

Outlier Detection using Graphical and Nongraphical Functional Methods in Hydrology

Insia Hussain¹

Department of Statistics
University of Karachi, Karachi
Pakistan

Abstract—Graphical methods are intended to be introduced in hydrology for visualizing functional data and detecting outliers as smooth curves. These proposed methods comprise of a rainbow plot for visualization of data in large amount and bivariate and functional bagplot and boxplot for detection of outliers graphically. The bagplot and boxplot are composed by using first two score series of robust principal component following Tukey's depth and regions of highest density. These proposed methods have the tendency to produce not only the graphical display of hydrological data but also the detected outliers. These outliers are intended to be compared with outliers obtained from several other existing nongraphical methods of outlier detection in functional context so that the superiority of the proposed graphical methods for identifying outliers can be legitimated. Hence present paper aims to demonstrate that the graphical methods for detection of outliers are authentic and reliable approaches compare to those methods of outlier detection that are nongraphical.

Keywords—Rainbow plot; bivariate bagplot; functional bagplot; bivariate boxplot; functional boxplot

I. INTRODUCTION

Methods of visualization help in exploring those characteristics which might have been neglected when using summary statistics and mathematical models. This area of study did not receive much attention when analysing data in a functional context.

Reference [1] bought this matter under consideration and introduced graphical tools for visualizing functional data and detecting functional outliers. These proposed graphical tools include three new methods the rainbow plot, the bagplot and the boxplot for displaying functional data graphically since it was aimed to make a contribution to the analytic toolbox of functional data. The bagplot and boxplot have the benefit of identifying outliers graphically which might not be visible through the original data plot. Curves of outliers may either located outside the data range and represent "magnitude outliers" or they may locate within the data range but having different shape compare to other existing curves and appear as "shape outliers" or they might exhibit due to these combined features. Outliers of all types should be dealt while attempting to identify curves of outliers. The proposed graphical methods were also legitimated as the better approaches for detecting outliers by comparing them with various other existing nongraphical methods of outlier detection.

The proposed graphical methods for data visualization and outlier detection were practised by [2-4] on hydrological data. Hence the work of [1, 2] is intended to be implemented on a daily flow series of Taunsa Barrage on Indus river in Pakistan.

II. LITERATURE REVIEW

Functional data are gaining immense importance in numerous fields, therefore latest statistical tools are required to be developed so that functional data can be analysed. Hence, the analysis of data comprises of the smooth curve are of interest in the present paper. Authors in [5-7] present surveys in detail regarding many nonparametric techniques to analyse functional data.

Most of the studies emphasis on functional data clustering, modelling and forecasting with visualization having a minor role to play. The phase-plane plot and the rug plot presented by [8] and [9], respectively, are of notable exceptions for highlighting characteristics of important distributions using first and second functional data derivatives. A plot of singular value decomposition introduced by [10] is another exception, which shows the latent component changes due to an increase in the dimensionality or sample size.

The work of [1] has been extended for identification of outlier curves with respect to shape or magnitude or combination of magnitude-shape features. Outliergram was introduced by [11] for detecting shape outliers, [12] presented functional outliers taxonomy by proposing methods of visualization for detecting outliers whereas [13] introduced plots for displaying outlier curves having combined magnitude-shape features. Several other methods have also been recently developed by [14-17] for data visualization and outlier detection in functional framework. The study conducted by [2] in hydrology has been employed for classification of hydrograph by [18] and for the purpose of streamflow forecasting by [19].

This paper is organized as follows: Section 2 contains the discussion about graphical and nongraphical functional methods. The description of employed hydrological data is presented in Section 3. Section 4 contains the application, discussion and results of the discussed methods on the employed hydrological data and the corresponding conclusion are presented in Section 5.

III. METHODOLOGY

This section represents a functional graphical method for visualizing data and also functional graphical and nongraphical methods for detecting and computing outliers, respectively. The rainbow plot provides data visualization graphically whereas bagplot and boxplot are graphical methods for outlier detection. Outliers are computed nongraphically using functional mathematical methods which include methods of likelihood ratio test, integrated error square, functional depth and robust Mahalanobis distances.

A. Functional Graphical Methods for Data Visualization and Outlier Detection

Outliers are observations which are unusual and required to be detected and treated [20]. A crucial phase before modelling in data analysis is to detect and treat outliers since outliers also affect the analysis, modelling and forecasting of data in hydrology, therefore detailed study in this regard was conducted by [2] in hydrology. Hence, the present section contains functional graphical methods for outlier detection to explore and analyse unusual observations.

Various methods were presented in some works of literature of functional context for outlier detection by [21-23]. However, [1] introduced new tools for detection of outliers which do not only identify outliers accurately and have fast computing speed but also provide a graphical presentation of outliers to observe the behaviour of unusual observations. These proposed graphical approaches contain rainbow plot for visualization of data pattern whereas bagplot and boxplot are the tools for identifying outliers graphically. The observations or curves which locate within or outside the range of data due to unusual structure or shape are identified as outliers. Practically, results may vary between the two methods proposed for outlier detection depending on the nature of employed data.

