Selection of best-fit probability distribution for at-site flood frequency analysis (FFA) in Pakistan

By

Azmat Ali

Department of Mathematics & Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan
2017

Probability distribution

Gumble distribution.

Selection of best-fit probability distribution for at-site flood frequency analysis (FFA) in Pakistan

By

Azmat Ali

Supervised by

Prof. Dr. Ishfaq Ahmed

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

2017

Selection of best-fit probability distribution for at-site flood frequency analysis (FFA) in Pakistan

By

Azmat Ali

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

Supervised by

Prof. Dr. Ishfaq Ahmed

Department of Mathematics &Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan
2017

Certificate

Selection of Best-Fit Probability Distribution for at-Site Flood Frequency Analysis

By

Azmat Ali

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. ______________________
Dr. Bahrawar Jan
External Examiner

2. ______________________
Dr. Farzana Akhtar Abbasi
Internal Examiner

3. ______________________
Dr. Ishfaq Ahmad
Supervisor

4. ______________________
Prof. Dr. Muhammad Sajid, T.I
Chairman

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

2017

Dedication

To Almighty Allah,

For blessing me the courage and spirit.

To my family,

For the endless support and patience.

To my Teachers,

For the constant source of Knowledge.

To my friends,

The ones that are close and the ones that are far.

Forwarding Sheet by Research Supervisor

The thesis entitled **"Selection of "best-fit probability distribution for at-site flood frequency analysis (FFA) in Pakistan"** submitted by **Azmat Ali** (Registration # 32-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: 28/04/2017

Supervisor

Prof. Dr. Ishfaq Ahmed

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

ACKNOWLEDGEMENT

In the name of Allah, We praise Him, seek His help and ask for His forgiveness. Whoever Allah guides none can misguide, and whoever He allows to fall astray, none can guide them aright. We bear witness that there is none worthy of worship but Allah Alone, and we bear witness that Muhammad (saws) is His slave-servant and the seal of His Messengers.

I express my gratitude to my supervisor **Prof. Dr. Ishfaq ahmed** for his regardless and inspirational efforts and moral support throughout my research carrier as well as the **Prof. Dr. Zahid Iqbal, Prof. Dr. Akbar Awan and Prof. Dr. Irshad Ahmad Arshad** for their sound advices and lots of good ideas were very helpful to me. May ALLAH bless them with all kinds of happiness and success in their life and life after here. I am also thankful to Chairman Department of Mathematics and Statistics **Dr. Arshad Zia** for providing such necessary facilities. I express my sincere appreciation to all the helpful staff of the Mathematics Department IIUI.

An expedition is easier when you work together; interdependence is certainly more valuable than independence. University is an enjoyable place to work and even more, especially when friends are around. My friends are source of laughter, joy and support. I am indebted to my friends and class mates for providing a stimulating and fun filled environment. My thanks go in particular to **Ubaid Ali, Rameez Malik , Shahid Rauf, Muhammad Fawad Ali,** especially MS class fellows being around and sharing several good times during my stay in the university. Thank you does not seem sufficient but it is said with appreciation and respect to all of them for their support, encouragement, care, understanding and precious friendship. 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.

Last, but not least my deepest gratitude goes to my beloved parents, brothers, sister and my family for their endless love, prayers and encouragement, for which my expression of thanks does not sufficient. I offer my great regards and blessings to everybody who was indirectly important to the successful realization of thesis; your kindness means a lot to me.

Azmat Ali

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 **Prof. Dr. Ishfaq Ahmed.** 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: ____________________

Azmat Ali
MS (Statistics)
Reg. No 32-FBAS/MSST/F13
Department of Mathematics and Statistics,
Faculty of Basic and Applied Sciences,
International Islamic University Islamabad, Pakistan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abbreviations

Words	Abbreviation
Annual Maximum peak	AMP
Flood Frequency Analysis	FFA
Normal distribution	NOR
Log normal distribution	LNO
Exponential distribution	EXP
Gumble distribution	GUM
Gen. Extreme value Dist:	GEV
Goodness of fit	GOF
Percentage	%age
Method of Linear moment	MLM
Method of likelihood estimation	MLE

Method of moment	MME
Akaike Information Criterion	AIC
Bayesian Information Criterion	BIC
Anderson Darling Criterion	ADC
Percentage	%age
Method of Linear moment	MLM
Method of likelihood estimation	MLE

Abstract

The current study is based on AMP flows data of 7 sites of river indus in Pakistan. the sites are Tarbela, Kalabagh, Chashma, Taunsa, Kotrri, Guddo and Sukker. The flood data is processed on the bases of three parameter estimation methods (MLM, MME and MLE) under three goodness-of-fit criteria (ADC,AIC and BIC). The basic and fundamental purpose of the study is to establish a probability distribution that will provide best prediction about upcoming extreme events. Flood is one of the most devastating factor that damages human life and property badly. The prevention of floods is not possible but some safety can be managed for the reduction of damages.

