Pol. J. Environ. Stud. Vol. 25, No. 3 (2016), 925-937
DOI: *10.15244/pjoes/61715*

Original Research

Finding Probability Distributions for Annual Daily Maximum Rainfall in Pakistan Using Linear Moments and Variants

Ishfaq Ahmad[1]*, Aamar Abbas[2], Aamir Saghir[3], Muhammad Fawad[1]

[1]Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan
[2]Department of Mathematics, University of Poonch Rawalakot, Pakistan
[3]Mir Pur University of Science and Technology, AJK, Pakistan

Received: 22 December 2015
Accepted: 7 February 2016

Abstract

In this study, at-site frequency analysis (AFA) of an annual daily maximum rainfall (ADMR) series was carried out using the method of linear moments (L-moments) and their variants such as trimmed linear moments (TL-moments) and higher order linear moments (LH-moments). The ADMR series we investigated was observed at 28 meteorological observatories across Pakistan as retrieved from the Pakistan Meteorological Department (PMD). The basic aim of the study was to find best-fit (i.e., the most suitable) probability distribution among the class of various probability distributions. Initially different goodness-of-fit (GOF) measures such as the Kolmogorov-Smirnov test (KST), Anderson-Darling test (ADT), root mean square error (RMSE) and L-moments ratio diagram (LRD) were applied to determine not only the best-fit distributions but also the best linear estimation method for AFA. We observed that no single probability distribution could be declared as the best-fit distribution for all the stations. Five distributions were found to be the most appropriate: generalized extreme value (GEV), three parameter lognormal (LN3), Pearson type III (P3), generalized logistic (GLO), and generalized pareto (GPA). The TL-moments method was also applied for parameter estimation to mitigate the effect of outliers on final estimates. LH-moments were used for estimating the upper part of probability distributions and larger events in the data samples. LH moments alleviate the unwanted affects due to small sample values that may be obvious during estimation of events related to larger return periods. Using different GOF tests, we observed that the L-moments method was best for eight stations, TL-moments with trimming (1, 0), and LH-moments with level $\eta = 2, 3, 4$ were best for six and 14 stations, respectively. A theoretical relationship between TL-moments and LH-moments was also revisited, which revealed that LH-moments are special cases of TL-moments when we are motivated to make trimming only from the lower side.

Keywords: best-fit distribution, LH-moments, relative root mean square error, TL-moments

*e-mail: ishfaq.ahmad@iiu.edu.pk

Introduction

Pakistan being a developing country, 70% of its economy is dependent on agriculture. Rainwater is of great need for the agricultural sector as well as for human life and other living organisms. Pakistan receives rainfall not only in summer but also in winter. In summer, rainfall occurs mostly during monsoon season (early July to September). July and August are the peak months for monsoon rainfall in Pakistan [1]. The most important source of water for agriculture worldwide is rainfall. Almost 60% of the total annual rainfall in Pakistan is due to rainfall in summer [2]. All Kharif crops are mostly dependent on the pattern of monsoon rainfall. Winter rainfall is very important for Rabi crops in the country [3]. In Pakistan rainwater is used mostly in production of agriculture and hydroelectricity. Agriculture is the backbone of Pakistan's economy, but unfortunately extreme rainfalls in the country cause loss of crops, lives, and infrastructure. Extreme environmental events can leave extensive impacts on society and the economy [4]. Extreme rainfall is one of the most disastrous environmental events with vulnerable consequences [5]. Among others, one reason is the improper system for storage of rainwater in the result of extreme rainfall. However, knowledge about the magnitude and frequencies of extreme rainfall is inevitable for different reasons such as planning for water-related emergencies, sustainable water resource management, and construction of different hydraulic structures [6].

Frequency analysis (FA) provides information for estimating how often a specified event will happen. It provides knowledge for the magnitude of extreme events to their frequency of happening through the use of probability distributions [7]. In other words, FA is to estimate the return period associated with a given magnitude of rainfall. In AFA, selection of probability distribution is of immense importance because the wrong selection may lead to significant bias and error in final estimates of design flood, especially at larger return periods. This can result in either under or over estimation, and may have serious implication in practice. [8]. For AFA, we need data from a reasonably large record period, since the available data of rainfall is of shorter length than return periods of interest. There are many studies available in literature on selection and comparison of different probability distributions for AFA, but due to the availability of the small length of observed data as compared to return periods of interest, this job always has been challenging and controversial [9].

This study investigates the AFA of ADMR series on 28 meteorological stations of Pakistan, and finds best-fit probability distributions and the best estimation method (among L-moments and their variants) using different GOF measures for each station. To accomplish this study, we applied robust estimation methods such as L-moments as introduced by Hosking [10], plus TL-moments introduced by Elamir and Seheult [11] and LH-moments introduced by Wang [12]. These techniques have many advantages over other estimation methods such as moments and maximum likelihood method (e.g., [10-22]), and many more. L-moments are analogous to conventional moments showing some advantages over conventional moments. The method of L-moments provides nearly unbiased estimates relative to the other estimation methods (less sample variances), and provides a better identification of the parent distribution for a given data sample and more robust results in the presence of outliers, especially in small sample studies [13, 23]. Since 1990, L-moments has been widely used by many researchers across the world in a variety of fields [24-27]. In order to avoid undue favor due to outliers in data, a more robust estimation method (i.e., TL-moments as introduced by Elamir and Seheult [11]) can be used. TL-moments assign zero weight to extreme values in the data and claim to be more robust than simple L-moments in the presence of outliers in the data [28-32]. It is common practice in FA to predict large return period events. Therefore, Wang [12] introduced higher order L-moments, i.e., LH-moments, which is an extension of L-moments. According to Wang [12], LH-moments are expected to describe the upper-part of the distribution as well as larger events more accurately. LH-moments have been proved to be superior to other estimation methods in avoiding undue influence that more frequent observations have on less frequent observations. It accentuates high upper quantiles of the distribution rather than lower quantiles [12]. Various studies are available in this respect, for example: Shabri [33] compare LH-moments and L-moments for GEV distribution and found LH-moments ratios better then L-moments for annual flood data. Gamage [34] considers a series of annual maximum rainfall and found GEV as a best distribution using L_1, L_2, and L_4 moments.

Murshed et al. [35] formulated LH-moments for a few distributions used in hydrology, namely generalized Gumbel distribution, three parameter kappa distributions, beta-p distribution, and beta-k distribution. Deka et al. [36] identified GPA distribution as best using LH-moments when the level of LH-moment is zero. It is worth noting that TL-moments and LH-moments are the same when we use TL-Moments using trimming from the lower side only and don't trim any observation from the upper side. In other words, the order of trimming from the lower side only, such as (1,0), (2,0), (3,0…) in TL-moments also shows the order of level of LH-moments. Ours is the first study in Pakistan that deals AFA of ADMR series using TL- and LH-moments.

This study also included PE3 and GNO distributions for TL- and LH-moments. Recently the TL-moments and LH-moments of PE3 and GLO have been derived by Jan and Shabri [37]. Previously, in spite of their importance in FA, these two distributions were not considered in TL- and LH-moment studies. This study is different from at-site flood frequency analysis carried out by Ahmad et al. [38] in Pakistan using L- and TL-moments in three aspects: first it uses the ADMR series; secondly it uses not only L- and TL-moments, but also LH-moments; thirdly it includes two extra distributions, namely GNO and PE3 (often used in hydrology and meteorology).

