as the secondary metric for ranking solvers, as they capture complementary aspects of reconstruction quality.

Earth Mover's Distance (EMD, primary metric). We compute the Wasserstein-1 distance between the collapsed maps $\bar{\mathbf{y}}_{\text{true}}$ and $\bar{\mathbf{y}}_{\text{pred}}$ using optimal transport [35]. Before transport, values below 25% of the maximum of each map are set to zero, and the remaining mass is normalized to sum to one. This suppresses low-amplitude diffuse tails that would otherwise dominate transport cost despite contributing little to the main reconstruction. The EMD then quantifies the minimum spatial "work" required to transform one thresholded source distribution into the other, jointly capturing localization error, missed mass, false positives, and extent mismatch.

Average Precision (AP, secondary metric). Average precision summarizes how well a solver ranks truly active dipoles above inactive ones across all score thresholds. The ground-truth support is binarized from $\bar{\mathbf{y}}_{\text{true}}$ using a relative threshold of 10% of its maximum. The prediction scores are the entries of $\bar{\mathbf{y}}_{\text{pred}}$. Dipoles are sorted by predicted score, a precision-recall curve is traced from that ordering, and AP is computed as the area under that curve. We prefer AP over the more commonly reported area under the ROC curve (AUC) because source imaging is an inherently imbalanced

classification problem: in a source space with $n = 1{,}284$ dipoles, a focal source activates only 1 dipole ($< 0.1\%$ positive rate), and even extended sources rarely exceed 3–5% of the source space. Under such extreme class imbalance, AUC is dominated by the true negative rate and can assign high scores to reconstructions that spread activity broadly, whereas AP focuses exclusively on the precision among predicted positives and is therefore more sensitive to false activations [36].

Mean Localization Error (MLE). MLE is computed from graph-local maxima rather than from centers of mass. Local maxima are detected separately in $\bar{\mathbf{y}}_{\text{true}}$ and $\bar{\mathbf{y}}_{\text{pred}}$ after thresholding each map at 10% of its maximum and comparing every candidate dipole against its neighbors in the cortical adjacency graph. At most five maxima are retained per map. The benchmark uses the default "DLE" aggregation mode implemented in the package: if D_{ij} is the Euclidean distance between true maximum i and predicted maximum j, then

$$\text{MLE}_{\text{DLE}} = \frac{1}{2}\left(\frac{1}{n_{\text{true}}}\sum_i \min_j D_{ij} + \frac{1}{n_{\text{pred}}}\sum_j \min_i D_{ij}\right). \tag{19}$$

This symmetric nearest-peak distance penalizes both missed true generators and spurious predicted ones.

Correlation. Correlation measures temporal fidelity on the original source time series rather than on the collapsed spatial maps. For multi-time-point data, a dipole is treated as active if its maximum absolute ground-truth amplitude reaches at least 10% of the global peak over all dipoles and times. For each active dipole, we compute the Pearson correlation between its true and predicted time courses and then average across active dipoles. By default the implementation uses *unsigned* correlation, i.e. it takes the absolute value of each Pearson coefficient before averaging, so sign-flipped but otherwise correct time courses are not penalized. For single-time-point data, the code falls back to a spatial Pearson correlation across dipoles.

Spatial Dispersion (SD). Spatial dispersion is reported as a blurring ratio based on a full-width-at-half-maximum-style proxy. For the collapsed true and predicted maps, nodes above 50% of the peak amplitude are identified, and a local extent measure is formed by summing distances to adjacent active nodes. SD is then the ratio of predicted extent to true extent, so values above 1 indicate a reconstruction that is spatially broader than the ground truth. We report SD descriptively but exclude it from the ranking procedure below.

6.2 Evaluation Scenarios

Different inverse methods are suited to different source configurations. We define four evaluation scenarios that span the range of conditions encountered in practice (Table 3):

1. **Focal Source**: A single point dipole at moderate-to-high SNR (3–5 dB). This baseline scenario tests pure localization accuracy and favors sparse/dipolar methods.

2. **Multi Source**: 2–4 simultaneously active focal dipoles at low-to-moderate SNR (1–3 dB). Tests the ability to resolve multiple concurrent sources without mutual interference.

3. **Extended Source**: 2–4 spatially extended contiguous Gaussian patches (spatial orders 1–3) at moderate SNR (1–3 dB). Uses the contiguous Gaussian patch model (Section 5.1) with

rank-1 or rank-2 temporal dynamics within each patch, producing irregularly shaped cortical activations.

4. **Noisy**: 1–3 sources with variable spatial extent (orders 0–2) under challenging low-SNR conditions (-5 to $0\,\mathrm{dB}$). Tests robustness to noise, which is the primary practical challenge in clinical and cognitive applications.

All four scenarios use the "realistic" noise model (Section 5.3) with a low-rank artifact dimension drawn uniformly from $[2, 8]$.

Table 3: Configuration of the four EEG benchmark scenarios in the frozen paper snapshot.

Parameter	Focal	Multi	Extended	Noisy
Sources	1	[2, 4]	[2, 4]	[1, 3]
Spatial model	Diffusion basis	Diffusion basis	Contiguous Gaussian	Diffusion basis
Order range	0	0	[1, 3]	[0, 2]
Patch rank	[1, 2]	[1, 2]	[1, 2]	[1, 2]
SNR (dB)	[3.0, 5.0]	[1.0, 3.0]	[1.0, 3.0]	[-5.0, 0.0]
Time points	200	200	200	200
Correlation mode	realistic	realistic	realistic	realistic
Noise low-rank dim.	[2, 8]	[2, 8]	[2, 8]	[2, 8]
Simulations	50	50	50	50

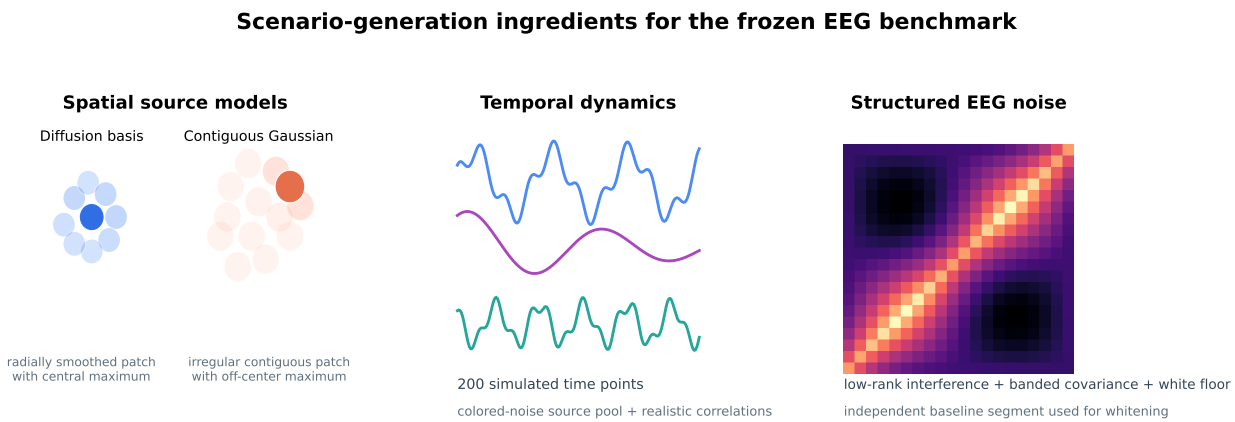

Figure 3: Scenario-generation overview for the frozen EEG benchmark. The benchmark combines two spatial source models, colored temporal dynamics, and structured EEG noise. The left panel illustrates the diffusion-basis and contiguous-Gaussian source constructions used in the simulation framework, without assigning them one-to-one to every scenario shown elsewhere in the manuscript.

6.3 Ranking Procedure

Solver rankings are computed from the frozen paper snapshot rather than re-derived manually in the manuscript. For each dataset, solvers receive dense ranks on four metrics: EMD, AP, MLE, and correlation. Spatial dispersion is reported but excluded from ranking. The per-dataset rank is the mean of those four metric ranks, and the global rank is the mean of the four per-dataset ranks. All ties and averages reported below follow that benchmark-runner definition.

6.4 Experimental Setup

The paper reports a frozen EEG benchmark snapshot containing 106 evaluated methods on a shared BioSemi-32 electrode montage with an `ico3` source space comprising $n = 1{,}284$ fixed-orientation dipoles. The snapshot covers 4 scenarios with 50 simulations per scenario. It therefore supports claims about this specific EEG benchmark family: synthetic EEG data, one 32-channel montage, one source space resolution, and one fixed-orientation forward-model configuration.

For each simulation, a noise covariance matrix is estimated from a separate 200-time-point baseline segment drawn independently from the same noise model (spatial covariance and temporal coloring) as the noise added to the signal. This simulates the common experimental practice of estimating noise statistics from an empty-room recording or pre-stimulus baseline. The empirical covariance is regularized with a fixed shrinkage toward a scaled identity ($\gamma = 0.05$). Each solver receives this estimated noise covariance and uses it to spatially whiten both the leadfield and the data via eigendecomposition-based whitening: the noise covariance is decomposed as $\mathbf{C}_n = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$, and the whitener $\mathbf{W} = \mathbf{\Lambda}^{-1/2}\mathbf{V}^\top$ is applied, with rank truncation to avoid amplifying near-zero-variance directions. This ensures that all solvers operate on data with approximately identity noise covariance, placing them on equal footing regardless of the original noise structure.

The `BenchmarkRunner` class automates the evaluation pipeline: it generates simulated data for each scenario, applies each solver with its recommended depth-weighting parameters, computes all metrics, and aggregates results. The novel solvers were not tuned on the exact simulation instances used in the frozen evaluation; new random draws were generated for benchmarking. However, development and evaluation share the same broad simulation assumptions, so external validity remains limited to nearby EEG regimes rather than being claimable as a universal solver ranking.

6.5 Benchmark Results

Figure 4 summarizes the trade-off between reconstruction quality (mean EMD) and inference time across all 106 methods. The full leaderboard and per-scenario tables are moved to the Appendix (Tables 5 to 9).

Overall ranking. The frozen snapshot yields a tie for best global rank between **FLEX-GreedyML** and **Hydra** (mean global rank 1.5), followed by **Subspace-SBL+** (4.0), **FLEX-SSM** (4.06), and **Subspace-SBL** (4.62). This top tier is dominated by flexible subspace scanners, hybrid switching methods, and Bayesian refinements built on subspace detection. The practical implication is not that a single universal winner exists, but that methods combining explicit source enumeration with flexible spatial extent modeling perform strongly across this benchmark family.

