REGIONAL FLOOD FREQUENCY ANALYSIS ON INDUS BASIN BARRAGES BY USING PARTIAL L-MOMENTS METHOD

By

Muhammad Waqas

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University,
Islamabad, Pakistan 2016

REGIONAL FLOOD FREQUENCY ANALYSIS ON INDUS BASIN BARRAGES BY USING PARTIAL L-MOMENTS METHOD

By

Muhammad Waqas

Supervised by

Dr. Ishfaq Ahmad

Department of Mathematics & Statistics

Faculty of Basic and Applied Sciences

International Islamic University,

Islamabad Pakistan 2016

Certificate

Regional Flood Frequency Analysis on Indus Basin Barrages by Using Partial L-Moments Method

By

Muhammad Waqas

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

We accept this dissertation as conforming to the required standard.

1 **Prof. Dr. Javed Shabir**
(External Examiner)

2 **Dr. Muhammad Akbar**
(Internal Examiner)

3 **Dr. Ishfaq Ahmad**
(Supervisor)

4 **Prof. Dr. M. Arshad Zia**
(Chairman)

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

2016

بسم الله الرحمن الرحيم

REGIONAL FLOOD FREQUENCY ANALYSIS ON INDUS BASIN BARRAGES BY USING PARTIAL L-MOMENTS METHOD

By

Muhammad Waqas

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 2016

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 FLOOD FREQUENCY ANALYSIS ON INDUS BASIN BARRAGES BY USING PARTIAL L-MOMENTS METHOD"** submitted by **Muhammad Waqas** (Registration # 40-FBAS/MSST/F13) 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 Mathematics & 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 I would like to extend my appreciation to **Dr. Akbar Awan** and **Dr. Zahid Shb** There sound advices and lots of good ideas were very helpful to me I would have been lost without them 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. Arshad Zia** for providing such necessary facilities

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

Muhammad Waqas

DECLARATION

I hereby declare that this thesis, neither as a whole nor a part of it, has been copied out from any source It is further declared that I have prepared this dissertation entirely on the basis of my personal efforts made under the supervision of my supervisor **Dr. Ishfaq Ahmad** No portion of the work, presented in this dissertation, has been submitted in the support of any application for any degree or qualification of this or any other learning institute

Signature: ____________________

Muhammad Waqas

MS (Statistics)

Reg. No 40-FBAS/MSST/F13

Department of Mathematics and Statistics,

Faculty of Basic and Applied Sciences,

International Islamic University Islamabad, Pakistan.

CONTENTS

List of Figures

List of Tables

Abstract

Regional frequency analysis based on Partial Linear Moments and Linear Moments has been utilized Both techniques are applied in this study and comparison made between these approaches The 10 barrages of Indus Basin Pakistan Barrages lying on the four Rivers the Ravi, Chenab, Jhelum and the Indus are selected for the study The data of annual maximum flows was retained for analysis The data ranging through the 30 years flow record to 93 years By applying several tests, the initial conditions of randomness, autocorrelation, homogeneity and equality of variance are satisfied for all barrages data Discordancy measure is calculated by using both Linear Moments (LM) and Partial Linear Moments (PLM), under these approaches all sites fulfilled the discordancy threshold Established on the physiographical positions of the barrages, a single homogenous region was formed and corroborated through the heterogeneity standards After ratification, of discordancy and Heterogeneity conditions, the Z-statistics and LMRD implied to select the best fitted distribution for described region GPA outperformed by PL-moments while GNO and GPA through L-moments GPA of PL and GNO, GPA of LM were selected for further analysis and to check robustness Quantile estimates and regional growth curves were constructed by using simulation studies Monte Carlo Simulation used for enactment assessment and repetitions made up to 10,000 The simulations results showed the absolute bias relative bias, root mean square error and error bounds which endorsed that GPA by PL method is robust distribution and overtake the GNO and GPA of L-moments for estimation of larger retune period events of 100, 500 and 1000 years That reinforce the performance of PLM over the LM for larger return periods

CHAPTER 1

1.1 Introduction

Climate over the globe is changing drastically which is the flaunting issue of today's world Assessment of atmospheric circumstances, established on long period of time, pertained as climate Global warming provoking to change in climatic conditions Climate changes evoking to extreme precipitations, hailing, thunderstorms alluvion, deluge inundation and floods These changing constraints confronted by every country with loss of immense amount of money resources and scathe of soft and hard infrastructure as well as the fearful number of loss of man power and human resources Advancing countries stricken by deleterious floods many times especially Pakistan

Precisely verbalizing causes of floods can be immoderate rains, river outpouring or fortified winds in maritime arena, dam breaking and snow melting As well major types of flood can be flash flood, rapid flood and slow flood Mostly, Pakistan faced rapid or slow flood Flash flood can occur for 2 to 6 hours aroused by extreme rain snow melting or dam break Rapid flood takes more time to rise than flash flood It can exist a day or two days and more destructive Slow flood can stay for few days, maybe weeks It usually appears in low lying plain areas spread over long arena and cause to damage all survivals in its way spread diseases malnutrition and insect's bites

Pakistan has diverse kind of land and the formation of climate is instable Naturally Pakistan has hot and dry climate but in preceding few decades it has revealed important palpable deviations The frequent reasons of flooding in Pakistan are the melting glacier and uttermost rainfalls Floods are mostly triggered by the intensive precipitation in the catchment areas of main and other rivers through monsoon spell which are often enlarged by melting of snow Floods are instigated mainly due to overflowing in major rivers and flash floods in tributaries including hill torrents cause to coastal and urban flooding (Annual Flood Report 2015)

After the birth (1947), Pakistan preoccupied by multiple number of perilous floods Pakistan had confronted floods of several degrees since 1950 to 2015 resultant into complaint of vast areas of the four provinces (Punjab, Sindh and Baluchistan & NWFP) including Gilgit-Baltistan FATA and Azad Jammu & Kashmir The country has agonized a cumulative economic loss of more than US$ 38.00 billion during the past 69 years (1947 to 2015) Around 11,939 deaths ruination of 192,596 villages and about 613,721 square km area was damaged due to 22 major floods The flood in 2010 was utmost destructive in which 1985 died, affected about 20 million population and drowned area of 160 000 square km The 2014 flood impaired 2 6 million people harvested area of 2 415 million acres, 4,065 villages (destructing 107 102 houses) and charged 367 deaths

Major five rivers, the Indus, Jhelum, and Chenab Ravi and Sutlej and their branches course through the realm's plains The Indus, Jhelum and Chenab are recognized as the Western tributaries and Ravi Beas and Sutlej named as the Eastern tributaries Aforementioned rivers spread water to the entire Indus Basin Irrigation System (IRSA) All these tributaries have their foundation in the higher altitudes and produce their flows primarily from snow melting and monsoon rains The Indus Basin catchment area is matchless that it comprehends seven highest-ranking mountaintops of the world beside Mount Everest These contain K-2 (28,253 feet) Nanga Parbat (26 660 feet) Rakaposhi (25,552 feet) etc The seven glaciers located in the Indus catchment namely Siachin, Hispar Biafo Batura, Baltoro Barpu and Hopper are among the leading in the world (NDMA-UNDP report 2010)

Indus river the main contributor to food basket of Pakistan has immense importance in Indus basin and one of the top gigantic rivers in the world It s the world s 12th largest river in the regard of drainage area that is 9.7×10^5 square km, on 7th in deltaic area 3×10^4 square km on 10th in water runoff having $2 \times 10^{11}\ m^3/year$ and on 6th in annual sediment ejection that is $2 \times 10^{11}\ kg/year$ Indus spreads from east to west in north of Pakistan The length of Indus river is 3180 km that announced it one of the longest river in the world and emptied into Arabian sea near Karachi Its originate from the mutual areas of Himalayas Hindu Kush and Karakoram These mountains shielded by snow almost all the year and contribute to the Indus in summer These areas not only comprise on the top most heights as well as the longest glaciers in the world (Pakistan Water Gateway 2008)

The Chenab adds considerable water into the Indus Basin Its total length is 770 miles (1 232 Km) Originating Himachal Pardesh India In the higher spreads Chandra and Bhaga are the two major brooks of Chenab, upsurge at an elevation of about 16000 feet AMSL These branches seam out at Tandi sited in Jammu and Kashmir The river after crossing mountain regions about 400 miles releases into the plain area near Akhnur, then Chenab invade in Sialkot district Chenab runs through sedimentary steppes of the Punjab chasing a expanse of about 5468km It unites with Jhelum river at Trimmu and lastly the Jhelum and Chenab after gathering Ravi and Sutlej rivers emptied into Indus at Mithankot Its catchment is about 26 079 square miles

The Jhelum is also main contributor having length 510 mile (816 km) which instigated from Pir Panjal and flows matching to the Indus at an elevation of 5500 feet. The river runs over Dal and Wular lakes. Near to Muzafarbad at Domel it junctures with the Nelum that is encompassed of about 2800 sq. mi. of mountainous ranges. Another tributary Kunhar links with Jhelum about 5 miles beneath Domel. Two other small rivers Kanshi & Punch connect with Jhelum between Domel and Mangla, and Punch enters into it about 07 miles above Mangla at Tangrot. Below Mangla, numerous flood water streams link with the Jhelum.

Ravi is also major supplier to Indus Basin. Its length is 550 miles (880 Km). The catchment of Ravi is 3100 sq. mi. The Ravi initiates from the Bangahal's basin and drains the southern slopes of the Dhanladhar. Through Lower Bangahal these streams run through the gorge of Chamba. Ravi omits the Himalayas at Basech. Its regular slop is 45 feet per mile. The Ravi enters to Pathankot at Chaundh and share a borderline between India and Jammu and Kashmir for 23 miles.

The capacity of flowing water in these tributaries is depend upon seasonal variations. It is decreases in winter and rises regularly with approaching to summer when snowpack in hilly catchment's begin to dissolve. In the starting of summer amount of water in the rivers fluctuates depend upon the height and condition with respect to the monsoons and the level of the snow melting. The level of water upsurges significantly in Indus, Jhelum and Chenab after March but in the eastern rivers this appearance is late. The Indus which rises early collects its water from two hefty glaciers the Hindu Kush and the Karakoram being larger than any other mountains outer the Polar region. The start of rainy spell at the completion of June or early July is noticed by a huge increment in course. The era of extraordinary flow ended in the Indus and Jhelum in September, but endures for next months in the eastern rivers. The reduction in flow occur sharply then the rise even months with high mean discharges are categorized by wide fluctuations in the daily releases. Floods often occurs in the early part of the rainy season in the western rivers and later in the eastern. Because 60% of the tide in the Indus system is engrossed in the 3 monsoon months. That's why dire requisite for artificial lake and dams to adjust the flow to lessen the floods and to reduce flow to the sea and deliver sufficient support to irrigation.

Implementing the project named as Indus Basin Project (IBP) launched after Indus water treaty (1960). Pakistan developed two dams Mangla, Tarbela, five new barrages and eight main link canals to exchange water between rivers. Four Barrages amended to meet the requirement of irrigation and flood control. Barrages Flood limits are designed for each station low, medium, high, very high and exceptionally high. Mangla and Rasul low limit at 7.5 ml cft sec and high 30 ml cft sec. Marala, Qadirabad and Khanki flood low limit at 10 ml cft sec while highest at 60 ml cft/sec Balloki low at 4 ml cft sec while high at 18 ml cft sec Sidhnai 3 ml cft sec and high at 12 ml cft/sec (The Indus Water Treaty 1960)

Table 1 1 Geographical Characteristics of Barrages, WAPDA Pakistan

Description of Barrage Characteristic						
River	**Barrage**	**Latitude**	**Longitude**	**Elevation**	**Length**	**Drainage Area**
Jhelum	Mangla	33 15	73 65	774	10,300	20712
	Rasul	32 68	73 5	699	3,209	24069
Chenab	Marala	32 68	74 43	702	4,472	31148
	Khanki	32 4	73 92	666	4,000	29732
	Qadirabad	32 33	73 73	639	3,373	25485
Ravi	Balloki	31 22	73 86	600	1,644	63712
	Sidhnai	30 58	72 07	426	712	4709
Sindh	Kotri	25 22	68 22	55	3,000	25485
	Guddu	28 3	69 5	222	3,840	25485
	Sukkur	27 72	68 79	201	4,490	42475

1.2 Regional Frequency Analysis

Regional Flood Frequency Analysis(RFFA) took remarkable importance in last few decades Regional frequency analysis definable as the use of same nature of data at various sites and forming the homogenous region by focusing on the site's hydrological and geographical characteristics in order to estimate the suitable frequency distribution Flood Frequency Analysis (FFA) are practiced for forecast design floods for stations along the river, that help out in better planning to deal with flood in advance and to decrease the loss of precious lives, properties, time, money and infrastructure The sole objective of the RFFA by practicing index flood process is to detect a most appropriate distribution, further which will be helpful in quantile estimation In RFFA by designing homogenous regions we can perform forecasting that will resultant to the least variability in predictions and actual happenings

