Regional Frequency Analysis of Annual Maximum Rainfall in Monsoon Region of Pakistan using *L*-moments

Amina Shahzadi
Department of Statistics
GC University Lahore 54000, Pakistan
aminashahzadi@gmail.com

Ahmad Saeed Akhter
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore 54000, Pakistan
asakhter@yahoo.com

Betul Saf
Pamukkale University
Civil Engineering Department
Hydraulic & Hydrology Division, 20070, Denizli, Turkey
betul.saf@gmail.com

Abstract

Estimation of magnitude and frequency of extreme rainfall has immense importance to make decisions about hydraulic structures like spillways, dikes and dams etc. This research involves the estimation of regional rainfall quantiles of 23 sites using *L*-moment based index flood regional frequency analysis. Initially, different tests are applied to check the assumptions of independence, stationarity and identical distribution. An *L*-moment based discordancy measure is used to detect discordant sites. Since in Pakistan, highly elevated area receive more rainfall. On the basis of this characteristic, the study region is divided into three regions which satisfy the *L*-moment based heterogeneity statistics using Monte Carlo simulations from Kappa distribution. The regional quantile estimates are obtained from GEV, GNO and GLO distributions which are found to be best choices for all three regions based on *L*-moment ratio diagram, Z-Statistics and average weighted difference values. For robust regional estimates, some accuracy measures are calculated using a simulation study of regional *L*-moment algorithm. On the basis of relative bias, relative absolute bias and relative RMSE, GNO is found be best robust for regional quantile estimation at lager return periods of 50, 100, 500 and 1000 and GEV at return periods of 1, 2, 5, 10 and 20 for all three regions.

Keywords: *L*-moments; Discordancy measure; Regionalization; Goodness-of-fit; Relative root mean square error (RMSE).

1. Introduction

Hydrologists are always short of information for making decisions about water resources structures like spillways, dikes, storm surge barriers and dams etc. The physical laws are inadequate to handle the inappropriate short data and significant changes in random process. The hydrometrological variables like extreme rainfall are difficult to describe because of random changes in weather and sampling error generated by limited data as being a small sample of unlimited population. In this application the recorded data at different sites of a well defined homogeneous region is used to estimate the extreme events expected to occur in 100 or 1000 years to reduce the uncertainties of rare events.

Regional rainfall frequency analysis is of paramount importance in civil structure designs as well as plays an important role in a diverse range of nonstructural problems involving natural hazards associated with extreme rainfall events, because the analyses provide the information about occurrence of rainfall amounts within a specified recurrence interval.

A lot of literature is available on the regional frequency analysis of extreme rainfall and precipitation. Early examples are Bilham (1936) for regional analysis of extreme rainfall in England and Wales, Ayoade (1976) for regional analysis of daily precipitation in Nigeria. Schaefer (1990), following the index flood methodology of Dalrymple (1960), found the extreme value II distribution to be the best distribution for annual precipitation data in Washington State. Cannarozzo et al. (1995) made a regional frequency analysis for Sicilian region using two-component extreme value distribution for annual maximum rainfall of various durations. The regional analysis of rainfall of Canada has been performed by Adamowski et al. (1996) with a conclusion of generalized extreme value (GEV) distribution giving the reliable quantile estimates of rainfall using *L*-moments.

Parida (1999) obtained the reliable quantile estimates for Indian summer monsoon rainfall at high and low return periods using four parameter kappa distribution. Koutsoyiannis and Baloutsos (2000) found the GEV distribution best for annual maximum rainfall in Greece for the prediction at large return periods. To obtain the estimates for various return periods in South Korea, Park et al. (2001) used Wakeby distribution with the method of *L*-moment estimates on the summer extreme rainfall at 61 stations. Smithers and Schulze (2001) made regional frequency analysis based on *L*-moments to estimate the short duration design rainfall in South Africa. Sveinsson et al. (2002) utilized regional frequency analysis of annual maximum precipitation for Northeastern Colorado. In Peninsular Malaysia, Zalina et al. (2002) found the GEV distribution to be best to obtain reliable and accurate maximum rainfall estimates. For rainfall in Korea, Lee and Maeng (2003) found the GEV and GLO distributions to be best for extreme rainfall using *L*-moment ratio diagram and Kolmogrov Smirnov test. Koutsoyiannis (2004 (a,b), 2007) concluded the EV2 distribution to be best for the behaviour of extreme rainfall. Other examples of regional frequency analysis of rainfall have been found in the works of Trefry et al. (2005), Yurekli (2005) for rainfall over Amasya province, Cavigli et al. (2006) and Tartaglia et al. (2006) for extreme rainfall in Tuscany, Lin et al. (2006) for rainfall data in Taiwan, Weaver (2006) for rainfall in the city of Charlotte and Mecklenburg, North Carolina, Kysely and Picek (2007) and Kysely et al. (2007) for extreme precipitation in Czech Republic and Norbiato et al. (2007) for extreme rainfall in eastern Italian Alps. Many more examples like Sankarasubramanain and Srinivas (2008), Yang et al. (2010), Anli et al. (2011), Ngongondo et al. (2011) can be found in literature.

2. *L*-moments

Historically the *L*-moments are defined as the linear function of the probability weighted moments (PWMs). PWMs are defined as

$$\beta_r = E\left[x\{F(x)\}^r \right]$$

The *r*th *L*-moment λ_r is related to the *r*th PWM (Hosking, 1990) through:

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

Therefore, the first four *L*-moments are:

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

The *L*-moments that are independent of units of measurement, called *L*-moment ratios (Hosking, 1990), are defined to the quantities

$$\tau = \lambda_2 / \lambda_1$$
$$\tau_3 = \lambda_3 / \lambda_2$$
$$\tau_4 = \lambda_4 / \lambda_2$$

where τ is *L*-coefficient of variation ($L - C_v$), τ_3 is *L*-coefficient of skewness ($L - C_s$) and τ_4 is *L*-coefficient of kurtosis ($L - C_k$). If the mean of a distribution exists, then all of the *L*-moments exists uniquely defining the distributions i.e., no two distributions have the same *L*-moments (Hosking and Wallis, 1997).

L-moments have superior abilities to conventional moments in discriminating between different distributions, because the *L*-moment ratio estimators of location, scale, and shape are nearly unbiased, regardless of the probability distribution from which the observations arise and efficient estimators of the characteristic of hydrologic data and of the parameters of the distribution. (Hosking, 1990, Stedinger et al., 1993; Hosking and Wallis, 1997, Singh, 1998; Zafirakou-Koulouris et al., 1998). These estimators are linear combinations of the ranked observations and thus are less sensitive to the largest observations in a sample than ordinary product moment estimators. They are particularly good at identifying the distributional properties of highly skewed data, whereas ordinary product moment diagrams are almost useless for this task (Vogel and Fennessey, 1993; Hosking and Wallis, 1997; Sankarasubramanian and Srinivasa, 1999; Ulrych et al., 2000).

In a developing country like Pakistan, because of the problem of sufficient information of recorded events of extreme rainfall, extrapolation is required. To overcome the lack of data and to model uncertainty, a regional frequency analysis has been performed for annual maximum rainfall (AMR) of monsoon region of Pakistan using the index flood method of Dalrymple (1960) using *L*-moments developed by Hosking (1990). The primary objective of the study is to develop a regional distribution for estimating extreme rainfall quantiles in the region of Pakistan more or less affected by monsoon. This approach will be new for providing the reliable and consistent design rainfall estimates to help the engineers in decision making concerning the design construction of hydraulic structures like barrages, reservoirs and dams etc. in the region.

3. Study Area and Data Description

Pakistan is basically a rainfall deficient country, except for the area above 32^0 latitude.

The study area lies within the active monsoon belt; monsoon belt is situated over the eastern and northeastern part of Pakistan including Kashmir. The 23 sites of study area are shown in Figure 1. The climatic conditions of the area are tropical in summer and temperate in winter. During spring and autumn, it is under transitional climatic zone. Spring season is pre-monsoon season during which the zone comes under unstable atmospheric conditions. On account of effect, the temperature of ground starts to increase while the temperature within low tropo spheric region is still under the impact of winter season. This situation is potentially unstable and causes large square dust and thunderstorm which occasionally showers. In contrast, the post monsoon season is dry since the monsoon has ended while the westerly waves have not yet started to affect the region.

