

Robust fuzzy clustering for high-dimensional multivariate time series with outlier detection

Ziling Ma ¹, Ángel López-Oriona ¹, Hernando Ombao ¹, Ying Sun ¹

SUMMARY: Fuzzy clustering provides a natural framework for modeling partial memberships, particularly important in multivariate time series (MTS) where state boundaries are often ambiguous. For example, in EEG monitoring of driver alertness, neural activity evolves along a continuum (from unconscious to fully alert, with many intermediate levels of drowsiness) so crisp labels are unrealistic and partial memberships are essential. However, most existing algorithms are developed for static, low-dimensional data and struggle with temporal dependence, unequal sequence lengths, high dimensionality, and contamination by noise or artifacts. To address these challenges, we introduce RFCPCA, a robust fuzzy subspace-clustering method explicitly tailored to MTS that, to the best of our knowledge, is the first of its kind to simultaneously: (i) learn membership-informed subspaces, (ii) accommodate unequal lengths and moderately high dimensions, (iii) achieve robustness through trimming, exponential reweighting, and a dedicated noise cluster, and (iv) automatically select all required hyperparameters. These components enable RFCPCA to capture latent temporal structure, provide calibrated membership uncertainty, and flag series-level outliers while remaining stable under contamination. On driver drowsiness EEG, RFCPCA improves clustering accuracy over related methods and yields a more reliable characterization of uncertainty and outlier structure in MTS.

KEY WORDS: fuzzy clustering; multivariate time series; dimensionality reduction; robust statistics; anomaly detection; neuroscience

¹King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division. Thuwal 23955-6900, Saudi Arabia. Correspondence: ziling.ma@kaust.edu.sa, angel.lopezoriona@kaust.edu.sa, hernando.ombao@kaust.edu.sa, ying.sun@kaust.edu.sa

1 Introduction

Time series data are ubiquitous across finance, medicine, engineering, and neuroscience, and modern acquisition technologies now generate massive repositories of recordings. These datasets are often heterogeneous in length, high-dimensional, and exhibit evolving temporal dynamics, making traditional mining tasks such as forecasting, classification, anomaly detection, and clustering challenging to scale and interpret (Esling and Agon, 2012; Ghassempour et al., 2014). The challenge is even greater for multivariate time series (MTS), where complex inter-channel dependencies must be preserved to meaningfully characterize system dynamics.

Clustering is a central tool for exploring and simplifying such data. By grouping MTS with similar temporal structures and cross-channel interactions (Aghabozorgi et al., 2015), clustering reduces data complexity, supports anomaly detection through deviations from cluster prototypes (Li et al., 2021; Shi et al., 2023), and facilitates predictive modeling tailored to homogeneous subgroups (Kavitha and Punithavalli, 2010; Jha et al., 2015). However, in real-world MTS applications such as EEG or sensor data, clustering performance is often undermined by noise, artifacts, and outliers. High dimensionality further exacerbates the problem: distance measures lose discriminative power due to the curse of dimensionality, and even mild contamination can distort cluster assignments (Rousseeuw and van Zomeren, 1991; Aggarwal and Yu, 2001; García-Escudero et al., 2010). Although robust clustering methods exist (Davé and Krishnapuram, 2002), they are typically computationally expensive and not well suited for large-scale or high-dimensional MTS.

Traditional clustering methods follow a hard (crisp) paradigm, where each element is assigned to exactly one group, whereas fuzzy clustering techniques offer a more flexible alternative. Many real datasets contain inherently ambiguous boundaries, making partial memberships more realistic (Hwang et al., 2007; Kruse et al., 2007). Beyond interpretability, fuzzy clustering also provides algorithmic advantages: (i) smoother membership updates

reduce sensitivity to poor initialization and local minima (Heiser and Groenen, 1997; Jayaram and Klawonn, 2013), (ii) optimization tends to be faster and more stable (McBratney and Moore, 1985; Suganya and Shanthi, 2012), and (iii) secondary memberships offer valuable diagnostic insights, highlighting borderline or anomalous samples that hard clustering ignores (Duran and Odell, 2013; D’Urso and De Giovanni, 2014; Ruspini et al., 2019). Despite these advantages, existing fuzzy clustering techniques rarely combine robustness, scalability, and the ability to handle high-dimensional MTS, leaving an important research gap.

In the previous literature on fuzzy clustering of MTS, D’Urso and Maharaj (2012) introduced a wavelet-based method that groups series by similarities in their time–frequency patterns. Izakian et al. (2015) extended fuzzy clustering to MTS using dynamic time warping (DTW) distances, and D’Urso et al. (2021) later proposed a robust trimmed version of DTW-based clustering. Li and Wei (2020) presented a fuzzy clustering approach with feature weights, emphasizing informative variables while down-weighting irrelevant ones. López-Oriona et al. (2022) proposed feature-based robust approaches for fuzzy clustering of MTS based on quantile cross-spectral density, enabling adaptive weighting of variables and distances. D’Urso et al. (2022) developed a spatially robust fuzzy clustering framework for COVID-19 time series, where B-spline smoothing captures temporal trajectories and spatial dependence enhances clustering performance while reducing noise and outlier effects. D’Urso et al. (2023) proposed a fuzzy c-medoids method that integrates DTW with an exponential transformation and entropy regularization, improving robustness to outliers and accommodating unequal-length series. D’Urso and Leski (2023) introduced a robust fuzzy clustering model for MTS based on a Student-t mixture, which down-weights the influence of outliers during clustering. Most recently, Ma et al. (2025) developed FuzzCoh, a robust spectral fuzzy clustering method using Kendall’s tau canonical coherence to capture frequency-domain connectivity, offering resilience to artifacts and strong interpretability for

brain data. Nevertheless, existing fuzzy clustering methods for MTS remain computationally demanding, insufficiently robust to outliers, and often unable to handle unequal-length, high-dimensional series.

To address these challenges, we propose RFCPCA, a robust fuzzy clustering framework grounded in common principal component analysis (CPCA). RFCPCA advances the state of the art in four key aspects:

- (1) It learns membership-weighted, cluster-specific low-rank subspaces via CPCA, linking soft memberships to the estimated bases while preserving cross-channel structure.
- (2) It naturally accommodates unequal-length and moderately high-dimensional series, avoiding restrictive preprocessing steps.
- (3) It integrates three complementary robustness strategies—exponential metric, trimming, and a noise cluster—that detect and mitigate outliers without sacrificing scalability.
- (4) It supports automatic hyperparameter selection, eliminating the need for ad hoc tuning.

In summary, RFCPCA is, to our knowledge, the first subspace-based fuzzy clustering framework for MTS that jointly achieves robustness, interpretability, and computational feasibility.

The remainder of the paper is organized as follows. Section 2 introduces the RFCPCA methodology. Section 3 evaluates its robustness on simulated contaminated EEG data. Section 4 investigates driver-drowsiness EEG data to assess how the learned structure aligns with levels of alertness and quantifies gains in accuracy, including principled series-level outlier handling. Finally, Section 5 concludes the paper.

2 RFCPCA: the robust fuzzy clustering method based on CPCA

Li (2019) proposed Mc2PCA, a method for MTS clustering based on CPCA. This approach was subsequently improved by Ma et al. (2024), who introduced ROBCPCA, a robust version of Mc2PCA that additionally accounts for the serial dependence inherent in MTS. Building

on the ROBPCPA framework, Ma et al. (2025) developed FCPCA, a fuzzy clustering method grounded in CPCA. Their work demonstrated that FCPCA offers advantages in terms of computational efficiency over ROBPCPA, as well as improved clustering accuracy compared to both ROBPCPA and Mc2PCA. In what follows, we first provide an overview of the FCPCA method before presenting our proposed robust framework, RFCPCA, which enhances FCPCA by effectively handling outliers in MTS fuzzy clustering.

2.1 Brief overview of FCPCA method

Consider a dataset $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$ consisting of N MTS, where each $\mathbf{X}_i \in \mathbb{R}^{T_i \times p}$ represents the i -th MTS object with length T_i and dimensionality (channels) p . Each series is assumed second-order stationary: it has a constant mean and an autocovariance that depends only on the lag $\ell \in \mathbb{Z}$. In line with Ma et al. (2025), we restrict attention to nonnegative lags up to $L = 2$ (i.e., $\ell \in \{0, 1, 2\}$).

The primary aim of FCPCA is to cluster the MTS dataset into S fuzzy clusters, where each MTS object can have partial memberships across clusters. Initially, fuzzy membership degrees u_{is} (for $i \in \{1, \dots, N\}$ and $s \in \{1, \dots, S\}$), representing the membership of the i -th object to cluster s , are randomly initialized, ensuring each object's memberships sum to one across all clusters, i.e., $\sum_{s=1}^S u_{is} = 1$.

FCPCA first computes cross-covariance matrices at each lag for every series. For each MTS $\mathbf{X}_i$, the lagged cross-covariance matrices $\hat{\Gamma}_i(l)$ are estimated, and corresponding block covariance matrices are constructed:

$$\hat{\mathbf{\Gamma}}_i(l) = \begin{pmatrix} \hat{\Gamma}_i(0) & \hat{\Gamma}_i(l) \\ \hat{\Gamma}_i(l)^\top & \hat{\Gamma}_i(0) \end{pmatrix}, \quad l = 1, \dots, L.$$

Subsequently, common weighted covariance matrices for each cluster s at lag l are calculated using fuzzy memberships:

$$\tilde{\Sigma}_s(l) = \frac{\sum_{i=1}^N u_{is}^m \hat{\Gamma}_i(l)}{\sum_{i=1}^N u_{is}^m}, \quad s = 1, \dots, S,$$

where $m > 1$ controls the fuzziness of the clustering, with a higher value of m gives more fuzziness.

Singular value decomposition (SVD) of each $\tilde{\mathbf{\Sigma}}_s(l)$ produces eigenvectors, from which the common projection axes $\mathbf{C}_s(l)$ are derived by selecting principal components that capture a substantial portion (e.g., 95%) of the variance.

Each object $\hat{\mathbf{X}}_i(l)$ is formed by concatenating two lagged segments of the original MTS $\mathbf{X}_i$: the past segment $\mathbf{X}_{i,t-l}^*$ and the future segment $\mathbf{X}_{i,t}^*$. This construction ensures conformable dimensions for projection. The resulting $\hat{\mathbf{X}}_i(l)$ is then projected onto, and reconstructed from, the subspace of each cluster:

$$\mathbf{Y}_i^s(l) = \hat{\mathbf{X}}_i(l) \mathbf{C}_s(l) \mathbf{C}_s(l)^\top, \quad s \in \{1, \dots, S\}.$$

The reconstruction error, squared Frobenius distance, for each series in cluster s across all lags is:

$$r_{is}^2 = \sum_{l=1}^L \left\| \hat{\mathbf{X}}_i(l) - \mathbf{Y}_i^s(l) \right\|_F^2.$$

The primary objective of FCPCA is to find the $N \times S$ matrix of fuzzy membership, $\mathbf{U} = (u_{is})$, and sets of common projection axes, $\mathbf{C}(l) = \{\mathbf{C}_1(l), \dots, \mathbf{C}_S(l)\}$ that minimize the total weighted reconstruction error:

$$\begin{cases} \min_{\mathbf{U}, \mathbf{C}(l)} & \sum_{i=1}^N \sum_{s=1}^S u_{is}^m r_{is}^2 \\ \text{subject to} & \sum_{s=1}^S u_{is} = 1 \quad \text{and} \quad u_{is} \geq 0. \end{cases} \quad (1)$$

The optimization problem is solved iteratively, and the membership degrees are updated based on reconstruction errors as:

$$u_{is} = \left[\sum_{s^*=1}^S \left(\frac{r_{is}^2}{r_{is^*}^2} \right)^{\frac{1}{m-1}} \right]^{-1}.$$

The algorithm alternates between updating fuzzy memberships and updating the common

projection axes, and stops when the total reconstruction error converges (absolute change $< 10^{-3}$) or when 1000 iterations are reached.

