World Environment 2018, 8(1): 1-14
DOI: 10.5923/j.env.20180801.01

Forecasting of Rainfall in Pakistan via Sliced Functional Times Series (SFTS)

Farah Yasmeen*, Shaheen Hameed

Department of Statistics, University of Karachi, Pakistan

Abstract In hydrological and climatological time series described over a varied range of time scales, the persistence is generally recognized. Among them, an important climatological variable is the amount of rainfall. Clearly, the rainfall analysis has a substantial role in the successful planning, development and implementation of water resource management to evaluate engineering projects and environmental problems. They include hydropower generation, reservoir operation, flood control and control of water quality. Therefore, an efficient study of the temporal rainfall behavior is considered to be critically important in hydrology. In this paper, a study is conducted across the country to model the rainfall trend in Pakistan over the past six decades. For this purpose, secondary dataset of average rainfall comprising 65 years for the period 1951 to 2015, are acquired from the World Bank website (www. http://sdwebx.worldbank.org). In Pakistan, adverse consequences of rainfall have already been observed. They are in the form of droughts and super floods which have badly affected human settlements, water management and agriculture. In this study, the data are analyzed through a sliced functional time series model, a relatively new method of forecasting. The results show a decreasing trend in average rainfall over the country. The monthly forecasts for the next ten years (2016-2025) are obtained along with 80% prediction intervals. These forecasts are also compared with the forecasts obtained from ARIMA and exponential smoothing state space (ETS) models.

Keywords Rainfall trend, Seasonality, Functional data analysis, Sliced functional time series, Forecasting, Forecast accuracy

1. Introduction

Rainfall is important for food production plan, water resource management and all activity plans in the nature. The occurrence of prolonged dry period or heavy rains at the critical stages of the crop growth and development may lead to significant reduction in crop yield. Pakistan is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural countries like Pakistan. Rainfall forecasting has been one of the most scientifically and technologically challenging problems around the world in the last century.

Rainfall prediction modeling involves a combination of probabilistic models, observation and knowledge of trends and patterns. Using these methods, reasonably accurate forecasts can be made up. Several recent research studies have developed rainfall prediction using various weather and climate forecasting techniques. Numerous studies have been conducted about rainfall forecasting in Pakistan and other countries.

* Corresponding author:
riazfarah@yahoo.com (Farah Yasmeen)
Published online at http://journal.sapub.org/env

Afsar et al. (2013) made a comparative study of temperature and rainfall fluctuation in Hunza Nagar District to develop and validate a forecast model that could predict temperature and rainfall. Regression and Time Series Stochastic modeling methods were used. It was observed that the rainfall increased with increasing temperature and that the trend in monthly mean max temperature indices increased from 2007 to 2011 while the amount of rainfall decreased. AR(1) model was found to be the most adequate to forecast temperature.

Yamoah et al. (2016) examined the rainfall pattern and fitted a suitable model for rainfall prediction in the Brong Ahafo (BA) region of Ghana for the period of 1975 to 2009. SARIMA $(0,0,0)\times(1,1,1)_{12}$ was identified as an appropriate model for predicting monthly average rainfall figures for the region. Yusof and Kane (2012) studied the rainfall time series of two selected weather stations in Malaysia. Seasonal ARIMA and ETS models were used and both found to be adequate for forecasting. The authors also evaluated the spatial autocorrelation of rainfall data in some locations of Peninsular Malaysia using geostatistical technique.

Ahmed et al (2015) investigated precipitation variability across 15 stations in the Swat River basin, Pakistan, for the period 1961–2011. The nonparametric Mann-Kendall (MK) and Spearman's rho (SR) tests were used to detect trends in monthly, seasonal, and annual precipitation, and the

trend-free pre-whitening approach was applied to eliminate serial correlation in the precipitation series. The results highlighted a mix of positive and negative trends in monthly, seasonal, and annual precipitation. The Saidu Sharif station revealed the highest positive trend (7.48 mm/year) in annual precipitation.

Faisal and Ghaffar (2012) determined an area weighted rainfall data set of Pakistan. Monthly rainfall data of 56 climatological stations, rain gauges for the period of 50 years (1961-2010) were used and a time series of weighted rainfall was developed using the Theissen Polygon Method. Kane and Yusof (2013) studied the persistence dependence of rainfall time series of Chui Chak, a station in peninsular Malaysia that observed the highest rainfall event for the period 1975-2008. The persistence dependence of rainfall time series modeled via fractional ARIMA model was added by the GARCH model. The Ljung –Box test for testing the autocorrelation showed that the combined ARFIMA-GARCH model captured well the temporal persistence behavior in the Chui Chak rainfall time series data.

Kambezidis et al. (2010) studied the spatial and temporal variability of the mean annual rain intensity in Greece during a 41-year period (1962–2002). In their study, the meteorological datasets concerned monthly rain amounts (in mm) and the respective monthly durations (in hrs) recorded at 32 meteorological stations, distributed uniformly on Greek territory. The spatial distribution of the mean annual rain intensity was studied using the Kriging interpolation method. The temporal variability concerning the mean annual rain intensity trends, along with their significance test was analysed using Mann-Kendall. The trend analysis revealed that statistically significant positive trends of the mean annual rain intensity appear in the wider area of Athens, the northern coastal areas of Crete Island, the Cyclades complex and the north-eastern sub-regions of Greece.

Sakellarian and Kambezidis (2003) investigated the total precipitation in the Athens area, Greece, using annual precipitation patterns averaged for five, ten and fifteen days. In another study, Sakellarian and Kambezidis (2004) obtained the prediction of the total rainfall amount during August and November in the Athens.

Menabde and Sivapalan (2000) proposed a new model for simulating rainfall time series, which showed that the intensity and duration of individual rainfall events can best be modeled by a Fat-tailed Levy-stable distribution. In contrast, a model based on a gamma distribution for rainfall intensity substantially underestimates the absolute values of extreme events and does not correctly reproduce their scaling behavior.

Muslehuddin and Faisal (2006) examined the relationship of Sindh monsoon rainfall with some of the important global and regional parameters. The Sindh Monsoon Rainfall Index (SMRI) was examined with the monthly mean values of SST (Sea Surface Temperature), IHP (Indian Ocean High Pressure), SOI (Southern Oscillation Index), NHT (Northern Hemisphere Temperature) and Pakistan regional data. The Sea Surface Temperature (SST) consisted of monthly mean temperature value for grid points in each 5 degree latitude and 5 degree longitude area with large regions Nino 1+2 (0-10S)(90W-80W) & Nino 4 (5N-5S) (160E-150W) for the period 1957 to 2003. The Southern Oscillation Index (SOI) was the monthly value (anomaly) of the difference in mean sea level pressure (MSL) between Tahiti (18°S, 15°W) and Darwin (12°S, 131°E). IHP represented the seasonal values of Indian Ocean High Pressure, whereas, the Northen Hemisphere Temperature (NHT) was the monthly mean temperature values of Northern Hemisphere. The rainfall data included of monthly total rainfall values (in mm) of eight different stations of Sindh for the years 1957-2003. These stations included Badin, Chhor, Hyderabad, Jacobabad, Karachi, Nawabshah, Padidan and Rohri. The data were obtained from the Computerized Data Processing Centre (CDPC) of Pakistan Meteorological Department (PMD). It was found that the SMRI is significantly correlated with SOI, IHP, mean temperatures of Baluchistan and Punjab provinces. A multiple regression equation was developed and applied in the period 1996-2003 to verify the results. In another study, Salma and Rehan (2012) studied the rainfall trend in different climate zones of Pakistan over the past three decades. The data were analyzed through ANOVA along Dunnet T3 and showed a decreasing trend (1.18mm/decade) all over the country.

Villarini, G. et al. (2010) modeled a record (1862–2004) of seasonal rainfall and temperature from the Rome observatory of Collegio Romano in a non-stationary framework by means of the Generalized Additive Models in Location, Scale and Shape (GAMLSS). The models were used to characterize non-stationarities in rainfall and related climate variables. In particular, the North Atlantic Oscillation is a significant predictor during the winter, while the Mediterranean Index is a significant predictor for almost all seasons.

Yusof et al. (2013) employed a procedure for estimating the fractional differencing parameter in semi-parametric contexts proposed by Geweke and Porter-Hudak (1983) to analyze nine daily rainfall data sets across Malaysia. The results indicating that all the data sets exhibit long memory. Furthermore, an empirical fluctuation process using the ordinary least square (OLS)-based cumulative sum (CUSUM) test for the break date was applied, however; the analysis showed a true long memory not due to structural break.

Zahid and Rasul (2011) analyzed the frequency of extreme temperature and precipitation events in Pakistan for the period 1965-2009. The statistical significance of the temperature extreme events was evaluated through F-Test and for the precipitation extreme events, the K-S Test was applied at 95% confidence interval. The extreme temperature events analysis showed that the frequency of maximum temperature events is increasing throughout the country.

