
ROBUST FUNCTIONAL WARD'S LINKAGES WITH APPLICATIONS IN EEG DATA CLUSTERING

Tianbo Chen and Hanbing Zhu

Anhui University

Hefei, China

E-mail: chentianbo@ahu.edu.cn; zhuhbecnu@163.com

ABSTRACT

This paper proposes two novel distance measures, called functional Ward's linkages, for functional data clustering that are robust against outliers. Conventional Ward's linkage defines the distance between two clusters as the increase in sum of squared errors (SSE) upon merging, which can be interpreted graphically as an increase in cluster diameter. Analogously, functional Ward's linkage defines the distance between two clusters as the increase in the width of the band after merging them. To address the limitations of conventional Ward's linkage in handling outliers, the proposed linkages focus exclusively on the most central curves by leveraging magnitude-shape outlyingness measure and modified band depth, respectively. Simulations and real-world electroencephalogram (EEG) data analysis demonstrate that the proposed methods outperform other competitive approaches, particularly in the presence of various types of outliers.

Keywords Functional data clustering; Functional Ward's linkages; Robustness; Magnitude-shape outlyingness; Modified band depth; EEG data analysis

1 Introduction

Functional data clustering has become an active research area in recent years, with applications across various scientific fields. One notable example is electroencephalogram (EEG) data analysis. For instance, Euán et al. (2018) applied the total variation distance (TVD) to estimated spectral densities to cluster resting-state EEG signals from different channels that are spectrally synchronized. Maadooliat et al. (2018) proposed a nonparametric collective spectral density estimation (NCSDE) algorithm for a collection of time series. The coefficients of the basis are used for clustering, thereby reducing the computational complexity. Chen et al. (2020) further extended NCSDE to spatial data, incorporating spatial dependence to construct spatially homogeneous clusters. Beyond EEG analysis, functional data clustering has been applied in other scientific fields. For example, Cheifetz et al. (2017) proposed two data-driven clustering methods for functional data, with applications in water demand modeling. A shape-based clustering approach using dynamic time warping (DTW) was proposed in Teeraratkul et al. (2017), applied to consumer demand response data, resulting in a 50% reduction in the number of representative groups and an improvement in prediction accuracy. For additional applications of functional data clustering, we refer readers to Jacques and Preda (2014); Wang et al. (2016); Zhang and Parnell (2023).

One critical issue in hierarchical clustering algorithms is to define the distance or linkage of two clusters. In addition to some well-known methods such as Ward's linkage, centroid linkage, and complete linkage, there are many functional data clustering methods. Liao (2005) concluded three approaches to cluster functional data: methods that depend on the comparison of the raw data, methods based on the comparison of models fitted to the data, and finally, methods based on features derived from the data. Jacques and Preda (2014) categorized several conventional functional data clustering methods into three group: dimension reduction before clustering, nonparametric methods using specific distances or linkages between curves, and model-based clustering methods. (Montero and Vilar, 2015) presented the **TSClust** package in R for time series clustering, offering a wide range of alternative linkage methods. Additional review articles can be found in Aghabozorgi et al. (2015); Chamroukhi and Nguyen (2019); Cheam and Fredette (2020).

Another crucial capability of clustering algorithms is robustness against outliers. A significant amount of research has focused on developing robust clustering algorithms for functional data. A widely used technique is trimming, where potential outlying data are deleted before clustering. Garcia-Escudero and Gordaliza (2005) adopted the cubic B-spline basis and a variant of the k -means algorithm, called the trimmed k -means algorithm, where in each iteration of the algorithm, a fixed number of the most outlying coefficient vectors are excluded from calculating the cluster centers. Rivera-García et al. (2019) proposed a robust, model-based clustering method based on an approximation to the "density function" for functional data. By incorporating a trimming step, the method reduces the effect of contaminated observations. Beyond trimming algorithms, Wu and Yu (2006) utilized independent component analysis (ICA) to obtain independent components for multivariate time series and developed a clustering algorithm called ICLUS to group time series according to the independent components. D'Urso et al. (2015) presented a robust clustering method based on autoregressive models, where a partition around medoids scheme is adopted and the robustness of the method comes from the use of a robust metric between time series. D'Urso et al. (2016) proposed a robust fuzzy clustering method for heteroskedastic time series based on GARCH models. Chen et al. (2021) developed a robust clustering algorithm for time series, where the distance measure is only based on the most central curve, with the centrality defined by the functional data depth (López-Pintado and Romo, 2009). For additional robust approaches, we refer to Banerjee and Dave (2012).

Our goal focuses on developing robust linkages for functional data clustering. The idea arises from Ward's minimum variance method, or Ward's linkage (Ward Jr, 1963), where the distance between two clusters is defined as the increase in the SSE when they are merged. In a graphical version, the linkage is the increment of diameter. We refer the readers to (Milligan, 1979; Strauss and Von Maltitz, 2017; Sharma et al., 2019; Großwendt et al., 2019) for further details and related works. Inspired by Ward's linkage, we propose functional Ward's linkage, where the distance of two clusters of curves is defined as the increased width of the band delimited by the merged clusters. To deal with the sensitivity of Ward's linkage to outliers, the two proposed functional Ward's linkages rely exclusively on the most central curves in a cluster. We rank the curves in a cluster from the center outward by leveraging magnitude-shape outlyingness measures (Dai and Genton, 2018b, 2019) and the modified band depth López-Pintado and Romo (2009); Sun and Genton (2011). By selecting the curves with high centrality values, the proposed linkages reduce the impact caused by contamination and outliers. The proposed linkages are applied to EEG data clustering tasks. Due to the influence of the skull and scalp on the transmission of electrical signals, the signals collected from the scalp are often contaminated with substantial noises. Moreover, external disturbances, such as those from experimental instruments and the environment, can lead to the contamination of certain channel data.

The rest of the paper is organized as follows. Section 2 introduces Ward's linkage and proposes two robust functional Ward's linkages. In Section 3, the performance of the proposed linkages is evaluated through two simulation experiments. The applications to the EEG data are presented in Section 4, and we conclude the paper in Section 5.

2 Methodology

2.1 Functional Ward's Linkage

Conventional Ward's linkage defines the squared distance between two clusters, denoted as \mathbf{C}_1 and \mathbf{C}_2 , as the increase in the SSE when they are merged into a new cluster $\mathbf{C}_1 \cup \mathbf{C}_2$:

$$D^2(\mathbf{C}_1, \mathbf{C}_2) = \text{SSE}(\mathbf{C}_1 \cup \mathbf{C}_2) - \text{SSE}(\mathbf{C}_1) - \text{SSE}(\mathbf{C}_2).$$

Intuitively, the SSE represents the diameter of clusters:

$$\text{SSE}(\mathbf{C}_i) = \text{dia}(\mathbf{C}_i) = \sum_{j=1}^{|\mathbf{C}_i|} (y_{ij} - \bar{y}_i)^\top (y_{ij} - \bar{y}_i) = |\mathbf{C}_i| \bar{d}_{\mathbf{C}_i}, i = 1, 2,$$

where $\text{dia}(\cdot)$ denotes the diameter, y_{ij} represents the j -th sample in cluster \mathbf{C}_i , \bar{y}_i is the centroid of cluster \mathbf{C}_i , $\bar{d}_{\mathbf{C}_i}$ is the average squared distance to the centroid, and $|\cdot|$ is the number of elements in A cluster. Then,

$$D^2(\mathbf{C}_1, \mathbf{C}_2) = |\mathbf{C}_1 \cup \mathbf{C}_2| \bar{d}_{\mathbf{C}_1 \cup \mathbf{C}_2} - |\mathbf{C}_1| \bar{d}_{\mathbf{C}_1} - |\mathbf{C}_2| \bar{d}_{\mathbf{C}_2}. \quad (1)$$

This implies that merging clusters with higher homogeneity results in a smaller cluster diameter compared to merging clusters with lower homogeneity.

Analogous to the concept of diameter in conventional Ward's linkage in (1), we define the functional Ward's linkage by using the increase in the width of the band to measure the distance between two clusters of functional data \mathbf{C}_1 and \mathbf{C}_2 . We suppose the width of the band delimited by clusters with high homogeneity is narrower:

$$D^2(\mathbf{C}_1, \mathbf{C}_2) = |\mathbf{C}_1 \cup \mathbf{C}_2| \cdot W(B(\mathbf{C}_1 \cup \mathbf{C}_2)) - |\mathbf{C}_1| \cdot W(B(\mathbf{C}_1)) - |\mathbf{C}_2| \cdot W(B(\mathbf{C}_2)), \quad (2)$$

where $W(\cdot)$ denotes the average width of the band, and $B(\cdot)$ represents the band delimited by a cluster of curves:

$$B(\{y_1(t), y_2(t), \dots, y_n(t)\}) = \{(t, x(t)) : t \in \mathcal{I}, \min_{r=1, \dots, n} y_r(t) \leq x(t) \leq \max_{r=1, \dots, n} y_r(t)\}.$$

Specifically, the average width of the band is computed by the area of the band divided by the Lebesgue measure $\lambda(\cdot)$ on \mathcal{I} :

$$W\{B(\{y_1(t), y_2(t), \dots, y_n(t)\})\} = \frac{1}{\lambda(\mathcal{I})} \int_{t \in \mathcal{I}} (\max_{r=1, \dots, n} y_r(t) - \min_{r=1, \dots, n} y_r(t)) dt.$$

Figure 1 illustrates both conventional and functional Ward's linkage, showing that the distance between \mathbf{C}_1 and \mathbf{C}_2 is smaller than that between \mathbf{C}_1 and \mathbf{C}_3 .

