Exploratory Functional Data Analysis

Zhuo Qu[1], Wenlin Dai[2], Carolina Euan[1], Ying Sun[1], and Marc G. Genton[1]

August 3, 2023

Abstract

With the advancement of technology, functional data are being recorded more frequently, whether they are over time or in different locations. Traditionally, functional data were assumed to be defined on common and finite coordinate grids. However, real-world functional data often exhibit irregular coordinate grids or multiple components. To adapt to the demands of practical applications, researchers have developed visualization tools, outlier detection techniques, and clustering/classification methods that can handle more general types of functional data. This paper offers a comprehensive overview of recent exploratory functional data analysis (EFDA). It begins by introducing fundamental statistical concepts, such as mean and covariance functions, as well as robust statistics like the median and quantiles in multivariate functional data. Then, the paper delves into the evolution of visualization methods, such as the rainbow plot, and various adaptations of functional boxplot. These modified versions of the functional boxplot are designed to accommodate the complexities of general functional data. In addition to visualization tools, the paper also reviews outlier detection technologies, which are commonly integrated with visualization methods to identify anomalous patterns within the data. Moreover, the paper explores the application of clustering and classification techniques tailored for functional data. In closing, the paper briefly discusses future directions for EFDA.

Keywords: Classification, Clustering, Data visualization, Exploratory data analysis, Functional data, Multivarite functional data, Outlier Detection

---

[1]Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: zhuo.qu@kaust.edu.sa, carolina.euancampos@kaust.edu.sa, ying.sun@kaust.edu.sa, marc.genton@kaust.edu.sa
This research was supported by the King Abdullah University of Science and Technology (KAUST).
[2]Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China. E-mail: wenlin.dai@ruc.edu.cn

1 Introduction

Exploratory data analysis (EDA, Tukey 1977, Martinez et al. 2017) serves as the primary step of data analysis since it intuitively explores the basic properties of the underlying dataset and provides diagnostics for statistical modeling. Tukey (1977) contrasts this with confirmatory data analysis (CDA, Tukey 1980), an area of data analysis that is mostly concerned with the techniques of statistical hypothesis testing, confidence intervals, estimation, etc. Overall, EDA can be categorized into data visualization and data mining. The data visualization tools include but not limited to the scatter plot, the histogram, the boxplot (Tukey 1977), and the quantile-quantile plot, whereas the data mining techniques include without limitation to the dimensionality reduction, data clustering/classification and smoothing.

When the data object turns from the univariate (multivariate) data to a real (multivariate) function of an index such as the time, the wavelength or the location index, we name the new data object as univariate (multivariate) functional data (Ramsay & Dalzell 1991, Ramsay & Silverman 2005). Common real-life examples of univariate functional data include raw cell cycle gene expression curves (Zhao et al. 2004), longitudinal study between relative diameter and relative height for trees (López-Pintado & Romo 2009), longitudinal height data for teenagers (López-Pintado & Romo 2009), petroleum level curves in an oil refinery (Ramsay et al. 2009), and daily temperature curves (Sun & Genton 2011); while common real-life examples of multivariate functional data comprise longitudinal hip and knee angle curves for children (Ramsay & Silverman 2005), daily temperature curves by the sensors located at different altitudes (Berrendero et al. 2011), coordinates of handwriting data (López-Pintado et al. 2014), hurricane trajectories (Yao et al. 2005, Harris et al. 2021), individual growth velocity curves of different body parts (Carroll et al. 2021), and the joint curves of stunted growth and prevalence of the low-birth weight for 77 countries (Qu & Genton 2022).

Mathematically, functional data are considered as a realization of a stochastic process taking values in a Hilbert space. Each subject from one realization of the above process is assumed to be independent and has a continuous sample path. Practically, we could never observe a function entirely on the whole domain but take records on some fixed or random discrete points, which are either the same or different between subjects. According to the sampling scheme, functional data can be classified as follows: 1) fully observed functions without noise at arbitrarily dense grid, see smoothed daily Canadian temperature curve; 2) densely sampled functions with noisy measurements (dense design), see original daily joint Canadian temperature and precipitation curves; and 3) sparsely sampled functions with noisy measurements (sparse design, Qu & Genton 2022), see univariate CD4 data and bivariate hurricane data. To demonstrate the application of exploratory analysis methods in diverse sampling schemes, we use the following three dataset representatives for illustration: 1) univariate sparse CD4 cell count data from the R package *refund*, see Figure 1 (a), 2) bivariate sparse hurricane trajectory data (downloaded online), see Figure 1 (b), and 3) bivariate dense Canadian daily temperature and precipitation curves from the R package *fda*, see Figure 1 (c).

On the one hand, functional data can be regarded as a natural extension of a vector from the finite dimension to the infinite one. On the other hand, with the development of the data collection techniques, functional observations present themselves more frequently. Hence, functional data analysis (FDA, Ramsay et al. 2009) include both an intrinsic and applied interest. The intrinsically infinite dimension of functional data poses challenges for the existing visualization tools as well as the exploratory analysis procedures for the data. During the past two decades, much effort has been made to find effective ways for exploratory

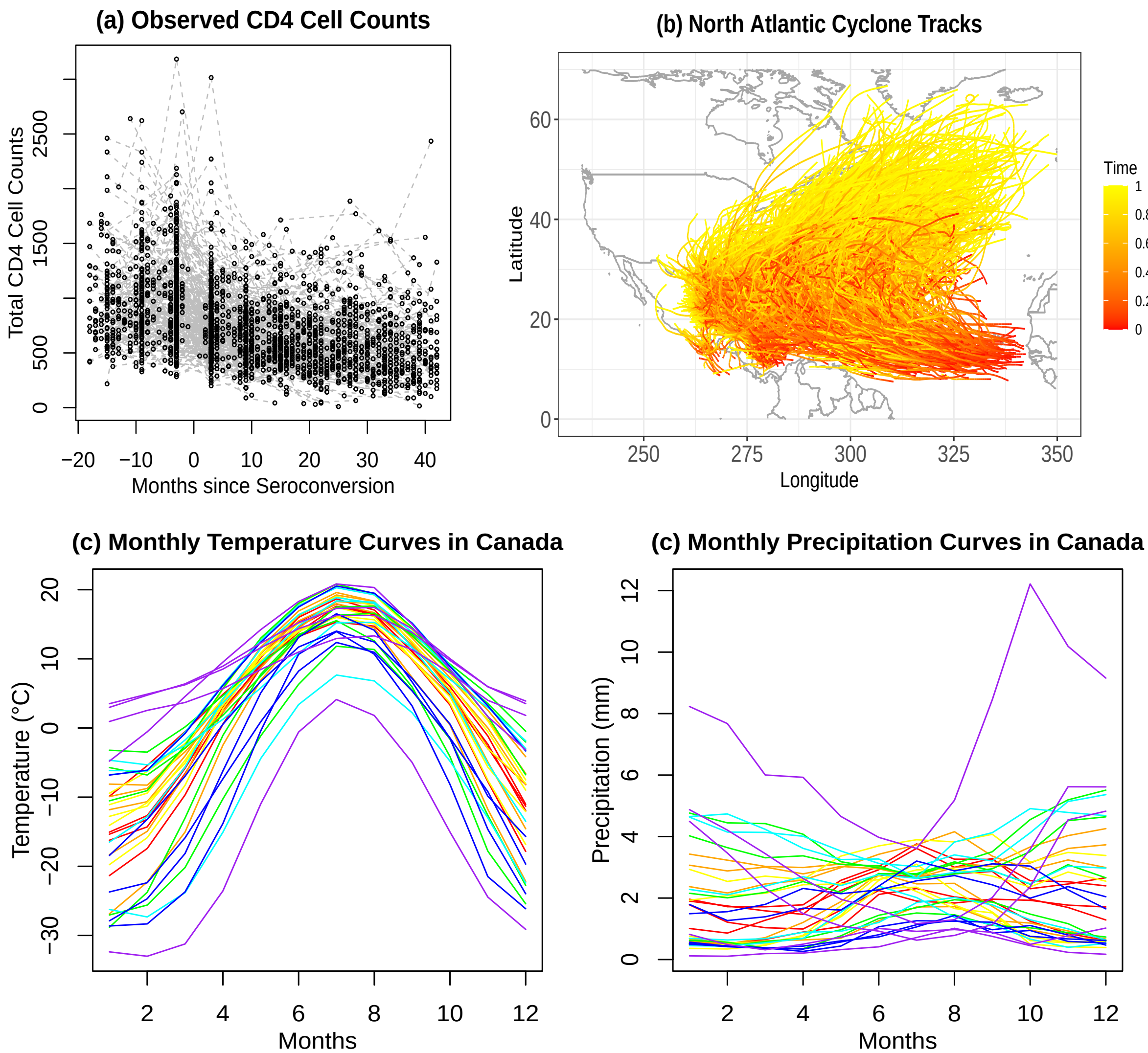

Figure 1: (a) shows observed CD4 counts for 366 subjects during months -18 and 42 since seroconversion, (b) shows 1873 North Atlantic cyclone tracks recorded from 1851 to 2021, and (c) shows the rainbow plots of monthly temperature and precipitation curves at 35 different locations in Canada averaged over 1960 to 1994. The orderings are based on the modified simplicial band depth (López-Pintado et al. 2014) of Canada temperature.

functional data analysis (EFDA), along with the estimations of mean and covariance functions (Yao et al. 2005, Wang et al. 2016, Happ & Greven 2018), and a series of methods and tools are developed, along with the blossom of statistical models and inference for functional data (Ramsay & Silverman 2005, Horváth & Kokoszka 2012, Wang et al. 2016).