1) *Rainbow plot*: The rainbow plot introduced by [1] and implemented by [2,4] in hydrology for the purpose of visualizing a complete set of employed data through a single plot having a distinct feature of displaying data curves using rainbow colour palette following data order. Two indices are used for data ordering that are functional depth and data density. The bivariate score depth and kernel density estimates are used for computation of the data ordering indices. The bivariate score depth is written as

$$OT_j = d(z_j, Z), \quad Z = \{z_k \in \mathbb{R}^2; k = 1, \dots, n\} \quad (1)$$

where $d(\cdot, \cdot)$ is the function of half-space depth introduced by [24]. Depth function by Tukey for bivariate score series z_j of first two principal components is defined as the amount of smallest observations present in closed half-space with z_j points on the boundary. The observations arrange decreasingly following depth values OT_j order. The first curve according to the depth order is the median curve having highest OT_j whereas curve having lowest OT_j is the outermost curve. The median curve based on Tukey's depth function is obtained using a bivariate score series z_j as observations of the curve which is adopted as an alternate approach to original data for

simplification. The series z_j as observations of functional data are also ordered using kernel density estimate [25] as follows:

$$OD_j = \hat{f}(z_j) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{z_j - z_k}{h_k}\right), \quad j \neq k, j = 1, \dots, n \quad (2)$$

where $K(\cdot)$ is a kernel function and smoothing parameter h_k is a bandwidth of k th point $\{z_k\}$. The observations are arranged decreasingly following OD_j order. Hence, the mode curve is the first curve containing the highest OD whereas the curve containing the lowest OD is an outlying curve. The curves ordered according to OT and OD are displayed according to order of rainbow colour in such a way that the curves closest to median or mode appear in red whereas outlying curves appear in violet colour.

2) *Functional bagplot*: The bivariate bagplot based on half-space depth function was introduced by [26]. This plot was employed in multivariate and functional context by [27] and [2,4] respectively in hydrology using streamflow data. The functional bagplot is obtained following the pattern of bivariate bagplot which is composed using the first two score series $z_j = (z_{j,1}, z_{j,2})$ of principal components in such a way that each bivariate bagplot point correspond to each functional bagplot curve. Three elements that are the central median curve surrounded by inner region enclosed in an outer region cause the formation of bivariate bagplot which also contribute to the composition of functional bagplot. The inner region captures total observation of 50 per cent and surrounded by 95 or 99 per cent of total observation bounded by the outer region which is formed by inflating inner region according to ρ factor. Author in [1] suggested ρ value 1.96 for 95 per cent inflation and 2.58 for 99 per cent inflation since the standard normal distribution is followed by score series z_j . Finally, the points appear outside the inflated outer region are outliers.

3) *Functional boxplot*: The bivariate boxplot based on highest density region is composed by the first two z_j score vector of the principal component was introduced by [28]. The pattern employed for the formation of bivariate boxplot is followed for the composition of a functional boxplot. The detailed study regarding functional boxplot was organized by [1] and practised by [2,4] in hydrology. Recently, applications of boxplot in functional context have been conducted by [29,30].

As discussed in subsequent section, three elements correspond to bivariate boxplot, combine to compose a functional boxplot such that mode curve is formed in the centre of inner region containing 50 per cent data observations enclosed in an outer region having 95 or 99 per cent of total observation while the observations not bounded by outer region and located outside the region are outliers.

The difference in outliers obtained by bagplot and boxplot are due to the approaches used for establishing inner and outer regions since median and depth function belong to bagplot whereas mode and density estimate belong to boxplot. Hence, outliers displayed by bagplot are unusual observations compared to median whereas boxplot display outliers which are unusual with respect to mode.

B. Detection of Outliers using Nongraphical Functional Methods

Reference [1] proposed the comparison of graphical methods (i.e. bagplot & boxplot) with various nongraphical methods for detecting functional outliers, performed on discretized functions. The graphical method has been legitimated better able to detect outliers than other nongraphical methods. The nongraphical methods for detecting functional outliers are as follows:

1) *Method of likelihood ratio test:* Reference [22] proposed a method for detecting outlier which calculates a statistic of likelihood ratio test for each $y_j(x)$ curve. An outlier is a value when most of the test statistics exceed a given c critical value. This outlier value is then removed and the new outlier is obtained from the remaining data. This process stops when no more values of outliers are detected. Following equation based on depth function is used for this test

$$o_j = \int D(y_j(x)) dx \quad (3)$$

A univariate measure of depth for a particular x value is $D(y_j(x))$. According to this definition, the curves are ordered increasingly by an order of $\{o_j\}$, such that a curve with the lowest depth function is the first curve whereas a curve with the highest depth function is the last curve.

2) *Method of integrated error square:* Author in [31] proposed a method for detecting outliers which contains the principal component analysis of robust function. Let the integration of error square for j th observation be

$$v_j(x) = \int_x e_j^2(x) g dx = \int_x (y_j(x) - \sum_{l=1}^L z_{j,l} \phi_l(x))^2 dx \quad (4)$$

where L is a number of pre-specified components; usually 2, $\{\phi_l(x)\}$ are the functions of principal components and $z_{j,l}$ are their related scores. This provides an accurate measure of the approximated principal component for j th observation. High integrated error square indicates points with high likelihood as outliers. If $e_j(x)$ is distributed normally, then $v_j(x)$ follows a chi-square (χ^2) distribution with $E(v_j(x)) = 0.5Var(v_j(x))$ then the probability that $v_j < c$, where $c = m + \lambda\sqrt{m}$ and $m = \text{median}(\{v_1, \dots, v_n\})$, is approximately $\Phi(\lambda/\sqrt{2})$, where $\Phi(\cdot)$ is the standard normal distribution function. For example, $\Phi(3.29/\sqrt{2}) = 99\%$ when $\lambda = 3.29$.

