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

By

Aamar Abbas

Supervised by

Dr. Ishfaq Ahmad

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2014

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

By

Aamar Abbas

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2014

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

By

Aamar Abbas

A Dissertation
Submitted in the Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
IN
STATISTICS

Supervised by

Dr. Ishfaq Ahmad

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan
2014

Certificate

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

By

Aamar Abbas

A DISSERTATION SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF THE MS IN STATISTICS

We accept this dissertation as conforming to the required standard.

1. ______________________
Prof. ~~Dr.~~ Javed Shabbir
(External Examiner)

2. ______________________
Dr. Zahid Iqbal
(Internal Examiner)

3. ______________________
Dr. Ishfaq Ahmad
(Supervisor)

4. ______________________
Dr. Irshad Ahmad Arshad
(Chairman)

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2014

Dedication

To my family,
For the endless support and patience.

To my Teachers,
For the constant source of Knowledge and
Inspiration.

To my friends,
The ones that are close and the ones that are far.

Forwarding Sheet by Research Supervisor

The thesis entitled **"Regional Frequency Analysis of Annual Total Rainfall in Pakistan Using L-moments"** submitted by **Aamar Abbas** (Registration # 09-FBAS/MSST/F12) in partial fulfillment of M.S degree in Statistics has been completed under my guidance and supervision. I am satisfied with the quality of his research work and allow him to submit this thesis for further process to graduate with Master of Science degree from Department of Mathematics and Statistics, as per IIU Islamabad rules and regulations.

Dated:____________________

Dr. Ishfaq Ahmad,
Assistant Professor
Department of Maths & Statistics,
International Islamic University,
Islamabad.

Acknowledgements

All the praises to Almighty ALLAH the most gracious, the most merciful and the creator of all the creature in the creature. Thanks to ALLAH and his Prophet (P.B.U.H). There are few people who made this journey easier for me with their encouragement and ideas.

I express my gratitude to my supervisor **Dr. Ishfaq Ahmad** for his regardless and inspirational efforts and moral support throughout my research carrier. His sound advices and lots of good ideas were very helpful to me. I would have been lost without him. May ALLAH bless him with all kinds of happiness and success in his life and may all his wishes come true. I am also thankful to chairman Department of Mathematics and statistics **Dr. Irshad Ahmad Arshad** for providing such necessary facilities. I also want express my unbounded thanks to all the faculty members of Department of Mathematics and statistics.

My Deepest gratitude to my family who are the real pillars of my life. They always encouraged me and showed their everlasting love, care and support throughout my life. Their continuous encouragement and, humble prayers, support (both financially and moral) from my father and mother is unforgettable. The love from elder brothers is priceless.
I am also thankful to my old teachers Dr. Tassidaq Hussain, Dr. Aamir Saghir, Dr. Kamran Abbas, Dr. Azhar Saleem and Atif Abbasi for their love, care and support in my life, which has been directly encouraging me for my study. I am also thankful to my friends Muhammad Mushtaq, Fowad khan, Sayyad Anwar, Shahid Ashraf, Irfan Aslam, Sajjad Shabbir and Thair Resham. I am grateful for all the emotional support, entertainment, love and care they provided.
Finally, I express my sincere appreciation to all the helpful staff of the Mathematics Department IIUI. I would like to thanks to everybody who was important to successful realization of this thesis as well as expressing my apology to those that I could not mention.

Aamar Abbas

Table of Contents

List of Figures

List of tables

Abstract

This study involves the estimation of regional rainfall quantiles of 30 sites with the help of L-moments based on index flood regional frequency analysis. Discordancy measure based on L-moments is used to screen the data of these 30 sites. Highly elevated areas of Pakistan receive more rainfall on the basis of this characteristic. The study area is divided into four different regions and L-moment based heterogeneity measure (H) was calculated for each region using four parameters Kappa distribution. The regional quantile estimates are obtained from person type III (PE3), generalized normal (GNO), generalized extreme value (GEV) and generalized logistic (GLO) distributions which are found to be suitable for all four regions based on Z-statistic and L-moment ratio diagram. Accuracy measures for the estimated regional growth curves and quantiles have been calculated for the four candidate distributions, using Monte Carlo simulations. On the basis of this simulation study, it was found that PE3 is most suitable choice for large return period for first three regions and for small return period GNO and GEV are most robust. Similarly for region IV, GEV is best for period of 1, 2, 5, 10 and 20 years return and for period of 50, 100, 500 and 1000 GNO gives most robust estimates.

CHAPTER 1

1.1 Introduction

The climate of any country is measured on the basis of continuing study of its weather conditions. Pakistan is located from Southwest to Northwest at 24^0 to 37^0 north latitude and 61^0 to 76^0 east latitude. Pakistan has been facing a lot of natural disasters in the form of droughts, floods and storms. These massive distributions are caused naturally and cannot be controlled anywhere but we can mitigate the losses caused by them by taking precautionary measure. Droughts and floods are a few of the most damaging factors for a country. Pakistan is an agricultural country with a wide range of irrigation system. This irrigation system mostly depends on rainfall. Whenever rainfalls remain below the required limits, then as a result, rivers and canals dry up. Water as the main source of hydroelectricity since the loss of water also affects energy production as well. The economy of a country mostly depends upon the energy sources. Droughts are negative factors in this regard as they minimize the economic activities. Rural areas and the industries are the most affected and ultimately the raw material provider also suffers a lot. Since the independence of Pakistan it has not seen any major period of droughts expect the drought of 1998-2002, the worst drought in the history of Pakistan. In the months from July to September there is a plenty of rainfall. Moreover, during the summer when temperature increases and snow melts on the mountains which results in the increase of water level in rivers. Consequently, both these factors result in excess of water and give rise to floods. Floods had been one of the most hazardous factors in our country Pakistan has seen many floods. The worst and destructive is the last flood of 2014. Flood of 1950 killed 2910 people, flood of 1977 killed 248 people flood of 1972 killed 1834 people, flood of 2003 due to monsoon rains killed 178 people in Pakistan, flood of 2010 kills 1463 peoples. More than half million people have been affected in Azad Jammu Kashmir and Punjab in flood of 2014. These floods also damaged the infrastructure of our country in billions. In Pakistan there is an interlude and dynamic network of canals and irrigation distributaries but on the other hand there is poor drainage system. The rivers and canals are not protected by the plantation and pavement of the banks. As a result, the soil erosion occurs continuously. Therefore, the capacity of rivers and canals decreases due to this soil deposition. These natural hazards like floods, droughts and storms occur naturally, but to some extent human activities are responsible for these. However, certain measures can be taken and prepared to control these distractions.

(a) Drought

(b) Flood

(c) Rainfall

(d) Wind speed

Fig.1.1: (a) Drought (b) Flood (c) Rainfall (d) Wind speed

The main sources of rainfall in Pakistan are the summer monsoon, the western depression and thunderstorms. Monsoon winds are seasonal winds which cause the rainfall mostly in Southeast Asia. These winds blow from sea to the land and are laid with water. Monsoon continues from July to September in Pakistan and during this period there is plenty of rainfall. In the subcontinent, the weather conditions remain warm, cloudy and moist. In Pakistan the monsoon winds enter after crossing India at tail end of the monsoon.

The second type of rainfall, which affects the country, is the western depression. These are the winds which enter through Iran and Afghanistan from the west side of Pakistan. As these winds face a long journey across land therefore they do not pass over any major bodies of water and they do not carry much moist. The western depression starts in December and continues up to March. Then it becomes slower and weaker. In the higher altitude these western depressions show their effect in the form of snowfall. The most areas in Pakistan have a dry climate, but in some parts of the south humid conditions also prevail. The central parts of the southern areas, Sindh and most regions of Punjab receive comparatively less rainfall annually.

Another factor which is responsible for the rainfall is the thunderstorms. They mainly occur from October to November and April to June. These two periods of the year suffer from the shortage of rainfall. In a moist and unstable atmosphere, these thunderstorms occur. These properties are found only in air masses over equatorial and tropical areas. Thunderstorms give rise to lightening, thundering, strong winds, heavy rains and also carry the hailstorms with them. The weather remains the driest between October and November. Only a few regions of the northern mountains receive a required amount of rainfall and in other areas this rate is lesser. Coastal areas of Sindh and Baluchistan have a moderate temperature and have almost constant humidity throughout the year. Owing to these thunderstorms, the rate of rainfall varies from year to year.

1.2 Regional Frequency Analysis

Frequency analysis is the assessment of how often a specific event will happen. There is a great importance of estimation of the frequency of these extreme events. As there are many sources of uncertainty about the physical processes that produce the observed events. A statistical approach to the analysis of data is often useful. The statistical methods recognize the existence of uncertainty and allow their effects to be quantified. The procedures for the analysis of statistical frequency of a simple set of data are established well. Mostly in this case many of the samples of related data are available for the analysis. For different observed quantiles if the frequencies of occurrence are similar, and therefore the more accurate conclusion can be achieved by analyzing all the data samples together using only single simple. This application is known as regional frequency analysis in environmental application, because the samples of data are analyzed in a way characteristic of observations of the same variable to a number of sites measured in a region are properly defined. The destination of regional frequency analysis is to obtain a useful approximation of the quantile Q_T of a repeating period of scientific relevance. May be this return period is design life of the structure (for floods and rainfall T=50 years) or just about legally mandated design period (in dam safety applications T=10000 years). More broadly, the destination may be to estimate Q_T for a range of return periods or to calculate the entire quantile function. Nonetheless, reliable appraisals of such extreme cases require very long station records, if station data are to be used. It gets a challenge to acquire such data in many regions of the world, but more significant in rising states like Pakistan.

L-moments are defined as the linear combination of probability weighted moments (PWM's) defined by Hosking (1990). The r^{th} L-moment defined by (Hosking, 1990) is given below

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

L-moments have many advantages over conventional moments such as probability distribution for L-moments is meaningful when the distribution has finite mean, requires finite variance for finite standard errors Hosking (1990). Asymptotic estimations applied to sampling distributions are more helpful for L-moments as compared to when used for ordinary moments (Hosking, 1990). L-moments provide a better tool to identify the parent distribution generating data sample (Hosking, 1990). Regional frequency analysis is based on four steps which are screening of data, Identification of homogeneous regions, Choice of a frequency distribution and estimation of the frequency distribution.

1.3 Objective of the study

- To form homogeneous regions from given stations.
- To select the suitable probability distributions for homogeneous regions.
- To predict the annual rainfall in Pakistan after different return periods using the probability model in the given homogeneous regions.
- To identify the destruction due to heavy rainfall in agriculture.
- To identify the region likely to receive minimum or maximum rainfall in future.
- To address precautionary measure in order to avoid any incident in the form of floods and droughts.

CHAPTER 2

Literature Review

The research work which has already been done by someone is based on some old theories, concepts and findings of the other researchers or scientists. The base of these findings may be randomized or observational studies which have been conducted in some particular regions or area. Many researchers of different areas including USA, UK, Malaysia, Pakistan, Iran, India, Turkey, China and many other countries worked on regional frequency analysis of maximum rainfall. In order to support this study, we reviewed the work of these researchers. In this chapter, we include the references starting from 1996 to 2014 which explain the methodology of regional frequency analysis based on L-moments.

Adamowski et al (1996) performed regional analysis of rainfall by using L-moments data across Canada. All of 320 stations of Canada may be considered as one region whose homogeneity was identified with the help of heterogeneity measure H. However, the L-coefficient of variation showed regional variability that was related to the mean annual precipitation (MAP). A regional parent distribution was identified as the general extreme value (GEV), with parameters depending on MAP and storm duration.