This paper aims to review the novel data mining methodology and visualization tools specifically for functional data analysis, which serves as a prior step before diving into the modelling and statistical inference analysis. Compared to a case study about the geometric features in internal carotid artery in Sangalli et al. (2009), we introduce the exploratory analysis with novel visualization tools and methods in clustering and classification. In addition, we use the univariate sparse CD4 data, the bivariate irregular North Atlantic cyclone track data, and the bivariate dense Canadian weather data as instances to illustrate different methods. In contrast to the review of Wang et al. (2016) targeting the general analysis of univariate functional data, we cover the exploratory data analysis of the p-dimensional ($p \in \mathbb{Z}_+$) functional data where the measurement index per subject can vary. Hence, univariate functional data correspond to the special case of $p = 1$, and samples with dense time grids correspond to the special case of identical measurement indexes per subject.

The rest of the paper is organized as follows. Section 2 summarizes descriptive statistics for functional data. Subsequently, Section 3 proposes current tools for visualizing the observed functional data intuitively. Then, Section 4 displays visualization tools featuring the descriptive statistics of functional samples. In Section 5, we present several methods for functional data clustering and classification for dense and sparse functional data, separately. Section 6 concludes the paper with a summary and discussion.

2 Notations and Descriptive Statistics

In this section, we focus on the mathematical definitions of functional data and basic descriptive statistics of functional data with emphasis on the case that the dataset is contaminated by some abnormal subjects.

2.1 Notation

A functional random variable $\boldsymbol{Y}$ (Hsing & Eubank 2015) is a random vector with values in an infinite-dimensional space. Specifically, we view the p-variate ($p \in \mathbb{Z}_+$) functional data as paths of a p-variate stochastic process taking values in some Hilbert space $\mathcal{H}$ such as the space of square-integrable functions defined on some bounded and closed interval $\mathcal{T}$. That is, $\mathcal{H} := L^2(\mathcal{T}) \times \cdots \times L^2(\mathcal{T})$. When $p = 1$, we go back to univariate functional data.

Without loss of generality, we allow each marginal random vector in a p-variate stochastic process $\boldsymbol{Y}(\boldsymbol{t})$ to be defined at different indexes, that is, $\boldsymbol{Y}(\boldsymbol{t}) = (Y^{(1)}(t^{(1)}), \ldots, Y^{(p)}(t^{(p)}))^\top$ with $\boldsymbol{t}^\top := (t^{(1)}, \ldots, t^{(p)}) \in \mathcal{T} := \mathcal{T}_1 \times \cdots \times \mathcal{T}_p$. Note that $\boldsymbol{t}$ is a p-dimensional vector, with its element $t^{(j)}$ being a random time and independent of all other random variables. Each element $Y^{(j)}(t^{(j)})$ ($j = 1, \ldots, p$) is defined on the domain $\mathcal{T}_j$, where the $\mathcal{T}_j$s are compact sets in $\mathbb{R}$ with finite Lebesgue measure. Briefly speaking, $Y^{(j)}(t^{(j)})$: $\mathcal{T}_j \rightarrow \mathbb{R}$ is assumed to be square-integrable in $\mathcal{T}_j$, expressed as $L^2(\mathcal{T}_j)$. Then, we consider the p-dimensional functional data $\boldsymbol{Y} = \{\boldsymbol{Y}(\boldsymbol{t})\}_{\boldsymbol{t} \in \mathcal{T}}$ as sample paths of stochastic process $\boldsymbol{Y}(t)$ and we have $\boldsymbol{Y} \in \mathcal{H}$, where the space $\mathcal{H} := L^2(\mathcal{T}_1) \times \cdots \times L^2(\mathcal{T}_p)$.

In the following, let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ be a set of independent observations of $\boldsymbol{Y}$. In practice, we observe the functions $\boldsymbol{Y}_i(\boldsymbol{t})$ ($i = 1, \ldots, N$) with error $\boldsymbol{\epsilon}_i(\boldsymbol{t}) = (\epsilon_i^{(1)}(t^{(1)}), \ldots, \epsilon_i^{(p)}(t^{(p)}))^\top$, and the element $\epsilon_i^{(j)}(t^{(j)})$ are i.i.d. random variables with zero means. Moreover, the functions $\boldsymbol{Y}_i(\boldsymbol{t}_i)$ are observed on irregular finite grids at the subject and element level, that is, the jth ($j = 1, \ldots, p$) element $t_i^{(j)}$ of $\boldsymbol{t}_i$ ($\boldsymbol{t}_i \in \mathcal{T}$) can vary per curve. Let the observed functions with measurement errors and the sparseness be $\widetilde{\boldsymbol{Y}}_i(\boldsymbol{t}_i)$ such that $\widetilde{\boldsymbol{Y}}_i(\boldsymbol{t}_i) = \boldsymbol{Y}_i(\boldsymbol{t}_i) + \boldsymbol{\epsilon}_i(\boldsymbol{t}_i)$.

2.2 Moment-based Methods

We consider a collection of functional data, $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$, consisting of N independent subjects observed at finite time points. The sampling schedule could vary in both location and number for each subject. Based on $\boldsymbol{Y}_i(\boldsymbol{t}_i)$ $(i = 1, \ldots, N)$, and $\boldsymbol{t}_i = (t_i^{(1)}, \ldots, t_i^{(p)})^\top \in \mathcal{T}$ define $\boldsymbol{\mu}(\boldsymbol{t}) := \mathrm{E}\{\boldsymbol{Y}(\boldsymbol{t})\}$ as the mean function $\boldsymbol{\mu}$ evaluated at time $\boldsymbol{t}$, with the element estimation $\hat{\mu}^{(j)}(t^{(j)}) = \mathrm{E}\{Y^{(j)}(t^{(j)})\} = \frac{\sum_{i=1}^N Y_i^{(j)}(t_i^{(j)})\mathbf{1}(t_i^{(j)}=t^{(j)})}{\sum_{i=1}^N \mathbf{1}(t_i^{(j)}=t^{(j)})}$ for $l = 1, \ldots, L$ and $j = 1, \ldots, p$. When functional data all have common and finite grid points, the number of observations at each grid point equals to the number of subjects. However, when functional data are observed on irregular grids, the number of observations at each grid point is various and imbalanced. It may be practical to count the number of observations at each bin rather than grid point. To get the whole curve, one can simply apply smooth interpolation or nonparametric smoothing methods, e.g., kernel smoothing (Wand & Jones 1995), local polynomial smoothing (Fan & Gijbels 1996) or spline smoothing (Wang 2011).

For $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{T}$, define the matrix of covariances $\boldsymbol{C}(\boldsymbol{s}, \boldsymbol{t}) := \mathrm{cov}\{\boldsymbol{Y}(\boldsymbol{s}), \boldsymbol{Y}(\boldsymbol{t})\}$ with elements $C_{ij}(s^{(i)}, t^{(j)}) := \mathrm{cov}\{Y^{(i)}(s^{(i)}), Y^{(j)}(t^{(j)})\}$ for $s^{(i)} \in \mathcal{T}_i$ and $t^{(j)} \in \mathcal{T}_j$. Likewise, the pointwise covariance function can be estimated as

$$\hat{C}_{ij}(t_k^{(i)}, t_l^{(j)}) = \frac{\sum_{n=1}^N \{X_n(t_n^{(i)}) - \hat{\mu}^{(i)}(t_k^{(i)})\}\{X_n(t_n^{(j)}) - \hat{\mu}^{(j)}(t_l^{(j)})\mathbf{1}(t_n^{(i)} = t_k^{(i)}, t_n^{(j)} = t_l^{(i)})\}}{\sum_{n=1}^N \mathbf{1}(t_n^{(i)} = t_k^{(i)}, t_n^{(j)} = t_l^{(i)})},$$

and the whole surface of covariance function can be obtained by smoothing the three dimensional scatterplot. Common smoothing methods for sparse functional data include multivariate functional principal components (MFPCA, Happ & Greven 2018) and tensor-product splines (Cai & Yuan 2010, Xiao et al. 2018, Li et al. 2020). Practically, these records are often assumed to be contaminated by measurement errors, and we refer the readers to Wang et al. (2016) for a comprehensive review of the estimation of mean and covariance functions in such a scenario.

2.3 Robust Methods

Functional data can be contaminated by abnormal subjects, also known as outliers, similar with univariate or multivariate data. Outliers may severely bias the aforementioned moment-based estimators and, consequently, lead to incorrect inference results. Hence, it is desired to develop methods that could get rid of the influence from outliers and summarize functional data robustly.

For univariate data, order-statistics and ranks induced naturally by the order of scalars on the real line, are popularly used to design robust analysis methods. Whereas for functional data, such a natural ranking is not available. During the past two decades, the idea of data depth, initially proposed to sort multivariate data, is generalized to functional data. Specifically, a functional depth, taking values in $[0, 1]$, maps functional data as scalars and assigns larger depth values to central ones and smaller depth values to those outward. Consequently, these scalars provide a ranking criterion for functional data from the center outward in the data cloud.

Popularly implemented depth notions for dense univariate functional data include but not limited to band depth and modified band depth (BD and MBD, López-Pintado & Romo 2009), half-region depth and modified half-region depth (HRD and MHRD, López-Pintado & Romo 2011), extremal depth (Narisetty & Nair 2016), functional tangential angle pseudo-depth (FUNTA, Kuhnt & Rehage 2016) and its robustified version, order extended integrated depth (Nagy et al. 2017), total variation depth (TVD, Huang & Sun 2019), and elastic depths (Harris et al. 2021). For dense multivariate functional data, available depth notions include combinations of univariate functional depth measures (Ieva & Paganoni 2013), simplicial band depth and modified simplicial band depth (SBD and MSBD, López-Pintado et al. 2014), multivariate functional halfspace depth (MFHD, Claeskens et al. 2014), and multivariate FUNTA pseudo-depth and its robustified version (Kuhnt & Rehage 2016).

For sparse univariate functional data, Lòpez-Pintado & Wei (2011) first proposed a model-based consistent procedure for estimating the depths based on the estimated curves on regular grids. Then, Sguera & López-Pintado (2021) proposed a new depth that allows the curve estimation uncertainty to be incorporated into the depth analysis. Those two depth notions are extended to sparse multivariate functional data by Qu & Genton (2022); they also compared their ranking performances with simulations. In a recent study, Qu et al. (2022) introduced a novel framework for multivariate functional depths, specifically designed for sparse multivariate functional data, eliminating the need for curve estimation. This new depth concept, termed "global dept", distinguishes itself from previous approaches by handling sparse functional data directly. The authors demonstrate how the procedures for multivariate functional halfspace depth and multivariate extremal depth (an extension of extremal depth) can be adapted to their global depth framework.