Linear Moments (LM) methodology extensively utilized in RFFA L-moments are the linear combination of Order statistics that is robust while outliers subsist in data, provides unbiased estimates subject to fitting on large or small sample data Outliers can efficaciously countered by using LM method LM established on Probability Weighted Moments (PWMs) provide more authentic results of fitted distribution and parameter estimates comparing to maximum likelihood technique provided by Hosking (1985b) Derivation, summarization and interpretation of results by LM is much easier, especially today due to the cushy access to latest tools and software LM must exist if mean of the distribution subsist where in Conventional Moments (CM) it's contrary to aforementioned methodology CM are highly effected by outliers, innately nonlinear declivity of accuracy in higher order and inconsistent in rapidly variating data, but LM can deal in all these situation (Hosking & Wallis 2005) Although the LM method is widely applied in regional flood frequency analysis and it is much effective in estimation of parameters but it is not much sufficient in predicting large return periods That's why, the use of lower bound censored samples is much suitable In estimating high return period floods the small one have least relevance to the larger one thus adding small floods might be cause to nuisance values Wang

(1990a), and censoring will make improvement in the estimation of large return period

Partial probability weighted moments (PPWM) are also stated as Partial Linear Moment (PLM) are parallel to Linear Moments method (LMM) PPWM are the extension of the Probability Weighted Moments (PWM) PLM accustomed to estimate a distribution from censored sample Censored samples extract high quantile estimates much efficiently as contrasting to uncensored samples acquaint by Wang (1990a) PLM expedient in estimating the upper quantiles of flood flows when someone concentrating on the right tail of the distribution and censoring of the few observations in the left tail is preferential

Floods appears on sundry phases like low, medium, high, very high and exceptional high Flood occurs umpteen times, but not always harmful that is marked as low level flood Occurrence of this type of flood is common but they don't exceed the threshold point Threshold point explicate that if flood surmounts the fixed level it can be catastrophic Data incorporating values above or below the applied analytical threshold point alluded as Censored data Censored data analysis implemented in the inspection of hydrology, environmental engineering, disaster management, biomedical and survival analysis Such data frequently dealt in quality and quantity superintended applications of air soil and water samples

Censored data classified as either type I censored, where the threshold is applied while the number of censored data varies, or type II censoring where the censored data series fixed and the pointed threshold varies Censored data also categorized as left, right and double(interval) censored Left censored in which data below a fix point is censored and remaining is used for analysis In RFFA we mostly perform left censoring, because in flood studies the level of water deviate sometime too low that undergoing the threshold point that it can be discarded Comprising on higher quantiles, lowest values seems to be nuisance that reinforce the use of left censored data Data censored above the fixed point, pertained as right censoring, mostly used in life testing events Data censored on an interval, the combination of left and right censoring is interval censoring its focus on some specific time line

Censoring also categorized as single stage and multiple stage censoring Censoring on data series left or right but on one level said as single stage censoring If data contain on one or more censoring levels, referred as multi stage censoring Multistage censoring used in environmental data, but as the censoring level increases, calculations become more difficult

1.3 Objective of the Study

- To screen out data
- To identify homogeneous regions
- To choose best fitted distribution
- To estimate parameter of fitted distribution
- To analyze the advantage of partial L moments over L moments
- To compare GEV, GLO and GPA distributions while applying partial L moments
- To utilize the results in designing of dams, culverts, bridges and different kinds of flood controlling devices for policy implication

CHAPTER 2

2.1 Literature Review

Wang (1990a) applied Partial Probability Weighted Moments (PPWM) on censored samples and estimated the GEV distribution Introduced the PPWM denotation of PWM PPWM applied on censored samples, explicated the unbiased estimator of PPWM, and derived estimator of parameters and quantiles of GEV by using PPWM Monte Carlo experiments were performed to canvas the properties of GEV distribution by PPWM method on censored samples and proved that lower bound censoring provide more accuracy in quantile estimation

Wang (1990b) plied unbiased facts about historical flood information and applications of PWM and PPWM to estimate the parameter of GEV distribution Monte Carlo Simulations executed to estimate parameter and quantiles findings showed that for (k = -0 2) the distribution plot has long thick tail and high quantile and large variance for small systematic record But the tail at $k = 0$ 0 becomes thin and historical record improve the high quantile estimation With the increment in k value the estimation of high quantiles is easier Simulation output for $k = 0$ 2 addressed that incorporating the historical information can cause to light adverse effect on high quantile estimation In this work Wang derived constraints for unbiased estimator of PWM and PPWM by using the recorded and historic flood information

Hosking and Wallis (1997) recommended four steps to carry out Rational Frequency Analysis of environmental extreme happenings, especially in hydrology Those steps designed as (1) Data Screening (2) formation of homogenous regions (3) selection of plummy distribution (4) parameter estimation of tailored distribution Each next stage can be covered after fulfilling the previous steps assumptions, plied by author

Wang (1997) by using higher probability weighted moments for flood frequency analysis, while exploring the trend that larger flows can show interpolation and extrapolation in large return periods In described situation lower bound censored samples can assist better and the application of higher PWM is proposed Higher moments gives more weightage to large flows, simulation studies also give support to HM to use for anticipating for large return periods

Koulouris *et al* (1998) worked on LM diagrams for censored data, data sets covered above or below a threshold point appointed as censored data A number of statistical techniques of censored data belongs to problem of moment, parameter, and quantile guesstimate methods Introduced LM figure for the assessment of suitable fit of left-censored data substitute to distributional hypothesis Simulated censored data sets also illustrated efficacy of LMRD

Bhattarai (2004) inquired censored flood samples by PLM method, made comparison of PL and LH moments techniques. by processing on real data of extreme flows of 25 rivers of Nepal as well as simulation studies executed Simulation studies analysis enkindle that censoring up to 30% level provided least bias value without increasing the RMSE On some occasion both demonstrated similar value of RMSE but for bias PLM revealed smaller value Equating LM, LHM and PLM for annual maximum flood flows LHM and PLM exhibited akin outcomes and surmounted on the LM But overall PLM resulted superiorly, although censoring level applied 10%, 205, 30% up to 60% with 10,000 replications

Hassan (2004) estimated PPWM for GPA distribution, analyzed their properties by Monte Carlo Experiments on censored samples Simulation for several sample sizes n=15,20,30 and 50 done Censoring engendered on various levels, for each group of values of n, k samples generated randomly about 1000 from GPA distribution α and k value derived under the double, left and right censoring PPWM

Anli (2006) utilized regional flood frequency estimation by using LM on the Goksu Basin, Turkey Data of ten stations used on several length of records of stream flows, discordancy, heterogeneity measured carried out and simulation studies, founded GEV as best fit for practicing area GNO & PE3 also suitable for the region as alternative Simulations studies 500 years applied, and quantile estimates based on fitted distribution are derived

Norbiato *et al* (2007) worked on extreme precipitations which cause to flash floods in east Italy, by RFA technique Extreme rains and large spatial variation boosted the flood RFA based on index flood and LM method applied to identify the storm location and short duration annual maximum rains Growth curves ground on kappa distribution applied sub-regions Severity graphs visualized the return period and variability for several rainfall durations within storm Outcomes illustrated that analyzing a single return period to a storm event is not realistic, severity of flash flood can't have captured by conventional rain gauge networks

Moisello (2007) analyzed hydrological extremes by PPWM, operate on 297 annual peak flow samples, compared PPWM and PWM, firstly on at-site data and ordinarily Monte Carlo analysis applied, divulged that partial PWM perform better than PWM by habituating Gumbel, Frechet, GEV and GPA distributions Three statistical tools adapted for analysis regression of quantile estimates, parametric one and method of index flood All methods engrafted PPWM as a effectual tool even though outlier exists but censoring make it trenchant

Wallis *et al* (2007) used regional flood frequency analysis and spatial mapping for 24 and 2-hour duration in Washington State Aforementioned state has high topography and climate with greater variability of precipitation ranging through 260 inches to 7 inches per year, reasoned through usurious altitudinous mountains ranges and plied too wet inclines as well as rain shadows falling in shortest range State divided into 12 similar regions on the classifying as 2 hours and 24 hour regions, on the basis of maximum rainfalls Fitted GEV distribution on all stations and found to be best fit

Saf (2008) applied LM in regional flood frequency analysis on the West Mediterranean Tract of Turkey By using the K-means clustering method for formation of similar tracts, then homogeneity tested through the use of Kappa distribution for simulations, PE3 performed best for Lower West area and GLO for Upper West region Accuracy of quantiles also estimated through the use simulation studies, on the basis of relative bias and relative root mean square error

Hussain and Pasha (2009) canvased RFFA of 7 stations of Punjab by LM Determine discordancy value to screen out each station, by using Kapa distribution's 500 simulations measured heterogeneity based on LM's Whole region acknowledged as homogenous Following three or above parameter distributions GLO, GNO, GEV, PE3, GPA and WAK used to choose smashing distribution for quantile estimates GNO, GPA and GEV selected as peachy distributions established on LMRD and Z statistic Quantile estimates and regional growth curve applied by using simulation studies on the behalf of selected distribution GNO performed as robust distribution, ordinarily GPA can be as alter

Shabri *et al* (2011) estimated GPA distribution from censored data by applying PLM script Compared PLM to LM for assessing hydrological immoderations, derived parameters of GPA according to PLM and tested by simulation studies Outcomes showed that censored samples up to 2% are same to LM Data of AMF on 32 stations of Golok River in Kelantan used further calculations and precised that performance of PLM is better than LM for large flows

Ahmad (2011) utilized the LM and TLM methodologies, on annual maximum stream flows and flood analysis on Negeri Sembilan, Malaysia data Estimated the most excellent fit for region, GLO founded as best on the basis of LM ratio diagram and mean square deviation index Used trimming on smallest values(TLMOM1)

Hussain (2011) practiced RFFA on upper basin and lower of River Indus, Pakistan Utilized AMP flow data of 7 sites, positioned at main Indus To test randomness and serial correlation, run test and lag-1 correlation employed, Grubbs test illustrated no major variation in data except low outlier To deal with nuisance of low outliers left censored PPWM applied Discordancy assessment supported to continue all the sites for further, measured heterogeneity pursuit to more than one region's development Development of sub-regions under Wards clustering by using site characteristic utilized Region (R1) contain on four and region(R2) consist on 3 showed different outcomes LMRD, Z statistic, and AWD engaged to select best fit R1 styled GNO, GLO, GEV and PE3 as suitable, and GLO for R2 Outcomes of growth curves and RAARB, RB, RRMS after simulations pronounced PE3 as robust for R1 and GLO for R2

Castellarin et al (2012) reviewed the applied statistical approaches for Regional Flood Frequency Analysis on Europe data Used the flood data from Norway, Poland, Lithuania, Italy, Greece, France, Finland, Germany, Cyprus, Bulgaria, Belgium, Austria, UK, Spain, Slovakia, Ireland, Denmark and GEV found as best fit but LN3, GLO stand as alternative LMRD for all mentioned states were applied

Zakaria and Shabri (2012) work on RFA of extreme precipitation by applying PLM methodology on 37 stations of Selangor catchment, Peninsular Malaysia LM, LH moments technique applied on the data and comparison made with PLM, as well new relationship for GLO and GPA induced grounded on the PLM, and practiced in RFA To ascertain peachy distribution PLM ratio diagram and Z statistic were utilized

Constituting on statistical technique for Selangor extreme precipitation and matching the three itinerary, GEV and GLO nominated as suitable distributions mutually coming from LM, LH and PLM Performance rating delineated that method of PLM surmounted on LM and LH for long perennial period events, at 3% censoring level($F_0 = 0.03$) PLM technique is more effective than LM and LH method for larger return periods, because in flood estimation the major concentration is on the right-hand tail estimation of the distribution and small floods have small relation to large floods while PLM centered on extreme events in data PLM can depress the unwanted influence of small values, on the estimation of long return periods Four parameter Kappa distribution were also employed to form similar region

Keshtkai (2012) dealt with flood flow data in arid and semi-arid arena of Iran by using LM method Data of 17 stations crest discharge taken, and outcome showed L-moments effectiveness over MLE and ordinary moments, concluded that PE3 as best fit for Iranian Central Platue

Zahana (2013) work on the maximum discharges data in Carpathians Tract, Romania, by using LM method Calculated discordancy statistic, developed homogeneous regions and derived most suitable distribution for region By fitting distributions on maximum discharge data strips found GPA, PE3 and Exponential distribution as good fit

Shahzadi et al (2013) delivered RFA of AMR by LM in monsoon area of Pakistan Index flood process induced on 23 stations of monsoon region to compute rainfall quantiles Discordancy measure and heterogeneity statistic value calculated, on the basis of region's geographic locations clustering of sub-regions done, as high elevated areas in Pakistan receives large rains contrary to low elevation areas GEV GLO and GNO flourished as best fitted and average weighted difference value, LMRD and Z test used for selection RB, RAB an RRMSE gave natality to GNO as robust among all GNO also effective for large return periods 50, 100, 500 and 1000 but GEV for 1,2,5,10 and 20 for all 3 regions

Dikbas *et al* (2013) prosecute K means clustering approach to delimit similar regions for stream flow in Turkey K means clustering sought to categorize the AMF and to generate hydrologically similar groups Data of 117 sites used in functioning to form homogenous plots, for mentioned purpose AMRF, variation and skewness coefficient of AMRF, latitude and longitude utilized After practice, obtained 7 groups with similar characteristics, and H statistic based on LM method proved to region's homogeneity, overall determine roots to use K means clustering for foundation of homogenous regions