Table 1. Basic information about meteorological stations used in the study.

Site Name	Province	Latitude	Longitude	Mean	L-CV	L-Skewness	L-Kurtosis
BADIN	SINDH	24.659	68.839	72.063	0.3613	0.1782	0.0974
CHHOR		24.943	68.278	72.967	0.3395	0.2108	0.1604
HYDERABAD		25.379	68.368	48.420	0.4009	0.1608	0.0674
JACOBABAD		28.276	68.451	59.453	0.5698	0.1608	0.0674
KARACHI		24.893	67.028	50.120	0.4170	0.1903	0.1055
MOHEN-JO-DARO		27.324	68.135	33.387	0.4089	0.3333	0.1355
NAWABSHAH		26.219	68.392	42.947	0.4531	0.2685	0.1355
PADIDAN		26.778	68.284	46.093	0.5051	0.4498	0.2925
ROHRI		27.679	68.899	46.640	0.4684	0.3334	0.2000
LASBELLA	BLOCHISTAN	25.837	66.522	48.323	0.3914	0.4895	0.4694
PASNI		25.266	63.466	34.760	0.4100	0.2989	0.2079
KHUZDAR		27.800	66.616	43.507	0.2846	0.4782	0.4780
ZHOB		31.350	69.450	37.547	0.2481	0.2970	0.3507
JIWANI		25.047	61.745	44.420	0.5033	0.3636	0.1938
PANJGUR		26.966	64.100	26.133	0.2978	0.2483	0.1014
Barkhan		29.897	69.527	47.597	0.2306	0.1715	0.1440
KHANPUR	PUNJAB	31.091	72.825	48.467	0.4612	0.3128	0.1380
FAISALABAD		31.418	73.077	61.820	0.2252	0.1860	0.1877
LAHORE		31.545	74.340	90.450	0.2343	0.2104	0.1539
JHELUM		32.933	73.720	100.05	0.2270	0.3202	0.1660
SIALKOT		32.497	74.536	72.063	0.2551	0.2746	0.1360
DI KHAN		31.823	70.909	59.963	0.2397	0.2572	0.2885
DROSH	KPK	35.550	71.797	52.717	0.1979	0.2838	0.3275
CHITRAL		35.839	71.780	52.253	0.2710	0.4265	0.2778
CHERAT		33.822	71.890	65.650	0.2693	0.4855	0.3509
BUNJI	NOTHERN AREAS	34.952	72.331	28.663	0.2758	0.1544	0.1351
CHILLAS		35.431	74.095	34.853	0.2885	0.3940	0.2841
GILGIT		34.952	72.331	25.457	0.3175	0.3094	0.0969

Data and Methodology

The data of daily rainfall from 28 meteorological observatories had been retrieved from PMD, Karachi. The data used in this study cover four provinces in Pakistan. Data have been selected following the standard criteria of length of the data, quality, urbanization, variability, and climate change. After that, ADMR series were abstracted from daily rainfall series. For any specific year and station, ADMR series is a single value in the whole year that is maximum among all the values of recorded daily rainfall. We use data of the same length of 30 years. The mean of ADMR series varies from 33 mm to 100 mm. The ADMR series of Jacobabad station has more relative variation, i.e., the values are more dispersed (sample L-CV) as compared to other data sets. One observation in the Jacobabad ADMR series is large enough compared to others, and may be a reason for this large L-CV. The Drosh station has less relative variation (sample L-CV) as compared to other stations in the study. In general, most stations in Sindh province have larger sample L-CV as compared Punjab, KPK, and northern areas. The ADMR series of three stations of Baluchistan (Padidan, Lasbella, Khuszdar) and one station of KPK province as Cherat are relatively more skewed as compared to other ADMR series. The basic information about all stations used in the study is given in Table 1. The geographical location of these stations is given in Fig. 1.

Fig. 1. Geographical locations of meteorological stations used in the study.

L-moments, TL-moments, and LH-moments

L-moments are analogous to conventional moments. L-moments are a function of linear ordered statistics. They provide summary statistics for probability distributions and data samples. These summary statistic are measurements of location, dispersion, skewness, and kurtosis.

Suppose we have a random sample of size "*n*" drawn from a distribution, with distribution function $F(y) = P(Y \leq y)$, the expectation of any i^{th} order statistics in a sample of size "*n*" can be defined as:

$$\mathrm{E}(X_{j:m}) = \frac{m!}{(j-1)!\,(m-j)!}\int_0^1 y(F)F^{j-1}(1-F)^{m-j}dF \tag{1}$$

The rth population L-moment can be explained as:

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\mathrm{E}(X_{r-k:r})$$
$$r = 1, 2 \ldots \tag{2}$$

Normally we need the first four L-moments for r = 1,2,3,4. L-moments can also be represented as the linear combination of probability weighted moments (PWMs) as given below:

$$\lambda_{r+1} = \sum_{k=0}^{r}\beta_k(-1)^{r-k}\binom{r}{k}\binom{r+k}{k} \tag{3}$$

In terms of PWMs, the first four L-moments are:

$$\lambda_1 = \beta_0 \tag{4}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{5}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{6}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{7}$$

L-CV, L-skewness, and L-kurtosis are represented by L-moment ratios τ, τ_3, and τ_4 respectively, which are given in Eqs. 8 to 10:

$$(\tau) = {\lambda_2}/{\lambda_1} \tag{8}$$

$$(\tau_3) = {\lambda_3}/{\lambda_2} \tag{9}$$

$$(\tau_4) = {\lambda_4}/{\lambda_2} \tag{10}$$

Sample L-moments are estimated from sample order statistics, which are defined by Asquith [32].

$$l_r = \frac{1}{r}\sum_{i=1}^{n}\left[\frac{\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-i}{j}}{\binom{n}{r}}\right]X_{i:n} \tag{11}$$

L-moments are less affected by extremes in data series and can model a wide range of theoretical distributions. L-moments is meaningful when the distribution has finite mean. L-moments provide a better tool to identify the parent distribution generating data sample [10].