Upon convergence, we obtain the membership matrix $\mathbf{U}$, the common projection axes $\mathcal{C}(l)$, and also the cluster prototype $\mathcal{P} = \{P_s(l)\}$, $s \in \{1, \dots, S\}$, $l \in \{1, \dots, L\}$, where $P_s(l) = \mathbf{C}_s(l) \mathbf{C}_s(l)^\top$.

2.2 Robust metric approach: RFCPCA-E

In FCPCA, the squared reconstruction error determines the membership degree of each series, with larger errors reducing its weight in the assigned cluster. However, due to the quadratic loss, the influence of large errors remains unbounded (Rousseeuw, 1984). Even with slightly reduced membership, a large squared error resulted by an outlier can dominate the objective and distort the common subspace. To address this, we replace the squared loss with an exponential loss:

$$\begin{cases} \min_{\mathbf{U}, \mathcal{C}(l)} & \sum_{i=1}^N \sum_{s=1}^S u_{is}^m [1 - \exp(-\beta r_{is}^2)] \\ \text{subject to} & \sum_{s=1}^S u_{is} = 1, u_{is} \geq 0, i = 1, \dots, N, \end{cases} \quad (2)$$

where $\beta > 0$ is a constant scale parameter controls how aggressively large reconstruction errors are downweighted in the exponential loss. Empirically, it can be selected as (D’Urso et al., 2015; López-Oriona et al., 2022):

$$\beta = \left(\frac{1}{N} \sum_{i=1}^N \min_{1 \leq s \leq S} r_{is}^2 \right)^{-1}. \quad (3)$$

The exponential distance is inherently more robust than the Euclidean distance, as it adaptively downweights outliers and gives greater weight to points near the bulk of the data (Wu and Yang, 2002; Lafuente-Rego et al., 2020).

The local optimal membership update for (2) is given as follows:

$$u_{is} = \left[\sum_{s'=1}^S \left(\frac{1 - \exp(-\beta r_{is}^2)}{1 - \exp(-\beta r_{is'}^2)} \right)^{\frac{1}{m-1}} \right]^{-1}. \quad (4)$$

Since $x \mapsto 1 - \exp(-\beta x)$ saturates at 1 as $x \rightarrow \infty$, outliers have bounded influence, unlike the unbounded penalty of the classical ℓ_2 loss. When an MTS is highly outlying, the exponential loss yields nearly equal costs across clusters, resulting in evenly distributed memberships.

2.3 Noise cluster approach: RFCPCA-N

The noise cluster approach introduces an additional cluster dedicated to outliers. In this framework, the first $S - 1$ clusters capture the genuine structure in the data, while cluster S serves as a noise cluster characterized by a fixed noise distance $\delta > 0$. The parameter δ defines a dissimilarity threshold: objects whose minimum reconstruction error across all regular clusters exceeds δ are considered too distant and therefore receive higher membership in the noise cluster. Conceptually, δ delineates the boundary between informative observations and potential outliers (Cimino et al., 2004).

This mechanism enables the algorithm to assign atypical series to the noise cluster rather than forcing them into an inappropriate group, thereby preventing outliers from distorting the estimation of cluster-specific subspaces.

The optimization problem is formulated as

$$\begin{cases} \min_{\mathbf{U}, \mathbf{C}^{(l)}} & \sum_{i=1}^N \sum_{s=1}^{S-1} u_{is}^m r_{is}^2 + \sum_{i=1}^N \delta^2 \left(1 - \sum_{s=1}^{S-1} u_{is} \right)^m \\ \text{subject to} & \sum_{s=1}^S u_{is} = 1, \quad u_{is} \geq 0, \quad i = 1, \dots, N. \end{cases} \quad (5)$$

The membership update for the regular clusters is

$$u_{is} = \left[\sum_{s'=1}^{S-1} \left(\frac{r_{is}^2}{r_{is'}^2} \right)^{\frac{1}{m-1}} + \left(\frac{r_{is}^2}{\delta^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \quad s = 1, \dots, S - 1. \quad (6)$$

The noise-cluster membership is then

$$u_{iS} = 1 - \sum_{s=1}^{S-1} u_{is}. \quad (7)$$

Suggested by Dave (1991), the noise distance δ is updated as

$$\delta^2 = \frac{\lambda}{N(S-1)} \sum_{i=1}^N \sum_{s=1}^{S-1} r_{is}^2, \quad (8)$$

where λ is a scale multiplier. Following Dave (1991); Rehm et al. (2007), λ can be selected by iteratively reducing its value and monitoring the proportion of objects assigned to the noise cluster. This proportion typically remains stable until λ becomes so small that genuine cluster members start being absorbed into the noise cluster, producing a distinct “elbow” in the curve. The value of λ at this elbow is taken as the optimal noise-distance threshold. However, as noted in Dave (1991); D’Urso et al. (2021), the results are generally not very sensitive to the choice of λ . Fig. 1 shows an example of the selection of λ . Starting from $\lambda = 1$, no MTS objects are identified as outliers. As λ decreases, the outlier proportion remains at 0% until about $\lambda = 0.0156$, where it increases to roughly 20% and then stabilizes. When λ is further reduced to approximately 0.0004, the proportion of outliers rises sharply to nearly 100%. The optimal λ is chosen at the elbow point, just before this abrupt increase, i.e., $\lambda \approx 0.0009$.

2.4 Trimming approach: RFCPCA-T

The trimming approach is a classical robust strategy designed to mitigate the influence of outliers. Instead of incorporating all series into the estimation, a fixed proportion of the objects with the largest errors is excluded from contributing to the update of the cluster subspaces. This ensures that the estimated projection axes and memberships are driven by the bulk of the data, rather than being distorted by a small number of atypical series.

Fix a trimming proportion $\alpha \in (0, 1)$ and let $H = \lfloor N(1 - \alpha) \rfloor$. Denote by $Y \subset \{1, \dots, N\}$ the index set of the H retained series, so that only a subset of size H of the N available

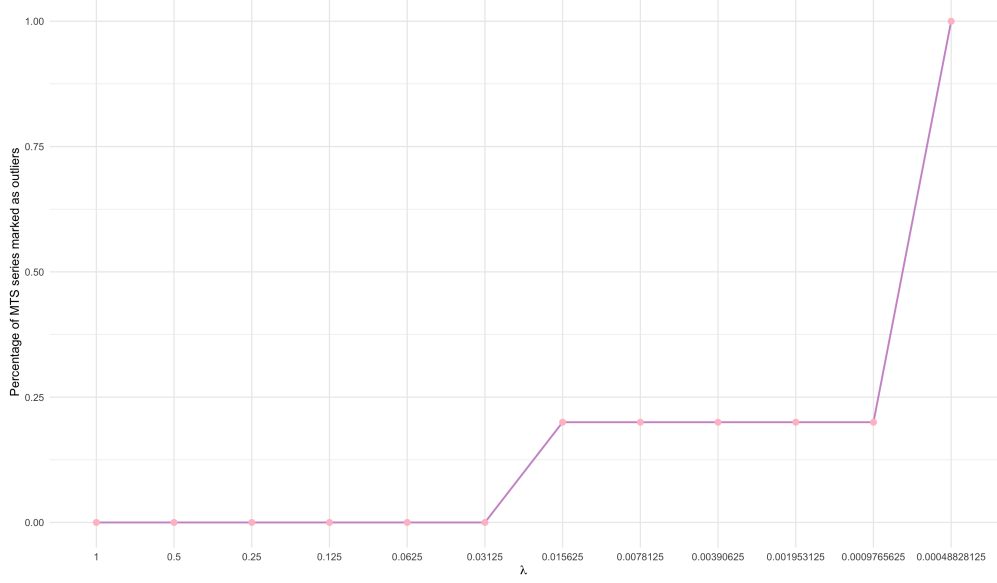


Figure 1: Example of outlier proportion versus λ , with the elbow point indicating the selected parameter.

MTS objects is retained in the optimization. The objective function is then given by

$$\left\{ \begin{array}{l} \min_{\mathbf{U}, \mathbf{C}(\ell), Y: |Y|=H} \sum_{i \in Y} \sum_{s=1}^S u_{is}^m r_{is}^2 \\ \text{subject to} \quad \sum_{s=1}^S u_{is} = 1, \quad u_{is} \geq 0, \quad i = 1, \dots, N. \end{array} \right. \quad (9)$$

The trimming set Y is selected adaptively: at each iteration, the H objects with the smallest reconstruction errors are retained, while the remaining $N - N$ series are excluded from updating the centroids and projection axes. In this way, the algorithm focuses on the bulk of the data while discarding potentially influential outliers.

When $\alpha = 0$, all series are retained and RFCPCA-T reduces to the standard FCPCA method. For $\alpha > 0$, the effective sample size is reduced to H , which increases robustness at the expense of efficiency.

2.5 Methodology comparison

The three above methods are designed to address the sensitivity of FCPCA to outliers in MTS. Each method operates through a distinct mechanism and presents unique advantages and disadvantages.

RFCPCA-E replaces the conventional L_2 norm with an exponential distance that progressively down-weights large errors (Banerjee and Dave, 2012). This soft penalization reduces the influence of outliers while retaining all data, ensuring that every series contributes to clustering. It offers continuous and adaptive robustness without requiring trimming rates or extra clusters, thereby preserving model simplicity. However, extremely deviant outliers may still exert residual influence, explicit outlier labeling is not available, and performance depends on careful tuning of the scale parameter.

RFCPCA-N introduces an additional cluster dedicated to noise or outliers (Rehm et al., 2007). All data are retained, but poorly fitting series are given higher membership in the noise cluster, allowing explicit identification of anomalies. This is particularly valuable for applications such as fault detection, neuroscience, finance, etc., where outlier behavior is informative. However, the approach increases model complexity and parameter tuning requirements, and borderline cases may receive ambiguous memberships, complicating interpretation.

RFCPCA-T excludes a fixed proportion of the most outlying series based on reconstruction errors (García-Escudero et al., 2008; García-Escudero and Mayo-Iscar, 2024). By removing these points entirely, the estimation of cluster parameters becomes more stable and resistant to extreme contamination. Trimming thus provides strong robustness and computational efficiency through reduced sample size, with clear separation between retained and discarded data. The drawback is that trimmed series are permanently excluded, which risks discarding meaningful signals, and the method does not adapt to varying levels of outlyingness.

