



# Visualization of rainfall data using functional data analysis

Mohammed Abduljabbar Hael<sup>1,2</sup> · Yuan Yongsheng<sup>1</sup> · Bassiouny Ibrahim Saleh<sup>3,4</sup> Received: 2 December 2019 / Accepted: 12 February 2020 / Published online: 21 February 2020  
© Springer Nature Switzerland AG 2020

## Abstract

In recent years, functional data analysis (FDA) has been used in many applications in order to analyze data that provide detail on curves, surfaces or other components of a continuum. The purpose of this research is mainly to incorporate and adapt visualization tools for the FDA of rainfall in the Taiz Region. Moreover, visualizing the temporal variations of rainfall in order to provide a clear understanding of the rainfall patterns and also to help in predicting the future. The current study has been conducted based on average monthly rainfall over the Taiz Region during the period of 1998–2018. The R software was used to process the data obtained from the Tropical Rainfall Measuring Mission in Taiz City. The functional rainfall data have been smoothed using penalized smoothing according to generalized cross-validation criteria. The results showed that the rainfall profiles in the Taiz Region depend significantly on their temporal locations due to the monsoonal influences, which reflect the distribution of rains in the spring and the summer seasons.

**Keywords** Functional data analysis · Taiz City · Rainfall data · Penalized smoothing · Visualization · Singular value decomposition

## 1 Introduction

Recently, the field of functional data analysis (FDA) has seen phenomenal growth in numerous areas of applications such as environmental science, medical, engineering and biomaterials in order to enhance the planning methods and increase the efficiency of services or products [1–3]. Therefore, the FDA method is considered to be one of the most advanced techniques using all available data as practical measurements and curves [4–6]. This is due to the emergence of many modern technologies and software that have made mathematical computing possible in analyzing data by using the FDA [7–10]. So, the quantitative structure of the FDA approach could be viewed as an acceptable methodology with a specific perspective compared to traditional statistical analyses [11–14]. Many contributions cover a wide range of statistical problems involving the FDA such as mathematical foundations,

covariance operator estimation, functional depth, functional autoregressive processes, linear regression [15,16], semiparametric regression, nonparametric regression, spatial functional statistics, robust functional data analysis as well as sparsity in FDA [17–19].

Ramsay and Silverman [20] offered an excellent overview of FDA bases and expanded the application of classical mathematical science, such as primary component analysis, regression analysis, linear models and confidence intervals, to the practical system. In another related book, Ferraty and Vieu [21] provided studies on multiple parametric and nonparametric methods in order to produce useful tools for the FDA, while Kokoszka and Reimherr [22] provided a detailed description of the current quantitative structure of the FDA. In order to improve the credit card payment system, Laukaitis [23] studied the automatic regression model to obtain the cash flow index and transaction intensity using the FDA. Through this study, it

✉ Mohammed Abduljabbar Hael, 2014mohanned@gmail.com; Mohammed2020@taiz.edu.ye | <sup>1</sup>College of Science, Hohai University, Nanjing 211100, China. <sup>2</sup>Department of Statistics and Informatics, Taiz University, Taiz 9674, Yemen. <sup>3</sup>College of Mechanics and Materials, Hohai University, Nanjing 211100, China. <sup>4</sup>Production Engineering Department, Alexandria University, Alexandria 21544, Egypt.



is possible to predict and then solve the continuous stochastic problem for a full period of time. In another related study, Hyndman and Shang [24] introduced new tools to form smooth curves of large amounts of visualization using the FDA. These tools included a bagplot and boxplot that use the first two rigorous factor scores: the data depth of Tukey and the regions with the highest density. According to their results, the proposed tools detect the extreme values at a high-accuracy mathematical speed with a graphic representation simultaneously. In another work, Ikeda et al. [25] studied how loud observation data could be appropriate for smooth and continuous functions that capture the main features of plankton volatility, without the need for explicit assumptions or distribution factors for distribution by using the FDA. According to their findings, the FDA approach meets the requirements of stability and smoothness, without sudden changes in the curves, and is not influenced by any asymmetrical measurements and missing observations. In the case of gas emissions considering vectors, which have been component gas concentration values for each observation, Torres et al. [26] suggested a solution to the problem of the finding of extreme values in urban gas emissions using functional data analysis. On the basis of their results, for a period of 2 months after the study, they achieved the obtained carbon emission extreme values. In another related work, Shang [27] proposed many methods for visualizing functional time series which are implemented by R add-on package software in order to analyze the fertility data and sea surface temperatures data. These techniques obviously have the advantages of identifying the characteristics of a functional time series and also useful tools to visualize trends based on his findings. Recently, Suhaila and Yusop [28] used some of the FDA techniques such as functional descriptive statistics and functional analysis of variance in order to describe the spatial and temporal rainfall variations. The functional results provided a detailed description of the information about differences and Variabilities in rainfall profiles. Therefore, the FDA approaches were able to extract additional insights included in the curves and functions, which could not be available from classical statistical methods.

Based on the advantages and facilities which presented by the FDA methodology during the previous years; it is obvious that this methodology is considered an effective technique in dealing with diverse data and advanced analyses compared to the traditional approaches. Therefore, The FDA approach has been used for the current work on the predictive data and adaptation to rainfall data in the Taiz Region in order to assist decision-makers in planning well for the management of rainwater conservation and to understand the temporal changes during the last two decades. The basic process and associated

statistical techniques are shown in the method section of the theoretical context of FDA methodologies. Results and discussion section present the findings of the case study to understand the rainfall measurements in Taiz City. Conclusions and summary are discussed at the end of this research.

## 2 Methods and materials

This section presents the mathematical background of the FDA methodology that will be applied in the current work. The Eqs. (1)–(12) are employed to complete the research steps. The major step after getting data is to convert the discrete data into functional data, the next step is to smooth the functional data and the last two steps are to obtain the functional results of rainfall data as shown in Fig. 1.

### 2.1 Constructing basis functions and Fourier system

Observations or measurements are usually measured in specific periods, such as days, months or years. In the first step of functional data analysis, observed discrete data are converted into functional data objects. Functional data objects are defined through the specification of a set of basis functions and a number of coefficients that describe a linear combination of basis functions. Suppose that the values of the discrete observations  $y_i(t_j)$  are fitted using the standard model [3, 20]:

$$y_i(t_j) = x_i(t_j) + \epsilon_{ij}, \quad i = 1, \dots, n; j = 1, \dots, T \quad (1)$$

where  $(\epsilon_{ij})$  are measurement errors, and the functions  $x_i(t)$  are linear combinations of basis functions  $\phi_k(t)$ . This permits to illustrate most of the variation included in the functional data.

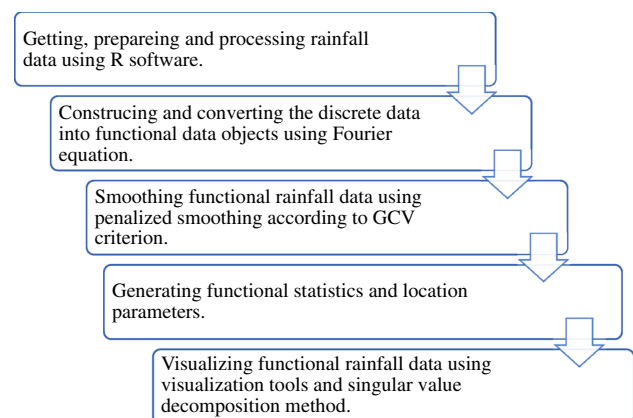


Fig. 1 Summary of the most important steps used in this research

A basis function system is a set of familiar functions  $\phi_k$  that are mathematically independent of each other; these functions have the attribute that converge well by taking a weighted sum or linear combination of a sufficiently big number  $K$  of basis functions.

