Robust two-layer partition clustering of sparse multivariate functional data☆

Zhuo Qu[a,∗], Wenlin Dai[b], Marc G. Genton[a]

[a] Statistics Program, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia
[b] Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing, 100872, China

ARTICLE INFO

ABSTRACT

A novel elastic time distance for sparse multivariate functional data is proposed and used to develop a robust distance-based two-layer partition clustering method. With this proposed distance, the new approach not only can detect correct clusters for sparse multivariate functional data under outlier settings but also can detect those outliers that do not belong to any clusters. Classical distance-based clustering methods such as density-based spatial clustering of applications with noise (DBSCAN), agglomerative hierarchical clustering, and *K*-medoids are extended to the sparse multivariate functional case based on the newly-proposed distance. Numerical experiments on simulated data highlight that the performance of the proposed algorithm is superior to the performances of existing model-based and extended distance-based methods. The effectiveness of the proposed approach is demonstrated using Northwest Pacific cyclone tracks data as an example.

1. Introduction

Functional data analysis (FDA, Ramsay and Silverman, 2005) is a branch of statistics that analyzes observations that can be regarded as curves, surfaces, or any object evolving over a continuum. It has wide applications in many fields, such as meteorology, medicine, finance, biology, and language. Real-life examples of the applications of FDA include the analyses of typhoon trajectories (Misumi et al., 2019), teletransmitted electrocardiograph (ECG) traces (Ieva et al., 2011), stock share prices over time (Horváth and Kokoszka, 2012), growth curves of different body parts (Sheehy et al., 2000), CD4 level counts (Yao et al., 2005), and character handwritings (Kneip et al., 2000). With the advances in data collection techniques, huge amounts of functional data are being recorded for various purposes and these data usually show heterogeneity. Hence, clustering these functional observations into homogeneous subgroups is crucial.

Overall, functional clustering can be achieved using four types of methods, according to Jacques and Preda (2014a): 1) raw data methods that involve clustering directly the curves on their finite set of points—for example, the piecewise constant nonparametric density estimation by Boullé (2012); 2) filtering methods that require smoothing curves into a basis of functions and clustering the resulting expansion coefficients in turn—for example, the principal points of curves (Tarpey and Kinateder, 2003), B-spline fitting (Abraham et al., 2003), and functional principal component analysis (Peng and Müller, 2008); 3) adaptive methods that perform the simultaneous clustering and expression of the curves into a finite dimensional

---

space—for example, basis expansion coefficients modelling (James and Sugar, 2003; Samé et al., 2011; Giacofci et al., 2013), and functional principal component analysis (FPCA, Chiou and Li, 2007; Bouveyron and Jacques, 2011; Jacques and Preda, 2013; Jacques and Preda, 2014b; Centofanti et al., 2021); and 4) distance-based methods where usual clustering algorithms are applied with specific distances for functional data. For example, the K-means algorithm (Hartigan and Wong, 1979) was recommended by Cuesta-Albertos and Fraiman (2007), Antoniadis et al. (2013), García et al. (2015) and Albert-Smet et al. (2022); an agglomerative hierarchical clustering algorithm (Day and Edelsbrunner, 1984) was considered by Liu et al. (2012); and the K-medoids method (Kaufman and Rousseeuw, 1990), which is a modification of the K-means method, was developed by Chen et al. (2017); in addition, Floriello and Vitelli (2017) extended the K-means, K-medoids, and hierarchical clustering algorithms to functional frameworks. Although the FPCA-based clustering method in Centofanti et al. (2021) can be applied to an irregular time grid, there may be computation issues due to the estimation of parameter sets in big functional data clustering.

For multivariate functional data, however, only few innovations exist for data clustering. From the viewpoint of generalization of adaptive methods, Jacques and Preda (2014b) generalized functional clustering from univariate to multivariate functional data via multivariate functional principal component analysis (MFPCA, Ramsay and Silverman, 2005). Furthermore, Schmutz et al. (2020) extended the work of Jacques and Preda (2014b), and their new method (funHDDC) is advantageous from two perspectives: for modelling all principal component scores whose estimated variances are non-null and for using the expectation-maximization (EM) algorithm to propose a criterion for selecting the number of clusters. Moreover, Misumi et al. (2019) proposed a multivariate nonlinear mixed-effects model to express multivariate functional data and applied a nonhierarchical clustering algorithm based on self-organizing maps to the predicted coefficient vectors of individual-specific random effect functions. With regard to distance methods, Ieva et al. (2011), Ieva et al. (2013), and Meng et al. (2018) proposed new distance measures and applied the K-means clustering algorithm to multivariate functional data. Other methods considered clustering curves while capturing the features of amplitude or phase variations (Marron et al., 2014). For instance, Sangalli et al. (2010) proposed a K-means clustering method to decompose amplitude and phase variations while assuming a linear warping time function; Park and Ahn (2017) proposed a conditional subject-specific warping framework and considered multivariate functional clustering with phase variation.

There are criteria to find the optimal number of clusters, i.e., K, for distance-based clustering methods, such as K-means, K-medoids, agglomerative hierarchical clustering, and density-based spatial clustering of applications with noise (DBSCAN, Ester et al., 1996). Generally, the optimal K can be determined from the elbow value (Shi et al., 2021), average silhouette (Struyf et al., 1997; Batool and Hennig, 2021), gap statistics (Tibshirani et al., 2001), model-based Akaike Information Criterion (AIC, Akaike, 1974), Bayesian Information Criterion (BIC, Schwarz, 1978), and integrated classification likelihood (Biernacki et al., 2000).

Two main characteristics are observed for the existing clustering methods: first, for sparse multivariate functional data, no specific distance measurements are defined and the model-based clustering methods, except for that of Misumi et al. (2019), cannot be applied directly; second, the above methods rarely consider noises or outliers, though real data are often corrupted by them, see Ronchetti (2021) for a recent review. Robust clustering methods for multivariate data to reduce the influence of outliers are acknowledged, see density-based spatial clustering of applications with noise (DBSCAN, Ester et al., 1996), robust spectral clustering (Li et al., 2007), improvements in K-means clustering (Wang and Su, 2011) and K-medoids clustering (Al Abid, 2014), and hierarchical clustering (Balcan et al., 2014; Gagolewski et al., 2016)). In addition, developments in outlier detection methods in multivariate functional data are, for instance, functional adjusted outlyingness and central-stability plots (Hubert et al., 2015), magnitude-shape plot (Dai and Genton, 2018), and directional outlyingness (Dai and Genton, 2019). However, robust clustering algorithms that correctly obtain clusters and detect outliers for sparse multivariate functional data are lacking.

Sparse (multivariate) functional data are defined as data objects with various time grids per subject. One common example of sparse data in practice is imbalanced data, where some objects may have a large number of measurements while others may have few measurements. Here, the measurement is assumed to be either observed or missing for all variables. The case where the measurements for one index are observed for some variables but missing for others is not considered. In addition, the location of missing observations is assumed to be random, and independent of the observed and missing values.

The objective of this work is to develop a method that can correctly detect clusters and flag outliers, if any, for sparse multivariate functional data. First, the concept of elastic time distance (ETD) is proposed, which is applicable to (multivariate) functional data with either identical or different time measurements per subject. Second, an outlier-resistant method to obtain clusters and remove outliers based on the ETD measures is presented. The rest of the paper is organized as follows. Section 2 proposes a novel robust clustering method based on a new distance measure, ETD. The proposed method is referred to as robust two-layer partition (RTLP) clustering. Section 3 describes numerical experiments performed to evaluate the clustering precision and outlier detection performance among RTLP clustering, one existing model-based method, and three distanced-based methods extended with ETD. Section 4 presents the results of application of RTLP clustering to Northwest Pacific cyclone tracks data to establish its efficiency. Section 5 concludes the paper with a summary and discussion. Proofs are collected in the Appendix.

2. Multivariate functional data clustering

A multivariate functional random variable of p dimensions is a p-variate random vector with values in an infinite-dimensional space. In a well-known model of multivariate functional data (Hsing and Eubank, 2015), the data are considered as sample paths of a stochastic process $\mathbf{Y} = \{\mathbf{Y}(t) := (Y^{(1)}(t), \ldots, Y^{(p)}(t))^\top\}_{t \in \mathcal{T}}$ taking real values in some Hilbert space, $\mathcal{H}$, of functions defined on some compact set $\mathcal{T} \subseteq \mathbb{R}$. Let $\mathcal{H}$ be square integrable functions of $\mathcal{T}$, written as $\mathcal{H} := L^2(\mathcal{T}) \times \cdots \times L^2(\mathcal{T})$.

2.1. Elastic time distance

The distance measure between $\mathbf{Y}_m$ and $\mathbf{Y}_n$ for $\mathbf{Y}_m, \mathbf{Y}_n \in \mathcal{C}^p(\mathcal{T})$ is introduced, where $\mathcal{C}^p(\mathcal{T})$ represents the p-vectors of continuous functions on $\mathcal{T}$. Suppose Assumption 1 below holds.

Assumption 1. Let the time/wavelength points come from a design density $g(u)$ such that $G(t) = \int_{-\infty}^t g(u)\mathrm{d}u$. Let the compact set $\mathcal{T} = [G^{-1}(0), G^{-1}(1)]$. It is assumed that g is differentiable and $\inf_{t \in \mathcal{T}} g(t) > 0$.

Instead of the existing L^p metric in the Hilbert space, $\mathcal{H}$, the supremum norm of point-wise L^2 norm between $\mathbf{Y}_m$ and $\mathbf{Y}_n$ is adopted to measure the similarity between $\mathbf{Y}_m$ and $\mathbf{Y}_n$:

$$d(\mathbf{Y}_m, \mathbf{Y}_n) = \sup_{t \in \mathcal{T}} \left[\sqrt{\sum_{l=1}^p \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right]. \tag{1}$$

Overall, $d(\mathbf{Y}_m, \mathbf{Y}_n)$ is an extended Chebyshev distance in the Hilbert space, with the point-wise distance being the L^2 norm for p-dimensional vectors. Other norms, such as L^1 norm and L^∞ norm, can be adopted as a point-wise distance measure in $d(\mathbf{Y}_m, \mathbf{Y}_n)$ in principle. With the L^2 norm as an example, $d(\mathbf{Y}_m, \mathbf{Y}_n)$ is proved to be a metric (Kelley, 1955, p. 119) satisfying the properties described in Theorem 1.

Theorem 1. For $\mathbf{Y}_m, \mathbf{Y}_n, \mathbf{Y}_o \in \mathcal{C}^p(\mathcal{T})$, d, as a metric, is a distance function:
   1. Non-negativity: $d(\mathbf{Y}_m, \mathbf{Y}_n) \geq 0$.
   2. Non-degeneracy: $d(\mathbf{Y}_m, \mathbf{Y}_n) = 0 \iff \mathbf{Y}_m = \mathbf{Y}_n$.
   3. Symmetry: $d(\mathbf{Y}_m, \mathbf{Y}_n) = d(\mathbf{Y}_n, \mathbf{Y}_m)$.
   4. Triangular inequality: $d(\mathbf{Y}_m, \mathbf{Y}_n) \leq d(\mathbf{Y}_m, \mathbf{Y}_o) + d(\mathbf{Y}_o, \mathbf{Y}_n)$.

Although the observations are supposed to be infinite-dimensional, in practice, only a finite number of discrete observations $\widehat{Y}_{n,k}$ ($n = 1, \ldots, N$) of each sample path $\widehat{Y}_n$ are evaluated at the set $\mathbf{t}_n = \{t_{n,k} := G^{-1}(k/T_n), k = 1, \ldots, T_n\}$. When the set of finite observations $S := \{\widehat{\mathbf{Y}}_1(\mathbf{t}_1), \ldots, \widehat{\mathbf{Y}}_N(\mathbf{t}_N)\}$ has the cardinality N, d usually cannot be applied to calculate the distance between different $\widehat{\mathbf{Y}}_m, \widehat{\mathbf{Y}}_n \in S$ due to various grid settings. To make d applicable to objects with different counts and locations of time points, a criterion to build a standard grid and interpolate observations at the standard grid for all curves is proposed. Then, the distance between $\widehat{\mathbf{Y}}_m$ and $\widehat{\mathbf{Y}}_n$ is approximated by the distance between the interpolated curves $\widetilde{\mathbf{Y}}_m$ and $\widetilde{\mathbf{Y}}_n$. The procedures are explained as follows and the derived distance is coined the elastic time distance (ETD) for sparse multivariate functional data.

Definition 1. (Standard grid $\mathbf{st}$). The standard grid $\mathbf{st}$ is defined as a set $\{st_k := \frac{k-1}{T-1}, k = 1, \ldots, T\}$ with $T = \max_{n=1}^N T_n$.

Definition 2. (Interpolated observation at $\mathbf{st}$). The interpolation $\widetilde{\mathbf{Y}}_n$ is defined as a set $\{\widehat{\mathbf{Y}}_n(\widetilde{t}_{n,k}), \widetilde{t}_{n,k} := \arg\min_{t \in \mathbf{t}_n} |t - st_k|, k = 1, \ldots, T\}$.

As shown in Fig. 1, because the maximum number of measurements T from the five samples $\{\widehat{\mathbf{Y}}_1, \ldots, \widehat{\mathbf{Y}}_5\}$ is 11, $\mathbf{st} = \{(k-1)/10, k = 1, \ldots, 11\}$. Then, the observation at any standard time st_k ($k = 1, \ldots, 11$) is the one with the closest time difference from $\mathbf{t}_n$ and st_k. Although the trend of the interpolated curve is slightly different from the original one (see the different measurements in [0.4, 0.6] before and after the interpolation), Definitions 1-2 provide a simple and feasible criterion in interpolations at the standard time grid. The interpolated curve is used to approximate the original curve and capture the similarity of the two curves with the farthest point-wise distance over the standard grid. Hence this simple interpolation criterion makes it feasible and effective to measure the similarity between curves.