In summary, trimming is most effective when extreme outliers must be eliminated, the noise cluster is preferable when retaining and labeling anomalies is important, and the robust metric is best suited when continuous robustness is desired without discarding data or modifying the clustering structure.

2.6 Selection of hyperparameters

To run the RFCPCA algorithm, the number of clusters S and the fuzziness parameter m must be specified (and, for the trimmed variant RFCPCA-T, the trimming proportion α). Because different hyperparameter choices can yield markedly different partitions and are rarely known in advance (Arbelaitz et al., 2013), a validation criterion is required to identify the optimal clustering (Kwon, 1998; Wang and Zhang, 2007; Wang et al., 2022).

Following Ma et al. (2025), we consider the generalization of the Xie-Beni index (Xie and Beni, 1991) as it directly uses the total weighted reconstruction error that FCPCA and RFCPCA present when the computation ends.

Denote by φ the loss function corresponding to the clustering method under evaluation (FCPCA, RFCPCA-E, RFCPCA-N, or RFCPCA-T). We define

$$J(\mathbf{U}, \mathbf{C}(l); \varphi) = \sum_{i=1}^N \sum_{s=1}^S u_{is}^m \varphi(r_{is}^2), \quad (10)$$

to provide a unified formulation and avoid redundancy.

Let

$$d_{\min} = \min_{s \neq s'} \sum_{l=1}^L \left\| \mathbf{P}_s(l) - \mathbf{P}_{s'}(l) \right\|_F^2, \quad s, s' \in \{1, \dots, S\}, \quad (11)$$

denote the minimum distance between the subspace prototypes of any two distinct clusters.

We then define the cluster validity index (CVI) as

$$\text{CVI}(\varphi) = \frac{J(\mathbf{U}, \mathbf{C}(l); \varphi)}{N d_{\min}}. \quad (12)$$

For every candidate tuple (S, m) (or triplet (S, m, α) in the case of RFCPCA-T), we run the algorithm until convergence and record $\text{CVI}(\varphi)$. The numerator measures the within-cluster reconstruction loss under the given method, while the denominator quantifies the separation between clusters at the end of the clustering process. Hence, lower values of $\text{CVI}(\varphi)$ are preferred.

3 Numerical results

We focus on the simulated contaminated EEG data to evaluate the robustness of RFCPCA. For completeness and reproducibility, we first specify the clean EEG generator and then the artifact models. Our method is also feasible

3.1 Clean EEG generator (oscillatory mixing)

Let $N = 20$ trials (10 per group), each an $n \times p$ matrix with sampling rate $f_s = 100$ Hz. For each trial we generate five latent AR(2) processes corresponding to five frequency bands $\{\delta, \theta, \alpha, \beta, \gamma\}$. For band b with peak frequency f_b and sharpness κ_b , the AR(2) coefficients are

$$\phi_1^{(b)} = \frac{2}{M_b} \cos\left(2\pi \frac{f_b}{f_s}\right), \quad \phi_2^{(b)} = -\frac{1}{M_b^2}, \quad M_b = \exp(\kappa_b),$$

with $(f_\delta, f_\theta, f_\alpha, f_\beta, f_\gamma) = (2, 6, 10, 22.5, 37.5)$ Hz and $(\kappa_\delta, \kappa_\theta, \kappa_\alpha, \kappa_\beta, \kappa_\gamma) = (0.05, 0.05, 0.05, 0.08, 0.10)$.

Each latent is standardized to unit variance and lightly band-pass filtered to its band. Let $Z(t) \in \mathbb{R}^{1 \times 5}$ stack the five latents at time t and $A \in \mathbb{R}^{5 \times p}$ be a nonnegative mixing matrix whose columns sum to one. Group 1 channels are dominated by $\{\delta, \theta, \alpha, \gamma\}$. Group 2 channels are dominated by $\{\theta, \beta\}$. Thus, the two groups are inherently fuzzy. Thus each trial is formed as

$$X(t) = Z(t) A \in \mathbb{R}^{1 \times p}, \quad t = 1, \dots, T.$$

3.2 Comparison methods and evaluation metrics

In this section, we present the baseline methods used for comparison and describe the evaluation metrics employed to assess clustering performance in the simulation studies.

3.2.1 Alternative methods

We compare RFCPCA with three widely used baselines for multivariate MTS as follows and we also evaluate their robust variants.

- *WWW* by Wang et al. (2007). WWW summarizes each univariate channel with a fixed-length vector of structure-based statistical features. The per-channel feature vectors are concatenated across channels and then clustered with standard fuzzy C -means.
- *QCD* by López-Oriona and Vilar (2021). Each MTS is summarized by quantile cross-spectral density estimates over a grid of frequencies and quantile levels. The resulting complex-valued summaries are split into real and imaginary parts and concatenated into a single feature vector per series. Clustering is then performed with fuzzy C -means in this feature space—often after a PCA step for decorrelation/denoising—using Euclidean distance to centroids. This representation captures serial and cross-component dependence and is robust to heavy tails and extremes.
- *MODWT* by D’Urso and Maharaj (2012). MODWT decomposes each channel via the MODWT and uses scale-specific wavelet variances (per channel) and wavelet correlations (across channel pairs) as features. The concatenated feature vector summarizes variability and inter-channel interaction across scales, and can be clustered with fuzzy C -means. MODWT retains time alignment, works for mean-nonstationary series, and makes inter-component relationships explicit.

The hyperparameters, namely β and δ , in the alternative methods are selected analogously to those in RFCPCA.

3.2.2 Evaluation metrics

For each scenario, we report (i) clustering accuracy on the non-outlier series (Acc) and (ii) the percentage of true outliers correctly detected (Out%). For RFCPCA-T we also report the average trimming ratio selected by the cluster validity index (CVI).

To compare with ground-truth hard labels, fuzzy memberships are converted to hard labels using the threshold 0.70 (Maharaj and D’Urso, 2011; López-Oriona and Vilar, 2021; Ma et al.,

2025). For $S = 2$,

$$\text{label}_i = \arg \max_{k \in \{1,2\}} u_{ik} \quad \text{if} \quad \max\{u_{i1}, u_{i2}\} \geq 0.70.$$

For each method, we present the outlier handling rules as follows.

- *Exponential metric (E)*. An observation is flagged as an outlier if it has no dominant membership across the two substantive clusters, i.e., $\max\{u_{i1}, u_{i2}\} < 0.70$. If this occurs for a non-outlier series, it is counted as a false positive for outlier detection.
- *Noise cluster (N)*. An observation is flagged as an outlier if its noise membership exceeds 0.50, i.e., $u_{i,S} \geq 0.50$. For observations not flagged as outliers, memberships are renormalized over the non-noise clusters,

$$\tilde{u}_{is} = \frac{u_{is}}{\sum_{s' \neq S} u_{is'}},$$

and the 0.70 threshold is then applied to $\{\tilde{u}_{ik}\}$. If no cluster attains the 0.70 threshold, the observation is counted as misclassified.

- *Trimming (T)*. An observation is flagged as an outlier if it is in the trimmed set selected by the cluster validity index (CVI). Clustering accuracy is then computed on the remaining (non-trimmed) series.

Metrics. Let $\mathcal{O}$ be the set of injected (true) outliers and $\hat{\mathcal{O}}$ the set flagged by a given method. We report

$$\text{Out\%} = 100 \times \frac{|\hat{\mathcal{O}} \cap \mathcal{O}|}{|\mathcal{O}|},$$

i.e., the true-positive rate (recall) for outlier detection. Clustering accuracy is computed on the remaining series $\mathcal{I} = \{1, \dots, N\} \setminus \hat{\mathcal{O}}$ by the Rand index (RI) between ground-truth labels $\{y_i\}_{i \in \mathcal{I}}$ and the hard labels obtained as above, and reported as Acc.

3.3 Simulated EEG data with transient electromyographic bursts

We evaluate on datasets with trial-level contamination. In awake EEG, scalp electromyographic (EMG) activity is a common artifact: it appears as brief, burst-like events with broadband spectra overlapping the β/γ range (overall $\sim 20\text{--}300$ Hz, with substantial power in $20\text{--}80$ Hz). These bursts are typically spatially localized (strongest at peripheral electrodes) yet can project more broadly across the montage; their amplitudes can exceed neural high-frequency EEG by orders of magnitude, yielding heavy-tailed residuals under Gaussian error models (Goncharova et al., 2003; Whitham et al., 2007). This setting stresses clustering methods and is well suited for assessing robustness.

In each group, we contaminate a fixed proportion $\rho = 0.20$ of trials at the trial level: specifically, $k_{\text{pg}} = \lceil \rho N_{\text{per group}} \rceil$ trials are selected uniformly at random without replacement and modified. All remaining trials stay clean.

For each contaminated trial, draw $n_{\text{burst}} \in \{1, 2, 3\}$. For each burst, sample: a start time $t_0 \in \{1, \dots, T_i - \tau\}$, a duration $\tau = \lfloor 0.25 f_s \rfloor$ (250 ms at sampling rate f_s), a center frequency $f^* \sim \mathcal{U}(30, 80)$ Hz, and a channel set $\mathcal{I} \subset \{1, \dots, p\}$ with $|\mathcal{I}| = \lceil 0.10 p \rceil$. With the Hann envelope $h_q = \frac{1}{2} \left(1 - \cos(2\pi q / (\tau - 1)) \right)$ for $q = 0, \dots, \tau - 1$, define

$$g_q = \sin\left(2\pi f^* \frac{q}{f_s}\right) h_q, \quad q = 0, \dots, \tau - 1.$$

Let $\hat{\sigma}_j$ denote the empirical standard deviation of channel j in that trial and set an amplitude multiplier $\eta = 5$. Update, for all $q = 0, \dots, \tau - 1$ and $j \in \mathcal{I}$,

$$X_{t_0+q,j} \leftarrow X_{t_0+q,j} + \eta \hat{\sigma}_j g_q.$$

Trials not selected remain clean.

Fig. 2 illustrates the simulated EEG data before and after the introduction of burst contamination. In total, 50 replications are used to obtain the results. Fig. 3 reports the average runtime of each method, where the noise-based variant is used as a reference since

robust approaches generally require more computation. The runtime of QCD at $p = 128$ is omitted due to memory limitations on the laptop. All computations are performed in R on a MacBook Pro (Apple M1 Max, 32 GB RAM).

From Fig. 3, we observe that when m and k are fixed, both FCPCA and RFCPCA-N are among the fastest methods, and employing the robust variant does not incur a substantial additional cost. The dashed blue and orange lines show the runtime when m is selected automatically from a grid $m \in \{1.1, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.5\}$, which increases the runtime but remains within a feasible range. Notably, the QCD method requires runtimes comparable to or even longer than those of FCPCA with a grid search over m , underscoring that the proposed RFCPCA methods are computationally efficient and scalable, even when automatic parameter selection is incorporated.