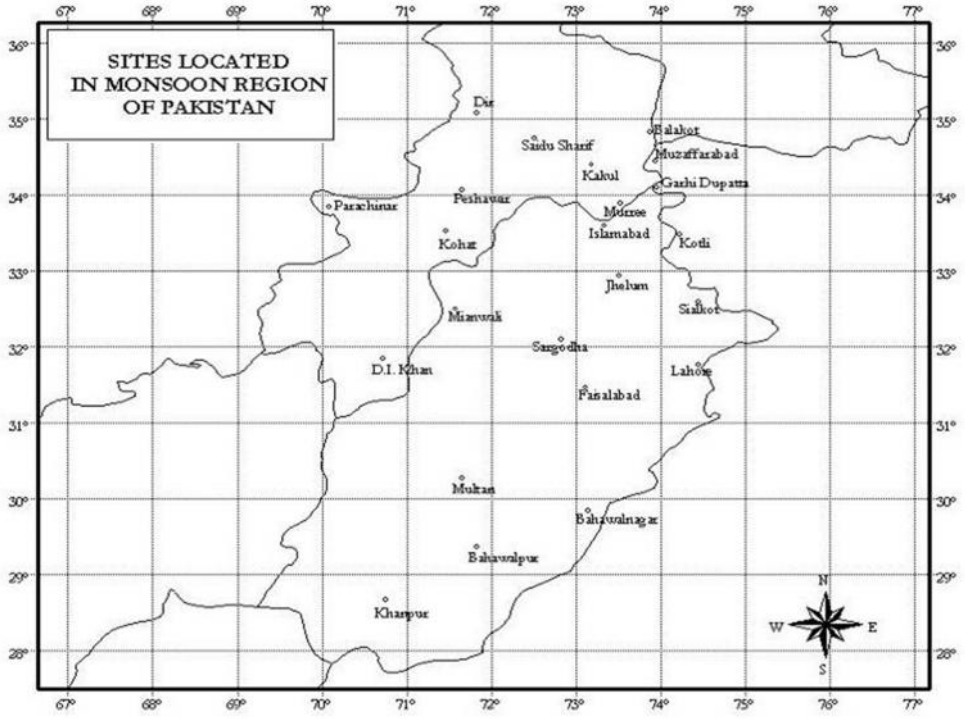

Figure 1: Sites of rainfall in Monsoon region

The rainfall is concentrated during three months of summer (July, August, and September) which is due to monsoon winds. Winter rainfall is due to western depression and is much smaller in amount. The northern mountainous region receives more rainfall. It is 1000mm or more yearly. The rainfall decreases sharply toward southern part of upper Indus plain. It is less than 100mm in south western areas. Annual average rainfall of the country is around 300mm out of which about 140mm rainfall occurs during the three monsoon months.

In this study, regional frequency analysis has been conducted on annual maximum rainfall measured at selected 23 gauged stations located in a region more or less affected by Monsoon (Figure 1). The daily rainfall amounts measured by rain gauges were provided by the Regional Meteorological Center (RMC) Lahore, which are measured in millimeters, from which annual maximum rainfall (AMR) series have been constructed for the proposed study. The record length of AMR series vary from 24 to 60 years.

4. Assumptions of regional frequency analysis

It is a usual practice to test the assumptions that the observations at various sites are stationary, independent and identically distributed. The different plots and tests have been applied on AMR series discussed below.

Time series plots

The time series plots of AMR depths measured in mm have been made out at the first stage of the study. Some sites like Balakot, Sialkot, Islamabad AP, Lahore and Kotli have greater variation in AMR events because of abrupt changes in climatic conditions. Islamabad SRRC has large variation because of extreme precipitation event of 591.90 mm in 10 hours which was the result of interaction of monsoon and western disturbance (Rasul et al. 2004). Overall some sites showed slight upward and some showed downward trends which were ignored because some slight trends exist for abrupt changes in climate.

Mann-Whitney test

One of the assumptions of regional frequency analysis is that the observations of series should follow the same distribution. For testing the hypothesis of same distribution, Mann-Whitney Rank sum test (1947) has been performed. Each AMR series has been divided into two parts to test the shift in means of two subgroups, i.e. whether the two parts are from the same distribution. All series showed insignificance. So it was concluded that the AMR series for all sites are identically distributed.

Kendall's tau test

A rank based Kendall's tau correlation method (Hirsch et al. 1993) has been applied to check the assumption of correlation of AMR series with time, which is also a trend test. AMR series for all sites did not reject the hypothesis that AMR do not change as a function of time. Being the trend test, there are insignificant trends in AMR series for all sites.

Ljung-Box Q-Statistics

To check the serial correlation of AMR series for all sites, an autocorrelation based Ljung-Box Q*-Statistics* developed by Ljung-Box (1978) has been applied. The hypothesis to be tested is that all autocorrelations, $\hat{\rho}_k$, up to certain lag are all equal to zero, with lag not be more than $n_i / 4$ (Box et. al. 1994) where n_i is the record length of ith site. All sites showed insignificant autocorrelations except Multan whose results are shown in Table 1.

At Multan, site 17, the correlation at lag 2 is highly significant. There is a tendency in the data for large and small values to alternate from 1974 to 1980. Multan is a rainfall deficient site where the monsoon has little effect. There were drought conditions in Multan for years 1975, 1977 and 1979. These abrupt changes in AMR produced significant autocorrelations at all lags but this site satisfies Mann-Whitney test and Kendall's τ test. And nonuniformity of the climatic conditions of the site suggests this site to be reasonable to work with.

From initial analysis, AMR series seemed to be appropriate for regional frequency analysis because all sites satisfied Mann-Whitney test, Kendall's τ test and $Q_{LB}-$ statistic. The patterns of AMR events are not uniform because of sudden changes in climatic conditions. It is reasonable to treat these patterns occurred by chance. According to Chaudhary and Rasul (2004), we could not find any significant correlations for internal annual-daily extreme events if correlations exist then it is just because of nonuniform climatic conditions.

Table 1: Correlogram, Autocorrelations and $Q_{LB}-$Statistic for 24h AMR at Multan (Site 18)

Correlogram	Lag	AC	Q-Stat	Prob
	1	-0.264	4.1720	0.041
	2	0.492	18.998	0.000
	3	-0.112	19.781	0.000
	4	0.144	21.099	0.000
	5	0.116	21.971	0.001
	6	-0.204	24.708	0.000
	7	0.315	31.400	0.000
	8	-0.308	37.902	0.000
	9	0.304	44.395	0.000
	10	-0.142	45.838	0.000
	11	0.102	46.599	0.000
	12	0.070	46.968	0.000
	13	-0.185	49.588	0.000
	14	0.218	53.295	0.000

Moreover, Potter (1979) described that the annual or monthly series are usually stationary, although when the series are not homogeneous this might involve many temporal effects.

5. Regional frequency analysis based on *L*-moments

The regional analysis methodology used for AMR is an index flood regional frequency analysis of Darlymple (1960) based on *L*-moments outlined by Hosking and Wallis (1997). The underlying concept supposes that data are available at N sites, with site i

having sample size n_i and observed data Q_{ij}, $j = 1, 2, \ldots, n_i$. Let $Q_i(F)$, $0 < F < 1$, be the quantile function of the frequency distribution at site i. The key assumption of the index flood procedure is that the sites are from a homogeneous region, i.e., the frequency distributions of the N sites are identical apart from a site-specific scaling factor called flood index. So it can be written $Q_i(F) = \mu_i q(F), \quad i = 1, 2, \ldots, N$.

Here μ_i is the index flood, site specific scale factor. Regional frequency analysis has the following steps:

- screening of data
- formation of homogeneous regions
- selection of regional distribution
- estimation of selected distribution
- decision about best robust distributions using assessment analysis

These steps are performed below.

5.1 Screening of data using discordancy measure

Hosking and Wallis (1993)'s discordancy measure for site *i* is defined as follows:

$$D_i = \frac{1}{3}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u})$$

where $u_i = \left[t^{(i)}, t_3^{(i)}, t_4^{(i)} \right]'$, $\bar{u} = N^{-1}\sum_{i=1}^{N} u_i$ and $S = \frac{1}{N-1}\sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T$

are vector $L-C_v$, $L-C_s$ and $L-C_k$, mean of vector u_i and covariance matrix of u_i. The measure D_i indicates how far u_i is from the centre of the region relative to the size of the region. Hosking and Wallis (1997) suggested that a site be regarded as discordant if its D_i value exceeds the critical value given in Table 3 at page 47 of Hosking and Wallis (1997).

Treating the entire set of 23 sites as a single region, to screen the discordant site, the discordancy measure, D_i has been calculated for each site in the Table 2. For 23 sites in the region, the critical value of D_i for all sites is 3. The site 20, Khanpur, has D_i=3.1460 greater than critical value of 3. The site has moderate *L*-skewness and *L*-kurtosis but very high *L*-CV that may cause the high D_i value.