Abbot and Marohasy (2012) studied the application of artificial intelligence to monthly and seasonal rainfall forecasting in Queensland, Australia. The data were assessed

by inputting recognized climate indices, monthly historical rainfall data, and atmospheric temperatures into a prototype stand-alone, dynamic recurrent time-delay artificial neural network. Outputs, as monthly rainfall forecasts 3 months ahead in the period 1993 to 2009, were compared with observed rainfall data using time-series plots, root mean squared error (RMSE) and Pearson correlation coefficients.

Nayak et al. (2013) provided a survey of available literature of some methodologies employed by different researchers to utilize Artificial Neural Network for rainfall prediction and found that rainfall prediction by using an Artificial Neural Network technique is more suitable than the traditional and numerical methods.

Bilgili and Sahin (2010) applied artificial neural networks to predict the long-term monthly temperatures and rainfalls at any target point of Turkey based on the observations from the neighboring measuring stations. Finally, the values determined by the artificial neural network model were compared with the actual data. Errors obtained in this model are well within acceptable limits.

This paper describes empirical method technique belongs to the functional time series approach which tries to make a short-term forecast of rainfalls over specified region of Pakistan. The paper is organized as follows. Section 1 is introductory and the relevant literature is reviewed. In section 2, we described the data and statistical methodology, whereas the results are presented in section 3. Finally some concluding remarks are given in section 4.

2. Material and Methods

Pakistan is situated in Asia between 23.30 degrees and 36.45 degrees Latitude (N) and 61 degrees and 75.45 degrees Longitude(E). Pakistan is bounded by the three world famous mountainous ranges, which play an important role for summer and winter precipitation. In the northwest lies the Hindukash Range, in north lies the central Karakaram Range and in northeast lies the Himalayan regions of Pakistan. The country has four seasons

i. Summer (May to mid-September)
ii. Autumn (Late September to November)
iii. Winter (December to February)
iv. Spring (March, April)

Climatic Regions of Pakistan

The climate of Pakistan on the whole is dry and extreme; summers are extremely hot and winters are extremely cold with little rainfall during the year. The climate varies from place to place, therefore, Pakistan may be divided into the following four distinct zones:

i. The North and North Western Mountainous Area

This Region consists of the North and the North-Western Mountainous areas. This region has a very severe winter and the temperature falls below freezing point. In this area the winter season extents from six to eight months. On the other hand, summers in this region are very pleasant.

ii. The Upper Indus Plain

Below the Northern Mountainous Area is the upper Indus Plain. In this area the summer is very hot. The months of May, June, and the first week of July are very hot because of absence of rainfall. However, the climate here becomes pleasant when rain falls in July. The winter season is very pleasant but it not last long.

iii. The Coastal Areas and Lower Indus Valley

The temperature in the coastal areas and the lower Indus Valley does not rise due to sea land breeze. In this region rain is scarce; however, during sea breeze conditions, humidity is found in the air. Sea breeze keeps the climate pleasant. There is not much difference in the temperature among the different months.

iv. The Plateau of Baluchistan and the Thar Desert

In summer, the temperature in the Plateau of Baluchistan and the Thar Desert rises and most of the Mountainous regions of Baluchistan are dry and hot. The winter season is very severe in Baluchistan and sometimes snow falls in certain parts. Pakistan is basically an agricultural country as most of its economy depends upon agriculture. The country has developed the world's largest contagious canal network. Monsoon precipitation is the lifeline of Pakistan which not only caters the national power supply and standing crops water demands but help gathers the reserve to meet the requirement of low flow period in the next 4-5 months. Rainfall plays an important role in agriculture so early prediction of rainfall is necessary for the better economic growth of the country. Prediction of rainfall contributes to crops development and harvesting sufficiency in water supply and water resource policy.

The climate of Pakistan is generally characterized by hot summers and cool or cold winters. Also it has wide variations between temperature extremes at certain locations. As said earlier, Pakistan has four seasons: a cool and dry winter from December through February; a hot and dry spring from March through May; the summer/rainy season from June through September; and the retreating monsoon period of October and November. The duration of these seasons vary somewhat according to location. Generally, there is little rainfall all over the country. Half of the annual rainfall occurs in July and August, while the remainder of the year has significantly less rain.

Figure 1 depicts the map of Pakistan with several regions according to rainfall (Source: http://www.globalcitymap.com/Pakistan). From this map, it is clear that most of the areas have little rain, and heavy rain occurs mostly in the Northern areas. Some parts of the provinces Sindh and Punjab and a huge part of Baluchistan are extremely dry with average rainfall of below 5 inch units (127 mm).

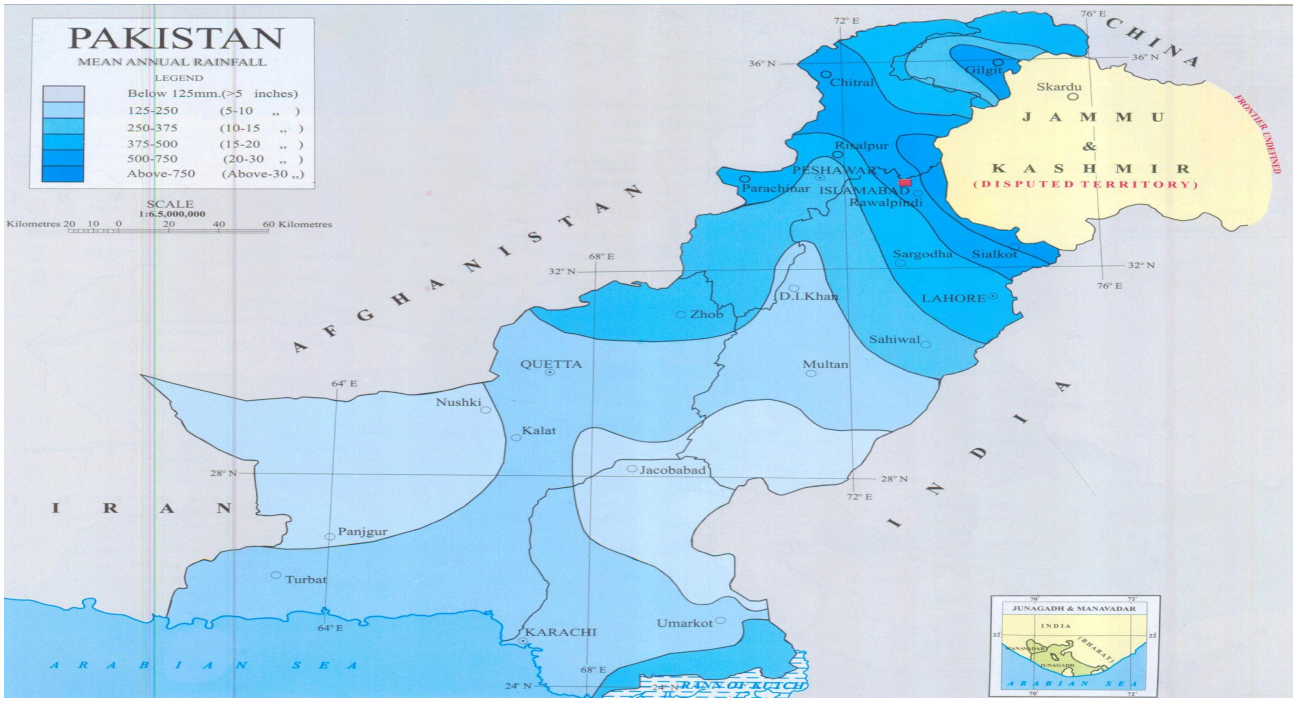

Figure 1. Map of Pakistan with different regions according to rainfall (Source: http://www.globalcitymap.com/pakistan)

In this research, secondary dataset of average rainfall comprising 115 years (1901 to 2015) were acquired from the World Bank website (www. http://sdwebx.worldbank.org). The data are transformed as sliced functional time series (SFTS) then several time series models were applied to forecast the average rainfall in Pakistan. The techniques used are described below:

Sliced Functional Time Series (SFTS) and Outlier Detection

The functional time series (FTS) models were first introduced by Hyndman and Ullah (2007). Initially, the aim was to model the log mortality and fertility rates. Using this method, the functional curves are observed (with error) at time t = 1,…,m, and we wish to forecast the functions for times t = m+1, …m+h. Let $[g_t(x_j)]$ denote the observed data, where j = 1,…p. We assume that there are underlying L_1continuous and smooth functions $[f_t(x)]$ such that

$$g_t(x_j) = f_t(x_j) + \delta_t(x_j)e_{i,j} \qquad (1)$$

where $[e_{i,j}]$ are independent and identically distributed random variables with zero mean and unit variance, and $\delta_t(x_j)$ allows for heteroskedasticity.

