Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences: 57 (3): 31-38 (2020)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Rainfall Pattern in Pakistan in the Perspective of Probability Distribution

Muhammad Fahim Akhter*, and Shaheen Abbas

Mathematical Sciences Research Centre, Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan

Abstract: Rainfall of Pakistan big cities (Karachi, Peshawar, Quetta, Lahore, and Islamabad) are investigated by the probability distribution. The time series range of monthly analysis (1961 to 2015) is used. In this study, Johnson SB, Gen. Pareto, Pareto 2, Power function and Weibull are fitted for rainfall. The significance of probability distribution is estimated by Kolmogorov-Smirnov, Anderson Darling and Chi-square statistical tests. The Johnson SB is the best-fitted distribution observed for each rainfall cities using Chi-square test, while the other two test shows the variation in fitted distribution. In our study of the distribution, all rainfall stations obtained (excess of Kurtosis > 0) is leptokurtic.

Keywords: Kolmogorov Smirnov, Anderson Darling, Chi-Square, Rainfall pattern, Probability Distribution.

1. INTRODUCTION

The weather has severe impacts on human life. Researchers have investigated rainfall in any country for different purposes such as for design purposes of regionalized detail that useful for planners and other users [1]. Application of probability distribution against rainfall data is applied in a different region of the world [2]. In the Indian region (Pantnagar) Sharma and Singh (2010) described the best-fitted distribution in comparison of 16 distributions for maximum daily rainfall data [3]. Similarly, 9 distantly stations in northeast India analyzed and fitted fived extreme value distribution [4]. In this study, five Pakistan regions (big cities) are considered. According to big cities of Pakistan is the hub of trade, many ruler areas trade cash on these cities. These cities are the economic hub and the climate of any cities has importance to the generation of financial resources. Water resources are also the problem of these cities, particularly Karachi, while drought and heavy rainfall can disturb these cities. For this purpose, rainfall analysis of these big cities is important to understand the dynamics of rainfall by data fitting technique.

The data fitting cycle is one of the strategies to estimate parameters in understanding the data sample. Numerous statistical techniques allow us to estimate the fitted parameters based on sample data. The advantage of data analyses and simulation with software is to understand the data pattern. In this paper, we utilized Easy-Fit software, which allowed numerous probabilistic distribution of the given data sample, best fitting sample and provided better decisions and implementation of results analysis. Easy Fit is an interactive software that determines the parameters of the dynamical system [5]. Easy Fit is friendly software that applied around 70 distributions on data and performs three tests (Kolmogrove, Anderson Darling and Chi-square) and using ranks as per goodness of fit [6 and 7]. The three data patterns involved here under fitting distributions of rainfall and their applicability defined by distributions.

2. MATERIALS AND METHODS

The study of rainfall is based on five stations in Pakistan, including Islamabad, Karachi, Lahore, Peshawar and Quetta. Islamabad is the capital of Pakistan and other stations are the provincial capital of Pakistan. In this study, we used the

Received: April 2019; Accepted: September 2020
*Corresponding Author: Muhammad Fahim Akhter <m_fahimakhter@yahoo.com>

monthly amount of precipitation (mm) records for each station. The records started from 1961 to 2015 which is 55 years' rainfall records and 55 years containing 660 months. In this paper, the purpose of the study to explore the hidden mathematical model by three statistical tests by following steps.

2.1. Fitting the Probability Distribution (Step 1)

We have applied the number of the probability distribution on each rainfall station. The best-fitted probability distributions are found in Johnson SB, Gen. Pareto, Pareto 2, Power function and Weibull. The list of best-fitted distribution using three statistical tests is given in Table 2.

2.2. Goodness-of-Fit Tests (Step 2)

For choosing the best probability distribution, we set goodness-of-fit tests by following null hypothesis:

H_0: rainfall stations follow the best-fitted distribution.
H_A: rainfall station doesn't follow the best-fitted distribution.