When our concern is extreme observations, we can prefer TL-moments assigning zero weight to these extreme values. They have some advantages over simple L-moments, for example they can exist even if the mean of probability distribution does not exist. The most common example is Cauchy distribution. The sample quantities are unbiased estimators to population quantities and are relatively more robust to outliers. In TL-moments, as compared to simple L-moments, the expectations of the order statistics are replaced by expectations of the order statistics of a larger size. The size is enlarged equal to the total amount of trimming. In other words, in TL-Moments the "$E(Xr\text{-}k{:}r)$" is replaced by "$E(Xr + t_1 - k{:}r + t_1 + t_2)$". The choice of trimming is important while using TL-moments. The choice of trimming may be determined by minimizing the sum of absolute differences between theoretical quantile function and its TL-moments representation. The advantage of this procedure is that it deals with the whole probability model. It is less sensitive for the choice of trimming for each parameter separately [11]. The r_{th}, TL-moments can be written as:

$$\lambda_r^{(t_1,t_2)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} E\left(X_{r+t_1-k:r+t_1+t_2}\right)$$
$$r = 1,2,\ldots \quad (12)$$

…where r and t represents the order and the level of trimming, t_1 the smallest and t_2 the largest values need to be trimmed. If level of trimming becomes zero (i.e., $t_1 = t_2 = 0$), the TL-moments becomes L-moments. In this study, we used trimming only from the lower side, i.e., $t_1 = 1$ and $t_2 = 0$. The first four TL-moments are given below:

$$\lambda_r^{(1,0)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} E\left(X_{r+1-k:r+1+0}\right)$$
$$r = 1,2,\ldots \quad (13)$$

$$\lambda_1^{(1,0)} = E\left(X_{2,2}\right) \quad (14)$$

$$\lambda_2^{(1,0)} = \frac{1}{2}E\left(X_{3,3} - X_{2,3}\right) \quad (15)$$

$$\lambda_3^{(1,0)} = \frac{1}{3}E\left(X_{4,4} - 2X_{3,4} + X_{2,4}\right) \quad (16)$$

$$\lambda_4^{(1,0)} = \frac{1}{4}E\left(X_{5,5}-3X_{4,5} + 3X_{3,5} - X_{2,5}\right) \quad (17)$$

The population TL-ratios for trimming of $t_1 = 1$, $t_2 = 0$ are defined as:

$$\tau^{(1,0)} = \frac{\lambda_2^{(1,0)}}{\lambda_1^{(1,0)}} \quad (18)$$

$$\tau_3^{(1,0)} = \frac{\lambda_3^{(1,0)}}{\lambda_2^{(1,0)}} \quad (19)$$

$$\tau_4^{(1,0)} = \frac{\lambda_4^{(t_1,t_2)}}{\lambda_2^{(t_1,t_2)}} \quad (20)$$

$\lambda_1^{(1,0)}$ is the measure of TL location, $\tau^{(1,0)}$ is a measure of TL-coefficient of variation (TL-CV), and $\tau_3^{(1,0)}$ and $\tau_4^{(1,0)}$ are measurements of TL-skewness and TL-kurtosis, respectively. Frequency analysis is conducted mostly for predicting the events of larger return periods.

The LH-moments are being used for depicting the upper part of distributions and larger events in the data samples. LH-moments are the extension or generalization of L-moments, which are defined as the linear combination of higher order statistics introduced by Wang [12]. Product moments and moment ratios have been found to be sensitive to the upper part of distributions and thus sample outliers [39]. The L moments are oversensitive to the lower part of distributions and give insufficient weight to large sample values that actually contain useful information on the upper distribution tail [12]. The LH moments mitigates the unwanted affects due to small samples that may be obvious during estimation of events for larger return periods.

The r_{th} LH-moments with order η *=1,2,3,4 and* for $r = 1,2,3,4$ as defined by Wang [12] are given as:

$$\lambda_r^\eta = \frac{(\eta + r)}{r!}\sum_{k=0}^{r-1} C_{r,k}E[X_{\eta+r-k:\eta+r}] \quad (21)$$

…where $C_{rk} = (-1)^k\binom{r-1}{k}$. As η increases, LH-moments explains more the characteristics of the upper part of distributions and larger events in data. LH-co-efficient of variation τ^η, LH Skewness τ_3^η, and LH kurtosis τ_4^η are given in Eqs. 22 to 24.

$$\tau^\eta = \frac{\lambda_2^\eta}{\lambda_1^\eta} \quad (22)$$

$$\tau_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta} \quad (23)$$

$$\tau_4^\eta = \frac{\lambda_4^\eta}{\lambda_2^\eta} \quad (24)$$

Relationship between TL-Moments and LH-Moments

It is worth noting that TL-moments and LH-moments are the same when we use TL-Moments using trimming from the lower side only and do not trim any observation from the upper side. In other words, the amount of trimming from the lower side only such as ((1,0),(2,0),(3,0) and (4,0) also equal LH-moments with order η = 1,2,3,4. In this section we show that trimmed L-Moments are special cases of LH-Moments.

Take a specific case with $\eta = 2$:

$$\lambda_1^{(\eta=2)} = E\left[X_{(\eta+1):(\eta+1)}\right] = E\left(X_{3,3}\right) = \lambda_1^{(2,0)} = E\left(X_{3,3}\right) \quad (25)$$

$$\lambda_2^{(\eta=2)} = \frac{1}{2}E\left[X_{(\eta+2):(\eta+2)} - X_{(\eta+1):(\eta+2)}\right] = \frac{1}{2}E\left(X_{4,4} - X_{3,4}\right) = \lambda_2^{(2,0)} = \frac{1}{2}E\left(X_{4,4} - X_{3,4}\right) \quad (26)$$

$$\lambda_3^{(\eta=2)} = \frac{1}{3}E\left[X_{(\eta+3):(\eta+3)} - 2X_{(\eta+2):(\eta+3)} + X_{(\eta+1):(\eta+3)}\right] = \frac{1}{3}E\left(X_{5,5} - 2X_{4,5} + X_{3,5}\right) = \lambda_3^{(2,0)} = \frac{1}{3}E\left(X_{5,5} - 2X_{4,5} + X_{3,5}\right) \quad (27)$$

$$\lambda_4^{(\eta=2)} = \frac{1}{4}E\big[X_{(\eta+4):(\eta+4)} - 3X_{(\eta+3):(\eta+4)} + 3X_{(\eta+2):(\eta+4)} - X_{(\eta+1):(\eta+4)}\big]$$
$$= \frac{1}{4}E\big(X_{6,6-}3X_{5,6} + 3X_{4,6} - X_{3,6}\big) = \lambda_4^{(2,0)}$$
$$= \frac{1}{4}E\big(X_{6,6-}3X_{5,6} + 3X_{4,6} - X_{3,6}\big) \quad (28)$$

Estimation of LH/TL-Moments:

$$\hat{\lambda}_1^{\eta} = \frac{1}{\binom{n}{\eta+1}}\sum_{i=1}^{n}\left\{\binom{i-1}{\eta}\right\}x_i \quad (29)$$

$$\hat{\lambda}_2^{\eta} = \frac{1}{2\binom{n}{\eta+2}}\sum_{i=1}^{n}\left\{\binom{i-1}{\eta+1} - \binom{i-1}{\eta}\binom{n-i}{1}\right\}x_i \quad (30)$$

$$\hat{\lambda}_3^{\eta} = \frac{1}{3\binom{n}{\eta+3}}\sum_{i=1}^{n}\left\{\binom{i-1}{\eta+2} - 2\binom{i-1}{\eta+1}\binom{n-i}{1} + \binom{i-1}{\eta}\binom{n-i}{2}\right\}x_i \quad (31)$$

$$\hat{\lambda}_4^{\eta} = \frac{1}{4\binom{n}{\eta+4}}\sum_{i=1}^{n}\left\{\binom{i-1}{\eta+3} - 3\binom{i-1}{\eta+2}\binom{n-i}{1} + 3\binom{i-1}{\eta+1}\binom{n-i}{2} - \binom{i-1}{\eta}\binom{n-i}{3}\right\}x_i \quad (32)$$

…where η = 0,1,2,3 and 4.