Table 2: *L*-moment Ratios and Discordancy Measures

Site No.	Site Name	Index	n_i	t	t_3	t_4	D_i	D_i
1	Dir	41508	40	0.1684	0.2337	0.1497	0.9160	1.0582
2	Saidu Sharif	41523	55	0.3346	0.2999	0.2161	0.6295	1.1233
3	Kakul	41535	54	0.2189	0.3698	0.2843	0.7370	0.7604
4	Balakot	41536	47	0.2239	0.2989	0.1481	0.9781	0.9224
5	Parachinar	41560	40	0.2283	0.2957	0.1947	0.2324	0.2111
6	Kohat	41564	53	0.2825	0.3136	0.1975	0.2504	0.3476
7	Islamabad AP	41571	48	0.2162	0.1742	0.0433	1.6263	1.5467
8	Murree	41573	47	0.1721	0.2878	0.2180	0.7387	0.9582
9	Islamabad SRRC	41577	24	0.3087	0.4865	0.3810	2.5160	2.4650
10	Mian Wali	41592	48	0.2381	0.1806	0.2411	1.8120	1.7187
11	Sargodha	41594	49	0.2073	0.1497	0.2032	1.6413	1.5886
12	Jhelum	41598	57	0.2090	0.2249	0.1877	0.2572	0.2621
13	Sialkot	41600	55	0.2891	0.2874	0.1428	0.6185	0.8041
14	DI Khan	41624	57	0.2586	0.1725	0.0990	0.5099	0.5844
15	Faisalabad	41630	56	0.2693	0.3883	0.3516	1.3836	1.3101
16	Lahore	41640	60	0.2792	0.2668	0.1966	0.0615	0.1607
17	Multan	41675	57	0.2714	0.2217	0.1843	0.2121	0.2847
18	Bahawalnagar	41678	35	0.3374	0.2855	0.1632	0.7512	1.3824
19	Bahwalpur	41700	47	0.2964	0.0944	0.0735	1.7075	2.0818
20	Khanpur	41718	55	0.4288	0.2517	0.1295	3.0203*	--------
21	Muzaffarabad	43532	52	0.1936	0.2640	0.1515	0.7651	0.7855
22	Garhi Dupatta	43533	52	0.1978	0.1835	0.2068	1.0041	1.0193
23	Kotli	43563	54	0.2301	0.3530	0.2969	0.6312	0.6246

The scatter plots for $L-C_s, L-C_v$ and $L-C_s, L-C_k$ have been drawn in Figure 2. In Figure 2, suggests the site 20 to be discordant with the pattern of *L*-moment ratios ($L-C_s$, $L-C_v$) of other sites and their weighted average but site 20 is emerged with the pattern of *L*-moment ratios ($L-C_s, L-C_k$) of other sites and not seems to be discordant. $L-C_v$. As discordancy measure is based on three coefficients $L-C_v, L-C_s$ and $L-C_k$, there are many sites with low value of $L-C_s$ and $L-C_k$ but site 20 has large value of $L-C_v$ which makes it discordant.

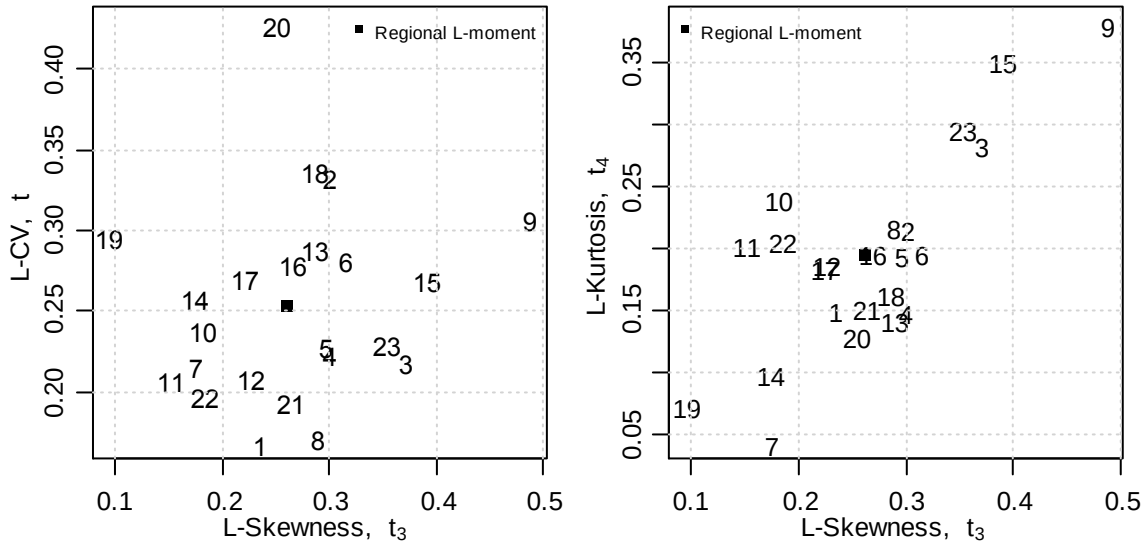

Figure 2: *L*-moment Ratios for 23 Sites

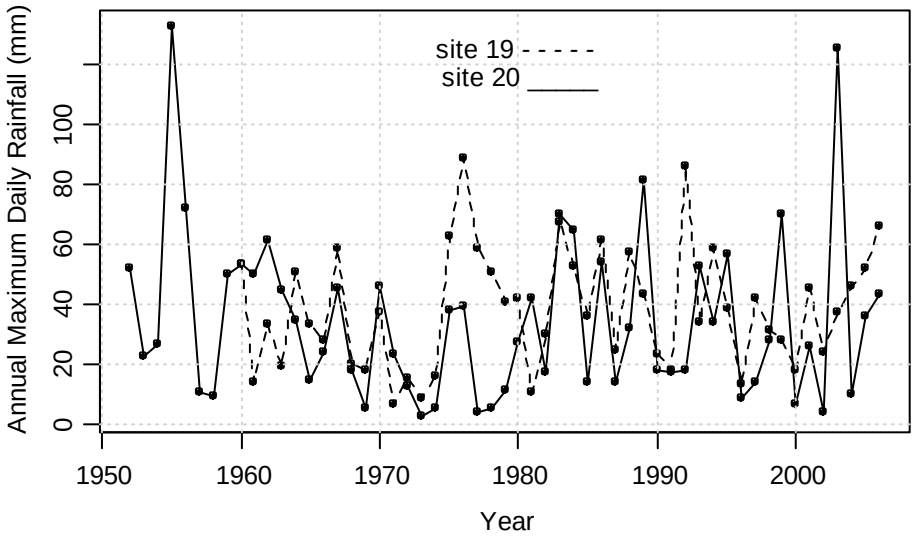

Figure 3: AMR Series for Bahawalpur (Site 19) and Khanpur (Site 20)

Site 20, Khanpur, is compared with the site nearest to it, site 19, Bahawalpur, in Figure 3. The pattern of changing the values for both sites seems to be approximately same, but extremes in 1955 and 2001 for site 20, Khanpur, are producing large value of $L-C_v$. Because of no clear reason to discard site 20, Khanpur, it has been decided to retain to see its effect on homogeneity of study area.

5.2 Regional heterogeneity and formation of homogeneous regions

The next step of identification of homogeneous regions is usually the most difficult and requires the greatest amount of subjective judgment. The homogeneity condition means that the sites have same frequency distributions.

Hosking and Wallis (1993) presented a heterogeneity measure to estimate the degree of heterogeneity in a group of sites and to assess whether they might reasonably be treated as a homogeneous region. The heterogeneity measure compares the observed and simulated dispersion of *L*-moments for N sites under consideration. For this purpose Monte Carlo simulation is made using four parameter Kappa distribution defined by

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

ε is a location parameter, α is a scale parameter, and k and h are shape parameters. The range of x is:

$$\begin{aligned}
\varepsilon+\alpha\left(1-h^{-k}\right)\Big/k \le x \le \varepsilon+\alpha/k \qquad & if\ h>0, k>0\\
\varepsilon+\alpha\log h \le x < +\infty \qquad & if\ h>0, k=0\\
\varepsilon+\alpha\left(1-h^{-k}\right)\Big/k \le x < +\infty \qquad & if\ h>0, k<0\\
-\infty < x \le \varepsilon+\alpha/k \qquad & if\ h\le 0, k>0\\
-\infty < x < +\infty \qquad & if\ h\le 0, k=0\\
\varepsilon+\alpha/k \le x < +\infty \qquad & if\ h\le 0, k>0
\end{aligned}$$

The reason to use Kappa distribution is that it is a generalized distribution which produces many distributions as special cases based on parameter values. For example, $h=1$ yields a generalized Pareto distribution, $h=0$ a GEV distribution, and $h=-1$ a generalized logistic distribution. An exponential distribution arises when $h=1$ and $k=0$, a Gumble distribution when $h=0$ and $k=0$, a logistic distribution when $h=-1$ and $k=0$, and a uniform distribution when $h=1$ and $k=1$. When $h=0$ and $k=1$, the four-parameter is reverse exponential distribution. *L*-moments of four parameter distribution covers a large area of $(\tau_3,\ \tau_4)$ plane.