For any $\widehat{\mathbf{Y}}_m, \widehat{\mathbf{Y}}_n \in S$ ($1 \leq n \leq m \leq N$), the finite-grid ETD between $\widehat{\mathbf{Y}}_m$ and $\widehat{\mathbf{Y}}_n$ is calculated as follows:

$$\begin{aligned}
ETD(\widehat{\mathbf{Y}}_m, \widehat{\mathbf{Y}}_n) := d(\widetilde{\mathbf{Y}}_m, \widetilde{\mathbf{Y}}_n) &= \max_{k=1,\ldots,T-1} \left(\max_{t \in [st_k, st_{k+1}]} \left[\sqrt{\sum_{l=1}^p \left\{ \widetilde{Y}_m^{(l)}(t) - \widetilde{Y}_n^{(l)}(t) \right\}^2} \right] \right) \\
&= \max_{k=1,\ldots,T} \left[\sqrt{\sum_{l=1}^p \left\{ \widetilde{Y}_m^{(l)}(st_k) - \widetilde{Y}_n^{(l)}(st_k) \right\}^2} \right].
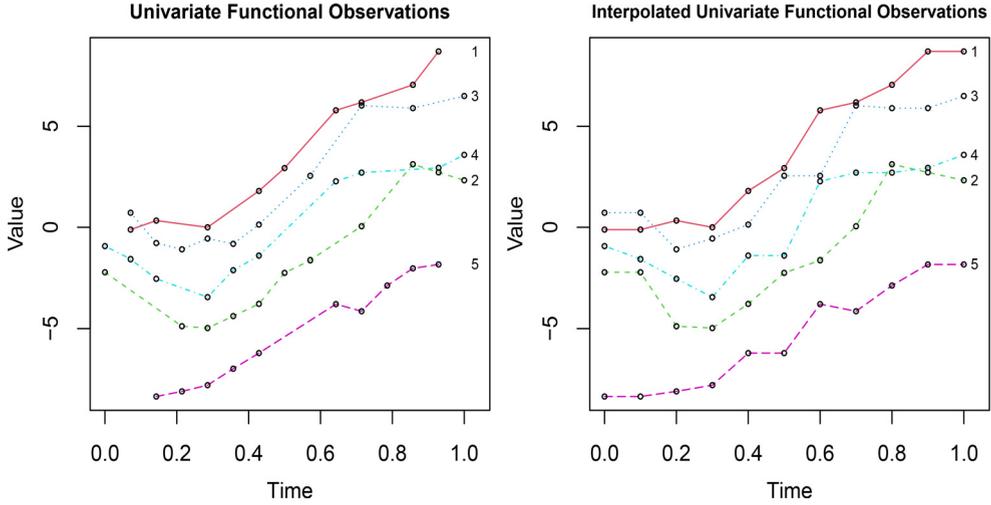\end{aligned} \tag{2}$$

Fig. 1. Example of five univariate functional samples evaluated on various time grids (left), and the corresponding interpolated univariate functional samples evaluated on the standard grid (right).

When all the curves are from common and equidistant time points, Definitions 1-2 make no difference between $\widehat{Y}_m$ and $\widetilde{Y}_m$ $(m = 1, \ldots, N)$. Then, $d(\widetilde{Y}_m, \widetilde{Y}_n) = d(\widehat{Y}_m, \widehat{Y}_n)$. When various time grids are allowed, $ETD(\widehat{Y}_m, \widehat{Y}_n)$ is a pseudo-metric (Kelley, 1955, p. 119). That is, Theorem 1 is satisfied except that $ETD(\widehat{Y}_m, \widehat{Y}_n) = 0$ does not lead to $\widehat{Y}_m = \widehat{Y}_n$. If $ETD(\widehat{Y}_m, \widehat{Y}_n) = 0$, then $\widetilde{Y}_m = \widetilde{Y}_n$ is obtained, but the original functions $\widehat{Y}_m$ and $\widehat{Y}_n$ may vary in terms of the location and number of measurement grids.

The consistency from the finite grid elastic time distance to the continuous version is satisfied given Assumption 1; see Theorem 2 below.

Theorem 2. *(Consistency). Given $\widehat{Y}_m, \widehat{Y}_n \in S$, $\widetilde{Y}_m, \widetilde{Y}_n \in C^p(\mathcal{T})$ from Definitions 1-2. Suppose Assumption 1 holds. When $T_n \to \infty$ $\forall n = 1, \ldots, N$, $|d(\widetilde{Y}_m, \widetilde{Y}_n) - d(Y_m, Y_n)| \xrightarrow{a.s.} 0$.*

The ETD provides a simple transformation that measures the similarity between available curves quickly. The procedure of calculating ETD in Eq. (2) is summarized in Algorithm 1. The existing distance-based methods, such as the DBSCAN,

---

Algorithm 1 Elastic Time Distance Algorithm

1: **Input:** $S = \{\widehat{Y}_1, \ldots, \widehat{Y}_N\}$, $t_1, \ldots, t_N$

2: **Initialize:** Zero matrix $D \in \mathbb{R}^{N \times N}$, $T = \max\limits_{n=1}^{N} T_n$, $\widetilde{Y}_n \leftarrow$ Definition 2 $\forall n = 1, \ldots, N$

3: **function** $ETD(\widehat{Y}_i, \widehat{Y}_j)$

4:        **return** $\max\limits_{k=1}^{T} \left\{ \sqrt{\sum_{l=1}^{p} (\widetilde{Y}_{i,k}^{(l)} - \widetilde{Y}_{j,k}^{(l)})^2} \right\}$

5: **end function**

6: **for** $i = 1$ to $N - 1$ **do**

7:      **for** $j = i + 1$ to N **do**

8:          $D_{i,j} \leftarrow ETD(\widehat{Y}_i, \widehat{Y}_j)$

9:          $D_{j,i} \leftarrow D_{i,j}$

10:      **end for**

11: **end for**

12: **Output:** D

---

agglomerative hierarchical clustering, and classical K-medoids, can be naturally combined with ETD. Then a distance-based clustering method is proposed with $ETD(\widehat{Y}_m, \widehat{Y}_n)$ as the measure in the numerical simulations. The point-wise L^2 norm, L^1 norm, and L^∞ norm all achieve identically excellent performances in the simulation trials yet to be reported, and D from Algorithm 1 is used to represent the result of the distance matrix.

2.2. Two-layer partition clustering algorithm

Based on the matrix of ETD, the neighbours of each curve and the core of a set are defined, which follow the style of DBSCAN (Ester et al., 1996) and depth-based clustering (Jeong et al., 2016). First, the first-layer partition is introduced.

However, the output of the first-layer partition may have disjoint sets that are from homogeneous groups. Then, another layer is implemented to make it possible to merge nearby groups into a cluster. In practice, the third-layer partition leads to one cluster that contains all curves. Hence, a two-layer partition clustering algorithm is proposed. Here, necessary concepts are introduced before describing the proposed clustering algorithm. Let $|A|$ be the cardinality of the set A, and $q_\theta(A)$ be the θ-quantile $(0 < \theta < 1)$ of the set A, and $D(S)$ be a set containing the upper triangular elements of $\boldsymbol{D}$ from Algorithm 1.

Definition 3 (Neighbours of a curve)**.** The set $nbr(\boldsymbol{Y}, S, \theta) := \{\boldsymbol{Z} \in S : ETD(\boldsymbol{Y}, \boldsymbol{Z}) < q_\theta(D(S))\}$ consists of neighbours of $\boldsymbol{Y}$ in the set S.

Here, θ is used to determine the number of neighbours per curve. Generally, the number of neighbours per curve is positively correlated with θ. The criterion for selecting the optimal θ is introduced in Section 2.4.

Definition 4 (Core and center of the set)**.** For any set A, its core is the element with the most neighbours, i.e., $core(A, \theta) = \arg\max_{\boldsymbol{Y} \in A} |nbr(\boldsymbol{Y}, A, \theta)|$. If A is a cluster, $core(A, \theta)$ is named as the center of A.

Definition 5 (Disjoint partitions)**.** The disjoint partition $\boldsymbol{P} = (P^1, \ldots, P^L)$ of S is defined as a separation of S that satisfies: $P^i \cap P^j = \emptyset$ $(1 \le i < j \le L)$, $P^1 \cup \cdots \cup P^L = S$, and $|P^1| \ge |P^2| \ge \cdots \ge |P^L|$. Here, L is the number of disjoint sets (also the cardinality of $\boldsymbol{P}$), and $L \le N$. For a disjoint partition $\boldsymbol{P}$ of S, let P^i be the i-th set in the partition $\boldsymbol{P}$.

The idea of the two-layer partition clustering algorithm is similar to the bottom-up approach in agglomerative hierarchical clustering. However, only two layers of hierarchy are implemented and not all curves are merged to one cluster in the two-layer partition method. In the first layer, each curve, as an isolated cluster, is merged into disjoint groups based on the concepts of neighbours and the core. Then, in the second layer, the first-layer disjoint groups are merged into a set of clusters. The criteria for determining the first and second layers of the partitions are proposed as follows.

Criterion 1. *(The first-layer partition $\boldsymbol{G}$). The first-layer partition $\boldsymbol{G}$ from the core and its neighbours is obtained. First, S is copied to the set $\widetilde{S}$, and $\boldsymbol{G}$ is initialized as the empty set, and $M = 1$. While $\widetilde{S}$ is not empty, implement the following three steps: define the curve with the most neighbours in $\widetilde{S}$ as a core ($\boldsymbol{c} = core(\widetilde{S}, \theta)$); define the set consisting of all its neighbours in $\widetilde{S}$ as G^M ($G^M = nbr(\boldsymbol{c}, \widetilde{S}, \theta)$); update $\widetilde{S}$ with $\widetilde{S} = \widetilde{S} \setminus G^M$ and increase M by 1. Let M represent the cardinality of $\boldsymbol{G}$.*

The intuition of Criterion 1 is as follows. Assuming the current set is a copy of S, the idea of disjoint groups $\boldsymbol{G}$ is: determining the core of the current set, finding the neighbours of the core in the current set as the disjoint group in $\boldsymbol{G}$, and removing such neighbours from the current set. For $G^i, G^j \in \boldsymbol{G}$ $(1 \le i < j \le M)$ in Criterion 1, $|G^i| \ge |G^j|$ is satisfied, where G^i is the prioritized group, and G^j is the subordinate group. The pairwise merging principle $merge(G^i, G^j)$ is as follows: a subordinate group G^j may be merged to a prioritized group G^i (i.e., $G^i := G^i \cup G^j$, $G^j := \emptyset$) if the core of G^j is a neighbour of a curve from G^i in S (i.e., $core(G^j, \theta) \in \cup_{\boldsymbol{Y} \in G^i} nbr(\boldsymbol{Y}, S, \theta)$), and if G^j was not merged to any previous prioritized group G^k $(1 \le k < i)$ (i.e., $G^j \ne \emptyset$) yet.

The pairwise merging principle is the core of the second-layer partition $\boldsymbol{C}$. Assuming the disjoint groups $\widetilde{\boldsymbol{G}}$ is a copy of the first-layer partition $\boldsymbol{G}$, the idea of merging disjoint groups is implemented in a decreasing order. While $\widetilde{\boldsymbol{G}}$ is not empty, implement the following three steps: find the most prioritized group $\widetilde{G}^1$ in $\widetilde{\boldsymbol{G}}$; for all its subordinate groups in $\widetilde{\boldsymbol{G}}$, check whether the subordinate group is a neighbour of any curve belonging to $\widetilde{G}^1$, if yes, merge $\widetilde{G}^1$ with the subordinate group, and remove such subordinate group in $\widetilde{\boldsymbol{G}}$; include $\widetilde{G}^1$ in $\boldsymbol{C}$, and remove G^1 from $\widetilde{\boldsymbol{G}}$. The first-layer (group) partition is named as $\boldsymbol{G}$, and the second-layer (cluster) partition is named as $\boldsymbol{C}$, and both satisfy Definition 5.

Criterion 2. *(The second-layer partition $\boldsymbol{C}$). To obtain the second-layer partition $\boldsymbol{C}$, first, the partition $\boldsymbol{G}$ is copied to $\widetilde{\boldsymbol{G}}$, and $\boldsymbol{C}$ is initialized as the empty set. While $\widetilde{\boldsymbol{G}}$ is not empty, implement the following three steps: separate partition $\widetilde{\boldsymbol{G}}$ into its most prioritized group $\widetilde{G}^1$ and its subordinate groups $\boldsymbol{R} := \widetilde{\boldsymbol{G}} \setminus \widetilde{G}^1$; for $A \in \boldsymbol{R}$, implement $merge(\widetilde{G}^1, A)$; include $\widetilde{G}^1$ in $\boldsymbol{C}$, and update $\widetilde{\boldsymbol{G}}$ as $\widetilde{\boldsymbol{G}} = \widetilde{\boldsymbol{G}} \setminus \widetilde{G}^1$. Let I represent the cardinality of the partition $\boldsymbol{C}$.*

Algorithm 2 is used to obtain clusters from two layers of hierarchy according to Definitions 1-5 and Criteria 1-2.

2.3. Cluster and outlier recognition algorithm

The output of Algorithm 2 may include clusters with a large number of curves and some with few curves because of the noise and outliers in the observations. To solve this issue, Algorithm 3 is proposed for recognizing the primary clusters and outliers using p_m and α. Here, p_m denotes the minimal required number of observations in a cluster divided by the sample size, and α $(0 < \alpha < 1)$ is a threshold that defines the α-quantile of the distances between curves inside a cluster and the cluster center. Let $\Psi_x(A)$ be the empirical cumulative distribution of x in the set A. The criterion for selecting the primary clusters and outliers is proposed as follows.

Criterion 3. *(The set of primary clusters $\boldsymbol{C}_p$ and the outlier set O). The set of primary clusters is initialized as $\boldsymbol{C}_p = \{C \in \boldsymbol{C}: |C| \ge Np_m\}$. Naturally, the potential outlier set is $O_p = \{\boldsymbol{X} \in C: |C| < Np_m$ and $C \in \boldsymbol{C}\}$. Let the outlier set O be empty. The potential outlier $\boldsymbol{X} \in O_p$ is labeled as an outlier ($\boldsymbol{X} \in O$), if for any $C \in \boldsymbol{C}_p$, the ETD between $\boldsymbol{X}$ and $core(C, \theta)$ is larger than the α-quantile of the set $D(C, core(C, \theta))$, which is the set of the ETD between each element in C and $core(C, \theta)$. Otherwise, $\boldsymbol{X}$ is classified into*

the set with minimum distance to the center distribution, $C_p^l = \arg\min_{C \in \boldsymbol{C}_p} \Psi_{ETD(\boldsymbol{X}, core(C, \theta))}(D(C, core(C, \theta)))$. In a mathematical notation, $\forall \boldsymbol{X} \in O_p$,

$$\boldsymbol{X} \in \begin{cases} O, & \text{if } ETD(\boldsymbol{X}, c) > q_\alpha(D(C, c)) \ \forall \ C \in \boldsymbol{C}_p, c := core(C, \theta), \\ C_p^l, & \text{if } C_p^l := \arg\min_{C \in \boldsymbol{C}_p} \Psi_{ETD(\boldsymbol{X}, c)}(D(C, c)), c := core(C, \theta). \end{cases}$$

Here, each element in $\boldsymbol{C}_p$ is a set containing at least Np_m number of curves, while each element in O is a detected outlier.

Algorithm 3 applies Criterion 3 and obtains a set of primary clusters $\boldsymbol{C}_p$ and outliers O. Let $p_m = 0.1$. It is recommended to use α between 0.85 and 0.9 based on the empirical simulations yet to be reported. Adoption of Algorithm 3 makes the two-layer partition clustering (Algorithm 2) robust to outliers and removes outliers. This method is referred to as RTLP clustering.