Parida (1999) made an attempt to model the random behavior of the Indian summer monsoon. The possibility of using a generalized four-parameter, Kappa distribution was explored. An L-Moment procedure was used for the estimation of reliable rainfall quantiles. As an application of this methodology for the estimation of quantiles based on the Indian summer monsoon rainfall data, observations at 50 stations across the country recorded over a common period (June–September) of 41 years between 1940 and 1980 was undertaken. Using these estimates, isopluvial maps were developed for some commonly used recurrence intervals 20, 50, 100, 200, 500 and 1000 years which could be used by operational hydro-meteorologists.

Smithers and Schulze (2001) identified 15 relatively homogeneous clusters in South Africa using information from 172 rainfall stations which had at least 10 years of short duration rainfall records. The used index storm was the mean of the annual maximum series and a relationship was derived between the index storm and MAP. The accuracy of design storms estimated using this technique was assessed at site not used in the regionalization procedure and error bounds were computed for the quantile growth curves for each of the clusters.

Park et al (2001) used the Wakeby distribution with the method of L-moments estimate on the summer extreme rainfall data at 61guaging stations over South Korea to obtain reliable quantile estimates for several return periods.

Lee Maeng (2003) used 38 Korean rainfall stations to derive the design rainfalls through the L-moment. Kolmogorov–Smirnov ($K-S$) test and L-moment ratio diagram was used to identified the appropriate distribution. Parameters of selected distribution were estimated using Monte Carlo techniques from the observed and simulated annual maximum daily rainfall. The L-moments were then used to design rainfalls.

Yurekli (2005) performed a regional frequency analysis of monthly rainfalls measured by using L-moment approach over Amasya province. Initially, two groups were formed in the province of rain gauge stations (sites) to satisfy the homogeneity condition. Thereafter, heterogeneity statistic (H) test was used to check weather proposed regions homogeneous or not and Z-statistic was used to choose the best fitted distribution.

Trefry et al (2005) applied a regional frequency analysis approach based on L-moments rather than a traditional at-site frequency analysis using method of moments estimator. Data was compiled from 76 hourly recording stations and 152 daily recording stations. The trend and outlier analyses were conducted on both annual maximum series and partial duration data.

Chen et al (2006) used the L-moment method to analysis the regional frequency of low flows. In this study, basin-wide analysis of low flows was conducted for Dongjiang basin using five distributions: generalized logistic, generalized extreme value, generalized lognormal, Pearson type III and generalized Pareto. Each of these had three parameters estimated by the L-moment method. The regional low-flow estimates for each return period were obtained by using the index flood procedure.

Koh et al (2008) derived the optimal regionalization of the precipitation data which could be classified on the basis of climatologically and geographically. In regional analysis of precipitation, 65 rain gauges were used in the study and homogeneous regions are identified by using K-means clustering method. By using this method five homogeneous regions for the precipitation were classified. The L-moment ratio diagram and Kolmogorov-Smirnov tests were applied to identify the regional probability distribution. Monte Carlo simulation technique was used to the design rainfall for regional and at-site analysis. (RRMSE) Relative root-mean-square error, (RR) relative reduction and (RBIAS) relative bias were computed and compared with those resulting from at-site Monte Carlo simulation.

Sabri et al (2009) performed a regional frequency analysis based on L-moment method to maximum daily rainfalls stations in Selangor and Kuala Lumpur. In this study several distributions were used, which include two-parameter normal, lognormal, three-parameter lognormal, logistic, generalized logistic, extreme value type I, generalized extreme value and generalized Pareto distribution. The best fitted distribution was determined on the basis of mean absolute deviation index, mean square deviation index and the L-moment ratio diagram.

Yurekli et al (2009) performed of L-moments method to estimate the regional PDmax of the Cekerek watershed in Turkey for 17 rainfall stations in the region. The discordant test shows that none of the station is discordant in the region. The homogeneity of regions was identified using homogeneity measure H. The Z statistic goodness-of-fit was applied to find the best fitted distribution.

Saf (2009) determined regional probability distribution for the annual maximum flood data observed at 45 stream flow gauging river basins in Turkey using index flood method. Seven sites out of 45were removed from the analysis because there was a trend in the data. A regional analysis was performed on the remaining 38 sites. Discordancy measure was used for screening of the data. The homogeneity of regions was identified using homogeneity measure H by carrying out 500 simulations by using Kappa distribution. A Monte Carlo simulation was used to estimate the results on the basis of the relative bias and relative root mean square error.

Hussain and Pasha (2009) carried out regional flood frequency analysis on the seven sites of the Punjab, Pakistan by using L-moment technique. The data of seven sites was screened with the help of discordancy measure. Homogeneity of the region was calculated using four parameter Kappa distribution. Z statistic and L-moment ratio diagram were used to find appropriate distribution for generalized logistic, generalized extreme-value, generalized normal, Pearson type III, generalized Pareto and five parameter Wakeby distributions

Ngongondo et at (2011) performed regional frequency analysis at 23 selected rainfall stations in Southern Malawi of 1, 3, 5 and 7 day annual maximum rainfall series. Three homogeneous regions were determined using cluster analysis. The regional index rainfall quantiles were derived by applying L-moments method. Three regions were identified as homogeneous on the basis of heterogeneity measure *H*. Heterogeneity measure was based on Monte Carlo simulations with regional average L-moment ratios fitted to the Kappa distribution.

Seckin et al (2011) applied index flood procedure to the annual flood peaks data taken at all stream-gauging stations in Turkey having at least 15-year-long records. The data was screened with the help of discordancy measure. Homogeneity of the total geographical area of Turkey was tested using the L-moments based heterogeneity measure using the four parameter Kappa distribution. The parameters of the distributions were estimated using L-moments. L-moment ratio diagrams and the Z statistic goodness-of-fit criteria were used for the identification of robust distribution.

Hassan and Ping (2012) examined the use of cluster analysis and L-moment methods together to quantify regional rainfall patterns of Luanhe basin with the help of annual rainfall of 17 stations for the period of 1932-1970. The cluster analysis followed "Ward's method" and showed seven regions of rainfall. The regional frequency distributions were selected for homogeneous regions using the goodness-of-fit test Z^{DIST}.

Hassan et al (2012) performed regional rainfall frequency analysis using the index flood procedure based on L-moments approach. Mean annual rainfall data observed at 15 gauged sites on the Sudan was investigated. The purpose of the study was to investigate and derive hydrological homogenous region or regions, to identify and establish the regional statistical distribution and to extend the methodologies to the case of ungauged site. For this purpose, five distribution functions generalized logistic, generalized normal, generalized extreme-value, Pearson type-3 and generalized Pareto distributions were used. Regional dimensionless growth curves for the identified regions were derived and results are evaluated on the basis of RMSE through the use of Monte Carlo simulation.

by heterogeneity measure H. The L-moment ratio diagram and Z statistic the goodness-of-fit test were used for the selection of robust regional frequency distribution.

Anli (2013) performed frequency analysis based on L-moment approach. The data of annual daily maximum rainfall was collected form 63 stations of Turkey belongs to Yesilirmak basin. According to discordancy measure six homogeneous groups were formed for the regional frequency analysis. The best fitted frequency distribution was selected by using Z statistic the goodness-of-fit measure based on L-moments approach. The criteria named as Z statistic was estimated for several candidate frequency distributions.

Shahzadi et al (2013) studied the estimation of regional rainfall quantiles of 23 sites using L-moment based index flood regional frequency analysis. An L-moment based discordancy measure was used to detect discordant sites. Since in Pakistan, highly elevated area receives more rainfall. On the basis of this characteristic, the study region was divided into three regions which satisfied the L-moment based heterogeneity statistics using Monte Carlo simulations from Kappa distribution.

Rostami (2013) used flood frequency analysis based on L-moment approach in west Azerbaijan province basins. Ward hierarchical cluster method was used for the identification of homogeneous regions. The west Azerbaijan province was divided into four regions. By the help of L-moment ratios the parameters of regional frequency distribution were estimated in these regions. For the selection of appropriate distributions L-moment diagram, Z statistic goodness-of-fit test and plotting position methods were used.

Dubey (2014) performed regional flood frequency analysis based on L-moments at Narmada basin. The parameters of Generalized Extreme value distribution are estimated by the method of L-moments. Generalized extreme value distribution was used to develop regional flood frequency relationship for the chosen basin. A relationship between catchment area and mean annual flood was obtained. This relationship was further employed on regional flood to generate formula for ungauged catchment of Narmada basin. By the help of this regional flood formula T-Year return period flood can be estimate, knowing only one parameter that is catchment area of ungauged watershed.

CHAPTER 3

Material and Methods

3.1 Study Area and Data

The daily rainfall amount taken from Pakistan Meteorological center Karachi is measured by rain gauges in millimeters, from which annual total rainfall (ATR) series have been constructed for the proposed study. The geographical location of these meteorological observatories is given in Fig.3.1. These sides include province Sindh, Punjab, Baluchistan, KPK and northern areas of Pakistan. The record length of ATR series varies from 29 to 51 years.

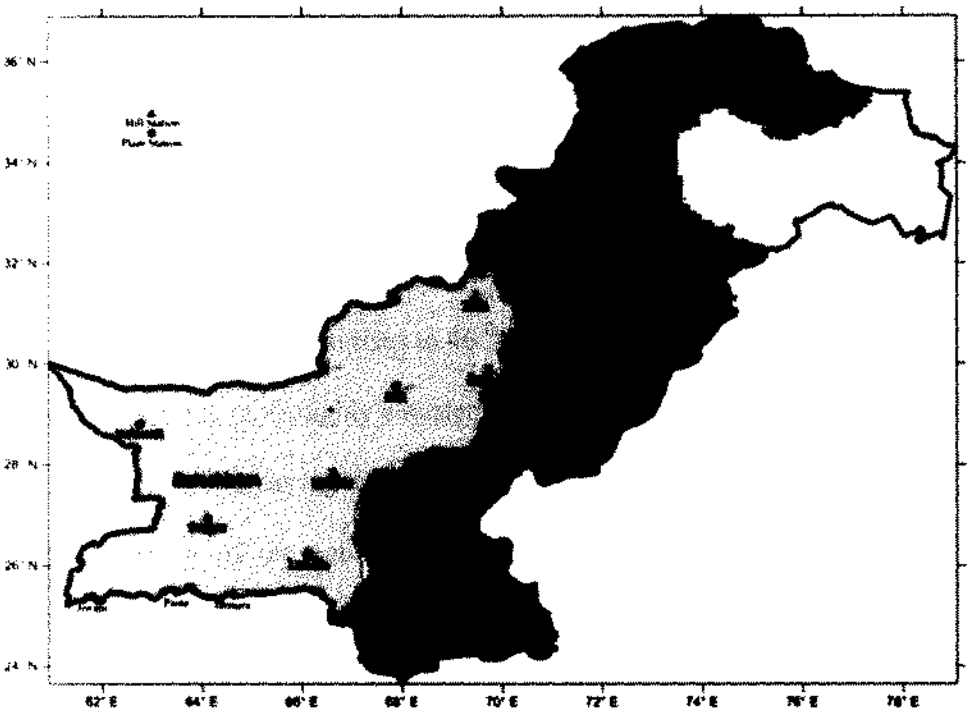

Fig. 3.1: Location of sites used in this study

Baluchistan is the largest province of Pakistan, which is situated in the southwest of Pakistan. The climate of highly elevated areas of Baluchistan is very hot in summer and cold in winter, especially northern districts of Baluchistan are extremely cold in winter. In these areas

Northern areas of Pakistan lie between Hindukush and Karakoram containing high mountains which are covered with snow. The climate of these areas varies from region to region, some areas like Gilgit and Chilas are very hot in summer during the day but cold at night. Some areas are cold during summer like Astor etc. In winter season monthly rainfall in these areas ranges from 25-75 mm whereas in summer it ranges from 10-20 mm.