To address the limitations of conventional Ward's linkage in handling outliers and contamination, the two proposed functional Ward's linkages rely exclusively on the most central curves in each cluster. Section 2.2 describes the two methods used to select these most central curves.

2.2 Two Robust Linkages

The two proposed linkages, magnitude-shape outlyingness linkage (MS-linkage) and band depth linkage (BD-linkage), are based on the MS-plot (Dai and Genton, 2018b) and modified band depth (MBD) (López-Pintado and Romo, 2009; Sun and Genton, 2011), respectively.

2.2.1 MS-Plot and MBD

The MS-plot measures both magnitude and shape outlyingness of functional data. It is constructed using directional outlyingness, which extends the conventional concept of outlyingness by incorporating direction. This approach recognizes that the direction of outlyingness is crucial for characterizing the centrality of functional data.

Suppose $y(t)$ is a stochastic process defined on a domain \mathcal{I} with distribution $F_{y(t)}$. Let $d(y(t), F_{y(t)}) : \mathbb{R} \rightarrow [0, 1]$ be a statistical depth function (Zuo and Serfling, 2000; Serfling, 2006) for $y(t)$ with respect to $F_{y(t)}$, and $v(t) =$

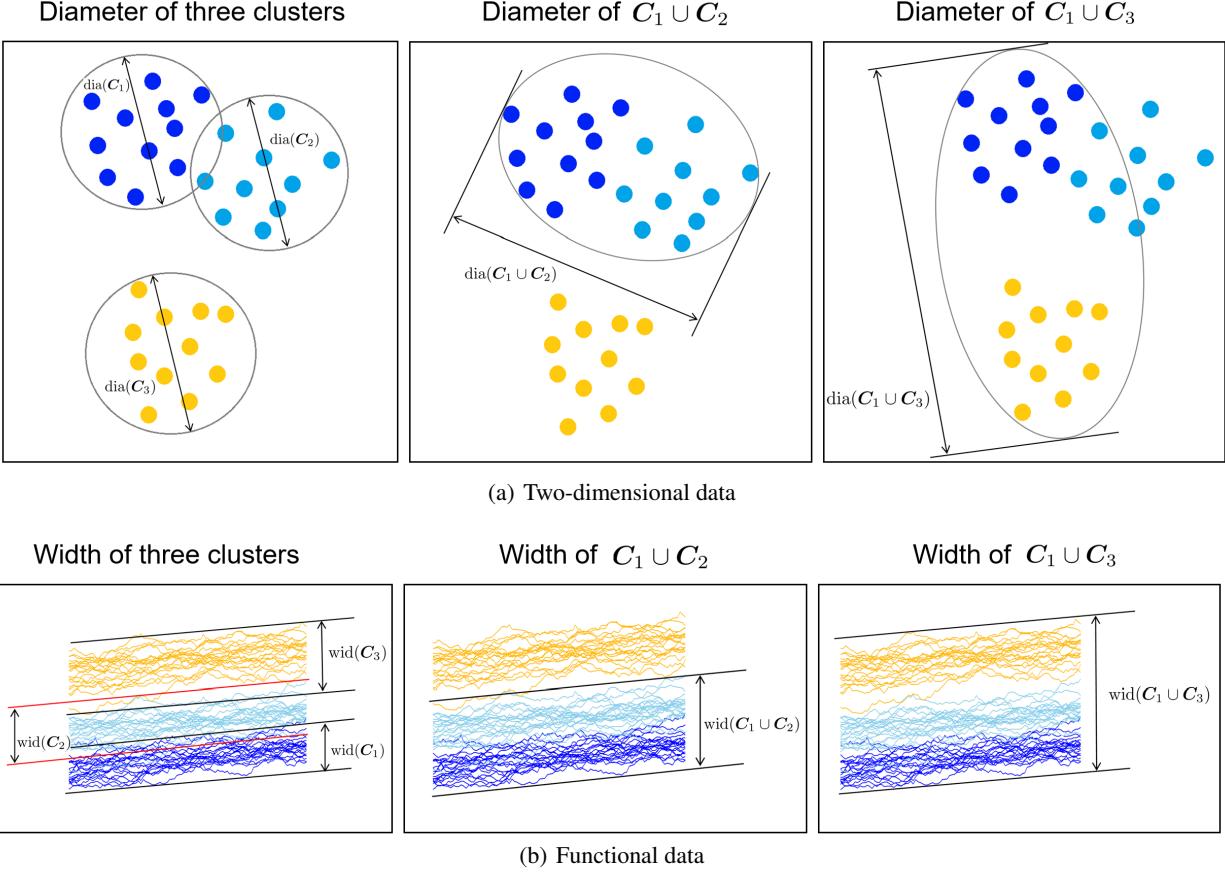


Figure 1: The Ward's linkage (a) and functional Ward's linkage.

$\{y(t) - z(t)\}/\|y(t) - z(t)\|$ is the spatial sign Möttönen and Oja (1995) of $\{y(t) - z(t)\}$, $z(t)$ denotes the unique median of the distribution $F_{y(t)}$ with respect to $d(y(t), F_{y(t)})$, and $\|\cdot\|$ denotes the L_2 norm. Then, the directional outlyingness is defined as:

$$\mathbf{O}(y(t), F_{y(t)}) = \{1/d(y(t), F_{y(t)}) - 1\} \cdot v(t),$$

and the two components of directional outlyingness are defined as:

1. Mean directional outlyingness (**MO**, or magnitude outlyingness),

$$\mathbf{MO}(y(t), F_{y(t)}) = \int_{\mathcal{I}} \mathbf{O}(y(t), F_y) w(t) dt;$$

2. Functional directional outlyingness (**VO**, or shape outlyingness),

$$\mathbf{VO}(y(t), F_{y(t)}) = \int_{\mathcal{I}} \|\mathbf{O}(y(t), F_{y(t)}) - \mathbf{MO}(y(t), F_{y(t)})\|^2 w(t) dt,$$

where $w(t)$ is a weight function defined on \mathcal{I} . Thus, the MS-plot is a scatter plot of points $(\mathbf{MO}, \mathbf{VO})^\top$ for a group of functional data.

MBD provides a center-outward ordering of the functional data. For n curves $\{y_1(t), \dots, y_n(t)\}$, let

$$\text{MBD}_n^{(j)}(y_i) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_j \leq n} \lambda_r \{A(y_i(t); y_{i_1}(t), \dots, y_{i_j}(t))\}$$

represent the proportion of $t \in \mathcal{I}$ that $y_i(t)$ is contained by $B(\{y_{i_1}(t), \dots, y_{i_J}(t)\})$, where

$$A(y_i(t); y_{i_1}(t), \dots, y_{i_J}(t)) = \{t : \min_{r=i_1, \dots, i_J} y_r(t) \leq y_i(t) \leq \max_{r=i_1, \dots, i_J} y_r(t)\},$$

and

$$\lambda_r(y_i(t)) = \frac{\lambda\{A(y_i(t))\}}{\lambda(\mathcal{I})}.$$

Then, the MBD is defined as:

$$\text{MBD}_{n,J}(y_i) = \sum_{j=2}^J \text{MBD}_n^{(j)}(y_i),$$

where J is a fixed value with $2 \leq J \leq n$.

This implies that a higher MBD value indicates a more central position for the curve $y(t)$, based on the fraction of bands containing it. Although the number of curves determining a band, j , can be any integer between 2 and J , the order of curves induced by band depth is sensitive to J . To reduce the computational cost, we choose $J = 2$. Sun and Genton (2011) proposed a graphical tool, called the functional boxplots, for visualizing functional data using MBD. Functional boxplot provides three key descriptive statistics: the envelope band of the 50% central region, the median curve, and the maximum non-outlying envelope. Additionally, outliers can be detected in a functional boxplot using the 1.5 times the 50% central region empirical rule. Extensions of functional boxplots include the surface boxplot Genton et al. (2014), multivariate functional boxplots (Dai and Genton, 2018a), trajectory functional boxplots (Yao et al., 2020), and sparse functional boxplots (Qu and Genton, 2022).

2.2.2 MS-Linkage

To identify the most central curves in a cluster determined by MS, denoted as $\mathbf{C}^* = \{y_{i_1}(t), y_{i_2}(t), \dots, y_{i_k}(t)\} \subseteq \mathbf{C} = \{y_1(t), y_2(t), \dots, y_n(t)\}$, the MS-Linkage focuses on the points in the bottom middle region of the corresponding MS-plot, which equals to the most central points in the union of the MS-plot $(\mathbf{MO}, \mathbf{VO})^\top$ and its symmetric counterpart $(\mathbf{MO}, -\mathbf{VO})^\top$. We apply the method proposed in Agarwal et al. (2022), which is capable of handling both non-Gaussian and non-convex data, to calculate the two-dimensional (2D) τ -quantile envelope of $(\mathbf{MO}, \mathbf{VO})^\top \cup (\mathbf{MO}, -\mathbf{VO})^\top$ and to identify the points that lie within it. Then, we replace all clusters within $B(\cdot)$ in (2) with their most central curves, denoted by $(\cdot)^*$, and obtain the MS-linkage:

$$D_{\text{MS}}^2(\mathbf{C}_1, \mathbf{C}_2) = |\mathbf{C}_1 \cup \mathbf{C}_2| \cdot W\{B((\mathbf{C}_1 \cup \mathbf{C}_2)^*)\} - |\mathbf{C}_1| \cdot W(B(\mathbf{C}_1^*)) - |\mathbf{C}_2| \cdot W(B(\mathbf{C}_2^*)). \quad (3)$$

Figure 2 illustrates the MS-linkage. The top two rows show an example with three clusters: \mathbf{C}_1 (blue) and \mathbf{C}_2 (skyblue) are similar, while \mathbf{C}_3 (yellow) is positioned far from \mathbf{C}_1 . The distance between \mathbf{C}_1 and \mathbf{C}_2 , represented by the width of the central band shown in Figure 2(c), is smaller than the distance between \mathbf{C}_1 and \mathbf{C}_3 , as indicated by the width of the central band in Figure 2(g). In the bottom two rows, we also illustrate the case in which each sample curve has a 20% probability of being contaminated by an outlier (see model 2 in Section 3). When only the most central curves are used, the width of the band delimited by these central curves shows only a slight difference (Figure 2(c, g) vs. Figure 2(j, n)). In contrast, the width of the band delimited by all curves exhibits a significant change (2(d, h) vs. Figure 2(k, o)). The MS-linkage effectively reduces the impact of the outliers.