Focal and multi-source scenarios. The focal scenario is deliberately easy for accurate sparse scanners: many methods achieve essentially perfect localization, so speed and robustness become the main differentiators. The multi-source scenario is more informative. Here **GreedyML** yields the best mean EMD (8.7 mm), followed by **FLEX-GreedyML** and **Hydra** (9.1 mm), **ExhaustiveML** (11.0 mm), and **Subspace-SBL+** (12.0 mm). This is the regime in which explicit source-enumeration methods separate most clearly from diffuse minimum-norm solutions.

Extended source scenario. For spatially extended sources, **Subspace-SBL+** leads (EMD 22.6 mm), followed closely by **FLEX-GreedyML** and **Hydra** (22.9 mm), **FLEX-SSM** (26.2 mm), and **Subspace-SBL** (26.5 mm). This scenario uses the contiguous Gaussian patch model rather than the diffusion basis used in focal and multi-source conditions, which helps test whether flexible

solvers generalize beyond a single source-construction rule. MSP also ranks well here, consistent with the advantage of distributed priors when the ground truth is genuinely extended. The qualitative comparison in Figure 1 illustrates the same point at the single-example level: for these moderately compact patches, subspace and Bayesian methods recover the bilateral generators more faithfully than the diffuse minimum-norm family. This figure should not be overgeneralized to all extended activity, however. If the true source prior were much broader or smoother than the contiguous Gaussian patches used here, methods such as MNE or eLORETA could become more competitive because their spatial smoothness assumptions would better match the generating process.

Noisy scenario. Under low-SNR conditions, **Subspace-SBL** achieves the best EMD (9.3 mm), closely followed by **Chimera** (9.7 mm), **FLEX-GreedyML** and **Hydra** (10.5 mm), and **FLEX-AP** and **FLEX-SSM** (10.9 mm). Bayesian methods generally maintain their performance better than beamformers under noise, consistent with the regularizing effect of their probabilistic priors.

Deep learning solvers. Three neural network solvers were included in the benchmark: CovCNN-KL, CovCNN, and FC. CovCNN-KL achieves the best performance among them (mean global rank 68.5, mean EMD 44.5 mm), placing it in the range of mid-tier classical beamformers such as DICS and LCMV. CovCNN (78.06, 52.0 mm) and FC (79.94, 55.6 mm) perform similarly to lower-ranked minimum norm methods. Notably, all three neural network solvers outperform the random-noise baseline and several classical methods, but fall substantially short of the top-performing classical solvers. The benchmark used reduced model capacity (24 dense units, 200 training epochs) to keep training tractable within the evaluation framework, resulting in approximately 10^6 trainable parameters per model. These ANN results should therefore be interpreted as constrained-budget baselines rather than as a definitive upper bound on deep-learning performance.

7 Usage Examples

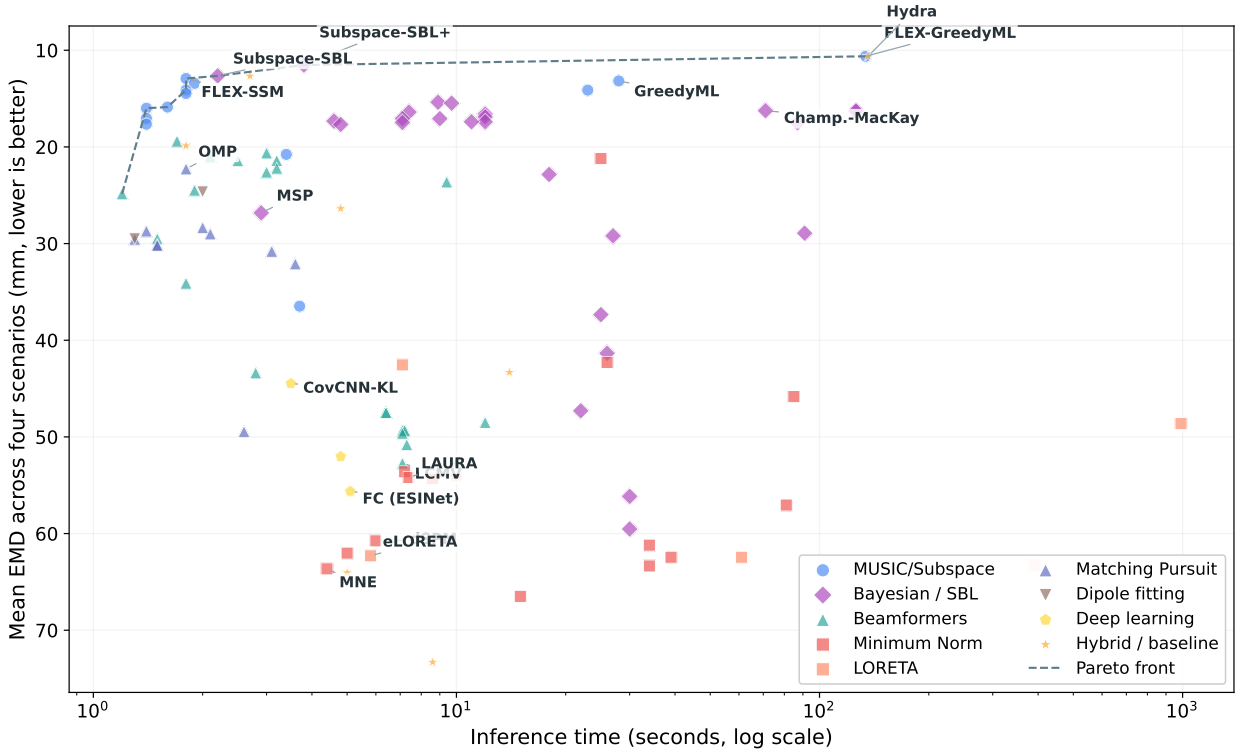

Figure 4: Performance–compute trade-off for the frozen EEG benchmark. The x-axis reports inference time only; neural-network training cost is discussed separately in the text. Labeled points include the Pareto front, canonical baselines, and the benchmarked neural-network baselines.

7.1 Basic Source Imaging

The following example demonstrates the core workflow:

```
import mne
from invert import Solver

# Load forward model and data
forward = mne.read_forward_solution("sample-fwd.fif")
evoked = mne.read_evokeds("sample-ave.fif")[0]

# Create solver and compute inverse operator
solver = Solver("eloreta")
solver.make_inverse_operator(forward, alpha="auto")

# Apply to data
stc = solver.apply_inverse_operator(evoked)

# Visualize
stc.plot(subject="fsaverage", hemi="both")
```

Switching to any other solver requires changing only the name string:

```
for method in ["mne", "dspm", "lcmv", "champagne", "music"]:
    solver = Solver(method)
    solver.make_inverse_operator(forward, alpha="auto")
    stc = solver.apply_inverse_operator(evoked)
```

7.2 Simulation and Benchmarking

The same API exposes the simulation machinery used to build the benchmark:

```
from invert.simulate import SimulationConfig, SimulationGenerator

# Configure simulation
config = SimulationConfig(
    n_sources=(1, 5),
    n_orders=(0, 2),
    snr_range=(-5, 5),
    batch_size=100,
)

# Generate data
gen = SimulationGenerator(forward, config)
x, y, info = next(gen.generate())

# x: sensor data [100, n_channels, n_timepoints]
# y: ground truth [100, n_dipoles, n_timepoints]
# info: metadata DataFrame
```

Figure 5 shows one representative sample from each benchmark scenario together with reconstructions from four solvers spanning different method families.

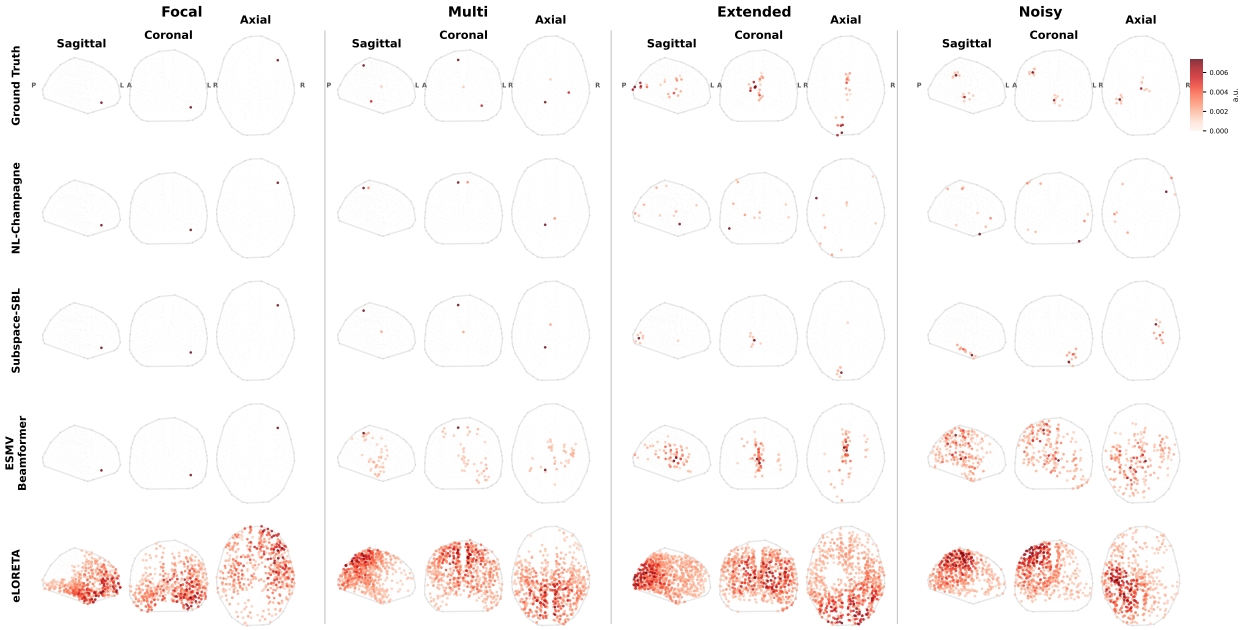

Figure 5: Qualitative reconstruction grid for one representative sample from each benchmark scenario. Columns show focal, multi-source, extended, and noisy simulations; rows show the ground truth together with reconstructions from NL-Champagne, Subspace-SBL, ESMV Beamformer, and eLORETA. The figure highlights distinct failure modes across scenarios: compact Bayesian solutions remain sparse in focal settings, ESMV is accurate for focal activity but becomes increasingly diffuse in the more challenging regimes, and eLORETA consistently produces broader spatial estimates.