In the study five probability distributions are involved which are GEV, GUM, EXP, NOR and LNO. The suitability of the AMP flows for independence, stationary and homogeneity is tested with the help of Lag-one-correlation, Ljung-Box and Mann-Whitney U test respectively, which revealed statistically insignificant results i.e. confirmed the suitability.

The study revealed the of the 5 applied distributions, the most appropriate one to be used for prediction is the LNO distribution. The LNO is best for 50.8% of the sites on overall basis. Under AIC, BIC and ADC the LNO is best for 47.6%, 52.4% and 52.4% respectively. The estimation method MME and MLE decided LNO most suitable distribution for 71.4% and 61.9% of the sites respectively.

The best estimation method under ADC is MLM which is proved best for 85.7% of the sites, whereas the combined results showed that MLM is best for 64.3% of the sites. the second method that is followed by MLM is MLE and the distribution followed by LNO is GEV.

CHAPTER 1

INTRODUCTION

There are so many factors which effect the environment and life structure of a locality. Some of them are directly related and some are indirectly concerned. Among these some are very dangerous which harm humanity as well as other creations. The prevention of these events is impossible but some arrangements can be made for protection. Floods, droughts, storms etc. are natural disasters, which devastate the life structure directly. Among them the most devastated and alarming disaster is flood, which damage the lives and property of millions of dollars on yearly basis.

The weather extremes in Pakistan include high and low temperatures, heaviest rainfall and flooding. The highest temperature that has been recorded in Pakistan was 53.5 degree centigrade in Mohenjo-Daro on 26th May 2010, which was not only the extreme temperature in the country , but also in the continent Asia as well and the 4rth highest temperature even measured on the earth. The heaviest rainfall of 620 mm on 24rth July 2001 was recorded in the capital (Islamabad) in the history of 100 years.

Pakistan has been suffering from the natural floods from so many decades, which are almost repeated in the season of monsoon when the rainfall is on its extreme. The estimates of financial loss for last six decades recorded by Federal Flood Commission (FFC) are in US$ 39 billion.

The Federal Flood Commission (FFC) declared the flood of 1950 as the deadliest disaster that caused 21,920 casualties, flooded 10,000 villages and affected 17,920 square km of area. The 2nd flood declared by (FFA) was that of 2010, which caused

1,985 casualties, affected twenty-one million people and 160,000 square km area with the destruction of 17,533 villages. The (FFC) reports showed that during 2010 to 2012 floods, the total human casualties were 3,072 with economic loss of US$ 16 billion. The natural disaster management authority (NDMA) reports revealed that 69 people killed in monsoon flood of 2013 of which 22 in Sind, 18 in Baluchistan, 15 in Punjab and 14 in Khyber Pakhtunkhwa.

One of the most notable points in the assessment of National Disaster Management Authority (NDMA) is that the damage caused by floods in the last five years is much greater than the damage caused in the past six decades. As the disaster management system (DMS) in Pakistan is not properly supervised, as a result the country suffers a loss of $ 800 billion on yearly basis. The ruins of these disasters can be faced in each and every field of life. Floods affect various sectors like, education, health, irrigation, water supply, communication, transportation etc. Of them the most effected sector is agricultural sector which is depended with 70% of the population. Therefore, it is of great importance to reduce the damages by designing the level of flood through fitting a best probability distribution.

In the September 2014 flood 360 people were killed, 646 injured, affected almost 2,523,681 people, destroyed 4,065 villages, damaged 2,416,558 acres crops and 8,957 cattle were washed away.

In the first week of Jun 2015, the heavy rainfall started which caused flood in the district of Khuzdar in the Baluchistan province. According to the early report, 9 were killed in the flood. The heavy rainfall, explosion from glacial lakes and snow melting again caused ruined of precious human lives and property and affected the economy of the concerned population. The Pakistan national disaster management authority (NDMA) reported on September 2015 the statistics of the flood damage. Accordingly 1,572,191 people were affected in about 4,111 villages with 238 reported casualties and 232 injuries. The crops in KPK province were affected badly in the consequences.

