Regional Frequency Analysis of Annual Total Rainfall in Pakistan Using L-Moments

Ishfaq Ahmad[1], Aamir Abbass[2], Muhammad Fawad[3], Aamir Saghir[4]

[1]Department of Mathematics and Statistics, International Islamic University Islamabad, Pakistan
[2]Department of Mathematics, Universty of Poonch, Rawalakot,AJK, Pakistan
[3]School of Mathematics and Statistics, Central China Normal University, Wuhan, China
[1]Department of Mathematics, Mirpur University of Science and Technology(MUST), Mirpur,AJK, Pakistan
Received: 10 September 2015 Accepted: 24 February 2017

Abstract

This study presents the estimation of regional quantiles of Annual Total Rainfall (ATR) for 30 meteorological stations using index flood procedure based on L-moments. Discordancy measure based on L-moments was used to screen the ATR. As a matter of the fact, highly elevated areas of Pakistan receive more rainfall, the study area was divided into four different regions. Further to justify the homogeneity of these regions, L-moment based heterogeneity measure (H) was calculated for each region using four parameter Kappa distribution. For each region, best distribution was found among Pearson type III (PE3), Generalized Normal (GNO), Generalized Extreme Value (GEV), Generalized Logistic (GLO) and Generalized Pareto (GPA) distribution using Z-statistic and L-moment ratio diagram. Regional quantiles on the basis of best-fit distribution for each region were determined and further for robustness, the accuracy measures for the estimated regional quantiles were calculated using Monte Carlo simulations. It was found that PE3 was most suitable choice for large return period for first three regions while for small return period GNO and GEV. Similarly, for region IV GEV was declared as the best fit for lower return period's up to 20 years while for a period of 50, 100, 500 and 1000 years GNO was the best one.

Keywords: Extreme events, L-Moments, Regional Frequency Analysis, Return periods, quantiles estimates

Introduction

The climate of any country is evaluated on the basis of its weather conditions. Pakistan is located from south-west to Northwest at 23-37degree north latitude and 61-76-degree east latitude. Pakistan has been facing the natural disasters among others such as droughts, floods and storms. These massive distributions are caused naturally and cannot be mitigated completely. Although the losses could be minimized to some or more extent by proper planning. The main sources of rainfall in Pakistan are the summer monsoon, the western depression and the thunderstorms. In July to September there is a plenty of rainfall. In Pakistan, there is an interlude and dynamic network of canals and irrigation distributaries but on the other hand, there is poor drainage system. Floods result in tremendous destruction in the country. Estimation of frequency and magnitude of the floods is of immense importance in policy implications in water resources management. Regional Frequency analysis is highly important for the construction of different hydrological structures in the country. A lot of literature is available on the regional frequency analysis of extreme rainfall. The regional frequency analysis of extreme rainfall in England by Bilham [1] and regional frequency analysis of daily precipitation in Nigeria by Ayoade [2] are the early examples. Adamowski et al [3] worked on the data collected from Canada for regional frequency analysis and identified GEV as a parent distribution. Parida [4] studied the random behaviour of the Indian summer monsoon using kappa distribution and reliable rainfall quantiles were estimated both for small and large return period. Smithers et al [5] identified 15 homogeneous regions and estimated short duration storms in South Africa and also computed error bounds for the quantile growth curves for each station of the clusters. Park et al [6] used Wakeby distribution to obtain reliable quantile estimates on summer extreme rainfall data by the method of L-moments. Lee et al [7] used L-moments ratio diagram and the Kolmogorov–Smirnov (K–S) test to judge appropriate distribution of annual maximum daily rainfall and found GEV and GLO to be the best. Yurekli [8] used discordancy measure, heterogeneity test and Z-statistic to perform a regional frequency analysis of monthly rainfall for two regions in Amasya province by using L-moments. Trefry et al [9] performed regional frequency analysis approach based on L-moments for annual maximum series and partial duration data. Koh et al [10] designed five regions using K-mean Clustering method and identified regional probability distribution using L-moments ratio diagram and the Kolmogorov–Smirnov (K–S) test and relative root mean square, relative bias. Shabri and Ariff [11] used different distributions for maximum daily rainfall data and determined the best distribution on the basis of mean absolute deviation index, mean square deviation index and L-moments ratio diagram. Hussain and Pasha [12] performed regional flood frequency analysis on different stations of Punjab in Pakistan and estimated quantiles for different return periods using L-moments technique. Ngongondo et al [13] used cluster analysis, method of L-moments and Monte Carlo simulations to study the regional frequency analysis of rainfall extremes in flood-prone regions of South Malawi. Hassan and Ping [14] used cluster analysis and L-moment approach to quantify regional rainfall pattern of Luanhe basin and apply the goodness-of-fit test to study regional frequency analysis. Ahmad et al [15] analysed the random behaviour of monsoon rainfall in Pakistan using Kappa distribution and calculated quantiles for different return periods from 2 to 500 years. Devi and Choudhury [16] performed extreme rainfall frequency analysis using L-moment approach and best-fitted distribution was selected on the basis of goodness-of-fit statistic and L-moment ratio diagram. and Similarly Shahzadi et al [17] also applied regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan, best-fitted distributions were

selected on the basis of L-moment ratio diagram, goodness-of-fit statistic. The study is different from the previous study such as [17] in the country in the sense we are using the data of ATR from all over the country. The comparison L-Moments with different estimation methods were also studied see for example [18-20]. The remaining of the paper is as follows: section 2 is about the methodology of regional frequency analysis, section 3 presents discussions and section 4 concludes the paper.

Methods and Materials

Data Description

The daily rainfall data, measured in millimetres, were taken from Pakistan Meteorological centre Karachi from which annual total rainfall (ATR) series, a total of 12 months' rainfall, were constructed for the proposed study. The geographical location of these meteorological observatories was given in Fig.1. These sides included the province Sindh, Punjab, Baluchistan and northern areas of Pakistan. The record length of ATR series ranged from 29 to 51 years.

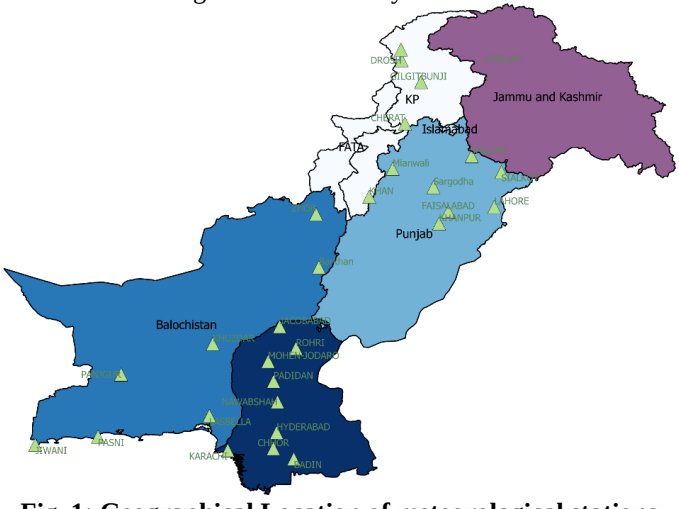


Fig. 1: Geographical Location of meteorological stations

Method of L-moments

L-moments are defined as linear combinations of ordered statistics that are used to summarize the data and the probability distributions. L- Moments can be written in the form of probability weighted moments (PMWs) where PWMs are defined as:

$$\beta_r = E[x\{F(x)\}^r] \quad (1)$$

The rth PWMs as given below.

$$\lambda_{r+1} = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \quad (2)$$

The first four PWMs and L-ratios are as follows

$$\lambda_1 = \beta_0 \quad (3)$$
$$\lambda_2 = 2\beta_1 - \beta_0 \quad (4)$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (5)$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (6)$$
$$\tau = \lambda_2/\lambda_1 \quad (7)$$
$$\tau_3 = \lambda_3/\lambda_2 \quad (8)$$
$$\tau_4 = \lambda_4/\lambda_2 \quad (9)$$

Hosking [21] defined sample L-moments, denoted by l_1,l_2,l_3,l_4 and sample L-moments ratios are denoted by t,t_3,t_4 where t is L-coefficient of variation, t_3 is L-skewness and t_4 L-kurtosis.