8 Discussion

8.1 Practical Guide to Solver Selection

A key goal of this benchmark is to provide practical guidance for practitioners. Based on the results across all four scenarios, we offer the following recommendations, organized by the expected source configuration. Table 4 summarizes these recommendations.

Table 4: Solver selection guide based on expected source configuration and computational budget. Inference times on a single CPU core (BioSemi-32, ico3 source space).

Scenario	**Recommended (fast, < 5 s)**	**Recommended (best, any time)**
Focal dipole(s)	FLEX-SSM, FLEX-MUSIC, OMP	FLEX-GreedyML, Omni-Champagne
Multiple sources	FLEX-SSM, SSM, Chimera	FLEX-GreedyML, Hydra
Extended patches	FLEX-SSM, FLEX-MUSIC	Subspace-SBL+, FLEX-GreedyML
Low SNR	Subspace-SBL, FLEX-SSM	Subspace-SBL, Chimera
Unknown/general	FLEX-SSM, Chimera	FLEX-GreedyML

When sources are expected to be focal. Many solvers achieve near-perfect performance on focal sources, so the differentiator is computational cost. Subspace methods (FLEX-MUSIC, FLEX-SSM, SSM) and matching pursuit methods (OMP) are fast (∼1–2 s) and highly accurate. Bayesian methods (Omni-Champagne, Flex-Champagne) achieve slightly better correlation but at higher computational cost (∼3–5 s).

When multiple sources are expected. Resolving multiple concurrent sources is the core strength of subspace scanning methods. FLEX-GreedyML and Hydra consistently lead, with SSM and ExhaustiveML close behind. Classical beamformers struggle in this regime due to signal cancellation between correlated sources.

When sources are spatially extended. Extended sources violate the sparsity assumption of many methods. Subspace-SBL+ and Subspace-SBL, which learn flexible source priors, excel here. Gamma-MAP-MSP, which employs a multiple sparse priors basis, also performs well (AP = 0.45). Pure sparse methods (OMP, CoSaMP, MCE) should be avoided.

When SNR is low. Subspace-SBL stands out under noisy conditions, consistently achieving the best rank across all metrics. Bayesian methods with noise learning (Champagne variants with homoscedastic or diagonal noise estimation) maintain reasonable performance. Beamformers and minimum norm methods degrade substantially as they lack the regularizing effect of structured source priors.

When no prior information is available. For a general-purpose solver that performs strongly across all scenarios in this benchmark, **FLEX-GreedyML** offers the best overall accuracy but at a higher computational cost (∼134 s per sample). For a fast default, **FLEX-SSM** (∼1.8 s) or **Chimera** (∼2.7 s) provide a good balance of performance and speed. Under this benchmark

configuration, classical minimum norm defaults such as MNE and eLORETA rank well below several subspace, hybrid, and Bayesian alternatives.

Inference time. Figure 4 shows the performance–compute trade-off across all solvers (inference time only; training time for neural networks is excluded). Solver inference times span three orders of magnitude. The fastest solvers (< 2 s) include most subspace scanners (SSM, FLEX-SSM, AP, RAP-MUSIC), matching pursuit methods (OMP, CoSaMP), and simple beamformers (SSP-ESMV, LCMV-MVPURE). Moderate-cost solvers (2–10 s) include most Champagne variants, the ESMV family, and basic minimum norm methods. The most expensive solvers (> 30 s) include iterative Bayesian methods (VB-SBL: 91 s, BCS: 87 s), L_1-based methods (MCE: 85 s), and hybrid methods (Hydra: 136 s, FLEX-GreedyML: 134 s). Total Variation (392 s) and SSLOFO (991 s) are the slowest solvers. Deep learning solvers have modest inference times (FC: $\sim$5 s, CovCNN-KL: $\sim$4 s) but require substantial upfront training ($\sim$250–1,900 s per scenario). The Pareto front (Figure 4) is dominated by subspace methods at the fast end and FLEX-GreedyML at the accurate end, with Subspace-SBL variants offering an attractive middle ground.

8.2 Related Software

Several established toolboxes provide M/EEG source imaging functionality. `MNE-Python` [14] includes MNE, dSPM, sLORETA, eLORETA, LCMV, DICS, MUSIC, and RAP-MUSIC. Brainstorm [15] (MATLAB) provides minimum norm variants, beamformers, and dipole fitting. FieldTrip [16] (MATLAB) focuses on beamforming methods. SPM/DAiSS [17] (MATLAB) offers Bayesian approaches including MSP and variational methods. ESINet [12] introduced deep learning for EEG source imaging but was limited to fully connected architectures.

`invertmeeg` differs from these tools in scope and methodology. With 118 registered implementations and 106 methods included in the frozen EEG snapshot, it enables the kind of large-scale systematic comparison presented here, which would be difficult when methods are scattered across different toolboxes with incompatible interfaces and data formats. The bundling of simulation, solving, and evaluation into one package ensures that all methods are compared under identical conditions—same forward model, same noise realizations, same metrics—eliminating confounds that arise from cross-toolbox comparisons.

8.3 Design Decisions and Limitations

All solvers in `invertmeeg` are re-implementations based on published algorithms, not wrappers around original authors' code. While we have validated the implementations against published results where possible, subtle differences may exist. Users who require exact reproduction of a specific paper's results should verify against the original implementation.

The benchmark is intentionally narrow. It is EEG-only, uses synthetic simulations only, and fixes a single BioSemi-32 montage, a single `ico3` source space (1,284 dipoles), and a fixed-orientation forward model. These results should therefore be interpreted as valid within this benchmark family rather than as universal rankings for all EEG inverse problems.

The four scenarios span more than one source model, but they are not assumption-free. The focal and multi-source conditions use diffusion-basis patches with different order ranges, whereas the extended-source condition uses contiguous Gaussian patches; the low-SNR condition is defined by its noise regime rather than by a separate spatial model in Figure 3. This reduces but does not eliminate the possibility that some solver families are advantaged by similarity between their internal priors and the simulation family.

The novel solvers were developed under the same broad simulation assumptions used in the benchmark, even though the evaluation instances themselves were regenerated with different random draws. This means the frozen snapshot is informative about within-family generalization, but external validity to different EEG montages, source spaces, or real data remains limited.

The package currently supports only fixed-orientation source models. Free-orientation forward models are automatically converted to fixed orientation, which may not be appropriate for all applications.

Computational cost varies substantially across solver categories. Linear solvers (MNE, eLORETA) compute an explicit inverse operator and are fast to apply; iterative solvers (Champagne, BCS) may require many iterations to converge; and deep learning solvers require training, which can take minutes to hours. The neural-network baselines reported here were trained with reduced model capacity and training budget to keep the benchmark tractable; stronger ANN performance may be achievable with larger models and longer training, but that was not evaluated in this frozen snapshot. `invertmeeg` does not currently provide GPU acceleration for non-neural-network solvers.

The package is licensed under the GPL-3.0-only license.

8.4 Future Work

Planned extensions include GPU-accelerated linear algebra for large-scale source spaces, time-frequency source imaging methods, connectivity-informed inverse solutions, and integration with parcellation-based analyses. We plan to extend the benchmark to MEG sensor configurations and higher-density EEG montages, and to include real-data validation on established phantom and auditory/visual paradigm datasets.

We also plan to analyze the role of AI coding agents in solver prototyping more systematically. Several agent-assisted solvers rank strongly within this benchmark, but that observation should be interpreted cautiously until it is tested under held-out simulation families and real-data validation. A companion paper dedicated to that development workflow is forthcoming.

9 Conclusion

We have presented a frozen EEG benchmark of 106 inverse solvers across 4 scenarios, together with `invertmeeg`, a unified Python package that currently exposes 118 solver implementations through the `MNE-Python` ecosystem. Within this benchmark family, solver choice has a large effect on reconstruction quality, and the best-performing methods depend on source geometry, source multiplicity, and noise level.

Key findings include: (1) flexible subspace scanners and hybrid switching methods occupy the top of the frozen global ranking; (2) Bayesian refinements such as Subspace-SBL remain especially competitive in the noisy and extended-source regimes; (3) classical minimum norm defaults rank below several alternatives under this specific EEG setup; and (4) the benchmarked deep-learning baselines remain well behind the strongest classical solvers under the constrained training budget used here.

Not every package solver is benchmarked in this paper, but all 118 registered implementations are available in `invertmeeg`, together with the simulation framework and evaluation tools needed to reproduce and extend the EEG benchmark. The source code is available at `https://github.com/LukeTheHecker/invertmeeg` under the GPL-3.0-only license and can be installed via `pip install invertmeeg`. Documentation and tutorials are available at `https://lukethehecker.github.io/invertmeeg/`.

References

[1] M. S. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413, 1993.

[2] S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30, 2001.

[3] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, and B. Vanrumste. Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 5(1):25, 2008.

[4] M. S. Hämäläinen and R. J. Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1):35–42, 1994.

[5] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18(1):49–65, 1994.

[6] R. D. Pascual-Marqui. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, 24(Suppl D):5–12, 2002.

[7] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880, 1997.

[8] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan. Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. *NeuroImage*, 49(1):641–655, 2010.

[9] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, T. Henson, G. Flandin, and J. Mattout. Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3):1104–1120, 2008.

[10] I. F. Gorodnitsky, J. S. George, and B. D. Rao. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and Clinical Neurophysiology*, 95(4):231–251, 1997.

[11] J. C. Mosher, P. S. Lewis, and R. M. Leahy. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Transactions on Biomedical Engineering*, 39(6):541–557, 1992.

[12] L. Hecker, R. Rupprecht, L. H. Thi, and S. Braun. ConvDip: A convolutional neural network for better EEG source imaging. *Frontiers in Neuroscience*, 15:569918, 2021.

[13] L. Hecker, R. Rupprecht, and S. Braun. Long short-term memory and convolutional neural networks for EEG source imaging with deep learning. *Unpublished manuscript*, 2022.

[14] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7:267, 2013.

[15] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy. Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011:879716, 2011.

[16] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:156869, 2011.

[17] V. Litvak, J. Mattout, S. Kiebel, C. Phillips, R. Henson, J. Kilner, G. Barnes, R. Oostenveld, J. Daunizeau, G. Flandin, W. Penny, and K. Friston. EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, 2011:852961, 2011.

[18] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67, 2000.

[19] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[20] R. D. Pascual-Marqui. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. *arXiv preprint arXiv:0710.3341*, 2007.