In the year 2016 two weeks of heavy rains caused the flood which was reported by national disaster management authority (NDMA). According to the report, nearly 121 people died, 127 injured, and about 857 buildings damaged. The rain intensity recorded between 12 and 13 March on the average was 51 mm and between 16 and 17 March it was 57 mm on the average.

The deadliest flood announce in the flood Wikipedia is that of China happened in 1931which caused 1,000,000-4,000,000 deaths. The list also ranked the flood that occurred in 1950 in Pakistan as 44, which caused 2,910 casualties. The second deadliest flood considered by the list is ranked on 65^{th} position, which caused 1,600-2,000 deaths.

In Australia in the time period of four years i.e. from 2010 to 2014, the floods damaged the economy of the country by billions of dollars and disturbed the living condition of urban and rural populations (Ayesha et al 2013). Floods directly affect the life of human and other living creations on the earth.

Flood frequency analysis is considered to be one of the most exciting problem in hydrology, also one of them which is filled with disagreements (Bobee et al. _ 1993). One of the wide research about flood frequency analysis is that the selection of suitable probability distribution and parallel parameter estimation method is one of the significant step in the construction of flood frequency analysis (Cunanne 1973; NERC 1975; Stedinger 1980; Stedinger et al 1992; vogel et al. 1993;Bobee et al. 1993

Marrkiewicz et al. 2006; Reibatet et al. 2007; Meshgi and Khalili 2009a,b; Femandes et al. 2010).

One of the significant task is estimating extreme values of flood for engineers and other responsible experts for the management of floodplain (Pearson & Davies, 1997). The selection of an appropriate probability distribution and sampling technique are considered great challenge in flood frequency analysis and other extreme events in the field of hydrology (B.K. Nagy et al 2017).

A lot of distributions are tested on the basis of various parameter estimation procedure and suitable distributions and methods are recommended all over the world as briefly summarized by Cananne (1989).

For the reduction of damage caused by floods and life protection, floods are modeled for various sites of the country under consideration to find flood estimates related with return periods of interest, this is termed as design of flood. There exists numerous methods to design flood estimation, of them the most direct and simple method is at-site flood frequency analysis (Ayesha et al 2013).

For flood frequency analysis Statistical distributions are usually selected on the basis of eminent tests provided by Statisticians or by graphs methods, and the convenience plays a vital role in the choice (Bobee et al._ 1993).

probability distributions that are used commonly in modeling the annual maximum flood frequency series include Extreme value type 1 (EV1), Generalized extreme value (GEV), Normal (NORM), Exponential (EXP), Lognormal(LN) ,Pearson type 3 (P3), Log Pearson type 3 (LP 3), Gamma (GAM), Weibull (WEI), Wakeby (WAK) and Generalized Pareto (GPA) (Cunnane 1989; Bobee et al. 1993).

A survey based on 54 agencies in about 28 countries, by Cunnane revealed that LN,EV1, EV2, LP3, GEV, and P3 were proposed for 8,10, 3, 7, 2 and 7 countries, respectively. The study involves five distributions namely, Normal (NORM), Lognormal (LN), Exponential (EXP), Gumbel (GUM) and Generalized Extreme Value (GEV) according to their importance. The parameter estimation methods commonly used are, method of maximum likelihood, method of moments, method of L moments, method of LQ moments and Bayesian method.

The underlying research involves the three estimation methods, namely, method of moments (MOM), method of maximum likelihood (MLE), and method of L-moments (MLM) with three goodness-of-fit (GOF) tests, namely Akaike Information Criterion (AIC), Basian Information Criterion (BIC) and Anderson- Darling Criterion (ADC).

The MLE involves the estimation of parameter value for which the probability distribution of sample data is maximized. According to Statisticians the MOM is less robust method as compare to MLE in most of the cases. MLE is considered providing estimators with better Statistical properties.

There are a number of limitations on MOM, one of them is that the product moments of data series are evenly influenced by commonly small values which are not consequently responsible for floods, as compare to huge amounts. Also the moments of high order namely coefficient of skewness and coefficient of variation are frequently influenced by extremes. On the other side, it is claimed the L moments are comparatively less influenced by extremes data points in the series (Hosking 1990).The basic and vital objective that the flood frequency analysis investigates the estimates of return periods related with the flood magnitude. One of the requirements of at-site flood frequency analysis is that the recorded stream flow data must be reasonably long period.