Functional depths provide a natural basis for defining median, extremes and quantiles for functional data. Fraiman & Muniz (2001) defined the functional median as the deepest observation, i.e., the sample with the largest depth value, denoted as $\boldsymbol{M} = \arg\max_{\boldsymbol{X}} D(\boldsymbol{X}, F_{\boldsymbol{X}})$, where $D(\boldsymbol{X}, F_{\boldsymbol{X}})$ denotes the depth of a random function $\boldsymbol{X}$ with respect to its distribution $F_{\boldsymbol{X}}$. The functional version of the α trimmed mean, $\boldsymbol{\mu}_\alpha$, is defined as the average of the deepest $1 - \alpha$ proportion of subjects,

$$\boldsymbol{\mu}_\alpha = \frac{E(\boldsymbol{X}\mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}, F_{\boldsymbol{X}})))}{E(\mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}, F_{\boldsymbol{X}})))},$$

where $\mathbf{1}_A(x) = 1$ for $x \in A$ and zero otherwise, and $E(\mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}, F_{\boldsymbol{X}}))) = 1 - \alpha$. The empirical definitions of the two statistics can be expressed as

$$M_N = \arg \max_{i=1,\dots,N} D(\boldsymbol{X}_i, F_{\boldsymbol{X},N}), \quad \text{and} \quad \hat{\boldsymbol{\mu}}_{\alpha,N} = \frac{\sum_{i=1}^N \boldsymbol{X}_i \mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}_i, F_{\boldsymbol{X},N}))}{\sum_{i=1}^N \mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}_i, F_{\boldsymbol{X},N}))}.$$

Similarly, the α trimmed covariance function can be defined as

$$C_\alpha(s,t) = \frac{E[\{\boldsymbol{X}(s) - \boldsymbol{\mu}_\alpha(s)\}\{\boldsymbol{X}(t) - \boldsymbol{\mu}_\alpha(t)\}\mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}, F_{\boldsymbol{X}}))]}{E(\mathbf{1}_{[\beta,\infty)}(D(\boldsymbol{X}, F_{\boldsymbol{X}})))},$$

and its empirical version can be derived by substituting the statistics with their respective estimators.

Another concept related to the ranking of functional data is central region (López-Pintado & Romo 2009, Sun & Genton 2011, Narisetty & Nair 2016, Myllymäki et al. 2017), which is defined as

$$C_{1-\alpha} = \{\boldsymbol{X} \in L_2(\mathcal{T}) : \; X_L^{(j)}(t) \le X^{(j)}(t) \le X_U^{(j)}(t), \forall t \in \mathcal{T}, j = 1, \dots, p\},$$

where $\boldsymbol{X}_L$ and $\boldsymbol{X}_U$ are lower and upper α-envelop functions, $\boldsymbol{X}_L = \inf\{\boldsymbol{X} \in L_2(\mathcal{T}) : D(\boldsymbol{X}, F_{\boldsymbol{X}}) > \alpha\}$ and $\boldsymbol{X}_U = \sup\{\boldsymbol{X} \in L_2(\mathcal{T}) : D(\boldsymbol{X}, F_{\boldsymbol{X}}) > \alpha\}$, respectively.

3 Simple Visualization

Visualization (Friedman & Stuetzle 2002) has long been a component of great importance to exploratory data analysis, and many tools are so widely utilized as routine steps of analyzing procedure. For instance, the histogram of a univariate dataset shows a rough sense of the density of its underlying distribution, the scatter plot of a bivariate dataset presents the locations of the data points on a two-dimensional plane to provide some intuition on the relationship between the two variables, and the heatmap shows the magnitude of an object as color in two dimensions. Similar demands also exist in functional data analysis, that motivate the researchers to develop new graphical tools. Here, we select several reasonably simple tools that have proven useful in the literatures (Hyndman & Shang 2010, Hubert et al. 2015, Wrobel & Goldsmith 2016). We consider the Canadian weather dataset from Figure 1 (c) as an instance. This dataset includes monthly recorded Canadian temperature and precipitation curves in 35 stations over 1960 to 1994.

3.1 Spaghetti Plot and Rainbow Plot

Spaghetti plot (Allen 2019) is a simple visualization that colors each subject a distinct color, which makes it easy to track movement for data with small samples. However, it may look messy when we visualize big functional data. Rainbow plot, proposed by Hyndman & Shang (2010), can be regarded as an improvement of the spaghetti plot. As a visualization of all the curves, it adds a feature of the data ordering and colors the samples based on its ordering in the rainbow palette. Such an order can be time, data depth, data density, or other indexes.

As shown in Figure 1 (c), we order the Canadian temperature curve with the extremal depth (Narisetty & Nair 2016) first from the median to the extreme, and label the curves from red to purple in the rainbow palette. We can see that the red group represents the median tendency of the temperature and precipitation over a year, while the purple group includes some stations with high temperature and precipitation during winter and some stations with low temperature and precipitation all the year.

3.2 Heatmap

A heatmap (Hubert et al. 2015) represents different values using a system of color-coding. In FDA, a $n \times m$ heatmap is suitable to show a functional dataset consisting of n subjects recorded on m common design points. We visualize the data with a heatmap in Figure 2. For instance, each cell in Figure 2 (a) represents the estimated temperature of one station in a specific month, each row represents the monthly temperature curve of a station, and each column represents the average temperature of 35 stations in a particular month. Some abnormal information can be easily detected through the heatmap. On the one hand, Figure 2 (a) shows that Victoria and Vancouver have the lasting high temperature between April and October, whereas Resolute and Iqaluit are less than 10 degree Celsius almost all the year. On the other hand, Pr. Rupert has a monthly precipitation up to 6mm except between April and August, while the other stations have the monthly precipitation less than 6mm.

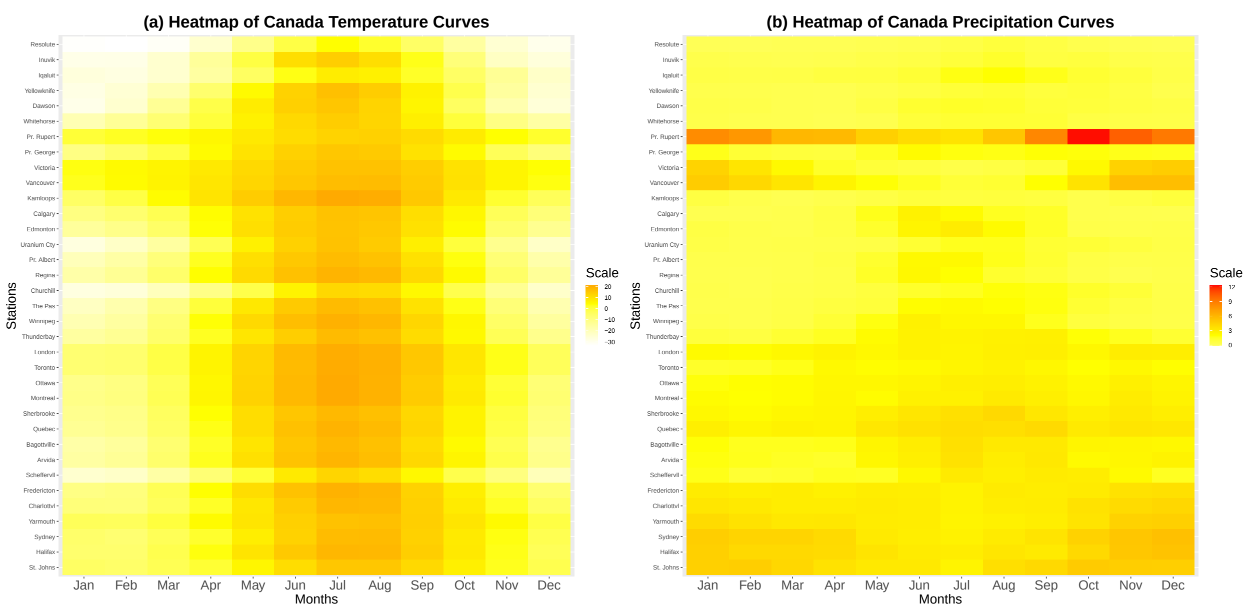

Figure 2: (a) is the heatmap of 35 Canada monthly average temperature curves over 1960-1994, and (b) is the heatmap of 35 Canada monthly average precipitation curves over 1960-1994.

3.3 Interactive Plots

Several packages have been developed to generate interactive visualization for functional data. An interactive plot retains the advantage of both visual and numerical illustration of data, i.e., it is intuitive as well as accurate. The interaction can be achieved in many ways, e.g., showing the associated records at the locations pointed by the cursor, zooming in or out, interacting between different plots. Wrobel et al. (2016) proposed the *refund.shiny* package that creates interactive graphics for functional data analysis. The *refund.shiny* package (Chang et al. 2015) relies on the *shiny* package to generate such an interactive user interface. Another commonly used tool is the *plotly* package (Sievert et al. 2018) that produces interactive plots with two or three dimensions in combination with a web portal.