The function is constructed based on a set of functional building blocks  $\phi_k, k = 1, \dots, K$  called basis functions expansion, which are combined linearly so that a function  $\mathbf{x}(\mathbf{t})$  defined in this method is expressed in the mathematical format as [3, 28]:

$$\mathbf{x}(\mathbf{t}) = \sum_{k=1}^K \mathbf{c}_k \phi_k(\mathbf{t}) = \mathbf{c}' \phi(\mathbf{t}) \tag{2}$$

where the parameters  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k$  are the coefficients of the expansion. The matrix expression  $\mathbf{c}' \phi$  uses  $\mathbf{c}$  to denote the vector of length  $K$  of coefficients  $\mathbf{c}_k$  and  $\phi$  to indicate a functional vector of length  $K$  comprising the basis functions  $\phi_k$ .

The Fourier basis is a system that is normally used for periodic functions and has the preference for high-speed computing and flexibility. The first function of the Fourier basis is the constant function. The remainders are pairs of sine and cosine with multiple integer base periods. The number of basis functions is always odd. The information that is required to define the Fourier basis system includes the number of basis functions  $K$  and the period  $T$ . The Fourier series can be written in the form of sine and cosine functions [3, 28]:

$$\hat{\mathbf{x}}(\mathbf{t}) = \mathbf{c}_0 + \mathbf{c}_1 \sin(\omega \mathbf{t}) + \mathbf{c}_2 \cos(\omega \mathbf{t}) + \mathbf{c}_3 \sin(2\omega \mathbf{t}) + \mathbf{c}_4 \cos(2\omega \mathbf{t}) + \dots, \tag{3}$$

Equation (3) is defined by the basis  $\phi_0(\mathbf{t}) = 1, \phi_{2r-1}(\mathbf{t}) = \sin(r\omega \mathbf{t}),$  and  $\phi_{2r}(\mathbf{t}) = \cos(r\omega \mathbf{t})$ . This basis is periodic, and the parameter  $\omega$  determines the period  $2\pi/\omega$ .

## 2.2 Penalized smoothing

The roughness penalty is determined through the establishment of a functional parameter object with a basis, smoothing parameter and penalized derivative. The next equation called the penalized sum square of error (PENSSE) is used to estimate the coefficient vector in order to minimize errors of curves [3, 28].

$$\text{PENSSE}_\lambda(\mathbf{x}) = \sum_j [\mathbf{y}_j - \mathbf{x}(\mathbf{t}_j)]^2 + \lambda \int [\mathbf{D}^4 \mathbf{x}(\mathbf{t})]^2 dt \tag{4}$$

where  $\mathbf{x}(\mathbf{t}) = \mathbf{c}' \phi(\mathbf{t})$ ; in the second term, the integrated fourth square derivative is called the roughness penalty, which calculates the overall curvature of the second derivative of  $\mathbf{x}$ . The  $\lambda$  is a parameter for smoothing; it measures

the exchange rate between the data normalization fit and smoothness. With the increase in  $\lambda$ , roughness is increasingly penalized and  $\mathbf{x}(\mathbf{t})$  will turn into linear, and with the decrease in  $\lambda$ , the penalty is reduced and allows  $\mathbf{x}(\mathbf{t})$  to fit the data better. The identical term for the vector of fits to the data is [20, 28]:

$$\mathbf{S}_\phi \mathbf{y} = \Phi (\Phi' \mathbf{W} \Phi + \lambda \mathbf{R})^{-1} \Phi' \mathbf{W} \mathbf{y} \tag{5}$$

where  $n$  by  $K$  matrix  $\Phi$  consists of the values of the  $K$  basis functions at the  $n$  sampling points,  $\mathbf{W}$  is a weight matrix to warrant for probable covariate structure among residuals and  $\mathbf{y}$  is the vector of discrete data to be smoothed. The  $\mathbf{S}_\phi$  is called the projection operator corresponding to the basis system  $\phi$ , and  $\mathbf{R}$  is known as the penalty matrix.

The generalized cross-validation measure (GCV) is designed to establish and choose the superior value for smoothing parameter  $\lambda$ . The criterion is given by [20, 28]:

$$\text{GCV}_\lambda = \frac{n^{-1} \sum [\mathbf{y}_i - \mathbf{x}(\mathbf{t}_i)]^2}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{S})]^2} \tag{6}$$

The degrees of freedom values are controlled by  $\lambda$ , and the measure of the smooth fit is  $df(\lambda) = \text{trace}(\mathbf{S}_{\phi, \lambda})$ , where the trace means the sum of the diagonal elements of  $\mathbf{S}_{\phi, \lambda}$ . The best choice of  $\lambda$  is associated with the minimum value of GCV.

## 2.3 Functional statistics and location parameters

### 2.3.1 Functions of mean and standard deviation

The traditional statistics for multivariate data are conventional to functional data. The mean function of curves is given as [3, 20, 28]:

$$\mu(\mathbf{t}) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i(\mathbf{t}) \tag{7}$$

Equation (8) describes the analysis of variance ( $\text{VAR}_X$ ), while Eq. (9) represents standard deviation (SD) [3, 20, 28]:

$$\text{VAR}_X(\mathbf{t}) = \frac{1}{N} \sum_{i=1}^n [\mathbf{x}_i(\mathbf{t}) - \bar{\mathbf{x}}(\mathbf{t})]^2 \tag{8}$$

$$\text{SD}_X(\mathbf{t}) = \sqrt{\text{VAR}_X(\mathbf{t})} \tag{9}$$

### 2.3.2 Functions of covariance and correlation

The functional covariance summarizes the dependence of measurements across different argument values and is estimated for all  $\mathbf{s}$  and  $\mathbf{t}$  by [3, 20, 28]:

$$\sigma(\mathbf{s}, \mathbf{t}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i(\mathbf{s}) - \bar{\mathbf{x}}(\mathbf{s}))(\mathbf{x}_i(\mathbf{t}) - \bar{\mathbf{x}}(\mathbf{t})) \quad (10)$$

The associated correlation function is

$$\rho(\mathbf{s}, \mathbf{t}) = \frac{\sum_{i=1}^N (\mathbf{x}_i(\mathbf{s}) - \bar{\mathbf{x}}(\mathbf{s}))(\mathbf{x}_i(\mathbf{t}) - \bar{\mathbf{x}}(\mathbf{t}))}{\sqrt{\sum_{i=1}^N (\mathbf{x}_i(\mathbf{s}) - \bar{\mathbf{x}}(\mathbf{s}))^2} \sqrt{\sum_{i=1}^N (\mathbf{x}_i(\mathbf{t}) - \bar{\mathbf{x}}(\mathbf{t}))^2}} \quad (11)$$

In multivariate data analysis, these are the functional analogs of the variance–covariance structures and correlation matrices, respectively. The variability of the functional data is visualized by plotting the surfaces of covariance and correlation as a function of  $\mathbf{s}$  and  $\mathbf{t}$  as well as the corresponding insights such as heat map, fields and contours.