The four parameter Kappa distribution is fitted to regional average *L*-moments ratios 1, t^R, t_3^R and t_4^R to simulate a large number N_{sim} of realizations of a region with N sites. The heterogeneity measure, H_j ($j=1, 2, 3$), is defined as

$$H_j=\frac{V_j-\mu_{V_j}}{\sigma_{V_j}}$$

where H_1 is heterogeneity measure based on observed weighted standard deviation of t values, V_1, H_2 is heterogeneity measure based on observed average of (t/t_3) distance, V_2 and H_3 is heterogeneity measure based on observed average of (t_3/t_4) distance, V_3. μ_{v_j} and σ_{v_j} are mean and standard deviation of N_{sim} simulated values of V_j.

H_j can be approximated by a normal distribution with mean zero and unit variance. Following Hosking and Wallis (1997), the region under analysis can therefore be regarded as "acceptably homogeneous" if $H_j<1$, "possibly heterogeneous" if $1\le H_j<2$, and "definitely heterogeneous" if $H_j\ge 2$.

For the present study, treating the entire set of 23 sites as a single region, Hosking and Wallis's heterogeneity statistics have been calculated including and excluding discordant site in Table 3.

Table 3: Heterogeneity Statistics

No. of Sites	H_1	H_2	H_3
23 sites with Discordant site	6.74*	1.86*	-0.06
22 sites without Discordant site	3.51*	1.23*	-0.02

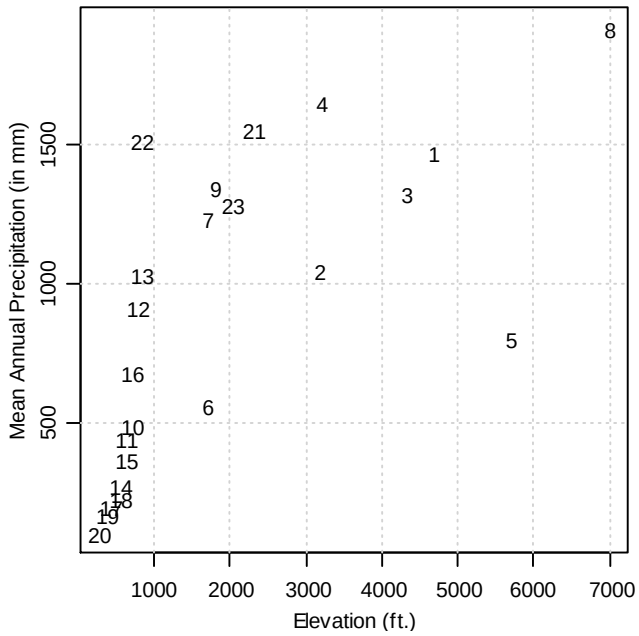

Figure 4: Scatter Plot of Elevation and Mean Annual Precipitation

According to the critical values of H_1, H_2, and H_3, the region of 23 sites appeared to be heterogeneous with and without discordant site 'Khanpur' but ignoring discordant site has reduced the values of heterogeneity statistics H_1 and H_2. According to Hosking and Wallis (1997), H_1 has better discriminatory power to detect the homogeneity of region than H_2 and H_3.

Formation of homogeneous regions

The study area has the characteristic that as highly elevated sites have high mean annual precipitation and as we move from north to south with low elevated sites (Indus Plain sites), the mean annual precipitation decreases. The scatter plot of elevation (ft) and mean annual precipitation is shown in Figure 4, numbers showing points (elevation, mean annual precipitation) of corresponding sites. It is clear from the Figure 4 that mean annual precipitation increases as elevation increases. The proposed region would be subdivided into homogeneous subregions using elevation and mean annual precipitation characteristics.

Table 4: Heterogeneity Statistics for Homogeneous Regions

	Site	Site name	D_i	Heterogeneity Statistics	Regional *L*-moments
Region 1	6	Kohat	0.50*	$H_1 =$ 0.95	t^R =0.2651
	10	Mian Wali	1.17	$H_2 =$ 0.55	t_3^R =0.2368
	11	Sargodha	1.08	$H_3 =$ 0.00	t_4^R =0.1870
	12	Jhelum	1.39		
	13	Sialkot	0.90		
	14	DI Khan	0.72		
	15	Faisalabad	2.05		
	16	Lahore	0.05		
	17	Multan	0.07		
	18	Bahawalnagar	1.27		
	19	Bahwalpur	1.80		
Region 2	7	Islamabad AP	1.21	$H_1 =$ 0.92	t^R =0.2199
	9	Islamabad SRRC	1.20	$H_2 =$ 0.98	t_3^R =0.2712
	21	Muzaffarabad	0.95	$H_3 =$ 1.47	t_4^R =0.1995
	22	Garhi Dupatta	1.22		
	23	Kotli	0.41		
Region 3	1	Dir	0.79	$H_1 =$ 0.34	t^R =0.1977
	3	Kakul	1.05	$H_2 =$ -0.59	t_3^R =0.3026
	4	Balakot	1.26	$H_3 =$ -0.57	t_4^R =0.2050
	5	Parachinar	0.58		
	8	Murree	1.32		

*Critical values for region 1 based on eleven sites is 2.63 and for region 2 and 3 each having five sites is 1.33

There are many sites in Figure 4 close to each other. The heterogeneity measures were calculated for various subgroups observed from Figure 4. In Table 4, the three regions, region 1 having eleven sites, region 2 and 3 each having five sites, have satisfied the heterogeneity measures. Site 2, Saidu Sharif, and site 20, Khanpur (discordant site) did not produce any group of sites to be homogeneous and have been removed for further analysis which might be studied with other sites not considered in this study.

Moreover, the values of discordancy measure for each site in three regions do not exceed the critical values given in Table 3.1, Hosking and Wallis (1997). The three homogeneous regions with their regional weighted *L*-moments have been shown in Table 4.

5.3 Selection of the best-fit regional distribution

Many distributions can be used for quantile estimation for regional data. The three–parameter distributions, generalized logistic (GLO), generalized extreme value (GEV), generalized normal (GNO), generalized Pareto (GPA), and Pearson type III (PE3) have been considered in this regional analysis. According to Hosking and Wallis (1997), two-parameter distributions might cause bias in tail quantile estimates if the shape of the tail of the true frequency distribution is not well approximated by the fitted distribution. The best fit distribution is one that gives robust estimates for regional growth curve as well as for at-site quantiles.

L-moment ratio diagram and $Z-statistic$ have been used as the best fit criteria to identify the underlying regional distribution for homogeneous regions. *L*-moments ratio diagram are constructed using the unbiased estimators of *L*-moments introduced by Hosking (1990) and recommended by Stendiger et al. (1993), Vogel and Fennessey (1993) and Hosking and Wallis (1995). The curves show the theoretical relationships between *L*-skewness and *L*-kurtosis of various candidate distributions. *L*-moment ratio diagrams have been suggested as a useful tool for discriminating between candidate distributions to describe regional data (Hosking 1990; Stedinger *et al*. 1993; Hosking & Wallis 1997). Numerous authors (for example Schaefer 1990; Pearson 1993; Vogel *et al*. 1993(a, b); Chow & Watt 1994; Onoz & Bayazit 1995; Vogel & Wilson 1996, Peel et al., 2001) have used *L*-moment ratio diagrams as part of their distribution selection process for regional data. According to Vogel et al. (1993a) and Hosking and Wallis (1995), the proximity of the sample average or the record length weighted average to a particular candidate distribution theoretical curve or point in $(L-C_s, L-C_k)$ space has been interpreted as an indication of the appropriateness of that distribution to describe the regional data.

Another criterion to select best fit distribution is $Z^{DIST}-statistic$ defined by Hosking and Wallis (1993) whose main aim to compare simulated $L-C_s$ and $L-C_k$ of a fitted distribution with the regional average $L-C_s$ and $L-C_k$ values obtained from observed data.