The Hyndman and Ullah (2007) technique uses non parametric smoothing on each curve $g_t(x)$ separately to obtain estimates of the smooth functions $[f_t(x)]$. Panelized regression splines (Wood 1994) are used for smoothing, and then a functional principal component approach (Ramsay and Silverman 2005) is used to decompose the time series of functional data into a number of principal components and their scores. The functional time series (FTS) model can be written as follows:

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \alpha_{t,k}\, \Psi k(x) + e_t(x) \qquad (2)$$

where $\Psi_k(x)$ is the k^{th} principal component, the set of coefficients $[\alpha_{1,k}, \ldots. \alpha_{m,k}]$ are the corresponding scores, $e_t(x)$ denotes the independent and identically distributed random functions with zero mean, and *K is the* number of principal components to be used.

Hyndman and Shang (2009) proposed three new graphical methods for plotting the functional time series. They include the rainbow plot, the functional bagplot and the functional highest density region (HDR) boxplot. A side benefit of their approach is the identification of outliers, which may not be obvious from the plot of the original data. These outliers are either magnitude outliers (i.e the curves lie outside the range of the vast majority of the data), or they may be the shape outliers (the curves that are within the range of the rest of the data) but they have different shape from other curves. It is also possible that the curves may exhibit a combination of these two features. The presence of the outliers may have serious effect on the modeling and forecasting series.

To detect the outliers from a functional time series, the first step is to obtain the functional curves, then the data are transformed into sliced functional time series (SFTS). For this, the entire data are sliced for each year as a function of 12 months. These curves are plotted in rainbow order with red for the earlier years and violet for the most recent years. The functional curves are then projected into a finite dimensional subspace. The subspace R^2 is chosen for simplicity. Based on Tukey's (1974) halfspace bagplot and Hyndman's (1996) HDR boxplot, each of the functional data point in R^2 are ordered by data depth and data density. Those curves that have either lowest depth or lowest density are considered to be the outliers (the reader is referred to Hyndman and Shang (2009) for details).

Functional Bagplot

The functional bagplot is based on the bivariate bagplot of Rousseeuw et al. (1999), applied to the first two principal component scores. It uses Tukey's (1974) halfspace location depths. The depth region R_k is the set of all θ, with $r(\theta,z) \geq k$. Since the depth regions form a series of convex hulls, we have $R_{k1} \subset R_{k2}$ for $k_2 > k_1$. The Tukey bivariate depth median is defined as the value of θ, which minimizes $r(\theta,Z)$ if there is such a unique θ, otherwise it is defined as the center of gravity of the deepest region.

Functional HDR boxplot

The functional HDR boxplot is based on the bivarate HDR boxplot (Hyndman, 1996), which is applied to the first two principal component scores. The bivariate HDR boxplot is constructed using a bivariate kernel density estimate f(z), which is defined as

$$f(z) = 1/n \sum_{i=1}^{n} k_{hi}(z - Z_i), \quad (3)$$

where Z_i represents a set of bivariate points, $K_{hi}(\cdot) = K(\cdot/hi)/hi$, K is the kernel function, and h_i is the bandwidth for the ith dimension. The bandwidths were selected using smoothed cross validation (Duong and Hazelton, 2005). Using the kernel density estimates, a HDR is defined as

$$R_\alpha = \{z: f(z) \geq f_\alpha\}, \quad (4)$$

where f_α is such that $\int_{R\alpha} f(z)dz = 1-\alpha$; that is, it is the region with probability of coverage 1- α, where all points within the region have a higher density estimate than any of the points outside the region, hence the name "highest density region".

3. Results and Discussion

Figures 2 and 3 represent the daily and the monthly plots of the series which show a clear seasonal pattern. The precipitations are higher for July and August but relatively less for the other months. It can also be observed that the amount of rainfall is continuously decreasing in the recent years, especially after 1980. In order to see a clearer picture of the series, this is divided into the groups of 30 years, as shown in Figure 4.

It is clear that the amount of rainfall was higher during 1901-1960, even for the months of Jan-March but not that high amounts in these months since last 50 years. In order to select a better forecasting time series model, we consider the data from 1951 to 2015. After the independence of Pakistan in 1947, most of the jungles were cut down; hence this may one of the reasons to the decline in rainfall.

Figure 5 depicts the amount of rainfall during (1951-2015). The months of July and August have heavy rain due to the monsoon whereas the other months are having relatively low precipitations. There is little rain in the months of May, June, October and November in northern areas with about no rain in Sind and Baluchistan in these four months. Hence the amount of average rainfall is lower in these four months while October is considered to be the driest month in Pakistan.

Next, the data are transformed into sliced functional time series. The first step is to obtain the functional curves. For this, the entire data are sliced for each year as a function of 12 months, as plotted in Figure 6. These curves are plotted in rainbow order with red for the earlier years and violet for the most recent year. We use R statistical software, with R package rainbow (Hyndman and Shang 2009) for this plot.

The functional curves are then projected into a finite dimensional subspace. The subspace R^2 is chosen for simplicity. Based on Tukey's (1974) halfspace bagplot and Hyndman's (1996) HDR boxplot, each of the functional data point in R^2 is characterized by data depth and data density. Those curves that have either lowest depth or lowest density are considered to be the outliers (the reader is referred to Hyndman and Shang (2009) for details).

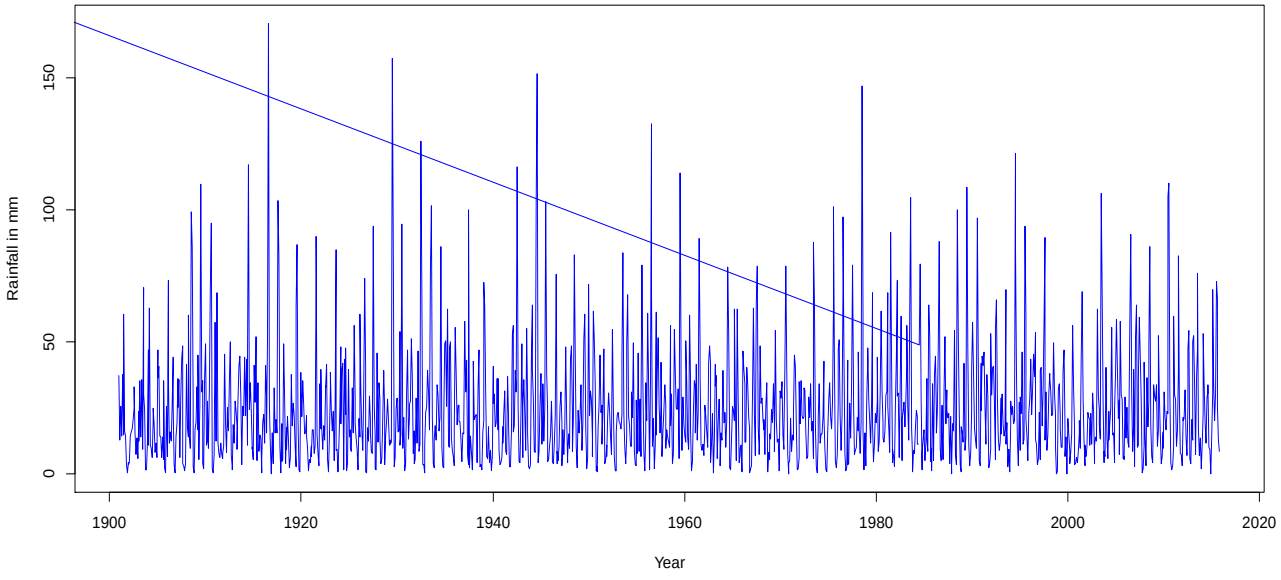

Figure 2. Plot of monthly rainfall in different areas of Pakistan during 1901-2015

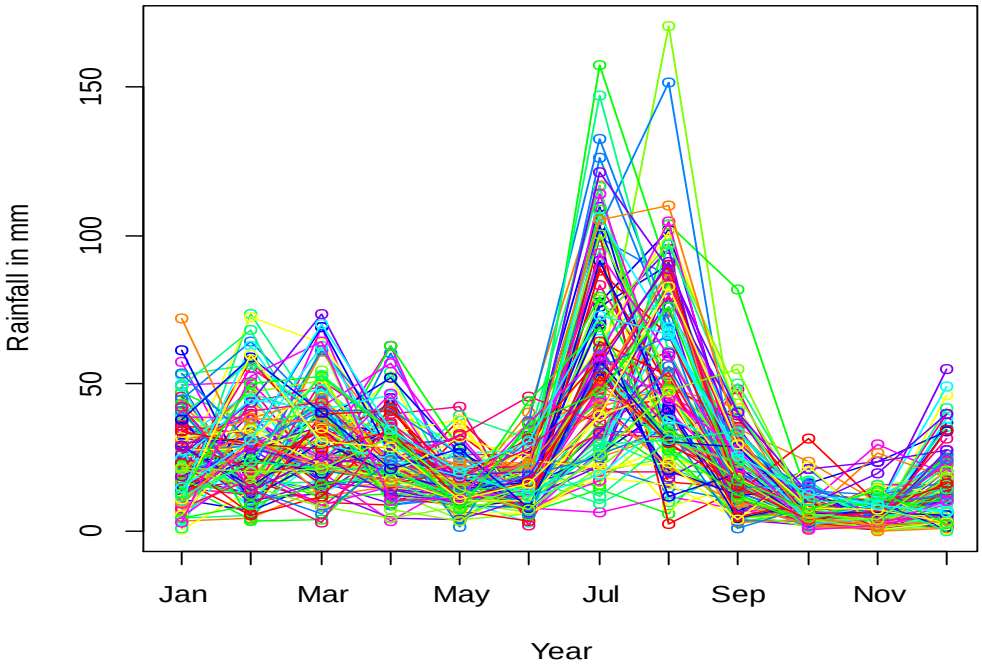

Figure 3. Plot of monthly rainfall in Pakistan during 1901-2015. These are 115 curves, each representing a single year