Table 1. Summary of Statistics for R.F station

Statistics	Islamabad	Karachi	Lahore	Peshawar	Quetta
Sample Size	660	660	660	660	660
Mean	100.04s	16.46	53.381	39.204	20.956
Variance	14810.0	1949.7	6656.5	2261.9	1167.1
Std. Deviation	121.7	44.155	81.587	47.559	34.163
Coef. Of variation	1.2165	2.6826	1.5284	1.2131	1.6302
Std. Error	4.7371	1.7187	3.1758	1.8512	1.3298
Skewness	2.1419	4.2879	2.8014	2.4889	2.3504
Excess Kurtosis	5.2018	22.991	10.17	9.5415	6.4372

Table 2. List of First Rank fitted Probability Distribution

Probability Density Function	Parameters
Johnson SB $f(x) = \frac{\delta}{\lambda\sqrt{2\pi}\ z(1-z)} \exp\left(-\frac{1}{2}\left(\gamma + \delta\ln\left(\frac{z}{1-z}\right)\right)^2\right)$ $\zeta \le x \le +\lambda$	γ- continuous shape parameter δ- continuous shape parameter (δ>0) λ- continuous scale parameter (λ>0) ζ- continuous location parameter
Gen. Pareto $f(x) = \begin{cases} \frac{1}{\sigma}\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}} & \text{for } k \ne 0 \\ 1 - \exp\left(-\frac{(x-\mu)}{\sigma}\right) & K = 0 \end{cases}$ $\mu \le x < +\infty, for\ K \ge 0\ \ \mu \le x \le \mu - \sigma/K\ \ for\ k < 0$	*K*- continuous shape parameter σ- continuous scale parameter (σ>0) μ- continuous location parameter
Pareto 2 $f(x) = \frac{\alpha\beta^{\alpha}}{(x+\beta)^{\alpha+1}}$ $\gamma \le x < +\infty$	α- continuous shape parameter (α>0) β- continuous scale parameter (β>0)
Power Function $f(x) = \frac{\alpha(x-a)^{\alpha-1}}{(b-a)^{\alpha}}$ $a \le x \le b$	α- continuous shape parameter (α>0) a,b- continuous boundary parameter (a<b)
Weibul $f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$ $0 \le x \le +\infty$	*α*- continuous shape parameter(α>0) β- continuous scale parameter (β>0)

2.2.1. Kolmogorov- Smirnov (K-S) Test

The Kolmogorov-Smirnov (K-S) test makes a comparison between the empirical distribution function (EDF) with the distribution function (DF) of hypothesized distribution [9 - 10]. The (K-S) test has a no better option to use for tail discrepancies, it is one of the weakest points of (K-S) test [8]. Mathematically,

The (K-S) test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function (ECDF). Assume that we have a random sample $x_1, ..., x_n$ from some distribution with CDF F(x). The empirical CDF is denoted by:

$$CDF = F_n(x) = \frac{1}{n}[Number\ of\ observation \leq x] \quad (1)$$

The Kolmogorov-Smirnov (K-S) test statistic (D) is the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF) [1& 3].

$$D = \max_{1 \leq i \leq n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \quad (2)$$

Xi = random sample. *i = 1,2,3…………,n.*

2.2.2. Anderson-Darling (A-D) Test

The Anderson-Darling (A-D) test has a better utility to define the tail discrepancies [1] than the Kolmogorov-Smirnov (K-S) test. Even though both tests are similar but (A-D) test gives more weight on the tails of the distribution. The (A-D) test doesn't rely on the number of intervals. The use of the (A-D) test applies only to the input sample data which is the weakest point of (A-D) test [8]. In our study, this test provides fluctuation of distributions in dynamical behaviour. More precisely, (A-D) test compares to fit an observed distribution cumulative distribution function (CDF) to an expected cumulative distribution function (CDF). Generally, the (A-D) test weight on the tail than (K-S) test and statistics denoted by A^2 [8, 9 and10], mathematically written as:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[lnF(X_i) + ln\left(1 - F(X_{n-i+1})\right) \right] \quad (3)$$

2.2.3. Chi-Squared (C-S) Test

The (C-S) test is applied to binned data, so the value of test statistic χ^2 depends on how the value of data is binned. The number of bins is calculated by this formula:

$$k = 1 + log_2 N \quad (4)$$

$$x^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

O_i = Observed frequency for bin *i*.
E_i = Expected frequency for bin *i*.
E_i is calculated by:

$$Ei = F(x2) - F(x1) \quad (6)$$

'i'= number of observations. (*i=1, 2………. K*)
K is a number of bins and N is a sample size.

We applied three tests in this paper, the (K-S) test not required the binned of data that has more advantage than (C-S) test. But (C-S) test helpful for peak data and it meets better (R.F) probability distribution discussed in the next section of the study.