The goodness of fit measure for a distribution $Z^{DIST}-statistic$ is defined by

$$Z^{DIST}=\left(\tau_4^{DIST}-t_4^R+B_4\right)\Big/\sigma_4$$

where τ_4^{DIST} is the *L*-kurtosis of fitted distribution, B_4 and σ_4 are simulated regional bias and simulated regional standard deviation of t_4^R where simulations are made from fitted kappa distribution to regional *L*-moments. The fit is good if $\left|Z^{DIST}\right|\le 1.64$ which corresponds to acceptance of the hypothesized distribution at a confidence level of 90%. A number of distributions may qualify this criterion; the most adequate is one that has $\left|Z^{DIST}\right|$ value close to zero.

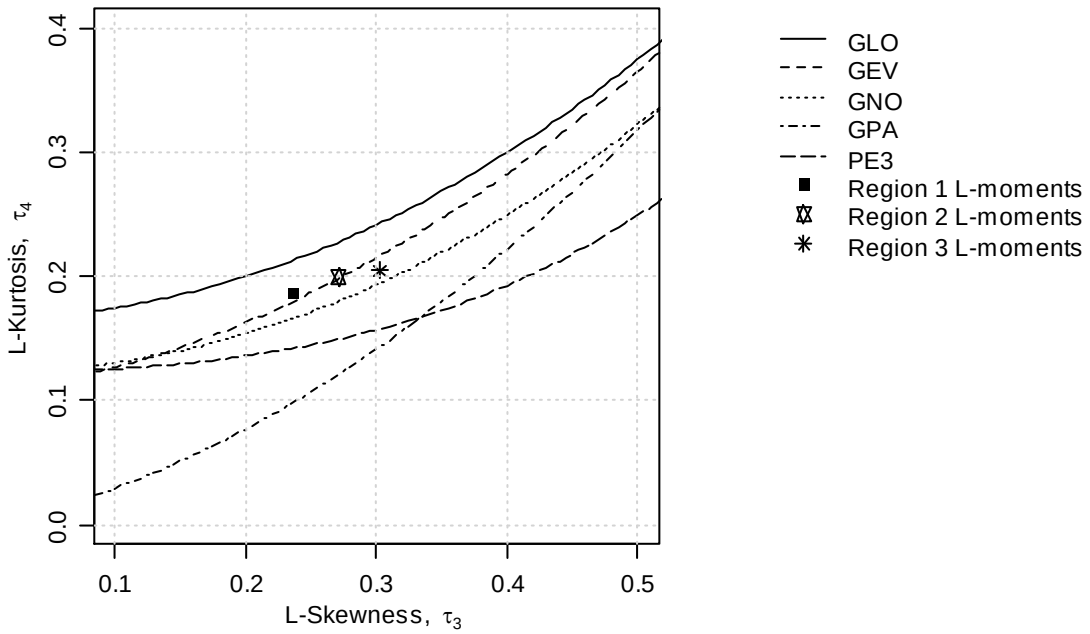

Figure 5: *L*-moment ratio diagram with regional *L*-moment ratio

The regional average *L*-moment ratios (t_3^R, t_4^R) are plotted on the theoretical *L*-moment ratios curves of candidate distribution in Figure 5 for all three regions. The point (t_3^R, t_4^R) for region 1 and 2 falls on GEV distribution curve, for region 3 it is somewhere between GEV and GNO distribution curves.

$Z^{DIST} - statistic$ has been calculated for proposed distributions for three homogeneous regions in Table 5, which shows that GPA and PE3 can be eliminated as unsuitable choices of distributions for three regions. Arranging the value of $\left|Z^{DIST}\right|$ for selected distribution in ascending order, GEV gives better fit for Region 1, 2 and 3 and GLO and GNO are the alternatives choices for all three regions.

Table 5: Z^{DIST} -Statistic for Various Distributions

Region	Z^{GLO}	Z^{GEV}	Z^{GNO}	Z^{GPA}	Z^{PE3}
1	1.16*	-0.49*	-1.14*	-4.53	-2.31
2	0.61*	-0.23*	-0.74*	-2.44	-1.65
3	0.92*	0.18*	-0.42*	-1.87	-1.66

*shows acceptance at 10% level of significance

5.4 Identification of results with AWD values

According to Kroll and Vogel (2002), a visual interpretation of *L*-moment ratio diagram is somewhat subjective. They defined a measure based on difference between sample and theoretical *L*-moment ratios average weighted distance (AWD) for three parameter distributions given by

$$\text{AWD} = \frac{\sum_{i=1}^{N} n_i d_i}{\sum_{i=1}^{N} n_i}$$

Where N is number of sites in region, n_i is the record length at site i and $d_i = \left|\tau_4\left[\tau_3^o(i)\right] - \tau_4^o(i)\right|$ with $\tau_4^o(i)$ to be observed or sample $L-C_k$ and $\tau_4\left[\tau_3^o(i)\right]$ to be the theoretical $L-C_k$ calculated from the distribution corresponding to a given sample $L-C_s$. The smaller difference between sample *L*-moment ratios and probability *L*-moment ratios produce the smaller value of AWD indicating the best choice for the describing the regional data. The AWD values have been calculated for all considered distributions in Table 6. GEV distribution has lowest AWD value among the other chosen distributions, GNO can be ranked as second alternative chosen distribution and GLO as third alternative for all three regions. GPA and PE3 have large AWD values and according to Z^{DIST} -Statistics, GPA and PE3 are not best choices for all three regions.

Table 6: AWD Values to rank the candidate distributions

Region	AWD Values				
	GLO	GEV	GNO	GPA	PE3
1	0.054	0.046	0.047	0.083	0.058
2	0.058	0.058	0.057	0.078	0.080
3	0.043	0.030	0.031	0.059	0.048

5.5 Estimation of regional growth curves

Regional *L*-moments algorithm outlined by Hosking and Wallis (1997) has been applied to estimate the regional frequency distributions. This procedure is to fit the distribution by equating its *L*-moment ratios λ_1, τ, τ_3 and τ_4 to regional average *L*-moments ratios l_1^R, t^R, t_3^R, t_4^R, where the averages are weighted proportionally to length of record of sites (n_i). The quantile function of the fitted regional frequency distribution is usually denoted by $\hat{q}(.)$. The quantile estimates at site i are obtained by combining the estimates of μ_i and $q(.)$. The estimate of the quantile with the nonexceedance probability F is

$$\hat{Q}(F) = l_1^{(i)} \hat{q}(F)$$

The regional quantiles estimates $\hat{q}(F)$, for various nonexceedance probabilities, have been shown in Table 8. Regional growth curves for three regions have been plotted in Figure 6. These can be interpreted as, for example, for region 1, $\hat{q}_{GEV}(0.99)$=2.7852 is the amount of rainfall which will occur once in 100 years and is 2.785 times larger than the average for all sites in region 1. The regional growth curves for all three regions are approximately same upto return period of 100. As return period exceeds 100, the GEV curve is between GNO and GLO curves. To select the robust regional quantiles estimates, there is always a need of some assessment analysis.

Table 8: Fitted Distributions and Regional Quantile Estimates for three Regions

Distribution		Parameters			Regional quantile estimates with nonexceedance probability F								
		$\hat{\xi}$	$\hat{\alpha}$	$\hat{k}$	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
Region 1	GEV	0.763	0.345	-0.101	0.500	0.888	1.309	1.621	1.949	2.417	2.805	3.836	4.345
	GNO	0.889	0.425	-0.491	0.500	0.885	1.321	1.637	1.958	2.399	2.750	3.633	4.045
	GLO	0.900	0.241	-0.264	0.500	0.896	1.285	1.583	1.917	2.440	2.917	4.392	5.233
Region 2	GEV	0.797	0.270	-0.151	0.546	0.887	1.274	1.572	1.894	2.369	2.774	3.903	4.484
	GNO	0.896	0.341	-0.565	0.544	0.883	1.287	1.591	1.907	2.352	2.712	3.640	4.083
	GLO	0.905	0.194	-0.271	0.544	0.894	1.253	1.537	1.861	2.381	2.866	4.413	5.321
Region 3	GEV	0.808	0.236	-0.195	0.626	0.898	1.219	1.475	1.758	2.189	2.566	3.664	4.252
	GNO	0.894	0.305	-0.631	0.626	0.898	1.232	1.494	1.774	2.176	2.507	3.379	3.803
	GLO	0.904	0.174	-0.301	0.624	0.904	1.203	1.446	1.729	2.193	2.633	4.083	4.957

