



Functional Data Visualization

By Marc G. Genton and Ying Sun

Keywords: functional boxplot, functional data, functional depth, functional outlyingness, image data, magnitude-shape plot, multivariate functional data, surface boxplot, surface data, trajectory functional boxplot, two-stage functional boxplot, visualization

Abstract: This article reviews tools to visualize functional data, that is, curves, surfaces/images, and trajectories. These tools are based on ranking functional data by means of notions of depth/outlyingness and make use of methods for functional outlier detections. For univariate functional data, the functional boxplot and surface boxplot are emphasized. For multivariate functional data, the magnitude–shape plot, the two-stage functional boxplot, and the trajectory functional boxplot are described. A bivariate functional dataset of the angles formed by the hip and knee of 39 children over their gait cycles is used throughout for illustration of the various visualization tools.

1 Introduction

Data in many fields are collected through a process naturally recorded as functional, such as curves, surfaces/images, and trajectories. Figure 1 shows one example of functional data introduced and discussed by Ramsay and Silverman^[1]. The data consist of the angles formed by the hip and knee of 39 children over their gait cycles. The cycle begins and ends when the heel under observation strikes the ground. Each curve is one functional observation measured from each child's gait cycle, and the time is translated into values in $[0, 1]$, proportional to the gait cycle. For such datasets, functional data analysis (FDA) provides a considerable number of methodologies and techniques and has become a popular branch of statistics. A broad overview of FDA and data examples can be found in Ramsay and Silverman^[1]. A more recent review of FDA and discussions can be found in Wang *et al.*^[2] and references therein.

Visualizing functional data can be a challenge. Genton *et al.*^[3] discussed and promoted dynamic visualization, coined visuanimation, including functional data visualization. Recently, Castruccio *et al.*^[4] presented a suite of apps for spatiotemporal data visualization using stereoscopic view and virtual reality on smartphones. In statistics, exploratory data analysis, as the first step of data analysis, requires informative visualization tools. Classical visualization tools designed for scalar data may not be appropriate for functional data. For example, the well-known boxplot is a graphical tool to summarize the distribution

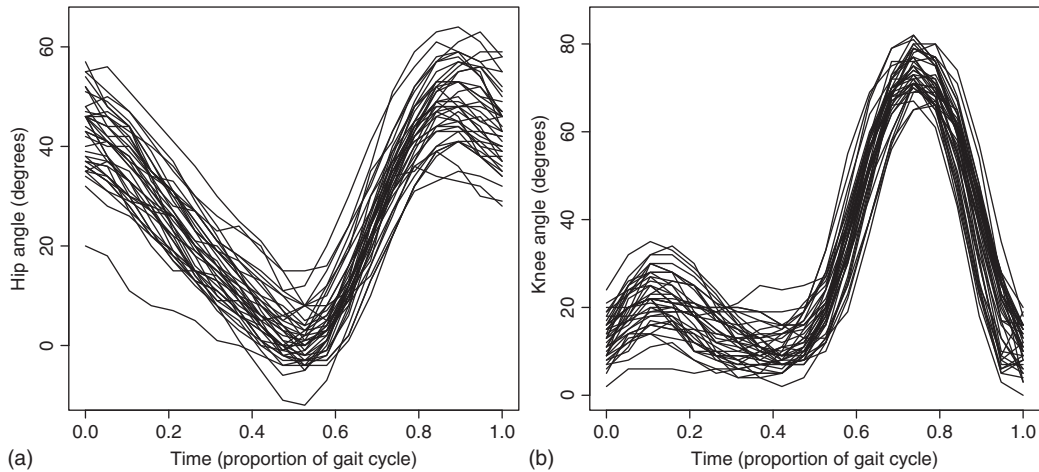


Figure 1. Functional data: the hip (a) and knee (b) angles of each of the 39 children as they go through their gait cycles. The time interval $[0, 1]$ represents a single cycle.

of a scalar random variable. Its construction is based on the ranks of the observations. However, for functional data, there is no unique ordering. Data depth is a key idea to generalize ranks to the functional setting. The depth value for each observation measures how central an observation is with respect to the entire sample. The modified band depth proposed by López-Pintado and Romo^[5] is one of the most popular depth notions for functional data due to its intuitive geometric interpretation. Other ways for ordering functional data exist, such as the tilting approach introduced by Genton and Hall^[6]. With a given ranking, Sun and Genton^[7] proposed the functional boxplot as a strong analog to the classical boxplot.

Functional data are often multivariate. For example, Figure 1 shows the hip and knee angles separately, and it does not indicate how the hip angle is paired with the knee angle from the same child. To be able to generalize the boxplot to the multivariate functional setting, multivariate depth is needed. Ieva and Paganoni^[8], Claeskens *et al.*^[9], López-Pintado *et al.*^[10], and Dai and Genton^[11] have proposed various methodologies for multivariate functional data ranking.