As it has experience on numerous occasions that the length records of the available gauged streams are not with accordance to the requirement, so extrapolation is performed. To select an appropriate probability distribution is extremely important, the wrong selection of probability distribution will cause under or over estimation, which will seriously devastate the results and the decisions on this basis must be misleading (Ayesha et al 2013).

McMahon and Srikanthan (1981) compared a number of probability distributions by using the moment ratio diagram on the data of 172 catchments in Australia. On the basis of research the result conducted was that LP3 distribution was most appropriate for the mentioned catchments in Australia. On basis of findings of the study, the LP3 distribution was recommended for forecasting the future flood indications by Rainfall and Runoff department of Australia in 1987.

Most of the statisticians give preference to MLE over MOM while estimating the distribution parameters (Bickel and Doksum 1977 ;Martins and stedinger 2000). In the past various probability distributions are compared, which we discuss below. Two extreme value distributions GPA and GEV distributions are fitted by Madsen et al. (1997) using MOM, MLE and MLM (method of L moments).

After detailed numerical work, the experiment revealed that for slightly shape parameter the MOM was not too bad with the GEV distribution, whereas, the MLE performed better for comparatively high positive shape parameter for GPA distribution. A model selection criterion presented by Laio et al. (2009), in which they used the three goodness of fit criterion, namely Basian Information Criterion (BIC), Akaike Information Criterion (AIC) and Anderson-Darling Criterion (ADC). After a heavy numerical struggle they concluded that the decisions made by these criterions in the recognition of correct distribution bears some variation from on case to another, however, it is recommended that these criterions are more effective for distributions with two parameters.

Table 1.1 List of floods ranked on the basis of death tolerance

Disaster type	Date	Death Toll
Flood	June 1977	10,354
Flood	1950	2,900
Flood	July and August 2010	1,645
Flood	September 1992	1,334
Flood	March 3, 1998	1,000

1.1 Definition and types of floods

The term flood in meant as the presence of water on land surface which extents a level that affect normal activities of life. Flood caused by the excess of rivers are termed as river floods and those arises from heavy rainwater for short interval are known flash floods. The floods arise from unusual inflow of marine water onto land is known as ocean floods. The flash and river floods are conclusion of irregularly high precipitation over a relative short time period. The floods caused due to rapidly melting of snow is called spring floods and that resulted by heavy rainfall in rainy seasons are termed as monsoon floods.

Seasonal floods are supposed to be very essential for agricultural productions as they are major source of nutrients. The deficiency of seasonal floods leads adversity in some areas of the world. If floods are properly managed, it will not be considered as a source of disaster but on the other hand it will bring a lot of prosperity in the country or region, so flood modeling also aim it.

According to the estimates published in the report of Pakistan National Disaster Management Authority (NDMA), more than 176,000,000 people affected by floods. Flood damaged ,bridges, hospitals, roads, buildings i.e. complete infrastructure of the locality, as a result millions of people became displaced, hundreds of thousands of livestock are killed, complete agriculture sector is destroyed and more than 20% of the country is inundated. Such type of disaster must be facilitated by the international community in the form of supplement of humanitarian relief.

Statement of the problem

One of the most notable points in the assessment of National Disaster Management Authority (NDMA) is that the damage caused by floods in the last five years is much greater than the damage caused in the past six decades. As the disaster management system (DMS) in Pakistan is not properly supervised, as a result the country suffers a loss of $ 800 billion on yearly basis. The ruins of these disasters can be faced in each and every field of life. Floods affect various sectors like, education, health, irrigation, water supply, communication, transportation etc. Of them the most effected sector is agricultural sector which is depended with 70% of the population. Therefore, it is of great importance to reduce the damages by designing the level of flood through fitting a best probability distribution.

CHAPTER 2

LITERATURE REVIEW

Tasker (1987) made a study of frequency analysis for round about 20 sites in Virginia using various estimation methods to 7-days minimum flow series. Amongst a number of distributions it was decided that Log Pearson type-111 (LP3) and Weibull type-111 (W3) are considered to be the suitable distributions for the above discussed data.

Karim and Chowdhury (1995) experienced a detailed study over FFA for 31 gauging sites in Bangladesh. The estimation method being used is L-moments and the distributions involved are GEV, LP3, Gumbel and Lognormal (LN). The suitability of the distribution s were compared by Root Mean Deviation Error (RMDE), PPCC and L-moments ratio diagram. On the basis of goodness of fit results it was established that the GEV is the best distribution for annual peak discharges in Bangladesh.