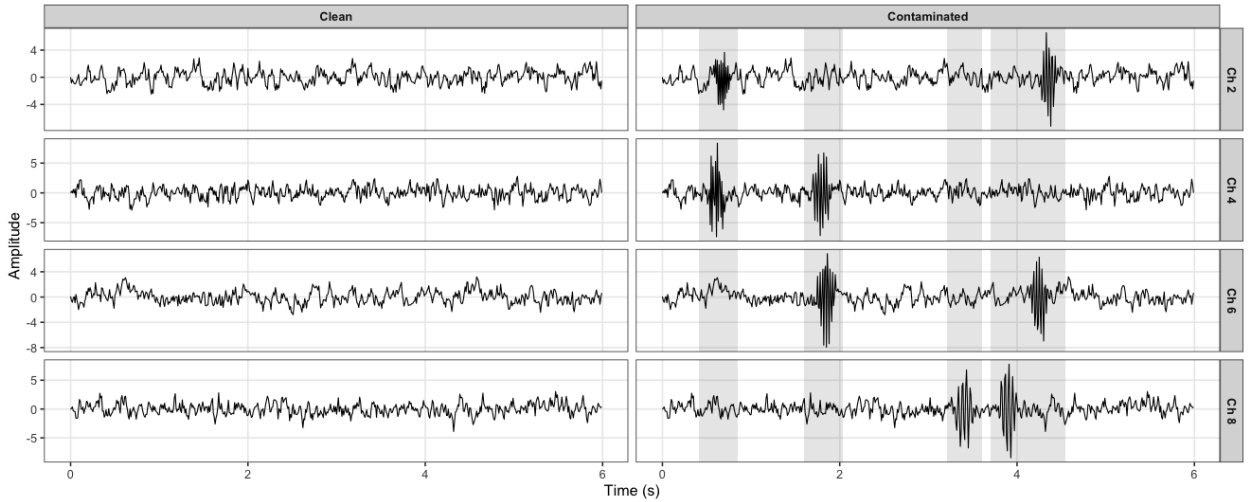


Figure 2: Simulated EEG data with burst contamination

Table 1 reports the results for FCPCA and the RFCPCA family with automatic selection of m and α for RFCPCA-T suggested by the CVI. For comparison, Tables 2 and 3 summarize the results for the alternative methods for which we did not implement automatic tuning. Performance is reported over manually specified grids of m and α . Across settings with a contamination rate of 20%, all methods achieve relatively high clustering accuracy.

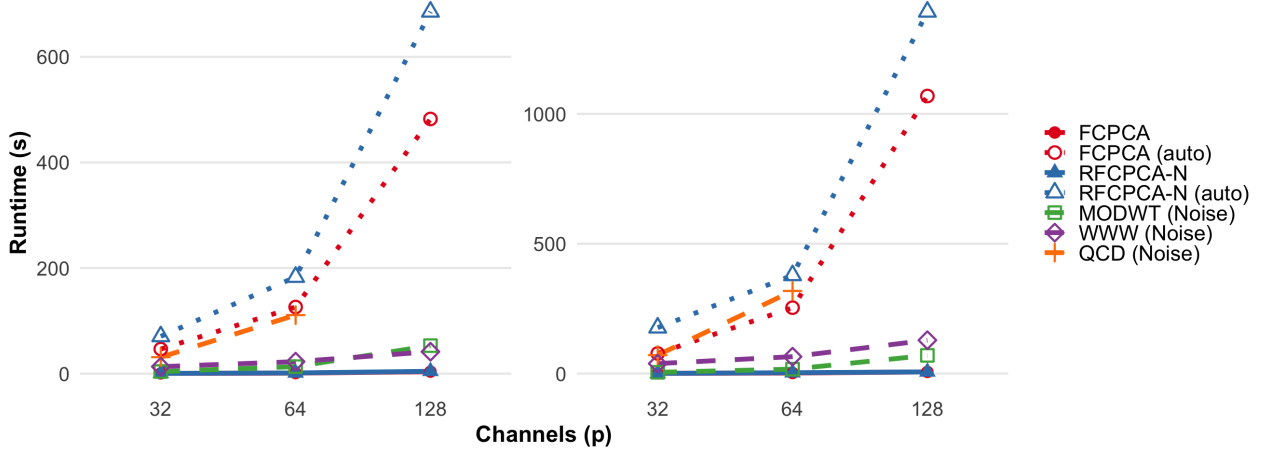


Figure 3: The run time comparison, the noise approach is used to represent the runtime of the robust methods. The left one is when $T = 400$, right one is when $T = 1000$.

RFCPCA delivers strong outlier handling without sacrificing accuracy. In particular, RFCPCA-T attains near-perfect detection ($\text{Out}\% \approx 0.95\text{--}1.00$) with $\text{Acc} = 1.00$ in all configurations, while RFCPCA-E also maintains $\text{Acc} = 1.00$ and detects a large fraction of bursts ($\text{Out}\% \approx 0.67\text{--}0.83$). The noise-cluster variant RFCPCA-N provides explicit anomaly labels and competitive detection ($\text{Out}\% \approx 0.68\text{--}0.78$), with a modest accuracy trade-off at larger p due to membership mass assigned to the noise component. By contrast, FCPCA attains high accuracy but detects virtually no outliers. Even with manual tuning, the comparison methods in Tables 2–3 detect far fewer outliers (typically 0–6% for E/N and 37–57% for trimmed variants with fixed α); notably, even a large fixed trimming level such as $\alpha = 0.40$ does not remove all contaminated trials. Overall, the results highlight the advantage of automatic tuning in RFCPCA for accurate clustering and effective outlier detection while eliminating manual parameter search.

3.4 Simulated EEG data with eye-blink contamination

EEG signals are often contaminated by artifacts from ocular, muscular, cardiac, and motion related activities. Among these, the electrooculogram (EOG) component (refer to eye-blink artifact hereafter) due to involuntary eyelid movements is the most prevalent. Eye blinks appear as high amplitude, spike-like deflections that contaminate EEG in both time and

Table 1: FCPCA and RFCPCA with automatic selection of m (and α for trimming).

Method	T	$p = 32$			$p = 64$			$p = 128$		
		Acc	Out%	α	Acc	Out%	α	Acc	Out%	α
FCPCA	400	1.000	0.000	—	1.000	0.000	—	1.000	0.000	—
	1000	1.000	0.000	—	1.000	0.000	—	1.000	0.000	—
RFCPCA-E	400	1.000	0.750	—	1.000	0.800	—	1.000	0.830	—
	1000	1.000	0.680	—	1.000	0.670	—	1.000	0.750	—
RFCPCA-N	400	1.000	0.750	—	0.980	0.780	—	0.970	0.730	—
	1000	1.000	0.750	—	0.950	0.680	—	0.960	0.700	—
RFCPCA-T	400	1.000	1.000	0.500	1.000	0.980	0.500	1.000	0.950	0.500
	1000	1.000	1.000	0.500	1.000	0.950	0.450	1.000	0.970	0.470

frequency domains, particularly in the low-frequency range (0–12 Hz) associated with motor control, attention, and drowsiness (Chang et al., 2016). As they can distort signals up to the alpha band, eye-blink artifacts are a major source of error in band-power analysis. Their removal is therefore essential to improve EEG quality and ensure reliable applications such as driver drowsiness detection, fatigue monitoring, and brain–computer interfaces (Maddirala and Veluvolu, 2021). Again, 10 MTS are generate each groups, each MTS is with length between 400 to 2000.

In each group, we fix a contamination proportion $\rho = 0.40$. For each contaminated trial, we draw the number of blinks $n_{\text{blink}} \in \{1, 2\}$. For each blink, we sample:

- a duration $\tau \sim \mathcal{U}(0.20, 0.40) f_s$, where τ denotes the blink length in samples (corresponding to 200–400 ms at sampling rate f_s),
- a start time $t_0 \in \{1, \dots, T_i - \tau\}$,
- a set of frontal channels $\mathcal{I} \subseteq \mathcal{F}$ with $|\mathcal{I}| = \lceil \rho_{\text{chan}} p_{\text{frontal}} \rceil$,

The blink waveform is modeled as a half-sine envelope,

$$b_q = \sin\left(\frac{\pi q}{\tau-1}\right), \quad q = 0, \dots, \tau - 1.$$

To mimic their large amplitude, we set an amplitude multiplier $\eta \sim \mathcal{U}(4, 8)$ and a random

Table 2: Comparison methods results for $T = 400$. Acc = accuracy, Out% = outlier fraction.

Method	m	$p = 32$		$p = 64$		$p = 128$	
		Acc	Out%	Acc	Out%	Acc	Out%
WWW-E	1.5	1.00	0.05	1.00	0.04	1.00	0.04
	1.8	1.00	0.06	1.00	0.04	1.00	0.05
	2.0	1.00	0.06	1.00	0.05	1.00	0.05
	2.2	1.00	0.06	1.00	0.05	1.00	0.06
WWW-N	1.5	1.00	0.00	1.00	0.00	1.00	0.00
	1.8	1.00	0.00	0.77	0.24	1.00	0.00
	2.0	1.00	0.00	0.79	0.24	1.00	0.00
	2.2	1.00	0.00	0.85	0.24	1.00	0.00
<i>Trimmed (T) with $\alpha = 0.20$</i>							
WWW-T	1.5	1.00	0.37	1.00	0.39	1.00	0.41
	1.8	1.00	0.37	1.00	0.39	1.00	0.41
	2.0	1.00	0.37	1.00	0.39	1.00	0.41
	2.2	1.00	0.37	1.00	0.39	1.00	0.41
<i>Trimmed (T) with $\alpha = 0.40$</i>							
WWW-T	1.5	1.00	0.51	1.00	0.56	1.00	0.54
	1.8	1.00	0.51	1.00	0.56	1.00	0.52
	2.0	1.00	0.51	1.00	0.56	1.00	0.50
	2.2	1.00	0.51	1.00	0.57	1.00	0.49
MODWT-E	1.5	1.00	0.05	1.00	0.04	1.00	0.04
	1.8	1.00	0.06	1.00	0.05	1.00	0.05
	2.0	1.00	0.06	1.00	0.05	1.00	0.05
	2.2	1.00	0.07	1.00	0.07	1.00	0.06
MODWT-N	1.5	1.00	0.00	1.00	0.00	1.00	0.00
	1.8	1.00	0.00	0.76	0.25	1.00	0.00
	2.0	1.00	0.00	0.78	0.25	1.00	0.00
	2.2	1.00	0.00	0.81	0.25	1.00	0.00
<i>Trimmed (T) with $\alpha = 0.20$</i>							
MODWT-T	1.5	1.00	0.39	1.00	0.37	1.00	0.39
	1.8	1.00	0.39	1.00	0.37	1.00	0.39
	2.0	1.00	0.39	1.00	0.37	1.00	0.39
	2.2	1.00	0.39	1.00	0.37	1.00	0.39
<i>Trimmed (T) with $\alpha = 0.40$</i>							
MODWT-T	1.5	1.00	0.55	1.00	0.58	1.00	0.52
	1.8	1.00	0.55	1.00	0.58	1.00	0.53
	2.0	1.00	0.55	1.00	0.58	1.00	0.54
	2.2	1.00	0.55	1.00	0.57	1.00	0.54
QCD-E	1.5	1.00	0.05	1.00	0.04	—	—
	1.8	1.00	0.06	1.00	0.04	—	—
	2.0	1.00	0.06	1.00	0.05	—	—
	2.2	1.00	0.06	1.00	0.05	—	—
QCD-N	1.5	1.00	0.00	1.00	0.00	—	—
	1.8	1.00	0.00	0.79	0.23	—	—
	2.0	1.00	0.00	0.79	0.24	—	—
	2.2	1.00	0.00	0.84	0.24	—	—
<i>Trimmed (T) with $\alpha = 0.20$</i>							
QCD-T	1.5	1.00	0.37	1.00	0.39	—	—
	1.8	1.00	0.37	1.00	0.38	—	—
	2.0	1.00	0.36	1.00	0.38	—	—
	2.2	1.00	0.37	1.00	0.39	—	—
<i>Trimmed (T) with $\alpha = 0.40$</i>							
QCD-T	1.5	1.00	0.51	1.00	0.56	—	—
	1.8	1.00	0.51	1.00	0.55	—	—
	2.0	1.00	0.52	1.00	0.53	—	—
	2.2	1.00	0.50	1.00	0.54	—	—

Table 3: Comparison methods results for $T = 1000$. Acc = accuracy, Out% = outlier fraction.