L-moments have many advantages over conventional moments. They are more robust especial in the presence of small samples and outlier. Asymptotic estimations applied to sampling distributions are more helpful for L-moments as compared to ordinary moments. L-moments provide a better tool to identify the parent distribution generating data sample [21].

Empirical Study using Regional Frequency Analysis

Before we analyze the data, initially we check the basic assumptions of regional frequency analysis which are stationary, independence, and homogeneity of data. First, we apply time series plot to identify the patterns in the ATR series. Generally, the ATR series of all 30 station showed that there is no systematic jumps or trend. Next for Stationarity, independence and homogeneity we apply Ljung-Box Q-statistics, lag-one correlation, Kendall's tau test and Mann- Whitney U tests respectively. The results of all the 30 station show Stationarity, independence and homogeneity.

All of the stations satisfied these tests for basic assumptions, so these stations were appropriate for regional frequency analysis. Regional Frequency Analysis was based on following four steps

1) Initial Screening of Data
2) Identification of homogeneous region
3) Choice of an appropriate probability distribution
4) Estimation and robustness of regional quantiles.

Initial Screening of Data

In any statistical analysis of data, the fundamental step was to check that whether the data were suitable for analysis or not. The purpose of initial screening of the data was to show and identify the discordant station. Discordancy measure of station i defined by Hosking and Wallis [22] was as follows:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (10)$$

Where

$$A = \sum_{i=1}^{N} (u_i - \bar{u})\,(u_i - \bar{u})^T$$
$$u_i = \left[t^{(i)}\ t_3{}^{(i)}\ t_4{}^{(i)}\right]^T$$
$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i$$

In screening of the data, N represents number of stations in the region and statistic D_i was calculated for each station. The values of D_i for all stations are smaller than the critical value, so we considered that there was no trend or no outlier in any station. The largest value of D_i was 2.53 with high L-CV and L-skewness. The L-CV and L-skewness of annual total rainfall of 30 stations in Pakistan are shown in table 1.

As we can see that discordancy measure is based on three coefficients L-CV, L-Skewness and L-Kurtosis, overall data depicting a correct sense and there was no evidence of gross errors. Station 12, Padidan has extremely high values of L-CV and L-skewness, but it was not discordant with other station. In Fig. 2 scatter plots are drawn for sample skewness versus sample CV and sample skewness versus sample kurtosis. Fig. 2

Table 1: L-moment Ratios and Discordancy Measures

Station No	Station Name	n_i	t	t_3	t_4	D_i
1	Badin	43	0.4089	0.2349	0.1764	0.44
2	Choor	43	0.3508	0.1321	0.1313	0.82
3	Barkhan	30	0.2135	0.1217	0.1471	0.15
4	Hyderabad	43	0.4480	0.2515	0.1104	0.83
5	Jacobabad	43	0.4635	0.3082	0.1414	0.84
6	Jiwani	43	0.4620	0.2857	0.1289	0.78
7	Karachi	43	0.4108	0.1535	0.0491	2.20
8	Khuzdar	43	0.2417	0.1180	0.2180	1.44
9	Lasbella	31	0.3511	0.1956	0.2010	0.71
10	Moen-jo-doro	30	0.4823	0.3402	0.1461	1.22
11	Nawabshah	43	0.4314	0.2403	0.1688	0.55
12	Padidan	30	0.5029	0.3719	0.2238	1.75
13	Panjgar	30	0.3369	0.2572	0.2271	0.95
14	Pasni	43	0.3819	0.1408	0.0771	1.39
15	Rohri	43	0.4590	0.2949	0.1979	0.86
16	Zhob	30	0.1776	0.0295	0.1540	1.23
17	Bunji	37	0.2250	0.1303	0.1400	0.13
18	Cherat	43	0.1613	0.0674	0.1748	0.78
19	Chilas	30	0.2315	0.2736	0.2354	2.53
20	Chitral	43	0.1632	0.0384	0.1693	1.18
21	DI Khan	43	0.2077	0.1460	0.1617	0.26
22	Drosh	43	0.1336	0.0383	0.1711	1.08
23	Faisalabad	43	0.2054	0.1288	0.1411	0.22
24	Gilgit	41	0.2058	0.1569	0.1350	0.53
25	Jhelum	30	0.1365	0.0362	0.0880	1.09
26	Khanpur	30	0.4080	0.2282	0.1474	0.31
27	Lahore	43	0.1912	0.1808	0.1429	1.10
28	Mianwali	29	0.1774	0.1017	0.1211	0.45
29	Sargodha	51	0.1621	0.0761	0.0477	2.18
30	Sialkot	43	0.1829	0.2063	0.1442	1.98

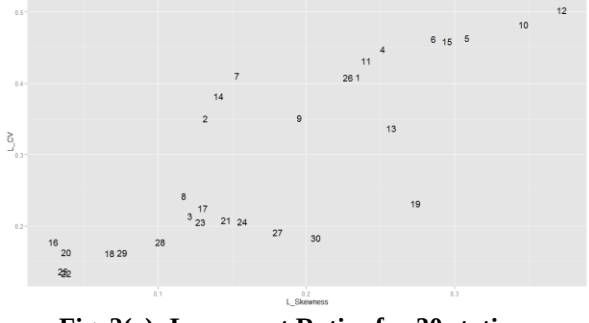


Fig. 2(a): L-moment Ratios for 30 stations

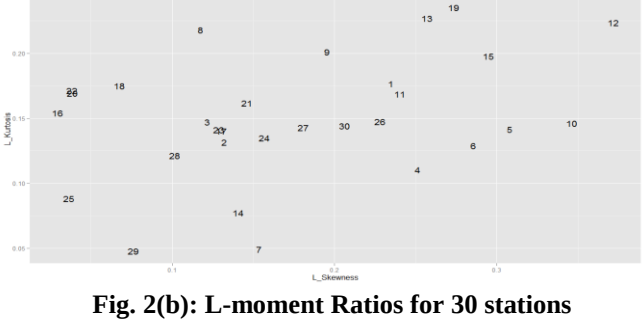


Fig. 2(b): L-moment Ratios for 30 stations

shows whether any station is discordant or not. From Fig. 2 station 12 Padidan appeared as the station having high L-skewness as compared to other stations. Similarly, station 19 Chilas appeared as the station having high L-kurtosis but these stations are not discordant.