[21] D. P. Wipf and S. S. Nagarajan. A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*, 44(3):947–966, 2009.

[22] C. Amblard, E. Lapalme, and J.-M. Lina. Biomagnetic source detection by maximum entropy and graphical models. *IEEE Transactions on Biomedical Engineering*, 51(3):427–442, 2004.

[23] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699, 2001.

[24] J. Vrba and S. E. Robinson. Signal processing in magnetoencephalography. *Methods*, 25(2):249–271, 2001.

[25] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[26] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[27] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, 2009.

[28] J. C. Mosher and R. M. Leahy. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47(2):332–340, 1999.

[29] S. Sommariva and A. Sorrentino. Sequential Monte Carlo samplers for semi-linear inverse problems and application to magnetoencephalography. *Inverse Problems*, 30(11):114020, 2014.

[30] L. Hecker, L. Tebartz van Elst, and J. Kornmeier. Source localization using recursively applied and projected MUSIC with flexible extent estimation. *Frontiers in Neuroscience*, 17:1170862, 2023.

[31] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.

[32] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[33] M. A. Lukas. Strong robust generalized cross-validation for choosing the regularization parameter. *Inverse Problems*, 26(4):045002, 2010.

[34] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, 1992.

[35] C. Villani. *Optimal Transport: Old and New*. Springer, 2009.

[36] T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3):e0118432, 2015.

[37] S. Yang, M. Jiao, J. Xiang, N. Fotedar, H. Sun, and F. Liu. Rejuvenating classical brain electrophysiology source localization methods with spatial graph Fourier filters for source extents estimation. *Brain Informatics*, 11(1):8, 2024.

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

[39] A. Gramfort, M. Kowalski, and M. Hämäläinen. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology*, 57(7):1937–1961, 2012.

[40] H. Liu, P. H. Schimpf, G. Dong, X. Gao, F. Yang, and S. Gao. Standardized shrinking LORETA-FOCUSS (SSLOFO): A new algorithm for spatio-temporal EEG source reconstruction. *IEEE Transactions on Biomedical Engineering*, 52(10):1681–1691, 2005.

[41] R. Grave de Peralta Menendez, S. Gonzalez Andino, G. Lantz, C. M. Michel, and T. Landis. Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topography*, 14(2):131–137, 2001.

[42] G. Backus and F. Gilbert. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, 16(2):169–205, 1968.

[43] S. Baillet and L. Garnero. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Transactions on Biomedical Engineering*, 44(5):374–385, 1997.

[44] S. Wang, C. Wei, K. Lou, D. Gu, and Q. Liu. Advancing EEG/MEG source imaging with geometric-informed basis functions. *arXiv preprint arXiv:2401.17939*, 2024.

[45] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992.

[46] M. Aud'hui, A. Kachenoura, L. Albera, I. Merlet, M. Yochum, P. Van Bogaert, and P. Benquet. SISSY: Structured iterative support selection with sparsity for EEG source imaging. *NeuroImage*, 2025. Early-access citation; volume and page details pending.

[47] A. Hashemi and S. Haufe. Improving EEG source localization through spatio-temporal sparse Bayesian learning. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1935–1939. IEEE, 2018.

[48] C. Cai, A. Hashemi, M. Diwakar, S. Haufe, K. Sekihara, and S. S. Nagarajan. Robust estimation of noise for electromagnetic brain imaging with the Champagne algorithm. *NeuroImage*, 225:117411, 2021.

[49] Z. Feng, C. Guan, and Y. Sun. Block-Champagne: a novel Bayesian framework for imaging extended E/MEG source. *IEEE Transactions on Medical Imaging*, 2025. Early-access citation; volume and page details pending.

[50] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[51] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, 2008.

[52] B. D. Van Veen and K. M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, 1988.

[53] J. Jurkowska, J. Dreszer, M. Lewandowska, and T. Piotrowski. Multi-source neural activity indices and spatial filters for EEG/MEG inverse problem: An extension to MNE-Python. *arXiv preprint arXiv:2509.14118*, 2025.

[54] Y. Jonmohamadi, G. Poudel, C. Innes, D. Weiss, R. Krueger, and R. Jones. Comparison of beamformers for EEG source signal reconstruction. *Biomedical Signal Processing and Control*, 14:175–188, 2014.

[55] S. E. Robinson and J. Vrba. Functional neuroimaging by synthetic aperture magnetometry (SAM). In *Recent Advances in Biomagnetism*, pages 302–305, 1999.

[56] K. Sekihara, S. S. Nagarajan, D. Poeppel, and A. Marantz. Asymptotic SNR of scalar and vector minimum-variance beamformers for neuromagnetic source reconstruction. *IEEE Transactions on Biomedical Engineering*, 51(10):1726–1734, 2004.

[57] J. L. Yu and C. C. Yeh. Generalized eigenspace-based beamformers. *IEEE Transactions on Signal Processing*, 43(11):2453–2461, 1995.

[58] A. S. Nunes, A. Moiseev, N. Kozhemiako, T. Cheung, U. Ribary, and S. M. Doesburg. Multiple constrained minimum variance beamformer (MCMV) performance in connectivity analyses. *NeuroImage*, 208:116386, 2020.

[59] A. Kuznetsova, Y. Nurislamova, and A. Ossadtchi. Modified covariance beamformer for solving MEG inverse problem in the environment with correlated sources. *NeuroImage*, 228:117677, 2021.

[60] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.

[61] M. Wax and A. Adler. Direction of arrival estimation in the presence of model errors by signal subspace matching. *Signal Processing*, 181:107900, 2021.

[62] A. Adler, M. Wax, and D. Pantazis. Brain source localization by alternating projection. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2022.

[63] M. Mishali and Y. C. Eldar. Reduce and boost: recovering arbitrary sets of jointly sparse vectors. *IEEE Transactions on Signal Processing*, 56(10):4692–4702, 2008.

[64] M. F. Duarte and Y. C. Eldar. Structured compressed sensing: from theory to applications. *IEEE Transactions on Signal Processing*, 59(9):4053–4085, 2011.

[65] C. Grova, J. Daunizeau, E. Kobayashi, A. P. Bagshaw, J. M. Lina, F. Dubeau, and J. Gotman. Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes. *NeuroImage*, 39(2):755–774, 2008.

[66] R. A. Chowdhury, J. M. Lina, E. Kobayashi, and C. Grova. MEG source localization of spatially extended generators of epileptic activity: comparing entropic and hierarchical Bayesian approaches. *PloS One*, 8(2):e55969, 2013.

[67] L. Albera, A. Ferréol, D. Cosandier-Rimélé, I. Merlet, and F. Wendling. Brain source localization using a fourth-order deflation scheme. *IEEE Transactions on Biomedical Engineering*, 55(2):490–501, 2008.

[68] C. Cai, H. Kang, A. Hashemi, D. Chen, M. Diwakar, S. Haufe, K. Sekihara, W. Wu, and S. S. Nagarajan. Bayesian algorithms for joint estimation of brain activity and noise in electromagnetic imaging. *IEEE Transactions on Medical Imaging*, 42(3):762–773, 2023.

[69] C. Cai, A. Hashemi, M. Diwakar, D. Chen, S. Haufe, K. Sekihara, and S. S. Nagarajan. Enhancing M/EEG source imaging with Champagne: model validation and novel contributions. *NeuroImage*, 2024. Early-access citation; volume and page details pending.

[70] R. Grave de Peralta Menendez, M. M. Murray, C. M. Michel, R. Martuzzi, and S. L. Gonzalez Andino. Electrical neuroimaging based on biophysical constraints. *NeuroImage*, 21(2):527–539, 2004.

[71] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[72] F.-H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hämäläinen. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, 31(1):160–171, 2006.

[73] J. P. Owen, D. P. Wipf, H. T. Attias, K. Sekihara, and S. S. Nagarajan. Performance evaluation of the Champagne source reconstruction algorithm on simulated and real M/EEG data. *NeuroImage*, 60(1):305–323, 2012.

[74] K. Uutela, M. Hämäläinen, and E. Somersalo. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage*, 10(2):173–180, 1999.

[75] H. Wu, S. Gao, J. Li, and X. Li. A new method for spatio-temporal EEG source reconstruction: standardized shrinking LORETA-FOCUSS (SSLOFO). *IEEE Transactions on Biomedical Engineering*, 52(11):1781–1792, 2005.

[76] Ü. Aydin, S. Rampp, A. Wollbrink, H. Kugel, J.-H. Cho, T. R. Knösche, C. Grova, and C. H. Wolters. Zoomed MRI-guided TRAP-MUSIC for MEG/EEG source analysis in presurgical evaluation of epilepsy patients. *NeuroImage: Clinical*, 17:566–575, 2018.

A Detailed Benchmark Results

The following tables provide per-metric results for each of the four evaluation scenarios. Values are reported as means across 50 independent simulations per scenario, sorted by EMD. Metrics: EMD (earth mover's distance in mm, primary), AP (average precision, secondary), MLE (mean localization error in mm), Corr (correlation).

Table 5: Earth mover's distance (EMD in mm↓, lower is better) across all four EEG benchmark scenarios for all 106 evaluated solvers, sorted by mean EMD. Means are computed across 50 simulations per scenario. Average inference time in seconds excludes ANN training. Best value per scenario in **bold**.