Goodness-of-fit Criteria

Initially, different probability distributions have been considered in this study such as three-parameter generalized extreme value (GEV), generalized normal (GNO), generalized Pearson 3 (PE3), generalized logistic (GLO), generalized pareto (GPA), lognormal (LN3), Pearson type III (P3), two-parameter exponential (E), Gumbel (G), logistic (LOG), and normal (N). All these distributions have been studied in many countries for at-site and regional frequency analysis [6, 8, 18, 37-38, 40]. After this, the next step was to select the most appropriate distribution for each site. Different methods have been used for the goodness-of-fit measure such as including the Chi-square method, quantile-quantile plots, mean absolute deviation index, LMRD, RMSE, KST, and ADT. These tests have been used by many authors [8, 16, 36, 41-47]. In this study we implemented three GOF tests – RMSE, KS, and AD – and one graphical method, i.e., LRD for visual inspection. All these GOF measures evaluate the agreement between theoretical probability distribution and observed random sample.

L-moments Ratio Diagram (LRD)

The simplest way to determine which distribution is appropriate for data is patterned by the L-moments ratio diagram. The graphical demonstration of L-skewness and L-kurtosis is known as the L-moments ratio diagram. Different probability distributions are plotted in an L-moment ratio diagram and distribution will be decided by which ratio of L-Skewness and L-Kurtosis is in closer agreement with the sample data.

Kolmogorov-Smirnov Test (KST)

To check whether the sample comes from hypothesized continuous distribution we apply KST. The KST is based on the empirical distribution function (ECDF). This is a step function that increases by 1/n at the value of each ordered data point. The largest vertical difference between the theoretical and ECDF is defined as (KST):

$$D = \max_{1\le i\le n}\left(F(X_i) - \frac{i-1}{n}, \frac{i}{n} - F(X_i)\right) \quad (33)$$

…where n denotes sample size and $F(X_i)$ is the cumulative distribution function. KST is an exact test that does not depend on the underlying cumulative distribution function being tested. If calculated value exceeds the critical value (calculated through software), we reject the null hypothesis that data follows some specific distribution mentioned in null hypothesis. One of the major limitations for the KST test is that distribution must be fully specified, but if parameters are estimated from sample data, the critical region is no longer valid. So the critical region is calculated through simulation.

Anderson-Darling Test (ADT)

ADT is considered to be a refinement of KST. It is considered to be more powerful compared to KST. It is used to test if the sample data came from a population with some specific distribution. The Anderson-Darling statistic (A^2) is defined as:

$$A^2 = -n - S \quad (34)$$

$$S = \frac{1}{n}\sum_{i=1}^{n}(2i-1)\big[lnF(X_i) + \ln\big(1 - F(X_{n-i+1})\big)\big] \quad (35)$$

…where n denotes the sample size $F(X_i)$, is distribution function, and X_i are ordered observations. At a given level of significance H_0 will be rejected if the calculated value of the above statistic is greater than the critical values. The AD test is the modification of (KS) that gives more weight to heavy tailed distributions with small size, which is mostly used in meteorological applications [45].

Table 2. Results for different tests for basic assumptions.

Sites Name	Mann Kendall Test		Mann-Whitney U test		Lag-1 correlation coefficient		Runs Test	
	Tau	P-value	Mann-Whitney U	P-value	r_1	P-value	No of Runs	P-Value
BADIN	-0.07	0.568	91	0.389	0.125	0.471	11	0.1071
CHHOR	-0.094	0.475	89	0.345	-0.256	0.141	14	0.8512
HYDERABAD	0.0805	0.544	91	0.389	0.087	0.618	13	0.3609
JACOBABAD	-0.012	0.943	92	0.412	0.106	0.544	10	0.3611
KARACHI	0.283	0.329	68	0.070	0.123	0.479	15	0.9255
MOHEN-JO-DARO	0.0138	0.928	98.5	0.567	-0.84	0.630	18	0.2982
NAWABSHAH	0.0484	0.721	111.5	0.967	0.027	0.878	15	1
PADIDAN	0.0138	0.928	90	0.367	0.307	0.078	8	0.4151
ROHRI	-0.083	0.532	76.5	0.137	-0.040	0.818	13	0.5658
LASBELLA	0.0624	0.642	105.5	0.775	-0.209	0.230	17	0.1170
PASNI	0.152	0.246	100	0.624	-0.024	0.888	15	1.0000
KHUZDAR	-0.032	0.816	83	0.233	0.077	0.658	13	0.7893
ZHOB	-0.078	0.555	84.5	0.250	-0.117	0.502	13	0.7270
JIWANI	-0.11	0.401	92.5	0.412	0.081	0.643	16	0.8225
PANJGUR	0.109	0.411	98.5	0.567	0.163	0.347	14	0.7251
BARKHAN	0.187	0.153	99.5	0.595	0.137	0.430	10	0.4140
KHANPUR	0.283	0.329	77	0.148	0.110	0.527	17	0.6759
FAISALABAD	0.295	0.341	70.5	0.079	0.073	0.673	16	1
LAHORE	0.0784	0.555	104	0.744	0.146	0.400	13	0.7270
JHELUM	-0.154	0.238	69	0.074	0.236	0.174	14	1
SIALKOT	0.0576	0.668	101.5	0.653	0.076	0.660	14	0.7251
DI KHAN	0.0437	0.748	100	0.642	-0.047	0.778	15	0.9696
DROSH	0.113	0.391	86	0.285	0.138	0.426	13	0.4607
CHITRAL	-0.021	0.886	100	0.624	-0.137	0.432	13	0.7893
CHERAT	0.0903	0.497	96.5	0.512	-0.221	0.203	15	0.9696
BUNJI	0.0576	0.668	111.5	0.967	-0.056	0.746	15	0.8481
CHILLAS	-0.161	0.218	84.4	0.240	-0.119	0.495	13	1
GILGIT	0.115	0.381	109	0.902	-0.068	0.695	15	1

Root Mean Square Error (RMSE)

Another GOF measure used in this study is RMSE, which is expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - xF_i}{x_i}\right)^2} \quad (36)$$

…where x_i and xF_i represent ordered set observation and computed observation values for a given value of F_i, respectively. F_i are calculated using the Hosking [10] plotting position formula:

$$F_i = \frac{(i - 0.35)}{n}, \qquad i = 1,2,\dots n \quad (37)$$

…where n denotes the sample size and i is the observation in ascending order. The smallest value of RMSE for a given distribution indicates that distribution is more suitable to the actual data. RMSE previously has been used in some studies such as [33].