2.3 BD-Linkage

To select the most central curves of a cluster determined by BD, denoted as $\mathbf{C}^{**} = \{y_{i_1}(t), y_{i_2}(t), \dots, y_{i_k}(t)\} \subseteq \mathbf{C} = \{y_1(t), y_2(t), \dots, y_n(t)\}$, the BD-linkage focuses on curves with large MBD values. We compute the MBD of each curve, yielding $\{\text{MBD}(y_1), \dots, \text{MBD}(y_n)\}$. The most central curves are then defined as $\mathbf{C}^{**} = \{y_{i_r} : \text{MBD}(y_{i_r}) \geq \text{MBD}_\tau\}$, where MBD_τ represents the τ -quantile of $\{\text{MBD}(y_1), \dots, \text{MBD}(y_n)\}$. Replacing all the clusters within $B(\cdot)$ in (2) with their most central curves $(\cdot)^{**}$, we obtain the BD-linkage:

$$D_{\text{BD}}^2(\mathbf{C}_1, \mathbf{C}_2) = |\mathbf{C}_1 \cup \mathbf{C}_2| \cdot W\{B((\mathbf{C}_1 \cup \mathbf{C}_2)^{**})\} - |\mathbf{C}_1| \cdot W(B(\mathbf{C}_1^{**})) - |\mathbf{C}_2| \cdot W(B(\mathbf{C}_2^{**})). \quad (4)$$

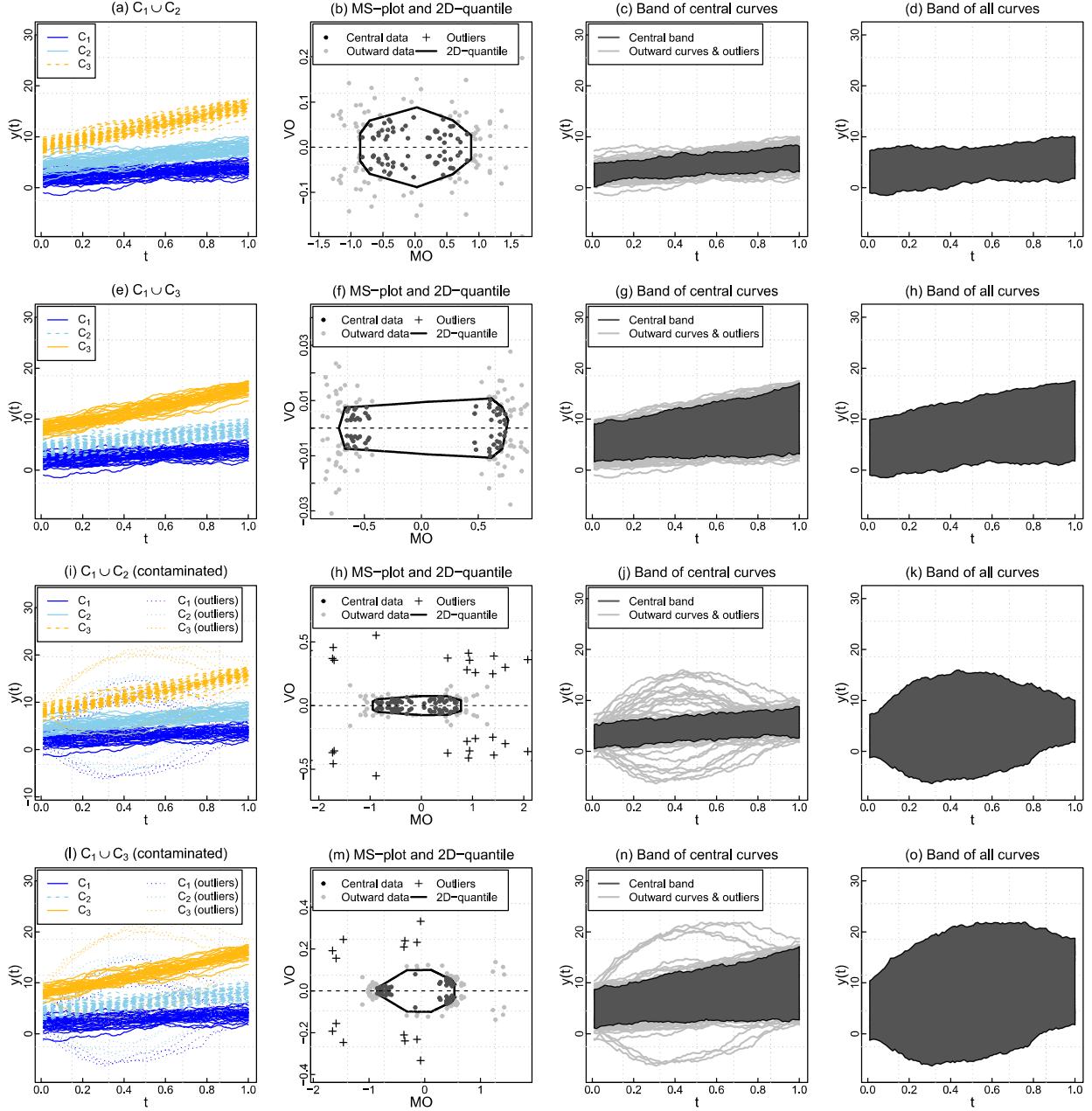


Figure 2: Illustration of the MS linkage. First row: the linkage of C_1 and C_2 ; second row: the linkage of C_1 and C_3 ; third row: the linkage of C_1 and C_2 with outliers; fourth row: the linkage of C_1 and C_3 with outliers. First column: the curves in each cluster; second column: the MS-plot $(MO, VO)^T \cup (MO, -VO)^T$ with the most central data selected by the 2D-quantile; third column: the band constructed by the most central curves; fourth column: the band constructed by all curves. The quantile level τ used is 0.5.

Figure 3 illustrates the BD-linkage using the same data as in Figure 2 using functional bospplots. In the first row, the distance between C_1 and C_2 , represented by the width of the central band shown in Figure 3(b), is smaller than the distance between C_1 and C_3 , represented by the width of the central band shown in Figure 3(d). The bottom row depicts the case where each sample curve has a 20% probability of being contaminated by an outlier. When only the

most central curves are used, the width of the band delimited by the most central curves shows only a slight difference (Figure 3(b, d) vs. Figure 2(f, h)). However, the width of the band delimited by all curves (i.e., the band delimited by the red dashed lines) changes significantly compared to the pink bands in Figure 2(f, h). This demonstrates that BD-linkage effectively reduces the impact of outliers.

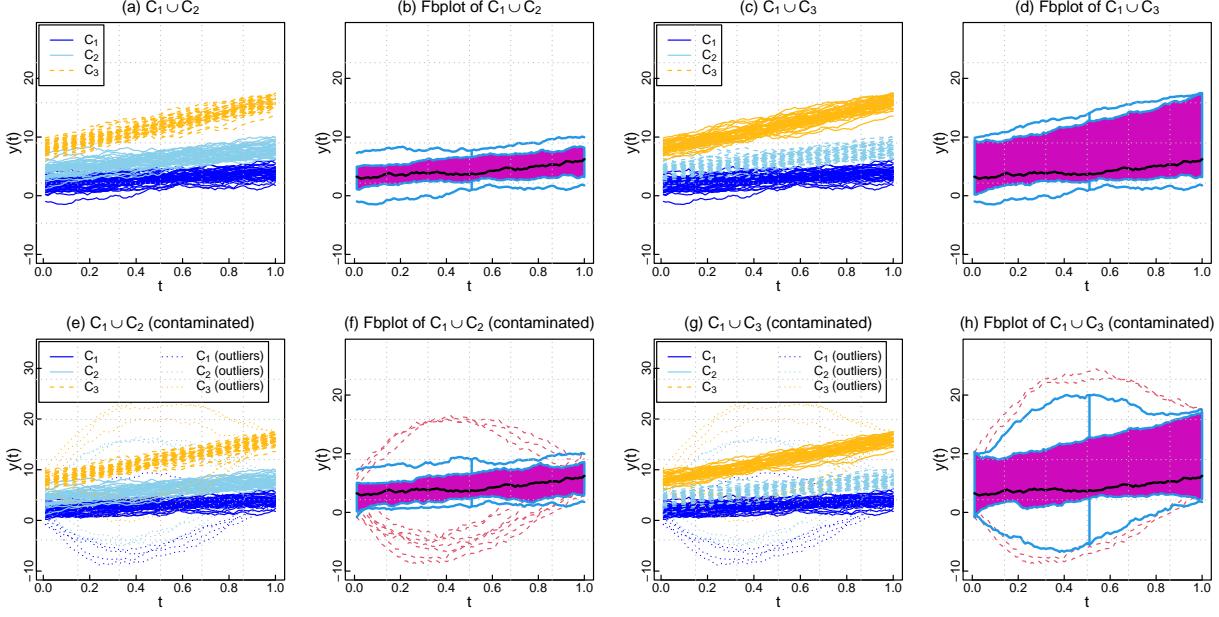


Figure 3: The illustration of the BD-linkage. Top row: no outliers; bottom row: outliers are included. First column: curves and outliers in C_1 and C_2 ; second column: the functional boxplot of $C_1 \cup C_2$; third column: curves and outliers in C_1 and C_3 ; fourth column: the functional boxplot of $C_1 \cup C_3$. The percentage of the most central curves τ is 0.5. In functional boxplot, the black curve denotes the functional median, the magenta band denotes the central region, the skyblue band denote the maximum non-outlying envelope, and the red dashed curves denote the detected outliers.