### 2.4 The singular value decomposition technique

Singular value decomposition (SVD) is a useful method in visualizing patterns of functional and multivariate data. This technique was proposed by Zhang et al. [29]; they employed the idea of projection pursuit by revealing just low-dimensional projections that show interesting features of the high-dimensional point. SVD technique decomposes high-dimensional smoothed multivariate data into singular rows, singular columns and singular values ordered according to the amount of interpreted variance. The SVD of  $\mathbf{f}(x_i)$  is defined as [27, 29]:

$$\mathbf{f}(x_i) = \mathbf{s}_1 \mathbf{u}_1 \mathbf{v}_1^T + \mathbf{s}_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \mathbf{s}_K \mathbf{u}_K \mathbf{v}_K^T \quad (12)$$

where the singular columns  $\mathbf{u}_1, \dots, \mathbf{u}_K$  form  $K$  orthonormal basis functions for the column space extended by  $\{\mathbf{c}_j; j = 1, \dots, n\}$ ;  $n$  is the number of curves; the singular

rows  $\mathbf{v}_1, \dots, \mathbf{v}_K$  form  $K$  orthonormal basis functions for the row space extended by  $\{\mathbf{r}_i; i = 1, \dots, p\}$ ; and  $p$  is the number of covariates. The vector  $\{\mathbf{u}_K\}$  is called singular column, and the vector  $\{\mathbf{v}_K\}$  is called a singular row; the scalars  $\mathbf{s}_1, \dots, \mathbf{s}_K$  are called singular values. The matrix  $\{\mathbf{s}_K \mathbf{u}_K \mathbf{v}_K^T; k = 1, \dots, K\}$  indicates the SVD component.

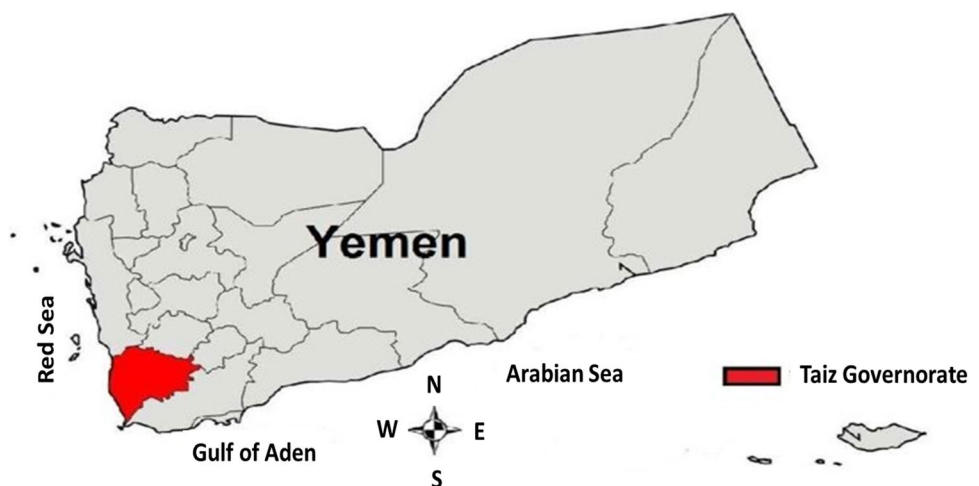
### 2.5 A case study

*Study area* Taiz City is one of the largest cities in Yemen and located in the southwest of Yemen as shown in Fig. 2 [30]. Its average height from sea level is 1311 m. The elevation is 1200 m above sea level with latitude N 13° 42' and longitude E 44° 55'. The mean annual rainfall and temperature are 588 mm and 24 °C, respectively, while the mean monthly evaporation is 140 mm [31].

*Data source* Average monthly precipitation over Taiz City, Yemen, region 13.5901 N and 44.2639 E of Taiz City during the period from January 1998 to December 2018 was obtained from the Tropical Rainfall Measuring Mission satellite data. In the present dataset, we have  $N = 21$  discrete measurements  $\mathbf{y}_i(\mathbf{t}_j), \mathbf{t}_j \in [1, 12], i = 1, \dots, 21$ ; the  $i$ th discrete observation  $\mathbf{y}_i(\mathbf{t}_j), j = 1, \dots, 12$  denotes rainfall records for the  $i$ th year that are converted into smoothed functions  $\mathbf{y}_i(\mathbf{t})$ . The discrete satellite data of average monthly precipitation  $\mathbf{y}_i(\mathbf{t}_j)$  are converted to smoothed curves  $\mathbf{y}_i(\mathbf{t})$  as continuous temporal functions by a base period of 12 months. This smoothing can be obtained by the predefine basis system.

*Used software* In the current study, the R software with the “FDA”, “rainbow”, “fields” and “plot3D” packages have been used to process the functional data [20, 27, 32, 33].

Fig. 2 Taiz governorate Map [30]



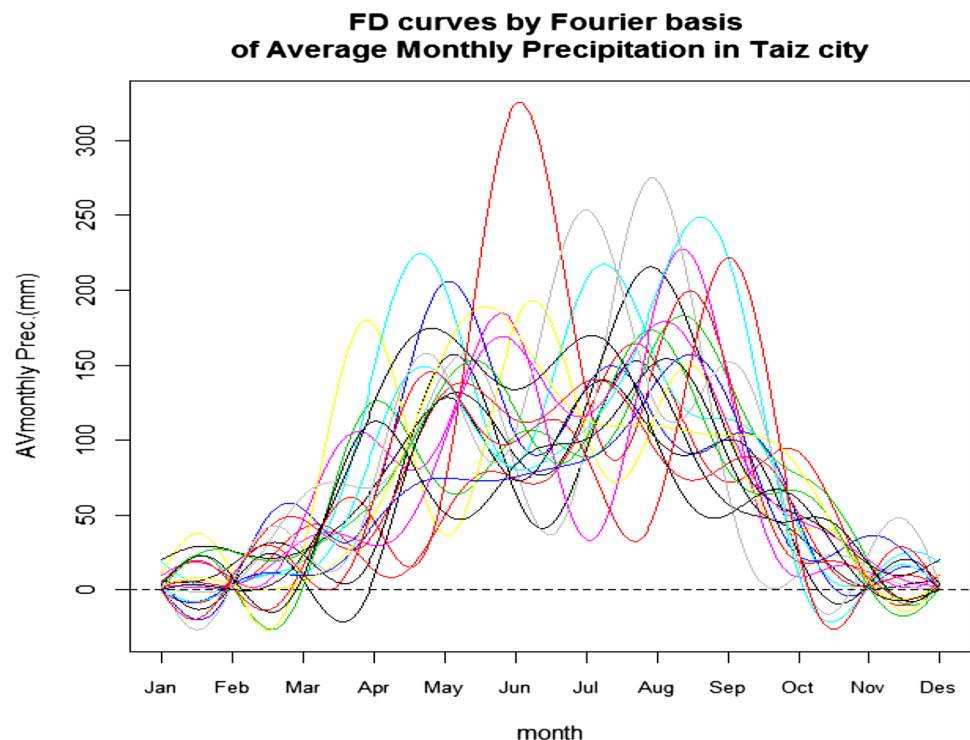
### 3 Results and discussion

This section is divided into four sections. Constructing functional data will be displayed in the first section, while penalized smoothing will be demonstrated in the second section. Then, functional statistics with visualization will be presented in the third section. The fourth section will display the results of singular value decomposition visualization.