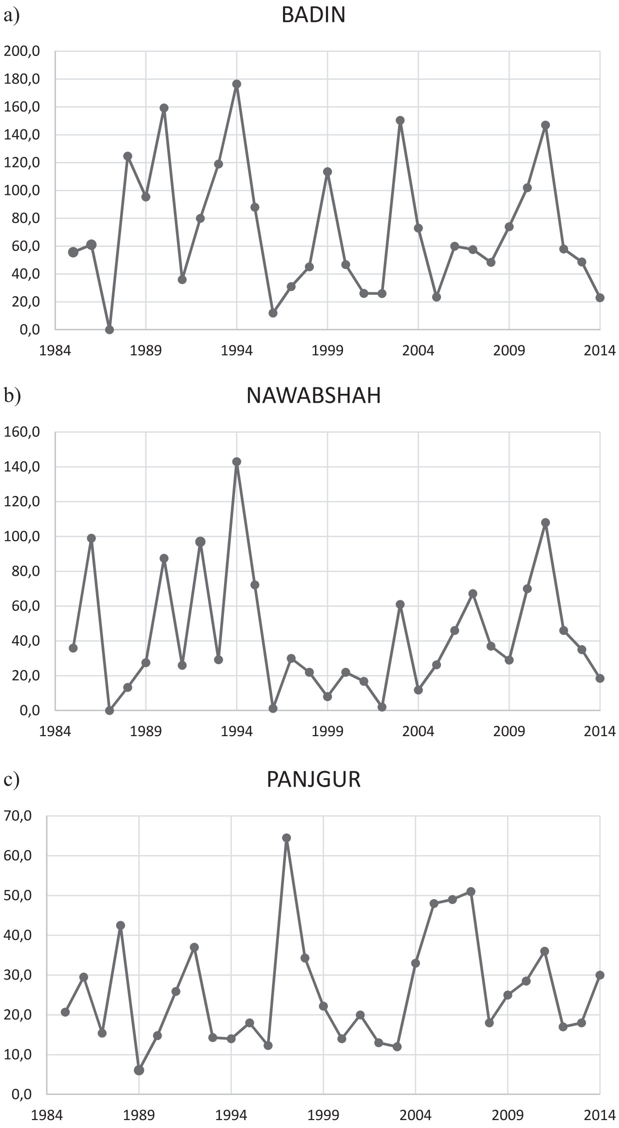

Fig. 2 (a-c). Time series plots for Badin, Nawabshsh, and Panjgur.

Results and Discussions

The basic assumption of hydrology data are randomness/stationarity, independence, and homogeneity. These assumptions are very important before further analysis; the final results may be misleading without fulfilling these assumptions. We tested these assumptions using such as Mann Kendall's test for trend detection and Mann-Whitney test for homogeneity with respect to location parameters (and sometimes also used for stationarity), and for independence we applied the Lag1 correlation test. Furthermore, to inspect randomness we also performed one sample runs test and also presented time series plots for visual inspection. Two sample runs tests can also be used for homogeneity assumption. In this study we used only one sample runs test. The results for these tests are given in Table 2. All of the assumptions are fulfilled, and the available data can be used for further FA. Time series plots of Badin, Nawabshah, and Panjgur sites are shown in Fig. 2.

L-moments and their variants are also useful in providing a graphical inspection in the form of ratio diagrams. In this study we included LRD. On the basis of LRD it was observed that GEV, GLO, GPA, P3, and LN3 distributions are in close agreement to the sample data as displayed in Fig. 3. In LRD, the relationship of L-skewness and L-Kurtosis of two parameter probability distributions are represented by dots, and three parameter probability distributions in the form of curves.

On visual inspection of LRD we found that GLO, GEV, GPA, PE3, and LN3 probability distributions are reasonable choices for the given data. No two parameter distributions show a close affiliation with data. Results of KST and ADT tests also indicate that mostly stations follow GEV, GPA, GLO, P3, and LN3 distributions using L-moments, TL-moments, and LH-moments as estimation methods. Results of all GOF measurements are given

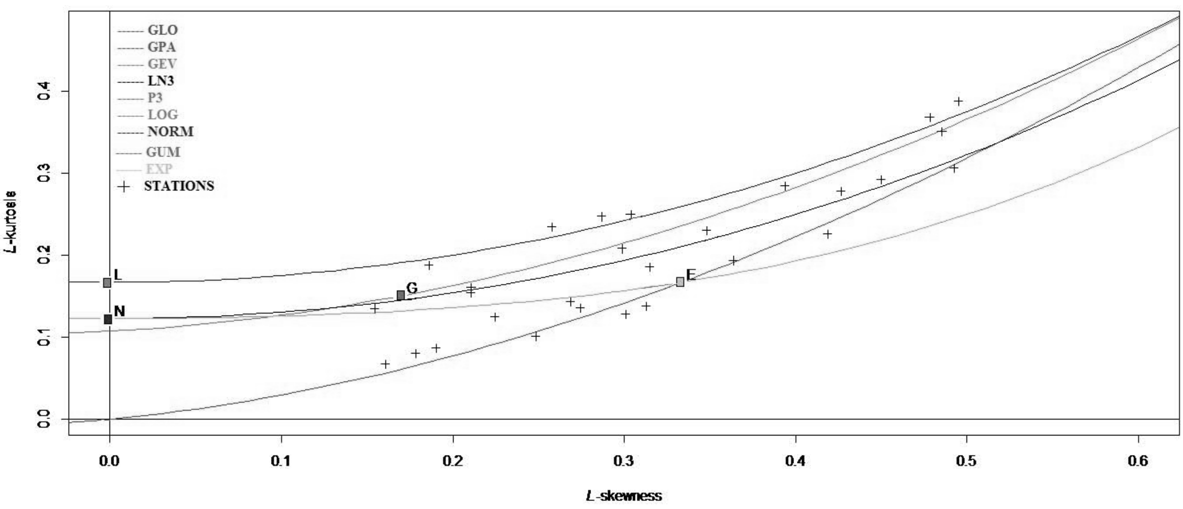

Fig. 3. L-moment ratio diagram for different distributions.

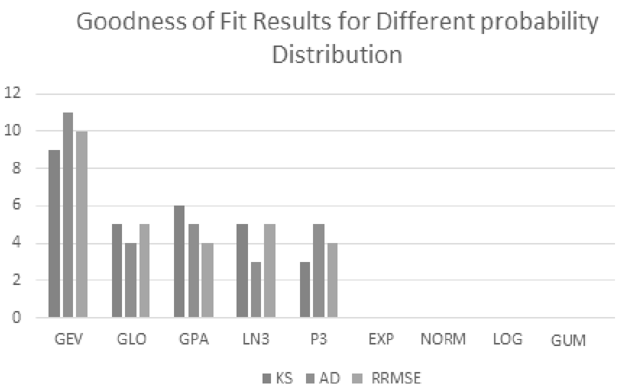

Fig. 4. Results of different goodness-of-fit tests.

in Table 3 using L-moments and their variants. Under KST, these results indicate that GEV can be declared as best fit distribution for the majority of stations, followed by GPA, GLO, LN3, and P3 distributions. Results of ADT and RMSE can also be explained in the same way. Different GOF measurement results are not exactly same. Comparison among three GOF measures is obvious in Fig. 4. From Fig. 4 it is not easy to find a best fit distribution for a particular station. For this purpose we present Table 3.

Different methods of parameter estimations change the results of the goodness-of-fit tests to some extent [8].

From Table 3 it is obvious that L-moments is not the only estimation method for all stations. TL- and LH-moments at different levels are also found to be most

Table 3. Results of different goodness-of-fit tests using L-moments and their variants.