In Fig. 3, station 12 Padidan was compared with its nearest station that was Moen-jo-doro (station 10). Due to extreme rainfall in the year 1992 and 1994 at station 12, the values of L-CV were large. So there was no clear reason to discard station 12.

Initial Screening of Data

The most difficult step in regional frequency analysis is the identification of homogeneous regions which demands the greatest interest of subjective judgment. By homogeneity condition, it is meant that stations have the same geographical characteristics and constitute a family of similar stations. This condition is achieved by dividing the stations into disjoint groups. The objective of heterogeneity measure is to calculate the degree of heterogeneity in a group and also to examine whether stations are considered as homogeneous or not. Heterogeneity measure,H_j (j=1,2,3), can be calculated as:

$$H_j = \frac{(V_j - \mu_{vj})}{\sigma_{vj}} \qquad (11)$$

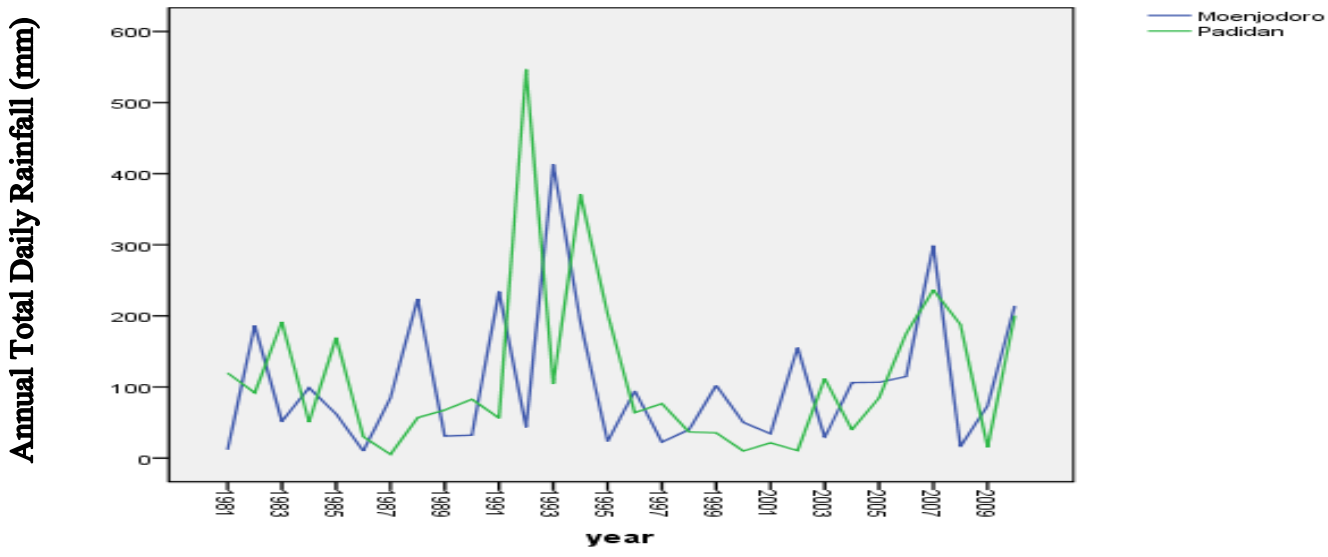


Fig. 3 Time Series data for Padidan (Station 12) and Moen-jo-doro (Station 10)

μ_v and σ_v are the mean and standard deviation of the value of V determined from simulations

Where

$$V = \left\{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^{R}\right)^2 / \sum_{i=1}^{N} n_i\right\}^{\frac{1}{2}} \quad (12)$$

The observed and simulated dispersion based L-moments ratios for different stations (under consideration) were compared by heterogeneity measure. For this purpose, Monte Carlo simulation was made using four parameter Kappa distribution. The Kappa distribution is defined as:

$$f(x) = \alpha^{-1}\left[1 - k(x-\varepsilon)/_{\alpha}\right]^{1/_{k-1}} [F(x)]^{1-h} \quad (13)$$

The location parameter for Kappa distribution is represented by ε, scale parameters by α, and shape parameters by k and h respectively. The possible variation of x is:

$\varepsilon + \alpha(1-h^{-k})/k \le x \le \varepsilon + \alpha/k \quad if\ k > 0, h > 0$

$\varepsilon + \alpha \log h \le x < +\infty \quad if\ h >, k = 0$

$\varepsilon + \alpha\,(1-h^{-k})/k \le x < +\infty \quad if\ h > 0, k < 0$

$-\infty < x \le \varepsilon + \alpha/k \quad if\ h \le 0, k > 0$

$\varepsilon + \alpha/k \le x < \infty \quad if\ h \le 0, k > 0$

The Kappa distribution is a flexible distribution which produces many distributions. In the formation of homogeneous regions, first of all, we considered a single region of 30 stations and the heterogeneity statistic was calculated. The value 0f〖 H〗_1 is 20.78 , as this value was greater than 2, so entire set of stations was definitely heterogeneous. So the possibility of treating a single region of 30 stations was rejected. Regional frequency analysis is not applied under this situation because we cannot fulfil the required criteria defined by [23]. This reference implies that region will be acceptably homogeneous if H_j<1, possibly heterogeneous if 1≤H_j<2 , and definitely heterogeneous if H_j>2. According to Hosking and Wallis [23] H_(1)is more efficient to detect homogeneity of region as the compare to H_2 andH_3. Table of heterogeneity statistic for a single region of 30 stations is given in table 2.

1) Formation of Homogeneous Regions

The study using the characteristic that a highly elevated station's receives high average annual rainfall and station's with low elevation receives low average rainfall. The scatter plot of elevation and average annual rainfall has been shown in Fig. 4 and the numbers in this plot indicate corresponding stations.

Table 2: Heterogeneity Statistics

No. of Stations	H_1	H_2	H_3
Heterogeneity statistic for 30 stations	20.78*	5.83*	0.48

The group of 30 stations are seemed to be subdivided into four homogeneous sub-regions by using these characteristics. We labelled these groups region I to IV. Now for each group the heterogeneity and goodness of fit measure were calculated.

Table 3 shows results of heterogeneity for the region I, II, III and IV. The heterogeneity statistics indicate that all these regions are homogeneous and none of any H_j>1.