#### 3.1 Constructing functional data

The first step is to transform the discrete rainfall details into continuous functions or curves. Seasonal fluctuations and periodicity of the rainfall data of the entire series are often shown throughout the annual cycle. Then, Fourier basis functions are preferred. The choice of  $k = 12$  can be justified to capture the precipitation variation within a month. In R software (FDA package), the discrete precipitation data were converted into functional data objects by using Fourier bases. All of the obtained functional data observations of the average monthly precipitation are represented in one plot, as shown in Fig. 3; this plot gives an idea that data are periodic. Therefore, this periodicity can justify the use of the Fourier basis functions, in order to mainly reach the peaks and circles. It is clear that certain basis functions are certainly too rough, especially for years with a much higher level of noise. Hence, as described in the next section, the penalized smoothing is used to improve functional results using general cross-validation (GCV) criteria.

**Fig. 3** Representation of all the functional data observations for average monthly rainfall using Fourier basis functions



#### 3.2 Penalized smoothing for functional precipitation data

In the previous section, functional data objects are explained and constructed by using the Fourier basis expansion. The goal of penalized smoothing is to eliminate the contribution of the errors and noise and obtain the best estimate of curves  $x(t)$ . The generalized cross-validation (GCV) criterion should be applied to choose the level of smoothing by arranging the harmonic acceleration operator with setting up a saturated basis that is capable of interpolating all the functional data of precipitation. The linear differential operator (LDO) is generated according to the formula  $(2\pi/12)^2$ , and selecting the smoothing parameter is controlled based on the minimized generalized cross-validation (GCV) estimate. It is important to choose a smoothing parameter that improves the quality of smoothing, reflects the structure of the data well and also captures a high percentage of illustrated variance. Therefore, the best-chosen smoothing parameter is 0.05, which minimizes generalized cross-validation, and the corresponding degree of freedom is 4. The connection between lambda with generalized cross-validation and lambda with a degree of freedom is described in Fig. 4. The smoothed precipitation curves for all years are visible as an assembly of the smoothed lines in Fig. 5. The penalized GCV smoothing indicates that the result of the curves well reflects the structure of the functional data and that the following functional results are dependent upon them.

### 3.3 Functional statistics and visualization

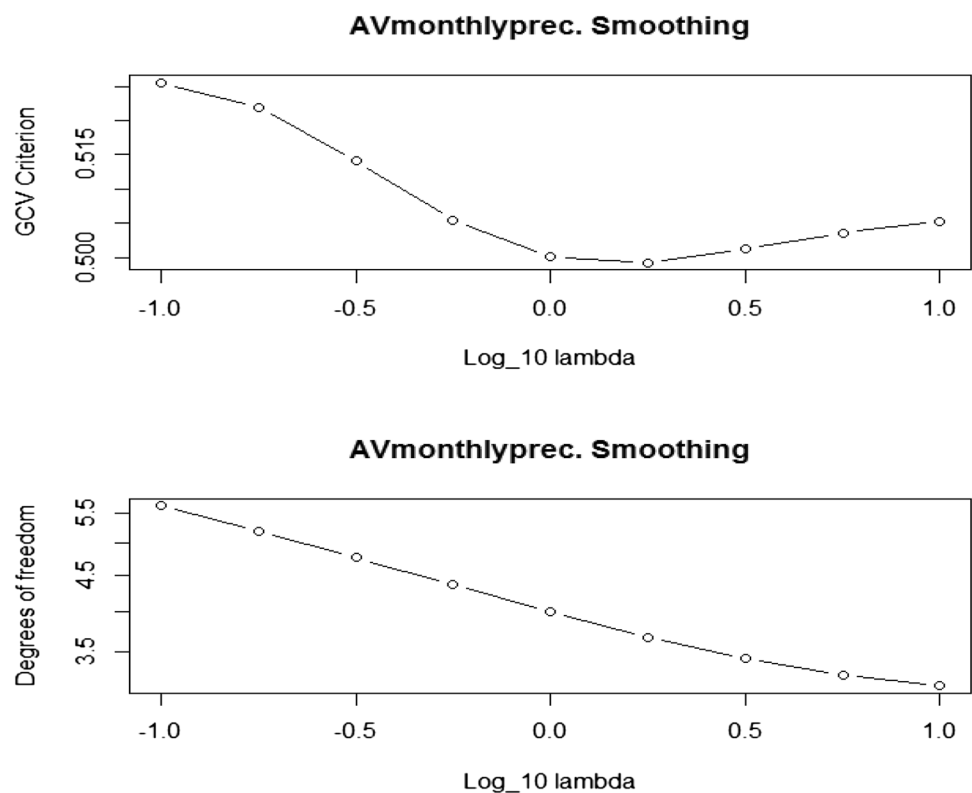
#### 3.3.1 Mean and standard deviation functions

The mean and standard deviation of functional data for average monthly precipitation are presented in Fig. 6a, the solid curve is the mean function of average monthly precipitation derived from the penalized smoothing and standard deviation curve is plotted in red dashed line. The confidence interval curves of the mean curve are presented in Fig. 6b, and the blue dashed lines represent 95% pointwise confidence intervals on the mean curve based on the penalized smoothing estimate of measurement plotted in Fig. 5 and standard deviation function plotted in Fig. 6a. The mean function provides a quick insight into the centering of functional data, and then the central curve of precipitation data can be determined. Generally, Fig. 6a shows that the maximum rainfall occurs in the summer season, especially in August and September. Besides, the maximum rainfall occurs in the spring season, especially in May and June, followed by a recession during the beginning and end of the year. The standard deviation gives us an idea about the typical variability of curves at any point of time  $t$ , but it gives no information on how the values of the curves at point  $t$  relate to those at point  $s$ ; the more in-depth analyses will be obtained in the next subsections.

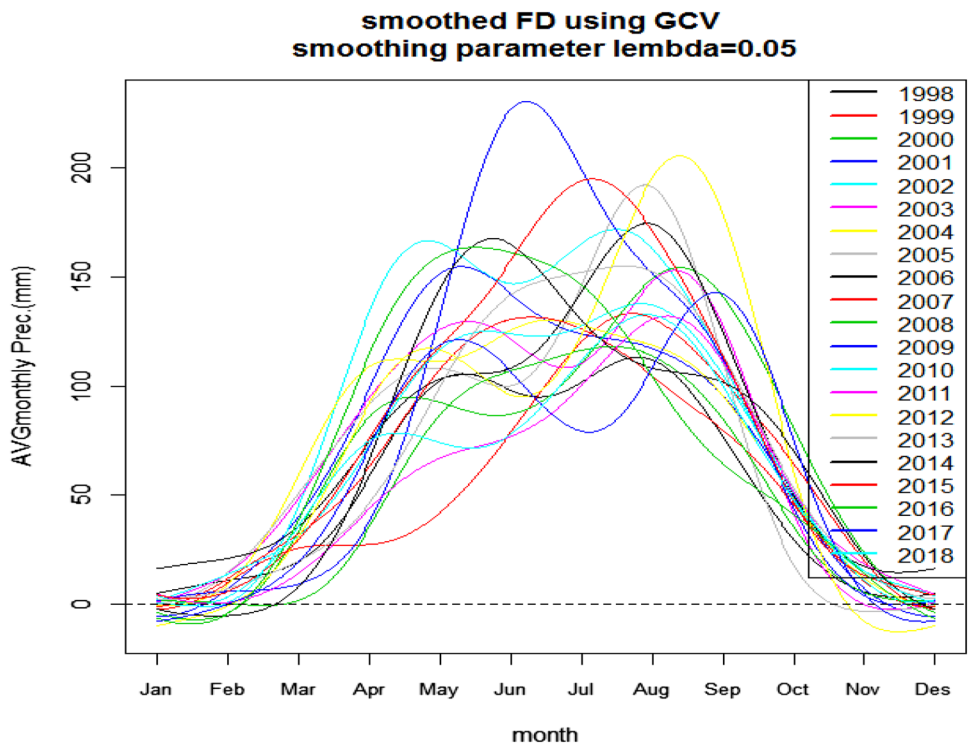
#### 3.3.2 Functional covariance and correlation with visualization