Algorithm 2 Two-Layer Partition Clustering Algorithm

1: **Input**: $S = \{\widehat{\boldsymbol{Y}}_1, \ldots, \widehat{\boldsymbol{Y}}_N\}$, $D(S)$ from Algorithm 1, θ
2: **function** $nbr(\boldsymbol{Y}, A, \theta)$
3:     **return** $\{\boldsymbol{Z} \in A \text{ if } ETD(\boldsymbol{Y}, \boldsymbol{Z}) < q_\theta(D(S))\}$
4: **end function**
5: **function** $core(A, \theta)$
6:     **return** $arg\,max_{\boldsymbol{Y} \in A}|nbr(\boldsymbol{Y}, A, \theta)|$
7: **end function**
8: **function** $merge(G^i, G^j)$
9:     **if** $|G^i| < |G^j|$ **then**
10:         **error**   The first parameter should be the prioritized group
11:     **end if**
12:     **if** $core(G^j, \theta) \in \cup_{\boldsymbol{Y} \in G^i} nbr(\boldsymbol{Y}, S, \theta)$ **then**
13:         $G^i \leftarrow G^i \cup G^j$
14:         $G^j \leftarrow \emptyset$
15:         **return** G^i, G^j
16:     **end if**
17: **end function**
18: **Initialize:** $\boldsymbol{G} \leftarrow \emptyset$, $M \leftarrow 1$, $\widetilde{S} \leftarrow S$
19: **while** $\widetilde{S} \neq \emptyset$ **do**
20:     $G^M \leftarrow nbr(core(\widetilde{S}, \theta), \widetilde{S}, \theta)$
21:     $\widetilde{S} \leftarrow \widetilde{S} \setminus G^M$
22:     $M \leftarrow M + 1$
23: **end while**
24: **Initialize:** $\boldsymbol{C} \leftarrow \emptyset$, $I \leftarrow 1$, $\widetilde{\boldsymbol{G}} \leftarrow \boldsymbol{G}$
25: **while** $\widetilde{\boldsymbol{G}} \neq \emptyset$ **do**
26:     $\boldsymbol{R} \leftarrow \widetilde{\boldsymbol{G}} \setminus \widetilde{G}^1$
27:     **for** A in $\boldsymbol{R}$ **do**
28:         $merge(\widetilde{G}^1, A)$
29:     **end for**
30:     $C^I \leftarrow \widetilde{G}^1$
31:     $\widetilde{\boldsymbol{G}} \leftarrow \widetilde{\boldsymbol{G}} \setminus \widetilde{G}^1$
32:     $I \leftarrow I + 1$
33: **end while**
34: **Output**: $\boldsymbol{C}$

2.4. Clustering execution

The clustering result may differ if different θ is applied. Thus, the optimal θ needs to be selected. Under $\theta \in [0, 1]$, implementing Algorithms 2–3 obtains a set of primary clusters $\boldsymbol{C}_{p,\theta} = (C_{p,\theta}^1, \ldots, C_{p,\theta}^{B_\theta})$ and the outlier set O_θ. Then, the average silhouette value to assess the goodness of clustering is implemented, which is the mean of the silhouette values (Rousseeuw, 1987) of all data objects. The silhouette value measures the resemblance of an object to its cluster compared to other clusters and ranges from -1 to 1. Here, a high value means that the object is well matched to its own cluster and poorly matched to neighbouring clusters, and a low value indicates that the clustering configuration may have an inappropriate number of clusters. The optimal θ is the one with the largest average silhouette value (Batool and Hennig, 2021). Thereafter, $\boldsymbol{C}_p$ and O under the optimal θ are adopted.

Z. Qu, W. Dai and M.G. Genton

Algorithm 3 Cluster and Outlier Recognition Algorithm

1: **Input:** $S = \{\widehat{Y}_1, \ldots, \widehat{Y}_N\}$, D from Algorithm 1, C from Algorithm 2, p_m, α
2: **Initialize:** $C_p \leftarrow \emptyset$, $O_p \leftarrow \emptyset$, $O \leftarrow \emptyset$, $B \leftarrow 0$
3: **for** i in $1, \ldots, |C|$ **do**
4:     **if** $|C^i| \geq N p_m$ **then**
5:         $B \leftarrow B + 1$
6:         $C_p^B \leftarrow C^i$
7:     **else**
8:         $O_p \leftarrow O_p \cup C^i$
9:     **end if**
10: **end for**
11: **if** $|O_p| > 0$ and $|C_p| > 0$ **then**
12:     **for** X in O_p **do**
13:         **if** $ETD(X, core(C, \theta)) > q_\alpha(D(C, core(C, \theta))) \; \forall C \in C_p$ **then**
14:             $O \leftarrow O \cup X$
15:         **else**
16:             $C_p^l \leftarrow arg\,min_{C \in C_p} \Psi_{ETD(X, core(C, \theta))}(D(C, core(C, \theta)))$
17:             $C_p^l \leftarrow C_p^l \cup X$
18:         **end if**
19:     **end for**
20: **end if**
21: **Output:** C_p and O

The original silhouette value is only defined when the object Y belongs to a cluster $C_{p,\theta}^b$ ($1 \leq b \leq B_\theta$). That is, if $Y \in C_{p,\theta}^b$, then the average distance within the cluster $a(Y, \theta) = \frac{1}{|C_{p,\theta}^b|-1} \sum_{Z \in C_{p,\theta}^b} ETD(Y, Z)$, the minimum distance across the clusters $b(Y, \theta) = \min_{\substack{j \neq b \\ 1 \leq j \leq B_\theta}} \frac{1}{|C_{p,\theta}^j|} \sum_{Z \in C_{p,\theta}^j} ETD(Y, Z)$, and the silhouette value of Y is $s(Y, \theta) = \frac{b(Y,\theta)-a(Y,\theta)}{\max\{a(Y,\theta),\, b(Y,\theta)\}}$. The silhouette value is also defined when the object is in the outlier set O_θ. That is, if $Y \in O_\theta$, then $s(Y, \theta) = 0$. It helps correctly recognize clusters and detect outliers if any.

Consider the silhouette value of a curve Y when it is correctly or wrongly detected, respectively. The first case is that the curve is an outlier ($Y \in O$). If an outlier $Y \in O$ is correctly detected ($Y \in O_\theta$), then $s(Y, \theta) = 0$. Otherwise, if an outlier $Y \in O$ is wrongly put into a cluster ($Y \in C_{p,\theta}^l$ and $C_{p,\theta}^l \in C_{p,\theta}$), usually the average distance within the cluster, $a(Y, \theta)$, is larger than the minimum distance across the clusters, $b(Y, \theta)$, hence $s(Y, \theta)$ is negative. The second case is that the curve belongs to a cluster, i.e., $Y \in C_p^b$ ($1 \leq b \leq B$). When Y is correctly labeled in its cluster $C_{p,\theta}^b$, its silhouette value, $s(Y, \theta)$, is positive because the minimum distance across the clusters, $b(Y, \theta)$, is larger than its average distance within the cluster, $a(Y, \theta)$. When Y is wrongly labeled in the outlier set O_θ, its silhouette value, $s(Y, \theta)$, turns to zero. When Y is wrongly labeled in any other clusters $C_{p,\theta}^w$ ($w \neq b$, $1 \leq w \leq B_\theta$), its silhouette value, $s(Y, \theta)$, becomes negative because the minimum distance across the clusters, $b(Y, \theta)$, is less than the average distance within the cluster, $a(Y, \theta)$. The optimal θ is defined as $\theta = \arg\max_{\theta \in [0,1]} \bar{s}(\theta)$, with the average silhouette value $\bar{s}(\theta) = \frac{1}{N} \sum_{k=1}^{N} s(Y, \theta)$. Hence, the idea of setting zero values for the set of outliers promotes the separation between clusters and outliers.

In practice, $\theta \in [0, 0.25]$, because all curves are merged into one group when $\theta \geq 0.25$ as the case in Fig. 2(b). The change of average silhouette value versus θ was explored when the number of clusters changed from two to five in the simulations yet to be reported, and the average silhouette value all dropped significantly to zero when $0.2 \leq \theta \leq 0.25$.

In principle, any distance measures suitable for multivariate functional data can be used in the RTLP clustering. The ETD is used as a building block as it is applicable for both complete and sparse multivariate functional data. Overall, the RTLP clustering is executed in several steps:

1) Compute ETD according to Algorithm 1.
2) Execute Algorithms 2–3 for $\theta \in [0.01, 0.25]$ and compute the average silhouette value $\bar{s}(\theta)$. The optimal θ is the one with the highest $\bar{s}(\theta)$.
3) Implement Algorithms 2–3 under the optimal θ from 2), and obtain the set of primary clusters C_p and the outlier set O.

To illustrate the application of RTLP, the stepwise results for a clustering scenario are presented. As shown in Fig. 2(a), 120 curves and four clusters with a cardinality of 30 are generated. Subsequently, random 12 out of 120 curves are replaced with outliers marked in dashed orange. Then, missing values are generated in random 96 out of 120 curves. Here, the standard time grid has 20 equidistant points from 0 to 1. For the curves with missing values, six values are missing randomly in the standard time grid on average, and the locations of the missing measurements are independent and random per subject. The transparency of the color increases with the time elapsed.
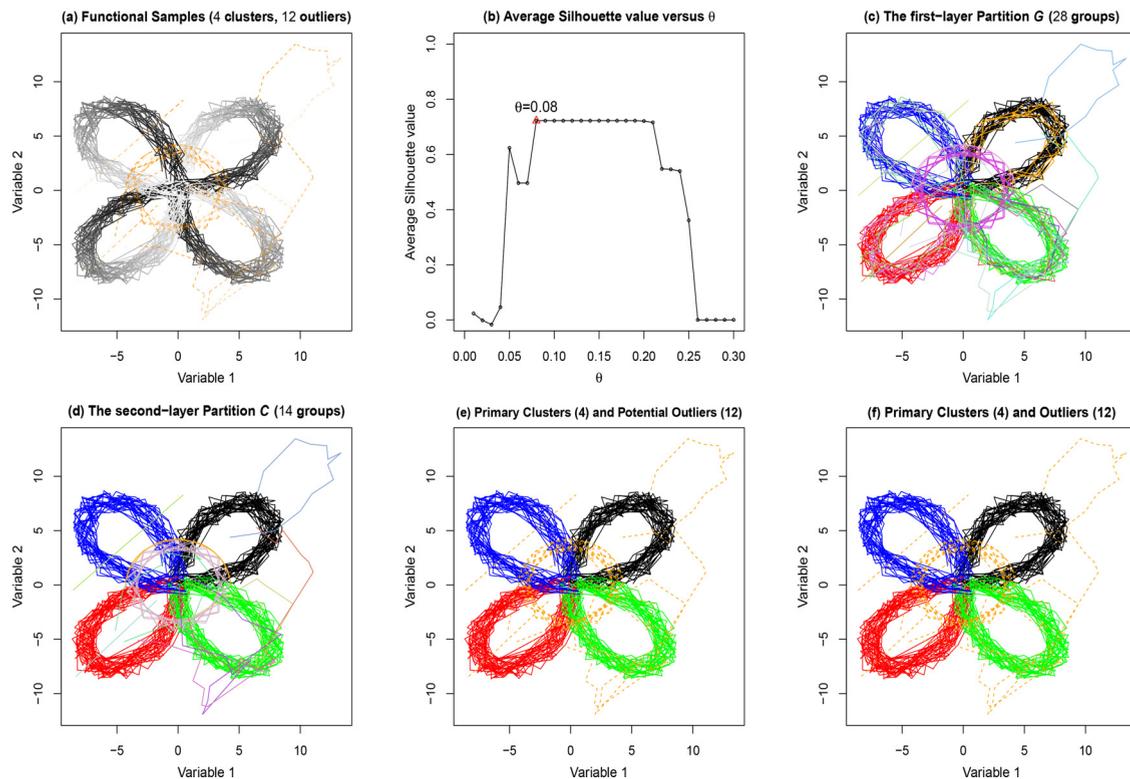
Fig. 2. Example of stepwise clustering results. (a) Original clusters in black (transparency increases with time elapsed) mixed with outliers in orange, and (b) changes in the average silhouette value vs. θ, (c) and (d) First- and second-layer partitions G and C from Algorithm 2, respectively, and (e) and (f) primary clusters C_p and outliers O before and after applying the detection criterion, respectively, in Algorithm 3.

Algorithms 2–3 are executed under $p_m = 0.1$, $\alpha = 0.85$ and $\theta \in [0, 1]$ after the calculation of the ETD. Fig. 2(b) only visualizes the change of average silhouette value versus θ when $\theta \leq 0.3$ since it already reaches zero when $\theta = 0.26$. In addition, Fig. 2(b) indicates an increase in $\bar{s}(\theta)$ for $\theta < 0.08$ when the number of clusters decreases significantly to four; a stable $\bar{s}(\theta)$ for $0.08 \leq \theta \leq 0.21$, corresponding to the correct number of clusters with a slight movement of the outlier set; and a decrease in $\bar{s}(\theta)$, when the number of clusters decreases from four to one. The optimal θ is 0.08, corresponding to the highest average silhouette value. Subsequently, the clustering result is that of Algorithms 2–3 under $\theta = 0.08$. Figs. 2(c) and (d) show the results of the first- and second-layer partitions, i.e., G and C, respectively, from Algorithm 2. The first four sets in G and C merge the majority of the curves and are successively marked in black, red, green, and blue. The remaining sets in G and C have a cardinality of either one or two with other distinct colors per set. Figs. 2(e) and (f) show identical results obtained before and after updating the primary clusters C_p via Algorithm 3. On applying Criterion 3, all potential outliers remain as outliers with all true outliers correctly detected, as seen in Fig. 2(f).

Algorithm 1 takes $ETD(\widehat{Y}_i, \widehat{Y}_j)$ as the basic operation. Algorithm 2 takes $G^M \leftarrow nbr(core(\widetilde{S}, \theta), \widetilde{S}, \theta)$ as the basic operation. Algorithm takes $ETD(X, core(C, \theta)) > q_\alpha(D(C, core(C, \theta)))$ as the basic operation. The computational complexity of the worst cases in the above algorithms take $\mathcal{O}(N^2)$, $\mathcal{O}(N)$, and $\mathcal{O}(|C_p| \times |O|)$, respectively. Here, $|C_p| \leq I$, $|O| \leq Np_m I$ and empirically, $I \leq c\sqrt{N}$ for some $c > 0$. Therefore, most of the running time is spent on Algorithm 1, and this situation can be improved by parallel computation (Rossini et al., 2007).