Choice of an appropriate Probability Distribution

1) Goodness-of-fit statistic

The fitting of a required regional frequency distribution for each region, comprising of some stations, can be measured from goodness-of-fit statistics. In this study as test defined by Hosking and Wallis [23] i.e. Z^DIST statistic has been used, given as:

$$Z^{DIST} = \frac{(\tau_4^{DIST} - t_4^R + B_4)}{\sigma_4} \quad (14)$$

B_4 and σ_4 denotes simulated regional bias and standard deviation of t_4^R respectively. These simulations are constructed from fitted Kappa distribution to regional L-moments. The Kappa distribution is used as it is more flexible distribution, which can take different forms of other probability distributions as well [22]. The goodness-of-fit statistic Z^{DIST} calculated for all regions. Three parameter distributions, that are generalizes logistic (GLO), generalized Pearson type III (PE3), generalized extreme value (GEV), generalized normal (GNO) and generalized Pareto (GPA) have been considered in this regional analysis. The aim is not only to identify the best fitted distribution, but also to identify the distribution that will provides accurate quantile estimates for each region. Distribution with two parameters may cause bias in tail quantile

Table 3: Heterogeneity Statistics for Region I, II, III and IV

Heterogeneity Statistics for region I	Heterogeneity Statistics for region II	Heterogeneity Statistics for region III	Heterogeneity Statistics for region IV
H1= 0.30	H1= 0.81	H1= 0.67	H1= 0.98
H2= -0.12	H2= -0.30	H2= -1.06	H2= -0.32
H3= -0.58	H3= -0.71	H3= -1.87	H3= -0.79

Table 4: ZDIST – Statistics for Various Distributions

Region	ZGLO	ZGEV	ZGNO	ZPE3	ZGPA
1	3.79	2.04	1.30 *	-0.06 *	-2.28
2	2.93	0.73 *	0.73 *	0.34 *	-3.90
3	0.32 *	-1.69	-1.40 *	-1.48 *	-5.59
4	-0.23 *	-1.37 *	-1.52 *	-1.92	-3.91

estimation if the shape of the tail of the true frequency distribution is not well approximated by the fitted distribution which was defined by [23]. Regions for which distributions gave acceptable fit are given in the table 4 that is $|Z^{DIST}| \leq 1.64$. In table 4 |ZDIST| - statistic has been calculated for four regions from which we conclude that GPA is unsuitable choice of distribution for these four regions so we eliminate this distribution. We declare the fit to be adequate if |Z^DIST | is sufficiently close to zero [23]. GNO and PE3 are best fitted the distribution for region -I and the most suitable is PE3 because its ZDIST is closer to zero. For region II GEV, GNO and PE3 are best fitted distributions but the most suitable is PE3 while for region III GLO, GNO and PE3 and for region IV GLO, GEV and GNO are the most suitable choice of distribution but GLO for both regions III and IV is the appropriate choice.

2) L-moment ratio diagram

L-moment ratio diagram is one of the most common methods used to measure the goodness-of-fit of distribution. The graphical representation of L-skewness versus L-kurtosis is known as L-moment ratio diagram. L-moment ratio diagram is a useful guideline for the selection of an appropriate distribution for describing a distinct relationship between L-moment ratios exists for each theoretical probability distribution and sample data defined by Hosking and Wallis [23]. L-moment ratio regional average L-skewness and L-kurtosis (AVG) lies closest to the PE3 distribution similarly for region II, III, IV regional average L-skewness and L-kurtosis (AVG) lies closer to PE3, GLO and GLO respectively.

Estimation of Homogeneous Region and regional growth curve

For a given station i the quantile estimates are obtained by changing the estimate of index flood μ_i and quantile function of $\hat{q}(.)$ The nonexceedance probability F for the estimate of the quantile is written as

$$\hat{Q}(F) = \mu_i \hat{q}(F) \qquad (15)$$

After the selection of best fitted regional distributions, we estimate regional quantile, for these four regions which are given in table 5. The values given in table 5a can be explained as for example for region I $\hat{q}_{GNO}(0.980) = 3.259$ is that amount of rainfall which will happen once in 50 years i.e. probability of occurrence in any year for the given return period of 50 years and is 3.259 time larger than its average for all stations in region I for the given return period. Estimated quantiles for all these regions can be interpreted in the same way.

In fig. 6 we construct regional growth curves for all these four regions. In regional frequency analysis, it was assumed that the stations have a common frequency distribution. One representation of this common distribution is the regional

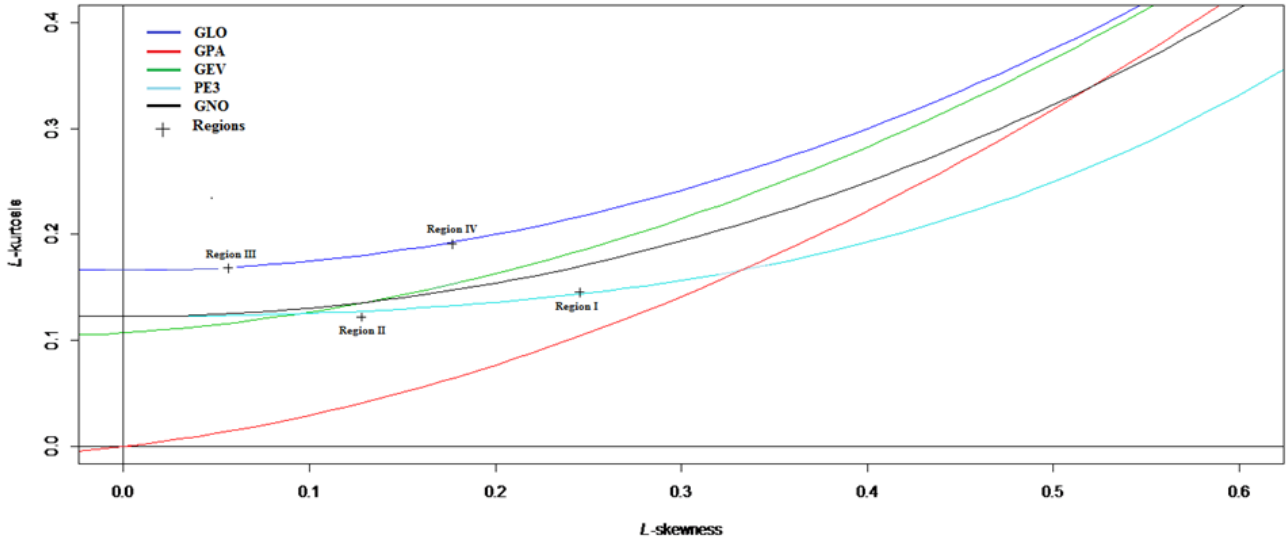


Fig. 4: L-moment ratio diagram for four regions

diagram for four regions was illustrated in Fig. 5. For region I growth curve, which specifies the quantiles that correspond to each non-exceedance probability. So the growth curve summarizes the common properties of the frequency distributions at all stations. The growth curve for the region I in fig. 6a is looking same up to the return period of 50 years, as we increase the return period up to 1000 years there is a small difference between these two distributions that are GNO and PE3. So we conclude that for the region I GNO and PE3 shows the reliable result for low return period 1, 2, 5, 10, 20 and 50 years. In fig. 6b regional growth curve for region II shows same values of quantiles up to return period of 1000. Fig. 6c for region III shows that all selected distributions show almost same results up to return period of 50 but as returned period increases up to 1000 the curve of GLO moves in an upward direction which means that the quantiles of GLO are high. Similarly, for region IV in fig. 6d set of all candidate distributions shows approximately same results up to return period of 50 when return period increases up to 1000 regional growth curve of GLO moves upward which indicates quantiles of GLO are high.