3.2 Application of Regional Flood Frequency Analysis

Regional Frequency Analysis is based on following steps

i) Initial Screening of Data

ii) Identification of homogeneous region

iii) Choice of an appropriate probability distribution

iv) Estimation of Homogeneous Region

3.3 Initial Screening of Data

In any statistical analysis of data, the fundamental step is to check that whether the data are suitable for analysis or not. There are two necessary conditions for frequency analysis of the data collected at a particular site. First collected data must truly reflect the quantity being measured and second, collected data must be derived from the same frequency distribution. The purpose of initial screening of the data is to prove that these conditions are obeyed. The nature of the problems that may have an influence on the collected data rely on which kind of data was taken under measurement. It means that the frequency distribution from which data were sampled is not constant over time and the frequency analysis of the data will not be a valid basis for estimating the probability distribution of future measurements at the site. Suppose a group consists of N sites. Let $u_i = [t^{(i)} \; t_3^{(i)} \; t_4^{(i)}]^T$ be a vector having t, t_3 and t_4 values for any site i. The superscript T represents the transposition of a vector. Let

$$\bar{u} = N^{-1}\sum_{i=1}^{N} u_i \qquad (3.3.1)$$

is the average of the group. The matrix of sum of square and vector products is defined as

$$A = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T \qquad (3.3.2)$$

Now the discordancy measure of site i is written as

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u}) \qquad (3.3.3)$$

Where site i is assumed to be discordant when D_i is large. Here, the term large is defined in terms of total number of sites in the group. Hosking and Wallis (1997) proposed that a site be treated as discordance if its D_i value is greater than a particular value called critical value as given below

Table (3.1) Critical values for the discordancy statistic D_i

Number of sites in region	Critical value
5	1.333
6	1.648
7	1.917
8	2.140
9	2.329
10	2.491
11	2.632
12	2.757
13	2.869
14	2.971
≥ 15	3.000

There are two applications of the discordancy measure. The first application involves large group of sites occupied by some large geographical area. Those sites which have gross error in their data values will be considered as discordant with respect to the sites. At this stage sites considered as discordant closely inspected for errors in the counting or recording of the data. The second application is that, when homogeneous regions are being temporarily identified, for each site, discordancy measure can be calculated in a selected region. It is possible to move one site to another region when any one site of the region is discordant with the region entirely.

3.4 Identification of Homogeneous Regions

The most difficult step in case of many sites for regional frequency analysis is the identification of homogeneous regions. It demands the greatest interest of subjective judgment. By homogeneity condition it is meant that the frequency distribution of sites is same. This condition is achieved by dividing the sites into disjoint groups. The most commonly grouping methods applied in regional frequency analysis for making similar sites groups are given below

- Geographical Convenience
- Subjective Partitioning
- Objective Partitioning
- Cluster Analysis

There are some other methods which can be used for making groups of similar sites. These methods carry out the same objective as discussed above. These are multivariate analysis methods.

Geographical Convenience: In this method regions are selected on the basis of administrative area and these selected regions are assumed to be the set of contiguous sites as Matalas et al

(1975). This approach looks arbitrary and subjective, and it gives the impression of physical integrity.

Subjective Partitioning: This method involves the investigation of site characteristics. It is actually a small scale study approach. Annual maximum precipitation data for sites in Washington State analysed by Schaeffer (1990) and formed regions by grouping together sites with similar values of mean annual precipitation. This method is based on subjective approach and heterogeneity measure is used to check the resulting region. This method is not considered better when someone uses at-site statistics instead of site characteristics.

Objective Partitioning: In this method sites are assigned one of the two possible groups to form regions depending on whether the characteristic of the chosen site does or does not cross some threshold value. The purpose of choosing a threshold value is to minimize heterogeneity within-group for example likelihood-ratio statistic, which is used by Wilshire (1985) or a variation of sample L-CV and L-skewness within-group used by Pearson (1991). The groups are further subdivided until to reach a set of homogeneous regions. There is no specific reason to select this approach, but it can be more helpful in conjunction with a later assessment of whether the final region is homogeneous.

Cluster Analysis: Based upon statistical multivariate analysis, cluster analysis is a technique used to divide the data into groups. This technique finds successful applications in the regional frequency analysis. It involves the formation of regions by assigning a data vector to each site and these sites are either divided or added into groups of identical vectors. The combination of at- site statistics and site characteristics or each of these two exclusively, but only site characteristics are preferably used in cluster analysis. For example, cluster analysis was used by

De Coursey (1973) to investigate site characteristics to form groups of sites with identical flood response. Regions for floods frequency analysis were derived by Burn (1989) using cluster analysis of at-site statistics. The most commonly used approach for making regions from large data is the cluster analysis of site characteristics. It proceeds with multiple major variants and subjective decision taken at different stages.

3.4.1 Heterogeneity Measure and Formal Definition

The objective of heterogeneity measure is to calculate the degree of heterogeneity in a group and also to examine whether sites are considered as homogeneous or not. In a homogeneous region population L-moment ratios for all sites are same. However due to sampling variability there sample L-moment ratios be different.

Let there be N sites in a proposed region, with a site i having record length denoted by n_i and sample L-moments ratios, $t^{(i)}$, $t_3^{(i)}$ and $t_4^{(i)}$. The t^R, t_3^R and t_4^R denote the regional average L-CV, L-skewness and L-kurtosis respectively, and their weights are proportional to the sites recorded length. For example

$$t^R = \sum_{i=1}^{N} n_i t^{(i)} \Big/ \sum_{i=1}^{N} n_i \qquad (3.4.1.1)$$

The following formula is used to calculate weighted standard deviation in case of sample L-CVs

$$V_j = \left\{ \sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2 \Big/ \sum_{i=1}^{N} n_i \right\}^{\frac{1}{2}} \qquad (3.4.1.2)$$

The observed and simulated dispersion of *L*-moments for different sites are compared by heterogeneity measure. For this purpose Monte Carlo simulation is made using four parameter Kappa distribution. The next step is to fit a Kappa distribution to the regional average L-moments ratios which are 1, t^R, t_3^R and t_4^R. At this stage simulate a large number denoted by N_{sim} of realization of a region having N sites and their frequency distribution is kappa distribution. These simulated regions are homogeneous with no serial-correlation or auto-correlation and both sites have same record length, and weighted standard deviation (V_j) is measured for all simulated regions. The mean and standard deviation of N_{sim} values of V_j are determined from simulations which are denoted by μ_{vj} and σ_{vj}. Heterogeneity measure H_j ($j = 1,2,3$) can be calculated as

$$H_j = \frac{(V_j - \mu_{vj})}{\sigma_{vj}} \qquad (3.4.1.3)$$

H_1, H_2 and $H_3$are heterogeneity measures based on observed weighted standard deviation of values $V1$, observed average of (t/t_3) distance $V2$ and observed average of (t_3/t_4) distance $V3$ respectively. According to Hosking and Wallis (1997), region is said to be heterogeneous if the value of $H_j \geq 2$, possibly heterogeneous if $1 \leq H_j \leq 2$ and "acceptably homogeneous" if $H_j < 1$. The Kappa distribution is defined as

$$f(x) = \alpha^{-1}\left[1 - k(x-\varepsilon)/_{\alpha}\right]^{1/_{k-1}} [F(x)]^{1-h} \qquad (3.4.1.4)$$

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

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

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

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

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

$$-\infty < x < +\infty \qquad if\ h \le 0, k = 0$$

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

The Kappa distribution is a generalized distribution and it produces many distributions, if its parameter values are changed. For example, when $h = 1$ it becomes generalized Pareto distribution, generalized Extreme value distribution if $h = 0$ and generalized Logistic distribution if $h = -1$. Kappa distribution becomes an exponential distribution if $h = 1$and $k = 0$, it becomes a Gumble distribution when $h = 0$ and $k = 0$, a Logistic distribution if $h = -1$ and $k = 0$, and similarly when $h = 1$ and $k = 1$ uniform distribution arises. Four-parameter reverse exponential distribution arises when $h = 0$ and $k = 1$.

3.5 Selection of Frequency Distribution

In a regional frequency distribution, there is no single true distribution that can be applied to each data site. The purpose of the use of regional frequency analysis is to find a distribution that can produce correct quantile estimates for all data sites rather than to identify a true distribution of these data sites. It is assumed that there may be a specific range of return periods for which quantile estimates are needed. When some extreme events such as floods, storms and droughts etc are analysed, quantile estimates have a particular interest under one tail of the distribution. There are multiple families of distributions that can be applied to a regional frequency analysis data set. Each family of this distribution can act as a candidate and the stability of such candidate can be estimated by accounting their ability to reproduce features of a given data that have a special importance in modeling. A distribution may have following important properties from application point of view.

It is physically impossible to estimate quantiles of 100,000 years return period, but possible for 100 years return period. The requirement of the accuracy of quantile estimates with real interest return period is that the distribution must have physical upper bound suitable for practical application. In an unbounded distribution the true distribution function is approximated by an unbounded distribution rather than by parametrizable bounded distribution. It is helpful to use a distribution having the capability of modeling bounded data when the upper bound of true distribution is judged by the observed data.

The considerations applied to the upper bound can also be applied to the lower bound in the same manner, whether the lower bound is not sufficient to approximate it explicitly. In contrast to the upper bound, sometime lower bound is known to be zero. In case of quantile of interest close to zero the requirement of the lower bound of the regional frequency distribution is worthwhile.

Upper tail estimation of frequency distribution has a special interest from application point of view. It has same interest in the case where the amount of data is not enough to find the shape of the upper tail with an accuracy level. The tail weight also called the behavior of the probability density function $f(x)$, as independent quantile (x) rises is important because it decides the rate of which quantiles approach to large values as the return period is generalized outside the boundary range of the data.

The conditions applied to the upper tail of the distribution are also applied to the lower tail of the distribution. There are many cases where it is better to count a set of range distribution with variant tail weights. The lower tail of the distribution is unsuitable where interests focus on the upper tail of the distribution. For annual maximum stream flow data upper tail quantiles of exponential distribution gives better results where the shape of the lower tail of the distribution accepts little similarity to that of the data.

There may be a number of zero values contained by precipitation total. In case of lower tail of the distribution, if estimates of quantile are considered to be important then the distribution allowing for a nonzero proportion of values to be zero should be used. A suitable distribution can easily be derived from standard distributions. When the assumed regional frequency distribution have the following form

$$F(x) = \begin{cases} 0, & x < 0, \\ p + (1-p)G(x), & x \geq 0, \end{cases} \qquad (3.5.1)$$

then a mixed distribution can be applied. Here p represents probability of a value to be zero and function $G(x)$ is known as cumulative distribution function of the values other than zero values. It is not necessary that this function has a lower bound of exactly zero value. The parameter p can be evaluated from the zero value present in the data collected from a given region and the function $G(x)$ can be checked using the nonzero values of regional L-moments.

3.5.1 Methods for Testing the Goodness of Fit Measure

There are different available methods which are used to test the goodness of fit of a given distribution for a single sample. These methods are quantile-quantile plots, chi- square distribution, Kolmogorov-Smirnov test and L-moment ratio diagram. There are also some other general tests and these are based either on conventional moments or L-moment statistic, some of these methods are used in regional frame work. The fitting of a required regional frequency distribution to the data of each site of the region can be measured from goodness-of-fit statistics obtained for each site. The statistic obtained in this way is then combined to form a goodness-of-fit statistics. L-moment ratio diagram is one of the most commonly method 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 used to choose a suitable distribution that will describe a set of variables as a distinguishable relation between L-moment ratios existing for all theoretical probability distribution prescribed by Hosking and Wallis (1997).