2.3. Significant Fit of Probability Distribution (Step 3)

The fitness of the probability distribution of each station is tested by three goodness-of-fit tests. Separately least value in all distributions concerning each test was selected. The particular distributions test statistic tested by critical value (α=0.01 & 0.05) and determined no rejection of the null hypothesis and shows rainfall fitness under particular distribution.

3. RESULTS AND DISCUSSION

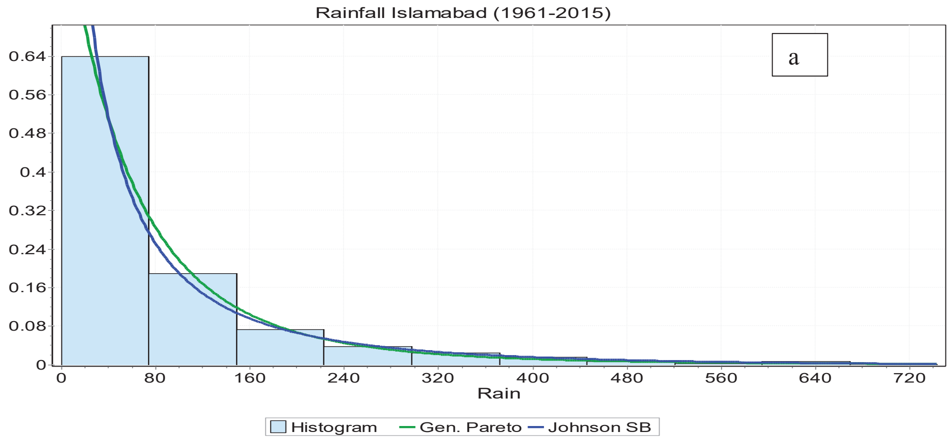
Pakistan is located in South Asia. The rainfall of Pakistan is different in each provincial capital and capital of Pakistan. Mostly hot weather is in surrounding of Pakistan. The heavy rainfall is mostly recorded in the monsoon period of Pakistan. In among all province capitals including the capital of Pakistan, the heavy rainfall recorded in Islamabad as compared to mean can be view in Table 1. In this study, we observed a mean monthly precipitation from 1961 to 2015. In that connection utilized numerous mathematical probability distributions to illustrate how the rainfall propagates in the capital province and the country capital of Pakistan. There is right-skewed observed in study stations of Pakistan, that is why the mean of each station is greater than the median. All stations provide positive Skewness and the bulk of data (peak) is on the left (Uni-model). The distribution is moderately skewed; its right tail is longer that is positively skewed of the distribution. The Kurtosis explained in terms of the central peak, higher values indicate the higher, sharper peak while lower values indicate the less distinct peak also mention the tails. Kurtosis is associated with the movement of the probability distribution on its centre and tails. In our study of the distribution, all stations obtained (excess of Kurtosis > 0) is called leptokurtic. Compared to a normal distribution, its tails are longer and fatter, and often its central peak is higher and sharper. The rainfall data of all stations are heavy-tailed relative to a normal distribution is obtained. The amount of variation relative to mean in the station is greater than 100%, the all related statistics of the rainfall station is depicted in Table 1.

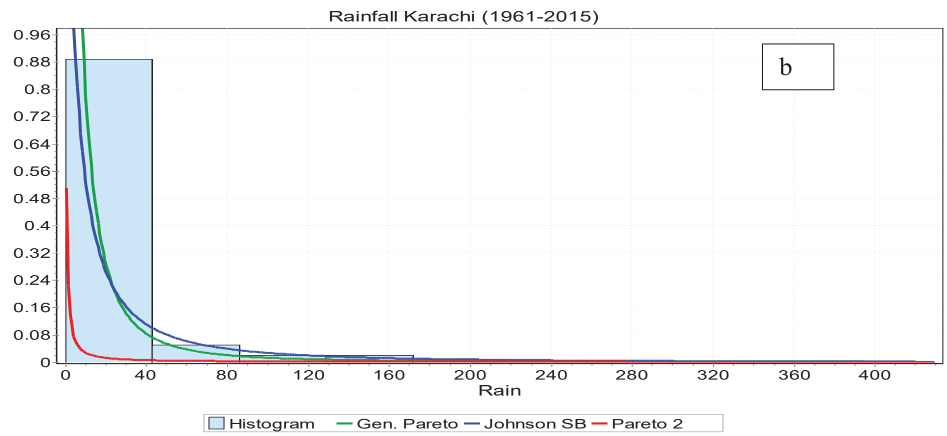
The list of probability distribution under three tests is depicted in Table 2. The Johnson SB and Gen. Pareto are observed best fitted for all capital provinces and country capital of Pakistan based on minimum statistics. With the help of the Chi-Squared test, the Johnson SB probability distribution was identified in all study stations of Pakistan. While the other two tests proposed Gen. Pareto, Pareto 2, Weibull and Power function distribution. The minimum statistic along with parameter of distribution for each rainfall station is depicted in table 3.