Site Names	KS ***	AD**	RMSE*	Ratio diagram	Best Distribution	Best method of estimation for η = 0,1 2, 3 and 4
BADIN	GEV	GEV	GEV	GPA	GEV	$\eta = 1$
BARKHAN	GEV	GEV	GEV	GEV	GEV	$\eta = 3$
CHHOR	GPA	GPA	GPA	GPA	GPA	$\eta = 0$(L-Moments)
HYDERABAD	GEV	GEV	GEV	GEV	GEV	$\eta = 4$
JACOBABAD	GLO	GLO	GLO	GLO	GLO	$\eta = 0$(L-Moments)
JIWANI	GLO	GLO	GLO	GLO	GLO	$\eta = 0$(L-Moments)
KARACHI	GEV	GEV	GEV	GPA	GEV	$\eta = 0$(L-Moments)
KHUZDAR	LN3	GEV	LN3	LN3	LN3	$\eta = 1$
LASBELLA	GLO	GLO	GLO	GLO	GLO	$\eta = 3$
MOHEN-JO-DARO	P3	P3	P3	P3	P3	$\eta = 2$
NAWABSHAH	GPA	GPA	GPA	GPA	GPA	$\eta = 4$
PADIDAN	P3	P3	P3	P3	P3	$\eta = 3$
PANJGUR	GEV	GEV	GEV	GEV	GEV	$\eta = 3$
LAHORE	GLO	GLO	GEV	GLO	GLO	$\eta = 4$
PASNI	LN3	LN3	LN3	LN3	LN3	$\eta = 0$
ROHRI	LN3	LN3	LN3	LN3	LN3	$\eta = 0$
ZHOB	GPA	GPA	GPA	GLO	GPA	$\eta = 2$
FAISALABAD	LN3	P3	P3	GPA	P3	$\eta = 1$
KHANPUR	P3	P3	GLO	P3	P3	$\eta = 1$
BUNJI	GPA	GPA	GPA	GLO	GPA	$\eta = 1$
CHERAT	LN3	LN3	LN3	LN3	LN3	$\eta = 1$
CHILLAS	GEV	GEV	GEV	GPA	GEV	$\eta = 0$
CHITRAL	GEV	GEV	GEV	GEV	GEV	$\eta = 0$
DI KHAN	GPA	P3	GPA	GPA	GPA	$\eta = 3$
DROSH	GEV	GEV	GLO	GEV	GEV	$\eta = 3$
GILGIT	GEV	GEV	GEV	GEV	GEV	$\eta = 2$
JHELUM	GPA	GPA	GPA	GPA	GPA	$\eta = 4$
SIALKOT	GLO	GEV	GLO	GLO	GLO	$\eta = 2$

Table 4. Estimated quantiles for different return periods.

Station Name	Etas	Best Dist.	0.100 1	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500	0.990 1000
BADIN	$\eta = 1$	GEV	38.21	65.82	93.72	113.2	132.8	159.3	180.1	231.7	255.4
BARKHAN	$\eta = 3$	GEV	36.37	44.69	54.31	61.79	69.94	82.11	92.61	122.3	137.8
CHHOR	$\eta = 0$	GPA	24.83	63.96	103.7	131.6	159.6	197.8	228.0	302.9	337.5
HYDERABAD	$\eta = 4$	GEV	35.2	44.49	55.94	65.31	75.94	92.57	107.6	152.8	178.0
JACOBABAD	$\eta = 0$	GLO	6.178	34.49	79.52	124.7	184.9	298.1	420.0	901.0	1241
JIWANI	$\eta = 0$	GLO	5.266	31.27	65.33	94.65	129.2	185.6	238.8	409.8	510.8
KARACHI	$\eta = 0$	GEV	8.191	43.07	77.24	100.5	123.4	153.8	177.1	233.1	258.1
KHUZDAR	$\eta = 1$	LN3	27.86	34.8	47.22	60.87	80.47	120.5	167.2	376.2	541.0
LASBELLA	$\eta = 3$	GLO	27.93	36.71	52.66	70.46	96.27	149.7	212.8	500.3	730.6
MOHEN-JO-DARO	$\eta = 2$	P3	14.9	28.47	44.34	56.8	70.49	91.1	109.0	160.4	187.6
NAWABSHAH	$\eta = 4$	GPA	27.47	39.72	52.66	62.06	71.73	85.31	96.32	124.8	138.5
PADIDAN	$\eta = 3$	P3	24.41	36.88	55.54	73.38	96.25	137.4	179.8	336.7	441.7
PANJGUR	$\eta = 3$	GEV	18.23	24.48	31.1	35.9	40.84	47.78	53.41	67.98	74.98
LAHORE	$\eta = 4$	GLO	69.28	85.21	103.3	117.1	132	153.9	172.6	224.3	250.9
PASNI	$\eta = 0$	LN3	8.464	27.62	50.23	68.13	87.90	117.8	144.1	220.0	260.6
ROHRI	$\eta = 0$	LN3	7.422	34.66	68.65	96.81	129.0	179.8	226.1	367.9	447.9
ZHOB	$\eta = 2$	GPA	26.8	32.46	41.51	50.65	62.89	86.01	111.0	210.4	281.3
FAISALABAD	$\eta = 1$	P3	42.89	57.03	73.2	85.7	99.22	119.3	136.4	184.6	209.5
KHANPUR	$\eta = 1$	P3	20.81	40.79	64.61	83.63	104.8	137.1	165.5	248.8	293.7
BUNJI	$\eta = 1$	GPA	18.16	26.58	35.31	41.55	47.9	56.69	63.72	81.57	90.00
CHERAT	$\eta = 1$	LN3	41.41	54.09	74.48	95.14	122.9	175.5	232.5	460.5	623.7
CHILLAS	$\eta = 0$	GEV	17.67	28.55	43.49	56.84	73.07	100.4	127.1	217.3	273.0
CHITRAL	$\eta = 0$	GEV	28.70	42.83	63.23	82.18	105.9	147.5	189.4	338.5	435.0
DI KHAN	$\eta = 3$	GPA	43.67	53.77	67.92	80.78	96.6	123.8	150.5	243	301.3
DROSH	$\eta = 3$	GEV	40.24	46.85	57.39	67.99	82.13	108.7	137.4	250.9	331.5
GILGIT	$\eta = 2$	GEV	15.88	23.24	31.34	37.42	43.84	53.12	60.86	81.74	92.18
JHELUM	$\eta = 4$	GPA	74.54	92.53	114.6	132.6	153	184.9	213.5	299.6	347.3
SIALKOT	$\eta = 2$	GLO	78.41	104.7	135.1	158.8	184.5	223.1	256.3	350.5	399.8

suitable for many stations. The results given in Table 3 show that by using different GOF measurements, including LRD, GEV is the most suitable probability distribution for nine stations, GPA for six stations, GLO for five stations, and LN3 and PE3 for four stations each. Out of 28 stations L-moments is found to be most suitable for eight sites when $\eta = 0$ and TL-moments is found to be most suitable for six sites, whereas higher order moments (LH-moments) are suitable for 14 sites with $\eta = 2$, $\eta = 3$, and $\eta = 4$.

The next step is estimating quantiles for different return periods T. The results of these quantiles for best fitted distributions are given in Table 4, which are important for policy implications in the country. They are also important for meteorologists and hydrologists in water resource management. Quantiles are estimated only for best fitted distributions at the selected level of trimming mentioned in Table 3. Extreme value plots were also drawn for best fitted distributions. The extreme value plot for only five stations are shown in Fig. 5 (a-e). It is observed from these plots that extreme rainfall (quantiles) calculated through best fitted distributions are in close agreement with observed ADMR series.