The bivariate temporal variance–covariance surface by contour map with fields is shown in Fig. 7, and perspective plot with heat map is shown in Fig. 8. These measures have the usefulness of supplying detailed information regarding the covariance structure, on the other side, when it takes place across temporal permutation. We observe that the main part of the variability occurs in the two periods of the year, and it is insignificant elsewhere. That is, the highest variability occurs completely in September and August and also in June and May. These periods correspond approximately to the highest rainfall. In the added fields onto the contour map, the black color indicates a very high covariance across years, and the red color indicates a slightly high covariance. In the heat map added onto the perspective plot, the red color indicates a strongly high covariance across years, and the blue color indicates a low covariance across years. The visualization for the bivariate temporal correlation surface using correlation functions is presented by contour map with fields as given in Fig. 9; it shows that there are strong correlations over average monthly precipitation across all years.

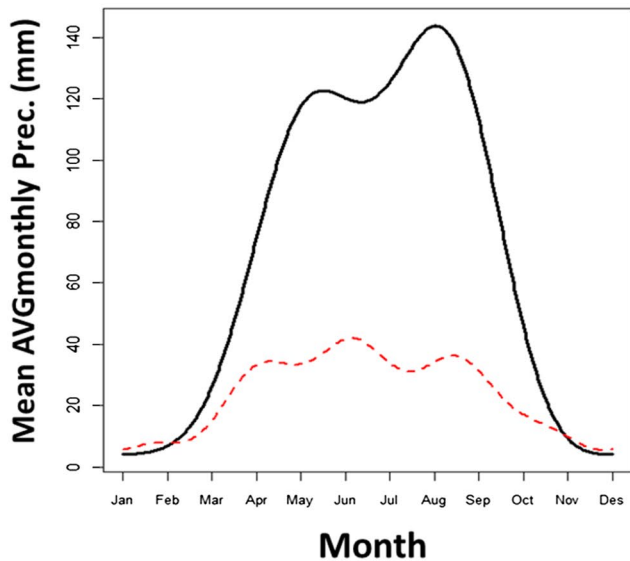
**Fig. 4** The above panel displays the relationship between the GCV criteria and the smoothing level lambda. The below panel displays the relationship between the degree of freedom and the smoothing level lambda



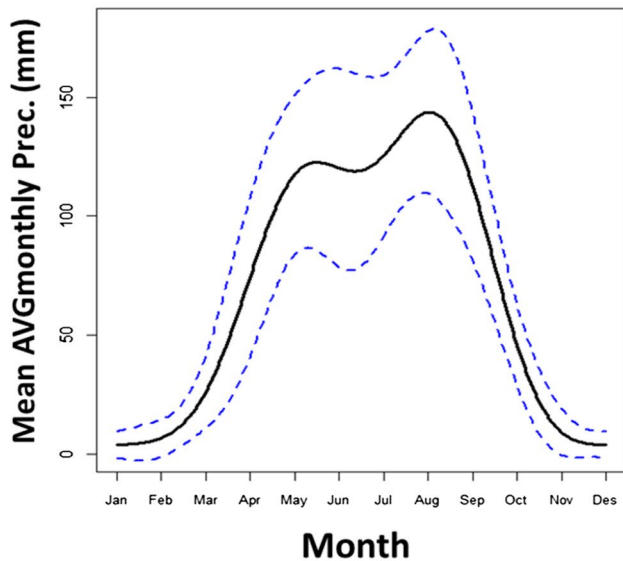
**Fig. 5** Plot of the smoothed functional data for average monthly rainfall using penalized smoothing with GCV; the smoothing parameter is 0.05



**(a) Mean function & Standard Deviation function**



**(b) Confidence Interval for mean function**



**Fig. 6** **a** Plot of the mean and standard deviation curves, **b** plot of confidence interval curves of the mean function

**3.4 Singular value decomposition visualization**

The main feature of the decomposition of a single value is to simultaneously view row information and column data from a two-sided matrix. It is intended to detect local variations and interactions in two-sided matrix years between

rows as months and columns. Moreover, it connects the corresponding curves to the matrix.

Figure 10 displays the singular value decomposition plot of average monthly precipitation over the years 1998–2018. The SVD1 component clearly captures the seasonal patterns, whereas the SVD2 and SVD3 components display the contrasts of average monthly precipitation

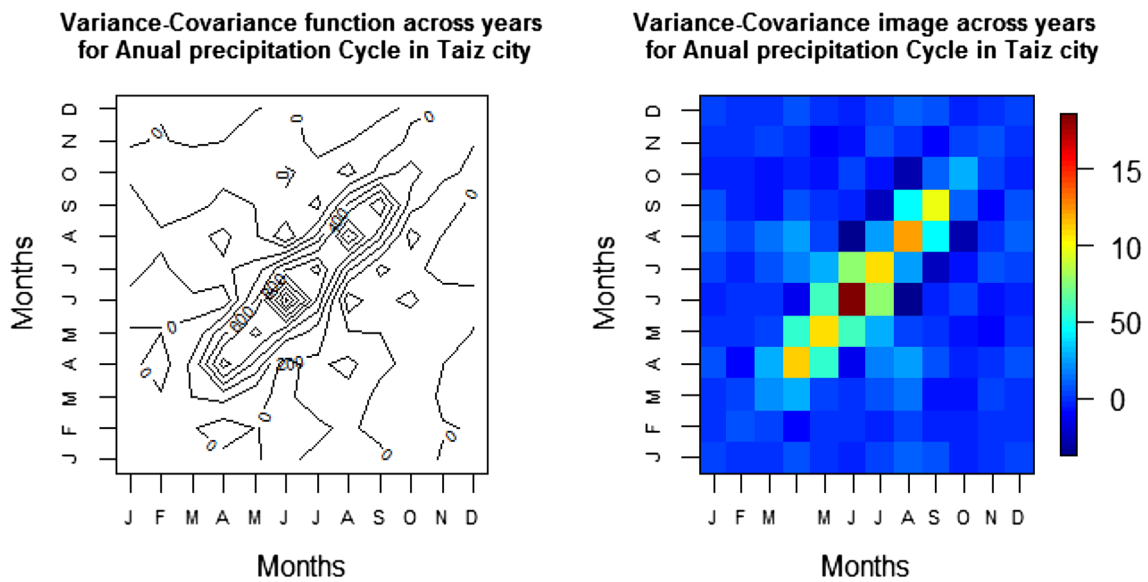


Fig. 7 Contour map on the left and fields plot on the right for estimated covariance surface from the rainfall data across all years