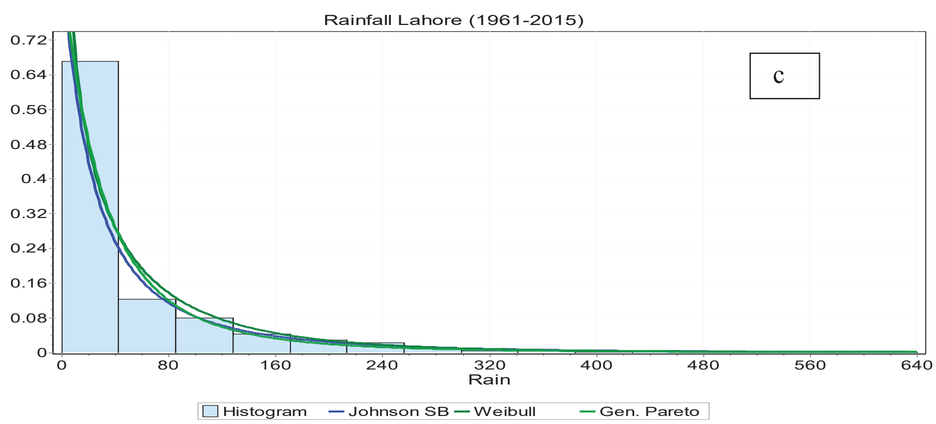
The minimum statistics results of Table 3 are also verified by two figures one is probability distribution Fig. and others are probability difference Fig. of significant models. The long right tail and significance of fitted distributions can be observed for each station Islamabad, Karachi, Lahore Peshawar and Quetta are furnished in Figure 1 (a-e). The highest peak of rainfall also identified rainfall 0 to 80mm has 64% in Islamabad, 0 to 40mm has 88% in Karachi, 0 to 40mm has 67%

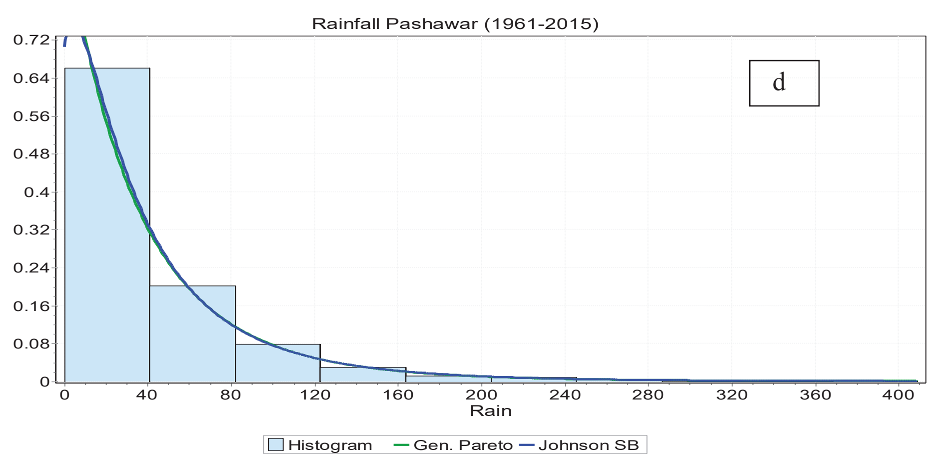
Table 3. The goodness of fit (GOF)- Summary of Statistics for Rainfall cycles