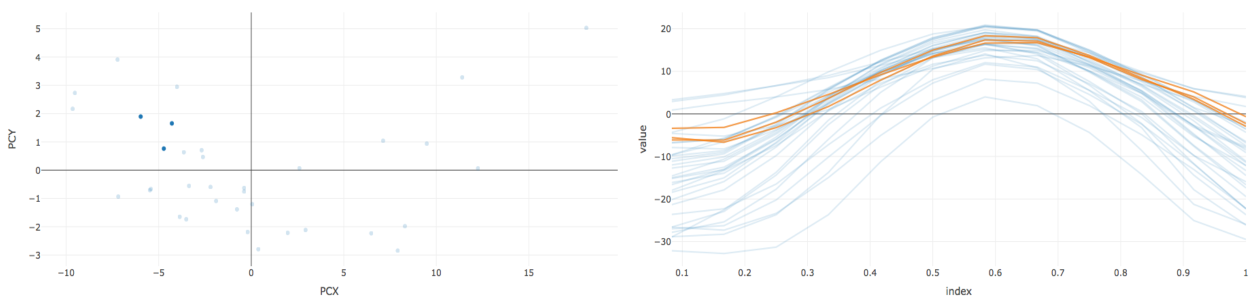

Figure 3: An illustration of the interactive functional principal component plot generated by the *refund.shiny* package. The left panel shows observed score scatterplot for selected FPCs of Canadian temperature, and the right panel shows fitted values for subjects, where subjects selected in the left panel are shown in orange, and the other subjects are shown in blue.

3.4 Animations

Animation or video is another powerful tool to enhance the still figures, that can visualize the dynamic evolution of the data. Genton et al. (2015) proposed the terminology *visuanimation* referring to visualization through animations, and explored the utility of animation in various perspectives of statistics. Castruccio et al. (2019) illustrate the predicted global temperature data, which can be regarded as functional data, varying spatially and temporally via a three-dimensional (3D) virtual-reality movie and developed a mobile application which allows users to watch the movie interactively.

4 Visualization with Structure Information

Many visualization tools for classical data have been developed featuring the descriptive statistics. For instance, the boxplot (Tukey 1977) of a univariate dataset illustrates the structure of this dataset by showing its descriptive statistics, e.g., median, quartiles, extreme values and possible outliers; and the bagplot (Rousseeuw et al. 1999) of a bivariate data presents the deepest data, the deepest 50% of the data points, and possible outliers under the ranks given by the halfspace depth (Tukey 1975). In functional data analysis, functional data can be transformed to notions of depth or outlyingness (Dai & Genton 2019) from the center outwards (see Subsection 2.3) for the visualization and outlier detection. Hence, we introduce visualization tools that contain structure information of raw data.

4.1 Visualization based on Ranking Information

Hyndman & Shang (2010) first proposed several visualization tools for smoothed functional data, such as functional bagplot and functional HDR boxplot available in R package *rainbow* (Shang & Hyndman 2019). Functional bagplot is based on the bivariate bagplot of Rousseeuw et al. (1999). Firstly, it applies the first two robust principal component scores to the bivariate bagplot as an auxiliary tool to rank the observations and detect outliers. Then, it displays the median curve, the 50% inner region and 99% fence. Curves that are partially outside these regions are identified as outliers. The functional highest density region (HDR) boxplot is a mapping of the bivariate HDR boxplot (Hyndman 1996) of the first two robust principal component scores to the functional curves. In contrast to the functional bagplot, this method displays curves with high HDR. Specifically, it focuses on curves whose first two principal component scores correspond to the 50% inner and 99% outer bivariate highest density regions (HDR). Additionally, it identifies outliers as points that are excluded from the 99% outer HDR. The visualization tools mentioned above use either time-ordering or the first two robust principal component scores to arrange the curves.

Functional boxplot, as proposed by Sun & Genton (2011), is a data visualization technique used to summarize the distribution and features of a set of functional data. It combines the functional depth and highlights the central quantiles and possible outliers. Analogous to the classical boxplot, there are four descriptive statistics in the functional boxplot (see Figure 4): envelope of the 50% central region, the median curve, outliers, and maximum non-outlying envelope. One observation is flagged as an outlier if its measurement at any grid point is outside a constant factor times the central region. The constant factor is set to be 1.5 under the assumption that observations at each index are independent and identically distributed and follow a normal distribution. The functional boxplot is generalized to other types of boxplots to suit functional data with additional characteristics. We categorize the variation of functional boxplots into four aspects: the one with more descriptive statistics, the one dealing with spatio-temporal data, the one with missing data, and the one with more general data objects.

In the first type, enhanced functional boxplot (Sun & Genton 2011), double-fence functional boxplot (Serfling & Wijesuriya 2017) and two-stage functional boxplot (Dai & Genton 2018a) were proposed to underline more features. For instance, the enhanced functional boxplot provided 25% and 75% central regions on the basis of functional boxplot, and two-stage functional boxplot. The double-fence functional boxplot includes an additional fence of 0.5 interquartile regions, enhancing its ability to identify specific shape and location outliers. The two-stage functional boxplot, implemented the directional outlyingness (Dai & Genton 2019) first and colored the detected outliers in green, and applied the remaining curves to the procedures in the functional boxplot.

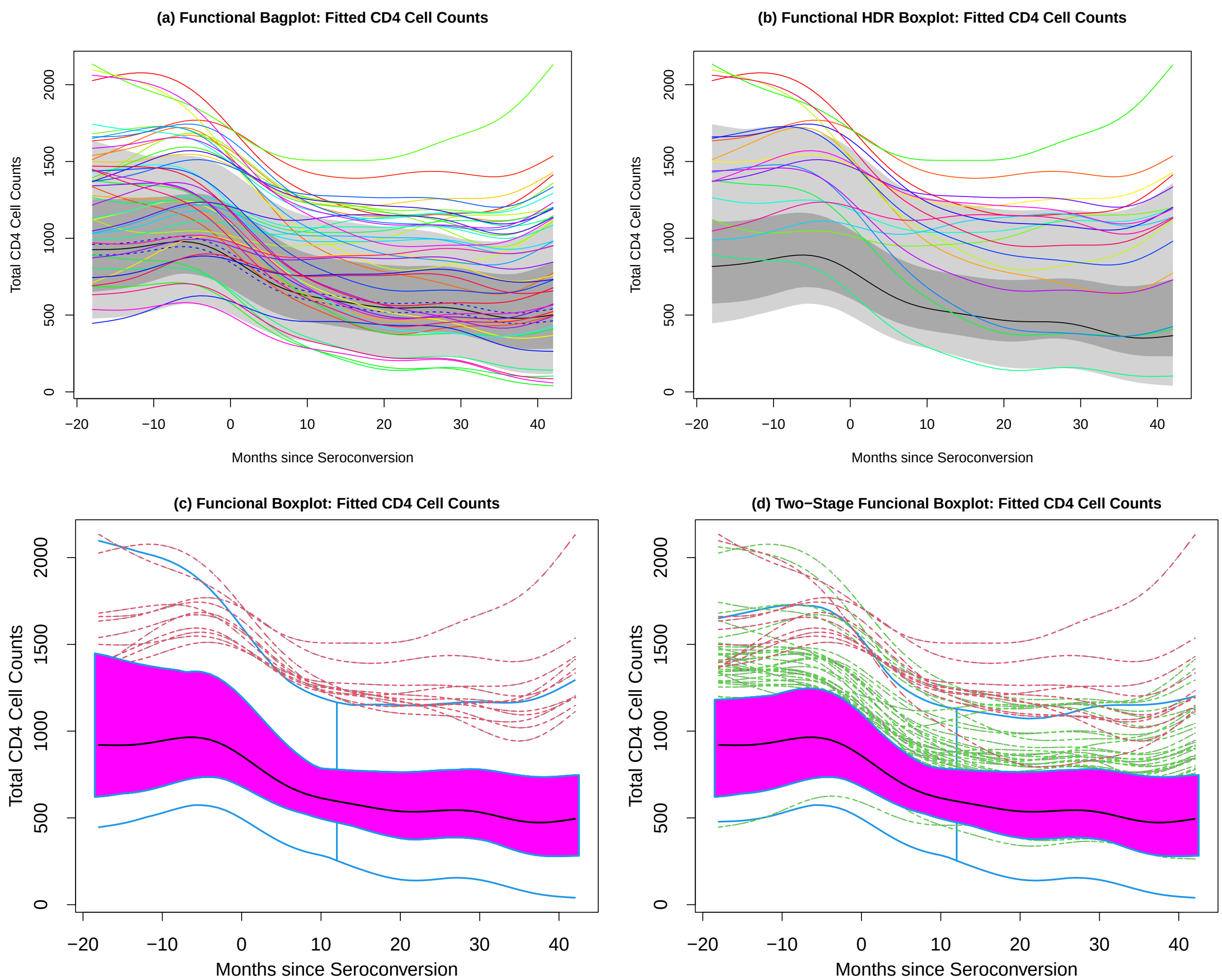

Figure 4: The comparisons between functional bagplot, functional HDR boxplot, and the functional boxplots of the fitted CD4 cell counts from bootstrap MFPCA (Qu & Genton 2022): (a) the functional bagplot, (b) the functional high density region (HDR) boxplot, (c) the functional boxplot, and (d) the two-stage functional boxplot.

In the second type, the adjusted functional boxplot (Sun & Genton 2012) and surface boxplot (Genton et al. 2014) were proposed for spatio-temporal data. The spatio-temporal data can be viewed as a temporal curve at each spatial location or a spatial surface at each time. In the former case, correlations need to be considered across locations. Hence, Sun & Genton (2012) proposed the adjusted functional boxplots, which flexibly selects the constant factor to control the probability of correctly detecting no outliers. In the mentioned work, Genton et al. (2014) extended the concept of modified band depth to modified volume depth specifically for image data. This extension allowed them to introduce a surface boxplot, which facilitates the visualization of image subjects based on the modified volume depth. Similarly, the same four descriptive statistics can also be established using the modified volume depth.