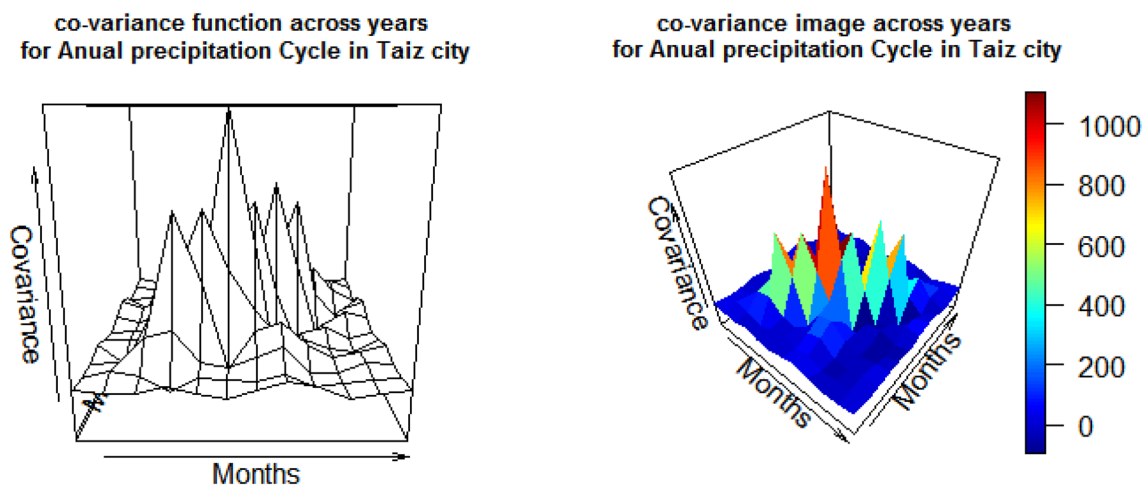


Fig. 8 Plot perspective on the left and heat map on the right for the estimated covariance surface from the rainfall data across all years

among different seasons and months. Residuals panel shows that the residuals are centered on zero as the number of singular value decomposition components increases, as shown in Fig. 10. Since the singular value decomposition visualization leads to the conclusion that there is evidence that the rainfall patterns oscillate, one might be interested in investigating which of the intersections between months and years is different from each other or which of them explains a greater variation in the amount of rainfall. Hence, the singular value decomposition image was used.

Figure 11 displays the singular value decomposition image for the average monthly precipitation of Taiz City between months and over the years 1998–2018. This plot

significantly captures the variability between months and years, with the appropriate adjustment for the degree of color; the black color represents a high variability. Indeed, the local variations and interactions between years and months are highlighted. One finds that rainfall data are clearly visualized by a singular decomposition technique, and variations are shown based on rainfall amounts. The current findings are consistent with the study developed by Shang [27] who applied the singular value decomposition technique on the sea surface temperature dataset, and therefore, these results are considered a basis for researchers to conduct future studies.

In comparison with previous studies, especially similar studies, this study has additional features such as the



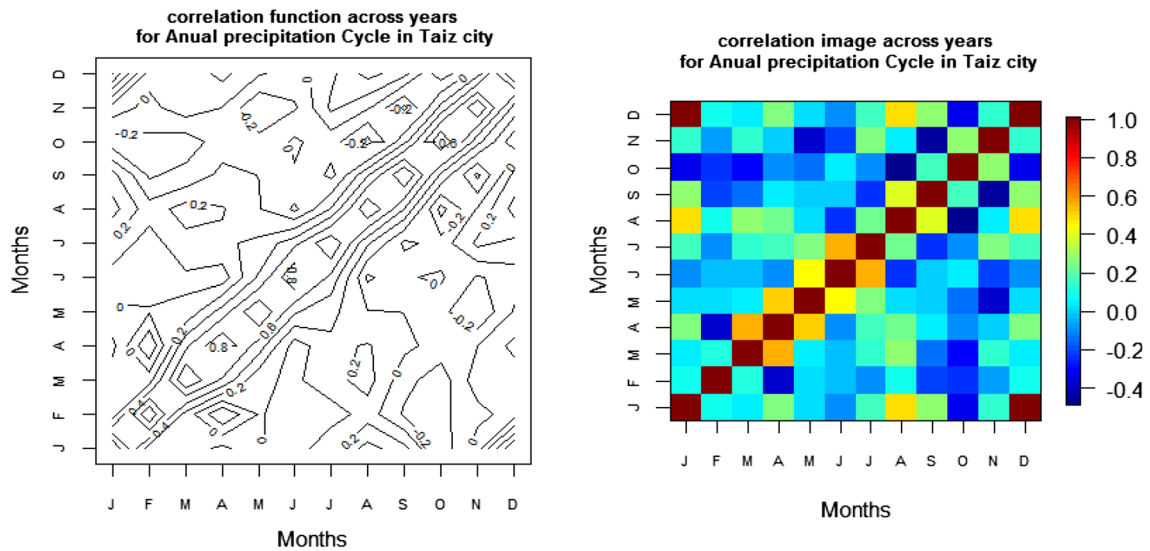


Fig. 9 Contour map on the left and fields plot on the right for correlation surface estimated from the rainfall data across all years

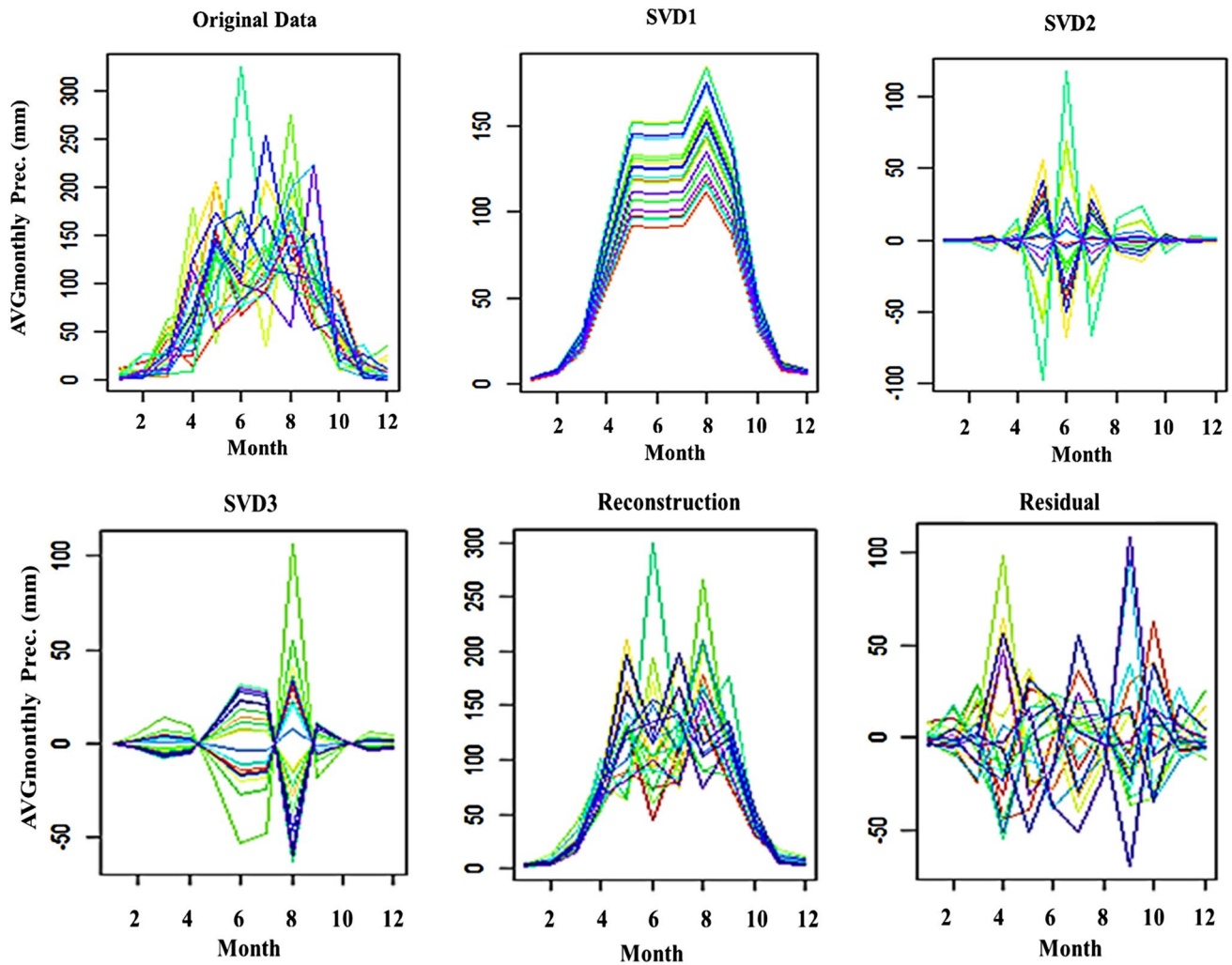


Fig. 10 Plot singular value decomposition for the average monthly rainfall of Taiz city over the years 1998–2018