Cycle	Kolmogorov Smirnov			Anderson Darling			Chi-Squared		
	Distribution	Statistics	Parameters	Distribution	Statistics	Parameters	Distribution	Statistics	Parameters
Islamabad	Johnson SB	0.02438	γ=1.873 δ=0.70726 λ=876.37 ξ=-3.7806	Johnson SB	0.3849	γ=1.873 δ=0.70726 λ=876.37 ξ=-3.7806	Johnson SB	1.6358	γ=1.873 δ=0.70726 λ=876.37 ξ=-3.7806
Karachi	Gen. Pareto	0.29627	k=0.72162 σ=5.0708 μ=-1.7556	Pareto	0.056212	α=0.18628 β=1.4421E-7	Johnson SB	32.466	γ=2.3147 δ=0.51009 λ=469.72 ξ=-4.316
Lahore	Weibul	0.12727	α=0.749 β=50.861	Gen. Pareto	0.26957	k=0.34594 σ=38.663 μ=-5.7321	Johnson SB	9.5319	γ=2.3223 δ=0.72886 λ=809.68 ξ=-8.82
Peshawar	Gen. Pareto	0.090402	k=0.11146 σ=38.056 μ=-3.6257	Gen. Pareto	3.4813	k=0.11146 σ=38.056 μ=-3.6257	Johnson SB	9.8268	γ=3.2744 δ=1.0657 λ=806.78 ξ=-11.228
Quetta	Power Function	0.18757	α=0.08557 a=7.1852E-15 b=405.7	Gen. Pareto	37.455	k=0.38712 σ=15.131 μ=-3.7326	Johnson SB	48.354	γ=1.9466 δ=0.68163 λ=258.21 ξ=-6.2001


Rainfall Islamabad (1961-2015)
a
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0
0
80
160
240
320
400
480
560
640
720
Rain
Histogram
Gen. Pareto
Johnson SB


Rainfall Karachi (1961-2015)
b
0.96
0.88
0.8
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0
0
40
80
120
160
200
240
280
320
360
400
Rain
Histogram
Gen. Pareto
Johnson SB
Pareto 2


Rainfall Lahore (1961-2015)
c
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0
0
80
160
240
320
400
480
560
640
Rain
Histogram
Johnson SB
Weibull
Gen. Pareto


Rainfall Pashawar (1961-2015)
d
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0
0
40
80
120
160
200
240
280
320
360
400
Rain
Histogram
Gen. Pareto
Johnson SB

