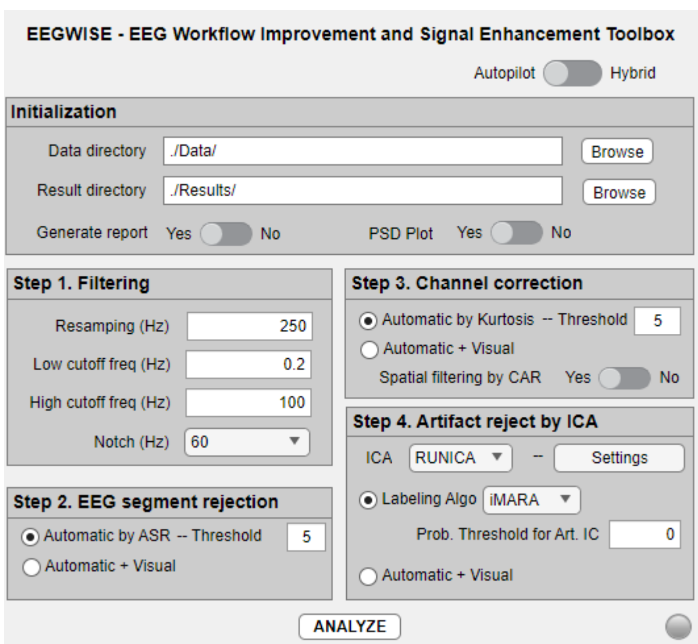
EEGWISE Operational Manual

Table of Contents

EEGWISE Overview....1
Repository....1
Input....2
Step 1: Resampling and Spectral Filtering....3
Step 2: Temporal Filtering....3
Step 3: Spatial Filtering....3
Step 4: Source Domain Rejection....3
Output....4
Questions and feedback....5

EEGWISE Overview

EEG Workflow and Signal Enhancement Toolbox (EEGWISE) features a GUI interface that enhances usability. The pipeline offers two modes: an Autopilot Mode, which provides a fully automated, hands-off experience, and a Hybrid Mode, which allows advanced users to manually adjust parameters for greater control. EEGWISE is built on EEGLAB and operates within the MATLAB environment.



EEGWISE GUI Interface

Repository

https://github.com/ranxiao/EEGWISE.git

Input

EEGWISE accepts data in the standard EEGLAB .set format, providing a single, unified entry point for recordings collected on any acquisition system once they have been converted to .set. When creating the .set file you must include a channel-location file (e.g., .ced, .locs, .sfp), because electrode coordinates are required for both spherical-spline interpolation of bad channels and for accurate spatial reconstruction during Independent Component Analysis (ICA) in EEGWISE pipeline.

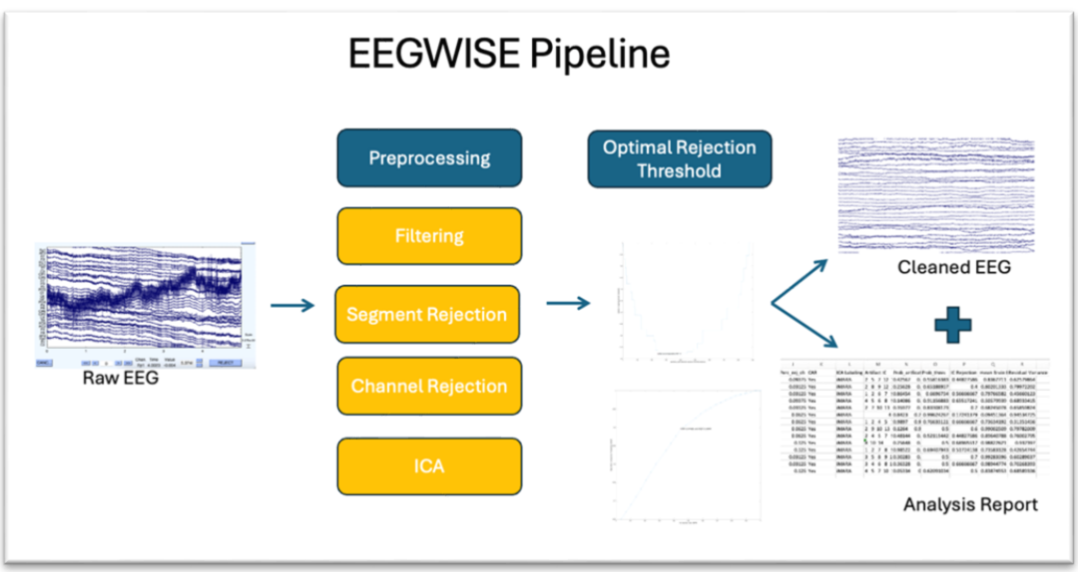
EEGWISE includes two MATLAB utilities, DataPreparation_Biosemi_perSub.m for a single participant and DataPreparation_Biosemi_batchAll.m for an entire cohort, that convert raw BioSemi recordings into EEGLAB .set files ready for analysis. Each script:

(1) reformats the data and attaches the correct channel-location file,

(2) re-references every channel to the T7/T8 average to improve signal-to-noise ratio, as recommended in the Biosemi manual,

(3) groups EEG files by trial type (i.e., baseline, reach, baseline+reach, SATCO, all) for the Reach study, and

(4) extracts absolute time stamps so that the pre-processed EEG can later be time-matched to external events.



The pipeline consists of four major steps, each addressing artifacts from a different perspective.

Step 1: Resampling and Spectral Filtering.

This step allows for both upsampling and downsampling to optimize processing efficiency, with a default sampling rate of 250 Hz. EEGWISE applies low-pass and high-pass filters to remove DC shifts, low-frequency noise, and high-frequency muscle artifacts (default cutoffs: 0.2 Hz for low-pass and 100 Hz for high-pass). Additionally, a notch filter removes powerline interference (default: 60 Hz for the U.S. power grid).