Consider three-parameter distributions in regional analysis that are generalized logistic, generalized normal, generalized extreme value, generalized Pareto and generalized Pearson type III. Each distribution is fitted to regional average L-moments denoted by τ_4^{DIST} where τ_4 is the L-kurtosis of fitted distribution and DIST denotes distributions that may be GNO, GP3, etc. Z^{DIST} statistic for a goodness of fit measure for a distribution is defined as

$$Z^{DIST} = {(\tau_4^{DIST} - t_4^R + B_4)}/{\sigma_4} \tag{3.5.2}$$

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R\right) \tag{3.5.3}$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{\sum_{m=1}^{N_{sim}} \left(t_4^{[m]} - t_4^R - N_{sim} B_4^2\right)\right\}\right]^{1/2} \tag{3.5.4}$$

B_4 and σ_4 denote simulated regional bias and standard deviation of t_4^R respectively. These simulations are constructed from fitted Kappa distribution to regional L-moments. The criterion for best fitted distribution is $|Z^{DIST}| \leq 1.64$. It is possible that many distributions meet this criterion, but the most suitable is one whose $|Z^{DIST}|$ is near to zero.

The following procedure is used for a homogeneous region. First of all Z statistic is estimated for all distributions. Mark all distributions for which following condition is satisfied $| Z^{DIST} | \leq 1.64$. The next step involves the construction of growth curve for best fitted distributions. Distribution are said to be adequate if these curves are nearly equal. In the case when growth curves are not nearly equal then a problem arises from the deficiency of data. If two models show differences it means they are statistically insignificant but important for operational point of view. In such a case when it is not possible to identify that which model is best, the role of robustness becomes important. It is better to chose four parameter Kappa or five parameter Wakeby distributions rather than three parameter distributions because these distributions are more robust in frequency distribution of a given homogeneous region.

3.6 Estimation of Homogeneous Regions

There are several techniques used to fit a distribution to the data collected from a homogeneous region. These methods are described as follows.

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

$$Q_i(F) = \mu_i q(F), \qquad i = 1,2, \dots \dots \dots . N$$

In above equation μ_i defines the site dependent scale factor, the index flood. Let $\hat{\mu}_i$ is denotes the estimate of scale factor at site i. The dimensionless rescaled data are $q_{ij} = {Q_{ij}}/{\hat{\mu}_i}$ where

$j = 1,2, \dots\dots n_i\,, i = 1,2, \dots .N$

The regional average mean is set to 1 that is $l_1 = 1$. The distribution is fitted by comparing the L-moment ratios denoted by λ, $\tau, \tau_3, \tau_4, \dots\dots$, of the distribution to the average L-moment ratios specified by $t^R, t_3^R, t_4^R, \dots\dots$.

$$t_r^R = \sum_{i=1}^{N} n_i t^{(i)} \Big/ \sum_{i=1}^{N} n_i \qquad r = 3,4, \dots\dots \tag{3.6.1}$$

The quantile function is denoted by $\hat{q}(.)$ for the fitted regional frequency distribution. For a given site i the estimates of μ_i and $\hat{q}(.)$ are combined to obtain the quantile estimates for the site. The nonexceedance probability F for the estimate of the quantile is written as

$$\hat{Q}(F) = l_1^{(i)} \hat{q}(F) \tag{3.6.2}$$

The regional quantile estimates measured from regional frequency analysis are not precise but reliable. When multiple regional distributions are chosen for the measurement of quantile estimation then the distribution is selected in such a way that it gives robust estimates. An assessment analysis is performed based on Monte Carlo simulation to measure the accuracy of quantile estimates. Hosking and Wallis (1997) have provided the regional L-moment algorithm for simulation. This Algorithm states that Monte Carlo simulation are performed for such a region in which there are identical characteristics that of actual region. The simulation procedure involves the calculation of quantile estimates for different nonexceedance probabilities. The site

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 repetitions M to get BIAS and relative RMSE as given by

$$B_i(F) = \frac{1}{M}\sum_{m=1}^{M}\frac{\{\hat{Q}_i^{[m]}(F) - Q_i(F)\}}{Q_i(F)} \qquad (3.6.3)$$

$$R_i(F) = \left[\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\left\{Q_i^{[m]}(F) - Q_i(F)\right\}}{Q_i(F)}\right\}^2\right]^{\frac{1}{2}} \qquad (3.6.4)$$

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

$$R^R(F) = \frac{1}{N}\sum_{i=1}^{N} R_i(F) \qquad (3.6.5)$$

$$B^R(F) = \frac{1}{N}\sum_{i=1}^{N} B_i(F) \qquad (3.6.6)$$

$$A^R(F) = \frac{1}{N}\sum_{i=1}^{N} |B_i(F)| \qquad (3.6.7)$$

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.

$$L_{0.05}(F) \le \frac{\hat{Q}(F)}{Q(F)} \le U_{0.05}(F) \qquad (3.6.8)$$

$$\frac{\hat{Q}(F)}{U_{0.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{0.05}(F)} \qquad (3.6.9)$$

The 90% statistical confidence interval gives amount of variation between true and estimated quantities. This interval is also known as error bound.

CHAPTER 4

Results and Discussion

4.1 Basic Assumptions of Regional Frequency Analysis

There are various basic assumptions those are taken into consideration in regional frequency analysis. The purpose of these assumptions is to test observations at any site are independent, stationary and identically distributed. These assumptions are checked for different kinds of data practically for annual totals or extreme data. In ATR series different plots and tests are applied which are given below.

4.1.1Time Series Plots

The time series is defined as a set of well-ordered numbers collected for regular durations of time. Initially, time series plots of annual total rainfall (ATR) in mm are made out on 30 sites of Pakistan to check variation or trend in the data. There are some sites that show large variations in ATR such as Choor, Karachi, Rohri and Chitral which means data of these sites is stationary. Fig. 4.1 shows variations in ATR in Choor and Karachi.

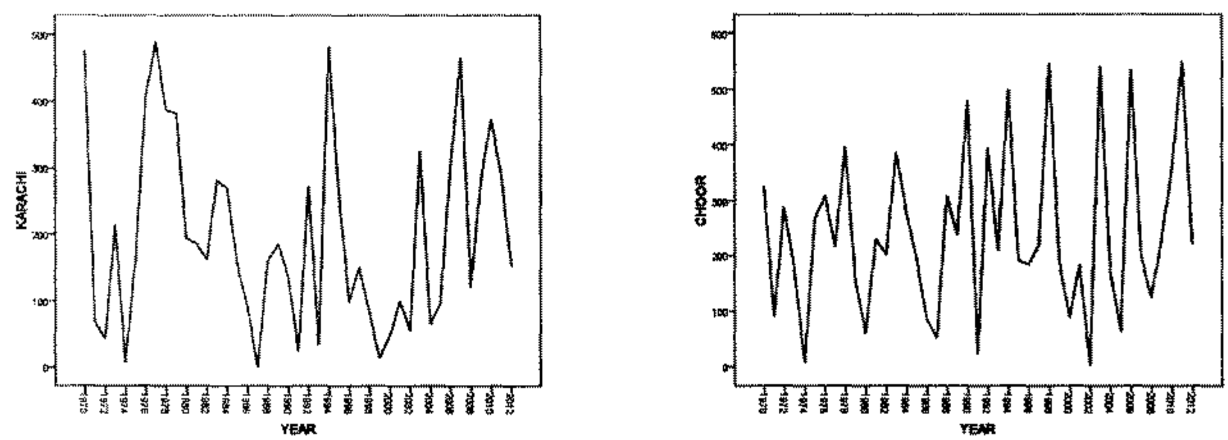

Fig. 4.1 ATR time series plots for Karachi and Choor

The variation in these sites is due to sudden changes in conditions of climate. In time series plots it is also noted that some sites show up and downward trends as shown in fig. 2 this trend is due to changes in climate conditions, which was ignored.

4.1.2 Mann-Whitney Test

Any series of observation follow the same distribution is one of the basic assumption of regional frequency analysis, which is tested by Mann-Whitney Rank sum test (1947). Mann-Whitney test is a nonparametric test employed to check whether two populations are identical or not. For this purpose the series of annual total rainfall ATR is divided into two different groups. In our analysis all of the 30 sites of Pakistan show insignificant result. So we can say that the series of ATR of all sites follows the same distribution or all these sites have identical distributions. The Mann-Whitney test statistic is given below

$$U = n_1 n_2 + [n_1(n_1+1)/2] - T \quad (4.1.2.1)$$

n_1 and n_2 denotes the size of first and second samples respectively where T denote the sum of ranks for the first sample

4.1.3 Kendall's tau Test

Kendall's tau rank correlation method (Hirsch et al.1993) is one of the nonparametric tests applied to estimate the influence of one estimated quantile on another and vice versa. Kendall's tau test is also a trend test. To check the assumption of correlation this method is applied to the series of ATR with time. All of the 30 sites of ATR in Pakistan did not reject the hypothesis which means that there is no trend in the series. We conclude that ATR series for all sites shows insignificant trends. The Kendall's tau test statistic is given below

$$\tau = 1 - \frac{4D}{n(n-1)} \qquad (4.1.3.1)$$

where D= number of discordant pairs and n is the sample size.

4.1.4 Ljung-Box Q-Statistics

Ljung-Box Q-Statistics which was developed by Ljung-Box (1978) is applied to the series of annual total rainfall for all 30 sites to check serial correlation. For this, hypothesis is defined as all autocorrelations are equal to zero up to certain lag, the lags are not more than $n_i/4$ (Box et al 1994) n_i is defined as the length of *ith* site. All 30 sites show insignificant correlation results. The test statistic is given below

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{p}_k^2}{n-k} \qquad (4.1.4.1)$$

where n represent the sample size, $\hat{\rho}_k$ represents sample autocorrelation at lag k and number of lags are denoted by h being tested.

From initial screening, we say that this data is suitable for regional frequency analysis. Reason is that all of 30 sites satisfy the conditions of Mann-Whitney test, Kendall's test and Ljung-Box-Q test. As we see that the series of ATR is non-uniform this is due to emergent changes in conditions of climate. Chaudhary and Rasul (2004) found no significant correlation for data collected under internal annual daily event, but correlation existence is possible only when climate conditions are non-uniform.

4.2 Screening of the Data Using Discordancy Measure

In screening of the data, we consider a single region consisting of 30 sites and discordancy statistic D_i was calculated for each site using Equation (3.3.3).The values of D_i for all stations are smaller than the critical value, so we say that there is no trend or no outlier in the region. The largest value of D_i is 2.53 with high L-CV (t) and L- skewness (t_3). The L-CV and L-skewness of annual total rainfall of 30 sites in Pakistan are shown in Table 4.1.

Table 4.1 L-moment Ratios and Discordancy Measures

Site No	Site 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	Padadin	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	Kanpur	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

As we know that discordancy measure is based on three coefficients L-CV, L-Skewness and L-Kurtosis. There are many sites with high L-CV and L-skewness, but the data is looking to be correct and there is no evidence of gross error. Site 12, Padidan has extremely high values of L-CV and L-skewness, but it does not discordant with other site. In Fig.-4.2 scatter plots are drawn for sample skewness versus sample CV and sample skewness versus sample kurtosis. The purpose of Fig. 4.2 is to check whether any site is discordant or not. From Fig. 4.2 site 12 Padidan appears as the site having high L-skewness as compare to other sites. Similarly site 19 Chilas appears as the site having high L-kurtosis but these sites are not discordant.