Rank-based methods are robust against outliers and can be used for outlier-detection purposes. The classical boxplot detects potential outliers by the 1.5 times IQR (interquartile range) empirical rule. For functional data, this empirical rule was generalized by Sun and Genton^[7]. However, outlier detection is much more complicated for functional data because the observations can be outlying either in magnitude or in shape. Huang and Sun^[12] proposed the total variation depth and its decomposition to visualize both magnitude and shape outliers. They compared the outlier-detection performance with several existing methods, including the functional boxplot with different depth notions for ranking, such as the modified band depth^[5] and the extremal depth^[13], the outliergram^[14], the functional outlier map (FOM) with the adjusted outlyingness proposed by Hubert *et al.*^[15], and the extended FOM with the “directional outlyingness” proposed by Rousseeuw *et al.*^[16]. Recently, Harris *et al.*^[17] proposed a notion of elastic depth to improve shape outlier detection.

In Section 2, we describe tools for univariate functional data visualization, whereas in Section 3 we focus on tools for multivariate functional data visualization. We end with some discussions in Section 4.

2 Univariate Functional Data Visualization

2.1 Functional Boxplots

The functional boxplot first proposed by Sun and Genton^[7] has been proven a valuable visualization tool for exploratory FDA. Inspired by the classical boxplot introduced as the “Box-and-Whisker plot” by Tukey^[18], the functional boxplot uses robust summary statistics to visualize the distribution of any given functional dataset. Instead of the original “Five Number Summary” in a classical boxplot, that is, the sample minimum, the first quartile, the median, the third quartile, and the sample maximum, the functional boxplot displays the functional median, the envelope of the 50% central region, and the maximum envelope.

To create a functional boxplot, the first step is to assign ranks to functional observations. Unlike for univariate scalar observations, the ranking for functional data is not unique. The functional boxplot by default uses the ranks induced by the modified band depth^[5], while the depth values calculated using other depth notions can be provided instead. Denote the order statistics by $y_{[1]}(t), \dots, y_{[n]}(t)$ according to decreasing depth values. Since the ordering is from the center outward, the first-order statistics, $y_{[1]}(t)$ or the functional observation with the largest depth value, has the most central position and thus called the *functional median*. The sample 50% central region is then defined as $C_{0.5} = \{(t, y(t)) : \min_{r=1, \dots, [n/2]} y_{[r]}(t) \leq y(t) \leq \max_{r=1, \dots, [n/2]} y_{[r]}(t)\}$, where $[n/2]$ is the smallest integer not less than $n/2$. The 50% central region indicates the variability of the central 50% of the data, and its border is called the *envelope* representing the box as in the classical boxplot. The functional median is displayed in the central region, which presents the centrality of the data. It is worth to point out that the functional median is one of the observations (or the average of the deepest curves if not unique). In contrast, the envelope of the 50% central region consists of pieces from different observations and is not an original observation anymore. The maximum envelope can be constructed in a similar way and is the analog to the whiskers in the classical boxplot.

In the functional boxplot, the maximum envelope is constructed after removing outliers by the 1.5 times the 50% central region empirical rule, where the fences are determined by inflating the envelope of the 50% central region by 1.5 times its range for each t . The constant factor F is set to be 1.5 by default as in the 1.5 times IQR rule in the classical boxplot. However, it can be modified by users. For instance, Sun and Genton^[19] proposed an adjusted functional boxplot for outlier detection in spatiotemporal data, where the factor F is adjusted according to the dependence among the functional observations.

To illustrate the construction of the functional plot, we consider the gait data, as shown in Figure 1. Figure 2 shows the functional boxplots of hip and knee angles for two choices of F for the illustration of outlier detection. Compared to the plots of original curves, the functional median curves for the hip and knee angles show more clearly the different phases of the data. For example, we see that at the beginning, the hip angle decreases from its maximum angle to its minimum near zero, while the knee angle first increases from zero to a local maximum near zero and then decreases to near zero. When the hip angle switches to a phase of increase, the knee angle increases sharply to its maximum as the leg prepares to leave the floor. During the last phase, the knee angle decreases rapidly to zero as the leg extends, and the hip angle levels off. The magenta area boxes the 50% central regions with the corresponding envelopes displayed in blue, which indicate the variability of these gait cycles among the 39 children. When $F = 1.5$, there were no outliers in both functional boxplots. For the hip angle data, one outlier is detected when decreasing F to 1.2, and for the knee angle data, when $F = 0.61$, one outlier is identified. These two outliers in red show slightly different features from the majority of the curves. Finally, the maximum envelopes in blue present the range of the nonoutlying observations. As pointed out by Sun and Genton^[19], the