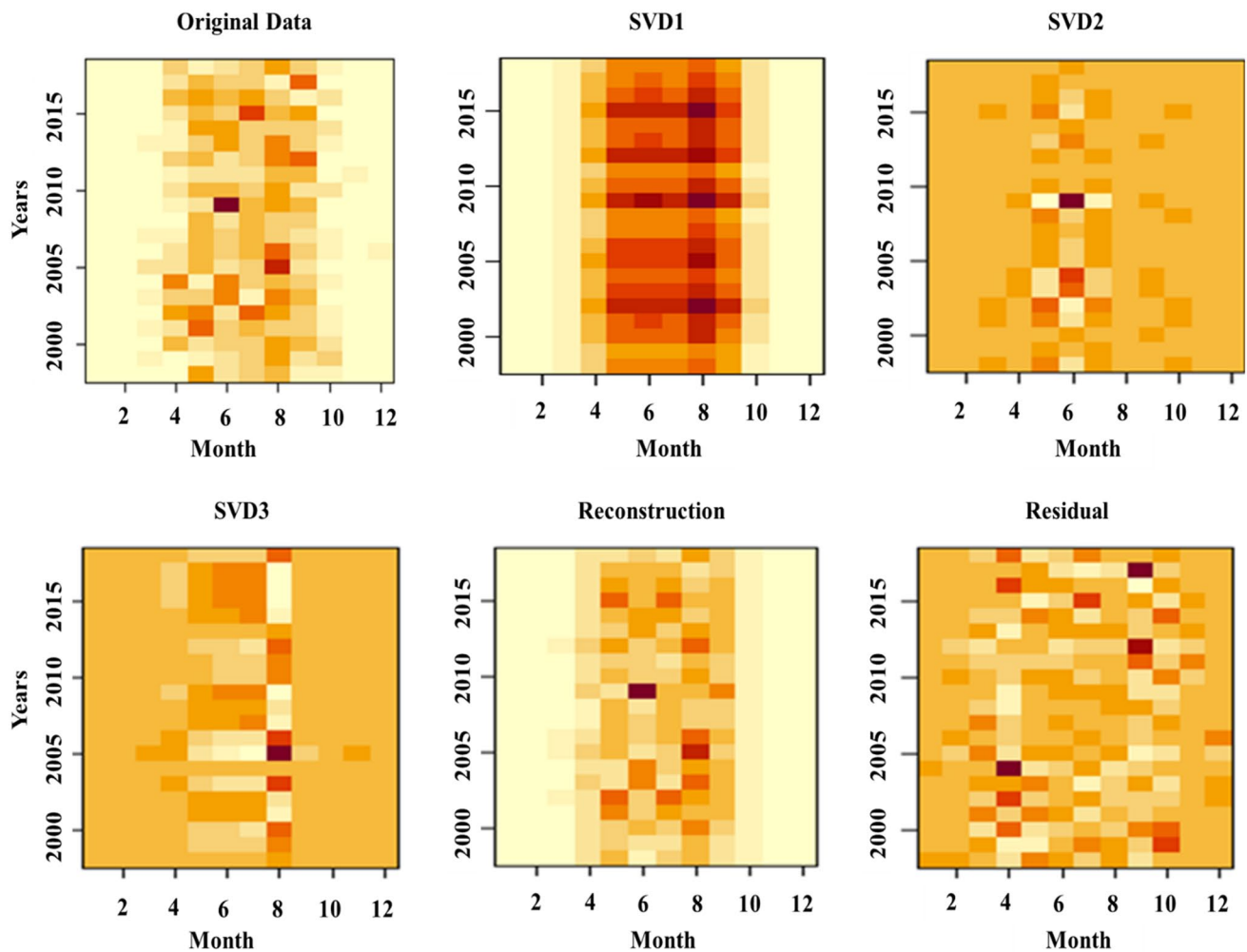


Fig. 11 Singular value decomposition image for the average monthly rainfall between months and over years 1998–2018

fields, heat maps and images as shown in Table 1 that are used for this study. Thus, it can be said that the visualization tools that were performed in this study, such as heat maps, fields and images, made it easier to understand the features of rainfall shapes and determine time variation. Moreover, these technologies have demonstrated the associated patterns of average monthly rainfall in the Taiz Region over the past two decades.

#### 4 Conclusions

This research analyzes and visualizes the average monthly rainfall throughout the last two decades for Taiz City. FDA approaches with emphasis on smoothing and visualization were modified and applied for the rainfall measurements as an important step in a full FDA. This methodology included penalized smoothing with generalized cross-validation criteria, functional parameters such as location

scales, variance–covariance surfaces and correlation functions with visualization and singular value decomposition technique. Based on the results, the following main conclusions can be drawn from this work:

1. The entire rainfall observations were treated by FDA techniques like function data shown by curves which represent the actual phenomena better and display many graphic displays of rainfall data.
2. The penalized GCV smoothing made it easy to choose the best smoothing parameter so that the average monthly rainfall functional data were determined. Therefore, the noise was decreased, and errors were eliminated.
3. More information about the rainfall and distribution of precipitation rates in the Taiz area was provided by means of continuous temporal directions in the location parameters and scale covariates.

**Table 1** Comparison of the methods used in the current study with the previous studies

Study	Applications	Smoothing techniques	Functional statistics	Visualization		Refs.
				Tools	Other techniques	
This research	Rainfall data	Penalized smoothing based on (GCV)	Location scales and confidence interval	Contour maps, perspective plot, fields, images and heat map	Singular value decomposition and images	–
Shang (2019)	Fertility rates	B-spline and spline	–	–	Phase plane plot	[34]
Kendrick et al. (2017)	Calorimetry data	B-spline bases	–	–	–	[35]
Suhaila and Yusop (2017)	Rainfall data	Roughness penalty	Location scales	Contour map and surface	–	[28]
Bellido et al. (2016)	Microbial data	Cubic spline	Location scales	Biplot	–	[36]
Bi and Kuesten (2013)	Sensory time–intensity data	Smoothing spline	Location scales and derivative curves	–	–	[37]
Chebana et al. (2012)	Flood flow data	Fourier bases	Location scales	Surface and contour map	Rainbow plots	[3]
Shang (2011)	Fertility rates and sea surface temperatures	B-spline and spline (GCV)	–	–	Singular value decomposition and rainbow plots	[27]
Suhaila et al. (2011)	Rainfall data	Roughness penalty	–	Scatter plot	–	[38]