Conclusions

In this study, AFA was carried out for ADMR series in Pakistan using data of 28 meteorological stations. The set

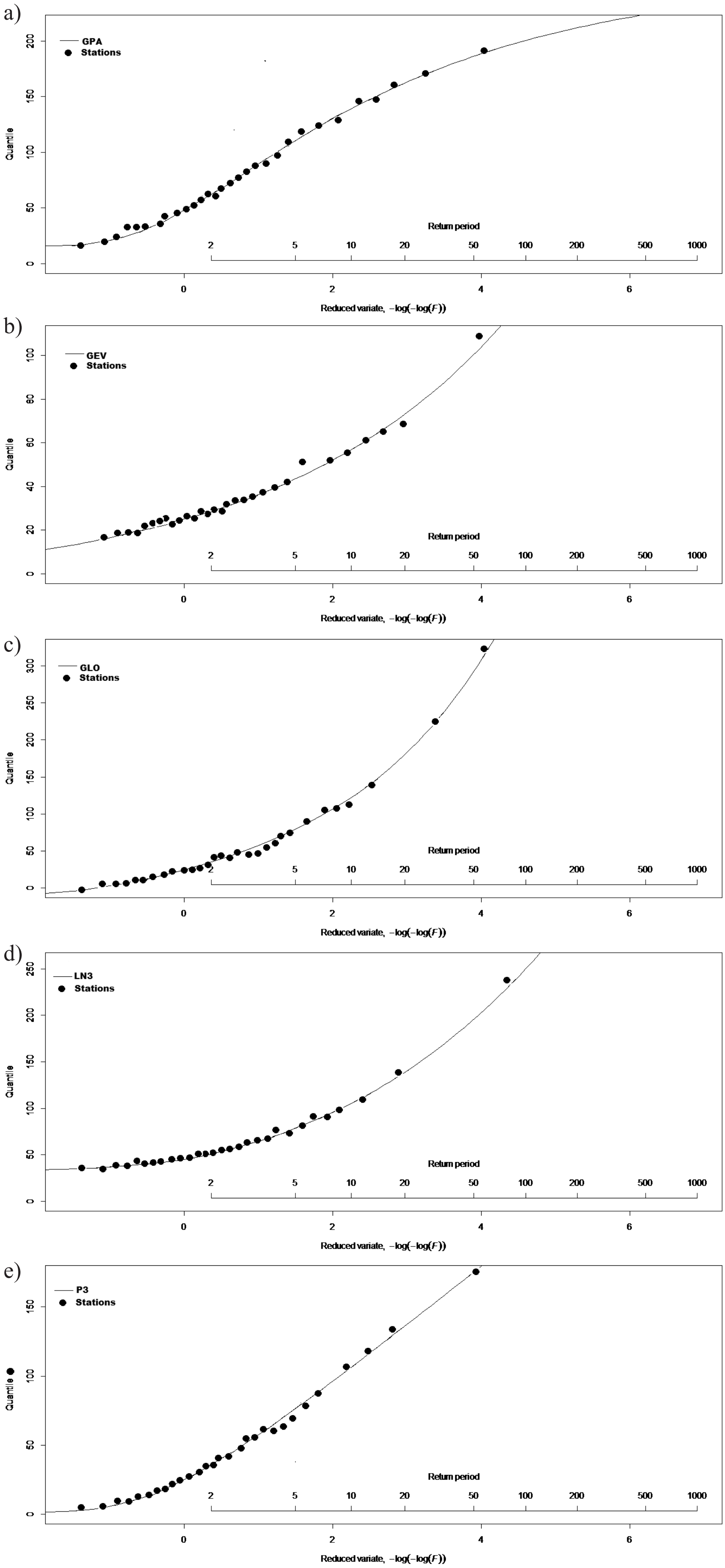

Fig. 5. Extreme value plot for: a) Chhor station, b) Chillas station, c) Jacobabad station, d) Cherat station, e) Kanpur station.

of most suitable distributions is selected on the basis of different GOF measurements such as KST, ADT, RMSE, and LRD. These GOF measurements not only identify the best probability distributions but also the paramount estimation method among L-Moments, TL-Moments, and LH-Moments. In practice, the purpose of analyzing ADMR series is to predict extreme rainfall for large return periods and to avoid the nuisance effect of smaller sample values in estimating upper quantiles. TL-moments and LH-moments should also be considered other than simple L-moments. Only five distributions (such as GEV, GLO, GPA, LN3, and P3) are declared the most suitable for different stations. L-moments, TL-moments with trimming (1, 0), LH-moments with level (η =2), (η =3), (η = 4) are suitable for 8, 6, and 14 stations, respectively. A theoretical relationship between TL-moments and LH-moments is also revisited, which persuades us that LH-moments are special cases of TL-moments, when trimming is performed only from the lower side. Awareness about rainfall modeling can be beneficial to take some precautionary measures to cope with such problems as loss of crops, human lives, and infrastructure due to the heavy rainfall. For any practical application in the future, such as planning for water related emergencies, sustainable water resources management and construction of different hydraulic structures, it is proposed that at least these five distributions should be considered and compared for final selection of best fit probability distribution for the ADMR series in Pakistan.

Acknowledgments

The authors thank to the Pakistan Meteorological Department Karachi for providing rainfall data for all possible sites. We are also grateful to the Higher Education Commission for providing us with financial assistance under project No. 20-3954R&D/HEC/14/305.