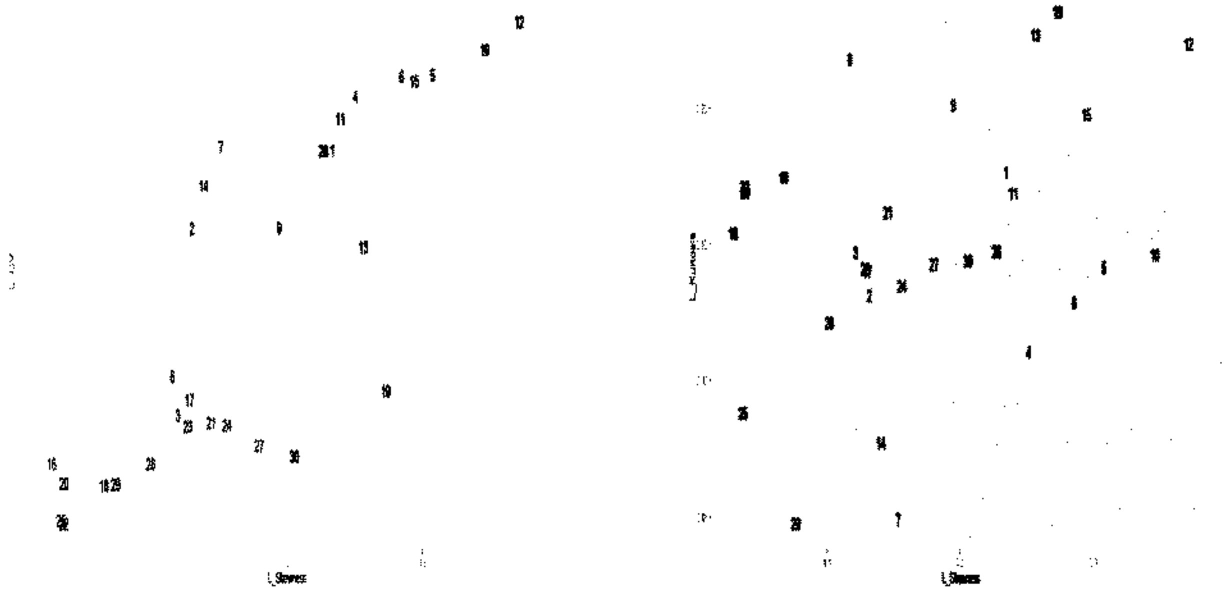

Fig. 4.2 L-moment Ratios for 30 sites

In Fig. 4.3, site 12 Padidan is compared with its nearest site that is Moen-jo-doro site 10. Due to extreme rainfall in the year 1992 and 1994 for site 12, the values of L-C_v are large. So there is no clear reason to discord site 12, Padidan.

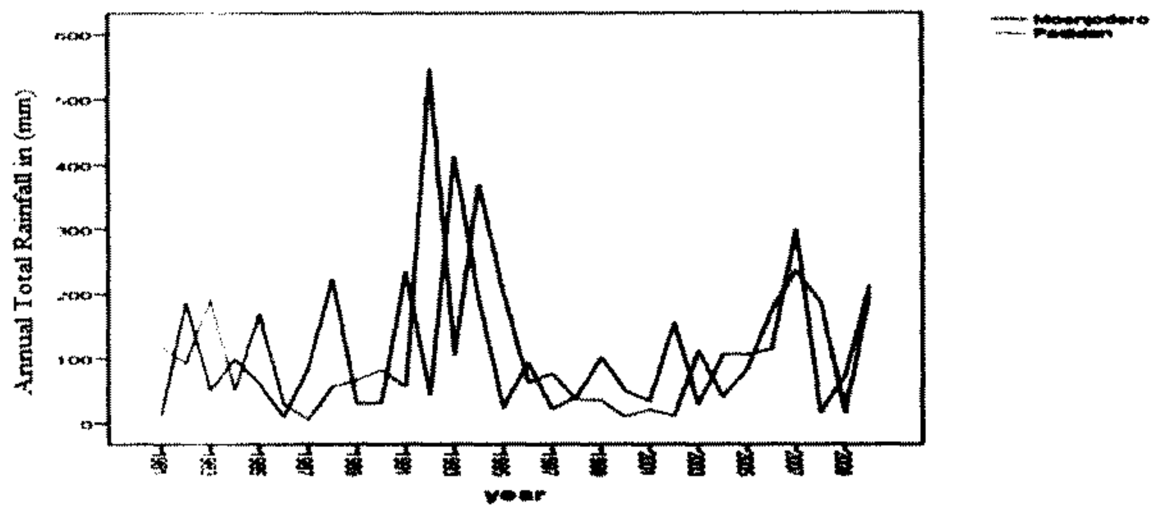

Fig. 4.3 Time Series data for Padadin (Site 12) and Moen-jo-doro (Site 10)

4.3 Formation of Homogeneous Regions and Regional Heterogeneity

Identification of homogeneous regions is the second step in regional frequency analysis. This step has many complexions and it needs special precautions. By homogeneity condition we mean that all the sites have identical frequency distribution. Heterogeneity measure presented by Hosking and Wallis (1993) is used to measure the degree of heterogeneity of sites included in groups to check whether group of sites is treated as homogeneous or not. In heterogeneity measure observed and simulated dispersion of L-moments are compared for N sites where N denotes the number of sites under study.

In the formation of homogeneous regions we consider a single region of 30 sites and the heterogeneity statistic was calculated by using Equation (3.4.1.3). The value of H_1 *is* 20.78 this value is greater than 2, so entire set of sites is decidedly heterogeneous. Possibility of regional frequency analysis is rejected treating a single region of all possible sites. Regional frequency analysis is not applied under this situation because we cannot fulfill the required criteria defined by Hosking and Wallis (1997). This reference implies that region will be acceptably homogeneous if $H_j < 1$, possibly heterogeneous if $1 \leq H_j < 2$, and definitely heterogeneous if $H_j > 2$. According to Hosking and Wallis (1997) H_1 is more efficient to detect homogeneity of region as compare to H_2 and H_3, the value of H_2 is greater than 1 which means region is definitely heterogeneous. Table of heterogeneity statistic for a single region of 30 sites is given in Table 4.2.

Table 4.2: Heterogeneity Statistics

No. of Sites	H_1	H_2	H_3
Heterogeneity statistic for 30 sites	20.78*	5.83*	0.48

4.3.1 Formation of Homogeneous Regions

The characteristics of the study are that a highly elevated site's receives high average annual rainfall and site's with low elevation receives low average annual rainfall. The plot for all these 30 sites of elevation and average annual rainfall is shown in Fig. 4.4 and the numbers in this plot indicate corresponding sites. Fig. 4.5 indicates the map of Pakistan with respect to Elevation of these 30 sites and the numbers in this plot indicate corresponding sites.

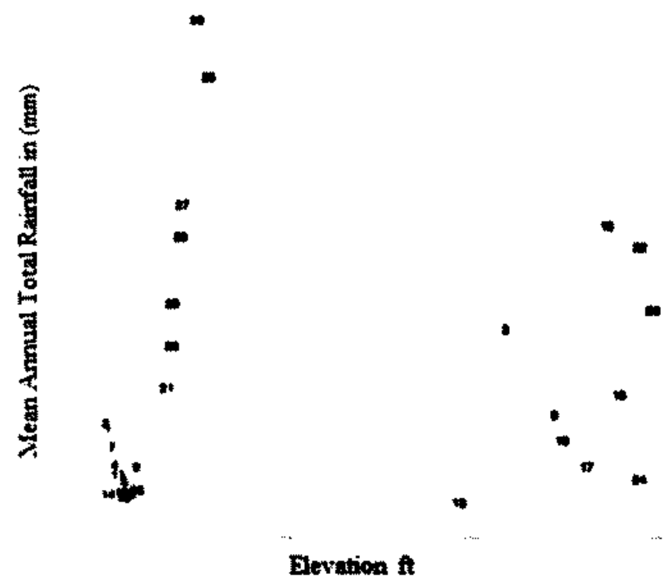

Fig. 4.4 Scatter Plot of Elevation and Mean Annual Total Rainfall

The group of 30 sites would be subdivided into four homogeneous sub regions by using these characteristics that are elevation and average annual rainfall. We name these groups region I to IV. Now for each group the discordancy, heterogeneity and goodness of fit measure were then calculated by using Equation (3.3.3), (3.4.1.3) and (3.5.2) respectively.

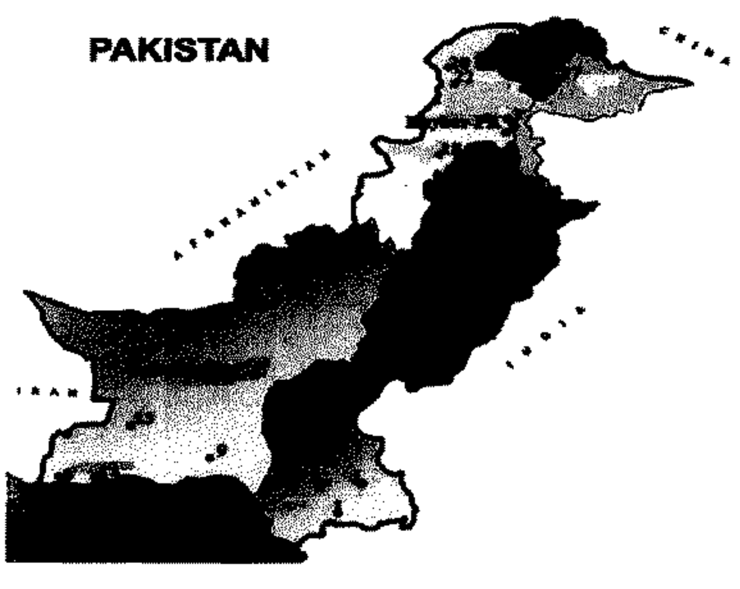

Fig. 4.5 Map of Pakistan with respect to Elevation of these sites

Table 4.3 (a) Heterogeneity Statistic for Homogeneous Region I

Sites	Site Name	D_i	Heterogeneity Statistic	Regional L-moments
1	Badin	0.26	H_1= 0.30	t^R= 0.4268
2	Choor	0.98	H_2= -0.12	t_3^R= 0.2403
4	Hyderabad	0.34	H_3= -0.58	t_4^R= 0.1429
5	Jacobabad	0.87		
6	Jiwani	0.37		
7	Karachi	1.71		
9	Lasbella	2.37		
10	Moen-Jo-Daro	1.51		
11	Nawabshah	0.80		
12	Padidan	1.70		
14	Pasni	0.89		
15	Rohri	1.02		
26	Khanpur	0.18		

The critical value for region I is 2.87.

Table 4.3 (b) Heterogeneity Statistic for Homogeneous Region II

Sites	Site Name	D_i	Heterogeneity Statistic	Regional L-moments
21	D.I.Khan	0.73	H_1= 0.81	t^R= 0.1822
23	Faisalabad	0.72	H_2= -0.30	t_3^R= 0.1291
25	Jehlum	1.76	H_3= -0.71	t_4^R= 0.1205
27	Lahore	0.39		
28	Mian Wali	0.14		
29	Sargodha	1.91		
30	Sialkot	1.35		

The critical value for region II is 1.92.

Table 4.3 (c) Heterogeneity Statistic for Homogeneous Region III

Sites	Site Name	D_i	Heterogeneity Statistic	Regional L-moments
16	Zhob	1.10	H_1= 0.67	t^R= 0.1674
18	Cherat	0.95	H_2= -1.06	t_3^R= 0.0676
20	Chitral	0.30	H_3= -1.87	t_4^R= 0.1616
22	Drosh	1.33		
24	Gilgit	1.32		

The critical value for region III is 1.33.