2.4 Clustering Algorithm

Suppose we have m initial clusters, the hierarchical clustering algorithm proceeds as follows:

Step 1: Let $\{C_1, C_2, \dots, C_m\}$ be the initial clusters.

Step 2: Compute the distance matrix based on linkages (3) or (4).

Step 3: Identify the two clusters with the smallest distance and replace them with a newly merged single cluster.

Step 4: Repeat Steps 2 and 3 until the number of clusters is reduced to 1 or reaches a prespecified number of clusters.

Algorithm 1 provides the pseudocode and summarizes the procedure for hierarchical clustering.

Remarks.

1. Sample sizes issue: If the number of curves is too small to reliably select the most central curves, one can manually set the number of curves in the initial clusters to a relatively large value, or consider alternative approaches. Further details are provided in Section 5.
2. Determining the number of clusters: The predefined number of clusters, p , can be chosen based on domain expertise or determined using data-driven methods such as the elbow method, the Calinski-Harabasz index (Caliński and Harabasz, 1974), or the approaches described in Chapter 17 of Gan et al. (2020).
3. Ratio of the most central curves: The clustering results depend on the ratio τ , we discuss this issue in Section 3.3.

Algorithm 1 : hierarchical clustering

Input: Initial clusters: $C = \{C_1, C_2, \dots, C_m\}$, prespecified number of clusters: p (optional).

```
for k = 1 to m - 1, do
    if length(C) == p then                                # Number of clusters reaches p
        break
    end if
    Compute distance matrix D with entry between  $C_i$  and  $C_j$ :  $D_{ij} = D_{\text{MS}}^2(C_i, C_j)$  or  $D_{\text{BD}}^2(C_i, C_j)$ 
     $(i_k, j_k) = \arg\min_{(i,j)} D_{ij}$                       # Find the closest clusters
     $C_{\text{new}} = C_{i_k} \cup C_{j_k}$                       # Merge the closest clusters
     $D^{\text{new}} = D \setminus \{D_{i_k \cdot} \cup D_{j_k \cdot} \cup D_{\cdot i_k} \cup D_{\cdot j_k}\}$       # Delete rows and columns  $i_k, j_k$ 
    for j = 1 to m - p - 1. do
         $D_{(m-k)j}^{\text{new}} = D_{j(m-k)}^{\text{new}} = D_{\text{MS}}^2(C_{\text{new}}, C_j)$  or  $D_{\text{BD}}^2(C_{\text{new}}, C_j)$       # Compute new distances
    end for
     $D = D^{\text{new}}$ ;  $C = (C \setminus \{C_{i_k}, C_{j_k}\}) \cup C_{\text{new}}$                                 # New matrix D and new clusters
end for
```

Output: C

3 Simulations

3.1 Experiment Design

For the two experiments, we consider 20 initial clusters, each containing 30 curves. There are $p = 4$ true clusters (ground truth), where each true cluster consists of 5 initial clusters. Each curve has a probability of being contaminated with an outlier, and we choose three contamination rate $c = 0.1, 0.15, 0.2$. The data is sampled at $T = 200$ time points over the interval $\mathcal{I} = [0, 1]$.

3.1.1 Experiment 1: Outlier Models

In the first experiment, we consider the outlier models used in López-Pintado and Romo (2009); Dai and Genton (2018b, 2019). Details of the models are described as follow, with illustration in Figure 4:

- **Model 1**

Main model: $y(t) = k + 2kt + e_1(t)$,
Contamination model: $y(t) = k + 2kt + 8U + e_1(t)$ with contamination probability c ,
for $0 \leq t \leq 1$, where $k = 1, \dots, 4$ represents the k -th true cluster, $e_1(t)$ is a Gaussian process with zero mean and covariance function $\gamma(s, t) = \exp\{-|s - t|\}$, and U takes values -1 and 1 with equal probability.
The contaminating curves shift up and down from the main model.

- **Model 2**

Main model: $y(t) = k + 2kt + e_1(t)$,
Contamination model: $y(t) = k + 2kt + Ug(t) + e_1(t)$ with contamination probability c ,
for $0 \leq t \leq 1$, where $g(t) = 60t^{1.5}(1-t)$. The contaminating curves alter the shape of the main model.

- **Model 3**

Main model: $y(t) = k + 2kt + e_1(t)$,
Contamination model: $y(t) = k + 2kt + e_2(t)$ with contamination probability c ,
for $0 \leq t \leq 1$, where $e_2(t)$ is a Gaussian process with zero mean and covariance function $\gamma^*(s, t) = 12\exp\{-|s - t|^{0.1}\}$. The contaminating curves share the same trend as the main model, but exhibit distinct covariance structures.

For model 3, we cluster the data in the spectral domain using smoothed log-periodograms to represent the spectral features of the time series. Although raw periodograms are unbiased estimators of the spectral density functions (SDFs), they suffer from their roughness and high variability. As a result, smoothing is required to produce consistent estimators. In this paper, the smoothing bandwidth is selected automatically using the gamma generalized cross-validation (gamma-GCV) method in Ombao et al. (2001). The log-periodograms introduce bias when estimating log-SDFs due to Jensen's inequality. To correct for this, we apply a bias adjustment by adding the Euler-Mascheroni constant $\gamma = 0.57721$ to the smoothed log-periodograms (Wahba, 1980).

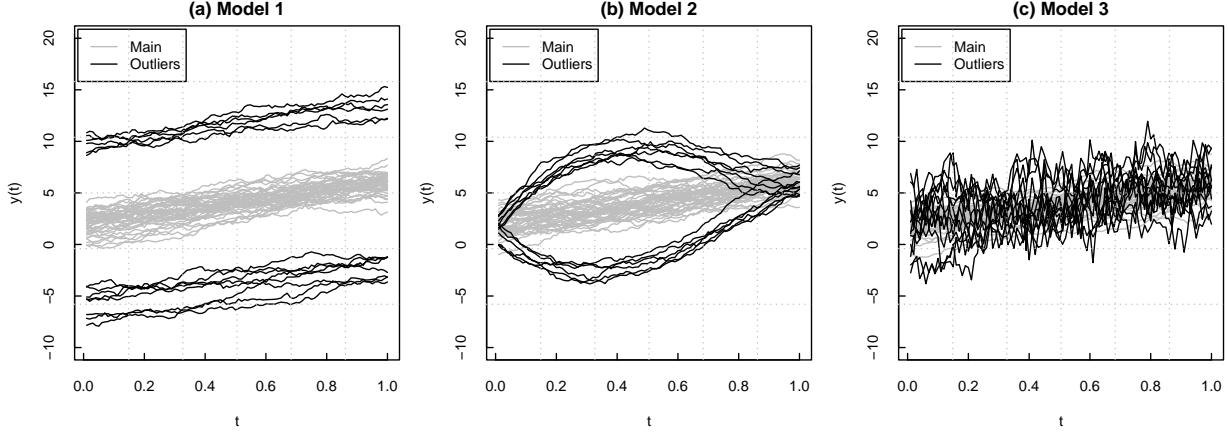


Figure 4: Illustrative examples of the three outlier models. Each plot contains 60 curves, with contamination rate $c = 0.2$.

3.1.2 Experiment 2: EEG Artifacts

In the second experiment, we simulate EEG data with artifacts and cluster the data in the frequency domain. EEG is widely used in both laboratory and clinical applications due to its non-invasiveness, affordability, portability, and high temporal resolution. We use a mixture of second-order autoregressive ($AR(2)$) models to capture the oscillatory activities of EEG signals:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t,$$

where w_t is the white-noise process with variance σ_w^2 . The characteristic polynomial of the process is $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$, and the roots z_1 and z_2 are complex conjugates. The magnitude $|z_1| = |z_2| = 1/r > 1$ ensures the causality. The AR coefficients ϕ_1 and ϕ_2 determine the peak frequency ω_0 in the SDF:

$$\phi_1 = \frac{2 \cos(2\pi\omega_0/F)}{M} \quad \text{and} \quad \phi_2 = -\frac{1}{M^2},$$

where F is the sampling frequency in Hertz, set to 1000 Hz in this study. M controls the peak sharpness, with the peak becomes narrower as $M \rightarrow 1^+$.

In EEG data analysis, clinically relevant frequency bands are commonly used to characterize brain activity. To simulate EEG signals, we generate five latent $AR(2)$ sources, each corresponding to a specific frequency band, by setting appropriate AR coefficients. The detail about the five latent sources is shown in Table 1. To illustrate the oscillatory patterns within each frequency band, Figure 5(a) depicts the log-SDFs and examples of the generated time series, where the oscillations become more rapid as the peak frequency approaches 50 Hz.