3) *Method of functional depth:* Author in [32] proposed a method for detecting outliers by employing functional depth including all curves as well as trimmed curves. The centre of a group of curves is measured using the functional depth. Hence, a group of curves ordered in the outward centre is obtained through depth. In fact, the curve having maximum depth can be considered as an estimate for the functional distribution centre. Therefore, depth is an inverse notion of outlyingness and curves of functional outliers which are supposed to be located far from the central region of the data,

are the curves with low depth. Hence, functional outlier curves are detected having comparatively lower depth.

4) *Method of robust mahalanobis distances:* The Robust Mahalanobis distance is a multivariate method for detection of outliers which is applicable on discretized curves $\{y_j(x_k); k = 1, \dots, p\}$. Suppose the functional data are obtained on an equal distance $\{x_1, \dots, x_p\}$ dense grid, the Mahalanobis distance square is defined as

$$r_j = (y_j(x_k) - \hat{\mu}(x_k))' \hat{\Sigma}^{-1} (y_j(x_k) - \hat{\mu}(x_k)), \quad k = 1, \dots, p \quad (5)$$

where the mean of a sample is $\hat{\mu}(x_k)$, $\hat{\Sigma}$ is a covariance matrix robust estimate of $\{y_j(x_k)\}$ which is assumed to be a positive definite, such that $\hat{\Sigma}^{-1}$ is nonsingular. The comparison of resultant distance with a critical value follows χ^2 distribution having degrees of freedom p . For $\alpha=99\%$, a predefined level, observations are considered to be outliers that contain Mahalanobis distance square greater than $\chi_{0.99,p}^2$ critical value. The Robust Mahalanobis distance variations are further discussed by [21,33].

IV. DATA DESCRIPTION

The major data source in hydrology is daily streamflow. Flows of water are also recorded instantly, hourly or on other time scales. The data series of the daily flow of Taunsa barrage are available from Sindh Irrigation department, Sindh Secretariat, Karachi, Pakistan.

Taunsa barrage is built in Taunsa Tehsil of Dera Ghazi Khan District located in Punjab province on the Indus River in Pakistan. It has a discharge capacity of up to 1,000,000 cusecs (i.e. approximately $28300 \text{ m}^3 \text{ s}^{-1}$). Fig. 1 indicates the geographical location of the Taunsa Barrage.

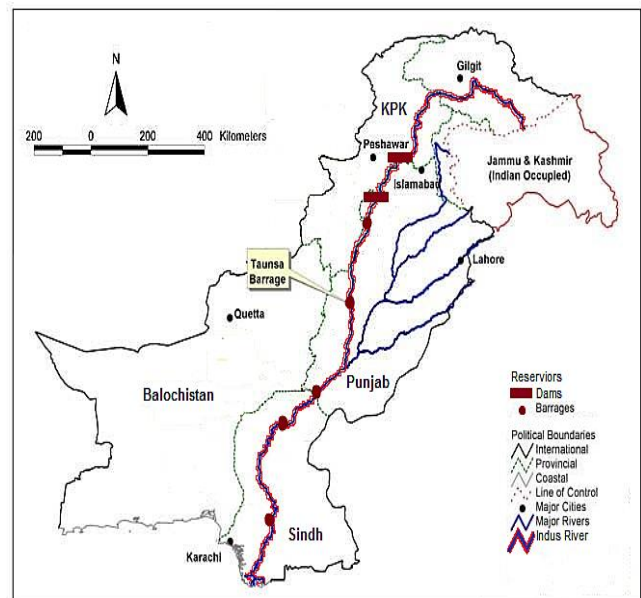


Fig. 1. Geographical Location of Taunsa Barrage

Some studies use data of entire year while some use data of only high flow season depending on what region is considered as an area for flood analysis. Data used in the present study are the observations of 6 months ($T = 183$ days) for the duration of years 1977 to 2017 ($n=41$ years) since high flow season is observed during April to September in Pakistan. The series of observations are $Y_j = (y_j(t_1), \dots, y_j(t_T))$, $j = 1, \dots, n$, $k = 1, \dots, T$, where $n=41$ years, $T = 183$ days and $y_j(t_k)$ is the recorded discrete flow value on t_k day in the j th year. Before any calculation is performed the values of streamflow which are recorded originally in cusec (a volume flow rate) are required to be converted into cubic meter per second ($m^3 s^{-1}$).

V. ANALYSIS

Specific formats of data ordering are followed by all the proposed graphical methods. Rainbow plot displays the curves of all the data in a single plot having a distinct feature of colour palette following the order of rainbow colours. By default, time order is followed by the curves of rainbow plot in such a way that the curve of the recent past appear in violet whereas curve belongs to remote past is appear in red colour.

The rainbow plot is the simplest way of displaying all the data in the functional format ordered with respect to time. Due to the overlapping of the number of curves, “median curve” or “mode curve” is not easily identified or it is difficult to locate where the majority of data curves present. Also, outliers are difficult to be identified if other curves obscure them. Therefore, there are also other options for ordering functional data using (1) and (2), respectively. This ordering is done using rainbow plot such that curves colour then follows the pattern of rainbow according to depth or density ordering. The curve near the core of data appears in red, whereas the curve in violet is an outlying curve. Since plotted curves follow either depth or density ordering, the red data curves are usually obscure, whereas data curves in violet are seen distinctly even if they are overlapped by other data curves.