Table 5(a): Regional Quantile Estimates for four regions

	Parameters			Regional quantile estimates with nonexceedance probability F							
Dist	ε	α	k	*0.100 **1	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500	0.999 1000
GNO	0.818	0.681	-0.498	0.173	0.819	0.819	2.042	2.557	3.259	3.814	5.195	5.837
PE3	1.000	0.807	1.448	0.168	0.812	1.564	2.078	2.570	3.199	3.664	4.719	5.166

Table 5(b): Regional Quantile Estimates for four regions

	Parameters			Regional quantile estimates with nonexceedance probability F							
Dist	ε	α	k	0.100 1	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500	0.999 1000
GEV	0.856	0.278	0.064	0.618	0.957	1.254	1.439	1.608	1.815	1.963	2.279	2.405
GNO	0.957	0.313	-0.26	0.617	0.958	1.253	1.436	1.604	1.814	1.967	2.312	2.459
PE3	1.000	0.329	0.786	0.615	0.957	1.257	1.440	1.605	1.805	1.949	2.262	2.391

Table 5(c): Regional Quantile Estimates for four regions

	Parameters			Regional quantile estimates with nonexceedance probability F							
Dist	ε	α	k	0.100 1	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500	0.999 1000
GLO	0.981	0.166	-0.067	0.642	0.981	1.223	1.375	1.523	1.721	1.877	2.264	2.444
GNO	0.979	0.294	-0.138	0.633	0.979	1.243	1.393	1.523	1.677	1.783	2.020	2.100
PE3	1.000	0.298	0.414	0.398	0.979	1.243	1.393	1.523	1.677	1.783	2.010	2.100

Table 5(d): Regional Quantile Estimates for four regions

	Parameters			Regional quantile estimates with nonexceedance probability F							
Dist	ε	α	k	0.100 1	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500	0.999 1000
GLO	0.930	0.236	-0.173	0.499	0.930	1.300	1.562	1.837	2.242	2.589	3.567	4.078
GEV	0.792	0.356	-0.005	0.496	0.923	1.329	1.599	1.859	2.197	2.452	3.043	3.299
GNO	0.923	0.417	-0.357	0.494	0.923	1.333	1.601	1.857	2.187	2.436	3.020	3.277

Estimation of Homogeneous Region and regional growth curve

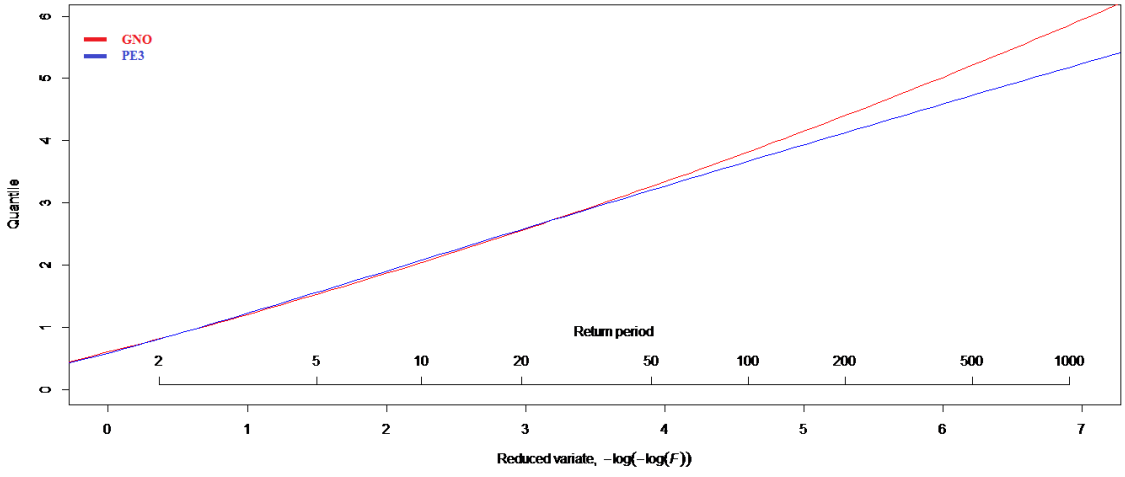
The results obtain from regional frequency analysis for the regional quantile estimates indicate that these estimates are uncertain and unreliable. An assessment analysis based on Monte Carlo Simulation is designed to check the accuracy of quantile estimates using an algorithm provided by [23]. This algorithm states that a region with similar station characteristic as of actual region that is, in terms of the number of stations, record length at each station and regional average L-moment ratios. In the simulation procedure, quantile estimates are estimated for various nonexceedance probabilities. The station i at the m^{th}repetition is $\hat{Q}_i^{[m]}(F)$ and F be the nonexceedance probability. The relative error for such an estimate is $\left\{\hat{Q}_i^{[m]}(F) - Q_i(F)\right\}/Q_i\,(F)$. This quantity is averaged over a total number of repetition [M] to get BIAS and relative RMSE as given by:

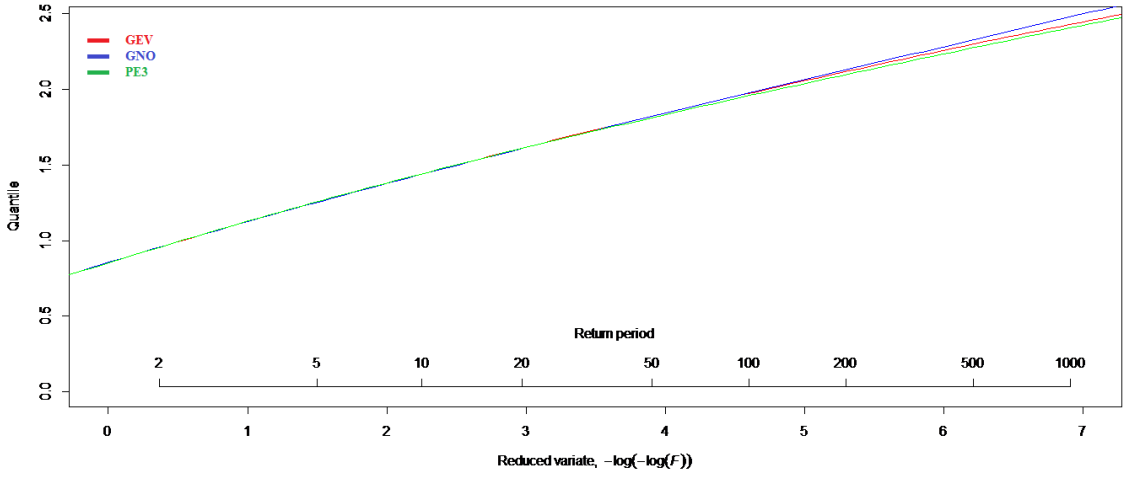
$$B_i(F) = \frac{1}{M}\sum_{m=1}^{M}\frac{\{\hat{Q}_i^{[m]}\,(F) - Q_i(F)\}}{Q_i(F)} \qquad (16)$$