The EEG data, simulated by mixture $AR(2)$ models, are the linear combination of the five latent sources. The log-SDFs of the four true clusters are:

$$f_1 = \log \left(\frac{4}{5} \tilde{f}_1 + \frac{1}{10} \tilde{f}_2 \right), \quad f_2 = \log \left(\frac{3}{5} \tilde{f}_1 + \frac{1}{10} \tilde{f}_3 \right),$$

Table 1: Five latent AR(2) sources used for EEG signals simulation.

Frequency Band	Frequency (Hz)	AR Coefficients (ϕ_1, ϕ_2)
Delta	0 – 4 Hz	(0.8, 0.1)
Theta	4 – 8 Hz	(0.9, -0.9)
Alpha	8 – 16 Hz	(-0.1, -0.9)
Beta	16 – 32 Hz	(-0.9, -0.9)
Gamma	32 – 50 Hz	(-0.8, -0.1)

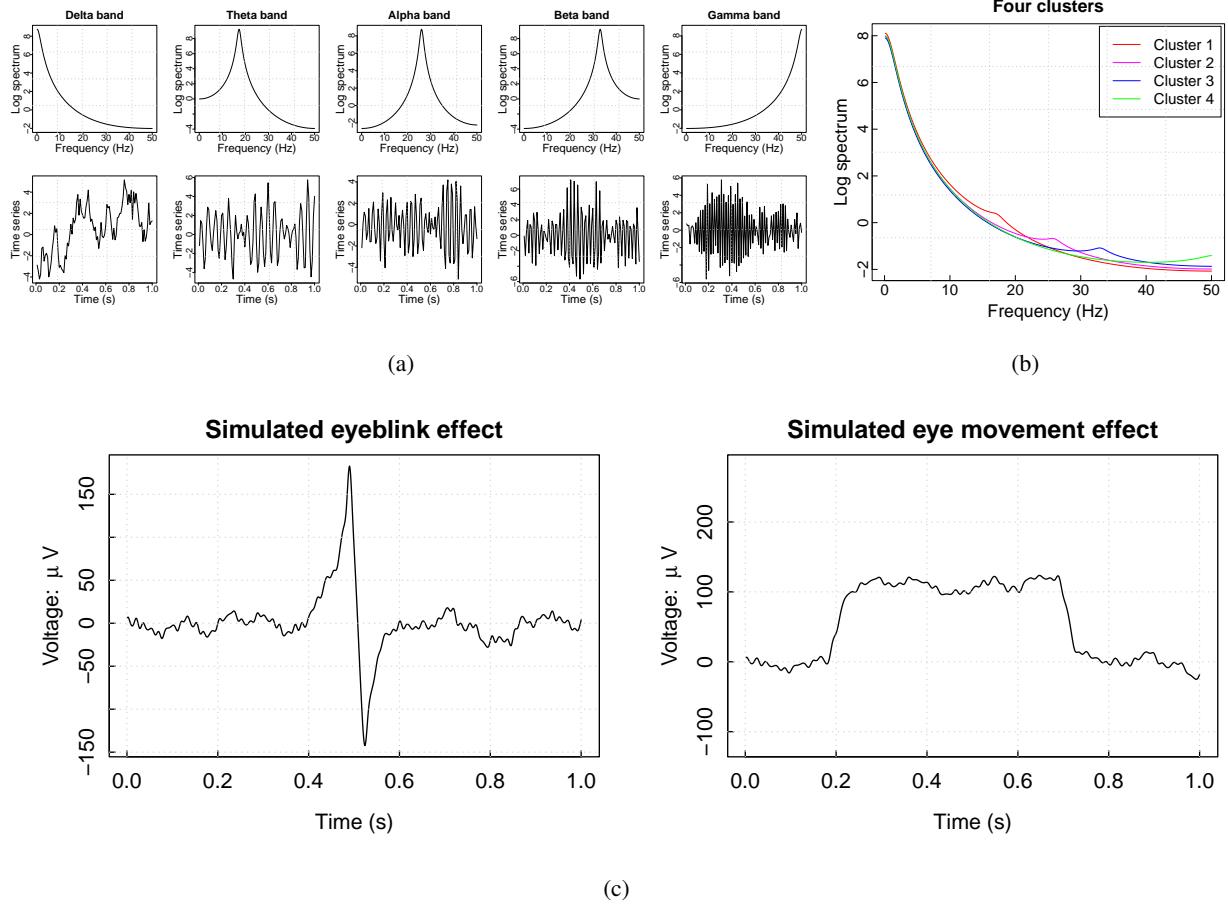


Figure 5: (a) The log-SDFs of the 5 latent sources (top) and the corresponding realizations (bottom). (b) The true log-SDFs of the four clusters. (c) The simulated artifacts.

$$f_3 = \log \left(\frac{2}{5} \tilde{f}_1 + \frac{1}{10} \tilde{f}_4 \right), \quad f_4 = \log \left(\frac{1}{5} \tilde{f}_1 + \frac{1}{10} \tilde{f}_5 \right),$$

where $\tilde{f}_i(\omega)$, $i = 1, \dots, 5$ are the log-SDFs of the five latent AR(2) sources. As illustrated in Figure 5(b), there are only slight differences among the four true clusters, which makes the clustering task more challenging.

We consider two types of clinical artifacts as contamination: eye-blink and eye-movement effects. EEG can be contaminated in the frequency or time domain by these artifacts, which result from internal sources such as physiological activities and subject movement, as well as external sources including environmental interference, equipment malfunction, and the movement of electrodes and cables. In our setting, eye-blanks are simulated by the

difference of two gamma functions, while eye-movements are generated by adding a wide peak to the EEG signal. The shape of the simulated artifacts is based on clinical EEG data in Mansor et al. (2011); Abo-Zahhad et al. (2015), and Kaya (2019); Sazgar et al. (2019); Mumtaz et al. (2021) provide comprehensive introduction of different types of EEG artifacts. The simulated artifacts are shown in Figure 5(c).

3.2 Measures of Quality

To evaluate the performance of the proposed linkages in discovering the correct labels in a more systematic framework, we use two popular measures: the adjusted Rand index (ARI) (Vinh et al., 2009) and SIM index in R package `TSClust` (Montero and Vilar, 2015), which are commonly used in the clustering evaluation literature.

The adjust Rand index is defined as

$$\text{ARI} = \frac{\sum_{i=0}^1 \sum_{j=0}^1 \binom{n_{ij}}{2} - [\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}] / \binom{m}{2}}{\frac{1}{2} [\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}] - [\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}] / \binom{m}{2}}.$$

To calculate the ARI, we compute the 2×2 confusion table, consisting of the following four cells:

- n_{11} : the number of observation pairs where both observations are comembers in both clusterings.
- n_{10} : the number of observation pairs where the observations are comembers in the one clustering but not in the other.
- n_{01} : the number of observation pairs where the observations are comembers in the second clustering but not in the other.
- n_{00} : the number of observation pairs where neither pair are comembers in either clustering results.

We also use the SIM index, which measures the amount of agreement between the ground truth $\mathcal{G} = \{\mathbf{G}_1, \dots, \mathbf{G}_p\}$ and the clustering results $\mathcal{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_p\}$. It is defined as:

$$\text{SIM}(\mathcal{G}, \mathcal{C}) = \frac{1}{p} \sum_{i=1}^p \max_{1 \leq j \leq p} \text{Sim}(\mathbf{G}_i, \mathbf{C}_j),$$

where $\text{Sim}(\mathbf{G}_i, \mathbf{C}_j) = \frac{|\mathbf{G}_i \cap \mathbf{C}_j|}{|\mathbf{G}_i| + |\mathbf{C}_j|}$, and $|\cdot|$ denotes the number of elements in the set. Both the ARI and SIM range from 0 to 1, with larger value indicating better clustering results.

3.3 Competitive Approaches and Simulation Results

For each experiment, we run 100 simulations and compute the averaged ARI and SIM index. In addition to the two proposed methods, we also include the following competitive approaches:

- Ordinary Ward's linkage;
- TCLUS: A trimming approach for robust clustering (Fritz et al., 2012), implemented in R package `tclust`. The trimming rate selected is $1 - \tau$, which equals to the ratio of the most central curves in the proposed linkages.
- HSM: Hierarchical spectral merger algorithm (Euán et al., 2018).

Note that HSM is only applied in **Experiment 2**, since the distance metric in this method is specifically designed for SDFs.

The results that compare the proposed linkages with the competitive ones are presented in Table 2 and 3. In **Experiment 1**, when there is no contamination present, all the approaches achieve high clustering accuracy. However, when different types of contamination are introduced, especially as the contamination rate increases, the proposed

Table 2: Clustering results based on 100 simulations. The mean of the ARI and SIM are reported.

Model	Measure	c	MS	BD	Ward	TCLUST
Type 1	SIM	0.1	0.95	0.99	0.70	0.82
		0.15	0.89	0.96	0.65	0.82
		0.2	0.83	0.92	0.62	0.81
Type 2	SIM	0.1	0.98	0.98	0.79	0.83
		0.15	0.97	0.96	0.69	0.83
		0.2	0.97	0.90	0.65	0.82
Type 3	SIM	0.1	0.82	0.86	0.77	0.59
		0.15	0.81	0.82	0.72	0.58
		0.2	0.81	0.78	0.68	0.58
Type 1	ARI	0.1	0.89	0.98	0.43	0.63
		0.15	0.80	0.91	0.33	0.63
		0.2	0.68	0.81	0.26	0.62
Type 2	ARI	0.1	0.96	0.96	0.79	0.64
		0.15	0.94	0.92	0.40	0.65
		0.2	0.94	0.77	0.31	0.64
Type 3	ARI	0.1	0.64	0.70	0.55	0.27
		0.15	0.64	0.63	0.45	0.27
		0.2	0.64	0.58	0.39	0.28

Table 3: Clustering results based on 100 simulations. The mean of the ARI and SIM are reported.