The simple rainbow plot follows time ordering as shown by Fig. 2(a), but some data sets required to be ordered according to the data value themselves. Therefore, two optional ordering methods stated earlier i.e. depth or density uses the first two scores of principal components for measuring either “depth” or “density” of data according to the order of data values.

Tukey’s depth is described as the number of smallest data values present in an enclosed half-plane. The data values are then increasingly ordered as the $OT_i = d(z_i, Z)$ distance, using (1). The curve we first obtain by this order is considered as a curve of the median, while the last order curve is the outlying curve among all curves where a distance from the median to each curve is defined as depth. Depth ordering rainbow plot is depicted in Fig. 2(b). The black curve in the centre represents the median, the curve in purple is an outlying curve and the curve in red, near the median, is not clearly seen due to the other data curves are overlapping.

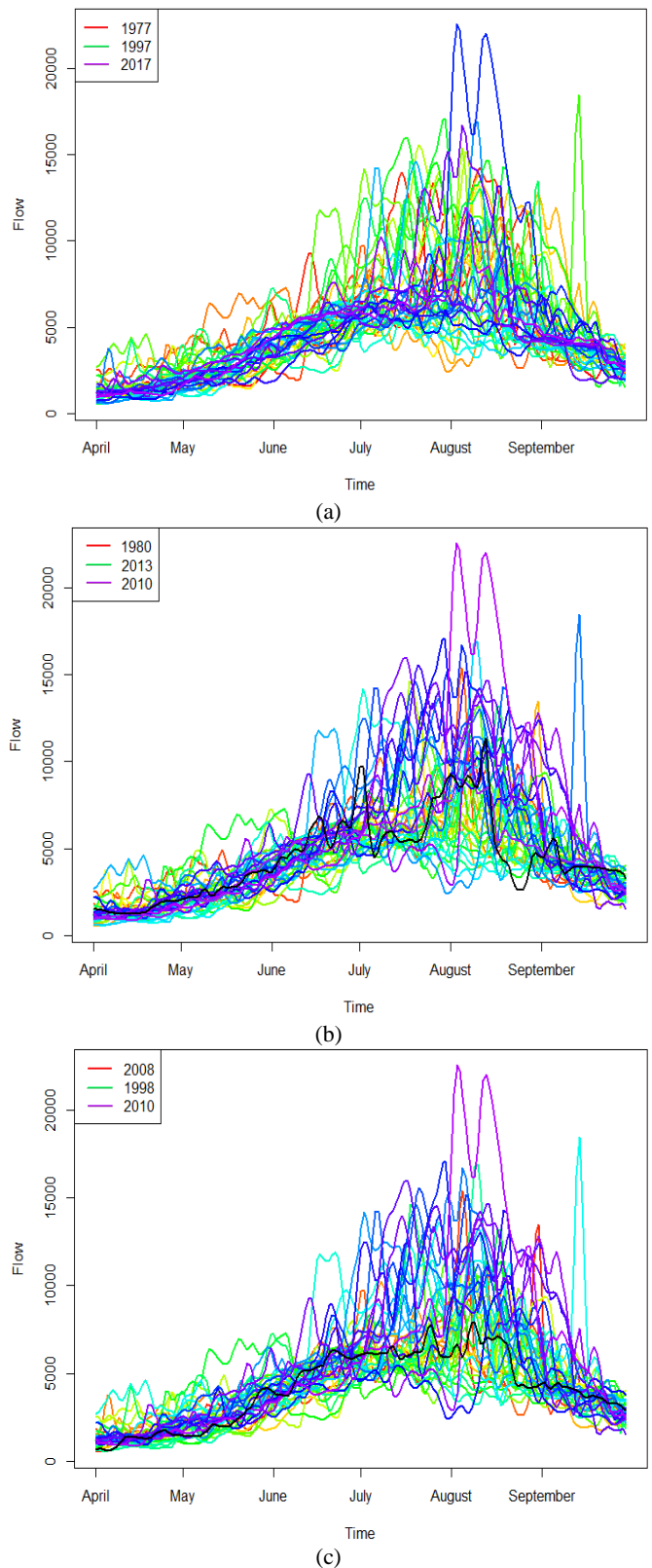


Fig. 2. Rainbow Plots with (a) Time (b) Depth and (c) Density Ordering for Years 1977-2017.

The third way for ordering data points is by the estimate of kernel density at each data value, using (2). The values of $\{OD_i\}$ decreasingly order the functional data, such that the first order curve is the curve of highest density and considered as “mode curve” whereas the last order curve is the curve of lowest density and considered as an unusual outlying curve. It can be noted that the values of the last curve obtained through this ordering is possibly remain the same compared to the other curves and its (z_1, z_2) bivariate scores might be in the centre of (z_1, z_2) scatter plot such that no other points are around it and hence having the value of lower density. Results reveal that both ordering methods may follow a similar order, especially for high ordering years. Density ordering rainbow plot is depicted in Fig. 2(c). The black curve in the centre represents a mode, the curve in purple is an outlying curve and the curve in red, near mode, is not clearly seen due to the other data curves are overlapping.

The data observations ordered by depth and density, lead to the formation of bivariate and functional bagplot and boxplot, as described in the subsections of 2.1. The association of bivariate and functional bagplot and boxplot to the first two scores of principal components for the probability coverage of 95% and 99% are illustrated in Fig. 3 and Fig. 4. The bivariate bagplot and boxplot exhibit median and mode, respectively, the inner region which covers 50% and an outer region which accumulate 95% or 99% of data values depending on the selected values of factor ρ i.e. either 1.96 or 2.58. The functional bagplot and boxplot contain curve of median and mode, respectively, area of the inner region and outer region. The functional bagplot and boxplot inner and outer region contain 50% and 95% or 99% of all the data curves, respectively.