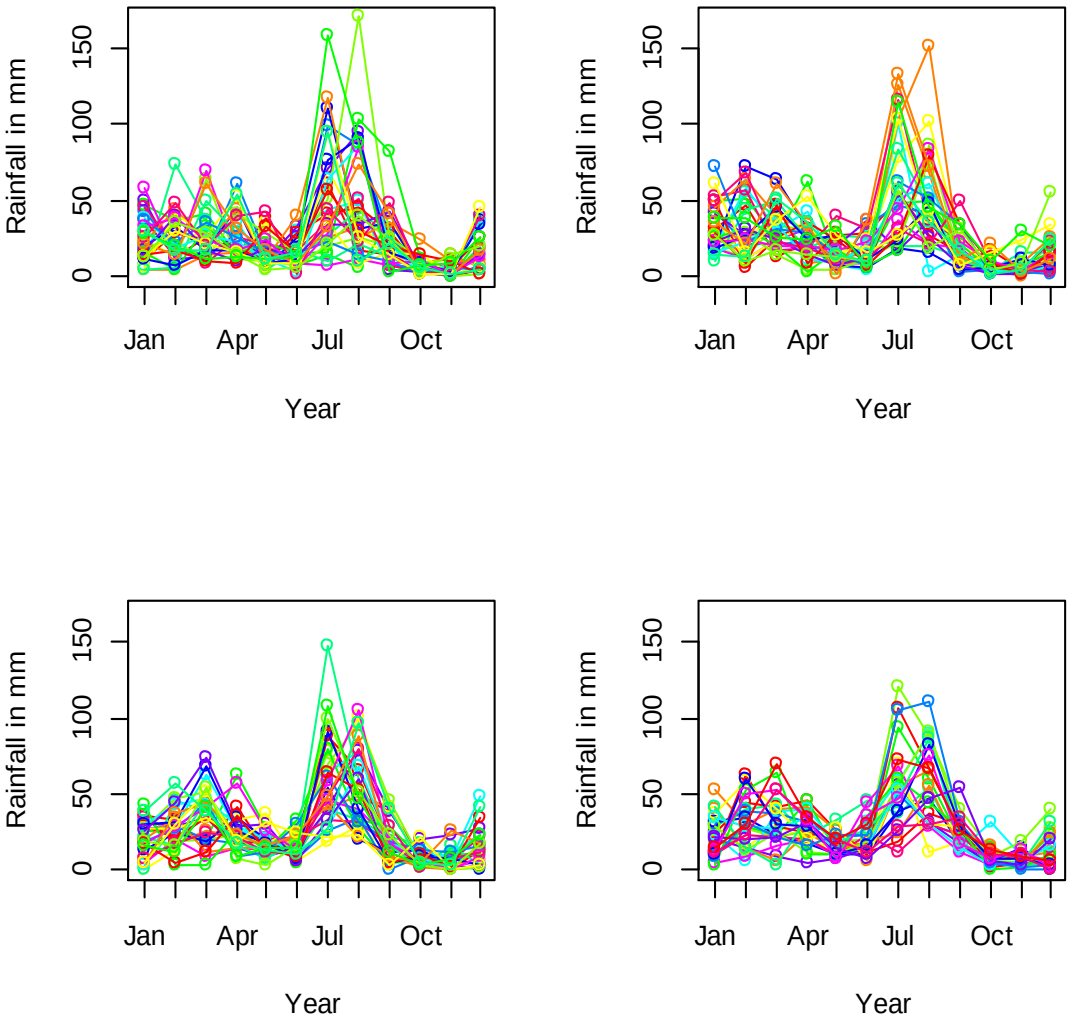

Figure 4. Season plots for rainfall during (1901-1930), (1931-1960), (1961-1990) and (1991-2015)

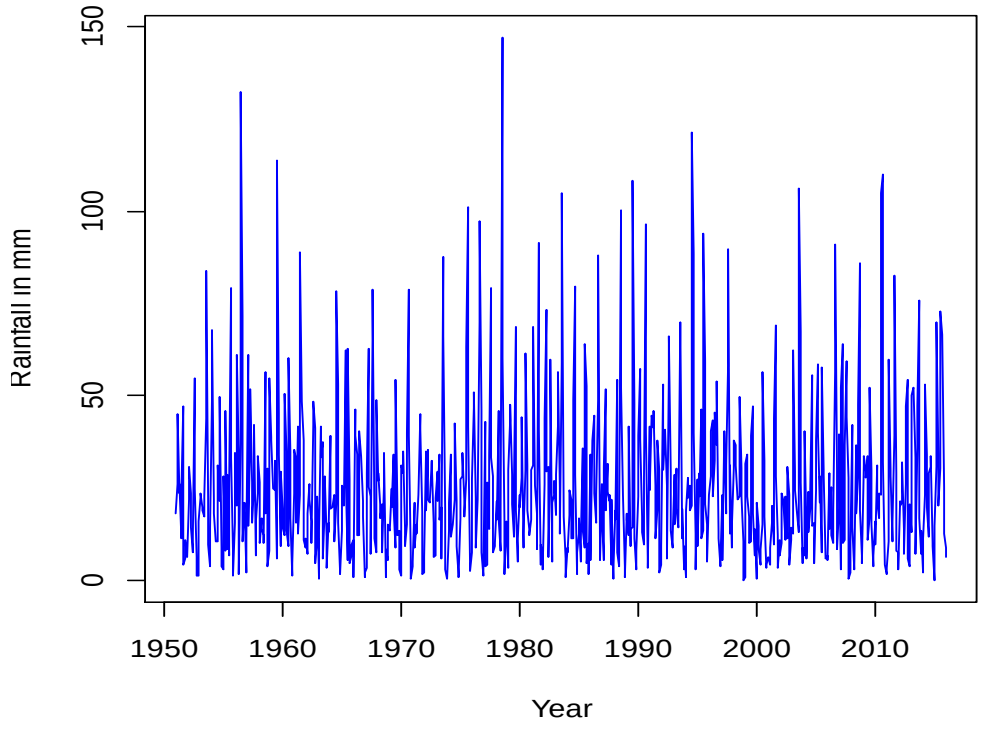

Figure 5. As in Fig. 2, but during 1951-2015

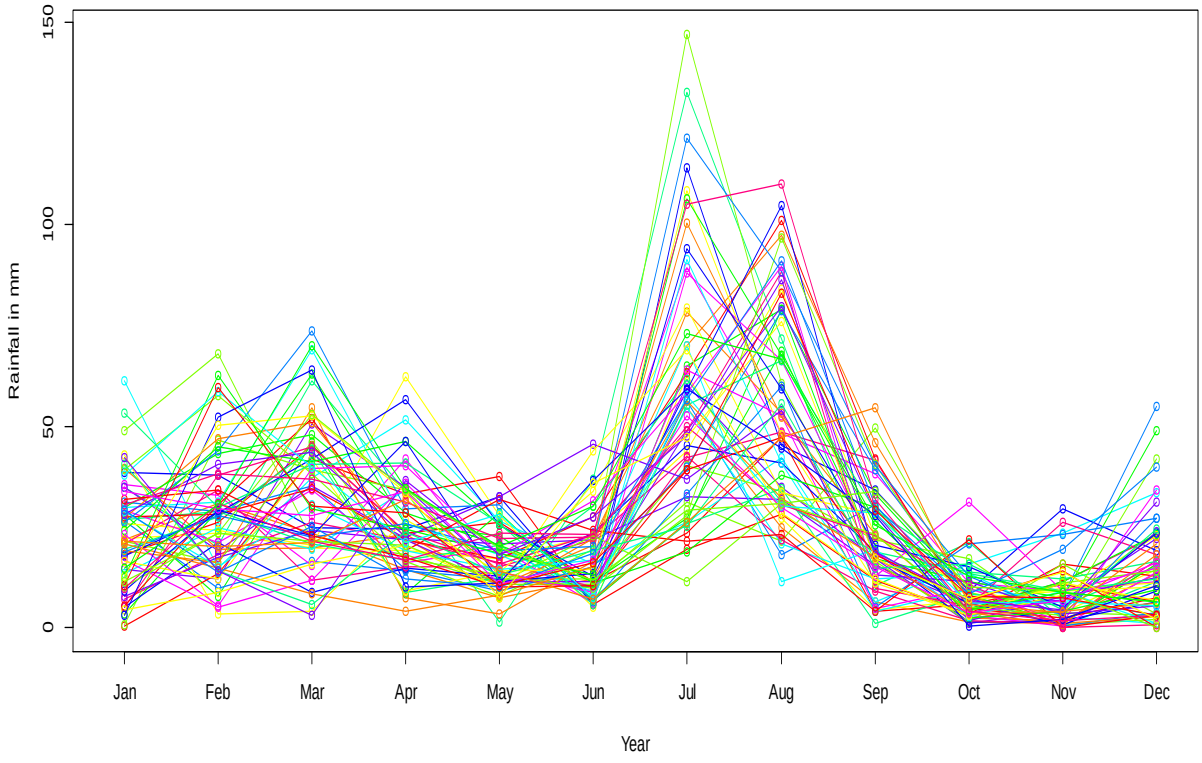

Figure 6. Plot of average rainfall during (1951-2015) as Sliced Functional Time Series. The curves for each of the 65 years are plotted in the colors of a rainbow, as function of the month