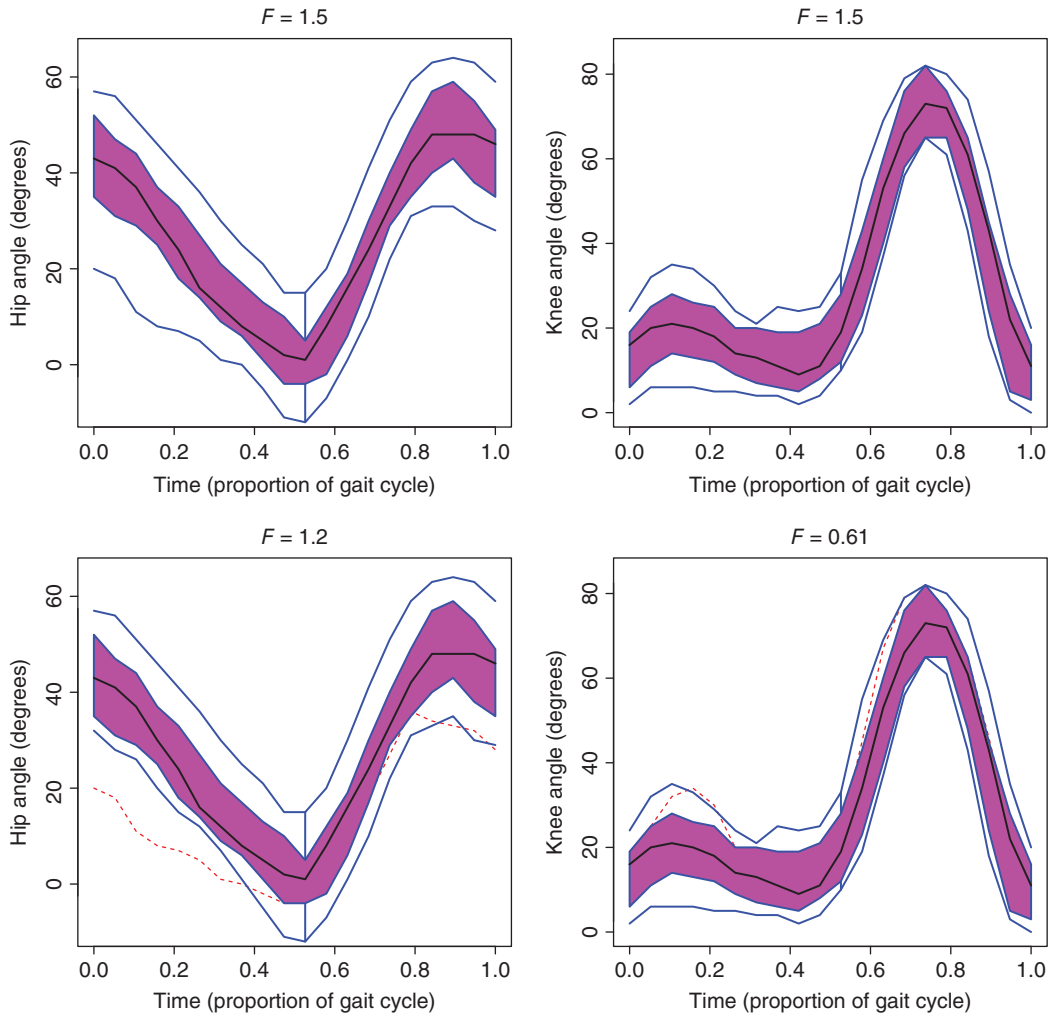


Figure 2. The functional boxplots for the hip and knee angles of each of the 39 children. The black curve in each functional boxplot is the median, the magenta box is the 50% central region, and the red dashed curves are the potential outliers identified by the F times the 50% central region rule.

adjustment of F is not so crucial for visualization purposes. However, it needs to be carefully selected for outlier detection. More discussion on the selection of the adjustment factor F and various examples can be found in Sun and Genton^[19].

Thanks to the fast algorithm developed in Sun *et al.*^[20], the functional boxplot has been widely used and applied to various types of functional datasets, from visualization, outlier detection to simulation, and model diagnostics. For instance, the functional boxplot was used to detect outliers in functional autoregressive models by Martínez-Hernández *et al.*^[21]; Huang and Sun^[22] proposed to assess spatiotemporal covariance properties using functional boxplots; Sun and Genton^[23] first used the functional boxplot to summarize simulation results of robust functional median polish for univariate ANOVA, and later it was extended to MANOVA by Qu *et al.*^[24]. It has also been employed to analyze and visualize the log periodograms of EEG time series data in the spectral domain by Ngo *et al.*^[25], and Euan and Sun^[26]

used functional boxplots to visualize clustering results for time series data and proposed the directional functional boxplot for visualizing the directional spectrum. Sun and Stein^[27] utilized the functional boxplot to assess the statistical properties of their spatiotemporal rainfall model, and La Vecchia and Ronchetti^[28] made use of the functional boxplot to examine the variability of the PP plot for comparing different methods when approximating the distribution of estimators and test statistics.