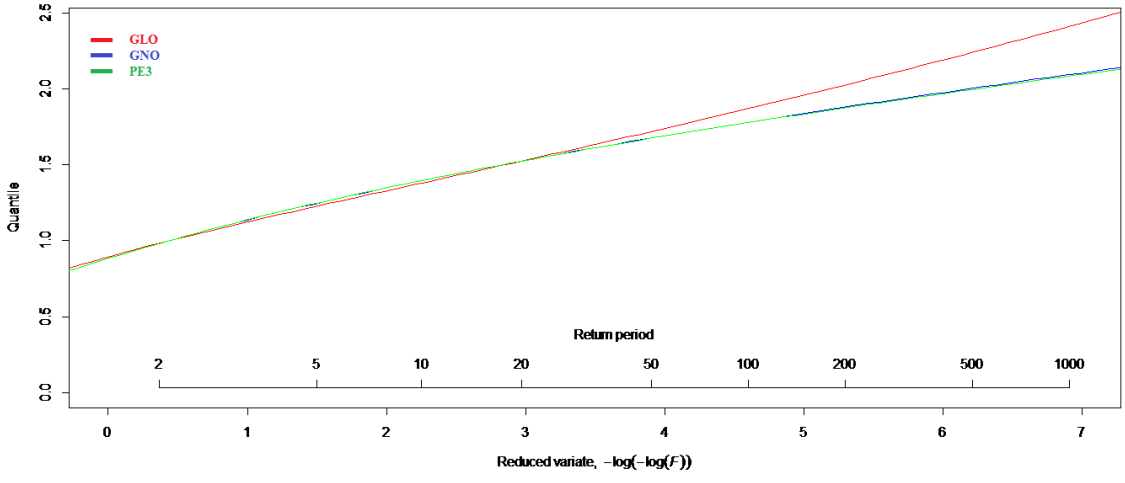
$$R_i(F) = \left[\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\left\{Q_i^{[m]}(F) - Q_i(F)\right\}}{Q_i(F)}\right\}^2\right]^{\frac{1}{2}} \qquad (17)$$

Relative RMSE, relative bias and absolute bias of the estimated quantiles are given below

$$R^R(F) = \frac{1}{N}\sum_{i=1}^{N} R_i\,(F) \qquad (18)$$







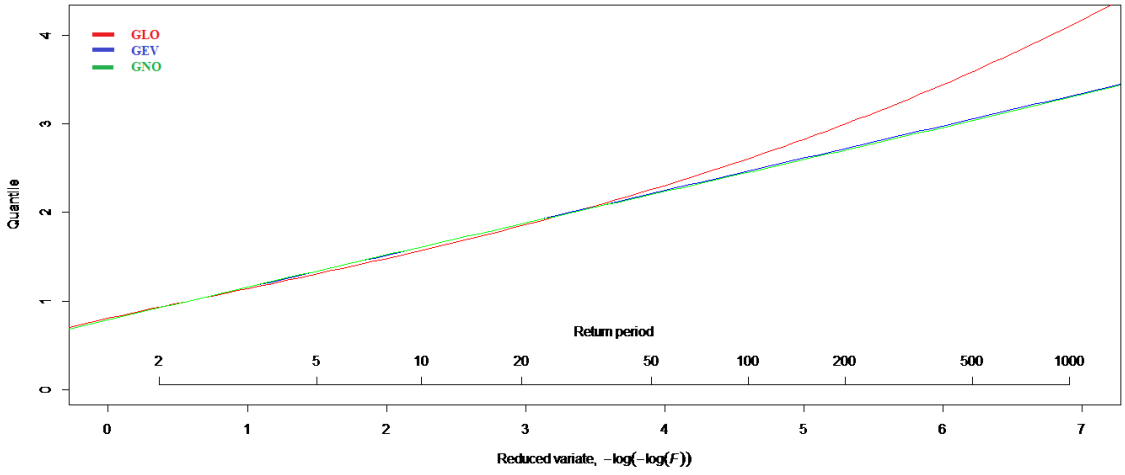


Fig. 4: L-moment ratio diagram for four regions

$$B^R(F) = \frac{1}{N}\sum_{i=1}^{N} B_i(F) \qquad (19)$$

$$A^R(F) = \frac{1}{N}\sum_{i=1}^{N} |B_i(F)| \qquad (20)$$

Empirical quantiles are also useful quantities for assessment analysis and these quantiles are obtained by measuring the ratio of estimated to true value $\hat{Q}_i(F)/Q_i$ for quantiles and $\hat{q}_i(F)/q_i\,(F)$ for the growth curve. The 90% of the regional growth curve lies within the interval if 5% of the simulated values lie below $L_{0.05}(F)$ and 5% lies above $U_{0.05}(F)$ respectively.

$$L_{0.05}(F) \le \frac{\hat{Q}(F)}{Q(F)} \le U_{0.05}(F) \qquad (21)$$

$$\frac{\hat{Q}(F)}{U_{0.05}(F)} \le Q(F) \le \frac{\hat{Q}(F)}{L_{0.05}(F)} \qquad (22)$$

The 90% error bounds give the amount of variation between true and estimated quantities. For the region, I two suitable choices of distributions are GNO and PE3 and on the basis of these two distributions, we perform a simulation study. It has been found that the correlation between stations varies from -0.003 to +0.79 and its average correlation is $\rho = 0.36$. We use an algorithm for simulation defined by Hosking and Wallis [23] in this step. There are 13 stations in the region I for which simulation procedure is applied. All these stations having same record lengths with L-Cv values varying from 0.388 to 0.4648 for GNO distribution. Root mean square, relative bias, absolute bias, lower and upper bounds are calculated for regional growth curve for different nonexceedance probabilities which are given in the table 6a. 10,000 repetitions and 500 simulations are set to perform this algorithm for each candidate distribution. Firstly, this procedure is performed for GNO distribution after then we perform it for PE3 distribution.
istribution.

In table 6a simulation results for the region, I show that the absolute bias for GNO at a return period of 2, 5, 10 and 20 are low as compared to large return period such as for years 50, 100, 500 and 1000. Relative bias for GNO shows that for low return period 1, 2, 5, 10 and 20 years the values are high as compared to large return period. As we know that in practice relative bias and absolute bias are not effective for quantities accuracy. Simulation results for GNO show that at low return period 1, 2, 5, 10 and 20 RMSE are low as compared to large return periods. The error bounds lower error bound and upper error bound (LEB, UEB) for GNO distributions are smaller at low return period whereas its error bounds for large return periods are high. For the region, I at return period 2, 5, 10 and 20 PE3 produce low absolute bias whereas for a large return period of 50, 100, 500 and 1000 its values are high. Relative bias for return period 2, 5, 10 and 20 is high as compared to large return period. Simulation results for PE3 show that at low return period 1, 2, 5, 10 and 20 RMSE are low as compared to large return periods. The error bounds for PE3 distributions are smaller at low return period whereas its error bounds for large return periods are high. At return period of 1, 5 and 20 years PE3 has a relatively low RMSE as compared to GNO at high return period and error bounds of PE3 distribution are narrower than GNO at high return periods. So we can say that PE3 distribution is best for quantile estimation in the region I at large return period.

In table 6b simulation results for region II shows that for GEV at low return period 1, 2, 5, 10 and 20 its RMSE and error bounds are low as compared GNO and PE3. On the other hand, for large return period, RMSE and error bounds of PE3 is narrow as compared to GNO and GEV distributions. So we conclude that for region II PE3 is the most suitable distribution for quantile estimation for large return period and GEV is most suitable choice for low return period.

In table 6c simulation results for region III shows that for GNO at low return period 1, 2, 5, 10 and 20 its RMSE and error bounds are low as compared GLO and PE3. On the other hand, for large return period, RMSE and error bounds of PE3 is narrow as compared to GNO and GLO distributions. So we conclude that for region III PE3 is the most suitable distribution

for quantile estimation for large return period and GNO is most suitable choice for low return period.