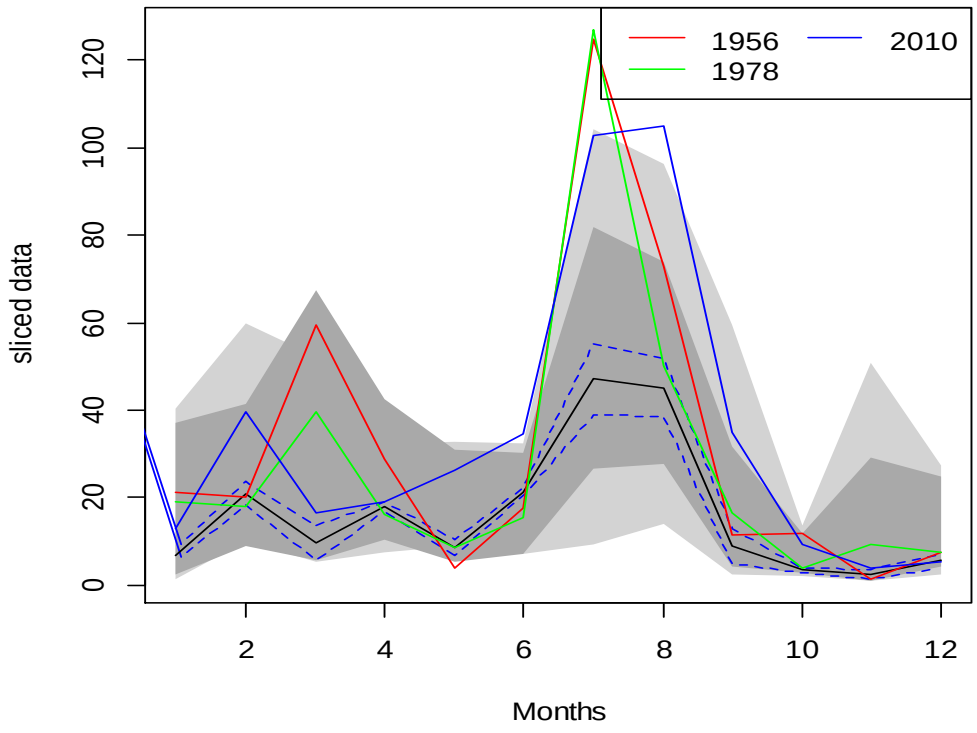

Figure 7. Functional bagplot for rainfall data. Median curve is denoted by black color, along with its confidence interval (blue dotted lines). The outliers are represented by red, green and blue colors. Inner and outer regions are plotted in dark grey and light grey color respectively

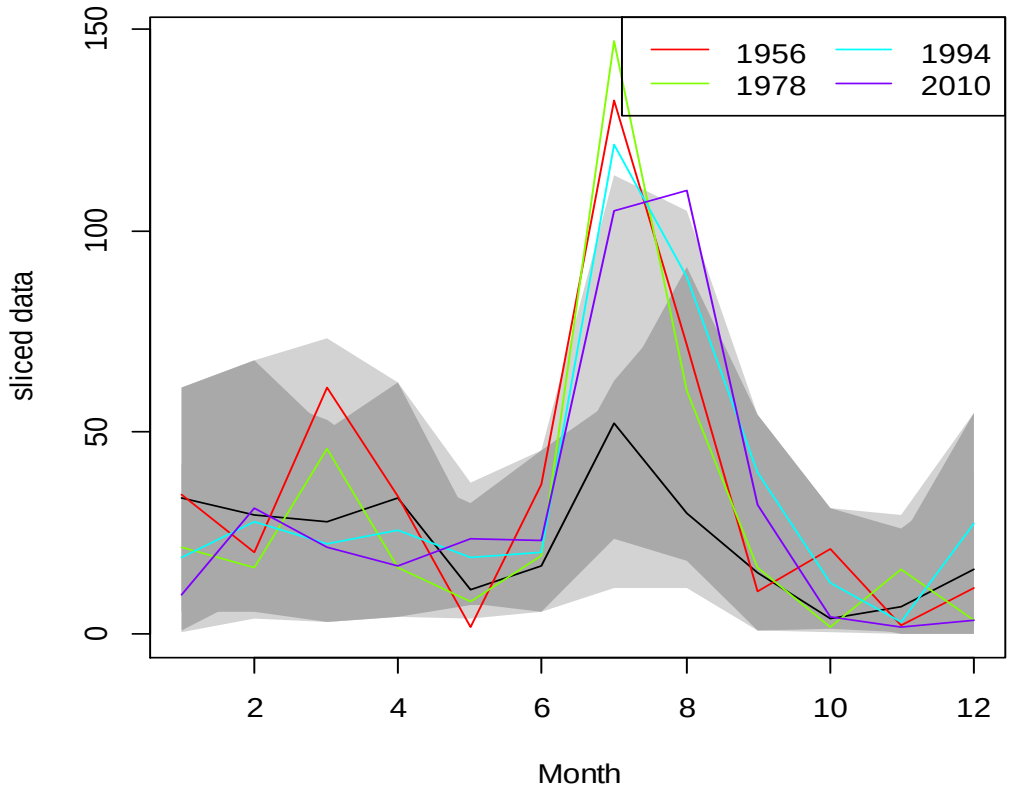

Figure 8. Functional HDR plot for rainfall data. Black represents the modal curve, whereas the outliers are represented by red, green and blue colors. Inner and outer regions are plotted by dark grey and light grey color, respectively

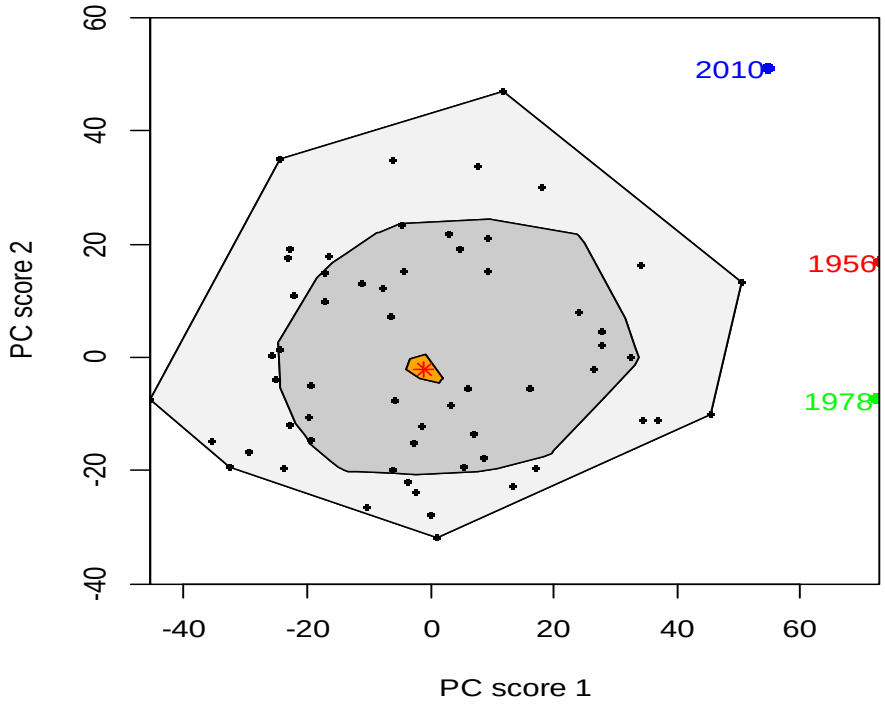

Figure 9. Functional bivariate plot for rainfall data with first two principal components being plotted. The Red asterisk is the sample median, whereas the inner and outer regions are plotted by dark grey and light grey color, respectively