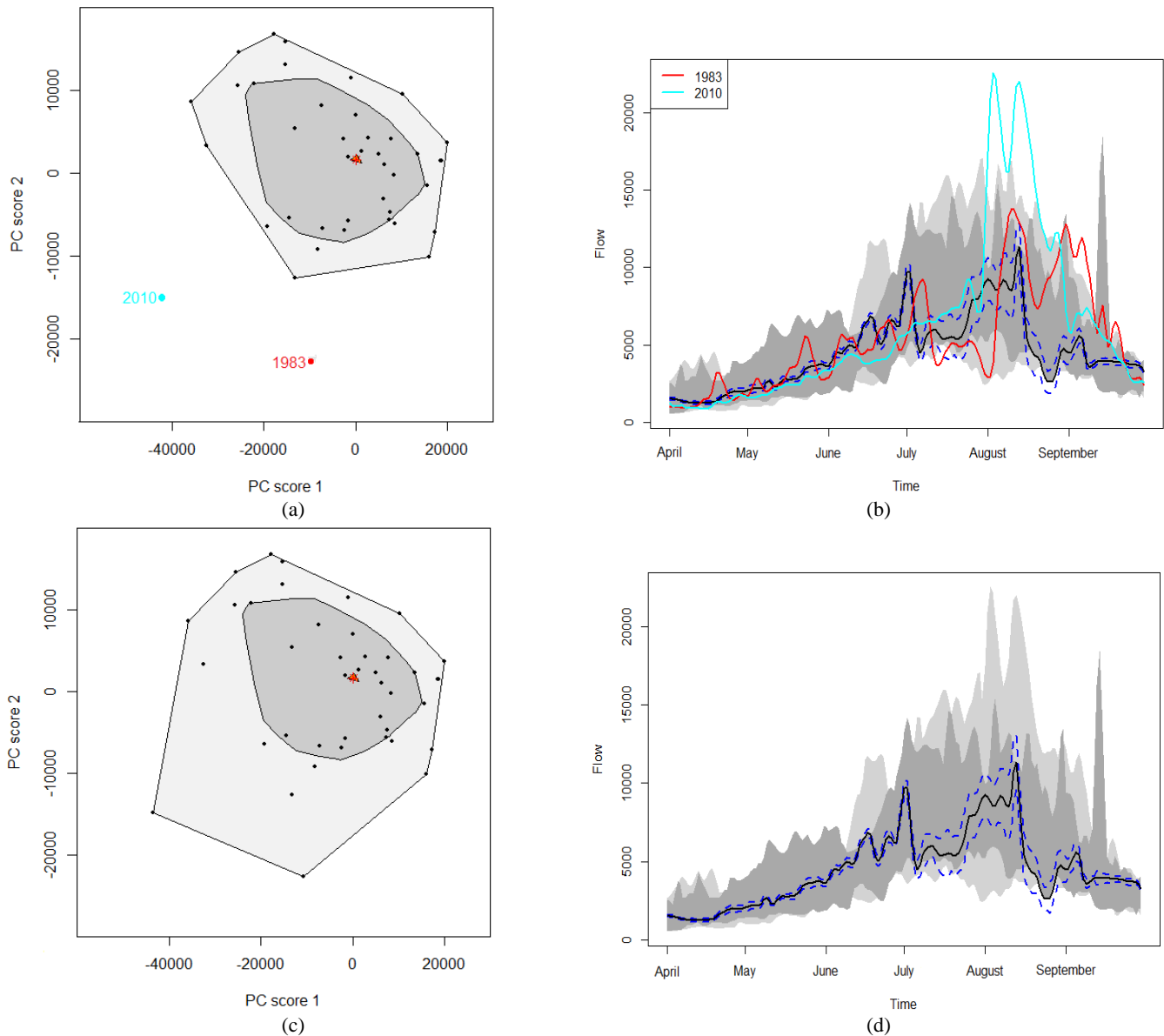


Fig. 3. (a) The bivariate and (b) Functional Bag Plot with 95%, whereas (c) The Bivariate and (d) Functional Bag Plot with 99% of Probability Coverage.