Onoz and Bayazit (1999) studied a number of probability distributions to the data based on low flows of about 16 European rivers. The performance of the distributions was compared by using PPCC test, which recommended the GEV as a best probability distribution the mentioned rivers.

W.G. Strupczewski et al (2001) discussed the statistical parametric techniques based on at-site flood frequency modeling to explore the time trend. The exploration involved the identification of an appropriate probability distribution and invention of a trend software. The parameter estimation method being used is maximum likelihood estimation method (MLE) and the performance of fit is decided by use of akaike information criterion (AIC). In the study a set of 56 distributions were compared while using discharge time series (1921-1990) of Annual peak flow from Vistula in Pulawy. The results revealed that Pearson type III is the most appropriate distribution.

M.D. Zalina et al (2002) discussed the maximum rain fall data from 17 gauging stations of Peninsular in Malaysia. The study was conducted on the basis of eight renowned candidate distributions with the goodness-of-test of these selected distributions is performed by commonly practiced tests say probability plot correlation coefficient (PPCC), relative root mean square error (RRMSE), root mean square error (RMSE) and maximum absolute deviation (MAD). The distributions involved in the discussion were Generalized Extreme Value (GEV), Generalized Normal (GN), Generalized Pareto (GPA), Gamma, Gumbel, Log Pearson type III, Pearson type III and Wakeby. The estimation of model parameters was performed by method of L-moment (MLM) with rainfall length varies in the range of 23-28 years. The goodness-of-fit tests concluded that GEV distribution is most valid distribution to be used for forecasting of maximum rainfall in Malaysia.

Bhattara (2005) compared simple L-moments, Partial L-moments and LH-moments by the performance of FFA using the Ireland river flow data. The GEV was fitted to AMP flow and the results were examined for comparison, which revealed that among the three estimation methods LH-moment and Partial L-moments were respectively considered to be the best estimation methods for FFA using the data.

Yurekli et al (2005) worked on FA for low flow data over a stream of Cekerek Bain in Turkey. The fundamental aim of the study was establish a probability distribution that is the best one for frequency analysis of seven days low flow data of three gauging stations of the Stream mentioned in the above lines. A number of distributions used in the study including GLO, LN, Logistic (L3), GPA, GEV, Extreme Value Type 1 (EV1) and LP3. The parameters of the distributions were estimated by the renowned estimation method named, method of L-moments. For the comparison of performance of probability distributions, MADI and MSDI goodness of fit tests were used. On the basis of these tests it was decided that the most suitable distribution for the stations of the stream is GPA.

Abida and Ellouze (2008) made a study over 42 stations of various zones in Tunisia, applying FFA based on renowned parameter estimation method, named as L-moment to the data of AMP flows. The distributions involved in the study are, GNO, GPA, GLO, GEV, and P3 . The results highlighted that GNO distribution was the best fitted to Northern Tunisia whereas GEV distribution was the most suitable for central/southern Tunisia.

Zin et al (2009) applied two parameter estimation methods i.e. LQ-moment and L-moment study the analysis of annual maximum rainfall data in Peninsular Malaysia for about 50 gauging stations. The distributions with three parameters, namely GEV, GNO, GPA, GLO and P3 were fitted to the data. The performance of the distributions was compared by the application of goodness of fit test. On the basis of announced results the GLO is stated as best for most of the stations and thus it is decided the best distribution. Moreover on the average L-moment was supposed better as compare to LQ-moment.

Deka et al (2009) applied the two estimation methods namely, L-moment and LQ-moment for the analysis of annual maximum daily rainfall data collected from nine stations in the locality of East India. The commonly used extreme value distributions were GLO, LN, GPA, GEV and P3. The comparison for best distribution was made by the use of goodness-of-fit tests i.e. Relative Root Mean Square (RRMS), Relative Mean Absolute Error (RMAE) and Probability Plot Correlation Coefficient (PPCC). The evaluation of the distributions was performed by the L-moment ratio diagram. The study finalized that for the annual maximum daily rain fall data the best distribution to be used is the Generalized Logistic distribution (GLO).