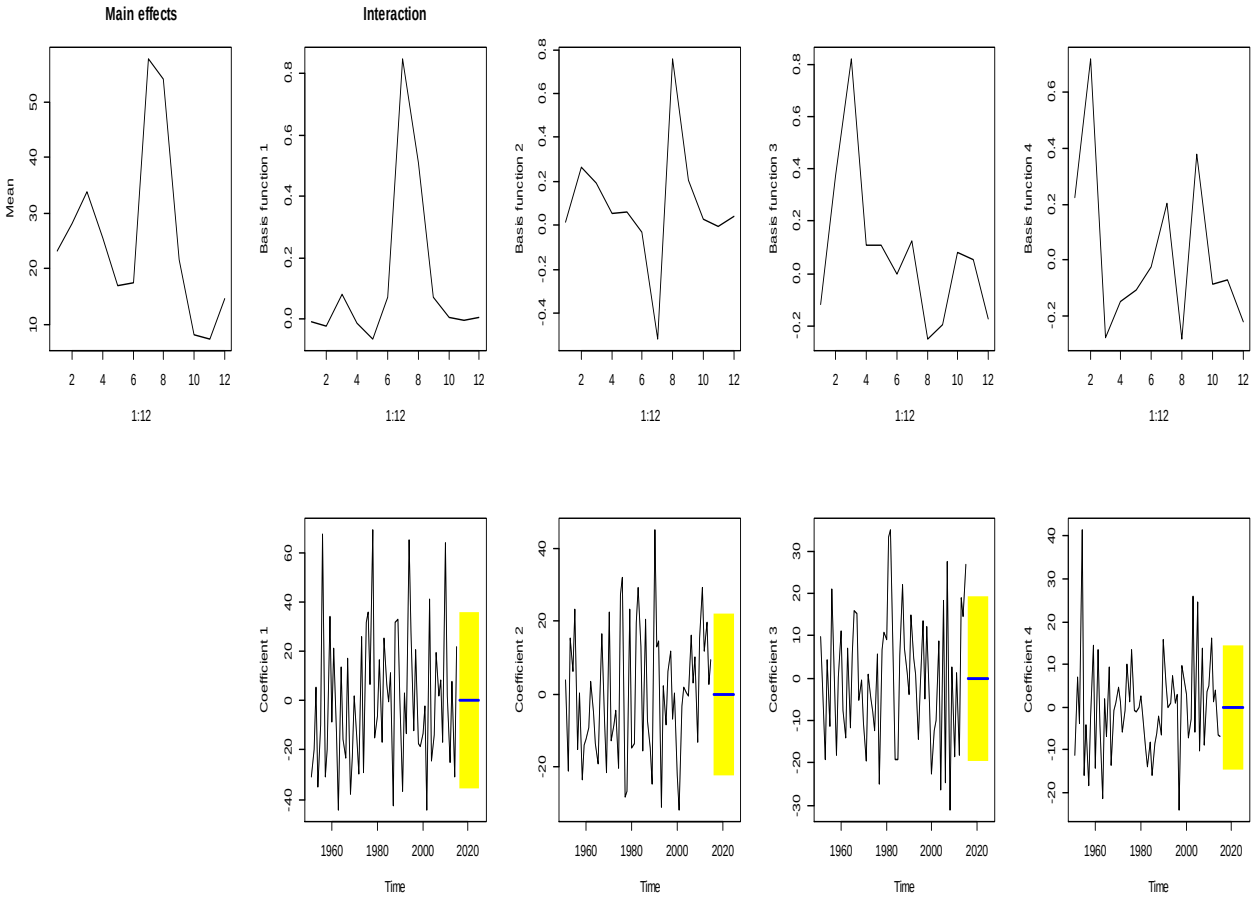

Figure 10. Different components of FTS models applied to the rainfall data, along with 10-year forecasts and 80% prediction intervals of the time series coefficients

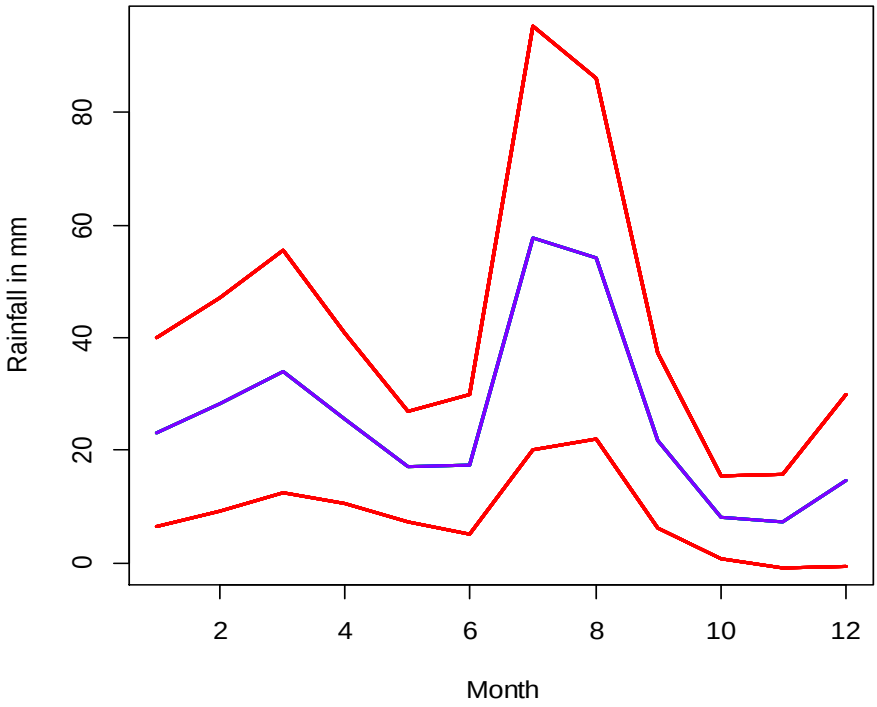

Figure 11. Sliced Functional Time Series Forecasts of monthly average rainfall in Pakistan (blue color), along with 80% prediction intervals (red color)

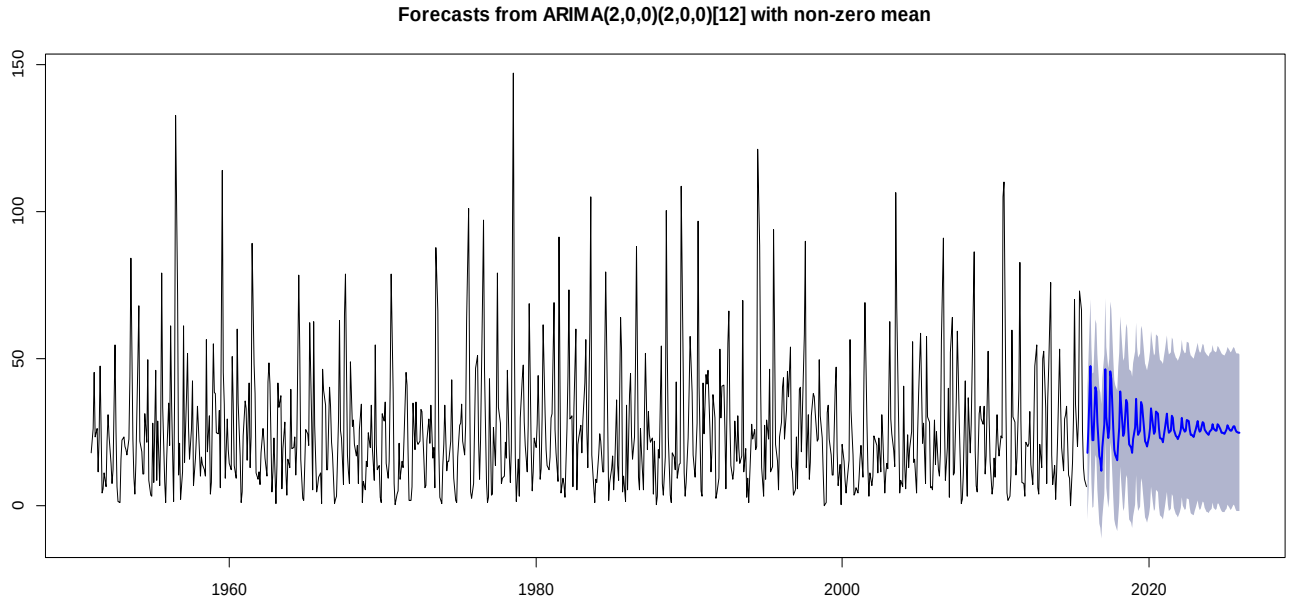

Figure 12. 10-year forecasts (2016-2025) of average rainfall in Pakistan using ARIMA model, along with 80% prediction intervals

Figures 7 and 8 represent the functional bagplot and functional HDR plot, whereas figure 9 depicts the functional bivariate plot. These plots showed that 1956, 1978 and 2010 are the outliers. Different components of FTS models and sliced functional time series forecasts are plotted in Figure 10 and 11 respectively. Figures 12 and 13 show the forecasts obtained from ARIMA and Exponential smoothing stat-space models.