3. Simulation study

The aim of the simulation is to construct several scenarios, including the corruption from both outliers and time sparseness, and to demonstrate the robustness of the proposed algorithm in comparison with existing methods. As to the assessment indexes, the adjusted Rand index (ARI, Hubert and Arabie, 1985; Ferreira and Hitchcock, 2009), percentage of correct outlier detection (p_c, i.e., the number of correctly detected outliers divided by the number of outliers), and percentage of false outlier detection (p_f, i.e., the number of falsely detected outliers divided by the number of nonoutliers) are adopted to assess the goodness of clustering and outlier detection. The settings with outliers and time sparseness are considered in Section 3.1, and the comparison results are given in Section 3.2. The RTLP, the ETD-based DBSCAN, agglomerative hierarchical clustering, and K-medoids methods, and the model-based funHDDC (Schmutz et al., 2020) are examined. The optimal K (or

θ) was determined by the average silhouette value for the ETD-based agglomerative hierarchical clustering and K-medoids (or ETD-based RTLP and DBSCAN) and by the BIC for funHDDC.

3.1. Simulation settings

Here, six types of scenarios are listed, and the contamination and the time sparseness are introduced in each scenario. The considered clustering scenarios are as follows: amplitude variation, phase variation, horizontal shift, clover petals as closed curves, cyclone tracks mimicked by the curves starting from a similar location and diverting to various destinations, and helixes with different radii.

For the sample f_i ($i = 1, \ldots, N$) evaluated at the common time grid $st = \{\frac{k-1}{T-1}, k = 1, \ldots, T\}$, $f_i = m_i + e_i$. For $t \in st$, $m(t) = (m^{(1)}(t), \ldots, m^{(p)}(t))^{\top}$, and $e_i(t) = (e^{(1)}(t), \ldots, e^{(p)}(t))^{\top}$. The measurement errors evaluated at the standard grid are denoted as $e = (e^1, e^2, \ldots, e^p)^{\top}$, where $e^k = \{e^{(k)}(t), t \in st\}$. It is assumed that e follows a normal distribution $\mathcal{N}_{pT}(0, \Sigma)$, where Σ is a covariance matrix consisting of blocks $\Sigma_{i,j} \in \mathbb{R}^{T \times T}$ for $1 \le i, j \le p$, and $\Sigma_{i,j}(s, t)$ denotes the covariance between $e^{(i)}(s)$ and $e^{(j)}(t)$. Assume that $\Sigma_{i,j}(s, t)$ follows the Matérn cross-covariance (Gneiting et al., 2010) function as follows:

$$\Sigma_{i,j}(s, t) = \begin{cases} \sigma_i^2 \mathcal{M}(|s - t|; \nu_i, \eta_i), & 1 \le i = j \le p, \\ \rho_{i,j} \sigma_i \sigma_j \mathcal{M}(|s - t|; \nu_{i,j}, \eta_{i,j}), & 1 \le i \ne j \le p, \end{cases}$$

where $\mathcal{M}(h; \nu, \eta) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\eta h)^{\nu} \mathcal{K}_{\nu}(\eta h)$, $h = |s - t| \in [0, 1]$, and $\mathcal{K}_{\nu}$ is a modified Bessel function of the second kind of order ν. Here, σ^2 is the marginal variance, $\nu > 0$ adjusts smoothness, and $\eta > 0$ is a scale parameter. Let $\nu_{i,j} = \frac{1}{2}(\nu_i + \nu_j)$, and ν_i is generated from the uniform distribution, $\nu_i \sim \mathcal{U}(0.2, 0.3)$. Let the matrix $(\beta_{i,j})_{i,j=1}^p$ be symmetric and nonnegative definite with diagonal elements $\beta_{i,i} = 1$ for $i = 1, \ldots, p$, and nondiagonal elements $\beta_{i,j} \sim \mathcal{U}(0, 1)$ for $1 \le i < j \le p$, and $\rho_{i,j} = \beta_{i,j} \frac{\Gamma(\nu_i + \frac{p}{2})^{1/2}}{\Gamma(\nu_i)^{1/2}} \frac{\Gamma(\nu_j + \frac{p}{2})^{1/2}}{\Gamma(\nu_j)^{1/2}} \frac{\Gamma\{\frac{1}{2}(\nu_i + \nu_j)\}}{\Gamma\{\frac{1}{2}(\nu_i + \nu_j) + \frac{p}{2}\}}$ for $1 \le i \le j \le p$.

Let the number of variables $p = 3$, and $T = 50$ equidistant points in [0,1]; the number of objects $N = 150$; the number of clusters $K = 3$. Here, $\sigma_1^2 = 0.05$, $\sigma_2^2 = 0.2$, and $\sigma_3^2 = 0.3$. Each cluster has an equal number of samples. For $k = 1, \ldots, K$ and $\nu = 1, \ldots, p$, $l \sim \mathcal{U}(1, p)$, and $r \sim \mathcal{B}(1/2)$, where $\mathcal{B}(b)$ represents a Bernoulli distribution with probability b. All six clustering scenarios are described as follows:

Scenario 1 (amplitude variation): $m^{(\nu)}(t) = 2\nu \cos\{(l + r\nu/4)\pi t\} + (-1)^r 3k\nu$.

Scenario 2 (phase variation): $m^{(\nu)}(t) = 2\nu \cos\{(l + r\nu/4)\pi h(t)\}$. Let $h(t) = \log_2(t + 1)$ when $k = 1$; $h(t) = t^2$ when $k = 2, \nu = 1$; $h(t) = 1 - \cos(\pi t/2)$ when $k = 2, \nu = 2$; $h(t) = \sin^2(\pi t/2)$ for $k = 2, \nu = 3$; $h(t) = t^3$ for $k = 3, \nu = 1$; $h(t) = \sin(\pi t/2)$ for $k = 3, \nu = 2$; and $h(t) = t$ for $k = 3, \nu = 3$.

Scenario 3 (shift variation): $m^{(\nu)}(t) = 6 \cos\{(l + \nu)\pi(t + 0.2lk)/2\}$.

Scenario 4 (clover petals): Let $m^{(1)}(t) = 5 \cos\{3w_k(t)\} \cos\{w_k(t)\}$, $m^{(2)}(t) = 5 \cos\{3w_k(t)\} \cdot \sin\{w_k(t)\}$, and $m^{(3)}(t) = 5 \cos\{3w_k(t)\}$, where $w_k(t) = 1.01(t + k - 1) + 0.548$.

Scenario 5 (cyclone tracks):
$$m^{(1)}(t) = \begin{cases} -5t - \sin(10\pi t), \\ 5\sin(10\pi t), \\ 5t + \sin(10\pi t), \end{cases} \quad m^{(2)}(t) = \begin{cases} 11t^2, \\ 7t, \\ 5\log_2(t + 1), \end{cases} \quad m^{(3)}(t) = \begin{cases} 7\log_2(t + 1), & k = 1, \\ 14\log_2(t + 1), & k = 2, \\ 21\log_2(t + 1), & k = 3. \end{cases}$$

Scenario 6 (helixes with variable radii):
$$m^{(1)}(t) = \begin{cases} 5\cos(10\pi t), \\ 5t\cos(20\pi t + 10), \\ 5t\sin(20\pi t + 10), \end{cases} \quad m^{(2)}(t) = \begin{cases} 5.5\sin(10\pi t), \\ 5t\sin(20\pi t + 10), \\ 5\log_2(t + 1), \end{cases} \quad m^{(3)}(t) = \begin{cases} 10t, & k = 1, \\ 10 - 10t, & k = 2, \\ 10t, & k = 3. \end{cases}$$

Scenarios 4 and 5 are representatives of closed and nonclosed curves. The visualization (Fig. 3) and the clustering results of Scenarios 4 and 5 are provided in this paper; the results for the remaining scenarios are provided in the supplementary material.

Next, three types of magnitude outliers (Contaminations 1-3), and three types of shape outliers (Contaminations 4-6) are introduced, with a 10% proportion of samples in the aforementioned clustering scenarios. The visualization of Scenario 4 with various outlier settings is shown in Fig. 4. The first and second rows in Fig. 4 reveal the outliers showing abnormalities in magnitude and shape, respectively. It is noticed that outliers in Contaminations 1-4 are scattered and those in Contaminations 5-6 are concentrated.

Contamination 1 (pure outlier): $\widetilde{m}_i^{(\nu)}(t) = m_i^{(\nu)}(t) + rh$. Here r is assigned to 1 if $\mathcal{U}(-1, 1) > 0$, otherwise r is -1, and $h = \max\{|\min(m_i^{(\nu)})(t)|, |\max(m_i^{(\nu)})(t)|\}/2$. Then, r and h below are generated in the same way as Contamination 1.

Contamination 2 (peak outlier): $\widetilde{m}_i^{(\nu)}(t) = \begin{cases} m_i^{(\nu)}(t) + rh, & t \in [st, st + 0.1], \\ m_i^{(\nu)}(t), & \text{otherwise}, \end{cases}$ where $st \sim \mathcal{U}(0, 0.9)$.

Contamination 3 (partial outlier): $\widetilde{m}_i^{(\nu)}(t) = \begin{cases} m_i^{(\nu)}(t) + rh, & t \in [st, 1], \\ m_i^{(\nu)}(t), & \text{otherwise}, \end{cases}$ where $st \sim \mathcal{U}(0, 0.5)$.
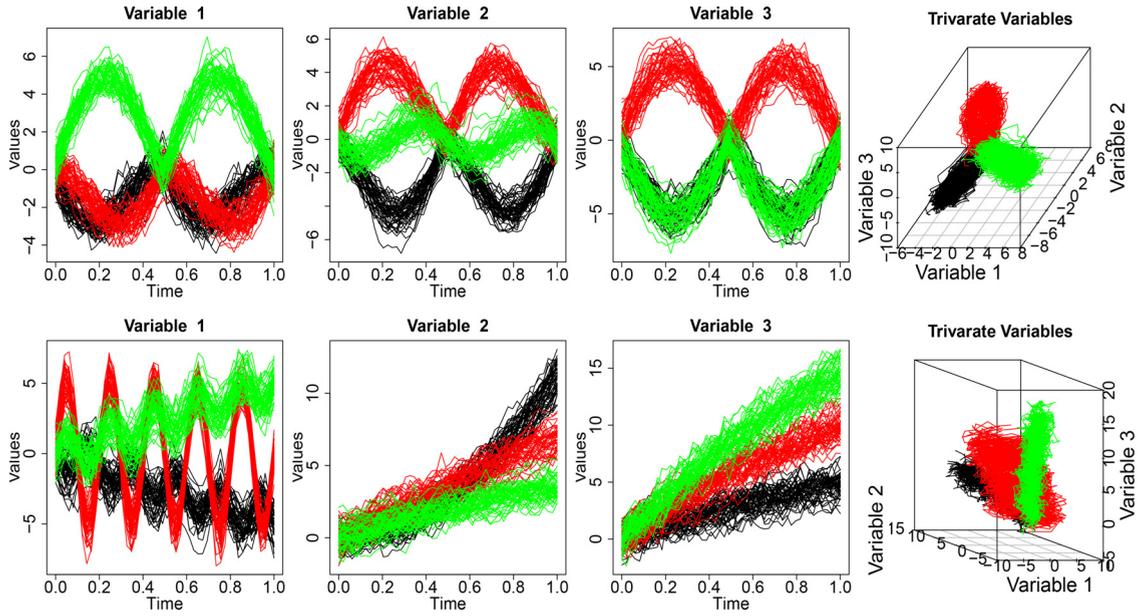
Fig. 3. Top and bottom rows represent samples from Scenarios 4 and 5, respectively. Three clusters are represented in black, red, and green.
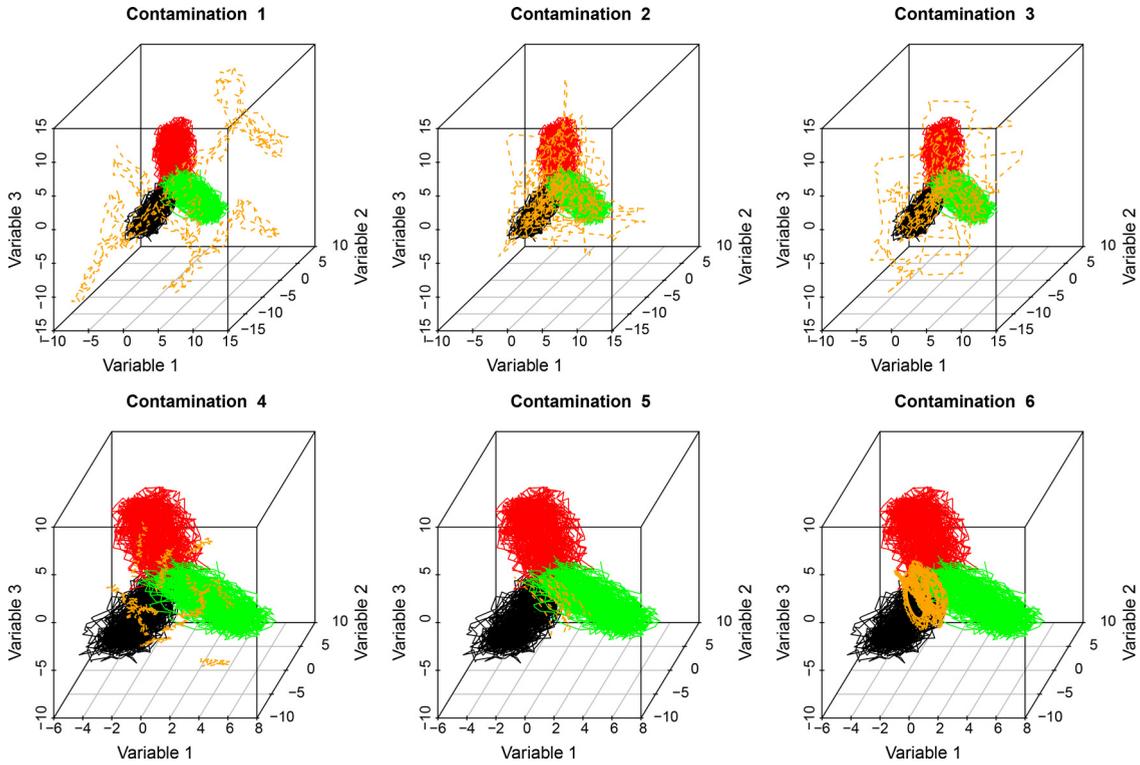

Fig. 4. Visualization of Scenario 4 with a proportion 10% of outliers. Three clusters are labeled in solid black, red, and green, whereas the outliers are labeled in dashed orange.

Contamination 4 (shape outlier I): $c \sim \mathcal{U}(\min(m_i^{(v)})(t)/2, \max(m_i^{(v)})(t)/2)$, $s \sim \mathcal{U}(-2v, 2v)$, $\widetilde{m}_i^{(v)}(t) = c + st + \epsilon_i(t)$, and $\epsilon_i(t) \sim \mathcal{U}(-0.3, 0.3)$.