*Return periods corresponding to nonexceedance probability

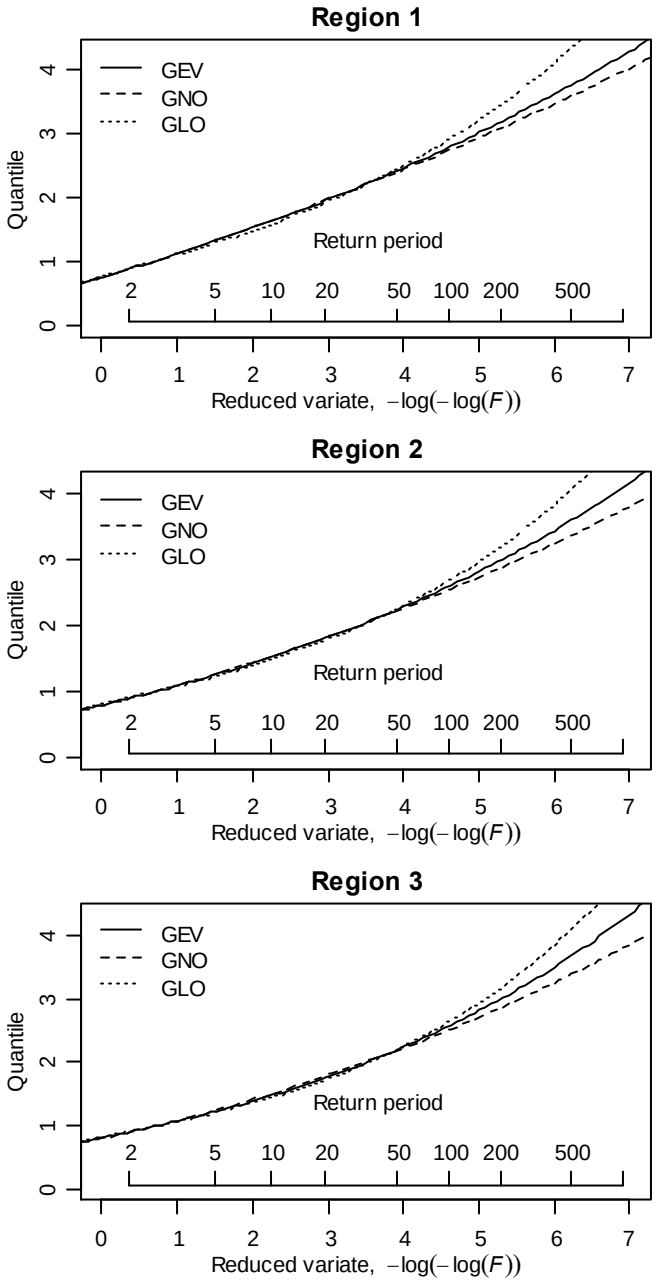

Figure 6: Regional Growth Curves for three Regions

5.6 Decision about best robust regional growth curves

The regional quantile estimates obtained by regional frequency analysis are uncertain and reliable. When more than one regional distribution is selected for quantile estimation then that distribution should be selected giving robust estimates. For this purpose an assessment analysis is made based on Monte Carlo simulation to estimate the accuracy of quantile estimates. An algorithm for simulation of the regional *L*-moment algorithm has been provided by Hosking and Wallis (1997) in section 6.4. According to this algorithm, Monte Carlo simulations are made from such a region which has similar characteristics as

of actual region i.e., having the same number of sites, same record length at each site and regional average *L*-moment ratios as original data. The region used for simulation should also take into account the possible heterogeneity in the region and inter-site dependence if exists (Hosking and Wallis, 1997). In the simulation procedure, quantile estimates are calculated for various nonexceedance probabilities. Let the site *i* quantile estimate, at the mth repetition, be $\hat{Q}_i^{[m]}(F)$ for nonexceedance probability *F,* then the relative error for this estimate is $\{\hat{Q}_i^{[m]}(F)-Q_i(F)\}/Q_i(F)$. This quantity can be averaged over all M repititions to obtain the BIAS and relative RMSE defined by

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

$$R_i(F)=\left[\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\{\hat{Q}_i^{[m]}(F)-Q_i(F)\}}{Q_i(F)}\right\}^2\right]^{1/2}$$

The regional average relative bias, absolute relative bias and relative RMSE of the estimated quantiles are

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

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

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

Other useful quantities, for assessment analysis, are the empirical quantiles of the distribution of estimates which can be obtained by calculating the ratio of estimated to true values, $\hat{Q}_i(F)/Q_i(F)$ for quantiles and $\hat{q}_i(F)/q_i(F)$ for regional growth curves. For a nonexceedance probability *F*, if 5% of the simulated values, $\hat{q}_i(F)/q_i(F)$ lie below some value $L_{0.05}(F)$ whereas 5% lie above some value $U_{0.05}(F)$. Then 90% of the regional growth curve lies within the interval

$$L_{0.05}(F)\le\frac{\hat{q}(F)}{q(F)}\le U_{0.05}(F)$$

Inverting the expression for $q_i(F)$, we have

$$\frac{\hat{q}(F)}{U_{0.05}(F)}\le q(F)\le\frac{\hat{q}(F)}{L_{0.05}(F)}$$

These error bounds have been considered as 90% statistical confidence interval giving the amount of variation between true and estimated quantities. These limits give a reasonable estimate of magnitude of errors expected in regional growth curves and estimated quantiles.

To quantify the accuracy of estimated regional growth curves for all three regions using above measures, we define three separate regions for use in simulation procedure. As GEV, GNO and GLO are suitable choices for region 1, so we perform simulations for region 1 based on all three distributions. For region 1, correlations between sites varies from -0.17 to 1.00 with an average of 0.19. So algorithm for simulation procedure of Table 6.1 of Hosking and Wallis (1997) has been used. The region used in simulation procedure has 11 sites with record lengths as for sites in region 1 having GEV distribution with L-C_v ranging from 0.2073 to 0.3374 for possible heterogeneity and L-C_s = 0.2368 and L-C_k =0.1870. Region1 has been simulated M=10,000 times and regional L-moment algorithm has been used to fit GEV distribution to the data generated for the estimation of quantiles and regional estimates.

The relative bias, relative absolute bias and relative RMSE for regional growth curves have been calculated for various nonexceedance probabilities in Table 7. Similarly simulations were made for region 1 having GNO and GLO distributions and accuracy measures were calculated. L-moment algorithm has also been applied to region 2 and 3 with average correlation between sites 0.31 and 0.30 respectively, having same number of sites, record lengths and L-moments ratios as of actual region 2 and 3. Number of repititions is set to M=10,000 and number of simulations is set to 500. Accuracy measures are shown in Table 8a, 8b and 8c. In Table 8a, simulation results for region 1, at return periods 1, 2, 5, 10 and 20 absolute bias for GNO is relatively high but low at large return periods of 100, 500 and 1000. Relative bias and relative absolute bias are not useful quantities in practice. Simulation results for GNO are producing low relative RMSE for large return periods of 100, 500 and 1000. At low return periods of 1, 2, 5, 10 and 20 GNO has relatively high RMSE than GEV and GLO distributions. Moreover, error bounds (LEB, UEB) for GNO regional quantiles are narrower than the error bounds for GEV and GLO at large return periods. Consequently, it may be concluded that GNO is best for region 1 for quantile estimation for large return periods and GEV for return periods of 1, 2, 5, 10 and 20 based on accuracy measures.

Simulation results for region 2 in Table 8b show that for return periods 50, 100, 500 and 1000, GNO is best and for return periods of 1, 2, 5, 10 and 20, GEV is best for quantile estimation on the basis of relatively low values of absolute bias, RMSE and narrower error bounds of regional quantiles estimated in Table 7.

In Table 8c, the absolute bias, RMSE and narrower error bounds exist for GNO distribution at return periods of 50, 100, 500 and 1000. At return periods of 2, 5, 10 and 20, accuracy measures are approximately same for GEV distribution. So GNO distribution is best choice for quantile estimation in region 3. In Figure 6, GEV and GNO are close at higher nonexceedance probabilities. But for large return periods GNO is best for quantile estimation for all three regions.