Shabri and Ariff (2009) experienced flood frequency analysis to maximum daily rainfall data of various stations of Kula Lumpur and Selangor on the basis of L-moment parameter estimation method. In the study distributions with two parameters, namely Normal (N2) and Lognormal (LN2), and distributions with three parameters, namely GLO, GEV, EV1, GPA, LN and LP3 were used. The performance of the above mentioned distributions was checked by applying the goodness-of-fit tests i.e. MADI and MSDI also the evaluation of the distributions was performed by the usage of L-moment ratio diagram. The study finalized the GLO distribution as the most suitable distribution to be used the above maximum daily rainfall data in the areas of Malaysia.

Gubareva and Gartsman (2010) applied MLE and L-moment estimation methods to estimate parameters of distributions used commonly in meteorology and hydrology, i.e. GEV, P3, GPA and LN. The study showed that as compare to method of likelihood estimation (MLE), L-moment results are more stable.

Ahmed et al (2011) applied FFA to the AMP flows data for Negeri Sembilan stations in Malaysia. The study consists of three extreme value distributions, namely GLO, GPA, and GEV. The estimation methods used were L-moments and TL-moments trimming in the interval (0,1). The distributions are evaluated for the performance by goodness-of-fit test MADI, L-moment ratios and TL-moment ratios. The results indicated that for both estimation methods the GLO distribution is considered best for AMP flows.

Ayesha et al (2013) applied three estimation methods namely method of moments (MOM), method of likelihood estimation (MLE) and method of L-moment (MLM), under four goodness-of-fit tests namely Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson Darling test and kolmogorov-Smirnov test for the aim to know about the distribution which best fit the data of Australian maximum flood. Further the L-moment ratio diagrams were used to facilitate the decision about best fit for the mentioned data. The results suggested the LP3, GEV and GPA were identified as the top three best distributions that fit the data of Australian maximum flood.

Salarpour et al (2013) experienced FFA using the data of annual peak flows across Johor river south station of Rantau Panjang in Malaysia. The study consisted of five distributions that are, GPA, Pearson, Beta, GEV angd G. The distributions were evaluated by Anderson Darling test and Chi-square test goodness-of-fit. The results indicated that on the basis of AD test GPA was best whereas Chi- square explains that GEV distribution was the best.

Galoie et al (2013) applied L-moment estimation method for the purpose to obtain the most appropriate probability distribution that best forecast the rainfall data of Schoeckelbach basin in Australia. The basic extreme value distributions used in the study are, LP3, GU, GEV and three parameter LN distributions. The performance of the distributions was evaluated by the usual goodness-of-fit tests such as C-S, K-S and RMSE. These tests recommended the GU distribution to be used for forecasting.

Izinyon and Ehiorobo (2014) established the FFA for the purpose of analysis AMP data in Benin Owena river in Nigeria on the basis of estimation method L-moment. The study was conducted on the basis of the extreme value distributions say GEV, GLO and GPA. The suitability of the distribution was evaluated by fundamental goodness-of-fit tests i.e. RMAE, PPCC and RRMSE, also the L-moment ratio diagram was used for the confirmation of best fitted distribution. On the basis of goodness-of-fit tests and diagram the decision was made that GPA must be used for further forecasting.

Gocic et al (2014) applied L-moment estimation method for the analysis of precipitation data consists of 29 stations in Serbia. The study discussed three main extreme value distributions, namely GEV, GPA and GLO under the goodness-of-fit criteria RRMSE, RMAE and PPCC as well as the L-moment ratio diagrams were established for the confirmation of best fitted distribution. The final results revealed that that GEV distribution is the best and must be applied in forecasting future happenings.

Fawad (2015) worked on estimation of parameters to establish the best fitted distribution using Annual Maximum Peak (AMP) flow data for eighteen sites in Pakistan. The study involve the three main and commonly used distributions for Food Frequency Analysis (FFA), namely, Generalized Pareto distribution (GPA), Generalized Logistic distribution (GLO) and Generalizes Extreme Value distribution (GEV). The three goodness-of-fit tests applied are Mean Absolute Deviation Index (MADI), Probability Plot Correlation Coefficient (PPCC) and Relative Root Mean Square Error (RRMSE). Most of the sites showed positive indication for Generalized Pareto Distribution (GPA) and finally it is decided that GPA is the best distribution for Annual Maximum Peak flow in Pakistan.

CHAPTER 3

METERIAL AND METHODOLOGY

3.1 Study area and Source of data

The data used in the study is of maximum peak flow based on 7 sites of river Indus in Pakistan. The flood data of the seven sites on annual basis is included in study. The data of the 7 sites has been obtained from a department of hydrology Water and Power Development Authority (WAPDA) in Federal Flood Commission (FFC) Islamabad. The annual maximum peak (AMP) flow flood data for specified sites have been documented in the season of monsoon i.e. July-September.