Other visualization tools exist for functional data, such as the functional bagplot and the functional highest density region boxplot^[29] based on the first two robust principal component scores, and boxplots for amplitude, phase, and vertical translation^[30] in the context of functional data registration.

2.2 Surface Boxplots

Sun and Genton^[7] presented an illustration of the surface boxplot as a potential extension of the functional boxplot when observations are surfaces or images. The surface boxplot can be constructed similar to the functional boxplot once the ranking of the sample surfaces is obtained. However, the 3D visualization is not trivial. A user-friendly visualization tool was developed by Genton *et al.*^[31] and illustrated on several real-world data examples. Yan *et al.*^[32] utilized this tool to visualize estimated covariance matrices from simulations by treating the matrices as images.

3 Multivariate Functional Data Visualization

In the gait cycle data example, the knee and hip angles of each child must be dependent and can be visualized jointly as multivariate functional data, that is, a random vector indexed by time. For multivariate functional data, the outlier detection is even more complicated because multivariate functional outliers are not necessarily marginal outliers, and the outlyingness could also occur in magnitude, shape, or both, as in the univariate case. There is no clear definition for each type of outliers. Dai *et al.*^[33] proposed a set of transformation operations on functional data for outlier-detection purposes, where potential outliers are either magnitude or shape outliers after suitable transformations. Next, we describe several tools for the visualization of multivariate functional data as well as their outlier detection.

3.1 Magnitude–Shape Plots

In order to rank multivariate functional data, Dai and Genton^[11] introduced the notion of directional outlyingness, a vector $\mathbf{O}(t)$ which measures at time t the centrality of functional data by assessing the level and the direction of their deviation from the central region together. The mean vector \mathbf{MO} and scalar variance VO of $\mathbf{O}(t)$ over time were defined to quantify the magnitude outlyingness and the shape outlyingness, respectively, of curves. Dai and Genton^[34] proposed a new graphical tool, the magnitude–shape (MS) plot of (\mathbf{MO}, VO) , for visualizing both the magnitude and shape outlyingness of multivariate functional data. They also provided a dividing elliptical curve (or ellipsoidal surface) to separate nonoutlying data from the outliers. They demonstrated through Monte Carlo simulations and data applications that the MS plot is superior to the existing tools for visualizing centrality and detecting outliers for functional data, such as the FOM of Hubert *et al.*^[15] and Rousseeuw *et al.*^[16], both of which fail to detect shape outliers. Dai and Genton^[34] also proposed the MS-plot array which displays marginal MS plots on the diagonal and pairwise bivariate MS plots on the off-diagonals.