Method	m	$p = 32$		$p = 64$		$p = 128$	
		Acc	Out%	Acc	Out%	Acc	Out%
WWW-E	1.5	1.00	0.03	1.00	0.04	1.00	0.04
	1.8	1.00	0.04	1.00	0.05	1.00	0.05
	2.0	1.00	0.04	1.00	0.05	1.00	0.05
	2.2	1.00	0.05	1.00	0.05	1.00	0.05
WWW-N	1.5	1.00	0.00	1.00	0.00	1.00	0.00
	1.8	1.00	0.00	0.74	0.24	1.00	0.00
	2.0	1.00	0.00	0.78	0.24	1.00	0.00
	2.2	1.00	0.00	0.83	0.24	1.00	0.00
<i>Trimmed (T) with $\alpha = 0.20$</i>							
WWW-T	1.5	1.00	0.36	1.00	0.40	1.00	0.39
	1.8	1.00	0.36	1.00	0.40	1.00	0.39
	2.0	1.00	0.36	1.00	0.40	1.00	0.39
	2.2	1.00	0.36	1.00	0.40	1.00	0.39
<i>Trimmed (T) with $\alpha = 0.40$</i>							
WWW-T	1.5	1.00	0.54	1.00	0.56	1.00	0.54
	1.8	1.00	0.54	1.00	0.56	1.00	0.54
	2.0	1.00	0.54	1.00	0.56	1.00	0.54
	2.2	1.00	0.54	1.00	0.57	1.00	0.54
MODWT-E	1.5	1.00	0.03	1.00	0.04	1.00	0.04
	1.8	1.00	0.04	1.00	0.05	1.00	0.06
	2.0	1.00	0.04	1.00	0.05	1.00	0.06
	2.2	1.00	0.06	1.00	0.05	1.00	0.06
MODWT-N	1.5	1.00	0.00	1.00	0.00	1.00	0.00
	1.8	1.00	0.00	0.76	0.25	1.00	0.00
	2.0	1.00	0.00	0.78	0.25	1.00	0.00
	2.2	1.00	0.00	0.81	0.25	1.00	0.00
<i>Trimmed (T) with $\alpha = 0.20$</i>							
MODWT-T	1.5	1.00	0.37	1.00	0.40	1.00	0.39
	1.8	1.00	0.37	1.00	0.40	1.00	0.39
	2.0	1.00	0.37	1.00	0.40	1.00	0.39
	2.2	1.00	0.37	1.00	0.40	1.00	0.39
<i>Trimmed (T) with $\alpha = 0.40$</i>							
MODWT-T	1.5	1.00	0.55	1.00	0.58	1.00	0.53
	1.8	1.00	0.55	1.00	0.58	1.00	0.53
	2.0	1.00	0.55	1.00	0.58	1.00	0.54
	2.2	1.00	0.56	1.00	0.57	1.00	0.54
QCD-E	1.5	1.00	0.03	1.00	0.04	—	—
	1.8	1.00	0.04	1.00	0.05	—	—
	2.0	1.00	0.04	1.00	0.05	—	—
	2.2	1.00	0.05	1.00	0.05	—	—
QCD-N	1.5	1.00	0.00	1.00	0.00	—	—
	1.8	1.00	0.00	1.00	0.00	—	—
	2.0	1.00	0.00	1.00	0.00	—	—
	2.2	1.00	0.00	1.00	0.00	—	—
<i>Trimmed (T) with $\alpha = 0.20$</i>							
QCD-T	1.5	1.00	0.36	1.00	0.40	—	—
	1.8	1.00	0.36	1.00	0.40	—	—
	2.0	1.00	0.37	1.00	0.40	—	—
	2.2	1.00	0.37	1.00	0.41	—	—
<i>Trimmed (T) with $\alpha = 0.40$</i>							
QCD-T	1.5	1.00	0.54	1.00	0.56	—	—
	1.8	1.00	0.53	1.00	0.56	—	—
	2.0	1.00	0.53	1.00	0.56	—	—
	2.2	1.00	0.55	1.00	0.56	—	—

polarity $\kappa \in \{-1, +1\}$. For all $q = 0, \dots, \tau - 1$ and $j \in \mathcal{I}$,

$$X_{t_0+q,j} \leftarrow X_{t_0+q,j} + \kappa \eta \hat{\sigma}_j b_q.$$

Thus each selected trial contains one or more transient half-sine deflections localized to frontal electrodes, reproducing the strong low-frequency transients characteristic of real eye-blink artifacts. Fig. 4 presents an example of the EEG data with eye-blink contamination.

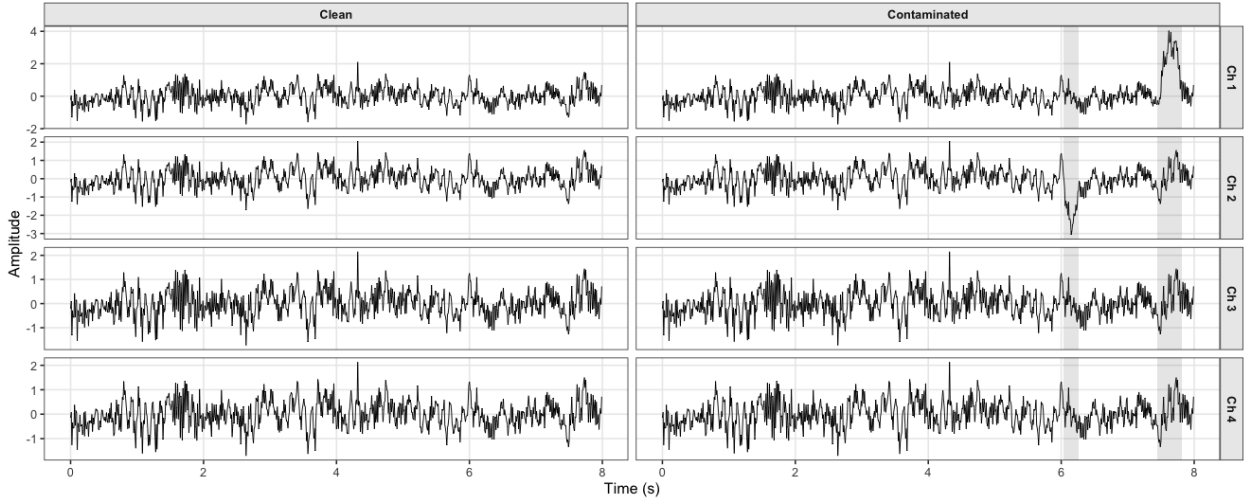


Figure 4: Simulated EEG with eye-blink contamination (one contaminated trial highlighted).

In Section 3.3, we already compared against competitors and found that, even with manual tuning, they detected substantially fewer outliers while matching or trailing our methods in clustering accuracy. This section will only focus on FCPCA and RFCPCA and do not repeat the other baselines. Our goal is to test the methods under a different contamination mechanism, and show the improvement of RFCPCA over FCPCA. As shown in Table 1, RFCPCA achieves the same clustering accuracy as FCPCA while providing substantially higher outlier detection, highlighting the advantage of robust variants under eye-blink contamination. At the same time, FCPCA also identifies a larger number of outliers in this scenario because the contamination level is as high as 40%, making outliers easier to detect.

From Table 4, across dimensions, the key difference lies in outlier detection: RFCPCA-E

Table 4: FCPCA and RFCPCA with automatic selection of m (and α for trimming). Trial-level contamination $\rho = 0.40$. All trials have variable lengths $T \in [400, 2000]$.

Method	$p = 32$			$p = 64$			$p = 128$		
	Acc	Out%	α	Acc	Out%	α	Acc	Out%	α
FCPCA	1.00	0.78	—	1.00	0.85	—	1.00	0.60	—
RFCPCA_E	1.00	0.98	—	1.00	1.00	—	1.00	1.00	—
RFCPCA_N	1.00	0.89	—	1.00	0.93	—	1.00	0.85	—
RFCPCA_T	1.00	1.00	0.43	1.00	1.00	0.50	1.00	1.00	0.46

and RFCPCA-T are essentially perfect (Out% ≈ 0.98 – 1.00), while RFCPCA-N is high but slightly lower, reflecting the conservativeness of a fixed noise threshold. Plain FCPCA flags only a fraction of contaminated trials (Out% = 0.60 – 0.85), consistent with its sensitivity to large low-frequency transients but lack of explicit robustness. The α values selected by RFCPCA-T (≈ 0.43 – 0.50) adapt closely to the true contamination level $\rho = 0.40$, with small deviations attributable to random blink count, timing, and channel selection. Overall, the robust variants reliably isolate eye-blink contamination under variable T_i without manual tuning, which is the intended operating regime for real EEG.

4 Exploring the mapping between EEG and alertness in a driving experiment

In this section, we consider a publicly available EEG data¹. The EEG Driver Drowsiness Dataset contains multi-channel EEG recordings from 11 subjects performing a sustained-attention driving task, with labels indicating alert and drowsy states. The data include 2022 3-second EEG samples (128 Hz, 30 channels). We evaluate clustering accuracy using three metrics: the RI, the adjusted Rand index (ARI), and the fuzzy Rand index (RIF). The same dataset is used here as Ma et al. (2025) to directly show the gains in robustness.

¹https://figshare.com/articles/dataset/EEG_driver_drowsiness_dataset/14273687?file=30707285

4.1 The clustering performance on all subjects

Table 5 shows that all three robust extensions generally outperform FCPCA across subjects.

We now focus on Subjects 6 and 9, which exhibit the most pronounced gains in accuracy under the RFCPCA variants.

Table 5: Comparison of clustering accuracy (RI / ARI / RIF) per subject. Bold indicates an RFCPCA variant strictly better than FCPCA on all three metrics.