The Pakistan rivers originate from two eminent mountains i.e. Karakorum and Himalaya. Most of the rivers pass through the province of Punjab. In pertain 'punj' is used for the meaning of five and 'aab' is used for water, so Punjab means five rivers. The Pakistan irrigation system is considered one of the widest system of the world. The irrigation system plays boundless role in agricultural sector of a country. In Pakistan the rivers are basic and vital source of watering.

The river Indus originates from a region of Himalayan that names Tibetan plateau which is near to Lake Mansarovar in China. Indus river is the major river in Asia, that flows through out Pakistan and is the longest Pakistani river. The length of river Indus is approximately 3180 kilometers. Balram, Gilgit, Tanubal, Astor, Kabul and Zanskar rivers are the branches of river Indus. Indus river comprises of 3 reservoirs, 16 barrages, 2 head works, 12 interlink canals, 2 siphons, 44 canal system: 2 in Baluchistan, 5 in KPK, 14 in Sind and 23 in Punjab .

3.2 The Assumptions

The Flood Frequency Analysis (FFA) has a number of necessary assumptions. The basic purpose of the assumptions is to examine the independence, stationary and Homogeneity of the data. Different types of data specially AMP flows, droughts, rainfalls etc. are applied to check these assumptions. The tests used in the underlying study are: Mann-Whitney U test, Ljung-box test and Lag-one correlation test discussed below.

3.2.1 Mann-Whitney U test

The Mann-Whitney U test is basically a non-parametric test which we use to make a comparison between two population means for the conclusion if they belong to the same parent population? It is alternative of independent samples t-test and applied to test the equality of means of two populations. Point to be noted that it is applied when sample sizes are same and the populations' medians are tested. The Mann-Whitney U test is applied for ordinal data situation. The Mann-Whitney U test, Kendall's and Wilcoxon rank sum are identical in tie case and also similar in such situation to chi-square test.

Mann-Whitney U test is basically non-parametric, thus don't depend on the distribution assumption. The particular assumptions of the test are as mentioned below:

- The selected sample must be random.
- The sample data must be of ordinal nature.
- There must be independence within the samples.

The calculation is based on the relation

$$U = n_1\, n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=n_1+1}^{n_2} R_i$$

Where

n_1 = size of sample one

n_2 = size of sample two

R_i = rank of data points in sample

U = the Mann-Whitney test statistic

This test is applied in every field particularly the frequent use is in business, medical science, nursing and psychology. In medical the test observe, if the two medicines used have same effect or not? The test play great role in business to see the preference of people with respect to brand, location, satisfaction etc.

3.2.2 Ljung-box test

The Ljung-box test (1978) is named for George E.P. Box and Greta M. Ljung, is a statistical test under the hypothesis that there is no autocorrelation in the data. The overall randomness for lags is tested instead, to test randomness at distinct lag. The test is also called the Ljung-Box Q test which is closely related to Box-Pierce test. This test has a wide application in time series analysis and econometrics. The null and alternative hypotheses under Ljung-Box test are formulated as:

H_0: The lake of fit is not exhibited by the model.

H_1: The lake of fit is exhibited by the model.

The test statistic for the test is defined by:

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

Where

k = number of lags in series, m = number of tested lags, n = data size and $\hat{r}^2$stands for autocorrelation estimated value for the series corresponding to k^{th} lag.

The rejection of the null hypothesis under the test means that there is a significant lake in fitting the distribution. The test may be rejected if $Q > \chi^2_{1-\alpha,h}$, Where

$\chi^2_{1-\alpha,h}$ stands for χ^2 tabulated value under α level of significance and h degree of freedom. Here, as the test deals with residuals, therefore for the model to be estimated, the value of degree of freedom "h" must be accounted for as:

h = m –p – q, where p and q are the ARMA(p,q) model parameters that fit the data.

3.2.3 Lag-one autocorrelation

The autocorrelation is being used for the two basic purposes:

- To identify non-randomness in the data.
- To detect a suitable model for time series for non-random data.

For measurements, $y_1, y_2, \dots y_n$ at time periods $x_1, x_2, \dots, x_n$, the lag k autocorrelation function is given by:

$$r_k = \frac{\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

For the lag-one autocorrelation put k = 1, we get

$$r_1 = \frac{\sum_{i=1}^{n-1}(y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

On the basis of the assumption that the observations are equally spaced, therefore the time variable 'x' is not involved in the formulation. Autocorrelation is correlation coefficient between the values of the variable at time periods x_i and x_{i+k} instead of corresponding values of two different variahles at same time period.