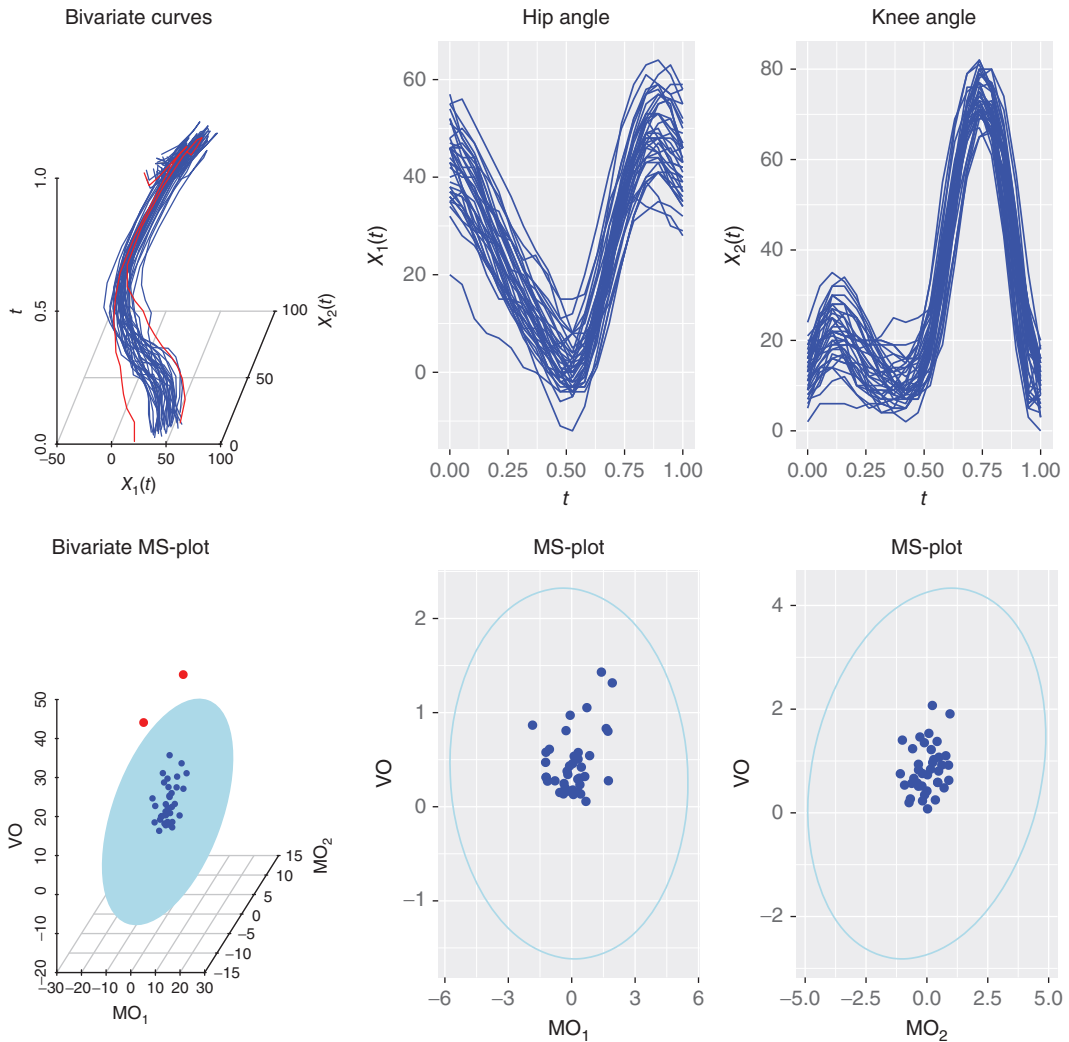


Figure 3. The bivariate and marginal MS plots for the hip and knee angles of each of the 39 children. The light-blue ellipsoid and ellipses represent the thresholds to flag potential outliers. The red curves/points are the potential outliers identified. They are joint outliers but not marginal outliers.

Figure 3 depicts the bivariate and marginal data of hip and knee angles of each of the 39 children, as well as the associated bivariate and marginal MS plots. The light-blue ellipsoid and ellipses represent the thresholds to flag potential outliers. The red curves/points are the potential outliers identified. They are joint outliers but not marginal outliers. Hence, it is important to apply joint rather than marginal methods in the multivariate case.

3.2 Two-Stage Functional Boxplots

Dai and Genton^[35] proposed a two-stage functional boxplot for the visualization and exploratory FDA of multivariate curves. They showed that the shape of the original functional boxplot is sensitive to shape

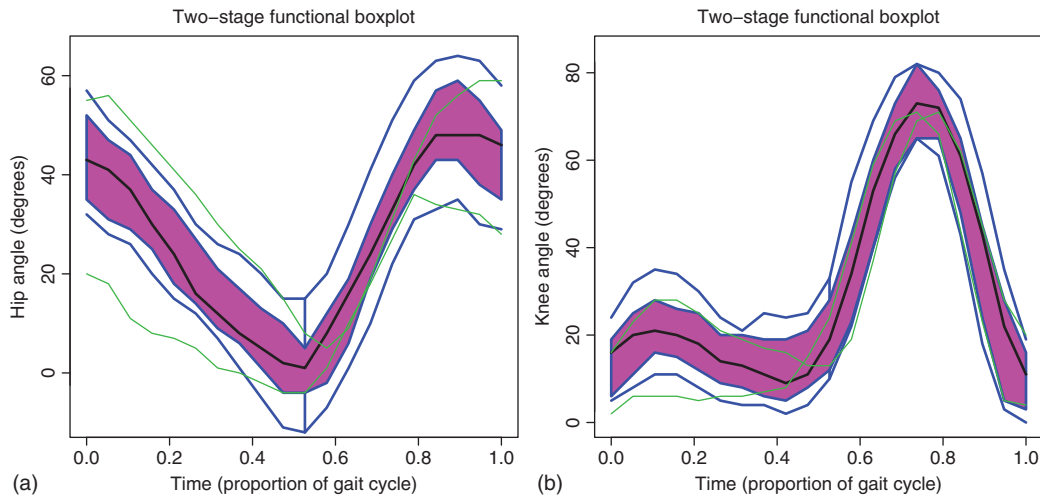


Figure 4. The two-stage functional boxplots for the hip (a) and knee (b) angles of each of the 39 children. The green curves are the potential bivariate shape outliers identified.

outliers and proposed to combine it with an outlier-detection procedure based on the aforementioned functional directional outlyingness in order to account for both the magnitude and shape outlyingness of functional data. In a first stage, shape outliers are identified, and in a second stage, the traditional functional boxplot is constructed on the remaining non-shape-outlying curves. This combination is robust to various types of outliers and allows to more precisely capture the data structures than does the functional boxplot alone. Moreover, it can provide both marginal and joint analyses of the multivariate curves.