Model	Measure	c	MS	BD	Ward's	TCLUST	HSM
Eye-blink	SIM	0.1	0.92	0.92	0.91	0.90	0.78
		0.15	0.94	0.85	0.84	0.74	0.76
		0.2	0.90	0.86	0.81	0.71	0.75
Eye-movement	SIM	0.1	0.94	0.92	0.91	0.96	0.80
		0.15	0.94	0.89	0.85	0.93	0.77
		0.2	0.93	0.87	0.82	0.86	0.76
Eye-blink	ARI	0.1	0.88	0.83	0.82	0.83	0.68
		0.15	0.88	0.71	0.70	0.62	0.65
		0.2	0.81	0.70	0.63	0.53	0.64
Eye-movement	ARI	0.1	0.88	0.84	0.85	0.87	0.71
		0.15	0.87	0.76	0.71	0.85	0.66
		0.2	0.87	0.73	0.67	0.77	0.65

methods consistently outperform the competitive approaches. Based on the fact that Model 1 involves purely magnitude contamination, BD-linkage demonstrates its effectiveness in handling this specific kind of outlier. Model 2, which alters the shape of the main model, is better addressed by MS-linkage. In **Experiment 2**, our methods also show robustness against the two types of artifacts. Figure 6 further verifies the robustness by presenting the distance matrix of the initial clusters in **Experiment 1** (Model 2 with $c = 0.1$). The distance matrices of the two proposed methods show much clearer block structures than those of conventional Ward's linkage.

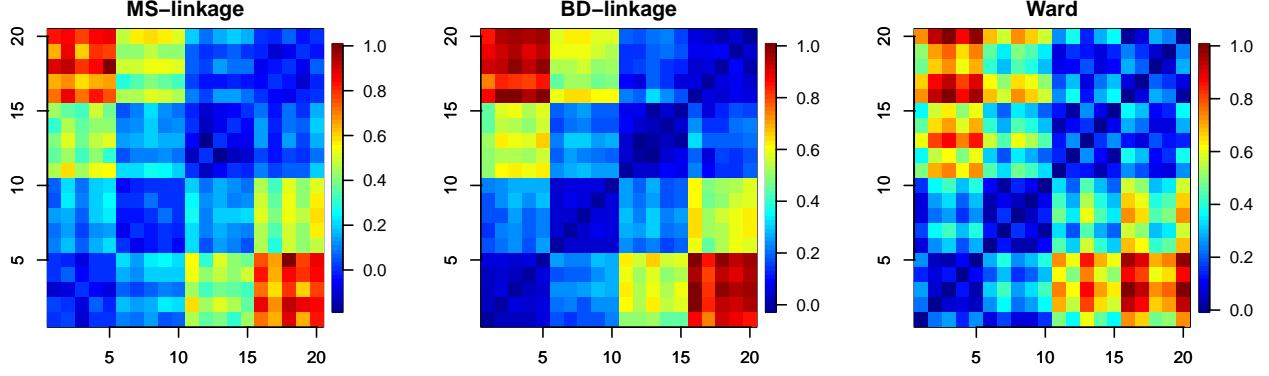


Figure 6: The distance matrix of the initial clusters using SIM.

We also examine how the central curve ratio τ affect the clustering accuracy. By setting $\tau = 0.3, 0.4, 0.5, 0.6$ and 0.7 in model 1 of **Experiment 1**, we obtain Figure 7, in which we plot clustering accuracy against τ . We observe that, the accuracy follows an inverted U-shaped pattern: it initially improves with increasing τ but then declines. The reason is that, larger τ incorporate more central curves, providing additional discriminative information. However, a large τ increases the probability of including outlier curves as central curves. The leftward shift of the optimal τ as c increases further supports this interpretation, demonstrating the trade-off between information utilization and outlier sensitivity.

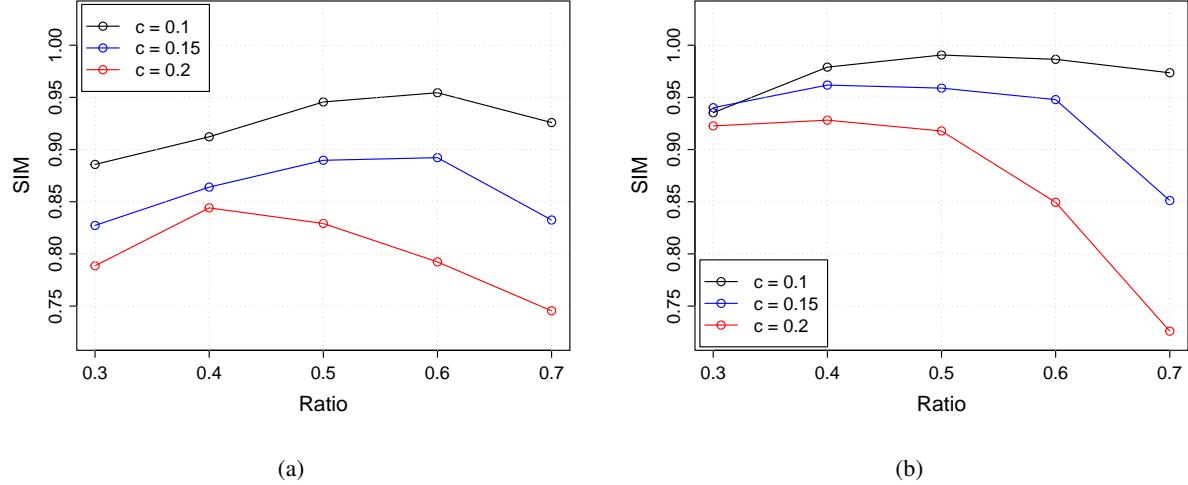


Figure 7: The clustering accuracy v.s. the ratio of the most central curves. (a) MS-linkage and (b) BD-linkage.

4 EEG Data Analysis

In this section, we apply our methods to two EEG data clustering applications. In Section 4.1, we cluster the resting-state EEG data on multiple channels to identify synchronized brain regions, where the true number of clusters p is unknown. In Section 4.2, we present four clustering tasks on single-channel epilepsy EEG data to divide patients into different groups, with the number of clusters p being prespecified. Both applications are analyzed in the frequency domain, and the SDFs are estimated using the same algorithm described in Section 3.

4.1 Resting-state EEG Data Analysis

The first application uses the resting-state EEG data from an experiment in Wu et al. (2014). The goal is to cluster EEG signals from different channels that are spectrally synchronized, i.e., channels that exhibit similar spectral properties. The EEG data was recorded from 256 channels on the brain scalp, as shown in Figure 8(a), with a millisecond resolution (1000 Hz). Of the 256 channels, 62 were excluded due to data quality issues. The total recording time for the EEG data is three minutes and we focus on the first minute and the last minute. Each second of recording is treated as a time series epoch, creating 60 epochs of data for a given minute. We treat the estimated SDFs of the 60 epochs in each channel as an initial cluster, and thus we generate 194 initial clusters. The optimal number of clusters p is selected by the elbow method implemented in the R package `factoextra`. We plot the total within-cluster sum of squares against the numbers of clusters (ranging from 1 to 12). As illustrated in Figure 8(b, c), the “elbow” appears at 4 clusters for both the first and the last minutes.

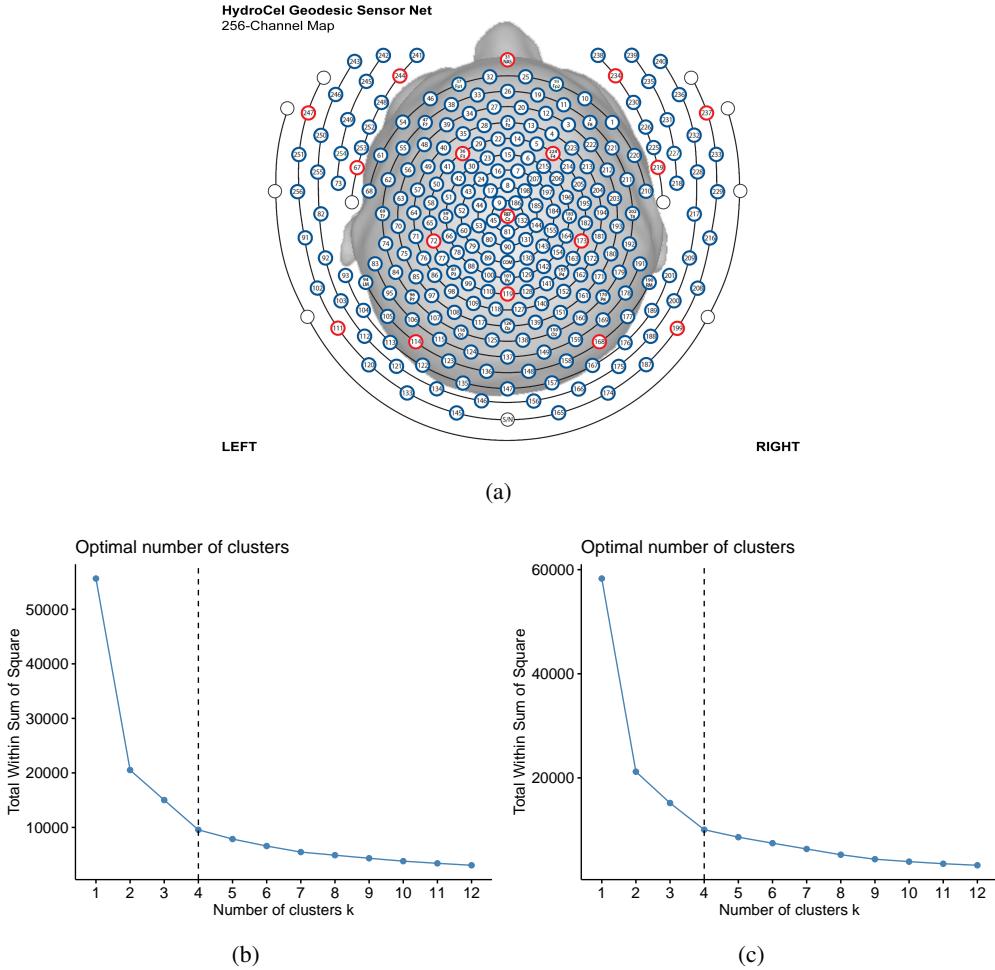


Figure 8: (a) Locations of the 256 channels on the scalp surface. (b) The scree plot of the first minute. (c) The scree plot of the last minute.