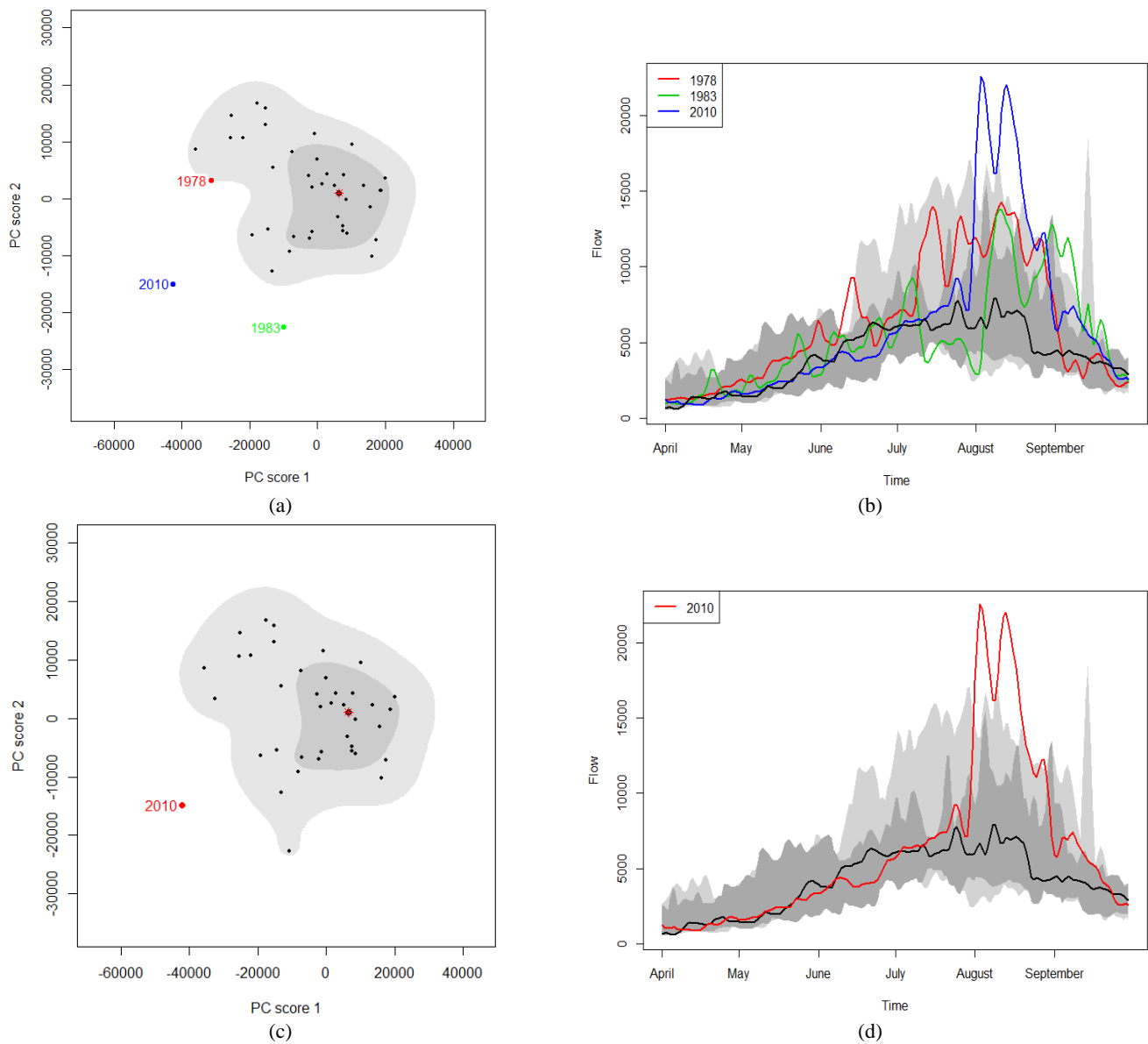


Fig. 4. (a) The Bivariate and (b) Functional Bagplot with 95%, While (c) The Bivariate and (d) Functional Bagplot with 99% of Probability Coverage.

The dark grey shaded area of bivariate bagplot and boxplot displays inner 50% region whereas area in light grey shade exhibits outer 95% and 99% region. These shaded areas correspond to the region of functional bagplot and boxplot with same shades. Points that appear beyond the outer region are detected as outliers. The bivariate bagplot and boxplot outliers appear in several colours match to the colour of functional bagplot and boxplot outlier curves. The median and mode display as the mark of a red asterisk (Fig. 3(a) and 3(c)) and (Fig. 4(a) and 4(c)), respectively, which corresponds to the black curve of median and mode appears in functional bagplot and boxplot (Fig. 3(b) and 3(d)) and (Fig. 4(b) and 4(d)), respectively.

It is clearly observed from Fig. 3 that the years 1983 and 2010 are identified as outliers when 95% region of both bivariate and functional bagplots are considered as displayed by Fig. 3(a) and 3(b). Besides this, no outlier is located outside

the 99% region of bivariate and functional bagplots, as exhibited by Fig. 3(c) and 3(d). Note that, usually when outliers are found near the median, it is difficult to be detected by bagplot [1]. Therefore, it is authentic to use more appropriate approach and that is boxplot. The years 1978 and 1983 appear beyond 95% outer region as shown by Fig. 4(a) and 4(b). The only outlier detected with the probability coverage of 99% is the year 2010 which is displayed outside the bivariate and functional boxplot outer region as displayed by Fig. 4(c) and 4(d). It can be deduced that the outlying year 1978 is close to the median, therefore, it was not detected by bagplot whereas the outlying years 1986 and 2010 are not close to the median. It can be concluded that the flow corresponds to the year 2010 is an outlier having different shape and magnitude compared to the curve of other years. The year 2010 is an authentic outlier identified through both bagplot and boxplot. Hence, boxplot is considered a more reliable approach for the detection of outliers than bagplot.

As discussed earlier in Section 2.1 that the bagplot and boxplot are the graphical methods for detecting and identifying functional outliers. These detected outliers are compared to the outliers detected by various nongraphical methods for detecting functional outliers, as discussed in Section 2.2. The purpose of conducting this comparison is to legitimate that the graphical methods are better able to detect outliers than nongraphical methods in a functional context. The nongraphical methods for detecting outliers were applied to the smoothed flow data for years 1977-2017. Outputs of nongraphical methods; likelihood, integrated error, depth and Robust method, and graphical methods; bagplot and boxplot for functional data are tabulated in Table I with the computation time of each method.