Table 8a: Simulation Results of regional growth curves for Region 1

Distribution	F	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	$B^R(F)$	0.020	0.003	0.004	0.005	0.005	0.004	0.004	0.001	-0.001
	$A^R(F)$	0.148	0.020	0.034	0.055	0.071	0.087	0.097	0.118	0.126
	$R^R(F)$	0.176	0.024	0.040	0.064	0.083	0.102	0.114	0.140	0.151
	LEB*	0.380	0.853	1.232	1.470	1.721	2.071	2.354	3.096	3.454
	UEB*	0.626	0.919	1.384	1.779	2.202	2.818	3.349	4.817	5.577
GNO	$B^R(F)$	0.051	0.000	0.005	0.008	0.010	0.013	0.014	0.017	0.018
	$A^R(F)$	0.223	0.007	0.047	0.064	0.076	0.087	0.094	0.105	0.109
	$R^R(F)$	0.278	0.009	0.054	0.074	0.088	0.102	0.110	0.123	0.128
	LEB*	0.317	0.872	1.218	1.459	1.704	2.047	2.313	2.977	3.289
	UEB*	0.674	0.899	1.431	1.828	2.232	2.793	3.235	4.361	4.898
GLO	$B^R(F)$	0.020	0.005	0.004	0.004	0.003	0.001	-0.001	-0.007	-0.010
	$A^R(F)$	0.149	0.020	0.032	0.054	0.071	0.090	0.102	0.127	0.137
	$R^R(F)$	0.178	0.024	0.037	0.062	0.083	0.105	0.120	0.152	0.164
	LEB*	0.378	0.859	1.212	1.441	1.696	2.091	2.450	3.523	4.125
	UEB*	0.626	0.925	1.353	1.732	2.166	2.869	3.521	5.631	6.894

$B^R(F)$=relative bias, $A^R(F)$=relative absolute bias, $R^R(F)$=relative RMSE, LEB=lower error bound, UEB=upper error bound

Table 8b: Simulation Results of regional growth curves for Region 2

Distribution	F	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	$B^R(F)$	0.023	0.005	0.005	0.005	0.004	0.003	0.002	-0.002	-0.003
	$A^R(F)$	0.152	0.027	0.034	0.061	0.084	0.110	0.126	0.163	0.179
	$R^R(F)$	0.180	0.033	0.040	0.071	0.090	0.129	0.150	0.199	0.221
	LEB*	0.405	0.835	1.185	1.392	1.614	1.922	2.182	2.853	3.160
	UEB*	0.694	0.930	1.361	1.757	2.215	2.928	3.578	5.593	6.699
GNO	$B^R(F)$	0.041	0.001	0.005	0.008	0.011	0.014	0.016	0.020	0.021
	$A^R(F)$	0.203	0.011	0.052	0.074	0.090	0.105	0.115	0.132	0.138
	$R^R(F)$	0.246	0.014	0.059	0.085	0.103	0.121	0.133	0.154	0.162
	LEB*	0.359	0.862	1.174	1.397	1.623	1.936	2.188	2.825	3.128
	UEB*	0.735	0.903	1.412	1.813	2.230	2.830	3.318	4.600	5.224
GLO	$B^R(F)$	0.023	0.008	0.005	0.004	0.002	-0.000	-0.003	-0.010	-0.013
	$A^R(F)$	0.154	0.028	0.037	0.063	0.085	0.112	0.131	0.173	0.191
	$R^R(F)$	0.181	0.034	0.042	0.072	0.099	0.132	0.156	0.211	0.235
	LEB*	0.401	0.840	1.170	1.368	1.589	1.928	2.236	3.184	3.709
	UEB*	0.697	0.937	1.332	1.715	2.187	2.963	3.735	6.476	8.171

Table 8c: Simulation Results of regional growth curves for Region 3

Distribution	F	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	$B^R(F)$	0.006	0.005	0.001	-0.003	-0.006	-0.011	-0.014	-0.018	-0.018
	$A^R(F)$	0.060	0.022	0.015	0.033	0.052	0.077	0.099	0.156	0.183
	$R^R(F)$	0.072	0.027	0.022	0.040	0.062	0.095	0.123	0.195	0.230
	LEB*	0.556	0.857	1.176	1.391	1.606	1.906	2.138	2.747	3.027
	UEB*	0.695	0.933	1.262	1.573	1.958	2.574	3.181	5.168	6.315
GNO	$B^R(F)$	0.007	0.005	-0.000	-0.003	-0.005	-0.007	-0.009	-0.010	-0.010
	$A^R(F)$	0.059	0.022	0.015	0.034	0.056	0.074	0.090	0.125	0.139
	$R^R(F)$	0.071	0.027	0.022	0.045	0.068	0.090	0.111	0.156	0.174
	LEB*	0.556	0.854	1.188	1.400	1.620	1.903	2.117	2.683	2.954
	UEB*	0.696	0.930	1.278	1.601	1.975	2.531	3.012	4.403	5.097
GLO	$B^R(F)$	0.008	0.007	0.002	-0.003	-0.008	-0.014	-0.019	-0.028	-0.305
	$A^R(F)$	0.065	0.025	0.016	0.033	0.053	0.081	0.105	0.167	0.195
	$R^R(F)$	0.075	0.030	0.026	0.045	0.069	0.100	0.130	0.206	0.241
	LEB*	0.554	0.860	1.163	1.363	1.578	1.899	2.189	3.043	3.500
	UEB*	0.697	0.944	1.241	1.544	1.934	2.602	3.305	5.876	7.600

6. Summary and Conclusions

There is an extreme variation in climate of Pakistan based on topography. The major part of Pakistan experiences dry climate. Humid conditions prevail over a small area in north. In Pakistan rainfall is caused by Monsoon and Western Depression. Monsoon takes place from July to September and Western Depression from December to March. The normal rainfall of the country is around 300mm of which 140mm rainfall occurs during Monsoon. Regional frequency analysis was performed on a region of Pakistan having 23 sites which are more or less affected by Monsoon.

Initially, the assumptions of regional frequency analysis were tested by time series plots, Mann-Whitney test, Kendall's tau test and Ljung-Box-Q-Statistics. All sites were satisfied by these tests. Site 17, Multan showed autocorrelations because of some drought years. Nonuniformity of climatic conditions did not allow removing this site from analysis.

Site 20, Khanpur was appeared to be discordant. Because of having no gross errors for this site, it was considered in testing heterogeneity of the region. A region of 23 sites did not satisfy Hosking and Wallis's heterogeneity statistics. Since in Pakistan, sites having high elevation receive more rainfall while the sites on plain areas receive less rainfall. The three regions were formed on the basis of mean annual precipitation and elevation

which are acceptably homogeneous. Region 1 were having 11 sites, region 2 and 3 each having 5 sites. Site 2 and 20 were discarded as being producing the high values of heterogeneity statistics. The further analysis was performed on three homogeneous regions.

The *L*-moment ratio diagram, Z-statistic and AWD values produced GEV, GNO and GLO to best for all three regions for quantile estimation. The growth curves for selected distributions were shown in Figure 6. At lower tails, GEV, GNO and GLO are approximately same but for large return periods there is difference between regional quantiles. The best choice is one which gives robust quantile estimates for region.

A regional *L*-moment algorithm of Hosking and Wallis (1997) was performed using Monte Carlo simulation. 10,000 runs of Monte Carlo simulations were performed to calculate the some accuracy measure like, relative bias, relative absolute bias, relative RMSE and error bounds for regional quantiles. On the basis of these measures, for all three regions, GNO was found to be best robust for quantile estimation at large return perionds of 50, 100, 500 and 1000 and GEV was found to be best robust for quantile estimation at return periods of 1, 2, 5, 10, and 20.

Since the actual purpose of frequency analysis of extreme events is the estimation of quantiles at large return periods, that is, quantiles in upper tail of distributions. It is concluded that GNO distribution is best choice for extreme rainfall events in long run.

Hussain and Pasha (2009) also concluded GNO distribution to be best for estimation of annual peak flows in Pakistan which justified the results for annual maximum rainfall.

Adamowski et al. (1996), Lee and Maeng (2003), Koutsoyiannis and Baloutos (2000) found in their studies GEV distribution best for quantile estimation of extreme rainfall.

In Pakistan, runoff occurs essentially on the account of rainfall and the major problem is to store this water for the purposes of electricity and irrigation. The at-site quantiles estimates or design rainfall estimates are very important for engineers who are concerned with the design of hydraulic structures like dam, reservoir, storm sewers etc. In Agromet sector, the effect of extreme rainfall events is of paramount importance especially for crop insurance. So the estimated extreme events are also important for this purpose.

Acknowledgment

Authors also present special acknowledgments for J. R. M. Hosking who guided step by step in this study. A special thank is presented to referees whose constructive comments and suggestions helped in producing a quality research.