Contamination 5 (shape outlier II): $\widetilde{m}_i^{(1)}(t) = s^{(1)}(t) + h\cos(0.5\pi t) + a$, $\widetilde{m}_i^{(2)}(t) = s^{(2)}(t) + h\sin(0.5\pi t) + a$, and $\widetilde{m}_i^{(3)}(t) = s^{(3)}(t) - h\cos(0.5\pi t) + a$, where $s^{(v)}(t) = \min(m_i^{(v)})(t)/2 + \max(m_i^{(v)})(t)/2$ $(v = 1, 2, 3)$, and $a \sim \mathcal{U}(-h, 0)$. The $s^{(v)}(t)$ below is generated in the same way.
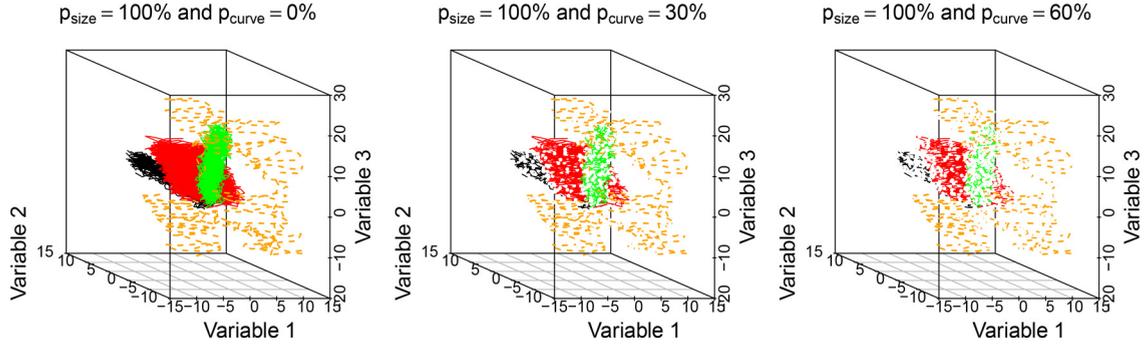
Fig. 5. Visualization of Scenario 5 with 10% pure magnitude outliers for $p_{curve} = 0\%, 30\%$, and 60% separately and $p_{size} = 100\%$ in all above settings. The three clusters are shown in black, red, and green, and the outliers are indicated in orange.

Contamination 6 (shape outlier III): $\widetilde{m}_i^{(1)}(t) = s^{(1)}(t) + h\cos(30\pi t) + a$, $\widetilde{m}_i^{(2)}(t) = s^{(2)}(t) + h\sin(30\pi t) + a$, and $\widetilde{m}_i^{(3)}(t) = s^{(3)}(t) - h\cos(30\pi t) + a$, where $a \sim \mathcal{U}(-h, 0)$.

In addition, the parameters p_{size} and p_{curve} are introduced to ensure that each curve has unique time measurements in order to verify that the RTLP clustering algorithm can handle sparse multivariate functional data. Here, p_{size} represents the number of curves with missing values divided by the total number of observed curves, and p_{curve} represents the number of missing points for curves with missing values divided by the cardinality of the standard time grid. The simulation for $p_{size} = 100\%$ and p_{curve} is conducted ranging from 0% up to 60%.

Fig. 5 shows the visualization of Scenario 5 with 10% proportion of pure magnitude outliers for different time sparseness $p_{curve} = 0\%, 30\%$, and 60% and $p_{size} = 100\%$, respectively. As data trajectories become sparser, the clustering patterns are harder to recognize, and the clustering is expected to become more challenging.

3.2. Clustering performance

The ARI, p_c, and p_f are applied to assess the goodness of clustering and outlier detection. ARI can be applied when the actual clusters are known. It estimates the matching between the actual partition and the estimated partitions, and it is zero in the case of random partitions and one in the case of perfect agreement between two partitions.

To illustrate the robustness of the proposed algorithm, ARI is compared among ETD-based RTLP, DBSCAN, agglomerative hierarchical clustering, K-medoids methods, and funHDDC. In particular, these five methods are applied to multivariate functional data and to each marginal functional data separately. The ARI for multivariables and the average ARI over marginal variables are obtained to evaluate the investigated method under the two situations. The method applied to multivariate (univariate) functional data is denoted as its multivariate (univariate) version. Note that the difference between ETD-based multivariate clustering methods and univariate clustering ones starts in the calculation of ETD; see step 1) in Section 2.4.

Figs. 6 and 7 show the boxplots of ARI^2 for Scenarios 4 and 5 under different outlier contaminations and missing time measurements. Overall, the ETD-based clustering methods perform better than funHDDC, and the multivariate RTLP exhibits the best performance in achieving the ARI with the highest mean and has a very low standard deviation in almost all cases. The results are viewed from horizontal and vertical perspectives.

From a horizontal perspective of Figs. 6 and 7, there are three summaries under the same time measurement setting. First, the multivariate ETD-based methods perform better than their corresponding univariate marginal methods in the clustering, as seen from the median of the boxplots. Second, the ETD-based clustering methods have better clustering precision than funHDDC, as seen from the differences among the boxplots in terms of the median and range. Third, outlier contaminations may lead to a drop in the clustering precision depending on the outlier types. For example, Contamination 5 in Scenarios 4 and 5 influence the clustering precision of all ETD-based methods, even when $p_{curve} = 0\%$. In most cases, RTLP is rarely influenced except for Contamination 5 in Scenario 4 and Scenario 5 with $p_{curve} = 60\%$.

From a vertical perspective of Figs. 6 and 7, two patterns are observed. First, the clustering pattern influences the clustering precision of the ETD-based methods when there are missing values. The precision of clustering in Scenario 4 remains almost identical when p_{curve} changes from 0% to 30% and then to 60%, except for Contamination 5; the multivariate RTLP exhibits the best performance, followed by multivariate DBSCAN, multivariate agglomerative hierarchical clustering, and multivariate K-medoids. However, the clustering precision of multivariate RTLP and multivariate DBSCAN in most contaminations in Scenario 5 with $p_{curve} = 60\%$ is not as good as multivariate agglomerative hierarchical clustering and multivariate K-medoids. From the pattern of Scenario 5 (Fig. 5), it is speculated that because of too much sparseness, ETD between a curve in the middle cluster and a curve at the edge of the remaining clusters gets small and the multivariate RTLP cannot separate curves into the correct three clusters regardless of θ. However, the multivariate agglomerative hierarchical clustering seeks to build a hierarchy of clusters from each observation as a separate cluster to all observations merged into one cluster, and the multivariate K-medoids considers the case of three clusters before determining the optimal number of clusters. Second,
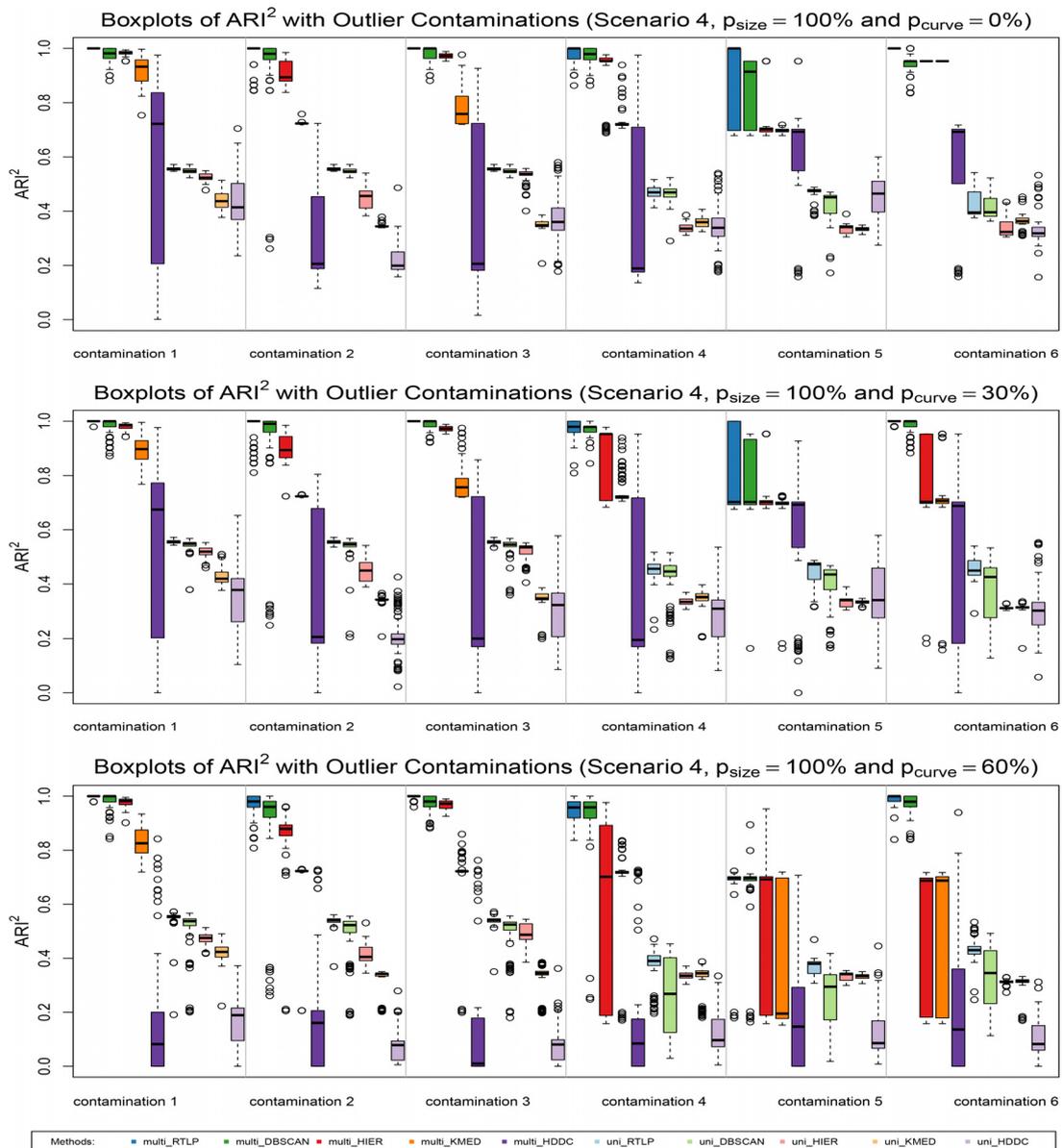
Fig. 6. Panels from the top to bottom display the boxplots of ARI^2 in Scenario 4 for $p_{curve} = 0\%$, 30% and 60%. Ten methods are compared under all settings with six contaminations. Here, $K = 3$, and there are 100 simulation replicates. The methods, from left to right, are the multivariate and average marginal univariate versions of RTLP, DBSCAN, agglomerative hierarchical clustering, K-medoids, and funHDDC methods.

multivariate and univariate funHDDC methods are easily affected by outliers when there are no missing values, and their clustering performances are worse when there are missing values, which can be seen from the movement of the boxplots vertically.

The multivariate RTLP usually behaves the best when the curve sparseness p_s is not greater than $50\% - 60\%$. Although the results are not reported in this paper, the running time is stable for all ETD-based clustering methods but shows a significant standard deviation for funHDDC. Additionally, the multivariate RTLP achieves the best clustering performance under scenarios with amplitude and phase variation. Although it works well for the considered misaligned data when phase/amplitude variations are present (see supplementary material), more complex misaligned data structures are deferred to future research to fully resolve the challenge of sparse multivariate functional data clustering specifically for misaligned data. It is currently believed that the RTLP method can deal with mildly misaligned data. Hence, multivariate RTLP is recommended when the data objects, i.e., multivariate functional data, are not too sparse, and it is acknowledged that the ETD-based multivariate DBSCAN and multivariate agglomerative hierarchical clustering are good alternatives for very sparse cases.
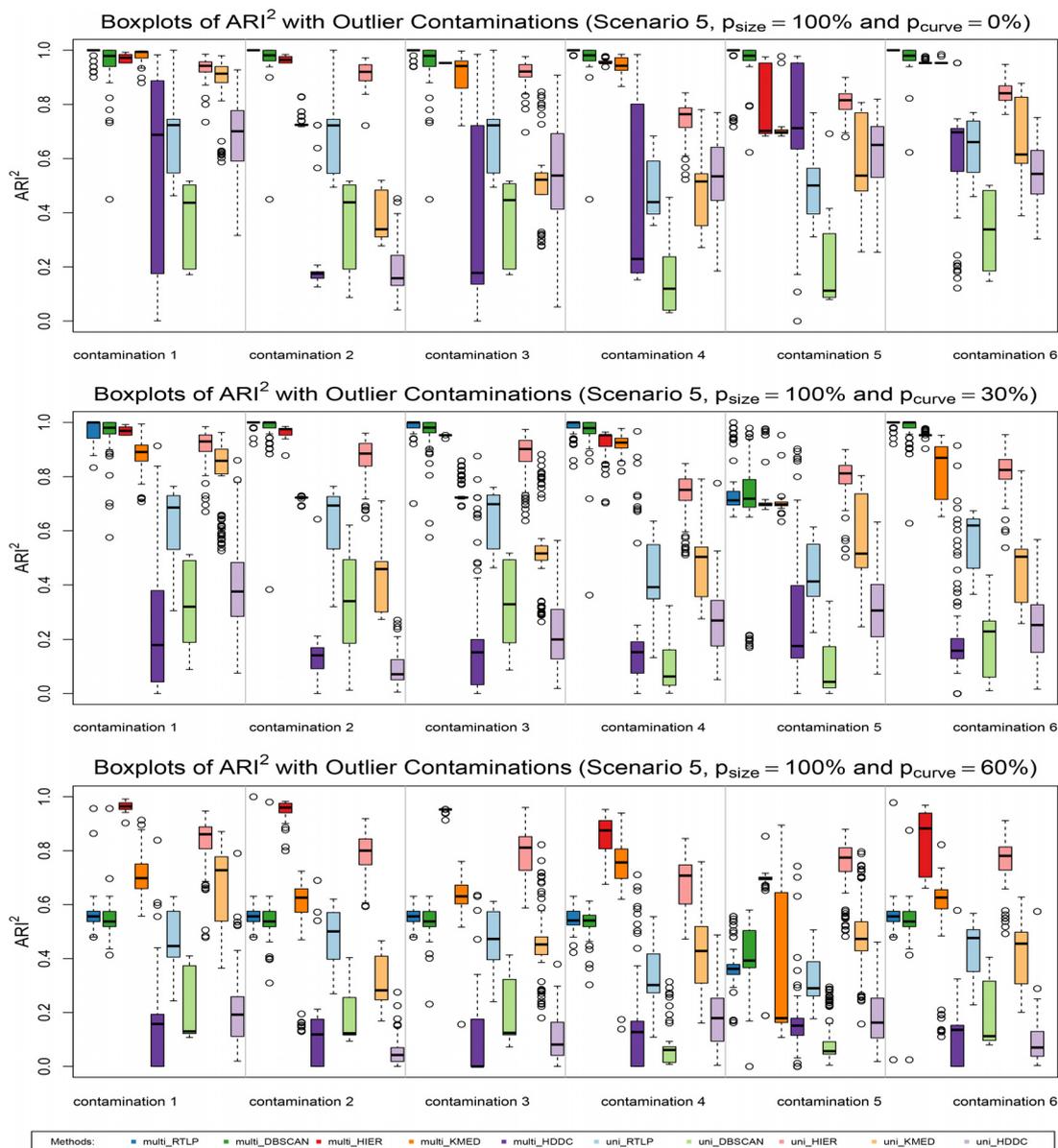
Fig. 7. Panels from the top to bottom display the boxplots of ARI^2 in Scenario 5 for p_{curve} = 0%, 30% and 60%. Ten methods are compared under all settings with six contaminations. Here, $K = 3$, and there are 100 simulation replicates. The methods, from left to right, are the multivariate and average marginal univariate versions of RTLP, DBSCAN, agglomerative hierarchical, K-medoids, and funHDDC methods.