VI. RESULTS

The results displayed by Table I demonstrates that likelihood method is unable to detect any outliers, the integrated method detected several outliers, a similar result is also produced by depth method while robust method provides a comparatively better result. Hence, the likelihood method produced the worst performance with respect to accuracy and computation time. The integrated method detected many years, among them only a few exist as outliers. The depth method is unable to encounter shape outliers according to [34] analysis, therefore results show false detection and slow computation time. In contrast, the highest accuracy and fastest time computation for detecting outliers are achieved by functional bagplot and boxplot, compare to other earlier implemented methods i.e. likelihood, integrated error, depth and Robust method.

The graphical methods to visualize functional data and identify functional outlier have advantages that they not only accurately detect outliers but also simultaneously give the graphical presentation to visualize those outliers. Hence, resultant values in Table I demonstrate that the performance of proposed graphical methods is explicitly better than existing nongraphical methods for detecting outliers. In spite of the fact that sometimes the method of bagplot could not detect those outliers which are around the curve of median and such outliers can be identified correctly by boxplot, the bagplot and boxplot are authentic methods for detecting correct functional outliers.

TABLE I. A COMPARISON OF THE OUTLIER DETECTION PERFORMANCES

Methods	Detected outliers	Time (mins)
Likelihood ratio test	None	1.54
Integrated squared error	1978, 1979, 1981, 1983, 1984, 1987, 1988, 1989, 1990, 1991, 1992, 1994, 1995, 1996, 1997, 1998, 2005, 2006, 2010, 2015	2.33
Functional depth	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017	2.19
Robust Mahalanobis Distance	2010	0.45
Functional bagplot	None	0.01
Functional HDR boxplot	2010	0.37

VII. CONCLUSION

In this paper, three graphical methods for visualization of functional data and identification of functional outliers have been employed. Ranking scores of principal components using depth or density of data is performed in a two-dimensional space so that inliers and outliers are separated. The graphical presentation is obtained by scores matching, displayed by both bivariate bagplot and boxplot followed by functional curves. The approaches employed in this paper have the benefits of accurately detecting outliers with a fast-paced computation while simultaneously representing it graphically.

The results obtained from the implementation of the proposed method on the data of Taunsa Barrage illustrate that the graphical methods are better in performance compare to existing nongraphical methods for detecting outliers, which either detect miss obvious or spurious outliers.

Besides this, the method of depth-based which contribute in the formation of bagplot is not able to detect outliers that are close to the curve of median whereas the method of density-based which contribute in the formation of boxplot can detect such outliers authentically. Hence boxplot is more reliable compared to bagplot nevertheless both the graphical approaches are authentic for displaying and detecting concrete outliers.

VIII. FUTURE SCOPE

Functional data modelling, analysing and forecasting are seriously affected due to the outlier presence. Statistical analysis can lead to invalid conclusions if the identification and treatment of outliers are ignored. Hence, a crucial phase before modelling is to detect and treat outliers so that outliers do not affect the analysis and forecasting of data in hydrology.

ACKNOWLEDGMENT

The author wants to express gratitude to Sindh Irrigation Department, Sindh Secretariat, Karachi, Pakistan for graciously catering with data of Taunsa Barrage for executing this study. The author also thanks the authors of Rainbow R packages [35].

REFERENCES

- [1] R. J. Hyndman and H. L. Shang, "Rainbow plots, bagplots, and boxplots for functional data," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 29–45, 2010. doi:10.2307/25651298.
- [2] F. Chebana, S. Dabo-Niang and T. B. Ouarda, "Exploratory functional flood frequency analysis and outlier detection," *Water Resources Research*, vol. 48, no. 4, W04514, 2012. doi:10.1029/2011WR011040.
- [3] L. Millán-Roures, I. Epifanio and V. Martínez, "Detection of anomalies in water networks by functional data analysis," *Mathematical Problems in Engineering*, vol. 2018, no.1, 2018. doi:10.1155/2018/5129735.
- [4] I. Hussain and M. Uddin, "Functional and multivariate hydrological data visualization and outlier detection of Sukkur Barrage," *International Journal of Computer Applications*, vol. 178, no. 28, pp. 20-29, 2019. doi:10.5120/ijca2019919097.
- [5] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2nd edition, 2005.
- [6] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, 2006 doi:10.1007/0-387-36620-2.