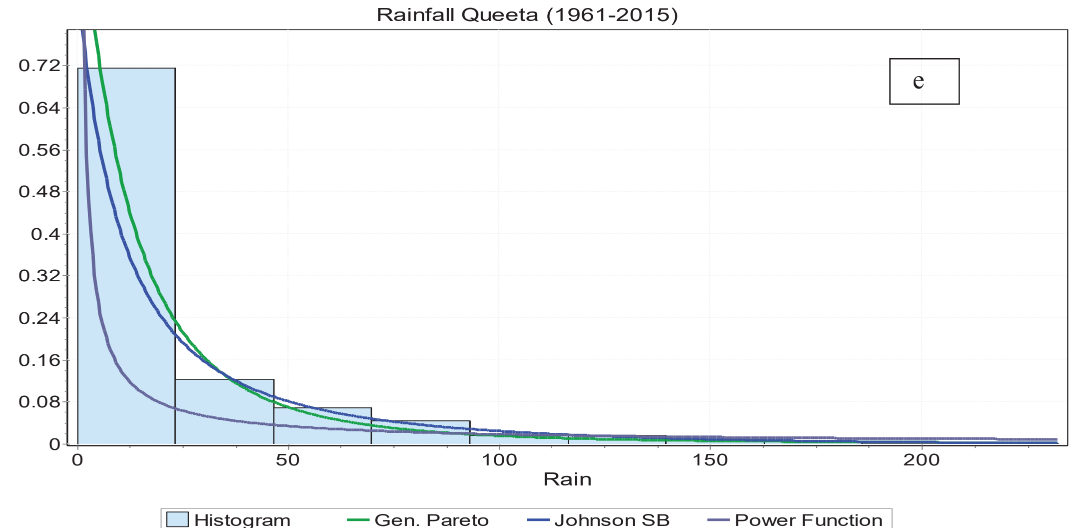
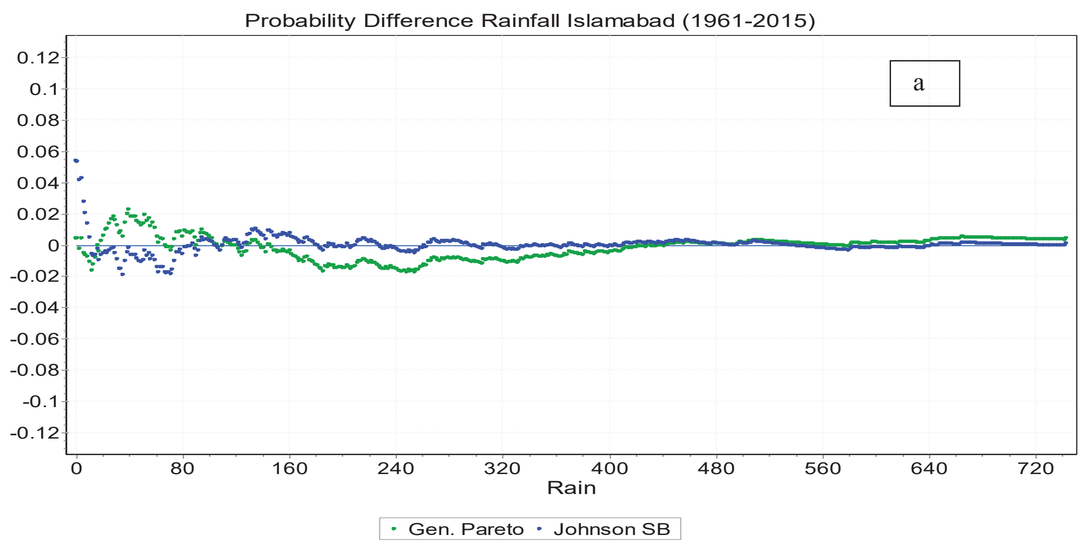
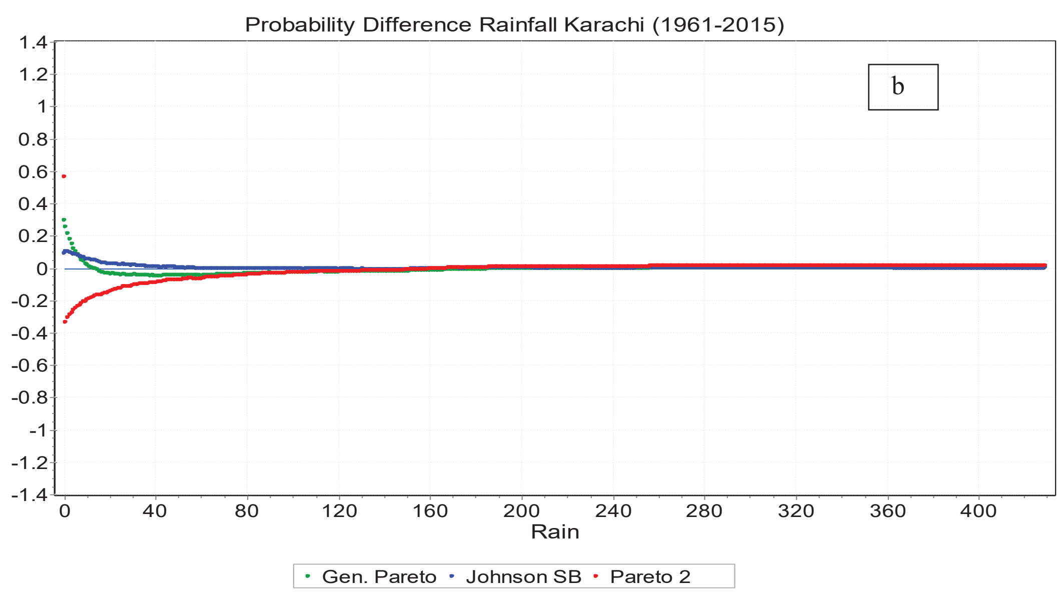