Solver	Cat.	Focal	Multi	Extended	Noisy	s	Solver	Cat.	Focal	Multi	Extended	Noisy	s
FLEX-GreedyML	MUS	**0.0**	9.1	22.9	10.5	134	Gamma-MAP-MSP	Bay	18.5	42.9	27.9	27.5	27
Hydra	HYB	**0.0**	9.1	22.9	10.5	136	ECD	DIP	**0.0**	45.1	43.8	29.1	1.3
Subspace-SBL+	Bay	**0.0**	12.0	**22.6**	11.5	3.8	LCMV-MVPURE	BF	11.5	38.8	38.3	29.4	1.5
Subspace-SBL	Bay	**0.0**	14.8	26.5	**9.3**	2.2	SP	MP	13.3	36.4	38.2	30.3	1.3
Chimera	HYB	**0.0**	14.4	26.6	9.7	2.7	FLEX-CoSaMP	MP	8.4	42.4	40.1	29.5	1.5
FLEX-SSM	MUS	**0.0**	14.6	26.2	10.9	1.8	CoSaMP	MP	8.4	40.4	41.0	31.0	1.5
GreedyML	MUS	**0.0**	**8.7**	27.2	16.8	28	ReMBo	MP	4.4	39.8	42.2	36.8	3.1
FLEX-AP	MUS	**0.0**	15.2	27.6	10.9	1.9	FLEX-ReMBo	MP	4.7	40.1	45.5	38.1	3.6
FLEX-MUSIC	MUS	**0.0**	14.4	30.2	11.9	1.8	ESMV-MVPURE	BF	22.6	41.1	40.4	32.3	1.8
ExhaustiveML	MUS	**0.0**	11.0	28.5	17.0	23	MUSIC	MUS	31.9	40.3	36.1	37.6	3.7
FLEX-TRAP-MUSIC	MUS	**0.0**	15.6	30.6	11.7	1.8	Source-MAP	Bay	34.2	41.2	34.3	39.7	25
Flex-NL-Champagne	Bay	**0.0**	19.2	27.9	14.4	8.9	Block-Champagne	Bay	36.5	42.7	35.0	51.2	26
Flex-NL-Champ.-Omix	Bay	**0.0**	20.2	27.2	14.5	9.7	L1/L2	MN	25.3	44.3	42.7	56.9	26
AP	MUS	**0.0**	14.4	31.3	17.8	1.6	SR-eLORETA	LOR	46.2	42.1	38.0	43.9	7.1
SSM	MUS	**0.0**	13.9	32.4	17.7	1.4	Champ.-MacKay-NL-F	Bay	18.9	45.2	48.2	61.0	14
Champ.-Conv.	Bay	0.3	16.6	27.9	20.2	126	DICS	BF	21.9	51.0	48.4	52.2	2.8
Champ.-MacKay	Bay	0.4	16.6	27.9	20.1	71	CovCNN-KL	NN	45.6	50.1	41.5	40.7	3.5
Champ.-LowSNR	Bay	0.4	15.8	28.1	20.9	126	MCE	MN	38.5	46.0	42.5	56.3	85
Champ.-MacKay-NL-H	Bay	1.8	14.8	28.2	20.8	7.4	Champ.-Conv.-NL-F	Bay	34.7	44.2	48.6	61.7	22
Champ.-Conv.-NL-H	Bay	2.0	15.2	28.3	20.9	12	HOCMCMV	BF	48.9	45.5	42.0	53.5	6.4
NL-Champagne	Bay	1.4	15.8	27.9	22.5	12	HOCMV	BF	48.9	45.5	42.0	53.5	6.4
RAP-MUSIC	MUS	**0.0**	12.9	35.8	19.5	1.4	ReciPSIICOS	BF	48.4	47.5	43.9	54.2	12
Champ.-MacKay-NL-P	Bay	1.9	16.2	28.4	21.8	7.1	SSLOFO	LOR	45.2	46.6	50.7	52.0	990
Champ.-MacKay-NL-D	Bay	1.9	16.2	28.4	21.8	9.0	Unit-Noise-Gain	BF	51.2	47.2	43.8	55.0	7.2
Flex-Champagne	Bay	**0.0**	23.6	29.5	16.2	4.6	WNMV	BF	51.2	47.2	43.8	55.0	7.1
Champ.-Conv.-NL-P	Bay	2.3	16.5	28.7	22.1	12	SAM	BF	52.5	47.4	43.8	53.7	7.2
Champ.-Conv.-NL-D	Bay	2.3	16.5	28.7	22.1	11	SubSMP	MP	48.9	52.6	46.4	49.9	2.6
Omni-Champagne	Bay	**0.0**	22.1	30.7	17.1	7.1	MVAB	BF	51.9	47.6	44.0	55.0	7.1
BCS	Bay	3.7	16.0	28.5	21.8	87	SMV	BF	53.4	48.7	45.3	55.8	7.3
TRAP-MUSIC	MUS	**0.0**	15.0	36.2	19.4	1.4	CovCNN	NN	59.9	52.5	43.6	52.1	4.8
Flex-Champ.-Omix	Bay	**0.0**	24.0	29.9	16.8	4.8	MCMV	BF	56.2	50.1	46.3	57.6	10
IR-ESMV	BF	**0.0**	27.1	29.0	21.5	1.7	LCMV	BF	57.2	49.5	46.3	57.9	7.1
SSP-IR-ESMV	BF	**0.0**	28.4	29.3	21.8	1.8	GFT-MNE	MN	60.2	51.3	46.5	56.1	9.3
Safe-Flex-ESMV	BF	**0.0**	27.2	34.2	21.1	3.0	EPIFOCUS	MN	60.2	51.2	47.1	55.9	7.2
EXSO-MUSIC	MUS	11.4	22.2	29.3	20.2	3.4	sLORETA	LOR	60.4	50.9	47.6	56.9	10
ESMV	BF	**0.0**	29.4	30.3	24.2	2.1	LAURA	MN	60.6	52.5	47.2	56.5	7.4
MxNE-BCD	MN	2.7	23.9	29.5	28.7	25	wMNE	MN	60.9	51.8	47.4	57.1	8.6
Deblur-Flex-ESMV	BF	**0.0**	30.1	35.1	20.4	3.2	FC (ESINet)	NN	61.7	52.2	49.1	59.5	5.1
Flex-ESMV	BF	**0.0**	30.1	35.2	20.4	2.5	Source-MAP-MSP	Bay	68.7	59.3	36.7	59.9	30
Adapt-Flex-ESMV	BF	**0.0**	28.5	34.1	26.2	3.2	GFT-L1	MN	63.1	53.7	50.5	60.9	81
OMP	MP	**0.0**	31.2	36.2	21.7	1.8	cMEM	Bay	61.8	59.1	54.5	62.7	30
Sharp-Flex-ESMV	BF	**0.0**	28.8	35.3	26.3	3.0	dSPM	MN	69.5	56.3	53.5	63.7	6.0
Gamma-MAP	Bay	0.1	35.6	28.8	26.9	18	APE-SISSY	MN	72.8	60.9	42.6	68.5	34
EBB	BF	**0.0**	30.1	38.5	25.9	9.4	GBF	MN	73.3	53.5	57.9	63.4	5.0
Flex-ESMV-MVPURE	BF	**0.0**	29.8	41.9	26.2	1.9	eLORETA	LOR	74.9	53.8	58.5	62.0	5.8
SESAME	DIP	0.4	34.5	40.2	23.5	2.0	LORETA	LOR	71.8	57.8	56.2	64.0	61
SSP-ESMV	BF	0.6	38.7	34.3	25.8	1.2	S-MAP	MN	71.8	57.8	56.2	64.0	39
APSF	HYB	13.2	30.5	35.3	26.5	4.8	TV	MN	69.0	56.9	63.9	63.5	392

Table 6: Results for the **Focal** scenario (means across 50 simulations, sorted by EMD). Best value per metric in **bold**.

Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑	Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑
FLEX-GreedyML	**0.0**	**1.00**	**0.0**	**1.00**	CoSaMP	8.4	0.88	6.4	0.73
Hydra	**0.0**	**1.00**	**0.0**	**1.00**	FLEX-SP	8.7	0.95	8.3	0.80
Subspace-SBL+	**0.0**	**1.00**	**0.0**	**1.00**	EXSO-MUSIC	11.4	0.81	5.2	0.60
Subspace-SBL	**0.0**	**1.00**	**0.0**	**1.00**	LCMV-MVPURE	11.5	0.65	11.4	0.58
Chimera	**0.0**	**1.00**	**0.0**	0.98	APSE	13.2	**1.00**	6.1	0.53
FLEX-SSM	**0.0**	**1.00**	**0.0**	**1.00**	SP	13.3	0.94	11.1	0.77
GreedyML	**0.0**	**1.00**	**0.0**	**1.00**	Gamma-MAP-MSP	18.5	0.29	15.6	0.19
FLEX-AP	**0.0**	**1.00**	**0.0**	**1.00**	Champ.-MacKay-NL-F	18.9	0.80	19.0	0.72
FLEX-MUSIC	**0.0**	**1.00**	**0.0**	**1.00**	MSP	21.4	0.16	20.2	0.12
ExhaustiveML	**0.0**	**1.00**	**0.0**	**1.00**	DICS	21.9	0.80	16.6	-0.01
FLEX-TRAP-MUSIC	**0.0**	**1.00**	**0.0**	**1.00**	ESMV-MVPURE	22.6	0.29	14.8	-0.03
Flex-NL-Champagne	**0.0**	**1.00**	**0.0**	**1.00**	L1/L2	25.3	0.51	25.3	0.45
Flex-NL-Champ.-Omix	**0.0**	**1.00**	**0.0**	**1.00**	MUSIC	31.9	0.18	22.0	0.21
AP	**0.0**	**1.00**	**0.0**	**1.00**	Source-MAP	34.2	0.51	15.4	0.24
SSM	**0.0**	**1.00**	**0.0**	**1.00**	Champ.-Conv.-NL-F	34.7	0.76	31.5	0.61
RAP-MUSIC	**0.0**	**1.00**	**0.0**	**1.00**	Block-Champagne	36.5	0.13	30.1	0.12
Flex-Champagne	**0.0**	**1.00**	**0.0**	**1.00**	MCE	38.5	0.29	34.7	0.25
Omni-Champagne	**0.0**	**1.00**	**0.0**	**1.00**	SSLOFO	45.2	0.41	18.5	0.14
TRAP-MUSIC	**0.0**	**1.00**	**0.0**	**1.00**	CovCNN-KL	45.6	0.41	27.7	0.28
Flex-Champ.-Omix	**0.0**	**1.00**	**0.0**	**1.00**	SR-eLORETA	46.2	0.64	16.2	0.14
IR-ESMV	**0.0**	**1.00**	**0.0**	**1.00**	ReciPSIICOS	48.4	0.72	13.3	0.18
SSP-IR-ESMV	**0.0**	**1.00**	**0.0**	0.99	HOCMCMV	48.9	0.23	22.7	0.09
Safe-Flex-ESMV	**0.0**	**1.00**	0.4	0.94	HOCMV	48.9	0.23	22.7	0.09
ESMV	**0.0**	**1.00**	**0.0**	**1.00**	SubSMP	48.9	0.51	24.9	0.49
Deblur-Flex-ESMV	**0.0**	**1.00**	**0.0**	0.98	Unit-Noise-Gain	51.2	0.38	15.8	0.09
Flex-ESMV	**0.0**	**1.00**	**0.0**	0.98	WNMV	51.2	0.38	15.8	0.09
Adapt-Flex-ESMV	**0.0**	**1.00**	0.5	0.94	MVAB	51.9	0.45	15.0	0.09
OMP	**0.0**	**1.00**	**0.0**	**1.00**	SAM	52.5	0.25	22.9	0.09
Sharp-Flex-ESMV	**0.0**	**1.00**	0.4	0.94	SMV	53.4	0.85	10.9	0.09
EBB	**0.0**	**1.00**	**0.0**	**1.00**	MCMV	56.2	0.06	25.6	0.05
Flex-ESMV-MVPURE	**0.0**	**1.00**	**0.0**	**1.00**	LCMV	57.2	0.18	22.8	0.08
FLEX-OMP	**0.0**	**1.00**	**0.0**	**1.00**	MEM	58.6	0.12	48.2	0.04
ISubSMP	**0.0**	**1.00**	**0.0**	**1.00**	CovCNN	59.9	0.21	39.1	0.10
ECD	**0.0**	**1.00**	**0.0**	**1.00**	GFT-MNE	60.2	0.05	35.0	0.06
Gamma-MAP	0.1	**1.00**	0.6	0.99	EPIFOCUS	60.2	0.11	38.9	0.08
Champ.-Conv.	0.3	**1.00**	0.7	**1.00**	sLORETA	60.4	0.43	21.5	0.08
Champ.-MacKay	0.4	**1.00**	0.9	**1.00**	LAURA	60.6	0.09	41.9	0.07
Champ.-LowSNR	0.4	**1.00**	1.6	0.99	wMNE	60.9	0.15	34.1	0.08
SESAME	0.4	0.97	0.4	0.97	FC (ESINet)	61.7	0.16	32.3	0.07
SSP-ESMV	0.6	0.97	0.5	0.97	cMEM	61.8	0.05	42.4	0.04
NL-Champagne	1.4	**1.00**	4.4	0.99	GFT-L1	63.1	0.02	44.1	0.03
Champ.-MacKay-NL-H	1.8	0.95	4.1	0.92	Source-MAP-MSP	68.7	0.07	50.3	0.06
Champ.-MacKay-NL-P	1.9	0.92	5.4	0.91	TV	69.0	0.04	47.8	0.03
Champ.-MacKay-NL-D	1.9	0.92	5.4	0.91	dSPM	69.5	0.03	42.6	0.04
Champ.-Conv.-NL-H	2.0	0.93	4.9	0.92	LORETA	71.8	0.03	50.9	0.02
Champ.-Conv.-NL-P	2.3	0.92	5.6	0.90	S-MAP	71.8	0.03	50.9	0.02
Champ.-Conv.-NL-D	2.3	0.92	5.6	0.90	APE-SISSY	72.8	0.03	57.6	0.03
MxNE-BCD	2.7	0.94	4.4	0.92	GBF	73.3	0.03	56.9	0.04
BCS	3.7	0.93	7.6	0.91	Backus-Gilbert	74.7	0.01	64.9	-0.00
ReMBo	4.4	0.99	12.5	0.97	eLORETA	74.9	0.03	64.0	0.04
FLEX-ReMBo	4.7	0.99	12.0	0.97	Random Noise	75.3	0.00	59.2	-0.00
VB-SBL	7.4	0.80	6.1	0.75	SISSY	76.4	0.01	63.6	0.03
FLEX-CoSaMP	8.4	0.88	6.4	0.73	MNE	78.0	0.03	63.6	0.04