3.3. Outlier detection performance

With regard to outlier detection, p_c and p_f are reported only for multivariate and marginal RTLP and DBSCAN methods because of their capabilities of recognizing outliers. Specifically, p_c (p_f) is obtained from the marginal RTLP (DBSCAN) for each variable and the average p_c (p_f) over different variables is used to represent the performance of univariate RTLP (DBSCAN) in the outlier detection.

Tables 1 and 2 list p_c and p_f in Scenarios 4 and 5 when p_{curve} increases from 0% to 30% and 60%. Three phenomena are worthy to point out. First, the multivariate DBSCAN usually exhibits the highest true outlier detection percentage, and the multivariate RTLP exhibits the lowest false outlier detection percentage. This is due to the difference between DBSCAN and RTLP. DBSCAN first regards all noncore observations as potential outliers. Then, some observations from potential outliers are labeled as boundary observations of clusters during the clustering process, and the remainings are outliers. Comparatively, RTLP did not obtain potential outliers after obtaining primary clusters. Then, RTLP used a criterion to check whether potential outliers remain as outliers. Hence, the probability of a curve to be falsely (correctly) detected as an outlier is higher

Table 1
Correct outlier detection percentage p_c (%) and false outlier detection percentage p_f (%) for multivariate and univariate RTLP and DBSCAN with different outlier contaminations for Scenario 4. Higher p_c and lower p_f values for each setting given the contamination and p_{curve} are indicated in bold. The proportion of outliers is 10%, and p_{size} is 100% for all settings. Simulations were conducted with 100 replicates.

p_{curve}	Methods	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multi_RTLP	**100.0** (0.0)	**0.0** (0.0)	92.4 (17.1)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)
	multi_DBSCAN	**100.0** (0.0)	0.8 (1.1)	92.4 (17.1)	0.6 (1.0)	**100.0** (0.0)	0.8 (1.1)
	uni_RTLP	**100.0** (0.0)	0.0 (0.1)	**99.9** (0.4)	0.0 (0.1)	**100.0** (0.0)	0.0 (0.1)
	uni_DBSCAN	**100.0** (0.0)	0.5 (0.4)	**99.9** (0.4)	0.5 (0.4)	**100.0** (0.0)	0.5 (0.4)
30%	multi_RTLP	**100.0** (0.0)	**0.0** (0.0)	93.8 (15.5)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)
	multi_DBSCAN	**100.0** (0.0)	0.7 (1.1)	93.8 (15.5)	0.5 (0.9)	**100.0** (0.0)	0.5 (0.7)
	uni_RTLP	**100.0** (0.0)	**0.0** (0.1)	99.5 (1.2)	0.0 (0.1)	99.4 (1.3)	0.0 (0.1)
	uni_DBSCAN	**100.0** (0.0)	0.6 (0.6)	**99.5** (1.1)	0.6 (0.6)	99.4 (1.3)	0.6 (0.7)
60%	multi_RTLP	**100.0** (0.0)	**0.0** (0.1)	89.9 (14.4)	**0.0** (0.1)	99.3 (2.0)	**0.0** (0.1)
	multi_DBSCAN	**100.0** (0.0)	0.7 (1.1)	89.9 (14.5)	0.5 (1.1)	**99.4** (1.9)	0.8 (1.1)
	uni_RTLP	99.8 (1.6)	0.1 (0.2)	94.4 (3.2)	0.1 (0.2)	94.4 (3.4)	0.1 (0.2)
	uni_DBSCAN	99.9 (0.7)	1.1 (1.1)	**94.5** (3.2)	0.9 (0.9)	94.9 (3.2)	1.0 (1.0)

p_{curve}	Methods	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multi_RTLP	**92.5** (11.3)	**0.0** (0.0)	**63.3** (48.7)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)
	multi_DBSCAN	93.3 (10.6)	0.4 (1.0)	0.0 (0.0)	0.4 (0.7)	18.4 (39.1)	0.8 (1.1)
	uni_RTLP	65.5 (11.2)	0.1 (0.1)	61.8 (11.8)	0.1 (0.1)	44.0 (16.3)	0.0 (0.1)
	uni_DBSCAN	67.9 (9.0)	0.4 (0.4)	4.1 (13.0)	0.5 (0.5)	7.8 (14.4)	0.2 (0.2)
30%	multi_RTLP	91.3 (11.5)	**0.0** (0.0)	30.0 (46.1)	**0.0** (0.1)	99.9 (0.9)	**0.0** (0.1)
	multi_DBSCAN	**92.9** (8.6)	0.2 (0.5)	0.0 (0.0)	0.2 (0.6)	99.9 (0.9)	0.6 (1.0)
	uni_RTLP	59.1 (12.4)	0.1 (0.2)	**56.3** (17.4)	0.1 (0.2)	62.6 (12.4)	0.1 (0.1)
	uni_DBSCAN	61.4 (9.6)	0.4 (0.6)	3.0 (9.6)	0.7 (0.7)	64.8 (12.8)	0.4 (0.4)
60%	multi_RTLP	**82.3** (14.5)	**0.0** (0.2)	0.0 (0.0)	**0.1** (0.3)	97.3 (6.8)	**0.0** (0.1)
	multi_DBSCAN	82.3 (16.1)	0.2 (0.5)	0.0 (0.0)	0.3 (0.8)	**97.1** (7.2)	0.7 (1.2)
	uni_RTLP	31.6 (13.2)	0.2 (0.3)	**22.3** (16.1)	0.3 (0.3)	54.3 (11.3)	0.2 (0.2)
	uni_DBSCAN	37.4 (11.3)	0.6 (0.7)	0.3 (3.3)	1.3 (1.5)	53.7 (9.4)	0.8 (1.3)

Table 2
Correct outlier detection percentage p_c (%) and false outlier detection percentage p_f (%) for multivariate and univariate RTLP and DBSCAN with different outlier contaminations for Scenario 5. Higher p_c and lower p_f values for each setting given the contamination and p_{curve} are indicated in bold. The proportion of outliers is 10%, and p_{size} is 100% for all settings. Simulations were conducted with 100 replicates.

p_{curve}	Methods	Contamination 1		Contamination 2		Contamination 3	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multi_RTLP	97.8 (5.0)	0.2 (0.5)	**100.0** (0.0)	**0.0** (0.0)	98.9 (2.5)	**0.2** (0.4)
	multi_DBSCAN	**100.0** (0.0)	2.2 (4.2)	**100.0** (0.0)	1.3 (3.6)	**100.0** (0.0)	2.2 (4.2)
	uni_RTLP	99.9 (1.6)	**0.2** (0.4)	99.3 (3.1)	0.2 (0.3)	**100.0** (0.0)	**0.2** (0.4)
	uni_DBSCAN	**100.0** (0.0)	0.4 (0.6)	99.3 (3.1)	0.4 (0.5)	**100.0** (0.0)	0.4 (0.6)
30%	multi_RTLP	96.4 (6.7)	**0.5** (0.8)	99.9 (0.9)	**0.1** (0.4)	98.7 (2.7)	**0.4** (1.4)
	multi_DBSCAN	**100.0** (0.0)	1.6 (3.9)	99.9 (0.9)	0.6 (1.0)	**100.0** (0.0)	1.5 (4.0)
	uni_RTLP	97.8 (4.8)	0.8 (2.0)	95.8 (5.3)	0.5 (1.4)	98.0 (2.4)	0.5 (1.4)
	uni_DBSCAN	99.6 (1.2)	1.2 (2.2)	96.1 (5.3)	0.7 (1.3)	98.2 (2.4)	0.6 (1.2)
60%	multi_RTLP	99.8 (2.0)	32.7 (4.5)	**100.0** (0.0)	33.0 (3.6)	**99.8** (1.1)	33.3 (1.2)
	multi_DBSCAN	**100.0** (0.0)	33.5 (4.1)	**100.0** (0.0)	33.8 (3.9)	**99.8** (1.1)	34.1 (2.2)
	uni_RTLP	98.5 (4.0)	**11.5** (0.9)	95.4 (3.5)	11.4 (1.0)	96.1 (3.5)	**11.3** (1.4)
	uni_DBSCAN	**100.0** (0.2)	11.6 (0.9)	95.5 (3.4)	**11.4** (0.6)	96.2 (3.5)	11.4 (1.2)

p_{curve}	Methods	Contamination 4		Contamination 5		Contamination 6	
		p_c	p_f	p_c	p_f	p_c	p_f
0%	multi_RTLP	**99.7** (1.3)	**0.0** (0.0)	89.1 (28.7)	**0.0** (0.0)	**100.0** (0.0)	**0.0** (0.0)
	multi_DBSCAN	**99.7** (1.3)	1.2 (3.6)	**95.6** (14.4)	0.7 (2.3)	**100.0** (0.0)	1.1 (2.4)
	uni_RTLP	37.5 (14.5)	0.2 (0.5)	38.2 (19.9)	0.3 (0.4)	91.7 (14.1)	0.2 (0.4)
	uni_DBSCAN	44.3 (13.3)	0.1 (0.2)	19.3 (14.2)	0.4 (1.0)	68.7 (8.1)	0.5 (0.7)
30%	multi_RTLP	97.7 (5.4)	**0.3** (0.8)	21.0 (34.7)	**0.4** (0.7)	**100.0** (0.0)	**0.2** (0.5)
	multi_DBSCAN	**97.9** (4.7)	1.0 (1.5)	29.1 (35.4)	1.1 (1.8)	**100.0** (0.0)	0.9 (2.3)
	uni_RTLP	36.0 (11.4)	4.1 (4.9)	**31.4** (14.7)	2.2 (3.7)	66.8 (2.3)	0.6 (0.9)
	uni_DBSCAN	39.2 (10.7)	4.8 (4.8)	21.3 (12.2)	3.0 (3.9)	66.7 (0.0)	1.3 (1.7)
60%	multi_RTLP	96.3 (9.0)	33.3 (1.2)	19.2 (31.3)	32.3 (5.4)	**100.0** (0.0)	32.6 (4.8)
	multi_DBSCAN	**97.9** (4.4)	33.9 (2.2)	**40.4** (36.7)	32.4 (5.5)	**100.0** (0.0)	33.4 (4.8)
	uni_RTLP	33.8 (12.7)	11.4 (0.7)	34.0 (16.2)	11.7 (0.8)	71.3 (7.6)	**11.5** (0.7)
	uni_DBSCAN	37.9 (11.4)	**11.2** (0.5)	22.6 (12.8)	**11.6** (1.0)	76.2 (8.2)	11.5 (0.9)

in DBSCAN. Second, when shape outliers are contaminated, the multivariate RTLP and DBSCAN have much better outlier detection performance compared to their marginal method, with a significantly higher p_c and lower p_f. Third, multivariate DBSCAN achieves $p_c = 0$ in Scenario 4 with Contamination 5 because it regards all outliers as one cluster instead of the outliers. Although DBSCAN and RTLP share p_m, DBSCAN uses p_m to determine core observations, while RTLP uses p_m to determine the primary clusters. From the empirical simulation, the RTLP is more robust in outlier detection compared to ETD-based DBSCAN under common θ and p_m.

Given the excellent performance of RTLP in the outlier-resistant clustering and outlier detection, and stable running time regardless of outlier settings and irregular time settings (running time is negatively proportional to p_{curve}), RTLP is highly recommended as an outlier-resistant clustering method for multivariate functional data when the average p_{curve} is small to moderate. Furthermore, RTLP can effectively block outliers in the primary clusters if there are concerns about outliers.
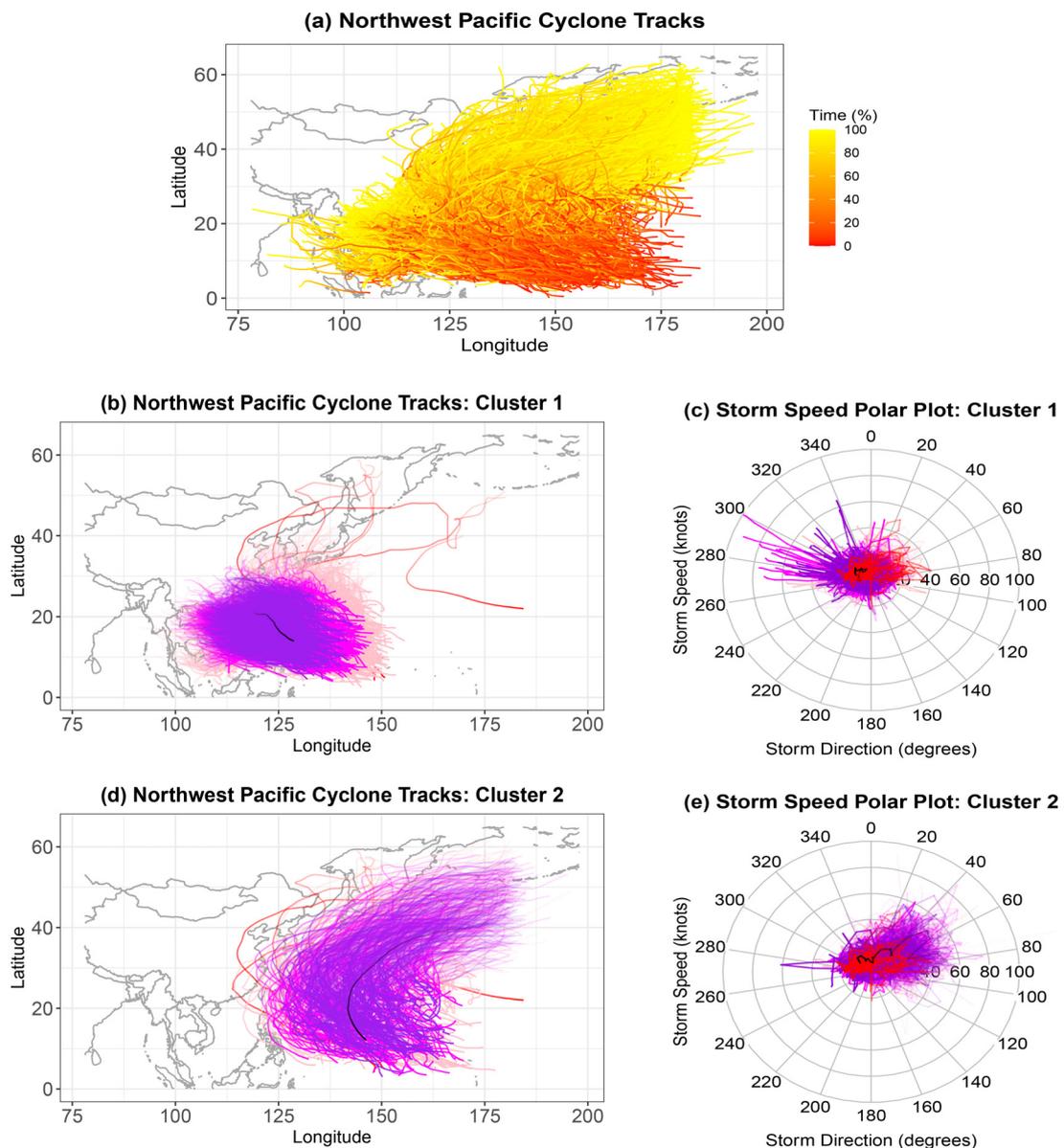