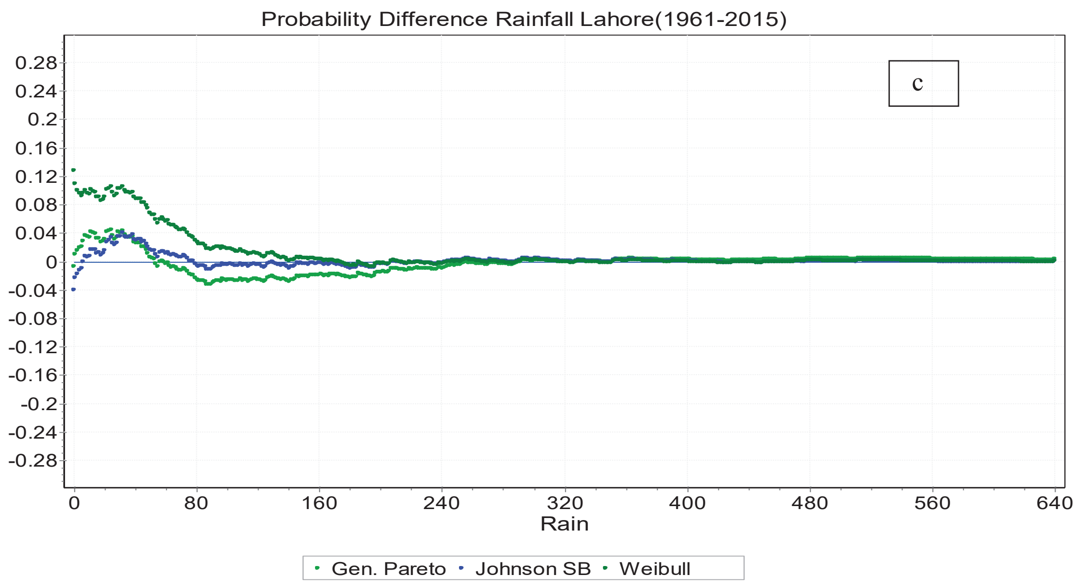
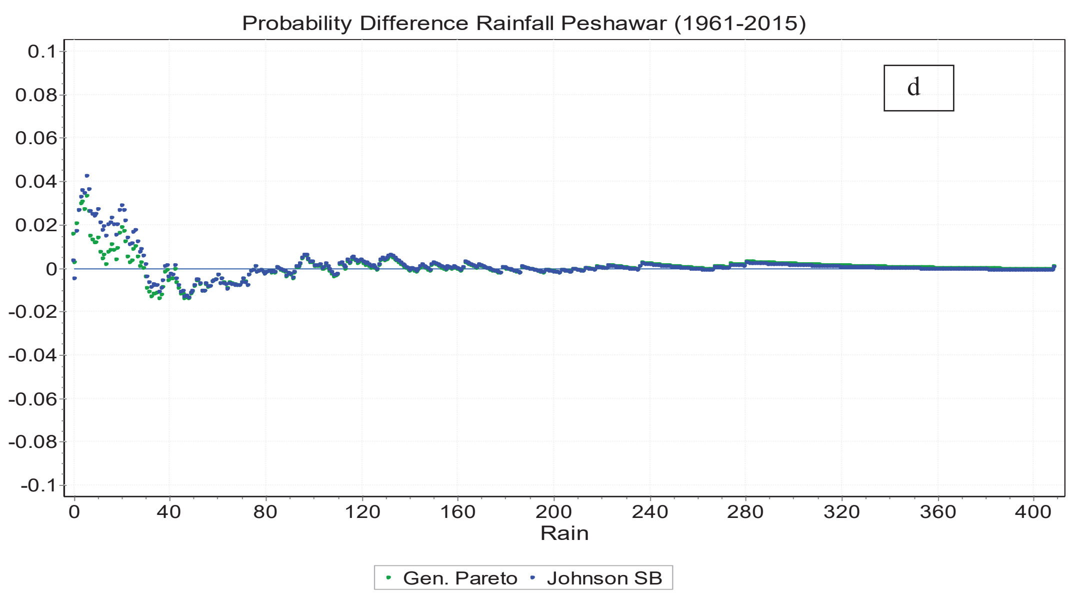