Al-Khodary *et al*(2014) performed double censoring to estimate generalized exponential distribution by PPWM Right and left censored samples were employed as special dually censored cases Comprehensive numerical analysis used to check the properties of freshly derived estimators and proved that censoring level have significant impact on the enactment of PPWM, but the bias and RMSE boost up as censoring level increases In addition, majority of the PPWM estimators tracks to Pearson I,IV and VI, but it follows to Pearson III especially when double censoring or left censoring is done at α= 5 or 1 Moreover right censoring effect on MSE is lesser than the effect of the left censoring, specifically in the large values of α

Malekinezhad (2014) used LM to estimate daily outermost rainfalls by Regional Frequency Analysis Data of 47 sits used from Golestan, Iran Divided area into five homogenous regions established on cluster analysis of same characteristics as well special geographical and spatial features supported to division Discordancy and heterogeneity outcomes plied way for further analysis, best fit is chosen and quantiles are estimated for each subdivided region

Memon (2014) report on extreme rains and flood in Pakistan by focusing on 2010, 2011,2012 flood and summarize results about previous floods Memon stated disastrous floods harmful effect on human lives and infrastructure, briefly described the loses in Pakistan especially concentrating on Sindh

Latt *et al* (2015) produced clusters of hydrological similitude regions and neural network depend upon index flood process for ungauged catchments on River Chindwin, Myanmar PCA and clustering techniques used for development of similar regions, K means clustering and Ward clustering employed based on physiographic and climatic features Preparation of index flood model done by ANN and regression established on the largest flow track, basin elevation, basin incline, soil erosion and AAP

Carvalho *et al* (2015) worked on variability in climatic factors temperature and precipitation by regionalization of Europe established on K-means clustering Studied the changing climate by earth system models and divide the spatial territory into same regions by k-means clustering Data of daily rainfall and temperature extending through 1986-2005 of Europe used for maneuver and prognosticated model by Coupled Model Inter-comparison Project(CMIP5) tested ranging from 2081 to 2100 Clustering method applied on univariate and multivariate levels Sensitivity test used to check the performance by scrutinizing consistency of outcomes

Ahmad et *al* (2015) dealt with annual extreme stream flows by utilizing robust estimation method on at site analysis, on 18 sites, Pakistan Made comparison of LM, TLM and MLE to select the most robust approach based on several distributions Suitable fit for each site gathered by applying goodness of fit and tests MADI, Anderson darling test, probability plot correlation coefficient, and LMRD GPA found to be best for maximum stations as well as GEV and GLO as alternatives LM approach proved to be robust, then TLM and MLE on last

Ahmad et al (2016) put efforts to find the probability distributions for annual daily maximum rainfall in Pakistan by plying linear moments and variants The study was at-site analysis by using LM, HLM and TLM on 28 stations they applied various nonparametric tests and LMRD to choose the best fit as well as the RMSE to judge the best fit in AFA Different distributions meet the criteria for differed stations over all the major five distributions meet the criteria that were GPA, GNO, PE3, GLO and LN3 TLM were used to cover the effect of outlier where the LHM to disclose the upper tail of distribution, for large sample and large return period Results exposed diverse method's efficiency for dissimilar stations as well as by theoretical reconsidering resolved that the LHM as a special case of TLM

CHAPTER 3

3.1 Material and Methods

A barrage is a gated hydraulic crosswise construction on a river or any other water course to control, adjust, and avert flows to canals or to facilitate navigation A headworks is a hydraulic edifice on a watercourse, slighter than a barrage, which sidetracks the water flow into canals Where another facilitator is reservoir which helps to storage of water. Irrigation, water designing, hydrologic power generation and soil conservation in flood stipulation Although reservoir has many types but most important is flood control reservoir Flood controller reservoirs are assembled for the tenacity of flood resistor, it shields the region lying on its downstream safe from the harms due to flood

Pakistan's irrigation and hydrologic structure is one of the stupendous and integrated irrigation system in the world that comprises on 3 main reservoirs Chashma, Mangla, and Tarbela and 18 barrages, 12 linking canals, 45 canals for irrigation, and more than 107,000 water courses with millions of field dykes The main canal system stretched up to 585000 km where the watercourses & field channels surpasses 1 62 million km Ten barrages of Indus basin are utilized in current study, whom about a brief description is given below,

Mangla Storage Mangla is the 12th largest earth-fill dam in world, constructed in 1961 and start working in 1967 The height of the dam is 380 feet Jhelum at Mangla has a catchment area of about 12,870 square miles Ability of running sluice is 1,100,000 cusecs where the disaster spillway can pour out 2,300,000 cusecs The purpose of edifice was water stowage for enhancing irrigation provisions and hydropower generation The main aim of this construction was agriculture and hydropower generation, therefore, no space is predominantly earmarked for flood control However, storage among the reservation levels 1202 feet and 1228 feet (1 5 MAF) is kept to achieve related protection in flood

Rasul Barrage Rasul Barrage constructed in 1968 on Jhelum, in between the Jhelum and Mandi Bahauddin Districts, located 72 km downriver to Mangla The length of barrage is 3209 feet with 2800 feet of strong water way to permit a extreme discharge of 850,000 cfs This construction habituate for management of irrigation system and flood flows in the Jhelum River, operational with its full capacity of venting of 204070 cubic meter per second At Rasul Barrage water is disported to River Chenab via link canal named as Rasul-Qadirabad link (RQ Link, concentrated capacity = 19000 cfs), with highest capacity of water emission to the link canal of Chashma-Jehulm and the crest level is 708 ft

Marala Barrage Marala Barrage also named as Marala Headworks developed in 1968 on the River Chenab, to source water to Upper Chenab Canal (UCC) and Qadirabad Balloki Link canal (Q-B canal) Marala giving birth to two main link canals, entitled as Upper Chenab and Marala-Ravi Link (MR Link, capacity 22,000 cft) Barrage is 4472 feet long to pass a maximum discharge of 11 lakh ft^3/s 66 bays with 10 inlets as underdrains and the crest level is 800 ft

Khanki Barrage Khanki Barrage 1889, located on Chenab River, listed as the oldest Barrage of Pakistan, used for irrigation and to manage the flood Khanki providing water to agricultural lands exceeding to 3 million acers through its major branch Lower Chenab Canal, and 59 other distributaries and the crest level is 721 ft

Qadirabad Barrage Qadirabad Barrage is situated 32km underneath Khanki Head Works on Chenab river Barrage is stretched on 3373 feet with 3000 feet of flawless water way to permit a extreme discharge of 900,000 cft, having 45 coves with five havens as under sluices The main purpose is to collect water from Rasul Qadirabad link canal and to sidetrack Qadirabad Balloki canal It has the crest level is 684 ft

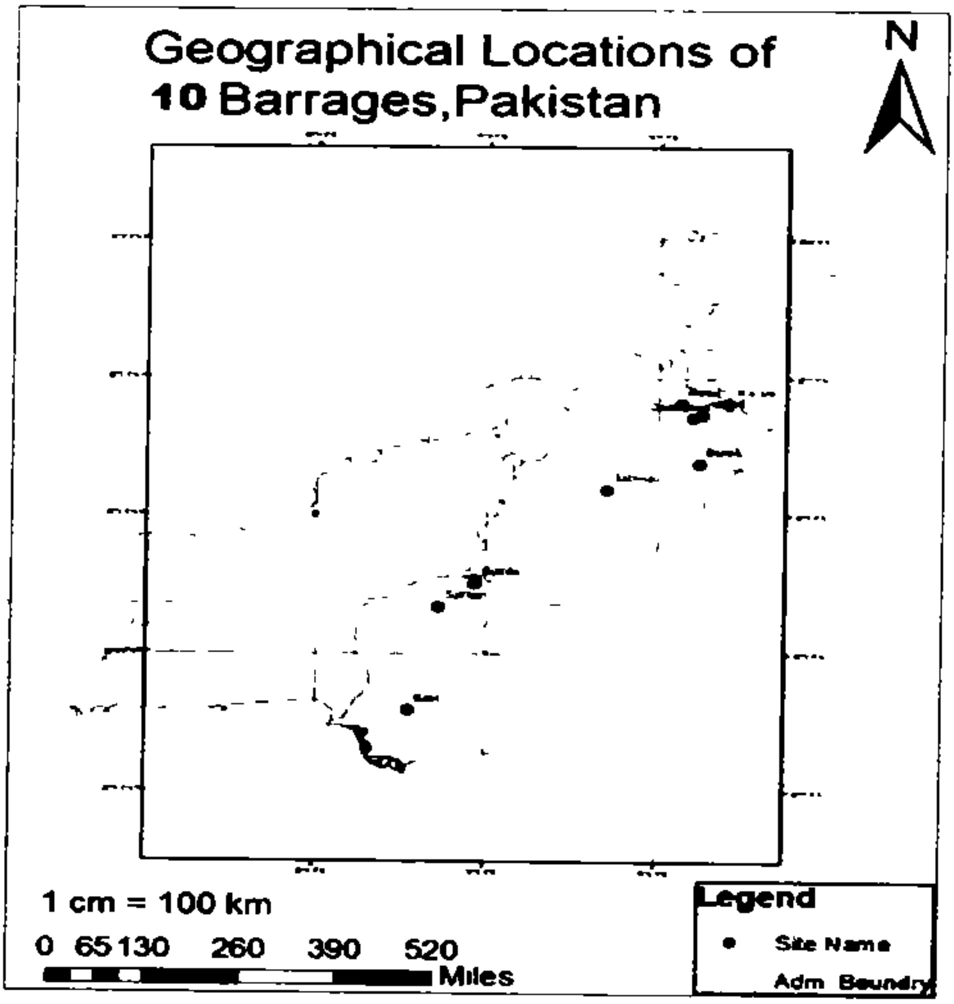
Balloki Barrage, Balloki Barrage is situated on Ravi, build in 1965 The main aim was to source water from Q-B to B-S Link and to provide water to Lower Bari Doab canal The length of barrage is 1647 feet, permitting a extreme release of 225000 cft, and the crest level is 622 ft

Sidhnai Barrage Sidhnai Barrage is positioned on Ravi, build in 1886 To fund water to Sidhnai feeder canal and to feed S-M link (10,100 cft) The length of barrage is 712 feet with 600 feet water way to allow a extreme discharge of 150,000 cft Is has14 havens with five havens under sluices and the crest level is 454 ft

Sukkur Barrage Sukkur Barrage constructed in 1932, mainly functioning to curb water flow for irrigation and reduce flood level Enabling a 9,923 km long canals network, verified as largest irrigation network in the world, feeding beaucoup 5 million acers, allowing 1,500,000 cft maximum acquittal and the crest level is 177 ft

Kotri Barrage Kotri Barrage developed in 1955, purposed for flood management and irrigating the nigh lands Kotri Barrage functions with 875000 cusecs discharge capacity, with 18 meter fanlike 44 bays Barrage alimenting Kolari, Fulleli and Pinyari Canals it can tackle up to 43 1 feet high flood and the crest level is 48 ft

Guddu Barrage Guddu Barrage ramped up in 1962, nigh to Kashmore Sindh, used for both irrigation and flood dealing, with capability of discharging 1 2 million cubic feet per second It has 64 bays each 18 meter broad, supplying water to three districts of Sindh Jacobabad, Larkana and Sukkur and one of the Baluchistan's, Naseerabad Major canals linked to it are Pat Feeder, Desert Feeder, Begari and Gohtki Feeder It can handle 26 feet high flood level and the crest level is 236 ft



[illegible]

The estimation of the amount of water which will pass through a specific site and to know about the probability of its happening is essential for Economic planning and the safe development of

- Bridges,
- Dam,
- Levees,
- Culverts,
- Sewage disposal plan,
- Other construction positioned along river and watercourse
- The effective management
- Flood plan,

- Flood defense scheme,
- To envisage the probable flood magnitude over a specific time period,
- To guesstimate the frequency with which overflows of a certain scale may occur,
- Familiarity of flood occurrence is necessary also to flood indemnification and flood

3.2 Regional Flood Frequency Analysis

Cunnane (1988) examined twelve divergent methods of RFA, and PWM algorithm ranked as best L moments proposed by Hosking (1986a, 1990) are the educed form of PWM, robust to outliers, easy to apply and more accurate Cunnane (1987), Wang (1990 a,b) acquainted the idea of Partial L moments for censored data, extension of Partial PWM to minimize the abdicable influence of small sample events on the derivation of larger return periods floods

3.2.1 Linear Moments Technique

PWM primarily defined by Greenwood (1979), later LM the linear combination of PWM by Hosking and Wallis (1990) The probability weighted moment of rth order was demarcated as

$$\beta_r = \int_0^1 x(F)F^r dF$$

Where $x(F)$ Is quintile function and $F = F(x)$ cumulative distribution function The first four L moments and PWM's are described below,

$$\lambda_1 = \beta_0 \qquad t_1 = \alpha_0$$

$$\lambda_2 = \beta_1 - \beta_0 \qquad t_2 = \alpha_0 - 2\alpha_1$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \qquad t_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \qquad t_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3$$