Table 4.3 (d) Heterogeneity Statistic for Homogeneous Region IV

Sites	Site Name	D_i	Heterogeneity Statistic	Regional L-moments
3	Barkhan	0.44	H_1= 0.98	t^R= 0.2481
8	Khuzdar	1.31	H_2= -0.32	t_3^R= 0.1734
13	Panjgar	1.32	H_3= -0.79	t_4^R= 0.1932
17	Bunji	0.60		
19	Chilas	1.33		

The critical value for region IV is 1.33.

Above tables show that regions I has 13 sites, region II has 7 sites and region III and IV have 5 sites respectively. The discordancy measure shows that none of the site from four regions exceed the critical value given in Table 3.1, Hosking and Wallis (1997). In region III and IV discordance statistic for site 22 Drosh and site 19 Chilas are equal to the critical value, but these sites do not exceed the critical value which means these sites are non discordant. The heterogeneity measures for all of the four regions along with regional weighted L-moments are shown in these tables. The heterogeneity statistics indicate that all these regions are homogeneous and none of any $H_j > 1$.

4.4 Selection of Best-fitted Regional Distribution

The goodness-of-fit statistic Z^{DIST} calculated for all regions using Equation (3.5.2). Three parameter distributions, that are generalizes logistic (GLO), generalized Pearson type III (PE3), generalized extreme value (GEV), generalized normal (GNO) and generalized Pareto (GPA) has 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 estimation if the shape of the tail of the true frequency distribution is not well approximated by the fitted distribution which was defined by Hosking and Wallis (1997). Regions for which distributions gave acceptable fit are given in the table 4.4 that is $|Z^{DIST}| \leq 1.64$. According to Hosking and Wallis (1997) the criteria $|Z^{DIST}| \leq 1.64$ is not an exact criterion of goodness of fit if serial correlation or cross correlation is present in the data.

Table 4.4 Z^{DIST} Statistic for Various Distributions

Region	Z^{GLO}	Z^{GEV}	Z^{GNO}	Z^{PE3}	Z^{GPA}
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

In Table 4.4 $|Z^{DIST}|$ - 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. Declare the fit to be adequate if $|Z^{DIST}|$ is sufficiently close to zero Hosking and Wallis (1997). GNO and PE3 are best fitted distributions for region I and the most suitable is PE3 because its Z^{DIST} is closer to zero. For region II GEV, GNO and PE3 are best fitted distributions but the most suitable is PE3 similarly 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 region III and IV is best.

4.4.1 L-Moments Ratio Diagram

L-moment ratio diagram is a useful guideline for the selection of an appropriate distribution for describing a set of variables as a distinct relationship between L-moment ratios exists for each theoretical probability distribution defined by Hosking and Wallis (1997). L-moment ratio diagram for four regions were illustrated in Fig. 4.6(a-d). For region I regional average L-skewness and L-kurtosis (AVG) lie closest to the PE3 distribution similarly for region II, III, IV regional average L-skewness and L-kurtosis (AVG) lie closest to PE3, GLO and GLO respectively.

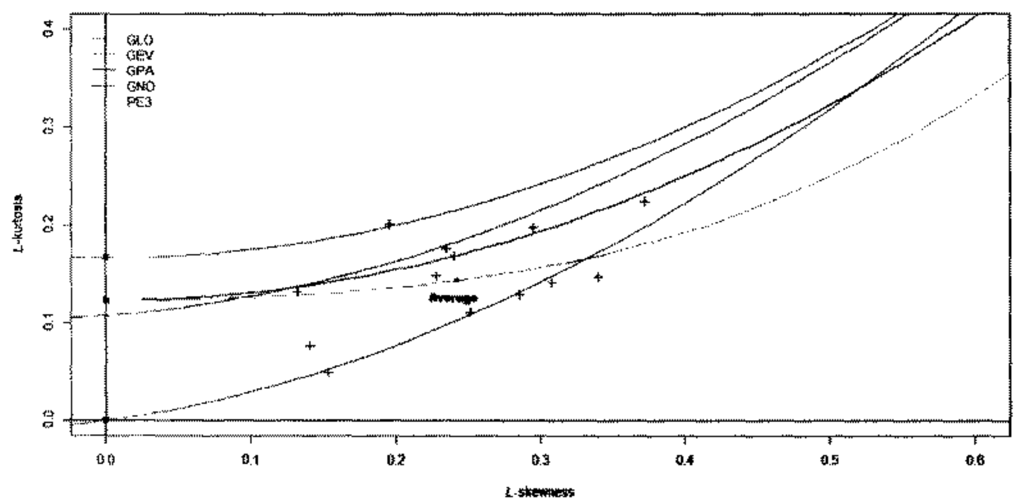

Fig. 4.6 (a) represents L-moment ratio diagram for region I

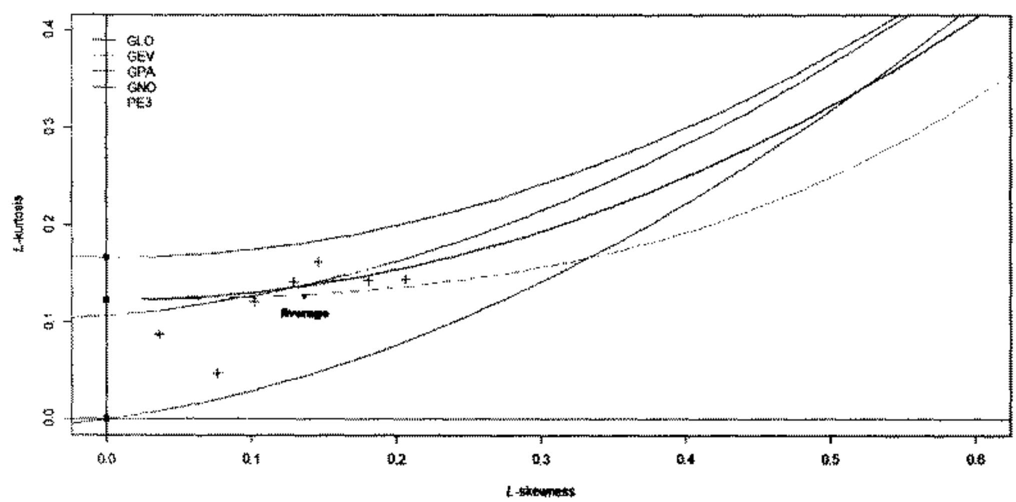

Fig. 4.6 (b) represents L-moment ratio diagram for region II

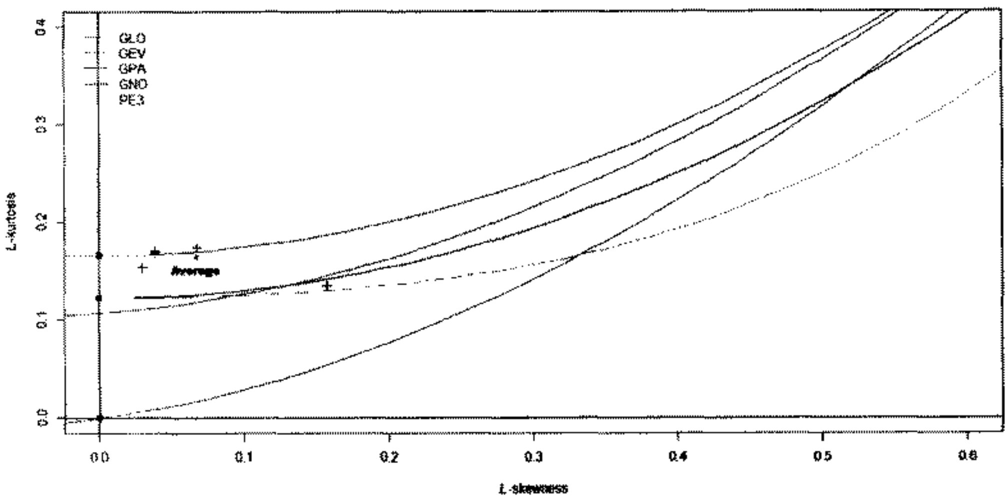

Fig. 4.6 (c) represents L-moment ratio diagram for region III

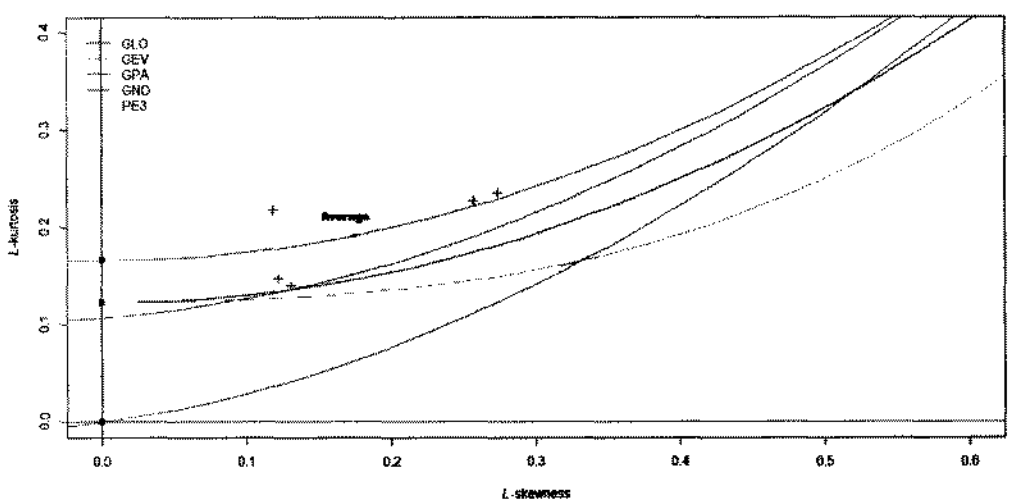

Fig. 4.6 (d) represents L-moment ratio diagram for regions IV

4.5 Fitted Distributions and Regional Growth Curves Estimation

After the selection of best fitted regional distributions we estimate regional quantiles for these four regions which are given in Table 4.5 (a-d). The values given in Table 4.5(a) 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 with exceedance probability of 0.02 and is 3.259 time larger than its average (135.38mm) for all sites in region I for the given return period. Estimated quantiles for all these regions can be interpreted in the same way. Return period which is also known as recurrence interval is an estimation of the likelihood of an event, such as rainfall, flood, storms or a river discharge flow to occur. Return period (T) is defined as $\frac{1}{P}$ where P is the probability of exceedance. Sometime we are interested to know that what the chance of extreme events over a given time period that will occur is known as the probability of occurrence or the exceedance probability. $P = \frac{1}{T}$ is the exceedance probability for example for 5 years $\frac{1}{.5} = 0.2$ is known as exceedance probability. For non exceedance probability we use relation that is $F = 1 - \frac{1}{T}$ for example for 5 years $1 - \frac{1}{5} = 0.80$ which is known as non exceedance probability. In Fig. 4.7 (a-d) we construct regional growth curves for all these four regions. In regional frequency analysis, it was assumed that the sites have a common frequency distribution. One representation of this common distribution is the regional 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 the sites.

The growth curve for region I in Fig. 4.7 (a) 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 region I GNO and PE3 shows same result for low return period 1, 2, 5, 10, 20 and 50 years. In Fig. 4.7 (b) regional growth curve for region II shows approximately same values of quantiles up to return period of 1000. Fig. 4.7 (c) 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 upward which means that the quantiles of GLO are high. Similarly for region IV in Fig. 4.7 (d) 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 in upward direction which indicates quantiles of GLO are high.