Subject #	Size	FCPCA			RFCPCA-E			RFCPCA-T			RFCPCA-N		
		RI	ARI	RIF	RI	ARI	RIF	RI	ARI	RIF	RI	ARI	RIF
1	188	0.740	0.240	0.510	0.820	0.390	0.700	0.780	0.320	0.550	0.810	0.350	0.640
2	132	0.890	0.620	0.820	0.890	0.600	0.810	0.910	0.660	0.890	0.890	0.610	0.820
3	150	0.550	0.010	0.890	0.550	0.010	0.900	0.690	0.140	0.950	0.630	0.060	0.910
4	148	0.570	0.020	0.580	0.600	0.040	0.590	0.690	0.130	0.650	0.700	0.160	0.660
5	224	0.880	0.560	0.740	0.880	0.570	0.880	0.880	0.570	0.720	0.880	0.560	0.730
6	166	0.510	0.000	0.520	0.840	0.450	0.940	0.990	0.880	0.980	0.710	0.170	0.640
7	102	0.580	0.020	0.660	0.640	0.090	0.690	0.590	0.040	0.670	0.580	0.010	0.650
8	264	0.660	0.090	0.610	0.690	0.120	0.680	0.690	0.140	0.680	0.650	0.080	0.600
9	314	0.750	0.260	0.840	0.880	0.550	0.850	0.920	0.680	0.850	0.920	0.690	0.890
10	108	1.000	1.000	0.920	1.000	1.000	0.930	0.990	0.960	0.940	1.000	1.000	0.940
11	226	0.930	0.750	0.860	0.930	0.720	0.950	0.960	0.850	0.970	0.930	0.710	0.870
Mean		0.730	0.320	0.720	0.790	0.410	0.820	0.830	0.490	0.800	0.790	0.400	0.760

4.2 Analysis of Subject 6

The ground-truth partition is simple: the first half of trials are alert and the second half are drowsy. However, the FCPCA membership matrix in Fig. 5 does not recover this step change: memberships drift across the sequence, so no clear block forms at the midpoint. A plausible reason is that the quadratic loss in FCPCA is dominated by many borderline/transitional windows and within-state heterogeneity. Suppose physiology shifts gradually while the labels switch at a hard boundary, numerous windows straddle the change and appear ambiguous. In that case, fitting them under squared loss pulls the two CPCA subspaces toward a common average, reducing separation. Consequently, the memberships in Fig. 5 are low-contrast and unstable.

Moving to RFCPCA-E, Fig. 6 shows that the exponential loss down-weights poorly fitted

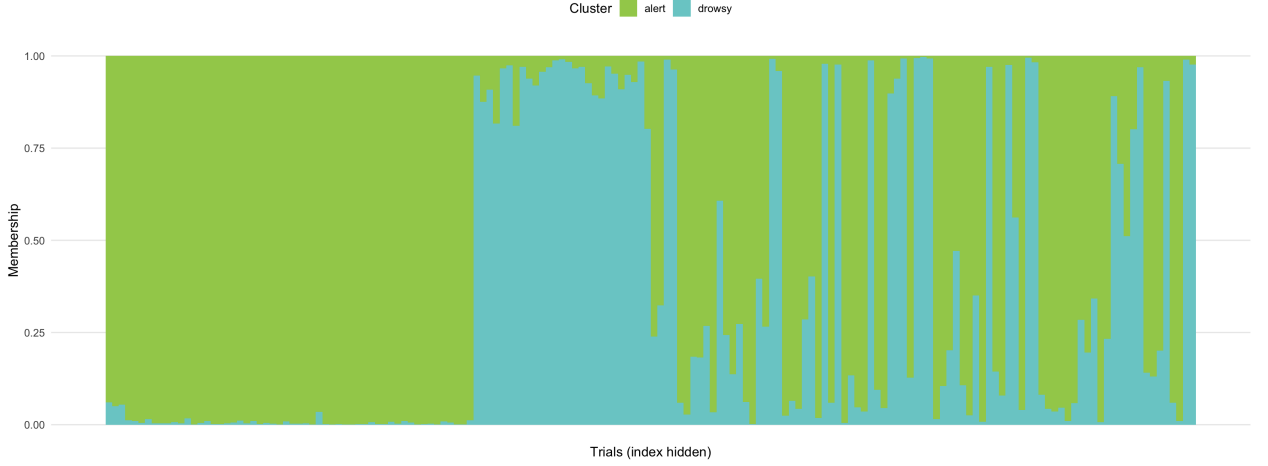


Figure 5: Membership matrix of Subj 6 using FCPCA

windows, so the subspaces are estimated mainly from the core trials. The result suggests a crisp, step-like partition with a clear split near the midpoint. Some short mistakes remain in the middle, a brief stretch that should be “alert” is assigned to “drowsy.” These misassignments are consistent with lingering transitional/heterogeneous windows: they are not gross outliers (as corroborated by the dispersed, low-noise mass under RFCPCA-N later), but still sit closer to the drowsy subspace after reweighting, so they flip labels locally, even though the global partition is much cleaner.



Figure 6: Membership matrix of Subj 6 using RFCPCA-E

With RFCPCA-N (Fig. 7), there is no obvious outlier trail. Noise memberships are small and dispersed. Accuracy improves because trials that assign some mass to the noise cluster

receive reduced effective weight when estimating the projection axes, so ambiguous or slightly outlying trials contribute less to the subspace estimation, yielding cleaner projection axes.



Figure 7: Membership matrix of Subj 6 using RFCPCA-N

As shown in Fig. 8, the trimmed windows (triangle markers) are concentrated in the middle portion of the session, precisely where RFCPCA-E and RFCPCA-N still make mistakes by assigning some “alert” trials to “drowsy.” By excluding these trials from the subspace estimation, RFCPCA-T achieves the highest clustering accuracy among the methods (near perfect). Notably, the same trial that carries a high noise membership under RFCPCA-N is included in the trimmed set of RFCPCA-T, indicating that the two robust strategies are aligned to identify difficult windows and that trimming resolves the ambiguity.

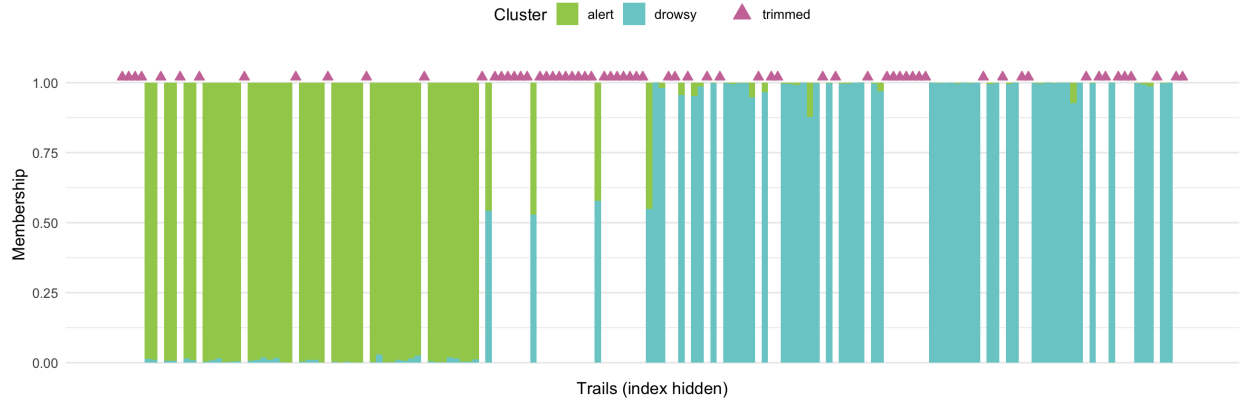


Figure 8: Membership matrix of Subj 6 using RFCPCA-T

4.3 Analysis of Subject 9

The membership matrix in Fig. 9 shows the membership matrix using FCPCA. Many windows carry mixed memberships, and the transition region is diffuse, indicating that the two CPCA subspaces are partially averaged when confronted with gradual state drift and within-state heterogeneity.

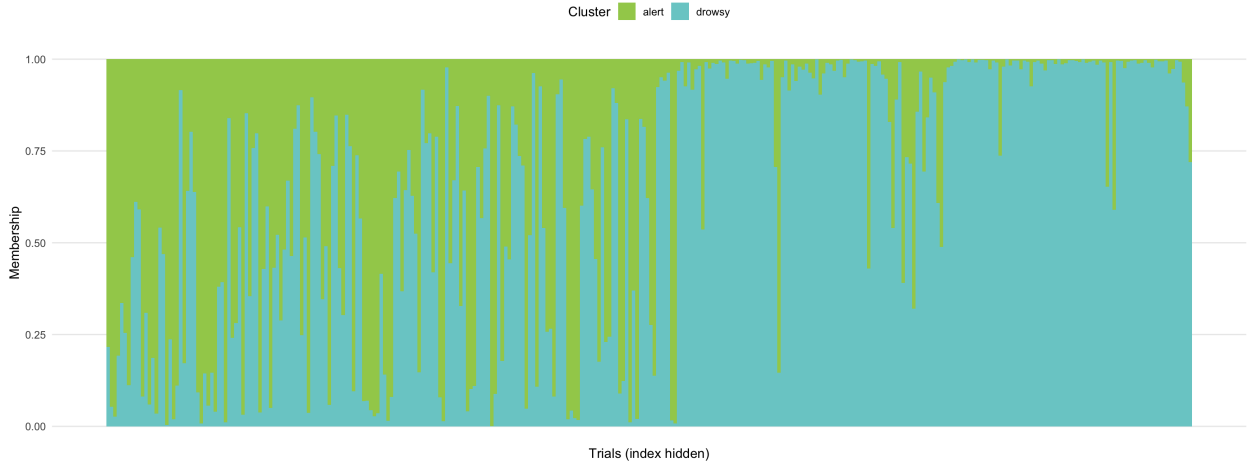


Figure 9: Membership matrix of Subj 9 using FCPCA

With RFCPCA-E (Fig. 10), down-weighting poorly fitted trials improves both brain states segment: most of the first part is now cleanly assigned to alert rather than being intermittently misclassified, and the second half forms a much clearer drowsy block with only occasional, narrow spikes. The transition between the two segments becomes markedly more step-like, indicating better separation of the subspaces.

In Fig. 11, trials flagged as outliers are frequent, especially in the first third. Several trials exhibit almost-crisp membership in the noise cluster, indicating activity that is highly inconsistent with both the alert and drowsy subspaces, either strong artifacts or an unmodeled brain state not captured by the two-class setup—and thus potentially meaningful for downstream analysis.