Finally the forecasting performance of the models will be measured by Mean Error, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE). These measures are defined as follows:

1. Mean Error (ME) = mean($|e_i|$) where e_i *is the forecast error;* $e_i = y_i - \hat{y}_i$
2. Root Mean Square Error (RMSE) = $\sqrt{mean(e_i^2)}$
3. Mean Absolute Error (MAE) = mean($|e_i|$)
4. Mean Percentage Error (MPE) = mean(p_i) where $pi = 100e_i/y_i$
5. Mean Absolute Percentage Error (MAPE) = mean($|pi|$)

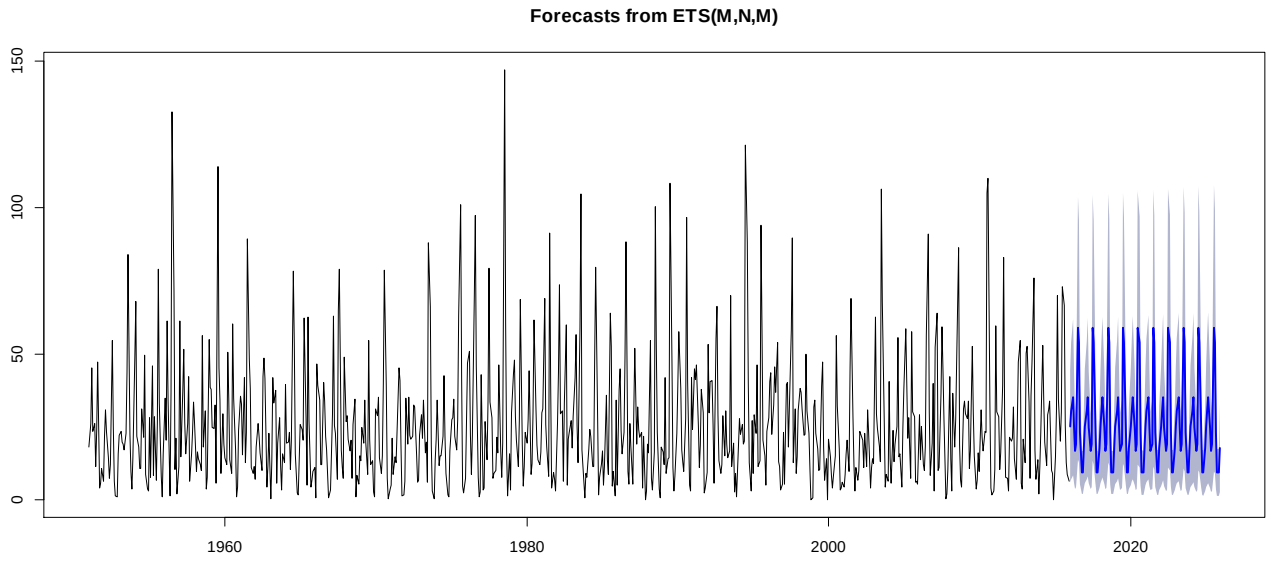

Figure 13. As in Fig. 12, but using Exponential Smoothing State Space model

Table 1. Out of sample forecasting performance of different models using mean error (ME), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE) and mean absolute percentage error (MAPE)

	ME	RMSE	MAE	MPE	MAPE
ARIMA/SARIMA	-0.7762	19.923	14.983	-290.986	316.75
ETS	-7.426	17.566	13.235	-249.697	261.963
Sliced FTS	0.2046	14.911	10.969	145.791	172.26

Table 2. Average rainfall forecasts obtained from Sliced Functional Time Series (SFTS), Exponential Smoothing State Space (ETS) and ARIMA models along with 80% prediction intervals. L_0 and H_0 are the lower and upper prediction limits

		SFTS			ETS			ARIMA		
	Time	**Forecast**	**L_0**	**H_0**	**Forecast**	**L_0**	**H_0**	**Forecast**	**L_0**	**H_0**
2016	**January**	23.16	6.31	40	25.33	6.061356	44.61467	17.83	4.671835	40.33958
	February	28.2	9.27	47.12	30.63	7.309544	53.96578	26.65	3.788073	49.51967
	March	33.96	12.49	55.43	35.13	8.359579	61.90637	47.34	24.44714	70.23694
	April	25.64	10.56	40.72	24.8	5.882769	43.69765	30.88	7.983067	53.78583
	May	17.03	7.24	26.81	16.74	3.961863	29.51913	22.21	0.697687	45.10508
	June	17.46	5.01	29.9	19.08	4.504264	33.66344	22.35	0.543824	45.25903
	July	57.68	20.01	95.34	58.8	13.83881	103.7448	40.24	17.34146	63.14432
	August	54.05	21.94	86.17	54.01	12.68029	95.3526	38.91	16.01111	61.81396
	September	21.73	6.17	37.3	23.14	5.418634	40.87252	28.33	5.433502	51.23636
	October	8.05	0.6034	15.5	9.48	2.214839	16.7581	16.92	5.97154	39.83132
	November	7.33	0.93	15.6	9.45	2.200347	16.70003	15.45	7.442604	38.36026
	December	14.55	0.7	29.8	17.85	4.146061	31.56516	11.73	11.16698	34.63588
2017	**January**	23.16	6.31	40	25.33	5.867251	44.80883	19.86	3.897922	43.63358
	February	28.2	9.27	47.12	30.63	7.074934	54.20047	27.59	3.804528	51.39075
	March	33.96	12.49	55.43	35.13	8.090652	62.17539	46.5	22.71112	70.3018
	April	25.64	10.56	40.72	24.8	5.693086	43.88739	30.02	6.226176	53.81785
	May	17.03	7.24	26.81	16.74	3.833823	29.64722	22.88	0.915795	46.67588
	June	17.46	5.01	29.9	19.08	4.358358	33.80939	26.23	2.440735	50.03241
	July	57.68	20.01	95.34	58.8	13.38949	104.1943	45.5	21.71106	69.30274
	August	54.05	21.94	86.17	54.01	12.26763	95.76539	42.96	19.17125	66.76293
	September	21.73	6.17	37.3	23.14	5.24188	41.04933	26.33	2.542413	50.13409
	October	8.05	0.6034	15.5	9.48	2.142423	16.83054	18.93	4.860238	42.73144
	November	7.33	0.93	15.6	9.45	2.128236	16.77217	17.17	6.618536	40.97314
	December	14.55	0.7	29.8	17.85	4.009863	31.70141	15.31	8.482458	39.10922

		SFTS			ETS			ARIMA		
	Time	**Forecast**	**L_0**	**H_0**	**Forecast**	**L_0**	**H_0**	**Forecast**	**L_0**	**H_0**
2018	**January**	23.16	6.31	40	25.33	5.674054	45.00209	21.42	4.112799	46.95849
	February	28.2	9.27	47.12	30.63	6.841418	54.43406	26.53	0.942679	52.12227
	March	33.96	12.49	55.43	35.13	7.822975	62.44315	38.73	13.1429	64.3313
	April	25.64	10.56	40.72	24.8	5.504282	44.07626	28.62	3.026021	54.21639
	May	17.03	7.24	26.81	16.74	3.706375	29.77471	23.72	1.871492	49.31888
	June	17.46	5.01	29.9	19.08	4.213125	33.95467	24.72	0.873199	50.31719
	July	57.68	20.01	95.34	58.8	12.94224	104.6417	36.09	10.50286	61.69325
	August	54.05	21.94	86.17	54.01	11.85686	96.1763	34.93	9.344451	60.53484
	September	21.73	6.17	37.3	23.14	5.065936	41.22533	26.73	1.140011	52.3304
	October	8.05	0.6034	15.5	9.48	2.070337	16.90265	20.85	4.735548	46.45484
	November	7.33	0.93	15.6	9.45	2.056453	16.84398	19.87	5.720138	45.47025
	December	14.55	0.7	29.8	17.85	3.874283	31.83703	18.11	7.48265	43.70774
2019	**January**	23.16	6.31	40	25.33	5.481732	45.19448	22.53	3.49034	48.56425
	February	28.2	9.27	47.12	30.63	6.608956	54.6666	26.54	0.504113	52.58634
	March	33.96	12.49	55.43	35.13	7.556505	62.70971	36.26	10.22496	62.30943
	April	25.64	10.56	40.72	24.8	5.316327	44.26427	27.93	1.897078	53.98206
	May	17.03	7.24	26.81	16.74	3.579498	29.90162	24.18	1.856169	50.22881
	June	17.46	5.01	29.9	19.08	4.068541	34.0993	25.58	0.46006	51.62492
	July	57.68	20.01	95.34	58.8	12.49698	105.0871	35.19	9.148114	61.2331
	August	54.05	21.94	86.17	54.01	11.44792	96.58537	34.02	7.977978	60.06296
	September	21.73	6.17	37.3	23.14	4.890772	41.40055	26.18	0.141818	52.2268
	October	8.05	0.6034	15.5	9.48	1.998571	16.97444	22.06	3.973978	48.111
	November	7.33	0.93	15.6	9.45	1.984987	16.91547	21.2	4.835593	47.24939
	December	14.55	0.7	29.8	17.85	3.739301	31.97206	20.09	5.951804	46.13318
2020	**January**	23.16	6.31	40	25.33	5.290254	45.38602	23.36	3.045384	49.78064
	February	28.2	9.27	47.12	30.63	6.377513	54.89811	26.19	0.229696	52.62006
	March	33.96	12.49	55.43	35.13	7.291199	62.9751	32.99	6.564868	59.41656
	April	25.64	10.56	40.72	24.8	5.129192	44.45147	27.28	0.856075	53.70819
	May	17.03	7.24	26.81	16.74	3.453173	30.02799	24.59	1.829088	51.02303
	June	17.46	5.01	29.9	19.08	3.924585	34.24331	25.32	1.103669	51.74845
	July	57.68	20.01	95.34	58.8	12.05365	105.5305	31.81	5.38466	58.23678
	August	54.05	21.94	86.17	54.01	11.04074	96.99268	31.09	4.66985	57.52197
	September	21.73	6.17	37.3	23.14	4.716362	41.57502	26.16	0.265421	52.5867
	October	8.05	0.6034	15.5	9.48	1.927112	17.04592	23.04	3.377612	49.47451
	November	7.33	0.93	15.6	9.45	1.913826	16.98665	22.47	3.947648	48.90447
	December	14.55	0.7	29.8	17.85	3.604895	32.10651	21.57	4.847806	48.00432