References

1. AHMAD I., MAHMOOD I., MALIK I.R., ARSHAD I.A., HAQ E., IQBAL Z. Probability analysis of monthly rainfall on seasonal monsoon in Pakistan. International Journal of Climatology, **34**, 827, **2014**.
2. ADNAN S., KHAN A.H. Effective rainfall for irrigated agriculture plains of Pakistan. Pak J Meteorology. **6**, 61. **2009**.
3. SHAMSHAD K.M. The Meteorology of Pakistan, First Edition, Royal Book Company Publishers. **1988**.
4. SHABRI A.B., DAUD Z.M., ARIFF N.M. Regional analysis of annual maximum rainfall using TL-moments method. Theor. Appl. Climatol, **104**, 561, **2011**.
5. KYSELY J., PICEK J., HUTH R. Formation of homogeneous regions for regional frequency analysis of extreme precipitaion events in the Czech Republic. Studia Geophysica et Geodaetica, **51**, 327, **2006**.
6. DURRANS S.R., KIRBY J.T. Regionalization of extreme precipitation estimates for the Alabama rainfall atlas. Journal of Hydrology, **295**, 101, **2004**.
7. NOTO L.V., LA LOGGIA G. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. Water resources management, **23**, 2207, **2009**.
8. RAHMAN A.S., RAHMAN A., ZAMAN M.A., HADDAD K., AHSAN A., IMTEAZ M. A study on selection of probability distributions for at-site flood frequency analysis in Australia. Natural hazards, **69**, 1803, **2013**.
9. BOBEE B., CAVADIAS G., ASHKAR F., BERNIER J., RASMUSSEN P. Towards a systematic approach to comparing distributions used in flood frequency analysis. Journal of Hydrology, **142**, 121, **1993**.
10. HOSKING R.M. L-Moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, Journal of Royal Statistical Society, **52**, 105, **1990**.
11. ELAMIR E.A.H., SEHEULT A.H. Trimmed L-Moments, Computational Statistics & Data Analysis, **43**, 299, **2003**.
12. WANG Q. J. Using Higher Probability Weighted Moments for Flood Frequency Analysis, Journal of Hydrology, **194**, 95, **1997**.
13. HOSKING J.R.M., WALLIS J.R. Regional frequency analysis: An approach based on L-moments, Cambridge University Press, UK, **1997**.
14. MESHGI A., KHALILI D. Comprehensive evaluation of regional flood frequency analysis by L-and LH-moments. I. A re-visit to regional homogeneity. Stochastic Environmental Research and Risk Assessment, **23**, 119, **2009a**.
15. MESHGI A., KHALILI D. Comprehensive evaluation of regional flood frequency analysis by L-and LH-moments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions. Stochastic Environmental Research and Risk Assessment, **23**, 137, **2009b**.
16. SHABRI A., ARIFF N. Frequency analysis of maximum daily rainfalls via l-moment approach. Sains Malaysiana, **38**, 149, **2009**.
17. DEKA S., BORAH M., KAKATY S.C. Distribution of Annual Maximum Rainfall Series of North-East India. European Water (EW) Publications **27**, 3, **2009**.
18. DAUD Z.M., KASSIM A.H.M., DESA M.N.M., NGUYEN V.T.V. Statistical analysis of at-site extreme rainfall processes in peninsular Malaysia. International Association of Hydrological Sciences. **274**, 61, **2002**.
19. HAKTANIR T., COBANER M., KISI O. Frequency analyses of annual extreme rainfall series from 5 min to 24 h. Hydrological processes, **24**, 3574, **2010**.
20. AL MAMOON A., JOERGENSEN N.E., RAHMAN A., QASEM H. Derivation of new design rainfall in Qatar using L-moment based index frequency approach. International Journal of Sustainable Built Environment. **3**, 111, **2014**.
21. ABDUL-MONIEM I., SELIM Y.M. TL-moments and L-moments estimation for the generalized Pareto distribution. Applied Mathematical Sciences, **3**, 43, **2009**.
22. AHMAD I., ABBAS A., ASLAM M., AHMAD I. Total Annual Rainfall Frequency Analysis in Pakistan using Methods of L-Moments and TL-Moments. Sci. Int. (Lahore), **27**, 2331, **2015**.
23. MURSHED M.S., AM SEO Y., PARK J.S. LH-moment estimation of a four parameter kappa distribution with hydrologic applications. Stochastic Environmental Research and Risk Assessment, **28**, 253, **2014**.
24. DURRANS S., TOMIC S. Comparison of Parametric tail Estimation for Low-Flow frequency Analysis. JAWRA Journal of the American Water Resources Association, 37 (5), 1203, **2001**.
25. ESLAMIAN S.S., GOHARI S.A., BIABANAKI M., MALEKIAN R. Estimation of monthly pan evaporation using artificial neural networks and support vector machines. J Appl Sci, **19**, 3497, **2008**.

26. TOLIKAS K., GETTINBY G.D. Modelling the distribution of the extreme share returns in Singapore. Journal of Empirical Finance, **16**, 54, **2009**.
27. MORGAN E.C., LACKNER M., VOGEL R.M., BAIS, L.G. Probability distributions for offshore wind speeds. Energy Conversion and Management, **52**, 15, **2011**.
28. BÍLKOVA D. Robust parameter estimations using L-moments, TL-moments and the order statistics. American Journal of Applied Mathematics, **2**, 36, **2014**.
29. ABDUL-MONIEM. L-moments and TL-moments estimation for the Ex-ponential distribution, Far East J.Theor. Stat, **23**, 51, **2007**.
30. ABDUL-MONIEM, I. B.TL-moments and L-moments estimation for the Weibull distribution. Advances and Applications in Statistics, **15**, 83, **2010**.
31. HOSKING J.R.M. Some theory and practical uses of trimmed L-moments. Journal of Statistical Planning and Inference. **137**, 3024, **2007**.
32. ASQUITH W.H. L-moments and TL-moments of the generalized lambda distribution. Computational Statistics & Data Analysis. **51**, 4484, **2007**.
33. SHABRI A. Comparisons of the LH Moments and the L Moments, Matematika, **18**, 33, **2002**.
34. GAMAGE W., S.H.P., HEWA G.A., SUBHASHINI W.H.C., DANIELL T.M., KEMP D. Modelling the extreme floods of South Australian catchments, MODSIM, 3435, **2009**.
35. MURSHED M.S., PARKB B.J., JEONGA B.Y., PARK J.S. LH-Moments of some distributions useful in Hydrology, Communications of the Korean Statistical Society, **16**, 647, **2009**.
36. DEKA S., BORAH M., KATATY S.C. Statistical Analysis of Annual Maximum Rainfall in North-East India: an approach of LH-moments, Theor Appl clim, **104**, 111, **2011**.
37. JAN N.A.M. SHABRI A. Estimating distribution parameters of annual maximum streamflows in Johor, Malaysia using TL-moments approach. Theoretical and Applied Climatology. **1**, **2015**.
38. AHMAD I., FAWAD M., MAHMOOD I. At-Site Flood Frequency Analysis of Annual Maximum Stream Flows in Pakistan Using Robust Estimation Methods. Polish Journal of Environmental Studies, **24**, 2345, **2015**.
39. VOGEL R.M., FENNESSEY N.M. L moment diagrams should replace product moment diagrams. Water Resources Research, **29**, 1745, **1993**.
40. CUNNANE C. Statistical distributions for flood frequency analysis, Operational hydrological Report No. 5/33, World Meteorological Organization (WMO), Geneva, Switzerland. **1989**.
41. CALENDA G., MANCINI C. P., VOLPI E. Selection of the probabilistic model of extreme floods: The case of the River Tiber in Rome. Journal of Hydrology. **371**, 1, **2009**.
42. HADDAD K., RAHMAN A. Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. Stoch. Env. Res. Risk Assess, **25**, 415, **2011**.
43. LAIO F., DI BALDASSARRE G., MONTANARI A. Model selection techniques for the frequency analysis of hydrological extremes. Water Resources Research, **45**, **2009**.
44. BESKOW S., CALDEIRA T.L., DE MELLO C.R., FARIA L.C., GUEDES H.A.S. Multiparameter probability distributions for heavy rainfall modelling in extreme southern Brazil. Journal of Hydrology: Regional Studies. **4**, 123, **2015**.
45. SHARMA M.A., SINGH J.B. Use of probability distribution in rainfall analysis. New York Science Journal, **3**, 40, **2010**.
46. AHMAD I., SHAH S.F., MAHMOOD I., AHMAD Z. Modeling of monsoon rainfall in Pakistan based on Kappa distribution. Sci. Int. (Lahore), **25**, 333, **2013**.
47. SALARPOUR M., YUSOP Z., YUSOF F. Comparison of Distribution Models for Peak flow, Flood Volume and Flood Duration. Research Journal of Applied Sciences, Engineering and Technology, **6**, 733, **2013**.