In the third type, sparse functional boxplot and intensity sparse functional boxplot (Qu & Genton 2022) were proposed for visualization. The data reconstruction is required with the multivariate functional principal component analysis (Happ & Greven 2018). In addition to the descriptive statistics in the functional boxplot, the sparse functional boxplot displays the smooth sparseness proportion within the 50% central region, and the intensity sparse functional boxplot displays the intensity of the smooth sparseness within the 50% central region. Usually, the directional outlyingness (Dai & Genton 2019) and sparse functional boxplots are combined to form the sparse two-stage functional boxplot and the intensity sparse two-stage functional boxplot for visualization and outlier detection; see Figure 5. Furthermore, sparse functional boxplots are extended to the simplified sparse functional boxplot (Qu et al. (2022)) and simplified intensity sparse functional boxplot without data reconstruction. The simplified visualization tools are established based on the global multivariate functional depths (Qu et al. 2022), which are applied to the sparse multivariate functional directly

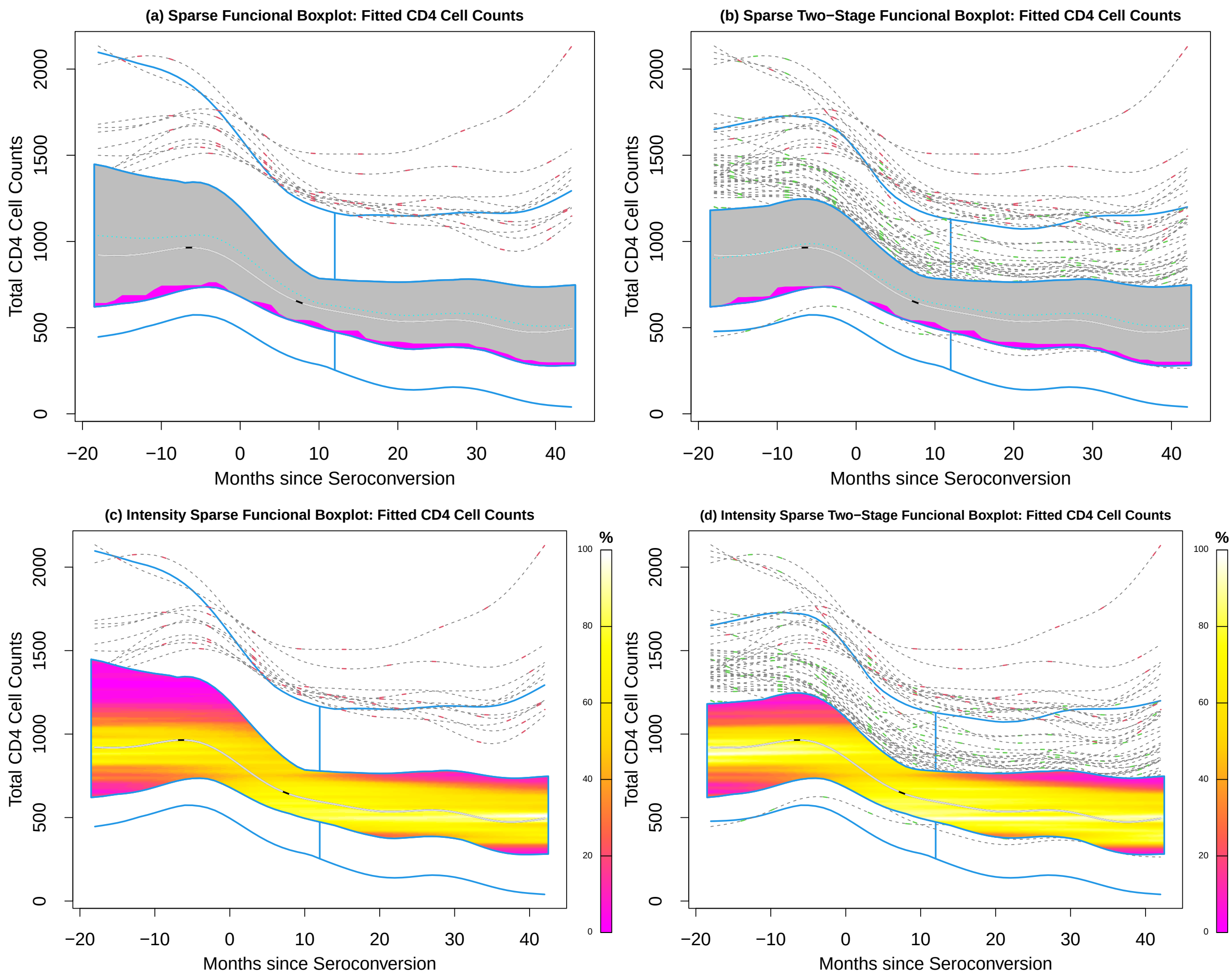

Figure 5: Functional boxplot and its variations when missing values exist, taking the instance of the fitted CD4 cell counts from bootstrap MFPCA (Qu & Genton 2022). The left column includes (a) the sparse functional boxplot, and (c) the intensity sparse functional boxplot. The right column includes (b) the sparse two-stage functional boxplot, and (d) the intensity sparse two-stage functional boxplot.

without data reconstruction.

In the fourth type, there are other natural extensions of the functional boxplot for the data expressed as sets, curves, paths, trajectories. Whitaker et al. (2013) defined the set band depth and introduced a contour boxplot for visualization and exploration of ensembles of contours or level sets of functions. Mirzargar et al. (2014) generalized the band depth for the curves and proposed curve boxplot. Hong et al. (2014) introduced a weighted functional boxplot when the data objects become shapes and images. Raj et al. (2017) proposed graph-simplex band depth and developed a visualization tool path boxplot. Yao et al. (2020) developed a trajectory boxplot (Figure 6) for visualizing and exploratory analysis for the trajectories which show the variation of the longitude and latitude through time.

However, there are shortcomings in the functional boxplot, such as losing the functional interpretation in the envelopes of the 50% central region and non-outlying region and not applicable to functional observations under hidden temporal warping variability. Therefore, Xie et al. (2017) decomposed observation variation in functional data into three main components: amplitude, phase, and vertical translation based on the curve registration (Srivastava et al. 2011). Then, they constructed different visualization for each element based on the median, two quartiles, and extreme observations. They also proposed identifying outliers based on those three components and visualizing the amplitude or phase outliers through the phase versus amplitude distance plot.

4.2 Centrality Decomposition Plots

Another set of visualizations, specifically for outlier detection, usually are based on ranking criteria such as statistical depths or outlyingness. The outlier detection visualization tools we introduce can be separated into the ones specifically for univariate functional data, and the ones for multivariate functional data (usually univariate functional data are special cases).

For univariate functional data, Arribas-Gil & Romo (2014) proposed outliergram to visualize and detect shape outliers in the functional data by exploiting the relationship between MBD and the modified epigraph index (MEI, López-Pintado & Romo 2011). Through a novel decomposition of the total variation depth, proposed by Huang & Sun (2019), we can easily detect shape outliers via the boxplot of the modified shape similarity (MSS). After that, the magnitude outliers can be seen among the remaining observations with the functional boxplots.

For multivariate functional data, Hubert et al. (2015) discussed amplitude and shape outliers and proposed various functional outlier maps based on the notion of outlyingness and depth in the multivariate functional data, e.g., adjusted outlyingness (AO) and skew-adjusted projection depth (SPD). Furthermore, they exploited the relationship between AO and SPD. They contructed the centrality-stability plot where the amplitude outliers lie in the upper right region and shape outliers lie in the right region. Rousseeuw et al. (2018) proposed a robust notion of outlyingness, directional outlyingness (DO), and can be applied in the univariate, multivariate setting. Based on the DO in the univariate setting, they also define the average of outlyingness as the functional directional outlyingness (FO) and measure the variability of its DO (VO). Then, they developed a graphical tool called the functional outlier map (FOM), which is a scatterplot of (FO, VO). Shift outliers, local outliers, and global outliers can be detected and displayed in different domains in FO. Based on the relationship between mean directional outlyingness (MO) and variation of directional outlyingness (VO). Dai & Genton (2018b) proposed a new graphical tool, the magnitude-shape (MS) plot to illustrate the centrality of curves comprehensively. They also generalized the outliergram to the bivariate outliergram for the outlier detection in bivariate functional data according to a quadratic relationship between functional directional outlyingness (FO) and MO. But the bivariate outliergram is limited to bivariate functional data and not as efficient as the MS plot in measuring the centrality of curves. The depth boxplot, introduced by Harris et al. (2021), is constructed on the elastic depths directly and serves as a half-boxplot. Its purpose is to identify potential amplitude and phase outliers. Yao et al. (2020) introduced wiggliness of directional outlyingness (WO) to detect outliers and constructed the WO-MSBD plot which can distinguish shape outliers and magnitude outliers. Ojo et al. (2023) proposed the magnitude-shape-amplitude plot (MSA-plot) based on fast massive unsupervised outlier detection (FastMUOD, Ojo et al. 2022).

5 Clustering and Classification

In the terminology of machine learning, functional data clustering is an unsupervised learning process, whereas functional data classification is a supervised learning procedure. Cluster analysis aims to group a set of data such that data objects within clusters are more similar than those across clusters with respect to a metric. In contrast, classification assigns a new data object to a predetermined group by a discriminant function or classifier. The assumption of functional clustering is the existence of inhomogeneous groups, and the goal is to estimate the membership of each group based on unlabeled observed data. Functional classification typically involves training data containing a functional predictor with an associated multi-class label for each data object. The discrimination procedure of functional classification is closely related to functional cluster analysis, even though the goals are different. When the structures or centers of clusters can be established in functional data clustering, the criteria used for identifying clusters can also be used for classification. Clustering and classification are both valuable tools for exploratory functional data analysis. The detection of meaningful and homogeneous clusters can improve the estimation of some descriptive statistics such as mean or median. Additionally, correctly assigning unlabeled functional data to predetermined classes enables better understanding and prediction of updated observations.

5.1 Functional data clustering

The range of applications for functional data clustering is vast, e.g., Abramowicz et al. (2017) applied a functional clustering method to study sediment data and to infer on past environmental and climate changes. Athanasiadis & Mrkvicka (2019) analyzed financial time series using functional clustering methods to identify different insurance penetration (IP) rate profile in the European markets. In general, the resulting clusters show high potential for data visualization and interpretation.