In table 6d simulation results for region IV shows that for GEV at low return period 1, 2, 5, 10 and 20 its RMSE and error bounds are small as compared GLO and GNO. On the other hand, for large return period, RMSE and error bounds of GNO is narrow as compared to GEV and GLO distributions. So we conclude that for region IV GEV is the most suitable distribution for quantile estimation for low return period and GNO is most suitable choice for high return period.

Table 6(a): Regional growth curves simulation results for region I

Distribution	F	0.100	0.500	0.800	0.900	0.950	0.980	0.990	0.998	0.999
		1	2	5	10	20	50	100	500	1000
GNO	RR(F)	0.3931	0.0355	0.0252	0.0393	0.0530	0.0696	0.0812	0.1061	0.1162
	BR(F)	0.0497	0.0065	0.0031	0.0004	-0.0020	-0.0049	-0.0069	-0.0108	-0.0122
	AR(F)	0.3041	0.0284	0.0205	0.0320	0.0428	0.0559	0.0651	0.0846	0.0924
	LEB*	0.5460	0.9480	0.9640	0.9390	0.9160	0.8890	0.8710	0.8340	0.8190
	UEB*	1.7960	1.0610	1.0440	1.0660	1.0870	1.1130	1.1320	1.1730	1.1900
PE3	RR(F)	0.3874	0.0366	0.0242	0.0401	0.0523	0.0643	0.0714	0.0838	0.0880
	BR(F)	0.0630	0.0036	0.0017	0.0007	0.0000	-0.0008	-0.0013	-0.0023	-0.0027
	AR(F)	0.2991	0.0292	0.0199	0.0325	0.0420	0.0514	0.0570	0.0668	0.0701
	LEB*	0.5800	0.9410	0.9630	0.9370	0.9180	0.8990	0.8880	0.8690	0.8620
	UEB*	1.8150	1.0610	1.0420	1.0690	1.0890	1.1100	1.1220	1.1430	1.1500

Table 6(b): regional growth curves simulation results for region II

Distribution	F	0.100	0.500	0.800	0.900	0.950	0.980	0.990	0.998	0.999
		1	2	5	10	20	50	100	500	1000
GEV	RR(F)	0.0490	0.0107	0.0162	0.0248	0.0337	0.0462	0.0562	0.0806	0.0915
	BR(F)	0.0007	0.0008	0.0003	-0.0001	-0.0005	-0.0007	-0.0007	0.0000	0.0006
	AR(F)	0.0403	0.0085	0.0133	0.0203	0.0274	0.0372	0.0450	0.0640	0.0725
	LEB*	0.9230	0.9830	0.9740	0.9600	0.9460	0.9260	0.9110	0.8770	0.8630
	UEB*	1.0830	1.0180	1.0270	1.0410	1.0560	1.0780	1.0960	1.1420	1.1630
GNO	RR(F)	0.0492	0.0112	0.0162	0.0254	0.0344	0.0457	0.0539	0.0721	0.0796
	BR(F)	0.0009	0.0008	0.0002	-0.0001	-0.0003	-0.0005	-0.0005	-0.0003	-0.0001
	AR(F)	0.0406	0.0090	0.0132	0.0208	0.0280	0.0370	0.0434	0.0578	0.0637
	LEB*	0.9230	0.9820	0.9750	0.9590	0.9450	0.927	0.9150	0.8890	0.8780
	UEB*	1.0830	1.0190	1.0270	1.0420	1.0570	1.077	1.0920	1.1250	1.1390
PE3	RR(F)	0.0491	0.0112	0.0162	0.0256	0.0341	0.0436	0.0500	0.0625	0.0672
	BR(F)	0.0015	0.0008	0.0001	-0.0002	-0.0005	-0.0009	-0.0011	-0.0016	-0.0018
	AR(F)	0.0405	0.0090	0.0132	0.0210	0.0277	0.0354	0.0404	0.0504	0.0541
	LEB*	0.9240	0.9820	0.9740	0.9590	0.9450	0.9300	0.9200	0.9000	0.8930
	UEB*	1.0830	1.0190	1.0260	1.0420	1.0560	1.0720	1.0830	1.1040	1.1120

Table 6(c): regional growth curves simulation results for region III

Distribution	F	0.100	0.500	0.800	0.900	0.950	0.980	0.990	0.998	0.999
		1	2	5	10	20	50	100	500	1000
GLO	RR(F)	0.0562	0.0115	0.0186	0.0276	0.0369	0.0502	0.0612	0.0899	0.1037
	BR(F)	0.0010	0.0009	0.0003	0.0000	0.0000	0.0000	0.0002	0.0016	0.0026
	AR(F)	0.0460	0.0092	0.0152	0.0226	0.0300	0.0405	0.0492	0.0715	0.0822
	LEB*	0.9120	0.9820	0.9710	0.9560	0.9410	0.9210	0.9050	0.8650	0.8480
	UEB*	1.0950	1.0200	1.0310	1.0460	1.0620	1.0850	1.1050	1.1590	1.1860
GNO	RR(F)	0.0542	0.0105	0.0182	0.0271	0.0353	0.0454	0.0528	0.0689	0.0756
	BR(F)	0.0015	0.0003	0.0001	0.0002	0.0005	0.0010	0.0015	0.0028	0.0034
	AR(F)	0.0445	0.0083	0.0150	0.0222	0.0288	0.0369	0.0426	0.0552	0.0603
	LEB*	0.9160	0.9830	0.9710	0.9570	0.9440	0.9290	0.9180	0.8950	0.8870
	UEB*	1.0930	1.0170	1.0300	1.0450	1.0600	1.0790	1.0900	1.1230	1.1360
PE3	RR(F)	0.0544	0.0105	0.0182	0.0274	0.0355	0.0447	0.0511	0.0642	0.0693
	BR(F)	0.0015	0.0003	0.0003	0.0003	0.0004	0.0005	0.0006	0.0009	0.0009
	AR(F)	0.0449	0.0083	0.0151	0.0225	0.0288	0.0362	0.0412	0.0516	0.0556

LEB*	0.9160	0.9830	0.9710	0.9570	0.9440	0.9300	0.9200	0.9000	0.8920
UEB*	1.0920	1.0170	1.0300	1.0460	1.0590	1.0760	1.0870	1.1110	1.1200