References

1. Adamowski, K., Alila, Y., Pilon, P.J., 1996. Regional rainfall distribution for Canada. Atmospheric Research 42, 75-88.
2. Anli, A. S., Ozturk, F. Apaydin, H., Yurekli, K. 2011. Regional frequency analysis of various rainfall data in Ankara province, Turkey, with Index storm method. EMS Annual meeting Abstracts, 8, EMS2011-77-2, 11th EMS/10th ECAM.
3. Ayoade, J.O., 1976. A preliminary study of magnitude of frequency and distribution of intense rainfall in Nigeria. Hydrological Sciences, XXI 3/9, 419-421.
4. Bilham, E.G., 1936. Classification of heavy falls in short periods. British Rainfall 1935, 262-280.
5. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. 1994. Time Series Analysis: Forecasting and Control, Pearson Education, Delhi.
6. Cannarozzo, M., D'asaro, F., Ferro, V., 1995. Regional rainfall and flood frequency analysis for Sicily using the two component extreme value distribution. Journal of Hydrological Sciences 40(1), 19-41.
7. Cavigli, E., Caporali, E., Petrucci, A., 2006. Regional frequency analysis of rainfall extremes in Tuscany. Geophysical Research Abstracts, Vol. 8, SRef-ID: 1607-7962/gra/EGU06-A-09322.
8. Chaudhary, Q. Z., Rasul, G., 2004. Agro-Climatic Classification of Pakistan. Science Vision, 9 (1-2) (Jul-Dec 2003) & No. 3-4 (Jan-Jun 2004), pp 59-66.
9. Chow, K. C. A. & Watt, W. E. (1994) Practical use of the *L*-moments. In: Stochastic and Statistical Methods in Hydrology and Environmental Engineering, vol.1 (ed. by K. W. Hipel), 55–69. Kluwer Academic Publishers, Boston, Massachusetts, USA.
10. Dalrymple, T., 1960. Flood frequency analyses. Water Supply Paper 1543-A, U.S. Geological Survey, Reston, Va.
11. Hirsch, R.M., Helsel, D.R., Chon, T. A., Gilory, E. J., 1993. Statistical analysis of hydrological data, In: D.R. Maidment (ed.) Handbook of Hydrology, McGraw Hill, New York.
12. Hosking, J.R.M, Wallis, J. R., 1993. Some statistics useful in regional frequency analysis. Water Resources Research 29, 271-281.
13. Hosking, J.R.M., 1990. *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal Royal Statistical Society, Series B 52, 105-124.
14. Hosking, J.R.M., 1995. The use of *L*-moments in the analysis of censored data. In Recent Advances in Life-Testing and Reliability, edited by N. Balakrishnan, CRC Press, Boca Raton, Fla, 545-564.
15. Hosking, J.R.M., Wallis, J. R., 1997. Regional Frequency Analysis: An approach Based on *L*-moments. Cambridge University Press, New York.
16. Koutsoyiannis, D., 2004a. Statistics of extremes and estimation of extreme Rainfall: I. Theoretical investigation. Hydrological Sciences 49(4), 575-590.

17. Koutsoyiannis, D., 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrological Sciences 49(4), 591-610.
18. Koutsoyiannis, D., 2007. A critical review of probability of extreme rainfall: Principles and Models, Advances in Urban Flood Management, edited by R. Ashley, S. Garvin, E. Pasche, A. Vassilopoulos, and C. Zevenbergen, 139-166, Taylor and Francis, London.
19. Koutsoyiannis, D., Baloutsos, G., 2000. Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences. Natural Hazards 29, 29-48.
20. Kroll, C. N., Vogel, R. M., 2002. The probability distribution of low streamflow series in the United States. Journal of Hydrologic Engineering. 7, 137-146.
21. Kysely, J. Picek, J., Huth, R., 2007. Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic, Studia Geophysica et Geodeatica 51, 327-344.
22. Kysely, J., Picek, J., 2007. Regional growth curve and improved design values of extreme precipitation events in the Czech Republic. Climate Research 33, 243-255.
23. Lee, S.H., Maeng, S.J., 2003. Frequency analysis of extreme rainfall using *L*-moments. Irrigation and Drainage 52, 219-230.
24. Lin, G.-F., Chen, Lu-H., 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing maps. Journal of Hydrology 324 (1-4), 1-9.
25. Ljung, G. M., Box, G. E. P. 1978. On a measure of a lack of fit in time series models, Biometrika, 65, 297-303.
26. Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18, 50-60.
27. Ngongondo, C. S., Xu, C.-Yu, Tallaksen, L. M. Alemaw, B., Chirwa, T., 2011. Regional frequency analysis of rainfall extremes in Southern malwai using index rainfall and *L*-moments approaches. Stochastic Environmental Research and Risk Assesment, 25(7), 939-955.
28. Norbiato, D., Borga, M., Sangati, M., Zanon, F., 2007. Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood. Journal of Hydrology 345(3), 149-166.
29. Onoz, B., Bayazit, M., 1995. Best-fit distributions of largest available flood samples. Journal of Hydrology 167, 195–208.
30. Parida B.P., 1999. Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution. International Journal of Climatology 19, 1389–1398.
31. Park, J.-S., Jung, H.-S., Kim, R.-S., Oh, J.-H., 2001. Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution. International Journal of Climatology 21(11), 1371-1384.

32. Pearson, C.P., 1993. Application of *L*-moments to Maximum River Flows. The New Zealand Statistician 28(1), 2–10.

33. Peel, M.C., Wang, Q.J., Vogel, R.M., McMahon, T.A., 2001. The utility of *L*-moment ratio diagrams for selecting a regional probability distribution. Hydrological Sciences 46(1), 147-155.

34. Potter, K. W., 1979. Annual precipitation in the northeast United States: long memory, short memory or no memory? Water Resources Researches, 15, 340-346.

35. Rasul, G., Chaudhry, Q.Z., Zhao, S.X., Zeng, Q.C., 2004. A diagnostic Study of heavy rain in twin cities Islamabad-Rawalpindi. Advances in Atmospheric Sciences 21(6), 976-988.

36. Sankarasubramanian, A., Srinivasan, K., 1999. Investigation and comparison of sampling properties of *L*-moments and conventional moments. Journal of Hydrology 218, 13-34.

37. Schaefer, M.G., 1990. Regional analyses of mean annual precipitation in Washington State. Water Resources Research 26(1), 119-131.

38. Singh, C.V., 1998. Long term estimation of monsoon rainfall using stochastic models. International Journal of Climatology 18, 1611–1624.

39. Smithers, J.C., Schulze, R.E., 2001. A methodology for the estimation of short duration design storms in South Africa using a regional approach based on *L*-moments. Journal of Hydrology 241, 42-52.

40. Stendinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events, In: D.R. Maidment (ed.) Handbook of Hydrology, McGraw Hill, New York.

41. Sveinsson, O.G.B., Salas, J. D., Boes, D.C., 2002. Regional frequency analysis of extreme precipitation in northeastern Colorado and Fort Collins flood of 1997. Journal of Hydrologic Engineering 7(1), 49-63.

42. Trefry, C.M., Watkins, Jr., Johnson, D., 2005. Regional rainfall frequency analysis for the State of Michigan. Journal of Hydrological Engineering 10(6), 437-449.

43. Ulrych T.J., Velis D.R., Woodbury A.D., Sacchi M. D., 2000. *L*-moments and C-moments. Stochastic. Environmental Research and Risk Assessment 14, 50-68.

44. Vogel, R.M., Fennessey, N.M., 1993. *L*-moment diagrams should replace product moment diagrams, Water resources Research 29(6), 1745-1754.

45. Vogel, R.M., McMahon, T.A., Chiew, F.H.S., 1993b. Floodflow frequency model selection in Australia. Journal of Hydrology. (Amsterdam) 146(a), 421–449.

46. Vogel, R.M., Thomas, W.O., McMahon, T.A., 1993a. Flood flow frequency model selection in southwestern U.S.A. Journal of Water Resources Planning and Management ASCE 119(3), 353–366.

47. Vogel, R.M., Wilson, I., 1996. Probability distribution of annual maximum, mean and minimum stream flows in the United States. Journal of Hydrologic Engineering 1(2), 69-76.

48. Weaver, J.C., 2006. Frequency of annual maximum precipitation in the city of Charlotte and Mecklenburg country, North Carolina, through 2004. Scientific Investigation Report 2006-5017. U.S. Department of Interior, U.S.

49. Yang, T., Wang, X., Hao, X., Li, H., 2010. Regional frequency analysis of rainfall extremes in the pearl river basin using *L*-moment method. Fuzzy Systems and Knowledge Discovery (FSKD), Seventh International Conference.

50. Yurekli, K., 2005. Regional analysis of monthly rainfalls over Amasya Province via *L*-moments method. GOÜ. Ziraat Fakültesi Dergisi 22(2), 51-56.

51. Zafirakou- Koulouris, A., Vogel, R.M., Craig, S.M., Habermeier, J., 1998. *L*-moment diagrams for censored observations. Water Resources. Research 34(5), 1241–1249.

52. Zalina, M.D., Nguyen, V.T.V., Hashim, M.K., 2002. Selecting a probability distribution for extreme rainfall series in Malaysia. Water Science and Technology 45, 63-68.