Lastly, in Fig. 12, the trimmed windows (triangle markers) concentrate in the early, more

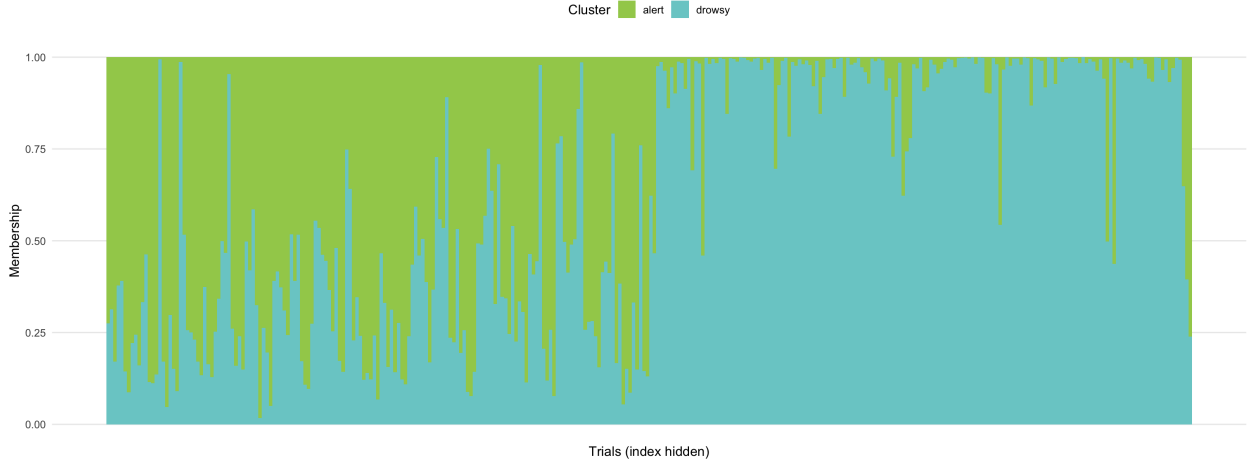


Figure 10: Membership matrix of Subj 9 using RFCPCA-E

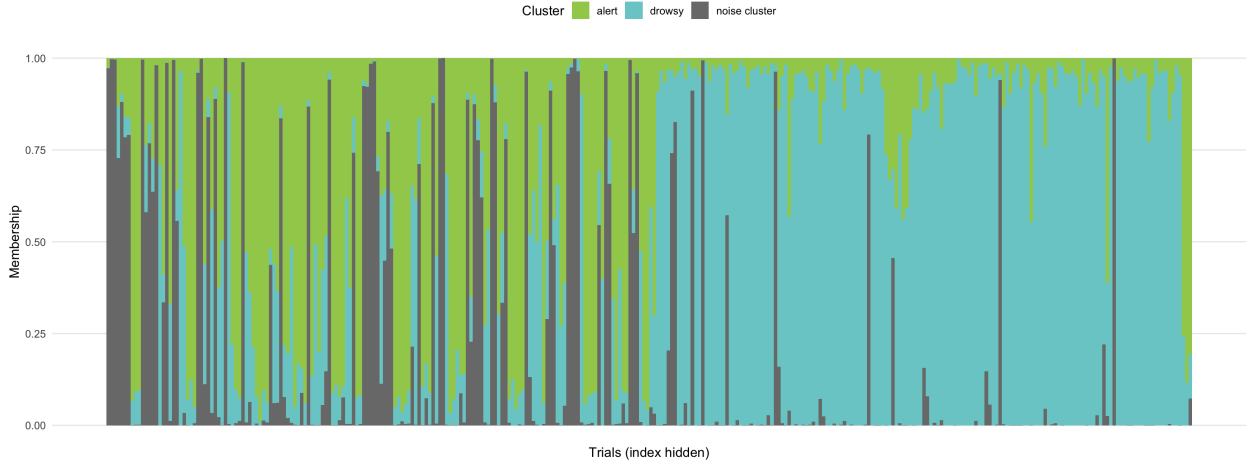


Figure 11: Membership matrix of Subj 9 using RFCPCA-N

variable portion and around the transition. Trimming removes these discordant windows from the subspace estimation and produces the cleanest two-block structure and the best accuracy. Quantitatively, RFCPCA-T trims about 30% of trials, whereas RFCPCA-N flags about 23% as noise. Interesting, all RFCPCA-N outliers are included in the RFCPCA-T trimmed set (100% overlap), while the remaining 7% trimmed windows are additional borderline cases that RFCPCA-N treats with small, dispersed noise mass. This alignment explains why both methods identify the same difficult region, and why trimming resolves the residual ambiguity.



Figure 12: Membership matrix of Subj 9 using RFCPCA-T

4.4 Discriminating subspaces and further insights with RFCPCA-N on Subject 9

Subject 9 contains a sizable fraction of trials with non-negligible noise membership under RFCPCA-N. Our goals here are twofold: (i) verify that the noise component corresponds to dynamics that are genuinely different from the normal physiological states; and (ii) measure the extent of overlap (shared axes) between the alert and drowsy subspaces and verify that they remain separable, as expected under fuzzy, graded transitions.

To quantify the degree of similarity and separation among the estimated subspaces, we first computed the principal angles between pairs of subspaces (alert, drowsy, and noise). Suppose $\mathcal{B}_a$ and $\mathcal{B}_b$ denote two sets of orthonormal basis vectors obtained from the projection axes of different clusters. The cosines of the principal angles are then given by the singular values of $\mathcal{B}_a^\top \mathcal{B}_b$. Small angles reveal overlapping directions (shared latent dynamics), while large angles indicate nearly orthogonal directions (distinct dynamics). Fig. 13 demonstrates that the alert and drowsy subspaces share several common axes, showing partially overlapping activity, whereas the noise subspace is much more different from both, indicating that the noise subspace captures dynamics distinct from both brain states. The separation is more pronounced at lag 0–2, suggesting that temporal dependencies enhance discrimination.

We then examined the contribution of individual EEG channels to the formation of each

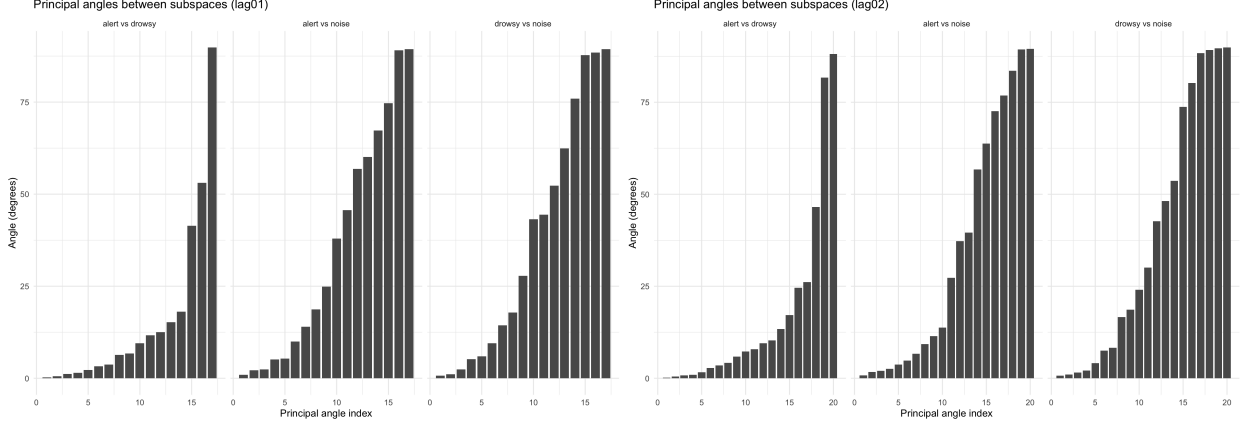


Figure 13: Principal angles (in increasing order) between alert, drowsy, and noise subspaces at lag 0–1 and lag 0–2

subspace. For channel j in cluster s , the contribution is defined as

$$\text{contrib}_{j,s} = \sum_k C_s(j, k)^2,$$

where $C_s(j, k)$ denotes the loading of channel j on the k -th retained basis vector of the subspace associated with cluster s . This measure reflects how strongly each electrode shapes the discriminating axes.

Fig. 14 and 15 display the contributions of all 30 EEG channels to the alert, drowsy, and noise subspaces using lag 0–1 and lag 0–2, respectively. The alert and drowsy maps share broadly similar patterns, further verifying separable but overlapping patterns consistent with transitional or mixed states. Despite this similarity, the two states also show differences in discrimination (e.g., FT7, F8, O1).

The noise subspace remains distinct across both lags, with higher weights concentrated at frontal/temporal sites (e.g., F3, F4, FT8), suggesting that it primarily captures artifacts or other unsystematic activity rather than task-related dynamics.

4.5 Summary of the EEG data: robust gains and fuzzy insights

Although the dataset is preprocessed and should not contain obvious outliers, RFCPCA still yields clear gains in clustering accuracy. This indicates that robustness is beneficial not only

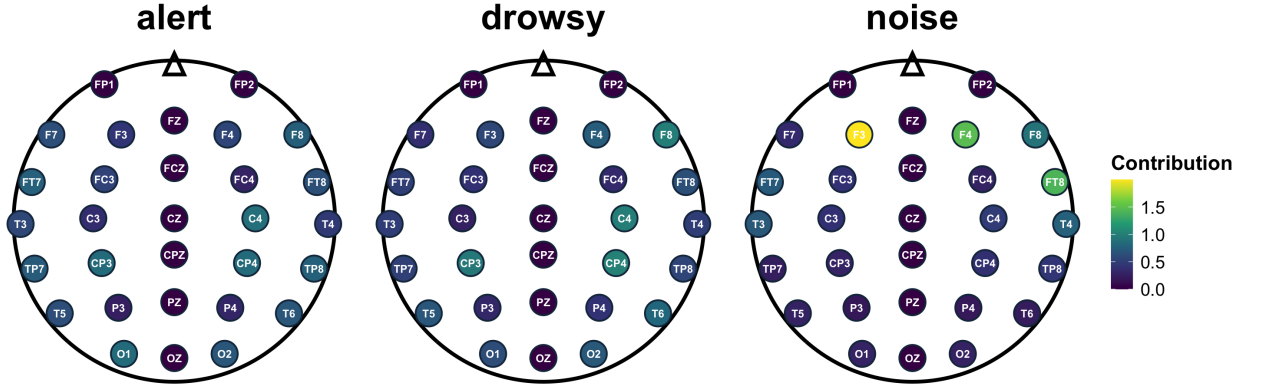


Figure 14: EEG Channel Contributions to alert, drowsy, and noise subspaces using lag 0–1

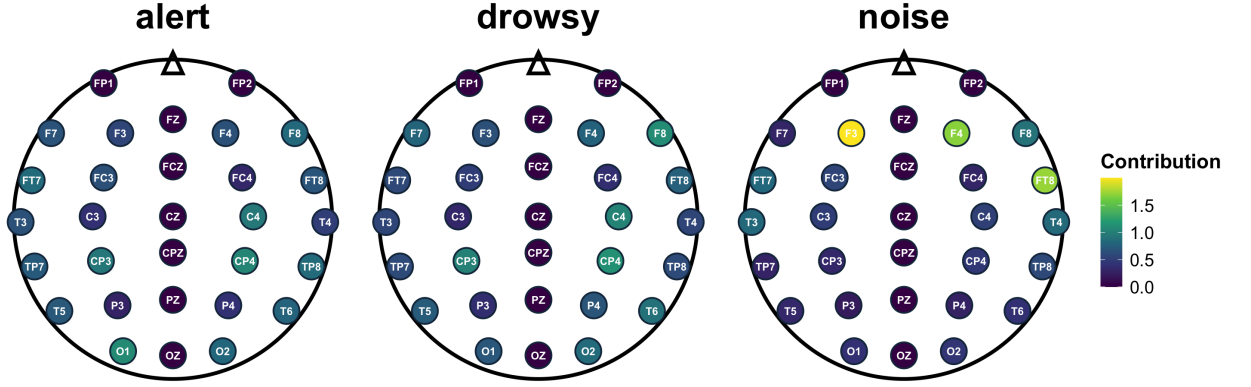


Figure 15: EEG Channel Contributions to alert, drowsy, and noise subspaces using lag 0–2 under heavy contamination but also for subtle irregularities and residual noise that persist after preprocessing.