Table 6(c): regional growth curves simulation results for region IV

Distribution	F	0.100	0.500	0.800	0.900	0.950	0.980	0.990	0.998	0.999
		1	2	5	10	20	50	100	500	1000
GLO	RR(F)	0.1103	0.0246	0.0255	0.0405	0.0572	0.0812	0.1008	0.1515	0.1755
	BR(F)	0.0054	0.0045	0.0014	-0.0008	-0.0031	-0.0058	-0.0076	-0.0103	-0.0108
	AR(F)	0.0906	0.0197	0.0210	0.0333	0.0465	0.0655	0.0808	0.1202	0.1385
	LEB*	0.8390	0.9630	0.9610	0.9350	0.9070	0.8700	0.8410	0.7730	0.7440
	UEB*	1.1950	1.0420	1.0430	1.0670	1.0930	1.1340	1.1690	1.2620	1.3070
GEV	RR(F)	0.1068	0.0211	0.0253	0.0399	0.0542	0.0744	0.0907	0.1317	0.1507
	BR(F)	0.0050	0.0024	0.0011	0.0000	-0.0009	-0.0017	-0.0020	-0.0009	0.0003
	AR(F)	0.0882	0.0168	0.0217	0.0330	0.0443	0.0600	0.0727	0.1046	0.1192
	LEB*	0.8450	0.9680	0.9590	0.9370	0.9130	0.8810	0.8570	0.8030	0.7800
	UEB*	1.1880	1.0370	1.0440	1.0660	1.0910	1.1260	1.1560	1.2340	1.2720
GNO	RR(F)	0.1054	0.0221	0.0254	0.0403	0.0546	0.0724	0.0853	0.1136	0.1253
	BR(F)	0.0069	0.0022	0.0004	-0.0003	-0.0008	-0.0012	-0.0012	-0.0008	-0.0004
	AR(F)	0.0873	0.0177	0.0211	0.0333	0.0447	0.0588	0.0689	0.0909	0.0999
	LEB*	0.8490	0.9650	0.9600	0.9350	0.9130	0.8850	0.8660	0.8270	0.8120
	UEB*	1.1880	1.0380	1.0420	1.0670	1.0920	1.1240	1.1490	1.2010	1.2240

Conclusions

Regional frequency analysis has been carried out for 30 meteorological observatories across the Pakistan. After fulfilment of initial assumptions, data found to be suitable for Regional Frequency Analysis. Discordancy measure exhibited that no station is discordant and all can be used for formation of homogenous regions. The region of 30 stations did not satisfy the heterogeneity statistic defined by [23]. On the basis of mean annual rainfall and elevation, four homogeneous regions were formed. On the basis of L-moment ratio diagram and Z-statistic PE3 is the most suitable choice of distribution for first two regions while GLO is for the third and fourth region. On the basis of simulation study using RR, BR and AR, it is found that PE3 the most suitable choice for large return period for first three regions while GNO and GEV for low return period. Similarly, for the fourth region, GEV is declared as best for period lower return periods up to 20 years while GNO for larger return periods. Evaluation of frequency and magnitude of ATR may be of great significance for making policies implication in the form of the measures for disaster prevention and mitigation. Regional quantiles estimates could also be used in economical designing and successful operation of different hydrological structures.

Acknowledgements

Authors are grateful to Pakistan Meteorological center Karachi for providing the required data. The authors are also grateful to Higher Education Commission for financial assistance under project No. 20-3954R&D/HEC/14/305.

REFERENCES

1. E.G. Bilham, “Classification of Heavy Falls of Rain in Short Periods”, British Rainfall, 1935, pp.262-280.
2. J.O. Ayoade, "A Preliminary Study of the magnitude, frequency and distribution of intense rainfall in Nigeria/Etude préliminaire de la grandeur, la fréquence et la distribution des précipitations fortes en Nigérie," Hydrological Sciences Journal, Vol. 21, No. 3, 1976, pp.419-429.
3. K. Adamowski, Y. Alila and P.J. Pilon, "Regional rainfall distribution for Canada," Atmospheric Research Vol. 42, No. 1, 1996, pp.75-88.
4. B.P, Parida,"Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution," International journal of climatology, Vol.19, No. 12, 1999, pp.1389-1398.
5. J.C. Smithers, and R. E. Schulze, "A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments," Journal of Hydrology, Vol.241, No. 1, 2001, pp.42-52.
6. J.S. Park, H.S. Jung, R.S. Kim, and J.H. Oh, "Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution," International Journal of Climatology, Vol.21, No. 11, 2001, pp.1371-1384.
7. S.H. Lee and S.J. Maeng, "Frequency analysis of extreme rainfall using L - moment," Irrigation and Drainage, Vol. 52, No. 3 ,2003, pp. 219-230.
8. K. Yurekli, “Regional analysis of monthly rainfalls over Amasya Province via L-moments method," GOÜ. Ziraat Fakültesi Dergisi, Vol.22, No. 2, 2005, pp.51-56.
9. C.M. Trefry, D. W. Watkins Jr, and D. Johnson, "Regional rainfall frequency analysis for the state of Michigan,"

10. Journal of Hydrologic Engineering, Vol.10, No. 6, 2005, pp. 437-449.
11. D.K. Koh, T.H. Choo, S.J. Maeng and C. Trivedi, "Regional Frequency Analysis for Rainfall using L-Moment," The Journal of the Korea Contents Association, Vol. 8, No. 3, 2008, pp. 252-263.
12. A. Shabri and N. Ariff. "Frequency analysis of maximum daily rainfalls via l-moment approach." Sains Malaysiana, Vol. 38, No. 2, 2009, pp. 149-158.
13. Z. Hussain and G. R. Pasha, "Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-moments," Water resources management, Vol.23, No. 10 ,2009, pp. 1917-1933.
14. C.S. Ngongondo, C.Y. Xu, L.M. Tallaksen, B. Alemaw, and T. Chirwa, "Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches, "Stochastic Environmental Research and Risk Assessment, Vol.25, No. 7, 2011, pp.939-955.
15. B.G. Hassan and F. Ping, "Regional Rainfall Frequency Analysis for the Luanhe Basin–by Using L-moments and Cluster Techniques," APCBEE Procedia, Vol. 1, 2012, pp. 126-135.
16. I. Ahmad, S.F. Shah, I. Mahmood, and Z. Ahmad. "Modeling of monsoon rainfall in Pakistan based on Kappa distribution," Sci. Int.(Lahore) Vol. 25, No. 2, 2013, pp. 333-336.
17. T.A. Devi and P. Choudhury, "Extreme Rainfall Frequency Analysis for Meteorological Sub-Division 4 of India Using L-Moments," World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering Vol. 7, No. 12, pp. 945-950.
18. A. Shahzadi, A.S. Akhter, and B. Saf, "Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments," Pakistan Journal of Statistics and Operation Research, Vol. 9, No. 1, 2013, pp. 111-136.
19. I. Ahmad, M. Fawad, and I. Mahmood, "At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods," Polish Journal of Environmental Studies, Vol.24, No. 06, 2015, 2345-2353.
20. I. Ahmad, A. Abbas, A. Saghir, and M. Fawad, "Finding Probability Distributions for Annual Daily Maximum Rainfall in Pakistan Using Linear Moments and Variants." Polish Journal of Environmental Studies, Vol.25, No.3, 2016, pp. 925-93.
21. I. Ahmad, M. Fawad, A. Abbas, and A. Saghir, "Probability Modeling of low flows at different sites of Indus basin in Pakistan using L-moments and TL-moments," Pakistan Journal of Science, Vol. 68, No. 1, 2016, pp.86-92.
22. J.R.M. Hosking, "L-moments: analysis and estimation of distributions using linear combinations of order statistics," Journal of the Royal Statistical Society. Series B (Methodological), Vol.52, No.1, 1990, pp. 105-124.
23. J.R.M. Hosking and J. R. Wallis, "Some statistics useful in regional frequency analysis," Water resources research Vol. 29, No. 2, 1993, pp. 271-281.
24. J.R.M. Hosking and J. R. Wallis, Regional frequency analysis: an approach based on L-moments. Cambridge University Press, 1997.