λ_1 *is measure of location*

λ_2 *is measure of scale*

Lcv (Coefficient of variation) $\tau_2 = \frac{\lambda_2}{\lambda_1}$

LCs (Coefficient of skewness) $\tau_3 = \frac{\lambda_3}{\lambda_2}$

LCk (Coefficient of kurtosis) $\tau_4 = \frac{\lambda_4}{\lambda_2}$

In praxis LM canvased by sample values, to use unbiased estimators is more appropriate b_r of PWM

b_r as

$$b_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(i-1)(i-2)\quad(i-r)}{(n-1)(n-2)\quad(n-r)} x_{i\,n}$$

Relationship of both LM and PWM is as

$l_{r+1} = \sum_{k=0}^{r} p^*_{r\,k} b_k \qquad r = 0,1,2, \qquad (n-1)$

For $r = 0,1,2,3$ Where b_0, b_1, b_2, b_3 are the values outturned by the b_r by putting the different values of r

$l_1 = b_0$

$l_2 = b_1 - b_0$

$l_3 = 6b_2 - 6b_1 + b_0$

$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$

3.2.2 Partial L moments

Wang (1990) yielded the idea of Partial Probability Weighted Moments (PPWM) on the basis of probability weighted moments (PWM)

$$\beta'_r = \frac{1}{1-F_0^{r+1}}\int_{F_0}^{1} x(F)F^r \mathrm{d}F$$

where F_0 is varying point, it can be at 10%, 20% or 30%

But if $F_0 = 0$, the partial linear moments becomes simple linear moments The unbiased estimates of this statistics can be defined as

$$b'_r = \frac{1}{(1-F_0^{r+1})n}\sum_{i=1}^{n} \frac{(i-1)(i-2)\quad(i-r+1)(i-r)}{(n-1)(n-2)\quad(n-r+1)(n-r)} x^*_{(i)}$$

For $x^*_{(i)} = 0$, $x_{(i)} \le x_0$, and if $x^*_{(i)} = x_{(i)}$, $x_{(i)} \ge x_0$

The level of censoring threshold expresses the number of samples data points to be censored

$F_0 = \frac{n_0}{n}$, where n is the length of unrestricted samples and n_0 is the quantity of those values which does not exceed from the censored data points

The first four partial L-moments are correspondent to first four L-moments

The PLCv PLCs, PLCk are defined as

$\varsigma_2 = \frac{\xi_2}{\xi_1}$ Partial linear coefficient of variation PLCv

$\varsigma_3 = \frac{\xi_3}{\xi_2}$ Partial Linear coefficient of skewness PLCs

$\varsigma_4 = \frac{\xi_4}{\xi_2}$ Partial linear coefficient of kurtosis

On the basis of LM theory, Hosking and Wallis (1997) offered RFA procedure elongated on four stages Zakaria and Shabri (2013) further utilized this idea to partial LM for censored data by following same steps Following are the four phases of RFA

(1) Data screening
(2) Formation of Homogeneous Regions
(3) Selection of most suitable frequency distribution
(4) Estimation of selected frequency distribution

3.3 Data Screening

Each statistical procedure possesses basic assumptions, mandatory to fulfill before advancing to further steps Data underneath the statistical operation must meet the essential conditions Primarily the data should show flawlessness or there should not be any gross error After fulfillment of basic assumptions, we can move for next steps In initial screening of sites, the sites stand out those disquieted in data collection, possessing trend and having gross error and expelled from bunch of sites Hosking and Wallis (1997) portrayed a discordance amount D_i on the basis of LM, to distinguish those sites that are utterly discordant with the group as a whole

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

where $S = \frac{1}{N-1}\sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T$ is matrix of sum of squares and cross products

where $u_i = [\ t^{(i)},\ t_3^{(i)}, t_4^{(i)}\ ]$

vector consisting sample LMR

$\bar{u} = N^{-1}\sum_{i=1}^{N} u_i$,

N = Total enumerate of sites

Hosking (1997) plied a touchstone for discordancy statistic, site's collection and relevant D_i brink point

Table 1 1 Critical Threshold for Discordancy Measure

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

If the value of D_i will increase by given criteria of sites, that site will be reckoned as discordant If the value of D_i of a site under the critical value that site will be considered for further analysis The discordancy value is helpful to assist in detecting those stations whose L-moment ratios are altered as compared to L-moment ratios of the collection of sites Hosking Explicitly, sites with a discordancy amount greater than said criteria are treated discordant relative to the collective behavior for the proposed grouping of sites Flagged data sets as discordant in the data screening practice should be revised by groping the probability-plot and time-series graphics to decide how many data values whose degree fluctuates markedly from the overall behavior of the dataset In the presence of discordancy, the record for that station should be checked out that if discordancy result of some error in the measurement, recording or data entry procedure or the data is valid Performance can have tested by comparing nearest sites or by using other sources By the screening of data, we will check out any inconsistency in data or any abrupt changes Zakaria and Shabri (2013) used this approach for partial LM

3.4 Recognition of Homogeneous Regions

1. **Geographical Convenience**
2. **Subjective Partitioning**
3. **Objective Partitioning**
4. **Cluster Analysis for Dissecting**

Geographical Convenience

Regions are selected on the basis of their geographical positions and administrative belt, because these are formed on the basis of already existing similar geographical features This method produces quick output, and support to homogenous regions, but not used extensively in RFA because it seems subjective approach and biased due to physical cohesion Matalas (1975)

Subjective Partitioning

Subjective partitioning enacted on small levels, formed on site characteristics like rainfall quantities and level of stream flow Attained by collating the sites by comparing their characteristics like mean value precipitations, happening of flood, extreme flows by considering time of occurrence as key factor While using at site statistics subjective technique is not useful Operating on rainfall data Schaeffer (1990) and on extreme flows Adamowski (1994) used subjective partitioning method

Objective Partitioning

This method is also used for partitioning of the region with alike attribute Objective partitioning process is in which sites divided into allocated two probable clusters to form regions conditional to whether the physiognomies of the selected sites cross some threshold value or not, objective to minimize heterogeneity exclusively inner the group for illustration of likelihood ratio indicators which is applied by Wilshire (1985) Also practiced by Pearson (1991), by checking disparity of sample L-cv L-skewness within group Furthermore, groups undisclosed till to attain a set of homogeneous regions There is no explicit reason to select this type of technique but it can be more supportive in accumulation with a later assessment of whether the finale region is homogenous

Cluster Analysis

Cluster analysis is a multivariate technique, used for formation of groups having least variability, matching characteristics and congruent features By allocating a data vector to each station and these stations either distributed or added into group of uniform vectors, formation of regions can have practiced At site characteristics are commonly used in cluster analysis to structure homogenous regions but additionally at site statistics can strengthen the process Characteristics of station can be latitude and longitude, annual average rainfall, level of elevation and drainage area can be added to construct cluster

K-Means Clustering

K means clustering procedure go through a simple way to categories data into fixed number of clusters assumed as k-clusters Ideas is that k centroids explained first, representing to each clusters Different locations provide different results that's why centroids should assigned by guileful way It's better to select centroids apart from each other, to attain better homogeneity Carvalho (2015) produced homogenous clusters by K means clustering working on the data temperature and precipitation of Europe Lin (2006) made comparison of SOM, ward clustering and K-means clustering Dikbas (2013) defined homogenous regions by K-means clustering technique on the data of stream flow, Turkey

3.4.1 Heterogeneity test

Homogeneity test explicated by Hosking and Wallis (1997) By using H test we approach to homogeneity of the stations in region, weather the region is homogenous or heterogeneous

$$H = \frac{(V - \mu_V)}{\sigma_V}$$

where V is standard deviation of sample Lcv $V = \{\frac{\sum_{i=1}^{n} n_i(t^i - t^R)^2}{\sum_{i=1}^{n} n_i}\}^{\frac{1}{2}}$

where t^R $Lcv\ of\ Regional\ Average$

$$t^R = \frac{\sum_{i=1}^{N} n_i\, t^{(i)}}{\sum_{i=1}^{N} n_i}$$

μ_V, σ_V Represents the mean and variance of population V

According to the criteria if $H < 1$ the region is homogeneous, H is in between 1 and 2 it can be considered as homogeneous, but not perfectly homogeneous $H \geq 2$ then the region will be considered as perfectly heterogeneous

The use of Kappa distribution for simulations is common because of its tremendous qualities, generated by emerging two gamma distributions, having four parameters (α, ξ, k, h) indicating scale, location, shape and redundant shape parameter, where values lower and upper as [$\xi + \frac{\alpha}{k}(1 - \frac{1}{h^k})$, $\xi + \frac{\alpha}{k}$] Kappa distribution is generalizing form of GEV distribution when h=0, generalized form of GPA if h=1, generalized form of EXP if h=1, k=0 and generalized form of GLO when h=-1

Its density, distribution and quantile functions are given below,

$$f(x) = \frac{(1-k(x-\frac{\xi}{\alpha}))^{(\frac{1}{k}-1)}}{\alpha}\{F(x)\}^{1-h}$$

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

$$x(F) = \xi + \frac{\alpha}{k}\{1 - (\frac{1-F^h}{h})\}$$

3.5 Selection of appropriate Frequency Distribution

Choice of apropos distribution is the next step after the test of homogeneity A number of parametric and assumption free tests are accessible to select the best distribution like as Chi-square fit, Anderson-Darling test, criterion of Akaike information Shapiro-Wilk test, Hosmer-Lemeshow test, Kolmogorov Smirnov method, Cramer Von Misses test and Z-statistic Graphical tactics are also accessible to arbiter goodness of fit, like as line fit passing through the data points, Quantile Plot, ratio diagrams of product moments, LM ratio diagram (LMRD) portrayed on L-CS and L-Ck but plotting methods plied subjective nature outcomes to support the final remarks Regional frequency studies exhibited that more than one distribution can be marked as suitable distribution, by pointing on the best output of quantile estimates based on fixed return periods

Hosking (1990) marked LMRD as powerful tool to identify the distribution for choice by depicting regional data Peel *et al* (2001) proposed the use of LMRD in the selection of best fitted distribution while establishing homogenous regions Sankarasubramanian and Srinivasan (1999) prove the superiority of LMRD over PMRD while working on the properties of LM and CM Madsen et al (1997) used LMRD for the recognition of suitable fit for both partial duration series (PDS) and annual maximum series (AMS) of flood data Fennessey (1993) revealed the benefits of LMRD over product moment ratio diagram on permanent basis, while dealing with Hydrological extreme events data

On the other side, the application of LMRD and Z-statistic advised by Hosking (1997) Outcomes based on the comparison of moments of the distribution and average moments of the regional data Aiming to select the best fit for witnessed data among the simulated distributions Depending on the support of PL-skewness and PL-kurtosis, regional average of skewness and kurtosis will describe about best fit

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

where $B_4 = \frac{\sum_{m=1}^{N\,sim}(t_4^{(m)} - t_4^R)}{N}$

$$\sigma_4 = \left[\frac{\{\sum_{m=1}^{N\,sim}\left(t_4^{(m)} - t_4^R\right) - N_{sim}B_4^2\}}{(N_{sim} - 1)}\right]^{\frac{1}{2}}$$

$t_4^{Dist} = LCk\ of\ fitted\ distribution$

B_4 is regional bias

σ_4 is regional standard deviation

N_{sim} is quantity of simulated regional data on the basis of kappa distribution

More than one distribution can stand out while practicing the preconditioned criteria by Hosking and Wallis (1997) Best fit can be claimed on the basis of lowest output of Z^{Dist} and some time closest to zero $|Z^{Dist}| \leq 1.64$ is the touchstone at 90% confidence level, Hosking's standard concentrate on this In the case, more than one candidatures, best fit will be judged on the base of nearest to zero or equal to zero Z-outturn

3.6 Estimation of Selected Frequency Distribution

By many ways a distribution can be fitted to data derived from homogeneous regions Index Flood Procedure (IFP) is a potent way to estimate quantiles For a given station i the quantile estimates can obtain by replacing the index flood estimates μ_i and quantile function of $q(\ )$ The probability F for the quantile estimates stated as

$$Q_i(F) = \mu_i q(F) \qquad i = 1,2,3 \ldots N$$

where μ_i is site dependent scale factor

$Q_i(F)$ refers to the at site quantile function

$q(F)$ refers to the regional quantile function

Let say $\hat{\mu}_i$ refers to the estimate of scale factor at site i Then the dimensionless rescaled data and suppose that the region has N sites, with site i having sample size n_i and observed data Q_{ij} then

$$q_{ij} = \frac{Q_{ij}}{\hat{\mu}_i}$$

where $i = 1,2,3, \ldots N, \quad j = 1,2,3, \ldots n_i$

Setting regional average mean is equal to 1 $(l_1 = 1)$ By comparing LMR's $\lambda, \tau, \tau_3, \tau_4$ to average LMR of region $t^R,\ t_3^R,\ t_4^R$

$$t_r^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i} \qquad r = 3,4,$$

For estimated regional frequency distribution $\hat{q}(\ )$ denote to quantile function for any site i Outturns of μ_i and $\hat{q}(\ )$ are conglomerated to deduct the quantile estimates for the i^{th} station For the quantile estimates of the non- exceedance probability F, quantile are explicated as