Beyond accuracy, the fuzzy formulation adds interpretability. First, fuzzy memberships capture ambiguous or transitional states between clusters—particularly relevant because brain activity seldom switches abruptly between alertness and drowsiness. Second, membership magnitudes quantify assignment uncertainty, highlighting series that, while not gross outliers, deviate from dominant patterns. Third, secondary memberships reveal overlap and shared structure across clusters, yielding a richer description than crisp partitions. Finally, the robust extensions detect partially outlying series along a continuum rather than forcing a binary inlier/outlier decision.

5 Conclusion

We presented a robust fuzzy subspace clustering framework for MTS that builds on FCPCA and introduces three complementary variants, RFCPCA-E (robust metric), RFCPCA-N (noise cluster), and RFCPCA-T (trimming), all with automatic selection of the number of clusters S , the fuzziness parameter m , and, when applicable, the trimming fraction α . The approach summarizes each object with covariance features, accommodates variable lengths T_i , and uses fuzzy memberships to quantify assignment uncertainty and handle atypical trials.

Across synthetic EEG-like settings with artifacts and in real EEG, the robust variants improve clustering accuracy relative to FCPCA while providing principled outlier handling. In Simulation 1 (burst contamination), RFCPCA consistently retained the clean structure and isolated contaminated trials. In Simulation 2 (eye-blink contamination), which stresses low-frequency, trial-level transients with varying lengths, RFCPCA again achieved perfect or near-perfect accuracy and markedly higher outlier detection than FCPCA under fully automatic tuning. On the real dataset, robustness remained beneficial even after standard preprocessing, and the fuzzy memberships improved interpretation by indicating gradual transitions between clusters; additionally, the estimated noise subspace was distinct from physiologic subspaces, supporting a meaningful separation between structured signal and contamination.

Despite these strengths, the framework is primarily second-order: covariance summaries emphasize amplitude cofluctuations and may miss discriminative phase or cross-lag structure when such information is essential. It also presumes a stationary core process under moderate contamination; performance can degrade when artifacts dominate many trials or persist for long durations. In addition, eigendecomposition of high-dimensional covariances becomes costly as p and the number of trials grow. Future work includes improving scalability

via shrinkage covariances, randomized eigensolvers, and GPU acceleration, and enhancing robustness to missing or noisy channels. Moreover, the robust fuzzy clustering-based MTS forecasting algorithm can also be considered.

Acknowledgments

This research was supported by King Abdullah University of Science and Technology (KAUST).

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code availability

All three RFCPCA methods are coded in R and are freely available on the following GitHub repository: <https://github.com/arbitraryma/RFCPCA.git>. For any questions or inquiries, please contact the corresponding author.

References

- Aggarwal, C. C. and Yu, P. S. (2001). Outlier detection for high dimensional data. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, pages 37–46.
- Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering—a decade review. *Information systems* **53**, 16–38.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., and Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern recognition* **46**, 243–256.
- Banerjee, A. and Dave, R. N. (2012). Robust clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2**, 29–59.
- Chang, W.-D., Cha, H.-S., Kim, K., and Im, C.-H. (2016). Detection of eye blink artifacts from single prefrontal channel electroencephalogram. *Computer methods and programs in biomedicine* **124**, 19–30.
- Cimino, M. G., Frosini, G., Lazzerini, B., Marcelloni, F., et al. (2004). On the noise distance in robust fuzzy c-means. In *International Conference on Computational Intelligence*, pages 361–364.
- Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters* **12**, 657–664.
- Davé, R. N. and Krishnapuram, R. (2002). Robust clustering methods: a unified view. *IEEE Transactions on fuzzy systems* **5**, 270–293.
- Duran, B. S. and Odell, P. L. (2013). *Cluster analysis: a survey*, volume 100. Springer Science & Business Media.
- D’Urso, P. and De Giovanni, L. (2014). Robust clustering of imprecise data. *Chemometrics and Intelligent Laboratory Systems* **136**, 58–80.
- D’Urso, P., De Giovanni, L., and Massari, R. (2015). Time series clustering by a robust

- autoregressive metric with application to air pollution. *Chemometrics and Intelligent Laboratory Systems* **141**, 107–124.
- D’Urso, P., García-Escudero, L. A., De Giovanni, L., Vitale, V., and Mayo-Iscar, A. (2021). Robust fuzzy clustering of time series based on b-splines. *International Journal of Approximate Reasoning* **136**, 223–246.
- D’Urso, P. and Leski, J. M. (2023). Owa-based robust fuzzy clustering of time series with typicality degrees. *Information Sciences* **651**, 119706.
- D’Urso, P. and Maharaj, E. A. (2012). Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems* **193**, 33–61.
- D’Urso, P., De Giovanni, L., and Massari, R. (2021). Trimmed fuzzy clustering of financial time series based on dynamic time warping. *Annals of operations research* **299**, 1379–1395.
- D’Urso, P., De Giovanni, L., and Vitale, V. (2022). Spatial robust fuzzy clustering of covid 19 time series based on b-splines. *Spatial Statistics* **49**, 100518.
- D’Urso, P., De Giovanni, L., and Vitale, V. (2023). Robust dtw-based entropy fuzzy clustering of time series. *Annals of Operations Research* pages 1–35.
- Esling, P. and Agon, C. (2012). Time-series data mining. *ACM Computing Surveys (CSUR)* **45**, 1–34.
- García-Escudero, L. A., Gordaliza, A., Matrán, C., and Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. *The Annals of Statistics* **36**, 1324–1345.
- García-Escudero, L. A., Gordaliza, A., Matrán, C., and Mayo-Iscar, A. (2010). A review of robust clustering methods. *Advances in Data Analysis and Classification* **4**, 89–109.
- García-Escudero, L. A. and Mayo-Iscar, A. (2024). Robust clustering based on trimming. *Wiley Interdisciplinary Reviews: Computational Statistics* **16**, e1658.
- Ghassempour, S., Girosi, F., and Maeder, A. (2014). Clustering multivariate time series

- using hidden markov models. *International journal of environmental research and public health* **11**, 2741–2763.
- Goncharova, I. I., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2003). Emg contamination of eeg: spectral and topographical characteristics. *Clinical neurophysiology* **114**, 1580–1593.
- Heiser, W. J. and Groenen, P. J. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika* **62**, 63–83.
- Hwang, H., DeSarbo, W. S., and Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika* **72**, 181–198.
- Izakian, H., Pedrycz, W., and Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence* **39**, 235–244.
- Jayaram, B. and Klawonn, F. (2013). Can fuzzy clustering avoid local minima and undesired partitions? In *Computational intelligence in intelligent data analysis*, pages 31–44. Springer.
- Jha, A., Ray, S., Seaman, B., and Dhillon, I. S. (2015). Clustering to forecast sparse time-series data. In *2015 IEEE 31st international conference on data engineering*, pages 1388–1399. IEEE.
- Kavitha, V. and Punithavalli, M. (2010). Clustering time series data stream-a literature survey. *arXiv preprint arXiv:1005.4270*.
- Kruse, R., Döring, C., and Lesot, M.-J. (2007). Fundamentals of fuzzy clustering. *Advances in fuzzy clustering and its applications* pages 3–30.
- Kwon, S. H. (1998). Cluster validity index for fuzzy clustering. *Electronics letters* **34**, 2176–2177.

- Lafuente-Rego, B., D’Urso, P., and Vilar, J. A. (2020). Robust fuzzy clustering based on quantile autocovariances. *Statistical papers* **61**, 2393–2448.
- Li, H. (2019). Multivariate time series clustering based on common principal component analysis. *Neurocomputing* **349**, 239–247.
- Li, H. and Wei, M. (2020). Fuzzy clustering based on feature weights for multivariate time series. *Knowledge-Based Systems* **197**, 105907.
- Li, J., Izakian, H., Pedrycz, W., and Jamal, I. (2021). Clustering-based anomaly detection in multivariate time series data. *Applied Soft Computing* **100**, 106919.
- López-Oriona, Á., D’Urso, P., Vilar, J. A., and Lafuente-Rego, B. (2022). Quantile-based fuzzy c-means clustering of multivariate time series: Robust techniques. *International Journal of Approximate Reasoning* **150**, 55–82.
- López-Oriona, Á. and Vilar, J. A. (2021). Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series. *Expert Systems with Applications* **185**, 115677.
- Ma, Z., Lopez Oriona, A., Ombao, H., and Sun, Y. (2024). ROBCPCA: A robust multivariate time series clustering method based on common principal component analysis.
- Ma, Z., Talento, M. S., Sun, Y., and Ombao, H. (2025). Fuzzcoh: Robust canonical coherence-based fuzzy clustering of multivariate time series. *arXiv preprint arXiv:2506.22861*.
- Ma, Z., Ángel López-Oriona, Ombao, H., and Sun, Y. (2025). FCPCA: Fuzzy clustering of high-dimensional time series based on common principal component analysis. *International Journal of Approximate Reasoning* **187**, 109552.
- Maddirala, A. K. and Veluvolu, K. C. (2021). Eye-blink artifact removal from single channel eeg with k-means and ssa. *Scientific Reports* **11**, 11043.
- Maharaj, E. A. and D’Urso, P. (2011). Fuzzy clustering of time series in the frequency domain. *Information Sciences* **181**, 1187–1211.

- McBratney, A. B. and Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and forest meteorology* **35**, 165–185.
- Rehm, F., Klawonn, F., and Kruse, R. (2007). A novel approach to noise clustering for outlier detection. *Soft Computing* **11**, 489–494.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association* **79**, 871–880.
- Rousseeuw, P. J. and van Zomeren, B. C. (1991). Robust distances: simulations and cutoff values. In *Directions in Robust Statistics and Diagnostics: Part II*, pages 195–203. Springer.
- Ruspini, E. H., Bezdek, J. C., and Keller, J. M. (2019). Fuzzy clustering: A historical perspective. *IEEE Computational Intelligence Magazine* **14**, 45–55.
- Shi, W., Karastoyanova, D., Ma, Y., Huang, Y., and Zhang, G. (2023). Clustering-based granular representation of time series with application to collective anomaly detection. *IEEE transactions on instrumentation and measurement* **72**, 1–12.
- Suganya, R. and Shanthi, R. (2012). Fuzzy c-means algorithm-a review. *International Journal of Scientific and Research Publications* **2**, 1.
- Wang, H.-Y., Wang, J.-S., and Wang, G. (2022). A survey of fuzzy clustering validity evaluation methods. *Information Sciences* **618**, 270–297.
- Wang, W. and Zhang, Y. (2007). On fuzzy cluster validity indices. *Fuzzy sets and systems* **158**, 2095–2117.
- Wang, X., Wirth, A., and Wang, L. (2007). Structure-based statistical features and multivariate time series clustering. In *Seventh IEEE international conference on data mining (ICDM 2007)*, pages 351–360. IEEE.
- Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., Broberg, M., Wallace, A., DeLosAngeles, D., Lillie, P., et al. (2007). Scalp electrical recording

during paralysis: quantitative evidence that eeg frequencies above 20 hz are contaminated by emg. *Clinical neurophysiology* **118**, 1877–1888.

Wu, K.-L. and Yang, M.-S. (2002). Alternative c-means clustering algorithms. *Pattern recognition* **35**, 2267–2278.

Xie, X. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 841–847.