Table 7: Results for the **Multi** scenario (means across 50 simulations, sorted by EMD). Best value per metric in **bold**.

Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑	Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑
GreedyML	**8.7**	**0.97**	**1.1**	**0.95**	ReMBo	39.8	0.19	32.6	0.25
FLEX-GreedyML	9.1	0.96	1.4	0.94	FLEX-ReMBo	40.1	0.18	33.5	0.25
Hydra	9.1	0.96	1.4	0.94	MUSIC	40.3	0.08	37.2	0.13
ExhaustiveML	11.0	0.87	2.3	0.87	CoSaMP	40.4	0.16	28.4	0.17
Subspace-SBL+	12.0	0.85	4.5	0.85	ESMV-MVPURE	41.1	0.07	30.0	0.02
RAP-MUSIC	12.9	0.85	4.6	0.85	Source-MAP	41.2	0.18	31.8	0.17
SSM	13.9	0.82	5.1	0.86	SR-eLORETA	42.1	0.14	32.0	0.13
Chimera	14.4	0.80	5.6	0.83	FLEX-CoSaMP	42.4	0.12	30.3	0.13
FLEX-MUSIC	14.4	0.81	5.7	0.81	FLEX-OMP	42.7	0.15	27.7	0.16
AP	14.4	0.75	6.4	0.77	Block-Champagne	42.7	0.05	33.8	0.07
FLEX-SSM	14.6	0.81	5.4	0.84	Gamma-MAP-MSP	42.9	0.09	31.5	0.08
Subspace-SBL	14.8	0.79	5.8	0.81	ISubSMP	43.0	0.12	27.6	0.13
Champ.-MacKay-NL-H	14.8	0.80	8.2	0.81	Champ.-Conv.-NL-F	44.2	0.40	30.1	0.38
TRAP-MUSIC	15.0	0.79	6.3	0.82	L1/L2	44.3	0.15	34.9	0.16
FLEX-AP	15.2	0.73	6.6	0.75	ECD	45.1	0.16	32.3	0.25
Champ.-Conv.-NL-H	15.2	0.80	8.2	0.81	Champ.-MacKay-NL-F	45.2	0.31	30.3	0.32
FLEX-TRAP-MUSIC	15.6	0.77	6.6	0.80	HOCMCMV	45.5	0.05	33.9	0.08
Champ.-LowSNR	15.8	0.79	9.9	0.80	HOCMV	45.5	0.05	33.9	0.08
NL-Champagne	15.8	0.74	10.5	0.78	MCE	46.0	0.08	38.4	0.09
BCS	16.0	0.78	8.8	0.79	SSLOFO	46.6	0.14	33.5	0.11
Champ.-MacKay-NL-P	16.2	0.77	9.7	0.78	Unit-Noise-Gain	47.2	0.09	30.9	0.08
Champ.-MacKay-NL-D	16.2	0.77	9.7	0.78	WNMV	47.2	0.09	30.9	0.08
Champ.-Conv.-NL-P	16.5	0.77	9.5	0.78	SAM	47.4	0.05	36.2	0.08
Champ.-Conv.-NL-D	16.5	0.77	9.5	0.78	ReciPSIICOS	47.5	0.15	27.5	0.09
Champ.-Conv.	16.6	0.76	8.8	0.76	MVAB	47.6	0.10	31.0	0.08
Champ.-MacKay	16.6	0.76	8.8	0.76	SMV	48.7	0.20	29.9	0.08
Flex-NL-Champagne	19.2	0.61	9.6	0.64	LCMV	49.5	0.05	34.6	0.07
Flex-NL-Champ.-Omix	20.2	0.60	9.9	0.63	CovCNN-KL	50.1	0.06	41.3	0.09
Omni-Champagne	22.1	0.59	11.4	0.60	MCMV	50.1	0.01	34.8	0.03
EXSO-MUSIC	22.2	0.60	8.9	0.61	sLORETA	50.9	0.12	35.2	0.08
Flex-Champagne	23.6	0.55	11.9	0.57	DICS	51.0	0.07	39.7	0.01
MxNE-BCD	23.9	0.53	16.3	0.58	EPIFOCUS	51.2	0.06	45.2	0.08
Flex-Champ.-Omix	24.0	0.53	13.0	0.55	GFT-MNE	51.3	0.03	39.0	0.06
IR-ESMV	27.1	0.73	11.4	0.60	wMNE	51.8	0.07	42.2	0.08
Safe-Flex-ESMV	27.2	0.73	12.2	0.54	FC (ESINet)	52.2	0.05	42.9	0.07
SSP-IR-ESMV	28.4	0.69	12.7	0.58	CovCNN	52.5	0.05	43.8	0.08
Adapt-Flex-ESMV	28.5	0.68	13.7	0.47	LAURA	52.5	0.04	44.4	0.07
MSP	28.5	0.15	20.6	0.17	SubSMP	52.6	0.09	35.2	0.07
Sharp-Flex-ESMV	28.8	0.69	13.6	0.47	MNE	53.4	0.04	46.1	0.07
ESMV	29.4	0.66	13.2	0.51	GBF	53.5	0.02	43.7	0.05
Flex-ESMV-MVPURE	29.8	0.85	9.1	0.73	GFT-L1	53.7	0.01	42.2	0.03
Deblur-Flex-ESMV	30.1	0.77	11.5	0.64	eLORETA	53.8	0.05	44.0	0.07
Flex-ESMV	30.1	0.77	11.5	0.64	dSPM	56.3	0.01	42.0	0.04
EBB	30.1	0.54	15.4	0.66	TV	56.9	0.02	45.8	0.03
APSE	30.5	0.52	16.4	0.35	LORETA	57.8	0.01	45.1	0.02
OMP	31.2	0.24	22.2	0.29	S-MAP	57.8	0.01	45.1	0.02
VB-SBL	34.3	0.48	21.0	0.47	Random Noise	58.8	0.00	48.3	0.00
SESAME	34.5	0.19	26.6	0.22	cMEM	59.1	0.02	44.4	0.03
Gamma-MAP	35.6	0.32	21.6	0.28	Source-MAP-MSP	59.3	0.03	47.0	0.04
SP	36.4	0.23	25.3	0.22	APE-SISSY	60.9	0.01	52.3	0.03
SSP-ESMV	38.7	0.33	24.4	0.40	Backus-Gilbert	61.8	0.01	55.9	0.00
FLEX-SP	38.8	0.18	28.0	0.17	SISSY	62.4	0.01	52.9	0.02
LCMV-MVPURE	38.8	0.10	30.0	0.13	MEM	76.1	0.03	55.2	-0.03

Table 8: Results for the **Extended** scenario (means across 50 simulations, sorted by EMD). Best value per metric in **bold**.

Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑	Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑
Subspace-SBL+	**22.6**	0.43	16.9	0.46	FLEX-SP	38.1	0.12	32.0	0.02
FLEX-GreedyML	22.9	**0.45**	**13.8**	**0.53**	SP	38.2	0.05	29.7	0.05
Hydra	22.9	**0.45**	**13.8**	**0.53**	LCMV-MVPURE	38.3	0.05	32.3	0.05
FLEX-SSM	26.2	0.40	17.8	0.47	EBB	38.5	0.08	27.6	0.16
Subspace-SBL	26.5	0.41	18.1	0.46	FLEX-CoSaMP	40.1	0.06	32.9	0.05
Chimera	26.6	0.34	19.4	0.41	SESAME	40.2	0.04	31.6	0.04
GreedyML	27.2	0.08	18.8	0.15	ESMV-MVPURE	40.4	0.04	32.8	0.01
Flex-NL-Champ.-Omix	27.2	0.38	20.9	0.41	CoSaMP	41.0	0.04	32.8	0.03
MSP	27.3	0.31	20.8	0.30	ISubSMP	41.1	0.10	31.0	0.04
FLEX-AP	27.6	0.36	19.6	0.42	CovCNN-KL	41.5	0.11	38.2	0.15
Flex-NL-Champagne	27.9	0.38	20.8	0.41	FLEX-OMP	41.6	0.11	30.5	0.04
Champ.-Conv.	27.9	0.10	20.8	0.17	Flex-ESMV-MVPURE	41.9	0.07	27.9	0.11
Champ.-MacKay	27.9	0.09	21.0	0.17	HOCMCMV	42.0	0.08	34.3	0.13
NL-Champagne	27.9	0.12	20.8	0.19	HOCMV	42.0	0.08	34.3	0.13
Gamma-MAP-MSP	27.9	0.40	18.9	0.32	ReMBo	42.2	0.04	36.1	0.06
Champ.-LowSNR	28.1	0.10	21.0	0.17	MCE	42.5	0.05	38.2	0.08
Champ.-MacKay-NL-H	28.2	0.10	21.5	0.17	APE-SISSY	42.6	0.11	39.5	0.18
Champ.-Conv.-NL-H	28.3	0.10	21.4	0.17	L1/L2	42.7	0.06	35.7	0.09
Champ.-MacKay-NL-P	28.4	0.10	22.1	0.17	CovCNN	43.6	0.09	40.8	0.13
Champ.-MacKay-NL-D	28.4	0.10	22.1	0.17	ECD	43.8	0.04	35.9	0.06
ExhaustiveML	28.5	0.07	19.4	0.13	Unit-Noise-Gain	43.8	0.09	33.2	0.14
BCS	28.5	0.11	21.9	0.17	WNMV	43.8	0.09	33.2	0.14
Champ.-Conv.-NL-P	28.7	0.10	22.4	0.17	SAM	43.8	0.07	36.7	0.13
Champ.-Conv.-NL-D	28.7	0.10	22.4	0.17	ReciPSIICOS	43.9	0.09	30.9	0.10
Gamma-MAP	28.8	0.14	21.4	0.16	MVAB	44.0	0.08	33.1	0.13
IR-ESMV	29.0	0.15	25.6	0.21	SMV	45.3	0.11	32.5	0.14
SSP-IR-ESMV	29.3	0.14	26.7	0.20	FLEX-ReMBo	45.5	0.04	38.7	0.04
EXSO-MUSIC	29.3	0.26	23.4	0.33	SISSY	46.0	0.09	41.2	0.15
Flex-Champagne	29.5	0.37	23.6	0.37	MCMV	46.3	0.05	34.4	0.08
MxNE-BCD	29.5	0.09	23.7	0.16	LCMV	46.3	0.07	35.0	0.12
Flex-Champ.-Omix	29.9	0.36	23.3	0.37	SubSMP	46.4	0.10	30.0	0.02
FLEX-MUSIC	30.2	0.35	20.5	0.41	GFT-MNE	46.5	0.11	38.5	0.17
ESMV	30.3	0.14	26.4	0.18	EPIFOCUS	47.1	0.08	45.5	0.14
FLEX-TRAP-MUSIC	30.6	0.35	21.2	0.40	LAURA	47.2	0.08	42.3	0.14
Omni-Champagne	30.7	0.24	23.5	0.29	wMNE	47.4	0.09	42.3	0.14
AP	31.3	0.06	23.6	0.10	sLORETA	47.6	0.10	36.7	0.14
SSM	32.4	0.06	23.5	0.12	Champ.-MacKay-NL-F	48.2	0.05	36.9	0.07
Adapt-Flex-ESMV	34.1	0.22	25.1	0.25	DICS	48.4	0.04	38.8	-0.00
Safe-Flex-ESMV	34.2	0.22	23.9	0.25	Champ.-Conv.-NL-F	48.6	0.06	37.7	0.08
SSP-ESMV	34.3	0.07	29.4	0.11	FC (ESINet)	49.1	0.11	40.3	0.18
Source-MAP	34.3	0.13	33.2	0.20	GFT-L1	50.5	0.09	40.8	0.14
Block-Champagne	35.0	0.25	29.5	0.34	SSLOFO	50.7	0.08	44.5	0.12
Deblur-Flex-ESMV	35.1	0.19	23.8	0.23	dSPM	53.5	0.04	41.5	0.07
Flex-ESMV	35.2	0.22	23.8	0.25	cMEM	54.5	0.06	43.1	0.08
Sharp-Flex-ESMV	35.3	0.20	25.1	0.22	LORETA	56.2	0.07	44.1	0.09
APSE	35.3	0.15	26.5	0.11	S-MAP	56.2	0.07	44.1	0.09
RAP-MUSIC	35.8	0.06	26.7	0.10	Random Noise	56.8	0.03	48.3	0.00
MUSIC	36.1	0.08	38.0	0.14	GBF	57.9	0.05	51.3	0.10
TRAP-MUSIC	36.2	0.06	27.2	0.10	eLORETA	58.5	0.04	56.6	0.06
OMP	36.2	0.04	28.1	0.07	Backus-Gilbert	59.9	0.04	54.8	0.01
Source-MAP-MSP	36.7	0.11	34.0	0.15	MNE	62.1	0.04	55.6	0.06
VB-SBL	37.6	0.15	27.4	0.20	TV	63.9	0.04	54.4	0.08
SR-eLORETA	38.0	0.10	33.4	0.16	MEM	72.6	0.06	54.2	-0.05

Table 9: Results for the **Noisy** scenario (means across 50 simulations, sorted by EMD). Best value per metric in **bold**.

Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑	Solver	EMD (mm)↓	AP↑	MLE (mm)↓	Corr↑
Subspace-SBL	**9.3**	**0.88**	**3.9**	**0.90**	FLEX-CoSaMP	29.5	0.22	22.4	0.19
Chimera	9.7	**0.88**	4.2	**0.90**	MSP	30.1	0.27	24.7	0.23
FLEX-GreedyML	10.5	**0.88**	4.0	**0.90**	SP	30.3	0.21	19.9	0.24
Hydra	10.5	**0.88**	4.0	**0.90**	CoSaMP	31.0	0.19	22.9	0.19
FLEX-SSM	10.9	**0.88**	4.0	**0.90**	ISubSMP	31.9	0.43	21.1	0.20
FLEX-AP	10.9	0.86	4.3	0.87	ESMV-MVPURE	32.3	0.10	23.8	0.03
Subspace-SBL+	11.5	0.87	5.7	0.87	VB-SBL	36.4	0.25	22.4	0.28
FLEX-TRAP-MUSIC	11.7	0.86	5.1	0.87	ReMBo	36.8	0.17	33.9	0.24
FLEX-MUSIC	11.9	0.87	4.8	0.87	MUSIC	37.6	0.07	33.6	0.14
Flex-NL-Champagne	14.4	0.81	8.0	0.80	FLEX-ReMBo	38.1	0.16	37.0	0.22
Flex-NL-Champ.-Omix	14.5	0.81	8.4	0.81	Source-MAP	39.7	0.13	30.7	0.17
Flex-Champagne	16.2	0.80	8.7	0.78	CovCNN-KL	40.7	0.10	31.4	0.13
GreedyML	16.8	0.36	9.1	0.54	SR-eLORETA	43.9	0.10	32.8	0.14
Flex-Champ.-Omix	16.8	0.79	9.0	0.78	SubSMP	49.9	0.27	32.8	0.06
ExhaustiveML	17.0	0.36	9.2	0.54	Block-Champagne	51.2	0.17	38.4	0.17
Omni-Champagne	17.1	0.66	10.1	0.68	SSLOFO	52.0	0.12	29.7	0.11
SSM	17.7	0.34	9.3	0.52	CovCNN	52.1	0.05	40.8	0.09
AP	17.8	0.34	9.6	0.52	DICS	52.2	0.09	38.0	-0.01
TRAP-MUSIC	19.4	0.33	11.4	0.52	HOCMCMV	53.5	0.07	33.1	0.09
RAP-MUSIC	19.5	0.34	11.3	0.52	HOCMV	53.5	0.07	33.1	0.09
Champ.-MacKay	20.1	0.35	12.2	0.49	SAM	53.7	0.05	38.0	0.09
Champ.-Conv.	20.2	0.35	12.3	0.49	ReciPSIICOS	54.2	0.11	28.3	0.07
EXSO-MUSIC	20.2	0.41	14.3	0.41	Unit-Noise-Gain	55.0	0.10	28.4	0.09
Deblur-Flex-ESMV	20.4	0.51	9.9	0.68	WNMV	55.0	0.10	28.4	0.09
Flex-ESMV	20.4	0.67	10.0	0.69	MVAB	55.0	0.11	27.8	0.09
Champ.-MacKay-NL-H	20.8	0.34	16.8	0.47	SMV	55.8	0.16	27.2	0.09
Champ.-LowSNR	20.9	0.36	14.0	0.49	EPIFOCUS	55.9	0.05	46.7	0.09
Champ.-Conv.-NL-H	20.9	0.34	17.1	0.47	GFT-MNE	56.1	0.07	38.9	0.09
Safe-Flex-ESMV	21.1	0.62	15.6	0.56	MCE	56.3	0.02	45.9	0.04
IR-ESMV	21.5	0.40	14.3	0.46	LAURA	56.5	0.05	45.5	0.08
OMP	21.7	0.26	15.4	0.40	L1/L2	56.9	0.06	42.8	0.10
Champ.-MacKay-NL-P	21.8	0.33	17.9	0.46	sLORETA	56.9	0.08	35.0	0.09
Champ.-MacKay-NL-D	21.8	0.33	17.9	0.46	wMNE	57.1	0.07	39.7	0.09
BCS	21.8	0.34	17.3	0.47	MCMV	57.6	0.02	33.3	0.04
SSP-IR-ESMV	21.8	0.39	15.2	0.45	LCMV	57.9	0.04	37.0	0.08
Champ.-Conv.-NL-P	22.1	0.33	18.8	0.46	FC (ESINet)	59.5	0.06	45.7	0.08
Champ.-Conv.-NL-D	22.1	0.33	18.8	0.46	Source-MAP-MSP	59.9	0.03	46.6	0.05
NL-Champagne	22.5	0.28	20.8	0.42	GFT-L1	60.9	0.03	45.5	0.05
SESAME	23.5	0.24	17.0	0.36	Champ.-MacKay-NL-F	61.0	0.07	42.6	0.10
ESMV	24.2	0.37	16.4	0.37	MNE	61.1	0.04	48.4	0.07
SSP-ESMV	25.8	0.29	19.8	0.40	Champ.-Conv.-NL-F	61.7	0.10	42.7	0.13
EBB	25.9	0.28	16.6	0.42	eLORETA	62.0	0.03	49.6	0.06
Adapt-Flex-ESMV	26.2	0.58	17.6	0.44	cMEM	62.7	0.04	45.2	0.05
Flex-ESMV-MVPURE	26.2	0.34	14.7	0.45	GBF	63.4	0.04	50.4	0.07
Sharp-Flex-ESMV	26.3	0.58	17.7	0.44	TV	63.5	0.04	50.9	0.05
APSE	26.5	0.38	17.9	0.29	dSPM	63.7	0.01	48.0	0.04
Gamma-MAP	26.9	0.29	19.4	0.30	LORETA	64.0	0.03	48.0	0.03
Gamma-MAP-MSP	27.5	0.41	21.6	0.26	S-MAP	64.0	0.03	48.0	0.03
MxNE-BCD	28.7	0.29	23.0	0.40	Random Noise	65.3	0.01	52.5	0.00
FLEX-OMP	29.1	0.45	20.4	0.21	APE-SISSY	68.5	0.02	59.8	0.04
ECD	29.1	0.23	20.5	0.37	SISSY	68.6	0.01	59.4	0.03
FLEX-SP	29.2	0.38	21.4	0.13	Backus-Gilbert	69.6	0.01	64.4	0.00
LCMV-MVPURE	29.4	0.16	23.0	0.19	MEM	85.9	0.02	68.6	-0.03