We illustrate the clustering results using 2D brain maps and the functional means of each cluster. Figure 9 presents the clustering results of the EEG records from the first minute. Out of the 194 channels, MS-linkage assigns 103, 38, 48, and 5 SDFs to the four clusters, respectively, while BD-linkage assigns 67, 36, 42, and 49 SDFs. For the last minute, as

illustrated in Figure 10, MS-linkage assigns 85, 58, 35, and 6 SDFs to the four clusters, respectively, while BD-linkage assigns 83, 35, 62, and 14 SDFs.

As shown in Figures 9 and 10, our findings align with the conclusions in Maadoliat et al. (2018):

- For both datasets (first and last minutes) and both linkages, the elbow method suggests using $p = 4$ clusters.
- All brain maps show relatively well-separated, spatially homogeneous, and symmetric regions.
- According to the brain maps and the corresponding functional means, the channels in the middle region of the brain exhibit lower density at 0–50 Hz than the front and back regions.
- There is no significant difference between the results in the first minute and the last minute. This is consistent with the fact that the data was recorded from a single subject during resting state.

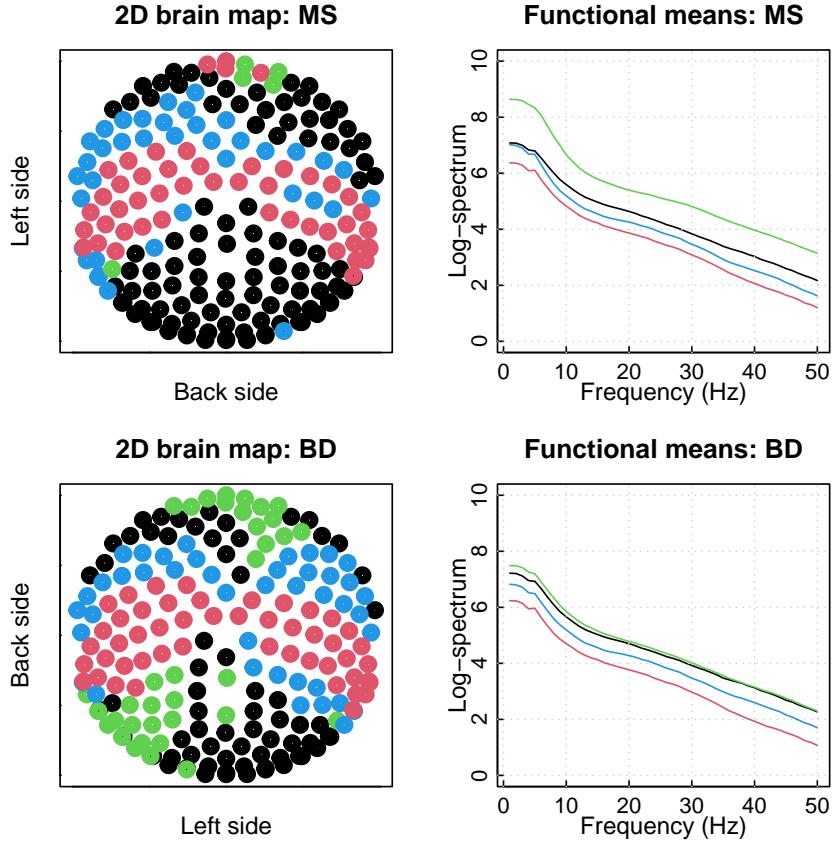


Figure 9: The clustering results of the first minute. Top row: MS-linkage; bottom row: BD-linkage

4.2 Epilepsy EEG data Analysis

In this application, we use the EEG segments described in Andrzejak et al. (2001) to cluster people to different groups. There are five sets of data (denoted as A, B, C, D and E) collected from individuals in different states. Each set contains 100 single-channel EEG segments of length 4098 in 23.6 seconds. The content of the five sets is:

- A: Healthy volunteers with eyes open;
- B: Healthy volunteers with eyes closed;

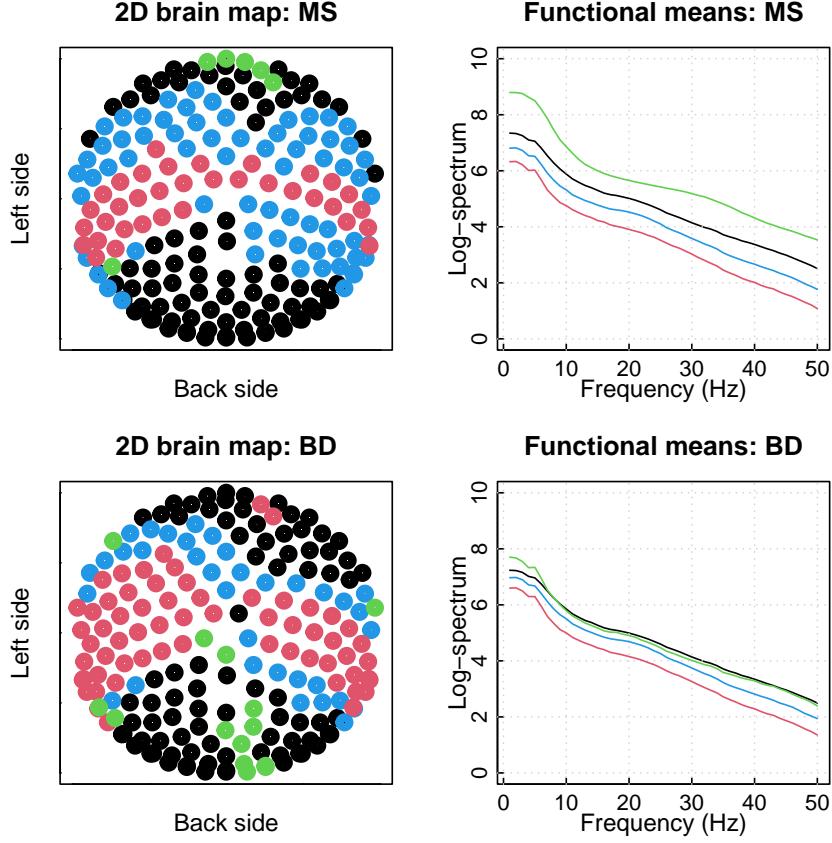


Figure 10: The clustering results of the last minute. Top row: MS-linkage; bottom row: BD-linkage

- C: Epilepsy patients in seizure-free intervals from epileptic hemisphere;
- D: Epilepsy patients in seizure-free intervals from opposite hemisphere;
- E: Epilepsy patients in seizure intervals.

We randomly split the 100 estimated SDFs in each set into five initial clusters, each containing 20 segments. Then, we use four different tasks conducted by Maadooliat et al. (2018), varying the level of complexity, to evaluate the performance of the proposed linkages in clustering different mixtures of healthy segments, epileptic seizure-free segments, and epileptic seizure segments.

- Easy Task (A vs. E): This task aims to distinguish between epileptic seizure patients (E) and healthy volunteers (A). Therefore, we cluster the 10 initial clusters into $p = 2$ true clusters.
- Intermediate Task (D vs. E): This task aims to distinguish between seizure-free epileptic patients (D) and the patients with seizure activity (E). Therefore, we cluster the 10 initial clusters into $p = 2$ true clusters.
- Hard Task (A vs. D): The aim of this task is to cluster the healthy volunteers (A) and the epileptic seizure patients in seizure-free intervals (D). Therefore, we cluster the 10 initial clusters into $p = 2$ true clusters.
- Challenging Task (A vs. D vs. E): This task requires clustering EEG signals from healthy volunteers (A), seizure-free epileptic patients (D), and patients experiencing seizures (E). Therefore, we cluster the 15 initial clusters into $p = 3$ true clusters.

We also compare our methods with two alternative approaches, NCSDE and GLK, as used in Maadoliat et al. (2018), using ARI and SIM as evaluation metrics. The results are presented in Table 4, where the results of the two proposed linkages are the averaged values of the 100 random split. As shown in all tasks, the proposed linkages achieve equal or higher clustering accuracy. Figure 11 shows the functional means of initial clusters in the first random split. We can see that set D and set E are difficult to separate, yet the ARI and SIM for both proposed methods are larger than 0.99.

Table 4: Clustering results based on 100 random split. The mean of the ARI and SIM are reported.