The rainfall data from January 1951 to December 2015 is divided into two subsets: a training set January 1951 – December 1995 and test set January 1996– December 2015. The training subset is used to fit the three models ARIMA, ETS and SFTS, and to estimate the parameters. The second subset is used for comparing the forecasts from different models. The results are shown in Table 1. Table 2 gives the average rainfall forecasts from 2016-2020 along with 80% prediction intervals.

4. Discussion

In Pakistan, adverse consequences of rainfall have already been observed. They are in the form of droughts and super floods which have badly affected human settlements, water management and agriculture. The majority of Pakistan's 180 million people live along the Indus River that is prone to severe flooding in July and August. Major earthquakes are also frequent in the mountainous northern and western regions. Measures to improve cultivation outputs and resilience to climate variation have long been underway, but a substantial push is still needed as amply demonstrated by the 2010 devastating floods. Priority areas for research and adaptation measures include the water, infrastructure, energy, and agriculture sectors, with particular attention to reducing vulnerability to flooding and improving water management in the Indus Basin.

In this paper, monthly rainfall data were analyzed through

the sliced functional time series (SFTS) model, a relatively new method of forecasting was introduced and the monthly forecasts for the next ten years (2016-2025) were obtained along with 80% prediction intervals. These forecasts were also compared with the forecasts obtained from Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing state space (ETS) models. It was found that the SFTS model performed better than standard ARIMA and ETS ones and the forecasts obtained from SFTS models are not only more accurate and reliable, but they also provide narrow prediction intervals as compared to other models.

ACKNOWLEDGEMENTS

The authors are thankful to the two anonymous referees for their valuable suggestions which improved the work considerably.

REFERENCES

[1] Abbot, J. and Marohasy, J. (2012) Application of artificial neural networks to rainfall in Queensland, Australia, 29(4), 717-730.

[2] Afsar, S., Abbas, N. and Jan, B. (2013) Comparative Study of Temperature and Rainfall Fluctuation in Hunza Nagar District, Journal of Basic and Applied Sciences, 9, 151-156.

[3] Ahmed, I., Tang, D., Wang, T., Wang, M. and Wagan, B. (2015) Precipitation Trends over Time Using Mann-Kendall and Spearman's rho Tests in Swat River Basin, Pakistan, Advances in Meteorology, http://dx.doi.org/10.1155/2015/431860.

[4] Ampow, E.M., Akuffo, B., Larbi, S.O. and Lartey, S (2013) Time Series Modeling of Rainfall in New Juaben Municipality of the Eastern Region of Ghana, The Special Issue on Contemporary Research in Business and Social Science, 4(8), 116-129.

[5] Bilgili, M., and Sahin, B (2010) Prediction of Long-term Monthly Temperature and Rainfall in Turkey, Energy Sources, 32(1), 60-71.

[6] Duong, T. & Hazelton, M. L. (2005) Cross-validation bandwidth matrices for multivariate kernel density estimation, Scandinavian Journal of Statistics 32(3), 485-506. http://www3.interscience.wiley.com/journal/118653028.

[7] Faisal, N. and Ghaffar, A (2012) Development of Pakistan's New Area Weighted Rainfall using Thiessen Polygon Method, Pakistan Journal of Metrology 9(17), 107-116.

[8] Geweke, J. and Porter-Hudak, S. (1983) The Estimation and Application of Long Memory Time Series Models. Journal of Time Series Analysis, 4, 221-237.

[9] Hyndman, R. J. (1996) Computing and graphing highest density regions, The American Statistician, 50(2), 120-126.

[10] Hyndman, R.J., and Shang, H.L (2010) Rainbow plots, Bagplots and Boxplots for functional data, Journal of Computational and Graphical Statistics, 19 (1), 29-45.

[11] Hyndman, R. J. & Ullah, M. S. (2007) Robust forecasting of mortality and fertility rates: A functional data approach, Computational Statistics & Data Analysis 51(10), 4942-4956. http://ideas.repec.org/a/eee/csdana/v51y2007i10p4942-4956.html.

[12] Kambezidis, H.D, Larissi, I.K. Nastos, P.T. and Paliatsos, A.G. (2010) Spatial variability and trends of rain intensity over Greece, Adv. in Geoscie. 26, 65-69.

[13] Kane, I. L. and Yusof, F. (2013) Assessment of Risk Rainfall Events with Hybrid of ARFIMA-GARCH, Modern Applied Sciences, 7(12), 78-89.

[14] Menabde, M., Harris, D., Seed, A., Austin, G. and Stow, D. (1997) Multiscaling Properties of Rainfall and Bounded Random Cascades Water Resources 33(12), 2823-2830.

[15] Muslehauddin,M. and Faisal, N. (2006) Long Range Forecast of Sindh Monsoon, International Journal of Metrology, 3(5), 35-44.

[16] Nayak, D.R., Mahapatra, A. and Mishra, P. (2013) A Survey on Rainfall Prediction using Artificial Neural Network, International Journal of computer Applications, 72(16), 32-40.

[17] Ramsay, J.O., Silverman, B.W. (2005) Functional Data Analysis. Second ed. Springer, New York.

[18] Rousseeuw, P, Ruts, I. and Tukey, J. (1999) The bagplot: A bivariate boxplot, The American Statistician, 53(4), 382-387.

[19] Sakellarian, N. K. and Kambezidis, H.D (2003) Total Precipitation in the Athens area, Greece: Annual precipitation patterns averaged for five, ten and fifteen days, Fres. Environ. Bull. 12(11), 1416-1420.

[20] Sakellarian, N. K. and Kambezidis, H.D (2004) Prediction of the total rainfall amount during August and November in the Athens area, Greece: Fres. Environ. Bull. 43(3B), 289-292.

[21] Salma, S. and Rehman, M.A., Shah (2012) The Rainfall Trend in Different Climate Zone of Pakistan, Pakistan journal of Metrology, 9(17), 37-47.

[22] Tukey, J.W (1974) Mathematics and the picturing of data, Proceedings of the International Congress of Mathematicians, Canadian Mathematical Congress, Montreal, 2, 523-532.

[23] Villarini, G., Smith, J.A., Napolitano, F. (2010) Non-stationary Modeling of a long record of Rainfall and Temperature over Rome, Advances in Water Resources Volume 33(10), 1256–1267.

[24] Wood, S.N., 1994. Monotonic smoothing splines fitted by cross validation. SIAM J. Sci. Comput. 15(5), 1126–1133.

[25] Yamoah, A., E., Bashiru I. Saeed and Karim A (2016) Sarima Modelling and Forecasting of Monthly Rainfall in the Brong Ahafo Region of Ghana, World Environment, 6(1), 1-9.

[26] Yusof, F. and Kase, I.L (2012) Modeling Monthly Rainfall Time Series Using ETS State Space and SARIMA Models, International Journal of Current Research 4(9), 195-200.

[27] Yusof, F. and Kane, I. L. (2012) Volatility Modeling of Rainfall Time Series, Theory and Applied Climatology 113, 247–258.

[28] Yusof, F., Kane, I. L. and Yusop, Z. (2013) Structural Break or Long Memory: An Empirical Survey on Daily Rainfall data sets across Malaysia, Hydrol. Earth Syst. Sci., 17(4), 1311-1318.

[29] Zahid, M. and Rasul, G. (2011) Frequency of Extreme Temperature & Precipitation Events in Pakistan 1965-2009, Science International 23(4), 313-319.