Fig. 8. (a) 4051 Northwest Pacific cyclone tracks from 1884 to 2021. Each curve represents a full cyclone process, and the color changes gradually from red to yellow with time. (b) and (d) Cyclone trajectories obtained using revised version of the trajectory boxplot (Yao et al., 2020). (c) and (e) Wind speed polar plot, with wind speed as the radius and wind speed direction as the angle. The black and red represent the median and outliers, respectively, and purple, magenta, and pink indicate the first, second, and third quartile curves, respectively.
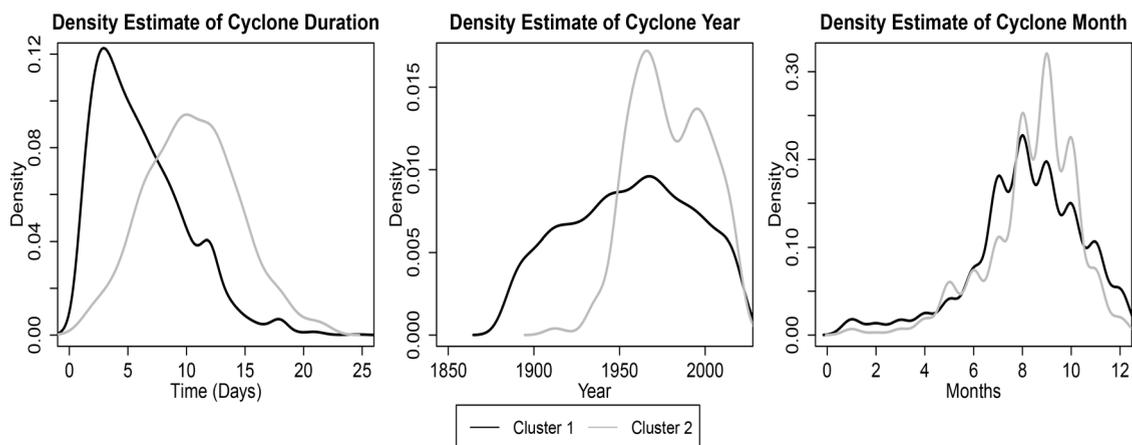
Z. Qu, W. Dai and M.G. Genton

Fig. 9. Extra time information from clusters 1 and 2 in the Northwest Pacific cyclones. The panels display the nonparametric density estimates of the lasting days, occurrence year, and occurrence month. The first cluster is labeled in black, and the second is in grey.

4. Application to northwest pacific tropical cyclone tracks

A cyclone (American Meteorological Society, 2000) is a large-scale air mass that rotates around an intense center of low-atmospheric-pressure center. The word "tropical" in the term "tropical cyclone" refers to the geographical origin of these systems, which form almost exclusively over tropical seas. The rising and cooling of the swirling air results in formation of clouds and precipitation. Although most cyclones have a life cycle of 3-7 days (Storm Science Australia, 2008), some systems can last for weeks if the environment is favorable. The global tropical cyclones tracks from version 4 of the International Best Track Archive for Climate Stewardship (IBTrACS v04, Knapp et al., 2010) are available online. 4051 Northwest Pacific (see Fig. 8(a)) cyclone tracks between 1884 and 2021 are extracted for clustering analysis. The data are usually measured every three hours. The time of generation and disappearance are aligned to be 0 and 1, respectively, for each trajectory.

Using RTLP, θ is obtained to be 0.22 according to the maximum silhouette value when θ changes from 0.01 to 0.25 in steps of 0.01 and with $p_m = 0.1$ and $\alpha = 0.85$. Two clusters were obtained according to the geographic trajectories of the cyclone centers; the first cluster has 3417 curves, and the second one has 627 curves. The clustering pattern is consistent with the pattern reported by Misumi et al. (2019). Cyclones in the first cluster (Figs. 8(b) and (c)) mainly originate from the ocean to the east of the Philippines islands and sweep into inland Asia. Usually, they end in either eastern or southwestern China, blocked by a high-pressure system from the Chinese mainland. Cyclones in the second cluster (Figs. 8(d) and (e)) barely pass by mainland Asia, move toward North America, and disappear somewhere in the North Pacific. Seven curves that cannot be recognized in any cluster remain as outliers. They mainly show abnormal patterns or originate from a location far from the Philippines.

Fig. 8 also maps the cyclone tracks and their storm speed information per cluster. In the storm speed polar plot, the storm direction is taken as the angle and the storm speed norm as the radius. The storm moves in the direction pointed by the vector, in degrees east of north from 0 to 360 degrees. In addition, different colors are used to denote the number of neighbours inside a cluster, with black for the median, and purple, magenta, and pink to indicate the first, second, and third quartile regions, respectively, and red to indicate outliers.

The medians show the typical tracks in a cluster. The median in the first cluster (see Fig. 8(b)) is directed towards the northwest (direction between 270 and 360 degrees) with a storm speed of almost 10 knots, which is common for most cyclones in the first cluster. Usually, the wind speeds (i.e., the radius of the polar plot) of cyclones traveling inland decreases slowly because of reduction in the power of the cyclones. However, the median in the second cluster (see Fig. 8(d)) first shifts northwest and then turns eastwards (direction between 0 and 180 degrees), with the storm speed (i.e., the radius of the polar plot) increasing after the cyclones steer eastward. The cyclones in the second cluster that do not move inland but move to the North Pacific are influenced by the northward shift of high pressure over the Pacific (Times, 2022).

Besides location and storm speed data, the lasting days and time of occurrence (Fig. 9) of the cyclones are used to analyze the difference between the clusters. The first group of cyclones mostly lasts less than ten days, and the second group mainly lasts between five and fifteen days. On average, cyclones from the first cluster (Fig. 8(b)) travel a shorter distance than those from the second cluster (Fig. 8(d)). The frequency of cyclones in both groups increased starting from the last half of the 1940s to around 1960 (AOKI, 1985), especially that of the cyclones that passed by Japan and Korea and reached the Sea of Okhotsk. A remarkable decreasing trend in the cyclone frequency is noted between the late 1960s to around 1980. Subsequently, the cyclone frequency remained constant since the 1980s. Fig. 9 shows that cyclones belonging to the second cluster mainly occurred after the 1950s; however, the ratio of the frequency of cyclones traveling to mainland Asia and directed to the North Pacific is around two versus one after 1950. High pressure (Times, 2022) in the Sea of Okhotsk in eastern Russia and low pressure to the east of the Philippines have induced shifts in the direction of cyclones

since the 1950s. Cyclones in the first cluster mainly occur from July to November, whereas those in the second cluster are concentrated between August and October. The increasing frequency of the cyclones is assumed to be related to global warming (Van Aalst, 2006), which has also influenced the rise of seawater levels and surging tides smacked in low-lying areas.

5. Discussion

An ETD measure was defined for complete and sparse multivariate functional data. With ETD, two algorithms (two-layer partition clustering, and cluster and outlier recognition) were proposed and summarized as robust two-layer partition (RTLP) clustering. In addition, classical clustering methods such as DBSCAN, agglomerative hierarchical clustering, and K-medoids were naturally extended to sparse multivariate functional cases based on ETD.

The idea of RTLP is as follows. Neighbours of a curve in the set and the core in the set given a θ-quantile are defined for the algorithm application. First, the two-layer partition clustering algorithm merges curves into disjoint groups based on the concepts of neighbours and the core, and merges these groups into a set of disjoint clusters. Next, the cluster and outlier recognition algorithm determines the final primary clusters and outliers. The time complexity of the whole algorithm majorly lies in the calculation of the distance across N curves, which requires $\mathcal{O}(N^2)$ time; the time complexity can be reduced through parallel computation. Regarding parameter selection, the optimal θ is selected as the parameter with the highest revised average silhouette value.

Using simulations, the clustering results of RTLP with three other ETD-based methods were compared: DBSCAN, agglomerative hierarchical methods, K-medoids, and the model-based funHDDC. The RTLP method exhibited excellent performance in terms of both clustering curves and outlier detection; the performance decreased in the following order: the RTLP method, DBSCAN, agglomerative hierarchical clustering and K-medoid clustering, and finally, the funHDDC algorithm. The running time of ETD-based methods was stable for a fixed sample size, and it decreased as the time grid became sparser. Overall, RTLP has a significant advantage in terms of clustering precision regardless of the outliers and the time sparseness settings.

In future works, the RTLP and current clustering methods can be applied to more general sparse (multivariate) functional data. First, the location of missing values can be not completely random and the observed time grid may carry information relevant to the phenomenon under observation (more types of sparseness are described in Qu and Genton (2022)); moreover, the measurement can be observed for some variables but could be missing for other variables for one index, which requires a more general distance measure. It should be noted that although the current framework transforms the curves under observation to the distance, the relevance between the missing locations and missing observations may be worth further analysis. In addition, the aforementioned clustering methods can be applied to multivariate nonfunctional data. The bridge between the RTLP clustering method and multivariate nonfunctional data can be the revised L^p metric such as $p = 1$, $p = 2$, and $p = \infty$. Another research direction is to cluster images or surfaces (Chen et al., 2005; Genton et al., 2014); this would require defining the distance between the images.

Declaration of Competing Interest

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.

Acknowledgments

Appendix

Proof of Theorem 1. 1. $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) \geq 0$ from Eq. (1).

2. If $\boldsymbol{Y}_m = \boldsymbol{Y}_n$, then $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) = 0$ from Eq. (1).

If $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) = 0$, i.e., $\sup_{t \in \mathcal{T}}\left[\sqrt{\sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_n^{(l)}(t)\right\}^2}\right] = 0$, then $\sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_n^{(l)}(t)\right\}^2 = 0 \ \forall t \in \mathcal{T}$. Hence $\boldsymbol{Y}_m(t) = \boldsymbol{Y}_n(t)$ for $t \in \mathcal{T}$.

3. $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) = d(\boldsymbol{Y}_n, \boldsymbol{Y}_m)$ from Eq. (1).

4. Write $\sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_n^{(l)}(t)\right\}^2 = \sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_o^{(l)}(t) + Y_o^{(l)}(t) - Y_n^{(l)}(t)\right\}^2 = \sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_o^{(l)}(t)\right\}^2 + \sum_{l=1}^{p}\left\{Y_o^{(l)}(t) - Y_n^{(l)}(t)\right\}^2 + \sum_{l=1}^{p} 2\left\{Y_m^{(l)}(t) - Y_o^{(l)}(t)\right\}\left\{Y_o^{(l)}(t) - Y_n^{(l)}(t)\right\} \ \forall \ t \in \mathcal{T}$. In addition, $\sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_o^{(l)}(t)\right\}\left\{Y_o^{(l)}(t) - Y_n^{(l)}(t)\right\} \leq \sqrt{\sum_{l=1}^{p}\left\{Y_m^{(l)}(t) - Y_o^{(l)}(t)\right\}^2} \cdot \sqrt{\sum_{l=1}^{p}\left\{Y_o^{(l)}(t) - Y_n^{(l)}(t)\right\}^2}$. Hence, it is obtained that

$\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2 \leq \sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_o^{(l)}(t) \right\}^2 + \sum_{l=1}^{p} \left\{ Y_o^{(l)}(t) - Y_n^{(l)}(t) \right\}^2 + 2\sqrt{\sum_{l=1}^{p} \left\{ Y_n^{(l)}(t) - Y_o^{(l)}(t) \right\}^2} \cdot$

$\sqrt{\sum_{l=1}^{p} \left\{ Y_o^{(l)}(t) - Y_n^{(l)}(t) \right\}^2}$. This equals to $\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2 \leq \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_o^{(l)}(t) \right\}^2} + \sqrt{\sum_{l=1}^{p} \left\{ Y_o^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right]^2$.

Hence, $\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \leq \sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_o^{(l)}(t) \right\}^2} + \sqrt{\sum_{l=1}^{p} \left\{ Y_o^{(l)}(t) - Y_n^{(l)}(t) \right\}^2}$. The inequality holds after taking the supremum norm. By combining the inequality relation with Eq. (1), it is obtained that $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) \leq d(\boldsymbol{Y}_m, \boldsymbol{Y}_o) + d(\boldsymbol{Y}_o, \boldsymbol{Y}_n)$. □

Proof of Theorem 2. From Assumption 1, $g(t)$ is differentiable in $\mathcal{T}$, then for each $l = 2, \ldots, T_n - 1$, there exists a $t \in [t_{n,l}, t_{n,l+1}]$ such that $G(t_{n,l+1}) - G(t_{n,l}) = g(t)(t_{n,l+1} - t_{n,l})$, which is equivalent to $\frac{l+1}{T_n} - \frac{l}{T_n} = g(t)(t_{n,l+1} - t_{n,l})$. Since $\inf_{t \in \mathcal{T}} g(t) > 0$, when $T_n \to \infty$, $t_{n,l+1} - t_{n,l} = \frac{1}{g(t) \cdot T_n} \to 0$. Sampling points grow dense when $T_n \to \infty$ for $n = 1, \ldots, N$. That is, $\max_{l=1,\ldots,T_n-1} |t_{n,l+1} - t_{n,l}| \to 0$. Recall that $d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) := \sup_{t \in \mathcal{T}} \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right]$, and $d(\widetilde{\boldsymbol{Y}}_m, \widetilde{\boldsymbol{Y}}_n) := \max_{k=1,\ldots,T} \left[\sqrt{\sum_{l=1}^{p} \left\{ \widetilde{Y}_m^{(l)}(st_k) - \widetilde{Y}_n^{(l)}(st_k) \right\}^2} \right]$.

Then, the proof is split into two parts.

First, prove $d(\widetilde{\boldsymbol{Y}}_m, \widetilde{\boldsymbol{Y}}_n) \xrightarrow{a.s.} \max_{k=1,\ldots,T} \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k) \right\}^2} \right]$; see part A.