We apply the two-stage functional boxplot to the hip and knee angles of each of the 39 children in Figure 4. The green curves are the potential shape outliers identified in the first stage. The shapes of the traditional functional boxplots in the second stage become slightly different from the shapes seen in the first row of Figure 2, in particular for the 50% central regions in magenta and the whiskers.

3.3 Trajectory Functional Boxplots

An important subset of multivariate functional data consists of the so-called trajectory functional data, which are viewed without a dimension/axis for time. Examples of trajectory functional data include the paths of hurricanes or migrating birds in the 2D plane. One could also view the gait data in the hip/knee 2D plane, see Figure 5a. Note that such data could be open or closed curves.

Mirzargar *et al.*^[36] introduced the curve boxplot as a generalization of the boxplot to an ensemble of curves. Yao *et al.*^[37] proposed two exploratory tools for trajectory functional data: the trajectory functional boxplot and the modified simplicial band depth (MSBD) of López-Pintado *et al.*^[10] versus wiggleness of directional outlyingness (WO) plot. They allow to visualize the centrality of trajectory functional data. Yao *et al.*^[37] defined the WO index, a measure of the roughness of curves, to effectively measure the shape variation of curves and to detect shape outliers. In addition, MSBD provides a center-outward ranking and identifies magnitude outliers. Using the measures of MSBD and WO, the functional boxplot of the

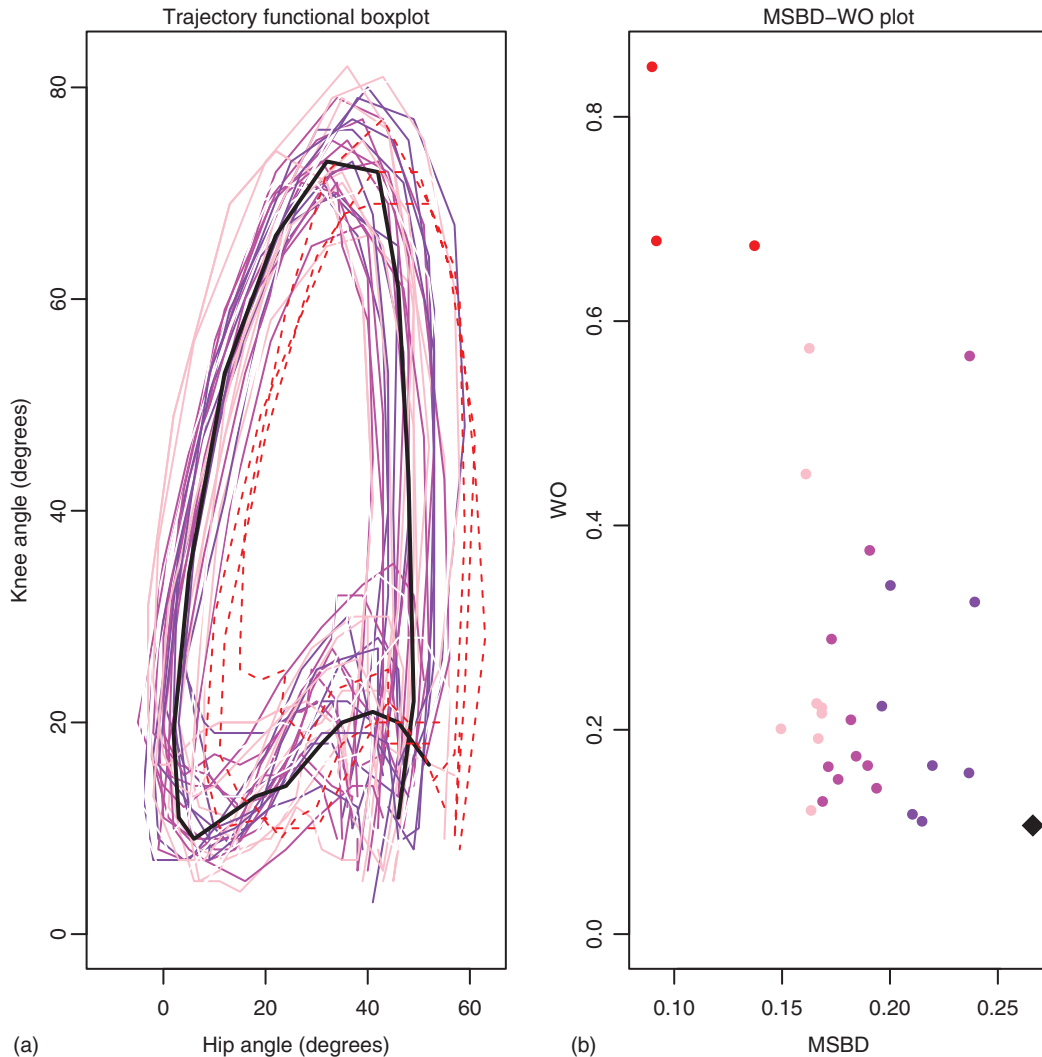


Figure 5. The trajectory functional boxplot (a) and the MSBD–WO plot (b) for the bivariate trajectories of hip and knee angles of each of the 39 children. The black curve/point represents the median. The purple/magenta/pink curves/points represent the 25%/50%/75% central regions. The three red dashed curves/points are the potential outliers identified.

trajectories reveals center-outward patterns and potential outliers using the raw curves. Moreover, the MSBD–WO plot illustrates such patterns and outliers in a space spanned by MSBD and WO.