B Complete Solver Reference

The complete package registry is summarized below and provided as a machine-readable CSV in `supplement/solver_reference.csv`.

Table 10: Complete solver reference (Part 1). A machine-readable version with module paths and provenance is provided in `supplement/solver_reference.csv`.

Category	Solver	Canonical id / aliases
Minimum Norm	mne	`mne`; none
	gft-mne	`gft-mne`; none
	wmne	`wmne`; none
	dspm	`dspm`; none
	MCE	`mce`; l1, fista, minimum current estimate
	l1l2	`l1l2`; mce-l1l2
	MxNE-BCD	`mxne-bcd`; mxne, mixed-norm-bcd
	gft-l1	`gft-l1`; gft-minimum-l1, gftmce
	laura	`laura`; laur
	backus-gilbert	`backus-gilbert`; b-g, bg
	s-map	`s-map`; smap
	tv	`tv`; total-variation, graph-tv
	sissy	`sissy`; none
	apesissy	`apesissy`; ape-sissy
	GBF	`gbf`; basisfunctions, basis-functions
	epifocus	`epifocus`; none
LORETA	self-regularized-eloreta	`self-regularized-eloreta`; sr-eloreta, sreloreta, selfregeloreta
	loreta	`loreta`; lor
	sloreta	`sloreta`; slor
	eloreta	`eloreta`; elor
	sslofo	`sslofo`; none
Bayesian / SBL	MacKay-Champagne	`champagne-mackay`; champagne, mackaychampagne, mackay-champagne, mcc
	Convexity-Champagne	`champagne-convexity`; convexitychampagne, convexity-champagne, coc, champagne-em, emchampagne, em-champagne, emc, mm-champagne
	TEM-Champagne	`champagne-tem`; temchampagne, tem-champagne, temc, t-em-champagne
	AR-EM-Champagne	`champagne-ar-em`; aremchampagne, arem-champagne, aremc, ar-em-champagne
	Low-SNR-Champagne	`champagne-low-snr`; lowsnrchampagne, low-snr-champagne, lowsnr-champagne, lsc
	MacKay-Champagne (homoscedastic noise)	`champagne-mackay-nl-homoscedastic`; none
	MacKay-Champagne (diag noise)	`champagne-mackay-nl-diag`; none
	MacKay-Champagne (full noise)	`champagne-mackay-nl-full`; none
	MacKay-Champagne (prec noise)	`champagne-mackay-nl-prec`; none
	Convexity-Champagne (homoscedastic noise)	`champagne-convexity-nl-homoscedastic`; champagne-em-nl-homoscedastic
	Convexity-Champagne (diag noise)	`champagne-convexity-nl-diag`; champagne-em-nl-diag
	Convexity-Champagne (full noise)	`champagne-convexity-nl-full`; champagne-em-nl-full
	Convexity-Champagne (prec noise)	`champagne-convexity-nl-prec`; champagne-em-nl-prec
	nl-champagne	`nl-champagne`; nlchampagne, nl-champagne, nlc
	omni-champagne	`omni-champagne`; omnichampagne, omni-champagne, omni champagne, oc
	flex-champagne	`flex-champagne`; flexchampagne, flex-champagne, fc-champ
	Flex-Champagne (ongoing-mixture tuned)	`flex-champagne-omix`; fc-omix
	flex-nl-champagne	`flex-nl-champagne`; flexnlchampagne, flex-nl-champagne, fnlc
	Flex-NL-Champagne (ongoing-mixture tuned)	`flex-nl-champagne-omix`; fnlc-omix
	Block-Champagne	`block-champagne`; blockchampagne, block-champagne, bc-champ
	gamma-map	`gamma-map`; gmap
	source-map	`source-map`; none
	gamma-map-msp	`gamma-map-msp`; none
	source-map-msp	`source-map-msp`; none
	msp	`msp`; multiple-sparse-priors
	mem	`mem`; none
	cmem	`cmem`; none
	subspace-sbl	`subspace-sbl`; ssm-nlc, subspacesbl
	subspace-sbl-plus	`subspace-sbl-plus`; ssm-nlc-plus, subspacesblplus
	vb-sbl	`vb-sbl`; vbsbl, variational-bayes-sbl
	bcs	`bcs`; none

Table 11: Complete solver reference (Part 2). A machine-readable version with module paths and provenance is provided in `supplement/solver_reference.csv`.

Category	Solver	Canonical id / aliases
Beamformers	mvab	`mvab`; none
	lcmv	`lcmv`; none
	lcmv-mvpure	`lcmv-mvpure`; mvpure-lcmv, mvpure
	dics	`dics`; dics-beamformer
	smv	`smv`; none
	wnmv	`wnmv`; wmnv
	hocmv	`hocmv`; none
	esmv	`esmv`; none
	esmv-mvpure	`esmv-mvpure`; mvpure-esmv
	mcmv	`mcmv`; none
	hocmcmv	`hocmcmv`; none
	ReciPSIICOS	`recipsiicos`; recipsiicos-plain
	recipsiicos-whitened	`recipsiicos-whitened`; none
	sam	`sam`; none
	ebb	`ebb`; empirical-bayesian-beamformer
	adapt-flex-esmv	`adapt-flex-esmv`; adaptflexesmv
	flex-esmv	`flex-esmv`; flexesmv
	flex-esmv-mvpure	`flex-esmv-mvpure`; flexesmvmvpure, flexesmv-mvpure, mvpure-flex-esmv
	deblur-flex-esmv	`deblur-flex-esmv`; deblurflexesmv
	safe-flex-esmv	`safe-flex-esmv`; safeflexesmv
	sharp-flex-esmv	`sharp-flex-esmv`; sharpflexesmv
	ssp-esmv	`ssp-esmv`; sspesmv
	ir-esmv	`ir-esmv`; iresmv
	ssp-ir-esmv	`ssp-ir-esmv`; sspiresmv
	unit-noise-gain	`unit-noise-gain`; ung, unig
MUSIC / Subspace	music	`music`; none
	RAP-MUSIC	`rap-music`; rap
	TRAP-MUSIC	`trap-music`; trap
	FLEX-MUSIC	`flex-music`; flex-rap-music, flex
	FLEX-TRAP-MUSIC	`flex-trap-music`; flextrap
	FLEX-SSM	`flex-ssm`; none
	SSM	`ssm`; none
	FLEX-AP	`flex-ap`; altproj
	AP	`ap`; none
	exso-music	`exso-music`; exsomusic
	ExhaustiveML	`exhaustive-ml`; exhaustiveml, subset-ml
	GreedyML	`greedy-ml`; greedyml
	FLEX-GreedyML	`flex-greedy-ml`; flexgreedyml, flex-greedyml
	FLEX-GreedyML-Diffuse	`flex-greedy-ml-diffuse`; flexgreedyml-diffuse, fgml-diffuse, flex-greedy-diffuse
Matching Pursuit	omp	`omp`; none
	FLEX-OMP	`flex-omp`; flexomp
	cosamp	`cosamp`; none
	FLEX-CoSaMP	`flex-cosamp`; flexcosamp
	rembo	`rembo`; none
	FLEX-ReMBo	`flex-rembo`; flexrembo
	sp	`sp`; none
	FLEX-SP	`flex-sp`; flexsp
	subsmp	`subsmp`; none
	isubsmp	`isubsmp`; none
Dipole Fitting	ecd	`ecd`; equivalent-current-dipole
	sesame	`sesame`; none
Deep Learning	fc	`fc`; fully-connected, fullyconnected, esinet
	covcnn	`covcnn`; cov-cnn, covnet
	covcnn-kl	`covcnn-kl`; cov-cnn-kl
	covcnn-kl-flexomp	`covcnn-kl-flexomp`; cov-cnn-kl-flexomp
	lstm	`lstm`; none
	cnn	`cnn`; none
Hybrid / Baseline	apse	`apse`; none
	MCA Smooth+Sparse (Full)	`mca-smooth-sparse-full`; mca-full, mca-smooth-full, mca-smooth+sparse-full
	Graph Wiener Smooth+Sparse (Adaptive EB)	`graph-wiener-smooth-sparse-adaptive`; gw-ssa, gwssa, wiener-smooth-sparse-adaptive
	GW-SSAR	`graph-wiener-smooth-sparse-adaptive-robust`; gw-ssar, gwssar, wiener-smooth-sparse-adaptive-robust
	GW-SR	`graph-wiener-salient-ridge`; gw-sr, gwsr, wiener-salient-ridge
	Sparse Seeded Support	`sparse-seeded-support`; sss, seeded-support, sparse-seeded
	chimera	`chimera`; none
	Hydra	`hydra`; none
	random-noise	`random-noise`; random, noise-baseline