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

Selection of the distribution for quantile estimates while working on regional frequency depending on the robustness of outcomes In the situation of more than one flagged distributions, reliability is the concluding point not the preciseness for picking of the best Monte Carlo simulations is the excellent way to result out quantile estimates LM algorithm plied by Hosking and Wallis (1997) is useful for this purpose Algorithm process by condition with the performance that simulations based on the likening characteristics of actual region Mechanism of simulation evolve the output of quantile estimates for different non-exceedance probabilities i^{th} site at m^{th} replication is $\hat{Q}_i^{[m]}(F)$ and F be the probability of nonexceedance $\frac{\{\hat{Q}_i^{[m]}(F)-Q_i(F)\}}{Q_i(F)}$ used to estimate relative error Mentioned quantity averaged on the total number of repetitions M to derive BIAS and RMSE given as,

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

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

Absolute bias, relative bias and relative root mean square error of estimated quantiles are

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

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

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

In the situation of distribution of estimates is skew the empirical quantiles can estimated by computing ratio of estimated true value $\frac{Q_i(F)}{Q_i}$ for quantiles and $\frac{q_i(F)}{q_i}$ to estimate growth curve On 90% level of regional growth curve fall within the interval 5% of the simulated returns fall below $L_{0\,05}(F)$ and 5% lies above $U_{0\,05}(F)$

$$L_{0\,05}(F) < \frac{\hat{Q}_i(F)}{Q_i} < U_{0\,05}(F)$$

$$\frac{\hat{Q}(F)}{U_{0\,05}(F)} < Q(F) < \frac{\hat{Q}(F)}{L_{0\,05}(F)}$$

90% Confidence interval provide observed and estimated quantities This interval denoted as error bound

CHAPTER 4

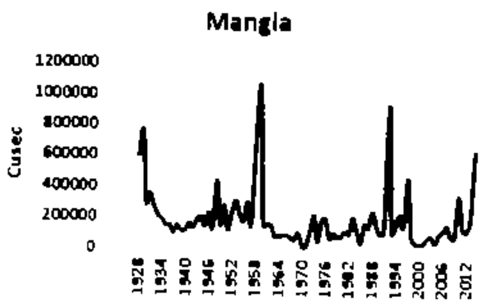
Results and Discussion

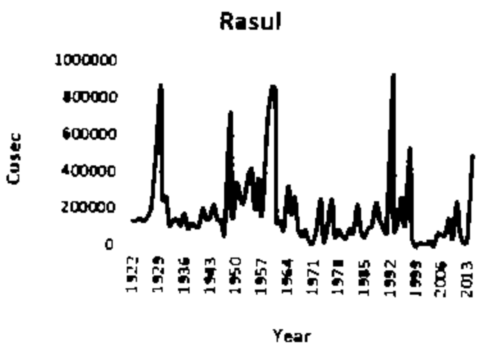
4.1 Elementary Assumptions of Regional Frequency Analysis

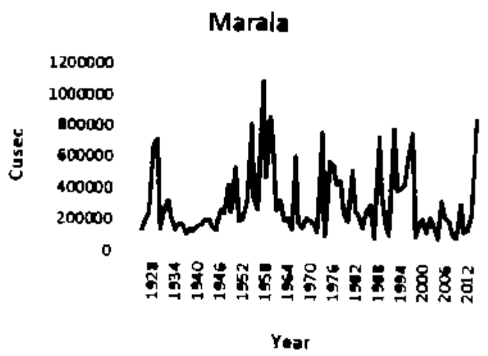
The data under deliberation must fulfil the basic assumptions of distributional homogeneousness, independency stationarity and consistency The mandatory assumptions should satisfy by the data of annual maximum stream flow recorded in cusecs Therefore time series graph and various nonparametric tests are applied to gratify these assumptions

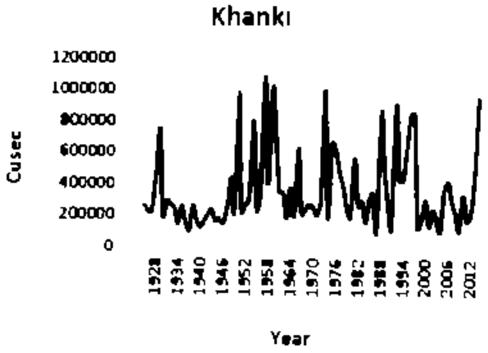
4.1.1 Time Series Diagram

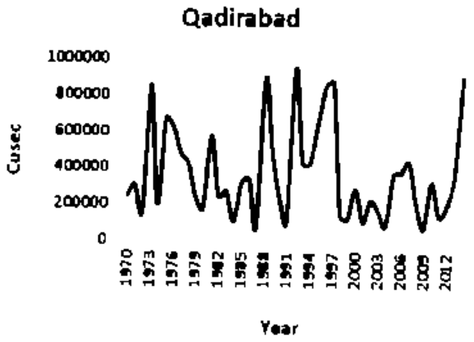
Stationarity over time is one of the basic assumption while working on the hydrological data Graph of the chronological ordered data on variable gives us better idea about stationarity Graph of 10 sites over time rambling from 1920 to 2014 are drawn Time styled on x-axis where extreme flow data on y-axis The diagrams of all the sites are given below.

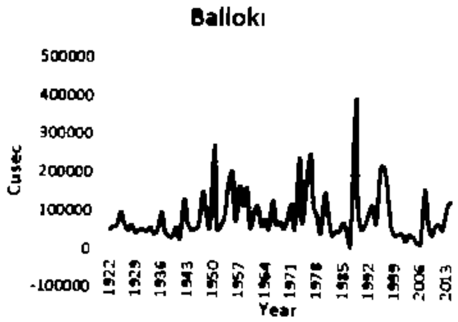


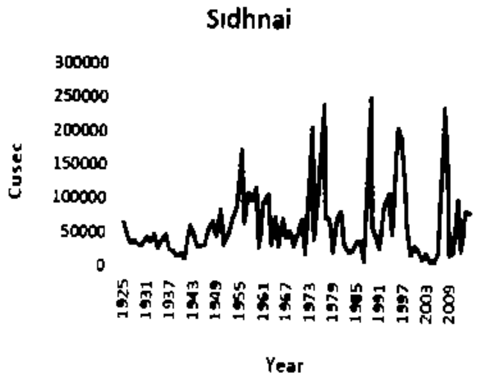


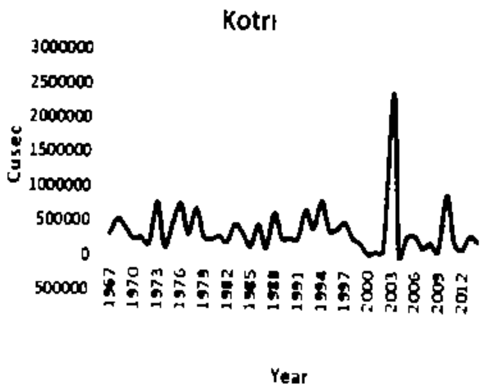

Marala
Cusec
Year

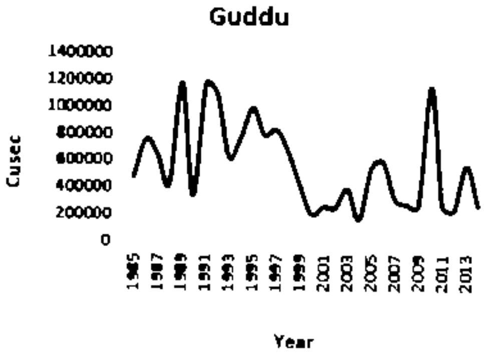

Khanki
Cusec
Year

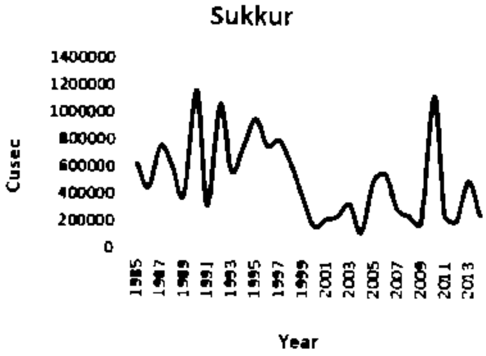

Qadirabad
Cusec
Year


Balloki
Cusec
Year


Sidhnai
Cusec
Year


Kotri
Cusec
Year





Above plots manifest that the data series of all 10 sites increasing and decreasing trend which bespeak about the existence of randomness in data. Some sites diagram explains a repeating behavior over a long time period but overall the graph patterns evident to stationarity.

4.1.2 Difference Sign Test for Randomness

The absence of patterns in data stated as randomness One of the Statistical assumptions is that data should not follow an intelligible pattern or combination The existence of randomization is key to reduction in bias Randomness is a degree of uncertainty of a data series as well as shows haphazardness which relates to concepts of chance, likelihood, probability

Hypothesis of the difference sign test is that

Ho The set of observations exhibiting randomness

Ha The data series has a sequential relationship

Table 4 1 Difference Sign Test

Site	p-value	Result	Conclusion
Mangla	.558	p>0.05	Accept Ho
Rasul	705	p>0.05	Accept Ho
Marala	.702	p>0.05	Accept Ho
Khanki	.251	p>0.05	Accept Ho
Qadirabad	.154	p>0 05	Accept Ho
Balloki	.708	p>0.05	Accept Ho
Sidhnai	702	p>0 05	Accept Ho
Kotri	796	p>0.05	Accept Ho
Guddu	278	p>0.05	Accept Ho
Sukkur	.718	p>0.05	Accept Ho

Difference sign test is applied to test the randomness on all stations data, output of p-value is greater than 05 (p>.05), that reinforced to accept null hypotheses Which's signposts the randomness of all data series

4.1.3 Ljung-Box Test

Autocorrelation is commonly tested in time series analysis Autocorrelation defined as the existence of correlation among the continual terms in time series data For detection of autocorrelation time series plot, lagged scatter plot and the auto correlation tests are used Box and Pierce in 1970 derived this test to check the presence of autocorrelation in chronic data, which is the extension of Q test developed by Bob and pierce

Test estates on the hypothesis that the autocorrelation up to certain lags should be zero Number of lags depend upon the length of the site, but lag's number should lesser than the $\frac{n_i}{4}$ where n_i represent site's length

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

n is the sample size

h is the number of lags to employ

$\hat{\rho}_k^2$ is the k^{th} lag autocorrelation

Table 4 2 Autocorrelation Test

Site	p-value	Result	Conclusion
Mangla	.005	$p<0.05$	Reject Ho
Rasul	.004	$p<0.05$	Reject Ho
Marala	.120	$p>0$ 05	Accept Ho
Khanki	.251	$p>0.05$	Accept Ho
Qadirabad	.589	$p>0.05$	Accept Ho
Balloki	.146	$p>0$ 05	Accept Ho
Sidhnai	025	$p<0.05$	Reject Ho
Kotri	.148	$p>0.05$	Accept Ho
Guddu	.185	$p>0.05$	Accept Ho
Sukkur	.199	$p>0.05$	Accept Ho

Out of 10 sites, 07 sites are considered to be independently distributed, they are not showing any serial autocorrelation Only the Mangla and Rasul's p-value is less than 05, it shows existence of serial autocorrelation but these sites will be further checked by applying various tests

4.1.4 Mann-Whitney U test

In 1947 Mann and Whitney established this test It's used to check the homogeneity assumption This test is non parametric test, alternate to T-test in parametric tests Its assumptions are soft every bit comparing to T-test Mann-Whitney test used to look into two populations, is there shape shows identical behavior or not, under the critic that both samples should be independent

If the sites under analysis antedate the same distribution, like to be undistinguishable, this will facilitate for advancement in Regional Frequency Analysis Null hypothesis for this test is that both populations are followed by the same distribution, though alternative hypothesis contrary to null

H_o Both samples emanated from same populations

H_a Both samples emanated from altered populations

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

n_1, n_2 Are the sample size

Table 4 3 Test for Identically Distributed based on Mean Value

Site	p-value	Result	Conclusion
Mangla	.0001	$p<0.05$	Reject H_o
Rasul	.0001	$p<0.05$	Reject H_o
Marala	.912	$p>0.05$	Accept H_o
Khanki	.396	$p>0.05$	Accept H_o
Qadirabad	.617	$p>0$ 05	Accept H_o
Balloki	.702	$p>0.05$	Accept H_o
Sidhnai	.648	$p>0.05$	Accept H_o
Kotri	.481	$p>0.05$	Accept H_o
Guddu	.273	$p>0.05$	Accept H_o
Sukkur	212	$p>0$ 05	Accept H_o

To practice the mentioned test, the data of each site was divided in to two sub parts The p-value and results are above All stations showed that the data originating from the identical distributions, except the station's Mangla and Rasul But additional Kruskal Wallis test is also practiced for further screening

4.1.5 Kruskal-Wallis test

Kruskal Wallis test is associated with its builder William Kruskal and Allen Wallis developed in 1952 This test is extension of Mann Whitney test, which serve to ascertain that both populations stretched from the same distribution It's also called one way ANOVA, because of its utilization to test several populations During the testing the hypothesis established as sample under consideration possess to same distribution, averse to that they are not from similar distribution This test is sturdy to outliers To comportment the test we first order the data in the pooled total sample from smaller to larger We also need to keep trajectory of the group assignments in the total sample

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1)$$

The critical region for the test is if H > $\chi^2_{critical\ value}$ then we will reject Ho