Figure 5 depicts the trajectory functional boxplot and the MSBD–WO plot for the 2D trajectories of hip and knee angles of each of the 39 children. The black curve/point represents the median. The purple/magenta/pink curves/points represent the 25%/50%/75% central regions. The three red dashed curves/points are the potential outliers identified. Note that the outliers detected with WO are not necessarily the same as those detected with directional outlyingness. Yao *et al.*^[37] showed that WO is a more effective measure for detecting outliers in trajectory functional data.

4 Conclusions

The visualization of functional data is essential for many applications. This article focused on visualization tools and outlier-detection techniques based on functional data ranking, from univariate to multivariate cases. The reviewed methods have been illustrated on a dataset consisting of the angles formed by the hip and knee recorded from 39 children over their gait cycles.

These rank-based methods are robust and do not require any model or distribution assumptions. However, the computational cost might limit their capability in practice. The functional boxplot is implemented by the user-friendly `fbplot` command available in the *fda* packages^[38] in R^[39] and in MATLAB^[40]. It is straightforward to extract important statistics from the `fbplot` command for future analysis and visualization purposes. Although it is possible to use depth values from other notions to rank functional data, the functional boxplot by default uses the fast algorithm of Sun *et al.*^[20] for calculating the modified band depth values, which makes many applications feasible.

We have only illustrated univariate and bivariate curves, as well as trajectories, in this article. With modern data collection instruments and various applications from different research fields, the observations can be even more complex functional objects. To understand the features of complicated data from all the aspects, visualization is particularly useful for exploratory FDA, yet challenging. There are many open problems related to robust statistics and computations that require further research. Moreover, we have only focused on discussing the visualization of functional observations. Many core methods in FDA have common graphics to visualize the results from standard analyses, such as scree plots for functional principal component analysis and coefficient function plots in regression models. Wrobel *et al.*^[41] illustrated how to create interactive graphics for functional data analyses with the *refund.shiny* package^[42] in R.

Acknowledgment

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

Related Articles

Boxplot; Exploratory Data Analysis; Functional data analysis – An Introduction; Functional Data Analysis – Theory; Multivariate Data Visualisation.

References

- [1] Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis*, Springer Series in Statistics, 2nd edn, Springer, New York.
- [2] Wang, J.L., Chiou, J.M., and Müller, H.G. (2016) Functional data analysis. *Ann. Rev. Stat. Appl.*, **3** (1), 257–295.
- [3] Genton, M.G., Castruccio, S., Crippa, P., *et al.* (2015) Visuanimation in statistics. *Stat.*, **4**, 81–96.
- [4] Castruccio, S., Genton, M.G., and Sun, Y. (2019) Visualizing spatiotemporal models with virtual reality: from fully immersive environments to applications in stereoscopic view. *J. R. Stat. Soc. A*, **182** (2), 379–387.
- [5] López-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *J. Am. Stat. Assoc.*, **104** (486), 718–734.
- [6] Genton, M.G. and Hall, P. (2016) A tilting approach to ranking influence. *J. R. Stat. Soc. B. Stat. Methodol.*, **78** (1), 77–97.
- [7] Sun, Y. and Genton, M.G. (2011) Functional boxplots. *J. Comput. Graph. Stat.*, **20** (2), 316–334.
- [8] Ieva, F. and Paganoni, A.M. (2013) Depth measures for multivariate functional data. *Commun. Stat. Theory Methods*, **42** (7), 1265–1276.