When dealing with functional data, similarities might take into account the characteristics of the curves such as shapes, magnitude, or derivatives (Hitchcock & Greenwood, 2015). Widely, we can classify the existing functional clustering methods as (Jacques & Preda, 2014): 1) Raw data methods, 2) Filtering methods, 3) Adaptive methods, and 4) Distance-based methods. Raw data method is a naive approach, and it might result in a high-dimensional vectorial clustering (Bouveyron & Brunet-Saumard, 2014). Filtering methods and adaptive methods use the basis expansion approach for functional data with a common basis for all data or common basis per group, respectively. Some examples of clustering methods that use B-splines, Fourier basis, or functional principal component analysis are detailed in Abraham et al. (2003); Serban & Wasserman (2005) and Shang (2014). Lastly, distance-based methods quantify the similarity between clusters by computing distances for functional objects. Here, we focus on distance-based methods. For a useful review on filtering and adaptive methods, we refer the reader to Jacques & Preda (2014) and Wang et al. (2016).

There are two main objects in a distance-based clustering method: similarity measure and clustering algorithm. We need to define a *similarity or dissimilarity measure* between curves which will be highly related to the interpretation of the clusters. Usually these measures are defined between two curves, $\{\boldsymbol{X}_i, \boldsymbol{X}_j\}$, where $\boldsymbol{X}_i = (X_i^{(1)}, \ldots, X_i^{(p)})^\top$ and $\boldsymbol{X}_j = (X_j^{(1)}, \ldots, X_j^{(p)})^\top$ are p-variate functional data, and we need a *clustering algorithm* to compute similarities across clusters, $C_1 = \{\boldsymbol{X}_1^1, \ldots, \boldsymbol{X}_{n_1}^1\}$ and $C_2 = \{\boldsymbol{X}_1^2, \ldots, \boldsymbol{X}_{n_2}^2\}$.

Usually, the similarity measure can be defined using a distance between functions, $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$. Natural choices for distance are the L_1, L_2 or L_∞ distances where

$$d_l(\boldsymbol{X}_i, \boldsymbol{X}_j) = \left(\frac{1}{p} \sum_{k=1}^{p} \int_{\mathcal{T}_k} (|X_i^{(k)}(t) - X_j^{(k)}(t)|)^l \mathrm{d}t \right)^{1/l}$$

for $l = 1, 2, \infty$. If we consider the L_1, L_2 or L_∞ distance, then the resulting clusters are built of functions with similar shape and magnitude. If there is no interest in similarity of magnitude, then the functions can be normalized and used the total variation (TV) distance (Alvarez-Esteban et al., 2016):

$$d_{TV}(\boldsymbol{X}_i, \boldsymbol{X}_j) = 1 - \frac{1}{p} \sum_{k=1}^{p} \int_{\mathcal{T}_k} \min\{X_i^{(k)}(t), X_j^{(k)}(t)\} \mathrm{d}t = \frac{1}{2p} \sum_{k=1}^{p} \int_{\mathcal{T}_k} |X_i^{(k)} - X_j^{(k)}| \mathrm{d}t.$$

These distances might be more complex if we also include information about the derivative curve and define a similarity measure as a weighted combination of them, $a_1 d(\boldsymbol{X}_i, \boldsymbol{X}_j) + a_2 d(\boldsymbol{X}'_i, \boldsymbol{X}'_j)$. Here, we assume that the curves $\boldsymbol{X}'_i s$ are all independent.

However, if the user is interested in clustering dependent curves, then a similarity measure can be proposed using the Spearman correlation or the rank correlation between functions (Heckman & Zamar, 2000). If these curves are linked to a time series trajectory, then a coherence based distance might be useful too (Euán et al., 2019). Under this setting, the resulting clusters are highly correlated within each group but low correlated across clusters.

Elastic time distance was proposed by Qu et al. (2023), which is applicable to (multivariate) functional data with either identical or different time measurements per subject.

The core idea is to build standard grid points and interpolate measurements at standard grid points with available observations. Assume curves $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are p-variate multivariate functional data, and $\widetilde{\boldsymbol{X}}_i$ and $\widetilde{\boldsymbol{X}}_j$ are their interpolated observations based on procedures in Qu et al. (2023), then

$$d_{ETD}(\boldsymbol{X}_i, \boldsymbol{X}_i) = \max_{k=1,\ldots,T} \sqrt{\sum_{l=1}^{p} \{\widetilde{X}_i^{(l)}(t_k) - \widetilde{X}_j^{(l)}(t_k)\}^2}, t_k \in \mathcal{T}.$$

Once the similarity measure is chosen, we use a clustering algorithm that selects the groups of functions that are more similar in an "optimally" manner, i.e., members within each group are highly similar, but members across groups are highly dissimilar. The most commonly used are the k-means and hierarchical clustering algorithms. Ferraty & Vieu (2006) introduced examples of hierarchical clustering using the L_2 distance between the functions and its second derivatives. Ieva et al. (2013) applied a k-means to identify clusters of electrocardiograph traces with a weighted distance between the curves and its first derivatives. Recently, Euán et al. (2018) proposed the hierarchical merger clustering algorithm. The main contribution to classical hierarchical algorithms is the use of a representative member for each cluster. Euán et al. (2018) proposed the use of the TV distance in a hierarchical merger algorithm to cluster spectral density functions from ocean waves time series. Euán & Sun (2019) extended this method for general 2D directional spectra functions. Moreover, Qu et al. (2023) combined the elastic time distance and the original robust two-layer partition (RTLP) clustering algorithm to cluster multivariate functional curves. They also compared RTLP clustering with other algorithms including the distance-based methods DBSCAN, k-means, k-median, and the model-based funHDDC algorithm (Schmutz et al. 2020).

Some real data applications might need more robust clustering algorithm, specially if

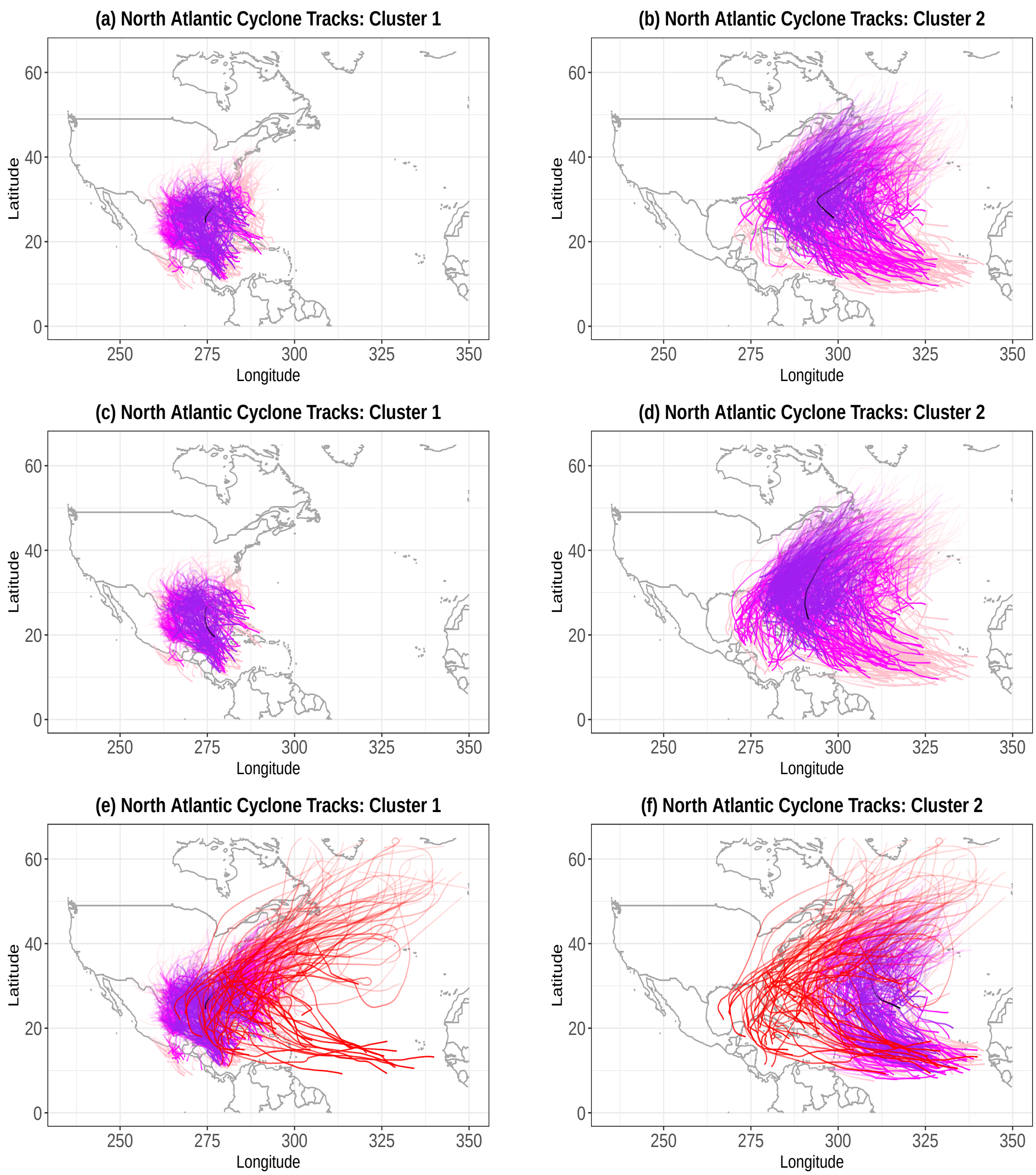

Figure 6: Cyclone trajectories obtained using our version of the trajectory boxplot (Yao et al. 2020). (a) and (b) are from k-medoids clustering, (c) and (d) are from hierarchical clustering, and (e) and (f) are from robust two-layer partition clustering. The black and red represent the median and outliers, respectively, and purple, magenta, and pink indicate the first, second, and third quartile curves, respectively.

the data has a high noise level. Although some of the previous methods described in this section might separate possible outliers as single clusters, this is not true for all methods. In the presence of potential outliers, Cuesta-Albertos & Fraiman (2007) proposed a trimmed k-means clustering that results in a robust cluster procedure for functional data. Also, Rivera-García et al. (2019) applied the trimming technique to introduce a robust model-based clustering method for functional data. When data is misaligned, the direct application of clustering methods might result in nonreasonable clustering structures. Sangalli et al. (2010) proposed an algorithm that considers the case when curves are misaligned. De Micheaux et al. (2021), based on the curve depth, employed the original clustering algorithm (Jörnsten 2004) with slight modifications for unparameterized curves. The robust two-layer partition clustering, introduced by Qu et al. (2023), utilizes both a two-layer partition algorithm and a modified silhouette index. This approach is effective for distinguishing clusters and identifying potential outliers in terms of their magnitude and shape.