Table 4 4 Parent Distribution Homogeneity Test

Site	p-value	Result	Conclusion
Mangla	.469	p>0 05	Accept Ho
Rasul	428	p>0.05	Accept Ho
Marala	.470	p>0.05	Accept Ho
Khanki	.469	p>0.05	Accept Ho
Qadirabad	.451	p>0.05	Accept Ho
Balloki	.470	p>0.05	Accept Ho
Sidhnai	.469	p>0 05	Accept Ho
Kotri	.473	p>0.05	Accept Ho
Guddu	.142	p>0.05	Accept Ho
Sukkur	.199	p>0.05	Accept Ho

To perform the test each data series was divided to apply the required procedure The outcomes (p>0 05) are in the favor of null hypothesis which was that the data samples follow the same distribution

4.1.6 Ansari Bradlay test

Ansari and Bradlay in 1960 derived this test, to test out the equality of variance This test is alternate to Siegel-Tukey test (1957) and Moses test (1963) This test is laborious as compare to aforementioned other two tests Ansari Bradlay test is designed to test the null hypothesis that two population distribution functions corresponding to the two samples are identical counter to the alternative they differ by dispersion This test result about the homogeneity of two populations by testing the variance

Combined the both samples from smaller to larger value and rank them Sum up all the rank values, titled as T Rank will be allotted according to this rule that smallest and largest one will possess Rank=1 Then second smaller value and second last largest value will achieve Rank=2 Then third smaller and third last largest value will gain Rank=3 This process will be continuing until all the values ranks tagged

The final remarks in Ansari Bradlay testing based on the critical value, which are defined in the table If $T < X_{1-\frac{\alpha}{2}}$ or $T \geq X_{\frac{\alpha}{2}}, n_1 n_2$ This will facilitate to decide that both samples are following homogenous population or inverse to it

Table 4 5 Variance Equality Test

Site	p-value	Result	Conclusion
Mangla	.147	p>0.05	Accept H_o
Rasul	.162	p>0.05	Accept H_o
Marala	.280	p>0.05	Accept H_o
Khanki	.494	p>0.05	Accept H_o
Qadirabad	.231	p>0 05	Accept H_o
Balloki	.109	p>0 05	Accept H_o
Sidhnai	.091	p>0 05	Accept H_o
Kotri	.206	p>0.05	Accept H_o
Guddu	.268	p>0.05	Accept H_o
Sukkur	399	p>0.05	Accept H_o

To test the equality of variance in the data each data series was divided into two sub-series and Ansari Bradlay Test was executed All the stations data showed the homogeneity in testing under the equality of variance assumption The results contributing to the null hypothesis which was marked as both samples follow the same distribution and homogeneity of the population under variance After the fulfillment of the basic assumptions, now we will move the further steps in the regional flood frequency analysis

4.2 Data Screening and Discordancy Measure by using L-moments

Ab initio passing through all the basic mandatory assumptions of under operation sites by graphs and several multidirectional nonparametric tests, the adjacent way is to move for data's screening by using its enunciated methodology of discordancy measure Discordancy measure fulfills two major purposes The rudimentary objective is the data screening, where it's also used to identify skeptical sites where the problem of data quality may be liable for the discordant comportment Another application is in performing a regional analysis is if the suggested region is outpoured to be heterogeneous as specified by a large value of the heterogeneity measure (H1), then the physical features of discordant stations may be cooperative in understanding the reason of heterogeneity and contribution in defining the course of action desired to develop a homogeneous region.

First step is to marking all the sites in a single homogeneous region Discordancy value D_i for 10 stations is calculated and gauged on the prescription furnished by the Hosking and Wallis (1997)

Table 4 6 Summary Statistics of annual maximum peak flow data

Site No	Site Name	N	$l1$	t	t_3	t_4	D_i
1	Mangla	87	194352	0 4226	0 4544	0 3467	0 77
2	Rasul	93	202372	0 4551	0 4412	0 2920	1 75
3	Marala	90	317614	0 3576	0 3659	0 1581	0 79
4	Khanki	90	360726	0 3520	0 3651	0 1775	0 47
5	Qadirabad	45	373400	0 3764	0 2526	0 0890	1 58
6	Balloki	93	90667	0 3421	0 3810	0 2242	1 12
7	Sidhnai	90	66534	0 4039	0 3871	0 2133	0 42
8	Kotri	48	382169	0 4048	0 4094	0 3255	1 56
9	Guddu	30	542060	0 3272	0 2522	0 0284	0 87
10	Sukkur	30	491594	0 3389	0 2596	0 0743	0 67

The discordancy testing fitted out the results coming forward, with the maximum value of D_i is 1 75 of Rasul, while all other 09 sites D_i value is less than that According to Hosking's Critical values table for 10 sites, if the D_i value is dandier than 2 491, then the site with greater value will be declared as discordant In present lot, any of the site is not exceeding 2 491, so no site will be considered as discordant Just one station flashes out its value, but that value is also under the recommended criteria, while all remaining bunched around 1 (approximately) Thus all sites are considered efficacious for farther procedure

4.2.1 Formatting Homogeneous Regions and heterogeneity Measure by L-Moments Method

Pakistan's physical structure has been ornamented by two main geomorphic progressions that have shaped two different physiographic entities The Indus Plains emanated from the displacement of sediments from the River Indus and its branches The Western Highlands formed by the highland building movement stretched from the Makran coast in the south to the Pamir Plateau in the extent end to north The Western Highlands shield most of Baluchistan, KPK, Gilgit Baltistan and upper parts of the Panjab Although, these regions can be further divided into many regions with minor alterations Our all stations lies in the Indus plains region

The next stride in RFA after ascertaining the discordancy value is to test out heterogeneity value of the under operation region Before ascending to next, the core aim is to shaping homogenous region, based on identical distribution Hosking's touchstone of heterogeneity plies a straight way to destination An Zero outcome of H_1 directs that the site-to-site variability in at-site LCv values for the defined region analogues to expected from a uniform region with the observed L-moment ratios as decanted out by a four-parameter Kappa distribution H_1 positive outturn specify the site-to-site variation of at-site LCv values is greater than the expected for a homogeneous region, and H_1 larger values indicate possible or likely heterogeneous region Contrariwise, H_1 negative results refer the site-to-site variability of at-site L-Cv values is fewer than predicted and the planned region would be recognized as homogeneous Considering all 10 sites as following to identical distribution, the H statistic is measured

Table 4 6(a) Heterogeneity Statistics H1 by LM Method

Number of simulations	500
Observed s d of group L-CV	0 0404
Simulation Mean of s d of group L-CV	0 0326
Simulation standard deviation of s d of group L-CV	0 0079
Heterogeneity measure H[1]	**0.98**

Table 4 6(b) Heterogeneity Statistics H2, by LM Method

Number of simulations	500
Observed s d of group L-CV/L-skew distance	0 0622
Simulation Mean of s d of group L-CV/L-skew distance	0 0604
Simulation standard deviation of s d of group L-CV/ L-skew distance	0 0137
Heterogeneity measure H[2]	**0.14**

Table 4 6(c) Heterogeneity Statistics H3 by LM Method

Number of simulations	500
Observed s d of group L-CV/ PL-K distance	0 0845
Simulation Mean of s d of group L-CV/ PL-K distance	0 0769
Simulation standard deviation of s d of group L-CV/ PL-K distance	0 0177
Heterogeneity measure H[3]	**0.43**

Hosking and Wallis (1997) gives three different aspects about heterogeneity values If $H<1$ then the region is utterly homogenous If $1 \leq H \leq 2$ the region may be homogenous But if $H \geq 2$ then region is prfectly hetrogeneus H_1 that is incorporated on L-CV is a strong nominator to grapple with homogeneity efficiently L-CV and L-skewness are used to calculate H_2, while L-Skewness and L-Kurtosis for H_3 As H_1 ground on L-Cv, it is stronger than both others because L-CV explicate more variation as compare to L-Skewness and L-Kurtosis H_1 = 0 98, H_2 =0 14, H_3=0 43 values indicate that no value is greater than 1,which meet the criteria of homogeneous region It validates the possibility of formation of single homogenous region comprise on all available sites

4.3 Data Screening and Discordancy Measure by Using Partial L-Moments Method

The main objective of the discordancy value is to recognize the barrages for that sample PLM are distinctly dissimilar comparing to other barrages Station with highest error will position out from other stations and be identified as discordant concerned by Hosking and Wallis (1997) criteria

Table 4 7 Summary Statistics for annual maximum peak flow data with censoring

Site No	Site Name	n	c	$l1$	t	t_3	t_4	D_i
1	Mangla	84	3	199414	0 4138	0 4662	0 2070	0 77
2	Rasul	90	3	208367	0 4421	0 4522	0 1453	1 99
3	Marala	90	0	317614	0 3576	0 3659	0 0299	0 38
4	Khanki	89	0	362860	0 3515	0 3618	0 0297	0 19
5	Qadirabad	45	0	373398	0 3764	0 2526	0 0028	1 44
6	Balloki	91	2	92263	0 3333	0 3968	0 0910	1 33
7	Sidhnai	79	11	73269	0 3700	0 4001	0 1128	0 55
8	Kotri	48	0	382169	0 4048	0 4094	0 2231	1 85
9	Guddu	30	0	542060	0 3272	0 2522	-0 0246	0 85
10	Sukkur	30	0	491594	0 3389	0 2596	0 0109	0 65

All discordancy outturns illustrated that Di fall in acceptance expanse, that's why no site discorded A peculiar divergence in discordancy value of LM and PLM moments is discernible, appraiser of discordancy is describing decreasing movement in partial linear moments as comparing to LM The rationality for this change might be due to the censoring of data Censoring threshold was arrayed on the data lower than the

25,000 cusecs Those values are censored which can have deliberated as nuisance in the data, which can mislead the results to biasness Another objective is to focus on the upper tail of the fitted distribution Censored data points on different sites varies and the better results of partial linear moments are evident in table in the form of discordancy and heterogeneity outcome in the favor of superiority of PLM Maximum number of values censored on the Sidhnai that's why its mean value raised up to 73269 from 66534 It is also due to our focus on the large return period by censoring nuisance values

4.3.1 Formation of Homogeneous Regions and Heterogeneity Measure by PL-Moments Method

Partial L moments are akin to L moments, and ongoing results also like PL-Cv, PL-Sk & PL-Ks provided by Hosking and Wallis (1997) same standards are shadowed here If $H<1$ then the region is absolutely homogenous If $1\leq H\leq 2$ the region may be heterogeneous But if $H \geq 2$ then region is perfectly heterogeneous

Table 4 7(a) Heterogeneity Statistics H1 by PLM Method

Number of simulations	500
Observed s d of group PL-CV	0 0368
Simulation Mean of s d of group PL-CV	0 0319
Simulation standard deviation of s d of group PL-CV	0 0080
Heterogeneity measure H[1]	**0.62**

Table 4 7(b) Heterogeneity Statistics H2 by PLM Method

Number of simulations	500
Observed s d of group PL-CV/ L-skew distance	0 0631
Simulation Mean of s d of group PL-CV/ L-skew distance	0 0594
Simulation standard deviation of s d of group PL-CV/ L-skew distance	0 0134
Heterogeneity measure H[2]	**0.28**

Table 4 7(c) Heterogeneity Statistics H3, by PLM Method

Number of simulations	500
Observed Standard Deviation of group PL-CV/ PL-K distance	0 0890
Simulation Mean of s d of group PL-CV/ PL-K distance	0 0756
Simulation standard deviation of s d of group PL-CV/ PL-K distance	0 0178
Heterogeneity measure H[3]	**0.75**

H_1 = 0 62, H_2 =0 28, H_3=0 75 values specify that no value is grander than 1, which is fulfilling the criteria of homogenous region Heterogeneity's H, depicting rearward trend similar to discordancy measure, as above values of heterogeneity of PLM are lower as corresponding to LM, it corroborates to move forward for selection of best fit distribution in said region and the efficiency of PLM over LM

4.4 Selection of Best Fitted Regional Distribution by using LM and PLM

Third phase in RFFA is the fitting of distributions and the choice of best fit distribution Hosking and Wallis (1997) ply the criterion to ascertain the foremost three parameters distributions like generalize Pareto (GPA), generalize logistic (GLO), generalize extreme value (GEV), generalize normal (GNO) and generalize Pearson 3 type (PE3) While commencing this, we keep in mind the both aims, premier is the nomination of the best distribution, ordinal is the estimation of quantiles of each region for several time periods But three distributions can be used for PLM, because GEV, GLO and GPA are the distributions only their partial L moments are derived till now (Zakaria 2013)

Hosking provided the two ways to approach the best distribution Main is the Z-fit while other one is the ratio plot Z- fit apply through the critical value if $|Z^{Dist}| \leq 1.64$ at level of Significance 10% It might be potential that more than one distribution strike to the said limits, than the distribution approaching to zero will be best considered as best fit