Task	Measures	MS	BD	NCSDE	GLK
Easy	ARI	1	1	0.98	0.36
A vs. E	SIM	1	1	0.99	0.80
Intermediate	ARI	1	0.99	0.88	0.28
(D vs. E)	SIM	1	0.99	0.97	0.75
Hard	ARI	0.85	0.67	0.85	0.43
(A vs. D)	SIM	0.95	0.90	0.96	0.83
Challenging	ARI	0.88	0.61	0.85	0.25
(A vs. D vs. E)	SIM	0.95	0.78	0.95	0.66

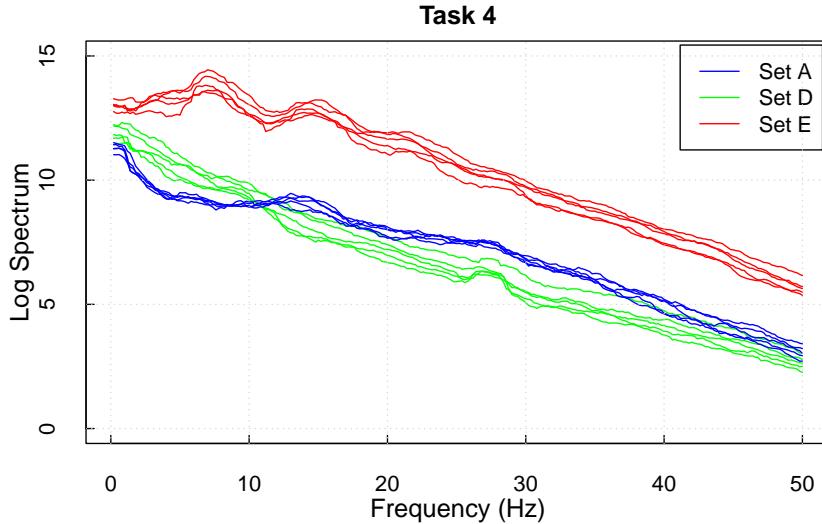


Figure 11: Functional means of each initial clusters in Task 4.

5 Conclusion

In this paper, we proposed two robust functional Ward's linkages, where the distance between two clusters is defined as the increase in the width of the band after merging them. By using magnitude-shape outlyingness and modified band depth, the two proposed linkages are determined solely by the most central curves. In such a way, the linkages reduce the influence of outliers and contamination. The robustness is verified in the simulations by introducing different types of outlier models in either temporal or spectral domains. Applications to two types of EEG data showed that the methods have wide applicability across various fields.

However, the linkages are not free from limitations, and further developments are needed. First, it is challenging to select the most central curves in a cluster when the number of curves is small. Based on data experiments, the BD-linkage requires at least 4 curves to identify the most central ones, while MS-linkage requires at least 12. Therefore, we set the number of curves in an initial cluster to a somewhat large value. One solution to this issue is that, when $|C_1 \cup C_2| \leq 12$, conventional Ward's linkage or the Euclidean distance between the functional medians of the two clusters can be used instead. Another approach is to segment the curve into shorter ones to increase the number of curves in the initial clusters. Second, the MS-linkage incurs a higher computational cost compared to conventional Ward's linkage. The main computational burden arises from computing the MS-plot and generating the 2D quantile contour. A possible solution is to simplify the contour polygon by reducing the number of arcs (e.g., approximating it with a rectangle). Although this may reduce the clustering accuracy, the trade-off between bias and variance can be balanced by incorporating a penalty term on the number of arcs.

Supplementary Materials

The R code to reproduce the results in this paper is available at <https://github.com/tianbochen1/Functional-Ward-s-Linkages>.

Acknowledgments

The authors thank Professor Jennifer Wu and Ralph G. Andrzejak for sharing the EEG data sets. We also appreciate the Researches in the First Affiliated Hospital of USTC for EEG artifacts simulations. This work was supported by the National Natural Science Foundation of China under Grant No.12301326, No.12301171 and No. 12201218; Anhui Provincial Natural Science Foundation under Grant No.2308085QA05 and No.2308085QA04; the University Natural Science Research Project of Anhui Province under Grant No.2023AH050099 and 2023AH033436.

References

- Abo-Zahhad, M., Ahmed, S. M., and Abbas, S. N. (2015). A new EEG acquisition protocol for biometric identification using eye blinking signals. *International Journal of Intelligent Systems and Applications*, 7(6):48.
- Agarwal, G., Tu, W., Sun, Y., and Kong, L. (2022). Flexible quantile contour estimation for multivariate functional data: Beyond convexity. *Computational Statistics & Data Analysis*, 168:107400.
- Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering—a decade review. *Information systems*, 53:16–38.
- Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907.
- Banerjee, A. and Dave, R. N. (2012). Robust clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):29–59.
- Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.
- Chamroukhi, F. and Nguyen, H. D. (2019). Model-based clustering and classification of functional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1298.
- Cheam, A. S. and Fredette, M. (2020). On the importance of similarity characteristics of curve clustering and its applications. *Pattern Recognition Letters*, 135:360–367.

- Cheifetz, N., Noumir, Z., Samé, A., Sandraz, A.-C., Féliers, C., and Heim, V. (2017). Modeling and clustering water demand patterns from real-world smart meter data. *Drinking Water Engineering and Science*, 10(2):75–82.
- Chen, T., Sun, Y., Euan, C., and Ombao, H. (2021). Clustering brain signals: A robust approach using functional data ranking. *Journal of Classification*, 38(3):425–442.
- Chen, T., Sun, Y., and Maadooliat, M. (2020). Collective spectral density estimation and clustering for spatially-correlated data. *Spatial Statistics*, 38:100451.
- Dai, W. and Genton, M. G. (2018a). Functional boxplots for multivariate curves. *Stat*, 7(1):e190.
- Dai, W. and Genton, M. G. (2018b). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4):923–934.
- Dai, W. and Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131:50–65.
- 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., De Giovanni, L., and Massari, R. (2016). Garch-based robust clustering of time series. *Fuzzy Sets and Systems*, 305:1–28.
- Euán, C., Ombao, H., and Ortega, J. (2018). The hierarchical spectral merger algorithm: A new time series clustering procedure. *Journal of Classification*, 35:71–99.
- Fritz, H., García-Escudero, L. A., and Mayo-Iscar, A. (2012). tclust: An r package for a trimming approach to cluster analysis. *Journal of Statistical Software*, 47:1–26.
- Gan, G., Ma, C., and Wu, J. (2020). *Data clustering: theory, algorithms, and applications*. SIAM.
- García-Escudero, L. A. and Gordaliza, A. (2005). A proposal for robust curve clustering. *Journal of classification*, 22(2):185–201.
- Genton, M. G., Johnson, C., Potter, K., Stenchikov, G., and Sun, Y. (2014). Surface boxplots. *Stat*, 3(1):1–11.
- Großwendt, A., Röglin, H., and Schmidt, M. (2019). Analysis of ward’s method. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2939–2957. SIAM.
- Jacques, J. and Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8:231–255.
- Kaya, I. (2019). A brief summary of EEG artifact handling. *Brain-computer interface*, (9).
- Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, 38(11):1857–1874.
- López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734.
- Maadooliat, M., Sun, Y., and Chen, T. (2018). Nonparametric collective spectral density estimation with an application to clustering the brain signals. *Statistics in medicine*, 37(30):4789–4806.
- Mansor, W., Rani, M., and Wahy, N. (2011). *Integrating Neural Signal and Embedded System for Controlling Small Motor*.
- Milligan, G. W. (1979). Ultrametric hierarchical clustering algorithms. *Psychometrika*, 44(3):343–346.
- Montero, P. and Vilar, J. A. (2015). Tsclust: An r package for time series clustering. *Journal of Statistical Software*, 62:1–43.
- Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, 5(2):201–213.

- Mumtaz, W., Rasheed, S., and Irfan, A. (2021). Review of challenges associated with the EEG artifact removal methods. *Biomedical Signal Processing and Control*, 68:102741.
- Ombao, H. C., Raz, J. A., Strawderman, R. L., and Von Sachs, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88(4):1186–1192.
- Qu, Z. and Genton, M. G. (2022). Sparse functional boxplots for multivariate curves. *Journal of Computational and Graphical Statistics*, 31(4):976–989.
- Rivera-García, D., García-Escudero, L. A., Mayo-Iscar, A., and Ortega, J. (2019). Robust clustering for functional data based on trimming and constraints. *Advances in Data Analysis and Classification*, 13(1):201–225.
- Sazgar, M., Young, M. G., Sazgar, M., and Young, M. G. (2019). EEG artifacts. *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees*, pages 149–162.
- Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1.
- Sharma, S., Batra, N., et al. (2019). Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. In *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 568–573. IEEE.
- Strauss, T. and Von Maltitz, M. J. (2017). Generalising ward’s method for use with manhattan distances. *PloS one*, 12(1):e0168288.
- Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*.
- Teeraratkul, T., O’Neill, D., and Lall, S. (2017). Shape-based approach to household electric load curve clustering and prediction. *IEEE Transactions on Smart Grid*, 9(5):5196–5206.
- Vinh, N. X., Epps, J., and Bailey, J. (2009). Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th annual international conference on machine learning*, pages 1073–1080.
- Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75(369):122–132.
- Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its application*, 3(1):257–295.
- Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244.
- Wu, E. H. and Yu, P. L. (2006). Iclus: A robust and scalable clustering model for time series via independent component analysis. *International journal of systems science*, 37(13):987–1001.
- Wu, J., Srinivasan, R., Kaur, A., and Cramer, S. C. (2014). Resting-state cortical connectivity predicts motor skill acquisition. *Neuroimage*, 91:84–90.
- Yao, Z., Dai, W., and G. Genton, M. (2020). Trajectory functional boxplots. *Stat*, 9(1):e289.
- Zhang, M. and Parnell, A. (2023). Review of clustering methods for functional data. *ACM Transactions on Knowledge Discovery from Data*, 17(7):1–34.
- Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics*, pages 461–482.