In general, a good strategy is to select the clustering method based on the research goal. We illustrate this using the bivariate hurricane trajectory data in North Atlanta (see Figure 1 (b)). Since the trajectory data have various observations per subject, we apply the elastic time distance mentioned in Qu et al. (2023) and consider following different clustering methods based on the interpolated data, 1) K-medoids clustering (Park & Jun 2009, see Figure 6 (a)-(b)); 2) Hierarchical clustering with average as linkage function, see Figure 6 (c)-(d); and 3) Robust two-layer partition clustering, see Figure 6 (e)-(f). Each clustering method generates two clusters, yet outliers are only introduced by the robust two-layer partition clustering.

5.2 Functional data classification

While clustering seeks to find homogeneous clusters without knowledge of the true clusters, functional classification assigns a group membership to a new data object with a discriminant function or a classifier. To construct such a classifier we assume that the observed data are $\{(\boldsymbol{X}_1, G_1), \ldots, (\boldsymbol{X}_n, G_n)\}$ where $G_i \in \{1, 2, \ldots, k\}$ is a categorical variable that indicates the class membership of curve $\boldsymbol{X}_i$. Then, for a new observed curve $\boldsymbol{X}_0$, the goal of functional data classification is to assign the unknown class membership G.

Broadly, we can consider three different approaches to build the classifier (Wang et al., 2016): 1) Regression-based functional classification, 2) Functional discriminant analysis, and 3) Depth-based classification. Regression-based classifiers assume that the response of the regression model is categorical, then the assigned class membership corresponds to the value with the maximal probability. For instance, using a baseline functional logistic regression,

$$\log \frac{\mathbb{P}(G_0 = g|\boldsymbol{X}_0)}{1 - \sum_{j=1}^{k-1} \mathbb{P}(G_i = j|\boldsymbol{X}_i)} = \gamma_{0,g} + \int_{\mathcal{T}} \sum_{l=1}^{p} X_0^{(l)}(t)\gamma_{1,g}(t)\mathrm{d}t, \qquad g = 1, \ldots, k-1,$$

where $\gamma_{0,g}$ is an intercept term and $\gamma_{1,g}(t)$ is the coefficient function of the predictor. Then, the label G_0 is the one that maximizes $\mathbb{P}(G_0 = g|\boldsymbol{X}_0)$.

Functional discriminant analysis is another popular method. Similarly to the multivariate case, the discriminant analysis model (Galeano et al. 2015, Chamroukhi & Nguyen 2019) assumes a set of prior probabilities of belonging to each class, π_i, where $\sum_{i=1}^{k} \pi_i = 1$. Given the conditional density $f_j(\boldsymbol{X}) = \mathbb{P}(G = j|\boldsymbol{X})$, the probability of a new object belonging to class g is

$$\mathbb{P}(G_0 = g|\boldsymbol{X}_0) = \frac{\pi_g f_g(\boldsymbol{X}_0)}{\sum_{i=1}^{k} \pi_i f_i(\boldsymbol{X}_0)}.$$

One limitation of these methods is that a model is assumed. The regression type classification assumes a linear model, and the discriminant type classification assumes a distributional assumption. The inference for these models could be affected if there are outliers in the data. Depth-based classification methods (Hubert et al. 2017) are more useful under the suspected presence of an outlier.

There are different alternatives to perform nonparametric classification (Ferraty & Vieu, 2006). A common criterion is the maximum depth procedure (Sguera et al., 2014; Cuevas et al., 2007). Without loss of generality, assume two groups ($k = 2$) then $\{\boldsymbol{X}_1^1, \ldots, \boldsymbol{X}_{n_1}^1\}$ and $\{\boldsymbol{X}_1^2, \ldots, \boldsymbol{X}_{n_2}^2\}$ are the observed curves within each class. For a new observation $\boldsymbol{X}_0$, we assign class 1 if $P_1 > P_2$, where P_i is the depth of observation $\boldsymbol{X}_0$ in the sample $\{\boldsymbol{X}_1^i, \ldots, \boldsymbol{X}_{n_i}^i\}$. This rule can be generalized to the case with k ($k > 2$) groups by identifying the depth of new observation $\boldsymbol{X}_0$ in each group, and assign the group with the highest depth to $\boldsymbol{X}_0$. This method can be used with any functional depth.

For multivariate functional data, Sguera et al. (2014) applied a kernel functional spatial depth for supervised classification. Dai & Genton (2018c) investigated supervised classification using directional outlyingness and the outlyingness matrix. Blanquero et al. (2019) have proposed the support vector machine (SVM) able to optimally select the most informative time instants in order to get optimal classification rates. De Micheaux et al. (2021) proposed a curve depth and built a DD-plot to classify unparameterized curves. The cluster and outlier recognition algorithm, introduced by Qu et al. (2023), can also be available in functional classification. The core idea is to compare any functional distance between $\boldsymbol{X}_0$ and the center in each group, and determine whether to assign $\boldsymbol{X}_0$ to any group based on its rank in the set of distances between curves within the group and its group center. If the distance is larger than a threshold quantile of the set of distances between curves within the group and its group center, then $\boldsymbol{X}_0$ is assumed to be an outlier.

6 Discussion

EFDA has broadened its scope from solely analyzing fully observed univariate functional data to now encompassing irregular multivariate functional data. By utilizing functional depths and distances, EFDA offers a wide array of tools for visualizing, detecting outliers, clustering, and classifying both dense and sparse multivariate functional data.

Functional depths play a pivotal role in establishing functional rankings, forming the foundation for generating functional boxplots and identifying outliers. These functional depths can be categorized into four types, as outlined in Zuo & Serfling (2000). The first type gauges the average closeness of the curve to random samples, exemplified by band depth and simplicial band depth. The second type measures the distance of the curve from random samples, represented by the L^p depth. The third type assesses the outlyingness of a point concerning the center of random samples, such as the elastic depth. As for the fourth type, it is an index related to the relative depth concerning the center of the distribution, known as the extremal depth. Applying functional depths to sparse functional data is more complex due to irregular coordinate grids. To address this, one approach involves estimating curves and their confidence bands, or alternatively, applying global functional depth to sparse functional data directly as proposed by Qu et al. (2022). The original functional boxplot serves to identify central tendencies and outliers. However, it has limitations in detecting certain shape outliers and may not be directly applicable to functional data with missing values. To address these limitations, variations of functional boxplots and other visualization tools have been proposed, enabling the detection of shape outliers and facilitating the application of sparse functional data.

Moreover, functional distances play a crucial role in facilitating functional clustering. Examples of such distances include L^p distance, total variation distance, and elastic time distance. By leveraging functional distances, a wide range of classical clustering algorithms, as well as novel ones, can be applied to functional data. To handle common noise present in real-world data, robust two-layer partition clustering techniques can effectively separate potential outliers from the clusters. In the context of functional classification, both functional depths and distances find utility. When comparing the distance or depth of a curve to all other groups, the group with the smallest distribution proportion of distances (or the maximal depth) is assigned as the label for that curve.

Although this review has extended its domain from classical functional data to sparse multivariate functional data, a wider area of functional data can be considered, which may also pose new challenges in the visualization, robust statistics, and clustering/classification problems. The new-generation functional data can be the interval-valued functional data (Nasirzadeh et al. 2022, see simultaneous systolic and diastolic blood pressure of subjects at different visit times), the longitudinal functional data in the clinical trial (see the medical imaging data of patients at different time points during a clinical study in Adeli et al. 2019 and Zhu et al. 2021), the spatial functional data (Delicado et al. 2010, see the longitudinal climate data from arrays of monitors in the nearby area), and wearable health data (see Smets et al. 2018).

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.

Abramowicz, K., Arnqvist, P., Secchi, P., Luna, S. S. d., Vantini, S., & Vitelli, V. (2017). Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. *Stochastic Environmental Research and Risk Assessment*, 31(1), 71–85.

Adeli, E., Meng, Y., Li, G., Lin, W., & Shen, D. (2019). Multi-task prediction of infant cognitive scores from longitudinal incomplete neuroimaging data. *NeuroImage*, 185, 783–792.

Allen, T. T. (2019). *Introduction to engineering statistics and lean six sigma: statistical quality control and design of experiments and systems*, volume 3. Springer.

Alvarez-Esteban, P. C., Euán, C., & Ortega, J. (2016). Time series clustering using the total variation distance with applications in oceanography. *Environmetrics*, 27(6), 355–369. env.2398.

Arribas-Gil, A. & Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4), 603–619.

Athanasiadis, S. & Mrkvicka, T. (2019). European insurance market analysis: A multivariate clustering approach. In *Proceedings of the 12th International Scientific Conference INPROFORUM* (pp. 328–333).

Berrendero, J. R., Justel, A., & Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9), 2619–2634.

Blanquero, R., Carrizosa, E., Jiménez-Cordero, A., & Martín-Barragán, B. (2019). Variable selection in classification for multivariate functional data. *Information Sciences*, 481, 445–462.

Bouveyron, C. & Brunet-Saumard, C. (2014). Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis*, 71, 52 – 78.

Cai, T. & Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *Technical report, University of Pennsylvania and Georgia inistitute of technology.*

Carroll, C., Müller, H.-G., & Kneip, A. (2021). Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics*, 77(3), 839–851.