Step 2: Temporal Filtering.

This step employs Artifact Subspace Reconstruction (ASR) to detect and correct transient, high-amplitude artifacts EEG artifacts (Eye blinks, Jaw clenching, Sudden head or cable movements, and muscle bursts). The algorithm segments EEG data into 0.5s segments with overlapping of 50% and compares it to a clean baseline (auto-scans the recording and stitches together ~ 60 s of the cleanest 0.5-s windows to build its reference covariance). For every window, the algorithm projects the data onto the calibration principal-component space and checks whether the variance along any component exceeds n standard deviations above baseline (default n = 5, a conservative cutoff). Windows that violate the threshold are rejected, yielding a cleaned EEG signal with minimal loss of true neural activity.

Step 3: Spatial Filtering.

EEGWISE performs automated channel screening through 1). Channel-wise Kurtosis test. For every channel EEGWISE computes the z-scored kurtosis of its time-series. Channels whose kurtosis exceeds n = 5 SD (user-settable) are flagged as artifactual. 2) Global-variance test. After kurtosis pruning the variance of each remaining channel is compared with the ensemble distribution; channels lying ±3 SD outside the mean (or perfectly flat) are also flagged.

All flagged channels are restored by spherical-spline interpolation from their surrounding electrodes, preserving the original montage. If the Automatic + Visual toggle is on, the cleaned data are displayed in a scrolling viewer, and the user may type additional channel indices to be interpolated.

When enabled, the data are finally re-referenced to the common average reference (CAR) to attenuate spatially widespread noise. Default thresholds: 5 SD for kurtosis, 3 SD for variance; CAR is enabled.

Step 4: Source Domain Rejection.

In this stage EEGWISE decomposes the multichannel EEG into statistically independent components (ICs) and removes those dominated by artefacts. Users may choose either RUNICA (default) or AMICA. EEGWISE automatically constrains the decomposition to the data's effective rank (channels minus previously rejected channels) to ensure numerical stability. Once ICA is complete, each IC is labelled by a machine-learning classifier—iMARA (default), MARA, or ICLabel, which assigns the probability that the component is artifactual (e.g., ocular, muscle, or movement–related).

EEGWISE then determines an optimal probability cutoff automatically. It scans threshold values from 0 to 1 in 0.01 steps, computing for each point (i) the IC-rejection ratio (proportion of ICs that would be discarded) and (ii) the mean brain-component probability of the ICs that would be kept. Both curves are normalized to [0, 1], and the threshold giving the shortest Euclidean distance to the ideal point (0 rejection, 1 brain probability) is selected, with an additional safeguard that the retained ICs' mean brain probability must be ≥ 0.80. Users may override this by entering a fixed threshold in the GUI. All ICs whose artefact probability exceeds the final cutoff, together with any extra ICs the user marks manually, are removed, and the cleaned data are saved.

Output

Analysis report. EEGWISE automatically compiles an Analysis Report that logs every key metrics and parameter settings used during processing, including the ASR and kurtosis thresholds selected, the IC-rejection ratio, the mean brain-component probability of the retained ICs, and the residual variance after artefact removal. Each time the pipeline reaches a checkpoint the current dataset, the associated metrics, and any error messages are appended to this report and written to the results directory, providing a transparent, fully reproducible audit trail that scales to large-cohort, batch analyses.

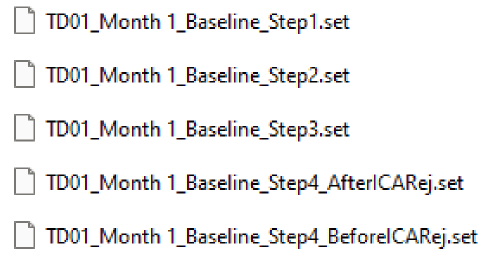
1	File Name	NoChannel	Filter_Lo_Hi_Notch_SampRate	Original Duration	ASR thres	Remaining Duration	Perc_rej_dur	Kurtosis thres	Bad Channels
2	TD01_Month	32	0.2,100,60,250	119.932	5	54.924	0.54204049	5	15 29

'erc_rej_ch	CAR	ICA Algorithm	ICA Parameters	ICA Labeling	Artifact IC	Prob_artficat		Prob_thres	IC Rejection Ratic	mean Brain IC prob	Residual Variance	Error Message
0.0625	Yes	RUNICA	{"lrate":0.00018755	iMARA	1 2 5 6 7 9 1(	0.55012	0.47141	0.46	0.533333333	0.847350517	0.34972316	

Visualization. For every major step, EEGWISE saves figures of the data, including EEG waveforms before and after cleaning, ICA topographies, and the probability-threshold optimization curves used for adaptive IC selection. These plots give users an immediate, visual record of how each operation has affected the signal and allow offline assessment.

- TD01_Month 1_Baseline.set_Step4_RelPower.jpg
- TD01_Month 1_Baseline_Step1.jpg
- TD01_Month 1_Baseline_Step2.jpg
- TD01_Month 1_Baseline_Step3.jpg
- TD01_Month 1_Baseline_Step4_AfterICARej.jpg
- TD01_Month 1_Baseline_Step4_ICATopo.jpg
- TD01_Month 1_Baseline_Step4_OptimalThres.jpg

Intermediate results. In addition to the final cleaned dataset, EEGWISE stores a complete series of intermediate .set files after each processing phase. Investigators can reload any of these checkpoints to re-run specific steps, perform alternative analyses, or verify the integrity of earlier processing decisions without repeating the entire pipeline.



Questions and feedback

All questions and feedback can be direct to Ran Xiao at ran.xiao@emory.edu