- [9] Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014) Multivariate functional halfspace depth. *J. Am. Stat. Assoc.*, **109** (505), 411–423.
- [10] López-Pintado, S., Sun, Y., Lin, J.K., and Genton, M.G. (2014) Simplicial band depth for multivariate functional data. *Adv. Data Anal. Classif.*, **8** (3), 321–338.
- [11] Dai, W. and Genton, M.G. (2019) Directional outlyingness for multivariate functional data. *Comput. Stat. Data Anal.*, **131**, 50–65.
- [12] Huang, H. and Sun, Y. (2019) A decomposition of total variation depth for understanding functional outliers. *Technometrics*, **61** (4), 445–458.
- [13] Narisetty, N.N. and Nair, V.N. (2016) Extremal depth for functional data and applications. *J. Am. Stat. Assoc.*, **111** (516), 1705–1714.
- [14] Arribas-Gil, A. and Romo, J. (2014) Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, **15** (4), 603–619.
- [15] Hubert, M., Rousseeuw, P.J., and Segaeert, P. (2015) Multivariate functional outlier detection. *Stat. Methods Appl.*, **24** (2), 177–202.
- [16] Rousseeuw, P.J., Raymaekers, J., and Hubert, M. (2018) A measure of directional outlyingness with applications to image data and video. *J. Comput. Graph. Stat.*, **27** (2), 345–359.
- [17] Harris, T., Tucker, J.D., Li, B. and Shand, L. (2019) Elastic depths for detecting shape anomalies in functional data. *arXiv:1907.06759*.
- [18] Tukey, J.W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, PA.
- [19] Sun, Y. and Genton, M.G. (2012) Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, **23** (1), 54–64.
- [20] Sun, Y., Genton, M.G., and Nychka, D.W. (2012) Exact fast computation of band depth for large functional datasets: how quickly can one million curves be ranked? *Stat*, **1**, 68–74.
- [21] Martínez-Hernández, I., Genton, M.G., and González-Farías, G. (2019) Robust depth-based estimation of the functional autoregressive model. *Comput. Stat. Data Anal.*, **131**, 66–79.
- [22] Huang, H. and Sun, Y. (2019) Visualization and assessment of spatio-temporal covariance properties. *Spat. Stat.*, **34**, 100272.
- [23] Sun, Y. and Genton, M.G. (2012) Functional median polish. *J. Agric. Biol. Environ. Stat.*, **17** (3), 354–376.
- [24] Qu, Z., Dai, W., and Genton, M.G. (2020) Robust functional multivariate analysis of variance with environmental applications. *Environmetrics*, in press. <https://doi.org/10.1002/env.2641>
- [25] Ngo, D., Sun, Y., Genton, M.G., et al. (2015) An exploratory data analysis of electroencephalograms using the functional boxplots approach. *Front. Neurosci.*, **9** (282), 1–18.
- [26] Euan, C. and Sun, Y. (2019) Directional spectra-based clustering for visualizing patterns of ocean waves and winds. *J. Comput. Graph. Stat.*, **28** (3), 659–670.
- [27] Sun, Y. and Stein, M.L. (2015) A stochastic space-time model for intermittent precipitation occurrences. *Ann. Appl. Stat.*, **9** (4), 2110–2132.
- [28] La Vecchia, D. and Ronchetti, E. (2019) Saddlepoint approximations for short and long memory time series: a frequency domain approach. *J. Econom.*, **213** (2), 578–592.
- [29] Hyndman, R.J. and Shang, H.L. (2010) Rainbow plots, bagplots, and boxplots for functional data. *J. Comput. Graph. Stat.*, **19** (1), 29–45.
- [30] Xie, W., Kurtek, S., Bharath, K., and Sun, Y. (2017) A geometric approach to visualization of variability in functional data. *J. Am. Stat. Assoc.*, **112** (519), 979–993.
- [31] Genton, M.G., Johnson, C., Potter, K., et al. (2014) Surface boxplots. *Stat*, **3**, 1–11.
- [32] Yan, Y., Huang, H.C., and Genton, M.G. (2020) Vector autoregressive models with spatially structured coefficients for time series on a spatial grid. *arXiv:2001.02250*.
- [33] Dai, W., Mrkvička, T., Sun, Y., and Genton, M.G. (2020) Functional outlier detection and taxonomy by sequential transformations. *Comput. Stat. Data Anal.*, **149**, 106960.
- [34] Dai, W. and Genton, M.G. (2018) Multivariate functional data visualization and outlier detection. *J. Comput. Graph. Stat.*, **27** (4), 923–934.
- [35] Dai, W. and Genton, M.G. (2018) Functional boxplots for multivariate curves. *Stat*, **7**, e190.
- [36] Mirzargar, M., Whitaker, R.T., and Kirby, R.M. (2014) Curve boxplot: generalization of boxplot for ensembles of curves. *IEEE Trans. Vis. Comput. Graph.*, **20** (12), 2654–2663.
- [37] Yao, Z., Dai, W. and Genton, M.G. (2020) Trajectory functional boxplots. *Stat*, in press. <https://doi.org/10.1002/sta4.289>
- [38] Ramsay, J.O., Wickham, H., Graves, S., and Hooker, G. (2018) fda: functional data analysis. R package version 2.4.8.
- [39] R Core Team (2019) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- [40] MATLAB (2018) version 9.4.0 (R2018a). The MathWorks Inc., Natick, MA.

- [41] Wrobel, J., Park, S.Y., Staicu, A.M., and Goldsmith, J. (2016) Interactive graphics for functional data analyses. *Stat*, 5 (1), 108–118.
- [42] Wrobel, J. and Goldsmith, J. (2016) refund.shiny: Interactive Plotting for Functional Data Analyses, R package version 0.3.0.