Castruccio, S., Genton, M. G., & Sun, Y. (2019). Visualizing spatiotemporal models with virtual reality: from fully immersive environments to applications in stereoscopic view. *Journal of the Royal Statistical Society: Series A*, 182(2), 379–387.

Chamroukhi, F. & Nguyen, H. D. (2019). Model-based clustering and classification of functional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), e1298.

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). Package 'shiny'. *https://cran.r-project.org/web/packages/shiny/shiny.pdf.*

Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505), 411–423.

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

Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3), 481–496.

Dai, W. & Genton, M. G. (2018a). Functional boxplots for multivariate curves. *Stat*, 7(1), e190.

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

Dai, W. & Genton, M. G. (2018c). An outlyingness matrix for multivariate functional data classification. *Statistica Sinica*, 28(4), 2435–2454.

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

De Micheaux, P. L., Mozharovskyi, P., & Vimond, M. (2021). Depth for curve data and applications. *Journal of the American Statistical Association*, 116(536), 1881–1897.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics: The official journal of the International Environmetrics Society*, 21(3-4), 224–239.

Euán, C., Ombao, H., & Ortega, J. (2018). The hierarchical spectral merger algorithm: A new time series clustering procedure. *Journal of Classification*, 35(1), 71 – 99.

Euán, C. & Sun, Y. (2019). Directional spectra-based clustering for visualizing patterns of ocean waves and winds. *Journal of Computational and Graphical Statistics*, 28(3), 659–670.

Euán, C., Sun, Y., & Ombao, H. (2019). Coherence-based time series clustering for statistical inference and visualization of brain connectivity. *Ann. Appl. Statist.*, 13(2), 990–1015.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

Fraiman, R. & Muniz, G. (2001). Trimmed means for functional data. *TEST*, 10, 419–440.

Friedman, J. H. & Stuetzle, W. (2002). John W. Tukey's work on interactive graphics. *The Annals of Statistics*, 30(6), 1629–1639.

Galeano, P., Joseph, E., & Lillo, R. E. (2015). The Mahalanobis distance for functional data with applications to classification. *Technometrics*, 57(2), 281–291.

Genton, M. G., Castruccio, S., Crippa, P., Dutta, S., Huser, R., Sun, Y., & Vettori, S. (2015). Visuanimation in statistics. *Stat*, 4(1), 81–96.

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

Happ, C. & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522), 649–659.

Harris, T., Tucker, J. D., Li, B., & Shand, L. (2021). Elastic depths for detecting shape anomalies in functional data. *Technometrics*, 63(4), 466–476.

Heckman, N. & Zamar, R. (2000). Comparing the shapes of regression functions. *Biometrika*, 87(1), 135–144.

Hitchcock, D. B. & Greenwood, M. C. (2015). Clustering functional data. In *Handbook of Cluster Analysis* chapter 13, (pp. 265–287). Chapman and Hall/CRC.

Hong, Y., Davis, B., Marron, J., Kwitt, R., Singh, N., Kimbell, J. S., Pitkin, E., Superfine, R., Davis, S. D., Zdanski, C. J., et al. (2014). Statistical atlas construction via weighted functional boxplots. *Medical image analysis*, 18(4), 684–698.

Horváth, L. & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

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

Huang, H. & Sun, Y. (2019). Visualization and assessment of spatio-temporal covariance properties. *Spatial Statistics*, 34, 100272.

Hubert, M., Rousseeuw, P., & Segaert, P. (2017). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification*, 11, 445–466.

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

Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, 50(2), 120–126.

Hyndman, R. J. & Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1), 29–45.

Ieva, F. & Paganoni, A. M. (2013). Depth measures for multivariate functional data. *Communications in Statistics-Theory and Methods*, 42(7), 1265–1276.

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. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3), 231–255.

Jörnsten, R. (2004). Clustering and classification based on the L1 data depth. *Journal of Multivariate Analysis*, 90(1), 67–89.

Kuhnt, S. & Rehage, A. (2016). An angle-based multivariate functional pseudo-depth for shape outlier detection. *Journal of Multivariate Analysis*, 146, 325–340.

Li, C., Xiao, L., & Luo, S. (2020). Fast covariance estimation for multivariate sparse functional data. *Stat*, 9(1), e245.

López-Pintado, S. & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486), 718–734.

López-Pintado, S. & Romo, J. (2011). A half-region depth for functional data. *Computational Statistics & Data Analysis*, 55(4), 1679–1695.

López-Pintado, S., Sun, Y., Lin, J. K., & Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification*, 8, 321–338.

Lòpez-Pintado, S. & Wei, Y. (2011). Depth for sparse functional data. In *Recent Advances in Functional Data Analysis and Related Topics* (pp. 209–212).: Springer.

Martinez, W. L., Martinez, A. R., & Solka, J. L. (2017). *Exploratory data analysis with MATLAB®*. Chapman and Hall/CRC.

Mirzargar, M., Whitaker, R. T., & Kirby, R. M. (2014). Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 2654–2663.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B*, 79(2), 381–404.

Nagy, S., Gijbels, I., & Hlubinka, D. (2017). Depth-based recognition of shape outlying functions. *Journal of Computational and Graphical Statistics*, 26(4), 883–893.

Narisetty, N. N. & Nair, V. N. (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association*, 111(516), 1705–1714.

Nasirzadeh, R., Nasirzadeh, F., & Mohammadi, Z. (2022). Some non-parametric regression models for interval-valued functional data. *Stat*, 11(1), e443.

Ojo, O. T., Fernández Anta, A., Genton, M. G., & Lillo, R. E. (2023). Multivariate functional outlier detection using the fast massive unsupervised outlier detection indices. *Stat*, 12(1), e567.

Ojo, O. T., Fernández Anta, A., Lillo, R. E., & Sguera, C. (2022). Detecting and classifying outliers in big functional data. *Advances in Data Analysis and Classification*, 16(3), 725–760.

Park, H.-S. & Jun, C.-H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2), 3336–3341.

Qu, Z., Dai, W., & Genton, M. G. (2022). Global depths for irregularly observed multivariate functional data. *arXiv preprint arXiv:2211.15125*.

Qu, Z., Dai, W., & Genton, M. G. (2023). Robust two-layer partition clustering of sparse multivariate functional data. *Econometrics and Statistics,* In press.

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

Raj, M., Mirzargar, M., Ricci, R., Kirby, R. M., & Whitaker, R. T. (2017). Path boxplots: a method for characterizing uncertainty in path ensembles on a graph. *Journal of Computational and Graphical Statistics*, 26(2), 243–252.

Ramsay, J., Hooker, G., & Graves, S. (2009). Introduction to functional data analysis. In *Functional data analysis with R and MATLAB* (pp. 1–19). Springer.

Ramsay, J. O. & Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), 539–561.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis* (second ed.). Springer.

Rivera-García, D., García-Escudero, L. A., Mayo-Iscar, A., & Ortega, J. (2019). Robust clustering for functional data based on trimming and constraints. *Advances in Data Analysis and Classification*, 13(1), 201–225.

Rousseeuw, P. J., Raymaekers, J., & Hubert, M. (2018). A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27(2), 345–359.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4), 382–387.

Sangalli, L. M., Secchi, P., Vantini, S., & Veneziani, A. (2009). A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. *Journal of the American Statistical Association*, 104(485), 37–48.

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(3), 1101–1131.

Serban, N. & Wasserman, L. (2005). Cats: Clustering after transformation and smoothing. *Journal of the American Statistical Association*, 100(471), 990–999.

Serfling, R. & Wijesuriya, U. (2017). Depth-based nonparametric description of functional data, with emphasis on use of spatial depth. *Computational Statistics & Data Analysis*, 105, 24–45.

Sguera, C., Galeano, P., & Lillo, R. (2014). Spatial depth-based classification for functional data. *TEST*, 23(4), 725–750.

Sguera, C. & López-Pintado, S. (2021). A notion of depth for sparse functional data. *TEST*, 30(3), 630–649.

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98, 121–142.

Shang, H. L. & Hyndman, R. J. (2019). Package 'rainbow'. *R packages*, 43.

Sievert, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2018). plotly for R.

Smets, E., Rios Velazquez, E., Schiavone, G., Chakroun, I., D'Hondt, E., De Raedt, W., Cornelis, J., Janssens, O., Van Hoecke, S., Claes, S., et al. (2018). Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *NPJ digital medicine*, 1(1), 67.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., & Marron, J. S. (2011). Registration of functional data using Fisher-Rao metric. *arXiv preprint arXiv:1103.3817*.

Sun, Y. & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2), 316–334.

Sun, Y. & Genton, M. G. (2012). Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, 23(1), 54–64.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2 (pp. 523–531).

Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Addison-Wesley. Reading, MA, USA.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(1), 23–25.

Wand, M. P. & Jones, M. C. (1995). *Kernel smoothing*. London: Chapman and Hall.

Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3, 257–295.

Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. New York: Chapman and Hall.

Whitaker, R. T., Mirzargar, M., & Kirby, R. M. (2013). Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2713–2722.

Wrobel, J. & Goldsmith, J. (2016). refund. shiny: Interactive plotting for functional data analyses. *R package version 0.3. 0, URL https://CRAN. R-project. org/package= refund. shiny*.

Wrobel, J., Park, S. Y., Staicu, A. M., & Goldsmith, J. (2016). Interactive graphics for functional data analyses. *Stat*, 5(1), 108–118.

Xiao, L., Li, C., Checkley, W., & Crainiceanu, C. (2018). Fast covariance estimation for sparse functional data. *Statistics and computing*, 28, 511–522.

Xie, W., Kurtek, S., Bharath, K., & Sun, Y. (2017). A geometric approach to visualization of variability in functional data. *Journal of the American Statistical Association*, 112(519), 979–993.

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6), 2873–2903.

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

Zhao, X., Marron, J., & Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica*, 14(2004), 789–808.

Zhu, Y., Kim, M., Zhu, X., Kaufer, D., Wu, G., Initiative, A. D. N., et al. (2021). Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data. *Medical image analysis*, 67, 101825.

Zuo, Y. & Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2), 461–482.