- [7] J.-L. Wang, J.-M. Chiou, and H.-G. Müller, "Functional data analysis," *Annual Review of Statistics and Its Application*, vol. 3, no. 1, pp. 257-295, 2016. doi:10.1146/annurev-statistics-041715-033624.
- [8] J. O. Ramsay, and J. B. Ramsay, "Functional data analysis of the dynamics of the monthly index of nondurable goods production," *Journal of Econometrics*, vol. 107, no. 1-2, pp. 327-344, 2002. doi:10.1016/S0304-4076(01)00127-0.
- [9] V. Hyde, J. Wolfgang, and G. Shmueli, "Investigating concurrency in online auctions through visualization," *The American Statistician*, vol. 60 no. 3, pp. 241-250, 2006. doi: 10.1198/000313006X124163.
- [10] L. Zhang, J. S. Marron, H. Shen, and Z. Zhu, "Singular value decomposition and its visualization," *Journal of Computational and Graphical Statistics*, vol. 16, no. 4, pp. 833-854, 2007. doi:10.1198/106186007X256080.
- [11] A. Arribas-Gil and J. Romo, "Shape outlier detection and visualization for functional data: the outliergram," *Biostatistics*, vol. 15, no. 4, pp. 603-619, 2014. doi:10.1093/biostatistics/kxu006.
- [12] M. Hubert, P. J. Rousseeuw, and P. Segaeert, "Multivariate functional outlier detection," *Statistical Methods and Applications*, vol. 24, no. 2, pp. 177-202, 2015. doi:10.1007/s10260-015-0297-8.
- [13] W. Dai and M. Genton, "Multivariate functional data visualization and outlier detection," *Journal of Computational and Graphical Statistics*, vol. 27, no. 4, pp. 923-934, 2018. doi:10.1080/10618600.2018.1473781.
- [14] W. Xie, S. Kurtek, K. Bharath, and Y. Sun, "A geometric approach to visualization of variability in functional data," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 979-993, 2017. doi:10.1080/01621459.2016.1256813.
- [15] F. Yu, L. Liu, L. Jin, N. Yu, and H. Shang, "A method for detecting outliers in functional data," *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, Beijing, pp. 7405-7410, 2017. doi: 10.1109/IECON.2017.8217297.
- [16] P. J. Rousseeuw, J. Raymaekers, and M. Hubert, "A measure of directional outlyingness with applications to image data and video," *Journal of Computational and Graphical Statistics*, vol.27, no. 2, pp. 345-359, 2018. doi:10.1080/10618600.2017.1366912.
- [17] W. Dai and M. Genton, "Directional outlyingness for multivariate functional data," *Computational Statistics & Data Analysis*, vol. 131, pp. 50-65, 2019. doi:10.1016/j.csda.2018.03.017.
- [18] C. Ternynck, M. A. B. Alaya, F. Chebana, S. Dabo-Niang, and T. B. M. J. Ouarda, "Streamflow hydrograph classification using functional data analysis," *Journal of Hydrometeorology*, vol. 17, no. 1, pp. 327-344, 2016. doi: 10.1175/JHM-D-14-0200.1.
- [19] P. Masselot, S. Dabo-Niang, F. Chebana, and T. B. Ouarda, "Streamflow forecasting using functional regression," *Journal of Hydrology*, vol. 538, pp. 754-766, 2016. doi: 10.1016/j.jhydrol.2016.04.048.
- [20] V. Barnett and T. Lewis, *Outlier in Statistical Data*, John Wiley, Chichester, U.K, 3rd edition, 1998.
- [21] J. Hardin and D. M. Rocke, "The distribution of robust distances," *Journal of Computational and Graphical Statistics*, vol. 14, no. 4, pp. 928-946, 2005. doi:10.1198/106186005X77685.
- [22] M. Febrero-Bande, P. Galeano, and W. González-Manteiga, "A functional analysis of NOx levels: location and scale estimation and outlier detection," *Computational Statistics*, vol. 22, no. 3, pp. 411-427, 2007. doi:10.1007/s00180-007-0048-x.
- [23] P. Filzmoser, R. Maronna, and M. Werner, "Outlier identification in high dimension," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1694-1711, 2008. doi:10.1016/j.csda.2007.05.018.
- [24] J. W. Tukey, "Mathematics and the picturing of data," *Proceedings of the 1975 International Congress of Mathematics*, vol. 2, pp. 523-531, 1975.
- [25] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley, New York, NY, USA, 1992.
- [26] P. Rousseeuw, I. Ruts, and J. W. Tukey, "The bagplot: A bivariate boxplot," *The American Statistician*, vol. 53, no. 4, pp. 382-387, 1999. doi:10.1080/00031305.1999.10474494.
- [27] F. Chebana and T. B. Ouarda, "Depth-based multivariate descriptive statistics with hydrological applications," *Journal of Geophysical Research*, vol. 116, D10120, 2011b. doi:10.1029/2010JD015338.
- [28] R. J. Hyndman, "Computing and graphing highest density regions," *The American Statistician*, vol. 50, no. 2, pp. 120-126, 1996. doi:10.2307/2684423.
- [29] W. Dai, T. Mrkvicka, Y. Sun, and M. G. Genton, "Functional outlier detection and taxonomy by sequential transformations," *arXiv e-prints*, 2018.
- [30] C. Zhao and J. Yang, "A robust skewed boxplot for detecting outliers in rainfall observations in real-time flood forecasting," *Advances in Meteorology*, vol. 2019, pp. 1-7, 2019. doi:10.1155/2019/1795673.
- [31] R. J. Hyndman and M. S. Ullah, "Robust forecasting of mortality and fertility rates: A functional data approach," *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 4942-4956, 2007. doi:10.1016/j.csda.2006.07.028.
- [32] M. Febrero-Bande, P. Galeano, and W. González-Manteiga, "Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels," *Environmetrics*, vol. 19, no. 4, pp. 331 - 345, 2008. doi:10.1002/env.878.
- [33] C. Becker and U. Gather, "The largest nonidentifiable outlier: a comparison of multivariate simultaneous outlier identification rules," *Computational Statistics & Data Analysis*, vol. 36, no. 1, pp. 119-127, 2001. doi:10.1016/S0167-9473(00)00032-3.
- [34] S. López-Pintado and J. Romo, "On the concept of depth for functional data," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 718-734, 2009. doi:10.1198/jasa.2009.0108.
- [35] H. L. Shang and R. J. Hyndman, *Rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data*, R package version 3.4, 2016.