Table 4.5 (a) Regional Quantile Estimates for region I

	Parameters			Regional quantile estimates with nonexceedance probability F								
Dist	ε	α	k	*0.10	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	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 4.5 (b) Regional Quantile Estimates for region II

	Parameters			Regional quantile estimates with nonexceedance probability F								
Dist	ε	α	k	*0.10	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	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 4.5 (c) Regional Quantile Estimates for region III

	Parameters			Regional quantile estimates with nonexceedance probability F								
Dist	ε	α	k	*0.10	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	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 4.5 (d) Regional Quantile Estimates for region IV

	Parameters			Regional quantile estimates with nonexceedance probability F								
Dist	ε	α	k	*0.10	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	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

* nonexceedance probability F; ** return periods

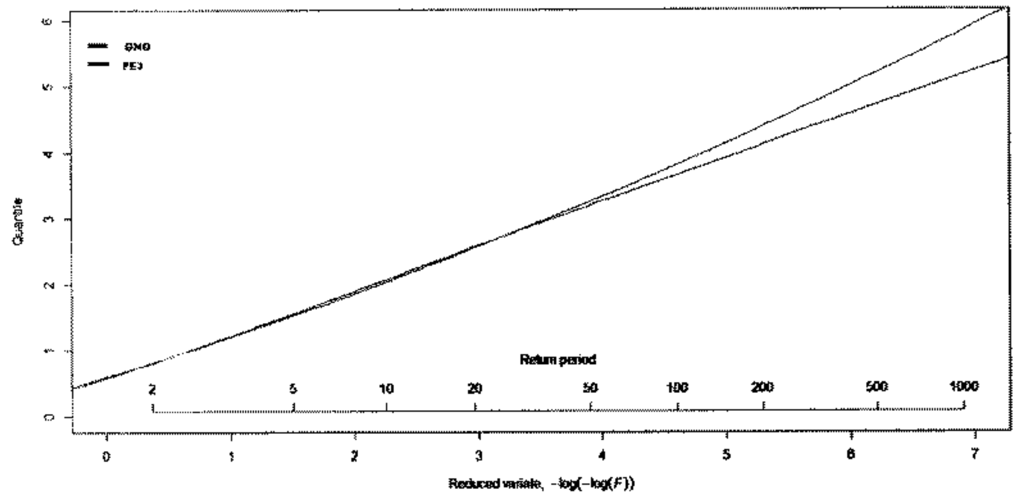

Fig. 4.7 (a) represents growth curves for regions I

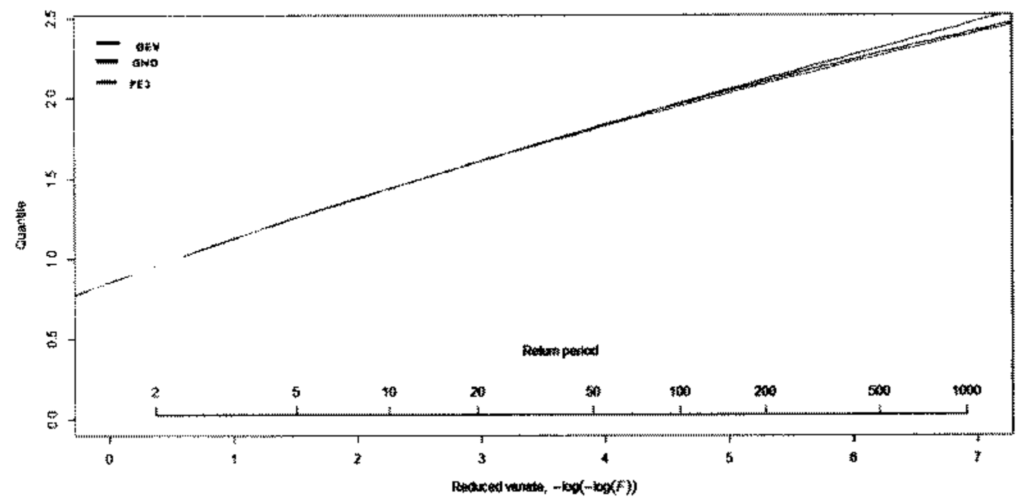

Fig. 4.7 (b) represents growth curves for regions II

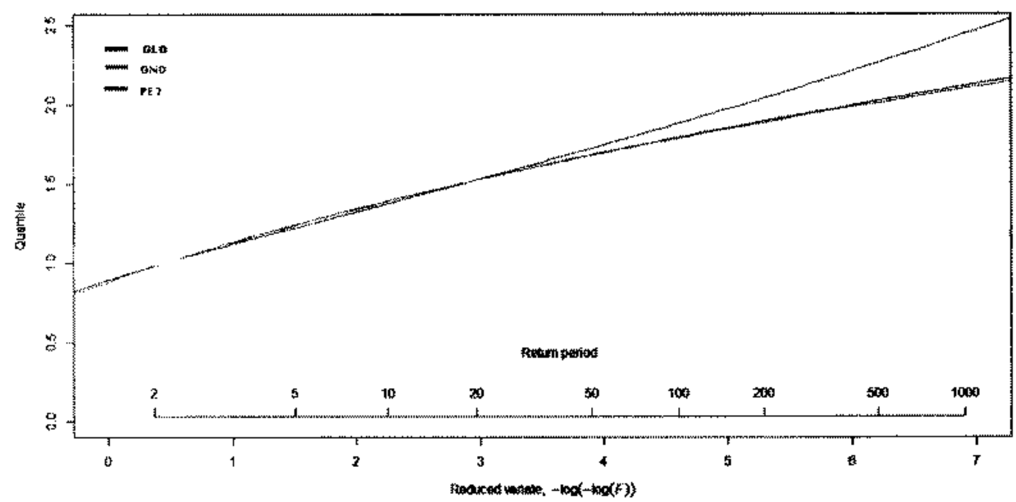

Fig. 4.7 (c) represents growth curves for regions III

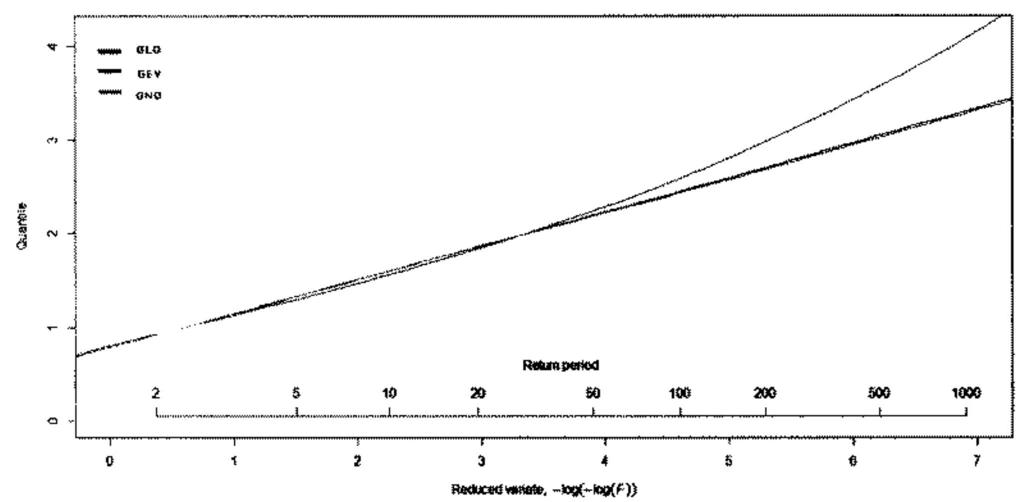

Fig. 4.7 (d) represents growth curves for regions IV

4.5.1 Decision about Regional Growth Curve and Estimated Quantiles

In this step we use Monte Carlo Simulation by the help of which quantile estimation accuracy is to be checked. In selected simulation procedure four different regions are defined to measure the accuracy level of estimated regional growth curves for these regions. For 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 sites 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 (1997) in this step. There are 13 sites in the region I for which simulation procedure is applied. All these sites having same record lengths as those for sites occupied by region 1 having L-C_v values varying from 0.388 to 0.4648 for GNO distribution. Root mean square error, relative bias, absolute bias, lowers and upper bounds are calculated for regional growth curve for several nonexceedance probabilities which are given in the Table 4.6 (a-d). 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.

In Table 4.6(a) simulation results for 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 and 10 years the values are high as compared to large return period. As we know that in practice relative bias and absolute bias are not affective for quantities accuracy. Simulation results for GNO shows that at low return period 2, 5, 10 and 20 RMSE are low as compared to large return periods. The error bounds lower error bounds and upper error bounds (LEB, UEB) for GNO distributions are smaller at low return period where as its error bounds for large return periods are high.

For 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 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 where as 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 region I at large return period.

In Table 4.6(b) simulation results for region II show that for GEV at low return period 1, 2, 5, 10 and 20 its RMSE and error bounds are low as compared to 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 the most suitable choice for low return period.

In Table 4.6(c) simulation results for region III show that for GNO at low return period 1, 2, 5, 10 and 20 its RMSE and error bounds are low as compared to 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 4.6(d) 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 compare 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 4.6 (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	$R^R(F)$	0.3931	0.0355	0.0252	0.0393	0.0530	0.0696	0.0812	0.1061	0.1162
	$B^R(F)$	0.0497	0.0065	0.0031	0.0004	-0.0020	-0.0049	-0.0069	-0.0108	-0.0122
	$A^R(F)$	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	$R^R(F)$	0.3874	0.0366	0.0242	0.0401	0.0523	0.0643	0.0714	0.0838	0.0880
	$B^R(F)$	0.0630	0.0036	0.0017	0.0007	0.0000	-0.0008	-0.0013	-0.0023	-0.0027
	$A^R(F)$	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 4.6 (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	$R^R(F)$	0.0490	0.0107	0.0162	0.0248	0.0337	0.0462	0.0562	0.0806	0.0915
	$B^R(F)$	0.0007	0.0008	0.0003	-0.0001	-0.0005	-0.0007	-0.0007	0.0000	0.0006
	$A^R(F)$	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	$R^R(F)$	0.0492	0.0112	0.0162	0.0254	0.0344	0.0457	0.0539	0.0721	0.0796
	$B^R(F)$	0.0009	0.0008	0.0002	-0.0001	-0.0003	-0.0005	-0.0005	-0.0003	-0.0001
	$A^R(F)$	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	$R^R(F)$	0.0491	0.0112	0.0162	0.0256	0.0341	0.0436	0.0500	0.0625	0.0672
	$B^R(F)$	0.0015	0.0008	0.0001	-0.0002	-0.0005	-0.0009	-0.0011	-0.0016	-0.0018
	$A^R(F)$	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 4.6 (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	$R^R(F)$	0.0562	0.0115	0.0186	0.0276	0.0369	0.0502	0.0612	0.0899	0.1037
	$B^R(F)$	0.0010	0.0009	0.0003	0.0000	0.0000	0.0000	0.0002	0.0016	0.0026
	$A^R(F)$	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	$R^R(F)$	0.0542	0.0105	0.0182	0.0271	0.0353	0.0454	0.0528	0.0689	0.0756
	$B^R(F)$	0.0015	0.0003	0.0001	0.0002	0.0005	0.0010	0.0015	0.0028	0.0034
	$A^R(F)$	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	$R^R(F)$	0.0544	0.0105	0.0182	0.0274	0.0355	0.0447	0.0511	0.0642	0.0693
	$B^R(F)$	0.0015	0.0003	0.0003	0.0003	0.0004	0.0005	0.0006	0.0009	0.0009
	$A^R(F)$	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 4.6 (d) 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	$R^R(F)$	0.1103	0.0246	0.0255	0.0405	0.0572	0.0812	0.1008	0.1515	0.1755
	$B^R(F)$	0.0054	0.0045	0.0014	-0.0008	-0.0031	-0.0058	-0.0076	-0.0103	-0.0108
	$A^R(F)$	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	$R^R(F)$	0.1068	0.0211	0.0253	0.0399	0.0542	0.0744	0.0907	0.1317	0.1507
	$B^R(F)$	0.0050	0.0024	0.0011	0.0000	-0.0009	-0.0017	-0.0020	-0.0009	0.0003
	$A^R(F)$	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	$R^R(F)$	0.1054	0.0221	0.0254	0.0403	0.0546	0.0724	0.0853	0.1136	0.1253
	$B^R(F)$	0.0069	0.0022	0.0004	-0.0003	-0.0008	-0.0012	-0.0012	-0.0008	-0.0004
	$A^R(F)$	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

CHAPTER 5

Summary and Conclusions

Regional frequency analysis was performed on 30 sites of Pakistan. The daily rainfall amount taken from Pakistan Meteorological center Karachi is measured by rain gauges in millimeters, from which annual total rainfall (ATR) series have been constructed for the proposed study. The record length of ATR series varies from 29 to 51 years.