- During specific periods, a singular value decomposition technique provided a deeper understanding of the variations in precipitation data over the Taiz Region and clearly showed intersections between years and months.
- Completely functional findings have shown that rainfall rates differ significantly. Generally, the variability was very high during the summer season, high in spring, moderate in autumn and less or none in winter; May and September also represented the most rainfall rate.
- Finally, based on findings, it is recommended to study future research that deals with other functional concepts such as functional principal component analysis, functional clustering, functional classification and functional regression. Furthermore, these techniques should be used in modeling and forecasting of precipitation with other associated variables.
- Ullah S, Finch CF (2013) Applications of functional data analysis: a systematic review. *BMC Med Res Methodol* 13(1):43
- Chebana F, Dabo-Niang S, Ouarda TBMJ (2012) Exploratory functional flood frequency analysis and outlier detection. *Water Resour Res* 48(4):1–20
- Sierra C, Flor-blanco G, Ordoñez C, Flor G, Gallego JR (2017) Analyzing coastal environments by means of functional data analysis. *Sediment Geol* 357:99–108
- Bur R, Neumann C, Saunders CP (2015) Review and application of functional data analysis to chemical data—the example of the comparison, classification, and database search of forensic ink chromatograms. *Chemom Intell Lab Syst* 149:47–106
- Müller H, Sen R, Stadtmüller U (2011) Functional data analysis for volatility. *J Econom* 165(2):233–245
- Beyaztas U, Yaseen ZM (2019) Drought interval simulation using functional data analysis. *J Hydrol* 579:124141
- Aristizabal J, Giraldo R, Mateu J (2019) Analysis of variance for spatially correlated functional data: application to brain data. *Spat Stat* 32:100381
- Lombardi JM, Bortolato SA (2018) Functional data analysis, a new approach to aligning three-way liquid chromatographic with fluorescence detection data. *Microchem J* 142(June):219–228
- Epifanio I, Ventura-Campos N (2011) Functional data analysis in shape analysis. *Comput Stat Data Anal* 55(9):2758–2773
- Dennis EB, Morgan BJT, Fox R, Roy DB, Brereton TM (2019) Functional data analysis of multi-species abundance and occupancy data sets. *Ecol Indic* 104(May):156–165
- Virta J, Li B, Nordhausen K, Oja H (2019) Independent component analysis for multivariate functional data. *J Multivar Anal* 176:104568
- Aitken C, Chang Y, Buzzini P, Zadora G, Massonnet G (2019) The evaluation of evidence for microspectrophotometry data using functional data analysis. *Forensic Sci Int* 305:110007
- Raj M, Whitaker RT (2018) Visualizing multidimensional data with order statistics. *Comput Graph Forum* 37(3):277–287
- El-Galy IM, Bassiouny BI, Ahmed MH (2018) Empirical model for dry sliding wear behavior of centrifugally cast functionally

**Acknowledgements** The authors would like to thank the Yemen Meteorological Authority in appreciation for the help they offered.

## Compliance with ethical standards

**Conflict of interest** We have no conflict of interest.

## References

- Seya H, Yoshida T, Tsutsumi M (2016) Ex-post identification of geographical extent of benefited area by a transportation project: functional data analysis method. *JTRG* 55:1–10

- graded composite based on pure Al/SiCp. *Key Eng Mater* 786:276–285
16. Saleh B, Jiang J, Xu Q et al (2020) Statistical analysis of dry sliding wear process parameters for AZ91 alloy processed by RD-ECAP using response surface methodology. *Met Mater Int.* <https://doi.org/10.1007/s12540-020-00624-w>
  17. Wang J, Chiou J, Muller H-G (2015) Functional data analysis. *Ann Rev Stat Appl* 3:257–295
  18. Aneiros G, Cao R, Fraiman R, Genest C, Vieu P (2018) Recent advances in functional data analysis and high-dimensional statistics. *J Multivar Anal* 170:3–9
  19. Ramsay AJO, Dalzell CJ (1991) Some tools for functional data analysis. *J Source Stat R Ser Soc* 53(3):539–572
  20. Ramsay JO, Silverman B (2005) *Functional data analysis*. Springer, New York. <https://doi.org/10.1007/b98888>
  21. Ferraty F, Vieu P (2006) *Nonparametric functional data analysis: theory and practice*. Springer, New York. <https://doi.org/10.1007/0-387-36620-2>
  22. Kokoszka P, Reimherr M (2017) *Introduction to functional data analysis*. Chapman and Hall/CRC, New York. <https://doi.org/10.1201/9781315117416>
  23. Laukaitis A (2008) Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes. *Eur J Oper Res* 185(2008):1607–1614
  24. Hyndman RJ, Shang HL (2010) Rainbow plots, bagplots, and boxplots for functional data. *J Comput Graph Stat* 19(1):37–41
  25. Ikeda T, Dowd M, Martin JL (2008) Application of functional data analysis to investigate seasonal progression with interannual variability in plankton abundance in the Bay of Fundy, Canada. *Estuar Coast Shelf Sci* 78(2):445–455
  26. Torres JM, Nieto PJG, Alejano L, Reyes AN (2011) Detection of outliers in gas emissions from urban areas using functional data analysis. *J Hazard Mater* 186:144–149
  27. Shang HL (2011) Rainbow: an R package for visualizing functional time series. *R J* 2(3):54–59
  28. Suhaila J, Yusop Z (2017) Spatial and temporal variabilities of rainfall data using functional data analysis. *Theor Appl Climatol* 129(1–2):229–242
  29. Zhang L, Marron JS, Shen H, Zhu Z (2007) Singular value decomposition and its visualization. *J Comput Graph Stat* 16(4):833–854
  30. Alharazi T, Alasbahi I, Saif N (2016) Study on cutaneous leishmaniasis among clinically suspected patients in Taiz City, Taiz Governorate, Yemen. *Int J Trop Dis Health* 19(January):1–5
  31. Taiz Monthly Climate Averages (2020) <https://www.worldweatheronline.com/lang/en-us/taiz-weather-averages/dhama/ye.aspx>. Accessed 20 Jan 2020
  32. Ramsay J, Hooker G, Graves S (2009) *Functional data analysis with R and MATLAB*. Springer, Berlin
  33. Shang HL, Hyndman R (2020) Package ‘rainbow’, UTC. <https://cran.r-project.org/web/packages/rainbow/rainbow.pdf>. Accessed 20 Jan 2020
  34. Shang HL (2019) Visualizing rate of change: an application to age-specific fertility rates. *J R Stat Soc A* 182(1):249–262
  35. Kendrick SK, Zheng Q, Garbett NC, Brock GN (2017) Application and interpretation of functional data analysis techniques to differential scanning calorimetry data from lupus patients. *PLoS ONE* 12(11):1–21
  36. Ruiz-Bellido MA, Romero-Gil V, García-García P, Rodríguez-Gómez F, Arroyo-López FN, Garrido-Fernández A (2016) Assessment of table olive fermentation by functional data analysis. *Int J Food Microbiol* 238:1–6
  37. Bi Jian, Kuesten Carla (2013) Using functional data analysis (FDA) methodology and the R package ‘fda’ for sensory time-intensity evaluation. *J Sens Stud* 28(6):474–482
  38. Suhaila J, Jemain AA, Hamdan MF, Wan Zin WZ (2011) Comparing rainfall patterns between regions in Peninsular Malaysia via a functional data analysis technique. *J Hydrol* 411(3–4):197–206

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.