Fig. 1. The significant probability distribution of rainfall along with comparison in different cities from 196-2015 in (a) Islamabad (b) Karachi (c) Lahore (d) Peshawar and (e) Quetta.









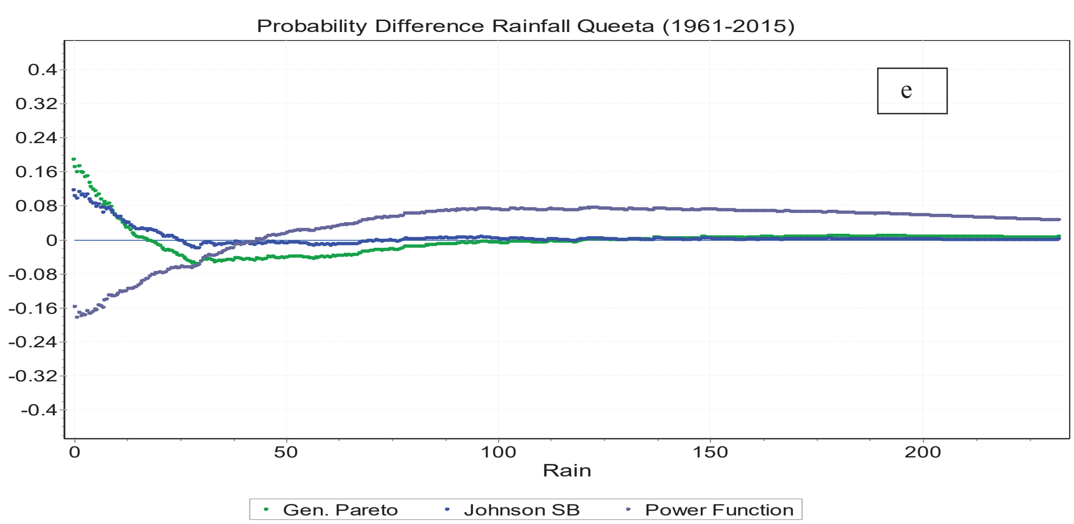


Fig. 2. Probability difference of significant distribution under three tests with comparison in (a) Islamabad(b) Karachi (c) Lahore (d) Peshawar (e) Quetta from 1961-2015.

in Lahore, 0 to 40mm has 66% in Peshawar and 0 to 40mm has 72% in Quetta can be seen in Figure 1 (a-e).

Further, we have checked the comparison of the obtained three tests distribution (Rank 1) with probability difference of distribution. The graph shows the acceptance of Johnson SB is more because less probability difference observed during study rainfall (mm). All significance of probability distribution is verified in Figure 2 (a-e). The similar regional difference shows the exponential decay of each station. We considered five big city rainfall stations in Pakistan. According to the higher concentration of rainfall, the probability difference was easy to visualize the probability differences indicates Johnson SB probability distribution.

4. CONCLUSION

The study was based on the generation of rainfall by the probability distribution. We utilized three tests confine on Johnson SB, Gen. Pareto, Pareto 2, Power function and Weibull distribution. But the comparison of among all three tests Chi-square produces a better option of peak data. Based on Chi-Square Johnson SB identified in each rainfall station, that is also verified with distribution probability differences. The probability differences of each station examine the less difference of probability which indicates Johnson SB distribution.

Further, we have observed that Islamabad is the heaviest rainfall station and Karachi was the lowest rainfall station but the behaviour of each is exponentially decayed. Based on results, we can conclude that in the future the generation of each big city of Pakistan rainfall follows a similar pattern of rainfall using Johnson SB distribution. This established probability distribution may represent the precipitation in hydrology, meteorology and others. The study parameters results may be applied for the identification of the best-fitted probability distribution of weather.

5. ACKNOWLEDGEMENT

Authors are thankful to Pakistan Meteorological depart for providing data and dedicate this work to (late) father of first author Muhammad Akhter Ahmed.

6. REFERENCES

1. T. Mayooran, and A. Laheetharan. The statistical distribution of annual maximum rainfall in Colombo district. *Sri Lankan Journal of Applied Statistics.* 15(2) 107-130 (2014).
2. H.I.D.E.O. Hirose. Parameter estimation in the extreme-value distributions using the continuation method. *Transactions of Information Processing Society of Japan.* 35(9): (1994).
3. M.A. Sharma, J.B. Singh. Use of probability distribution in rainfall analysis. *New York Science Journal.* 3 (9) 40-49 (2010).
4. S. Deka., M. Borah., S.C. Kakaty. Distributions of annual maximum rainfall series of north-east India. *European Water*, 27(28) 3-14 (2009).
5. K. Schittkowski. EASY-FIT: a software system for data fitting in dynamical systems. *Structural and Multidisciplinary Optimization* 23(2)153-169(2002).
6. G. K. Muralee, and I. Muthuchamy. Use of Easy Fit Software for Probability Analysis of Rainfall of Lower Bhavani Project command, Tamil Nadu. *Trends in Biosciences.* 7(19) 3053-3056(2014).
7. E.M. Masereka., G.M. Ochieng, and J. Snyman. Statistical analysis of annual maximum daily rainfall for Nelspruit and its environs. Jàmbá: *Journal of Disaster Risk Studies* 10(1) 1-10(2018).
8. C. Walk. Handbook on statistical distributions for experimentalists (2007).
9. M.F. Akhter., S. Abbas, and D. Hassan. Study of Coronal Index Time Series Solar Activity Data in the Perspective of Probability Distribution. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences,* 55(1) 27-33 (2018).
10. M.A. Hussain. Mathematical Aspects of The Impact of Urban Greenhouse Gas Emissions on Global Warming (Doctoral dissertation, Federal Urdu University of Arts, Science and Technology) (2006).