Initially the basic assumptions of regional frequency analysis are tested by different tests that are time series plots, Mann-Whitney test, Kendall's tau test and Ljung-Box-Q-Statistics. All sites satisfied these tests which are performed to observations at any site are independent, stationary and identically distributed.

The regional frequency analysis is based on four steps. The first step is initial screening of data in which none of 30 sites is discordant. It means that there is no gross error in the data.

Second step is the formation of homogeneous regions which is the most difficult step in regional frequency. Initially all 30 sites are treated as a single region but it does not satisfy the heterogeneity statistic defined by Hosking and Wallis's. The study area has the characteristic that a highly elevated site receives high average annual rainfall and sites with low elevation receive low average annual rainfall. On the basis of this characteristic four regions were designed such that all of these regions were acceptably homogeneous.

The next step is the selection of best fitted distribution for all four regions which was selected on the basis of L-moment ratio diagram and Z-statistic. On the basis of L-moment ratio diagram and Z-statistic it was found the PE3 is most suitable choice of distribution for first two regions and

GLO suitable choice for region III and IV respectively. Its means these suitable distributions that are PE3 and GLO are best for quantiles estimation.

The final step is the estimation of homogeneous region and regional growth curve. The growth curves for selected distributions are shown in Fig. 4.6 (a-d). At the lower tail for region I GNO and PE3 are same but there is a small difference between these two distributions at large return period. For region II GEV, GNO and PE3 are same up to the return period of 200 years but for large return period the quantiles of GLO are high. At the lower tail for region III GLO, GNO and PE3 are same but for large return period quantiles of GLO are high. Similarly for region IV quantiles of GLO, GEV and GNO are same at lower tail but at upper tail quantiles of GLO are high.

Ten thousand runs of Monte Carlo simulation is made to calculate some accuracy measure such as RMSE, relative bias, relative absolute bias, lower error bound and upper error bound for the estimated regional growth curve for these regional quantiles. On the basis of this simulation study it was found that PE3 is most suitable choice for large return period for first three regions and for small return period GNO and GEV are most robust. Similarly for region four GEV is best for period of 1, 2, 5, 10 and 20 and for period of 50, 100, 500 and 1000 GNO gives most robust estimates. The basic objective of frequency analysis of extreme events is to estimate quantiles for upper tail of the distribution. So we conclude that PE3 and GNO are most robust for large return period.

Extreme events such as rainfalls, storms, high wind speeds, extreme temperatures, floods etc in many areas can have harmful consequences. In these conditions we use extreme value theory to develop appropriate probability modeling to mitigate risks caused by these extreme events.

For region I, II and III PE3 distribution is best fitted for large return period and theoretically PE3 distribution is also a heavy tailed distribution on right side. For low return period GNO is best fitted distribution for region I and III and GEV for region II.

Similarly for region IV GNO distribution is best fitted for large return period and theoretically GNO is a heavy tailed distribution on both sides and for low return period GEV distribution is best fitted distribution.

Recommendations for future study

1. The study can be conducted using other estimation methods like LQ-moments, LH-moments and trimmed L-moments, and then estimated quantiles can be compared.

2. The at-site quantiles can be estimated using trimmed L-moments and L-moments, then estimated quantiles can be compared.

3. The quantile can be estimated by four parametric Kappa and five parametric Wakeby distribution and compared with those obtained from three parametric distributions.

References

1. Adamowski,K.,Alila,Y.,Pilon,P.,(1996)."Regional rainfall distribution for Canada", Atmospheric Research; 42 (1&4):75-88.
2. Ahmad,I., Shah,S.F., Mahmood,I., Ahmad,Z., (2013). "Modeling of monsoon rainfall in Pakistan based on Kappa distribution", Science International. (Lahore); 25 (2), 333-336.
3. Anli,A.S., (2013). "Frequency analysis of annual maximum rainfall series using L-moment methods on Yesilirmak Basin in Turkey", Journal of Food, Agriculture and Environment; 11 (3&4): 1763-1767.
4. Benabdesselam,T., Amarchi,H., (2013). "Regional approach for the estimation of extreme daily precipitation on North-east area of Algeria", International Journal of Water Resources and Environmental Engineering; 5 (10): 573-583.
5. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994). "Time Series Analysis: Forecasting and Control, Pearson Education", Delhi.
6. Burn, D. H., (1989). "Cluster analysis as applied to regional flood frequency", Journal of Water Resources Planning and Management; 115: 567-82.
7. Chaudhary, Q.Z., Rasul, G., (2004). "Agro-Climatic Classification of Pakistan", Science Vision, 9 (1-2) (Jul-Dec 2003) & N0. 3-4 (Jan-Jun 2004): 59-66.
8. Chen,Y.D., Huang,G., Shao,Q., XU,C.-Y., (2006). "Regional analysis of low flow using L-moments for Dongjiang basin, South China", Hydrological Sciences Journal; 51(6): 1051-1064
9. De Coursey, D. G., (1973). "Objective regionalization of peak flow rates. In Floods and Droughts", Proceedings of the Second International Symposium in Hydrology, Fort

Collins, Colorado, edited by E. L. Koelzer, V. A. Koelzer, and K. Mahmood, Water Resources Publications, Fort Collins, Colo; 395-405

10. Devi, T.A., Choudhurt. P., (2013). "Extreme Rainfall Frequency Analysis for Meteorological Sub-Division 4 of India Using L-Moments", International Journal of Environmental, Earth Science and Engineering; 7(12): 138-143.
11. Dubey,A., (2014). "Regional Flood Frequency Analysis Utilizing L-Moments: A Case Study of Narmada Basin", International Journal of Engineering Research and Applications; 4(2): 155-161.
12. Eslamian,S.S., Feizi,H., (2007). "Maximum monthly rainfall analysis using L-moments for an arid region in Isfahan province, Iran", Journal of Applied Meteorology and Climatology; 46:494-503.
13. Hassan,B.G.H., Ping,F., (2012). "Formation of Homogeneous Regional for the Luanhe Basin-by Using L-moments and Cluster Techniques", International Journal of Environmental Science and Development; 3(2): 205-210.
14. Hassan,B.G.H., Atiem,T.A., Jianzhu,L., Ping,F.,(2012). "At site and regional frequency analysis for Sudan annual rainfall by using the L-Moments and nonlinear regression technique", International Journal of Engineering Research and Development; 3 (6): 13-19.
15. 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.
16. Hosking, J.R.M., Wallis, J.R., (1993). "Some Statistics Useful in Regional Frequency Analysis", Water Resource Research; 29: 271-281.

17. Hosking, J.R.M., (1990). "Analysis and estimation of distributions using liner combination of order statistics", Journal of the Royal Statistical Society, Series B; 52,105-124.

18. Hosking, J.R.M., Wallis, J.R., (1997). "Regional Frequency Analysis: An approach Based on L-moments", Cambridge University Press, New York.

19. Hussain, Z., Pasha, G.R., (2009). "Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-moments", Water Resource Manage; 23: 1917-1933.

20. Koh, D.K., Choo, T.H., Maeng, S.J., Trived, C., (2008). "Regional Frequency Analysis for Rainfall using L-Moment", The Journal of the Korea Contents Association; 8 (3): 252-263.

21. Kysely, J., Picek, J., (2007). "Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic", Climate Research; 33: 243-255.

22. Lee, S.H., Maeng, S.-J., (2003). "Frequency analysis of extreme rainfall using L-moment", Irrigation and Drainage; 52: 219–230.

23. Ljung, G. M., Box, G. E. P., (1978). "On a measure of a lack of fit in time series models", Biometrika; 65, 297-303.

24. 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.

25. Matalas, N. C., Slack, J. R., Wallis, J. R., (1975). "Regional skew in search of a parent", Water Resources Research; 11: 815-26.

26. Ngongondo,C.S., Xu,C.Y., Tallaksen,L.M., Alemaw,B., Chirwa,T., (2011). "Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and

L-moments approaches", Stochastic Environmental Research and Risk Assessment; 25 (7): 939–955

27. Norbiato, D., Barga, 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.
28. Parida, B.P.,(1999). "Modeling of Indian summer monsoon rainfall using a four-parameter Kappa distribution", International Journal of Climatology; 19: 1389-1398.
29. Park, J.S., Jung, H.S., Kim, R.S., Oh, J.H., (2001). "Modeling summer extreme rainfall over the Korean peninsula using Wakeby distribution", International journal of climatology; 21 (11): 1371-1384.
30. Pearson, C. P., (1991). "New Zealand regional flood frequency analysis using L-moments", Journal of Hydrology; 30: 53-64.
31. Rahman, M.M., Sarkar, S., Najafi,M.R., Rai, R.K., (2013). "Regional extreme rainfall mapping for Bangladesh using L-moment technique", Journal of Hydrologic Engineering, 18 (5): 603-615.
32. Rostami, R., (2013). "Regional flood frequency analysis based on L-moment approach (case study: West Azarbaijan Basins)", Journal of Civil Engineering and Urbanism, 3(3):107-113.
33. Sabri, A., Ariff, N.M., Ani Shabri., (2009). "Frequency analysis of maximum daily rainfalls via L-moment approach", Sains Malaysiana; 38 (2):149-158.
34. Saf, B., (2009). "Regional flood frequency analysis using L-moments for the Buyuk and Kucuk Menderes of Turkey", Journal of Hydrologic Engineering; 14(8), 738-794.

35. Schaefer, M. G., (1990). "Regional analyses of precipitation annual maxima in Washington State", Water Resources Research; 26: 119-31.

36. Seckin,N., Haktanir,T., Yurtal,R., (2011). "Flood frequency analysis of Turkey using L-moments method", Hydrological Processes; 25(22), 3499-3505.

37. Shahzadi, A., Akhtar, A.S., Saf, B., (2013). "Regional Frequency Analysis of Annual Maximum Rainfall in Monsoon Region of Pakistan using *L*-moments", Pakistan Journal of Statistics and Operation Research; IX (1): 111-136.

38. 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 (1&2): 42–52.

39. Trefry, C.M., Jr, D.W.W., Johnson, D., (2005). "Regional rainfall frequency analysis for the state of Michigan", Journal of Hydrological Engineering; 10 (6), 437-449.

40. Wiltshire, S. E., (1985). "Grouping basins for regional flood frequency analysis", Hydrological Sciences Journal; 30: 151-9.

41. Yang,T., Wang, X., Hao, X., Li,H., (2010). "Regional frequency analysis of rainfall extremes in the Pearl River Basin using L-moments method", Fuzzy system and knowledge discovery (FSKD), seventh international conference; 6: 2647 – 2651.

42. Yurekli, K., (2005). "Regional Analysis of Monthly Rainfalls over Amasya Province via L-Moments Method", Ziraaat Fakultesi Dergisi; 22 (2): 51-56.

43. Yurekli, K., Modarres, R., Ozturk,F., (2009). "Regional daily maximum rainfall estimation for Cekerek Watershed by L-moments", Meteorological Application; 16:435-444.