Second, prove $\left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k) \right\}^2} \right] \xrightarrow{a.s.} \max_{t \in [st_k, st_{k+1}]} \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right]$ for $k = 1, \ldots, T$; see part B.

Then, it is easy to show $d(\widetilde{\boldsymbol{Y}}_m, \widetilde{\boldsymbol{Y}}_n) \xrightarrow{a.s.} d(\boldsymbol{Y}_m, \boldsymbol{Y}_n)$ under Assumption 1 and $T_n \to \infty$ for $n = 1, \ldots, N$.

$d(\boldsymbol{Y}_m, \boldsymbol{Y}_n) = \max_{k=1,\ldots,T-1} \left(\max_{t \in [st_k, st_{k+1}]} \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right] \right)$ because $\mathcal{T}$ is a compact set.

A. Assumption 1 implies $|t_{n,j+1} - t_{n,j}| \to 0$ as $T_n \to \infty$ $\forall$ $j = 1, \ldots, T_n - 1$. For $n = 1, \ldots, N$, and for $k = 1, \ldots T$, there exists $j = 1, \ldots, T_n - 1$ so that $st_k \in [t_{n,j}, t_{n,j+1}]$. According to Definition 2, $\widetilde{t}_{n,k} = \arg\min_{l=j,j+1} f(t_{n,l}) := \{t_{n,l} : |t_{n,l} - st_k|\}$. Then, $|\widetilde{t}_{n,k} - st_k| \leq |t_{n,j+1} - t_{n,j}|$. According to the continuity of $\boldsymbol{Y}_n$, for $l = 1, \ldots, p$, $|\widetilde{t}_{n,k} - st_k| \to 0$ leads to $|\widetilde{Y}_n^{(l)}(st_k) - Y_n^{(l)}(st_k)| = |Y_n^{(l)}(\widetilde{t}_{n,k}) - Y_n^{(l)}(st_k)| \to 0$ as $T_n \to \infty$. Similarly, for $l = 1, \ldots, p$, $|\widetilde{t}_{m,k} - st_k| \to 0$ leads to $|\widetilde{Y}_m^{(l)}(st_k) - Y_m^{(l)}(st_k)| = |Y_m^{(l)}(\widetilde{t}_{m,k}) - Y_m^{(l)}(st_k)| \to 0$ as $T_m \to \infty$. Then, $\sum_{l=1}^{p} \{\widetilde{Y}_m^{(l)}(st_k) - \widetilde{Y}_n^{(l)}(st_k)\}^2 = \sum_{l=1}^{p} \{\widetilde{Y}_m^{(l)}(st_k) - Y_m^{(l)}(st_k) + Y_m^{(l)}(st_k) - \widetilde{Y}_n^{(l)}(st_k) + Y_n^{(l)}(st_k) - Y_n^{(l)}(st_k)\}^2 = \sum_{l=1}^{p} \left[\{\widetilde{Y}_m^{(l)}(st_k) - Y_m^{(l)}(st_k)\}^2 + \{\widetilde{Y}_n^{(l)}(st_k) - Y_n^{(l)}(st_k)\}^2 + \{Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k)\}^2 + 2\{\widetilde{Y}_m^{(l)}(st_k) - Y_m^{(l)}(st_k)\}\{\widetilde{Y}_n^{(l)}(st_k) - Y_n^{(l)}(st_k)\} + 2\{\widetilde{Y}_m^{(l)}(st_k) - Y_m^{(l)}(st_k)\}\{Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k)\} + 2\{\widetilde{Y}_n^{(l)}(st_k) - Y_n^{(l)}(st_k)\}\{Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k)\} \right] \xrightarrow{a.s.} \sum_{l=1}^{p} \{Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k)\}^2$.

B. $\sum_{l=1}^{p} \{Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k)\}^2 = \sum_{l=1}^{p} \{Y_m^{(l)}(st_k) - Y_m^{(l)}(t) + Y_m^{(l)}(t) - Y_n^{(l)}(t) + Y_n^{(l)}(t) - Y_n^{(l)}(st_k)\}^2 = \sum_{l=1}^{p} [\{Y_m^{(l)}(st_k) - Y_m^{(l)}(t)\} + \{Y_n^{(l)}(t) - Y_n^{(l)}(st_k)\} + \{Y_m^{(l)}(t) - Y_n^{(l)}(t)\}]^2$ for $t \in [st_k, st_k + 1]$. Since $\mathcal{T}$ is a compact set, $|t - st_k| \leq |st_{k+1} - st_k| := |\mathcal{T}|/T_n \to 0$ as $T_n \to \infty$ for $n = 1, \ldots, N$. Then, under the continuity of $\boldsymbol{Y}_m$ and $\boldsymbol{Y}_n$, $|Y_m^{(l)}(st_k) - Y_m^{(l)}(t)| \to 0$ and $|Y_n^{(l)}(t) - Y_n^{(l)}(st_k)| \to 0$. Next, $\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k) \right\}^2} \to \sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2}$ for $t \in [st_k, st_k + 1]$. Hence, $\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(st_k) - Y_n^{(l)}(st_k) \right\}^2} \to \max_{t \in [st_k, st_{k+1}]} \left[\sqrt{\sum_{l=1}^{p} \left\{ Y_m^{(l)}(t) - Y_n^{(l)}(t) \right\}^2} \right]$, which concludes the proof. □

References

Abraham, C., Cornillon, P.-A., Matzner-Løber, E., Molinari, N., 2003. Unsupervised curve clustering using B-splines. Scandinavian Journal of Statistics 30 (3), 581–595.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Al Abid, F.B., 2014. A novel approach for PAM clustering method. International Journal of Computer Applications 86 (17), 1–5.

Albert-Smet, J., Torrente, A., Romo, J., 2022. Band depth based initialization of k-means for functional data clustering. Advances in Data Analysis and Classification to appear.

American Meteorological Society. (2000). Glossary of meteorology: Cyclonic circulation. https://glossary.ametsoc.org/wiki/Cyclonic_circulation. Accessed: 27.08.2021.

Antoniadis, A., Brossat, X., Cugliari, J., Poggi, J.-M., 2013. Clustering functional data using wavelets. International Journal of Wavelets, Multiresolution and Information Processing 11 (01), 1350003.

AOKI, T., 1985. A climatological study of typhoon formation and typhoon visit to Japan. Meteorology and Geophysics 36, 61–118.

Balcan, M.-F., Liang, Y., Gupta, P., 2014. Robust hierarchical clustering. The Journal of Machine Learning Research 15 (1), 3831–3871.

Batool, F., Hennig, C., 2021. Clustering with the average silhouette width. Computational Statistics & Data Analysis 158, 107190.

Biernacki, C., Celeux, G., Govaert, G., 2000. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (7), 719–725.

Boullé, M., 2012. Functional data clustering via piecewise constant nonparametric density estimation. Pattern Recognition 45 (12), 4389–4401.

Bouveyron, C., Jacques, J., 2011. Model-based clustering of time series in group-specific functional subspaces. Advances in Data Analysis and Classification 5 (4), 281–300.

Centofanti, F., Lepore, A., Palumbo, B., 2021. Sparse and smooth functional data clustering. arXiv preprint arXiv:2103.15224.

Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., Xu, X., Pei, F., 2017. Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based K-medoids method. Landscape and Urban Planning 160, 48–60.

Chen, Y., Wang, J.Z., Krovetz, R., 2005. Clue: Cluster-based retrieval of images by unsupervised learning. IEEE Transactions on Image Processing 14 (8), 1187–1201.

Chiou, J.-M., Li, P.-L., 2007. Functional clustering and identifying substructures of longitudinal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (4), 679–699.

Cuesta-Albertos, J.A., Fraiman, R., 2007. Impartial trimmed K-means for functional data. Computational Statistics & Data Analysis 51 (10), 4864–4877.

Dai, W., Genton, M.G., 2018. Multivariate functional data visualization and outlier detection. Journal of Computational and Graphical Statistics 27 (4), 923–934.

Dai, W., Genton, M.G., 2019. Directional outlyingness for multivariate functional data. Computational Statistics & Data Analysis 131, 50–65.

Day, W.H.E., Edelsbrunner, H., 1984. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification 1 (1), 7–24.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. Density-based spatial clustering of applications with noise. In: Int. conf. knowledge discovery and data mining, vol. 240, p. 6.

Ferreira, L., Hitchcock, D.B., 2009. A comparison of hierarchical methods for clustering functional data. Communications in Statistics-Simulation and Computation 38 (9), 1925–1949.

Floriello, D., Vitelli, V., 2017. Sparse clustering of functional data. Journal of Multivariate Analysis 154, 1–18.

Gagolewski, M., Bartoszuk, M., Cena, A., 2016. Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. Information Sciences 363, 8–23.

García, M.L.L., García-Ródenas, R., Gómez, A.G., 2015. K-means algorithms for functional data. Neurocomputing 151, 231–245.

Genton, M.G., Johnson, C., Potter, K., Stenchikov, G., Sun, Y., 2014. Surface boxplots. Stat 3 (1), 1–11.

Giacofci, M., Lambert-Lacroix, S., Marot, G., Picard, F., 2013. Wavelet-based clustering for mixed-effects functional models in high dimension. Biometrics 69 (1), 31–40.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. Journal of the American Statistical Association 105 (491), 1167–1177.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1), 100–108.

Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications, vol. 200. Springer Science & Business Media.

Hsing, T., Eubank, R., 2015. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, vol. 997. John Wiley & Sons.

Hubert, L., Arabie, P., 1985. Comparing partitions. Journal of Classification 2 (1), 193–218.

Hubert, M., Rousseeuw, P.J., Segaert, P., 2015. Multivariate functional outlier detection. Statistical Methods & Applications 24 (2), 177–202.

Ieva, F., Paganoni, A.M., Pigoli, D., Vitelli, V., 2011. Multivariate functional clustering for the analysis of ECG curves morphology. In: Cladag 2011 (8th international meeting of the classification and data analysis group), pp. 1–4.

Ieva, F., Paganoni, A.M., Pigoli, D., Vitelli, V., 2013. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. Journal of the Royal Statistical Society: Series C (Applied Statistics) 62 (3), 401–418.

Jacques, J., Preda, C., 2013. Funclust: A curves clustering method using functional random variables density approximation. Neurocomputing 112, 164–171.

Jacques, J., Preda, C., 2014a. Functional data clustering: A survey. Advances in Data Analysis and Classification 8 (3), 231–255.

Jacques, J., Preda, C., 2014b. Model-based clustering for multivariate functional data. Computational Statistics & Data Analysis 71, 92–106.

James, G.M., Sugar, C.A., 2003. Clustering for sparsely sampled functional data. Journal of the American Statistical Association 98 (462), 397–408.

Jeong, M.-H., Cai, Y., Sullivan, C.J., Wang, S., 2016. Data depth based clustering analysis. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 1–10.

Kaufman, L., Rousseeuw, P.J., 1990. Partitioning around medoids (program pam). Finding Groups in Data: An Introduction to Cluster Analysis 344, 68–125.

Kelley, J.L., 1955. General Topology. D. Van Nostrand.

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J., 2010. The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. Bulletin of the American Meteorological Society 91 (3), 363–376.

Kneip, A., Li, X., MacGibbon, K.B., Ramsay, J.O., 2000. Curve registration by local regression. Canadian Journal of Statistics 28 (1), 19–29.

Li, Z., Liu, J., Chen, S., Tang, X., 2007. Noise robust spectral clustering. 2007 IEEE 11th International Conference on Computer Vision 1–8. doi:10.1109/ICCV.2007.4409061.

Liu, X., Zhu, X.-H., Qiu, P., Chen, W., 2012. A correlation-matrix-based hierarchical clustering method for functional connectivity analysis. Journal of Neuroscience Methods 211 (1), 94–102.

Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2014. Statistics of time warpings and phase variations. Electronic Journal of Statistics 8 (2), 1697–1702. doi:10.1214/14-EJS901.

Meng, Y., Liang, J., Cao, F., He, Y., 2018. A new distance with derivative information for functional K-means clustering algorithm. Information Sciences 463, 166–185.

Misumi, T., Matsui, H., Konishi, S., 2019. Multivariate functional clustering and its application to typhoon data. Behaviormetrika 46 (1), 163–175.

Park, J., Ahn, J., 2017. Clustering multivariate functional data with phase variation. Biometrics 73 (1), 324–333.

Peng, J., Müller, H.-G., 2008. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. The Annals of Applied Statistics 2 (3), 1056–1077.

Qu, Z., Genton, M.G., 2022. Sparse functional boxplots for multivariate curves. Journal of Computational and Graphical Statistics 31 (4), 976–989.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer Series in Statistics. Springer Science & Business Media.

Ronchetti, E., 2021. The main contributions of robust statistics to statistical science and a new challenge. Special Issue on the Centenary of Metron. Metron: International Journal of Statistics 79 (2), 127–135.

Rossini, A.J., Tierney, L., Li, N., 2007. Simple parallel statistical computing in R. Journal of Computational and Graphical Statistics 16 (2), 399–420.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.

Samé, A., Chamroukhi, F., Govaert, G., Aknin, P., 2011. Model-based clustering and segmentation of time series with changes in regime. Advances in Data Analysis and Classification 5 (4), 301–321.

Sangalli, L.M., Secchi, P., Vantini, S., Vitelli, V., 2010. K-mean alignment for curve clustering. Computational Statistics & Data Analysis 54 (5), 1219–1233.

Schmutz, A., Jacques, J., Bouveyron, C., Cheze, L., Martin, P., 2020. Clustering multivariate functional data in group-specific functional subspaces. Computational Statistics 35, 1101–1131.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Sheehy, A., Gasser, T., Molinari, L., Largo, R.H., 2000. Contribution of growth phases to adult size. Annals of Human Biology 27 (3), 281–298.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., Liu, J., 2021. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. EURASIP Journal on Wireless Communications and Networking 2021 (1), 1–16.

Storm Science Australia. (2008). Tropical cyclones. https://www.ausstormscience.com/tropical-cyclones/. Accessed: 27.08.2021.

Z. Qu, W. Dai and M.G. Genton

Struyf, A., Hubert, M., Rousseeuw, P., 1997. Clustering in an object-oriented environment. Journal of Statistical Software 1 (4), 1–30.

Tarpey, T., Kinateder, K.K.J., 2003. Clustering functional data. Journal of Classification 20 (1), 093–114.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B 63 (2), 411–423.

Times, N. Y. (2022). Nature, science, animals - land, weather, volcanoes and earthquakes. https://factsanddetails.com/japan/cat26/sub160/item856.html. Accessed: 03.01.2022.

Van Aalst, M.K., 2006. The impacts of climate change on the risk of natural disasters. Disasters 30 (1), 5–18.

Wang, J., Su, X., 2011. An improved K-means clustering algorithm. 2011 IEEE 3rd International Conference on Communication Software and Networks 44–46.

Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association 100 (470), 577–590.

Yao, Z., Dai, W., Genton, M.G., 2020. Trajectory functional boxplots. Stat 9 (1), e289.