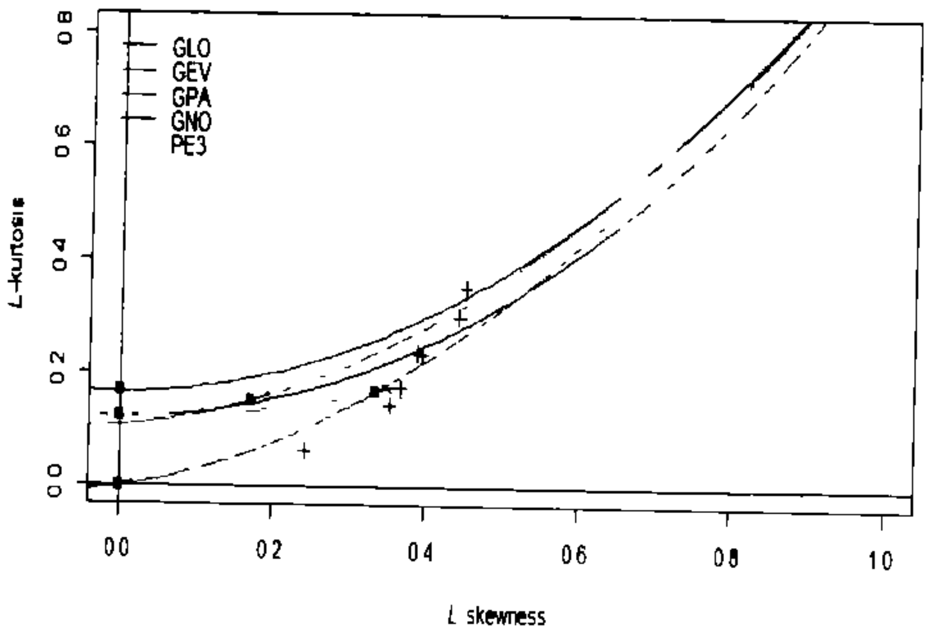
Table 4 8 Z-Statistics for Best Fitted Distribution

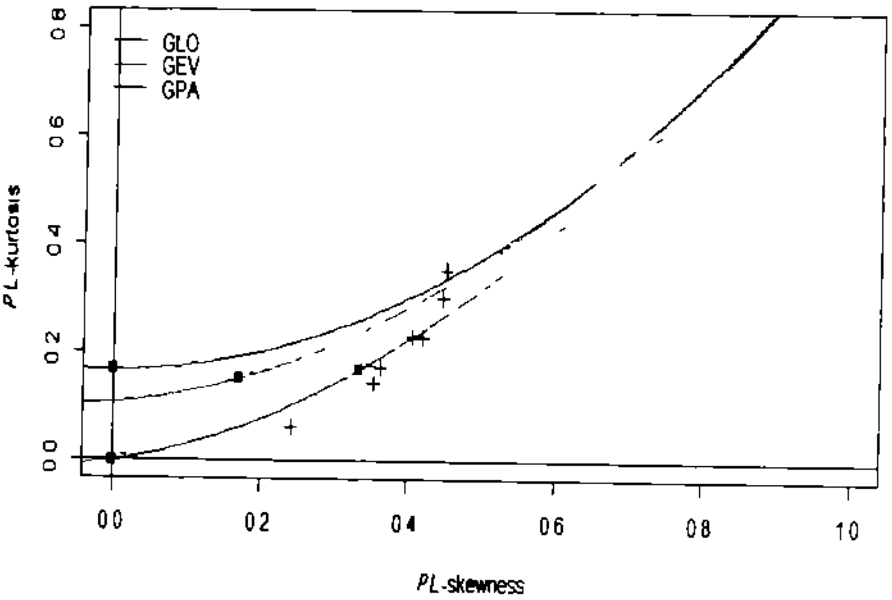
Z-Statistic Value					
Method	GNO	Pearson3	GEV	GLO	GPA
L moments	**0.71**	-1 76	2 14	3 04	**-0.83**
PL moments			2 29	3 15	**-0.53**

Above table evinced the result of the three parameter distributions for the formatted region undergoing the both procedures According to condition of best fit the distribution with minimum value or approximately zero value should be selected Evaluations under narrated criteria reveals that GNO confirm the standards as most appropriate distribution out of five and GPA as alternative one, while practicing L-moments method The estimate of GNO marked as lowest comparing to others Where the GPA emulated for the said region under the PL-moment procedure Both distributions GNO and GPA are nominated as best fit in the region under LM and PLM methods

4.4.1 L-moments and PL-moments Ratio Diagram

L-Moments Ratio Diagram (LMRD), delivers direction to choose a suitable distribution to articulate a collection of variables into a discrete relationship of LM-Ratios which already exist for each probability distribution clarified by Hosking and Wallis (1997) Already stated that partial linear moments are correspondents to linear moments, so the partial linear moment ratio diagram (PLMRD) is also comparable to LMRD LMRD and PLMRD are delineated for the described region L- skewness and L-kurtosis used for LMRD PL-skewness and PL- kurtosis used for PLMRD Regional averages of L-skewness and L- kurtosis fall contiguous to the GNO where regional average of PL-skewness and PL-kurtosis fall nearest to GPA




GLO
GEV
GPA
PL-kurtosis
PL-skewness
0.0
0.2
0.4
0.6
0.8
1.0

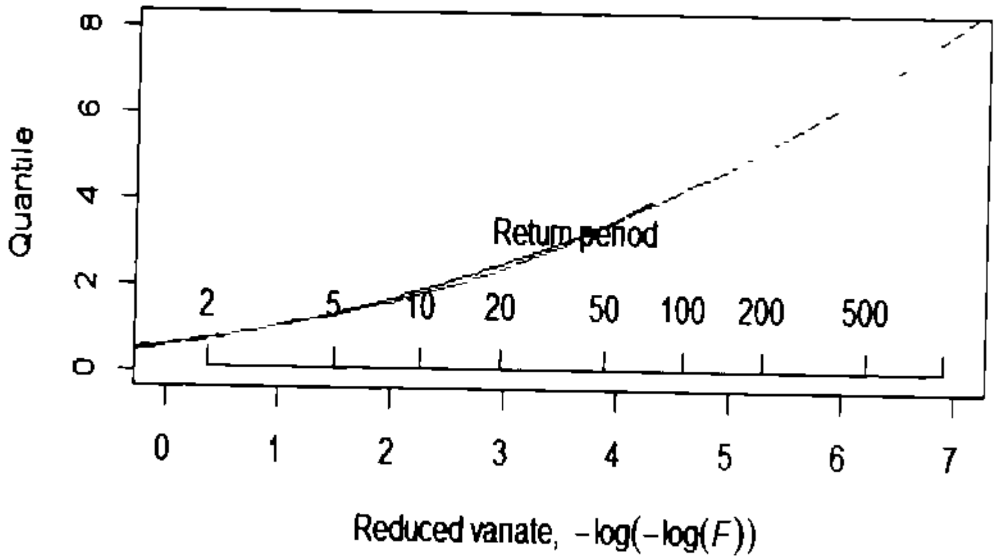
4.5 Fitted Distribution and Regional Growth Curve Estimation

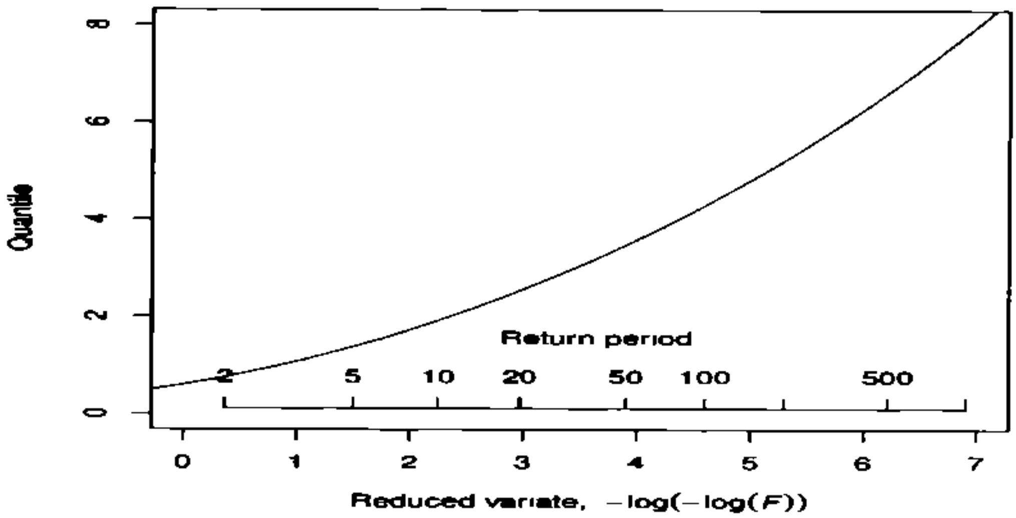
After the selection of best possible fit the next step in regional flood frequency analysis is to find out the quantile estimates on various return periods and plotting of the growth curves by plying the chosen distribution A return period can be defined as the average recurrence interval of an event such as a floods, droughts, streamflow, rainfalls or earthquake Return period of time T can be termed as $\frac{1}{P}$ with its exceedance probability P Probability of occurrence or exceedance is the chance of occurring of an event in a specific time span i e $P = \frac{1}{T}$ probability of occurrence For instance in situation of 20 years ($\frac{1}{20} = 0.05$) is definable as the chance of exceedance, where ($1 - \frac{1}{20} = 0\ 95$) is the probability of non-exceedance

The peak flow quantile estimates of prescribed region for 2, 5, 10, 20, 50, 100, 500 and1000 year are illustrated in Table (4 9) We may drive quantile estimate for each i^{th}site in the said region for a particular return period Let consider the station Guddu which has on the average annual maximum peak flow 542,060 cusecs We can attain for Guddu, by multiplying the regional quantile estimate to mean flow of the relevant site As the table values, $\hat{q}_{GNO}(0\ 980)$=3 473, interpretable as 542060*3 437=1863060 2 cusecs is the amount of extreme flow once in coming 50 years (for given return period) with non-exceedance probability 0 98 and exceedance probability 0 02 at Guddu All other stations can be interpreted in the similar way

If a homogeneous region fulfill the criteria that all sites within the region describable by a single probability distribution possessing common distribution parameters after the rescaling of site data by their at-site mean, this rescaled dimensionless probability distribution is named as regional growth curve The presentation of quantile estimate by growth curve is illustrated in fig-5 3 (LM) and fig-5 4 (PLM), by depicting quantiles on y-axis where non-exceedance probability on x-axis Growth curve displays relationship among the quantiles and the several return periods and provide the roots for frequency analysis at sites While analyzing extreme events such as floods, quantile estimates in one tail of the distribution will be of special interest The fig-5 3 in which the GNO and GPA are portrayed is showing the same behavior up to 50 years As the increment in years occur the GPA showed more upward up to 200 return period as comparing to GNO, but for larger return periods GNO showing higher quantile estimates In fig-5 4 GPA showed lower quantile for lower return period and higher quantile estimates for larger return period Further the accuracy of these curves will be checked by the error bounds and RMSE

Method	Dist.	Parameters			Quantile Estimates											
		b	a	k	Q_{10}	Q_{20}	Q_{50}	Q_{100}	Q_{200}	Q_{500}	Q_{800}	Q_{900}	Q_{950}	Q_{980}	Q_{990}	Q_{999}
LM	GNO	0.7540	0.5207	-0.8013	0.205	0.230	0.278	0.337	0.435	0.754	1.380	1.919	2.532	3.473	4.295	7.834
	GPA	0.2687	0.6604	-0.0970	0.275	0.282	0.303	0.339	0.418	0.742	1.419	1.972	2.564	3.411	4.103	6.766
PLM	GPA	0.2894	0.6337	-0.1083	0.296	0.302	0.322	0.357	0.433	0.740	1.404	1.946	2.532	3.376	4.073	6.801




Quantile
0
2
4
6
8
Return period
2
5
10
20
50
100
500
0
1
2
3
4
5
6
7
Reduced variate, −log(−log(F))

4.6. Decision about Regional Growth Curve and Estimated Quantiles

The Monte Carlo simulation procedure based on simulations to check the accuracy of estimated quantiles and growth curves is used which was designed by Hosking (1997) These calculations are impossible to execute analytically because the regional L-moment quantile estimation process is too intricate Therefore, Monte Carlo simulation procedure is suitable The repetitions in this process must be larger enough so that the RMSE and bias values are contiguous to the actual RMSE and bias Therefore, ten thousand repetitions are performed

The biases, RMSE and error bounds of the estimated growth curves are measured The outcomes are depicted in the Table 4 10 BR(F) regional average relative bias calculates the trend of quantile estimates to be uniformly too larger or smaller through the entire region AR(F) regional average absolute bias tell us about the trend of the quantile estimates that consistently high at some points and low at others This happen in the situation of heterogeneous region. when sometime regional growth curve tends to overestimate a few at-site growth curve and underestimates to others RMSE demonstrate about the overall variation in estimated quantiles from actual quantiles By this criteria we are much focused in determining that which estimation procedure is superior to other, Hosking (1997)

Both techniques linear moments and partial linear moments are used on the formed region Relative mean square error values delivered by GNO of LM are low for smaller return periods 2, 5,10 and 20 where these are higher for larger return period Relative bias continually decreasing as the return periods increasing, meanwhile the absolute bias increasing over the increase in the return period that is definable as smaller for lower return periods 2, 5, 10, 20 and 50 and high for larger return periods 100, 500 and 1000

RMSE by GPA of L moments is low for lower return periods 2, 5, 10, 50 and 100 and it is large for larger return period 500 and 1000 Relative bias declining up to the 100 return period and then rising at the 500 and 1000 return period Absolute bias showed lower trend for smaller return period 2, 5, 10, 50 and 100 and higher values at larger return periods 500 and 1000 GNO perform better than the GPA under LM method

RMSE provided by GPA of PL moments is lower at smaller return periods where its higher for the larger return period 500 and 1000 but it is lower than the RMSE of GNO and GPA of LM method Relative bias by GPA of PLM showed constantly decreasing behavior over the increment in return periods while it is also lesser than the biases of GNO and GPA of LM

Absolute bias by GPA, (PL) is low at the return period of 2, 5, 10, 20, 50 and 100 where its higher at the return period of 500 and 1000 By comparing PLM and LM, PL moments showed lower absolute bias as paralleling to LM

Error bounds provided by GPA of PL method are more narrow as compared to the error bound provided by GNO and GPA of LM method, where GNO and GPA by LM both are satisfactory up to smaller return periods The performance of GNO and GPA distributions by LM is satisfactory up to the 50 (F=0 98), but for larger return periods 100,500 and 1000 GPA by PL is much robust As the main objective of the RFFA is to estimate the quantiles in upper tail of the distribution and to know about the larger return periods, therefore we can say that GPA distribution by PLM has more robust results These results also reinforce the superiority of PLM over LM and the efficiency of PLM to estimate large return periods

Table 4 10 Showing LM PLM RMSE, Absolute Bias, Relative Bias and Lower and Upper Bounds

Method	Dist	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
LM	GNO	$R^R(F)$	0 2500	0 0600	0 0340	0 0560	0 0800	0 1110	0 1320	0 1800	0 2000
		$B^R(F)$	0 0456	0 0079	0 0020	-0 0001	-0 0016	-0 0029	-0 0034	-0 0035	-0 0030
		$A^R(F)$	0 2027	0 0483	0 0281	0 0473	0 0659	0 0887	0 1049	0 1398	0 1538
		LEB	0 7240	0 9080	0 9490	0 9130	0 8750	0 8320	0 8030	0 7470	0 7260
		UEB	1 5040	1 1040	1 0570	1 0930	1 1360	1 1910	1 2310	1 3210	1 3590
	GPA	$R^R(F)$	0 2470	0 0550	0 0380	0 0560	0 0760	0 1070	0 1330	0 2030	0 2380
		$B^R(F)$	0 0456	0 0054	0 0017	0 0004	-0 0006	-0 0012	-0 0007	0 0043	0 0086
		$A^R(F)$	0 2004	0 0447	0 0311	0 0476	0 0636	0 0869	0 1066	0 1588	0 1835
		LEB	0 7300	0 9160	0 9420	0 9160	0 8830	0 8370	0 8020	0 7230	0 6900
		UEB	1 4970	1 0960	1 0640	1 0930	1 1280	1 1840	1 2340	1 3750	1 4460
PLM	GPA	$R^R(F)$	0 2010	0 0450	0 0330	0 0500	0 0670	0 0890	0 1080	0 1570	0 1800
		$B^R(F)$	0 0320	0 0060	0 0031	0 0007	-0 0018	-0 0053	-0 0076	-0 0113	-0 0119
		$A^R(F)$	0 1649	0 0364	0 0271	0 0425	0 0555	0 0723	0 0862	0 1234	0 1411
		LEB	0 7680	0 9340	0 9520	0 9250	0 8970	0 8580	0 8300	0 7620	0 7340
		UEB	1 3920	1 0810	1 0560	1 0830	1 1090	1 4800	1 1800	1 2690	1 3130

CHAPTER 5

5.1 Summary and Conclusions

This study delivers an assessment of the PLM and LM by revising regional frequency analysis established on the LM by Hosking and Wallis (1997) Both techniques were applied on 10 barrages, stationed in Indus Basin, Pakistan All the barrages positioned on different branches of Indus River The annual maximum flow data is collected from Pakistan Flood Commission and Indus River System Authority Annual maximum flow (AMF) described in cusecs and the data under analysis is ranging from 30 years to 93 years

The basic assumptions of regional frequency analysis are verified through the various tests that are the time series plotting, sign test to check randomness, Ljung-Box test to check autocorrelation, Mann Whitney-U test to check the homogeneity by utilizing mean value, Kruskal Wallis test to check the Homogeneity of the parental distribution based on variance and Ansari Bradlay test to check the equality of variances The data of all sites satisfied the mentioned tests which were implemented to check that data at any site should be independently distributed, stationary and follow the identical distribution

Regional frequency analysis procedure based on the four stages The first one is the scrutiny of the data by means of discordancy measure, in which all stations satisfied the defined criteria under the both techniques PLM and LM Which indicates that there is no major gross error in the data Second step in the regional frequency analysis is the construction of the homogeneous region, which is deliberated as the most difficult step in regional frequency analysis At start the 10 barrages were considered that they are following the single region The heterogeneity measure is calculated under the both methods PLM and LM and both meet the criteria of homogenous region defined by the Hosking and Wallis (1997), because all barrages located in same region due to the physiographical division of Pakistan Under both performance's Partial LM showed lower values of heterogeneity as compared to LM So, this region is considered as the suitable for further analysis

The third step in the regional frequency analysis is the selection of the best fitted suitable distribution for the formed region which is selected on the basis of linear moment ratio diagram and Z statistics On the basis of L-moments ratio diagram and

Z-statistics it was established that GNO is ultimately suitable and GPA as an alternative for the region under LM method GPA outperformed by plying the PLM and selected for quantile estimation The selection of appropriate distributions provides a way to further analysis of quantile estimates The last stage is the assessment of the quantiles and regional growth curves Quantile estimates are presented in table (4 9) and growth curves for the assured distributions are revealed in table (4 10)

Monte Carlo Simulations are performed up to 10,000 repetitions to compute the degree of precision such as RMSE, relative bias, relative absolute bias lower error bound and upper error bound for the regional growth curves showed in Table 4 10 Based on simulations results, the GPA gives most robust and accurate estimates under the PLM method as comparing to GNO and GPA of LM The primary objective of the frequency analysis of the extreme events is the approximation of quantiles for upper tail of the distribution, and PLM provided most robust results for upper tail which is helpful in estimation of extreme flood Because PL moments reduced the undesirable influence of small sample events on the estimation of large return period events That's support the supremacy of the PLM over LM for the estimation of the larger return periods

5.2 Recommendations

- The study can be extend to all barrages of Pakistan
- Comparison can be performed by using PLM, TLM, HLM and LM by estimation of quantiles on flood data
- This study can perform on the rainfall data and other data required to test quality performance on different levels like air, temperature etc
- These results can be used in the development of different level architects to reduce the harmful effects of extreme quantity of water

5.3 References

- Ahmad, I , Fawad, M , Mehmood, I , (2015) "At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods", Pol J Environ Study, 24 11-21
- Ahmad, I , Fawad, M , Abbas, I , Saghir, A , (2016) "Probability modeling of low flows at different sites of Indus basin in Pakistan, using L-moments and TL-moments", Pakistan Journal of Science, 68 86-92
- Ahmad, I , Abbas, A , Saghir, A , Fawad, A , (2016) "Finding probability distributions for annual daily maximum rainfall in Pakistan, using L-moments and variants", Pol J Environ Study, 25(3)1-13
- Anli, S A , Apaydin, H , Faz, O , (2006) " Regional flood frequency estimation for the Goksu River Basin through Linear moments", River Basin Flood Management, 424-438
- Annual Flood Report (2015), Govt of Pakistan, Ministry of Water and Powers
- Bhattari KP (2004) Partial L moments for the analysis of censored flood samples Journal of Hydrol Sci 49(5)
- Cunnane, C , (1987) "Review of statistical models for flood frequency estimation In Hydrological Frequency Modelling", Singh VP (ed) Reidel Publishing By Dordrecht, 49–95
- Carvalho, M J , Goncalves, P , Teixeira, J C , Rocha, A , (2015) "Regionalization of Europe based on the k-means cluster analysis of the climate change of temperature and precipitation", Chem Earth, 1-18
- Castellarin, A , Kohnova, S , (2012) "Review of applied statistical methodsfor flood frequency analysis in Europe", Center for Ecology and Hydrology.
- Dikbas, F , Firat, M , Koc, A , Gungor, M , (2013) "Defining homogeneous regions for stream flow processes in Turkey using k-means clustering method", Arab J Sci Eng, 38 1313-1319
- Hosking J R M, Wallis J R, (1986) " Estimation of generalized extreme value distribution by the method of probability weighted moments", Technometrics, 27 251-261

- Hassan, A S , (2005) "Partial probability weighted moments for generalized Pareto distribution", Faculty of Commerce J , 30 30-41
- Hosking, J R M , Wallis, J R , (1997) "Regional Frequency Analysis An Approach Based on L-moments", Cambridge University Press Cambridge, 244
- Hosking, J R M , (1990) "L-moments Analysis and estimation of distributions using linear combinations of order statistics", Journal of Royal statistical Society, Series B, 52, 105-124
- Keshtkar, A A , Kabir, M J , (2012) "Flash flood preparedness in Golestan province of Iran", Am J Disaster Med, 4 197-214
- Khodary, E H , Allam, S A , (2014) "Double Censoring Partial Probability Weighted Moments Estimation of the Generalized Exponential Distribution", SSRN Electronic Journal, 2-20
- Koulouris, Z A , Vogel, M R , (1998) "L moment diagram for censored observations", Water Resources Research, 34 1241-1249
- Keshtkar, A R , Salagegheh, A , Najafi, M , (2012) "Flood flow frequency model selection using L-moments method in arid and semi-arid regions of Iran", DESERT, 17 41-48
- Lin, G F , Chen, L H , (2006) "Identification of homogenous regions for regional frequency analysis using the self-organizing map", Journal of Hydrology, 324 1-9
- Latt, Z , Urban, B , "Clustering Hydrological Homogeneous Regions and Neural Network Based Index Flood Estimation for Ungauged Catchments an Example of the Chindwin River in Myanmar", Water Resource Management, 29 913-928
- Malekinezhad, H , Garizi, A , (2014) "Regional frequency analysis of daily rainfall extremes using L-moments approach", Atmosfera, 27 411-427
- Memon, I , (2014) "Gender and social vulnerability to climate change", Social Policy and Development Center Report, Karachi
- Matalas, N C , Slack, J R , Wallis, J R , (1975) "Regional Skew in search of a parent", Water Resources Research, 11 815-26

- Moisello, U , (2007) "On the use of partial probability weighted moments in the analysis of hydrological extremes", Willy Inter Science, J Hydro Processes, 21 1265-1279
- 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, 345 149-166
- Pakistan floods (2010) Impact Assessment Report JHS-PAK-WHO Report
- Saf, B , (2008) "Regional flood frequency analysis using L-moments for Medtierranean region of Turkey", Water Resour Manage, 10 1007/s11269-008-9287-z
- Shahzadi, A , Akhter, A S , Saf, B , (2013) "Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments", Pak j stat oper res, 1 111-136
- Shabri, A , Zakaria, A Z , (2011) "Flood frequency analysis of annual maximum stream flows using LM and TLM approach", Applied Mathematical Science, 5 243-253
- Stedinger, J R , Vogel, R M , Foulfoula, G E , (1993) 'Frequency analysis of extreme events", Handbook of hydrology, Mc Graw-Hill, New York
- Sankarasubramanian, A , Srinivasan, K , (1999) "Investigation and comparison of sampling properties of L-moments and conventional moments", Journal of hydrology, 21 13-34
- The Indus Water Treaty (1960) Report issued by World Bank
- Wallis, J R , Schaefer, M G , Barker, B L , Taylor, G H , (2007) "Regional frequency analysis and spatial mapping for 24-houe and 2-hour durations for Washington State", Hydrol Earth Syst Sci, 11 415-442
- Wang, Q J , (1990a) "Estimation of GEV for censored samples by method of partial probability weighted moments", J Hydrology, 120 103-110
- Wang, Q J , (1990b) "Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimation of GEV distribution", J Hydrol, 120 115-124

- Wang, Q J , (1997) "Using partial probability weighted moments to fit the extreme value distribution on censored samples", Water Resource Res, 32(6) 1767-1771
- Hussain, Z , (2009), "Regional flood frequency analysis of the seven sites of Punjab, Pakistan, Using L moments", Water Resource Manage 23 1917–1933
- Hussain, Z , (2011) "Application of the regional flood frequency analysis to the upper and lower basin of Indus River, Pakistan", Water Resour Manage, 25 2797-2822
- Zakaria, A , Shabri, A , Nadiah, U , (2011) "Estimation of generalized distribution from censored flood samples using partial L-moments", Journal of mathematics research, 3 112-120
- Zakiria, A , Shabri, A , Nadiah, U , (2012) "Regional frequency analysis of extreme rainfalls in the west coast of Peninsular Malaysia using partial L-moments", Water Resource Manage, 26 4417-4433
- Zakaria, A , Shabri, A , (2013) "Regional frequency analysis of extreme rainfalls using Partial L-moments methods", Theor Appl Climatol, 113 83–94
- Zaharia, L , (2013) "L-moments and there use in maximum discharges analysis in curvature Carpathians region", Advances in Meteorology, 119-126