The autocorrelation used to identify the non-randomness, on the basis of first lag, is called lag-one autocorrelation. The autocorrelation used to identify an appropriate model for time series data, is usually based on many lags.

The questions to be answered by the use of autocorrelation function are given below:

- Was the sample data is generated by a random process?
- Would a time series model or non-linear model be more suitable model for the

data as compare to simple constant and error model?

For the statistical process with one variable to be in control, randomness is the key assumption. When the assumption of randomness, constant scale and location and fixed distribution are reasonable, then the model for univariate can be used as:

$y_i = a_0 + e_i$, where e_i is random error.

3.3 Estimation Methods

3.3.1 Method of Moments

3.3.1.1 Introduction

Let $X_1, X_2, \ldots$ are independent random variahles with common distribution having mean μ_X. As number of observations increase to some extent, the sample means approach to mean of parent distribution. i.e.

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mu_X \text{ as } n \rightarrow \infty \qquad \text{i)}$$

For the explanation of the procedure of method of moments, distribution with one parameter is discussed first. Let m_1, m_2, ... are independently selected random variables with corresponding probability density $f_M(m/\theta)$ associated to unknown parameter θ. The mean of m_i. μ_m, is a function $k(\theta)$ of θ i.e.

$$\mu_m = \int_{-\infty}^{\infty} m f_M(m/\theta) dm = k(\theta) \qquad \text{ii)}$$

The large number law states that

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mu_X \text{ as } n \rightarrow \infty \qquad \text{iii)}$$

For large sample size, i.e. $n \rightarrow \infty$, the mean of distribution

$$\bar{M} \approx k(\theta)$$

The relation can be given for estimator $\hat{\theta}$ as

$$\bar{M} = k(\hat{\theta}) \qquad \text{iv)}$$

And then we solve for $\hat{\theta}$.

3.3.1.2 The procedure

More often, for independent random variates X_1, X_2,....... Selected with accordance to probability distribution with parameter θ and t as a function of real values, if $k(\theta)= E_{\theta m}(X_i)$, then

$$\frac{1}{n}\sum_{i=1}^{n} m(X_i) \rightarrow k(\theta) \quad \text{as} \quad n \rightarrow \infty \qquad \text{v)}$$

The moment method results from the choice $m(x) = x^m$. We can write

$$\mu_m = EX^m = k_m(\theta) \quad \text{vi)}$$

For the m^{th} moment.

The estimation procedure follows the steps for linking sample moments and population moments.

Step 1. Consider the function with d parameters, then we compute the above functions k_m given in equation vi) for first d moments,

$$\mu_1 = k_1(\theta_1, \theta_2, \dots, \theta_d), \quad \mu_2 = k_2(\theta_1, \theta_2, \dots, \theta_d), \dots, \quad \mu_d = k_d(\theta_1, \theta_2, \dots, \theta_d) \quad \text{vii)}$$

And thus d equations are obtained in d unknowns.

Step 2. then these equations are solved simultaneously for parameters as function of d moments.

$$\theta_1 = g_1(\mu_1, \mu_2, \dots, \mu_d), \quad \theta_2 = g_2(\mu_1, \mu_2, \dots, \mu_d), \dots, \quad \theta_d = g_d(\mu_1, \mu_2, \dots, \mu_d). \quad \text{viii)}$$

Step 3. On the basis of sample data X = $(x_1, x_2, \dots, x_d)$ the first d sample moments are calculated, as given,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \dots, \quad \overline{x^d} = \frac{1}{n}\sum_{i=1}^{n} x_i^d \quad \text{ix)}$$

According to law of large numbers we know that sample moments approximates to population moments i.e. $\mu_m \approx \overline{x^m}$ for m = 1, 2, …, d.

Step 4. In this step the population moments μ_m are replaced by sample moments $\overline{x^m}$,

Then using the solution given in viii) provide us the relation for moment estimators $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_d)$. For the above mentioned data X, the estimates are

$$\hat{\theta}_1(X) = g_1(\bar{x}, \overline{x^2}, \dots, \overline{x^d}), \hat{\theta}_2(X) = g_2(\bar{x}, \overline{x^2}, \dots, \overline{x^d}), \dots, \hat{\theta}_d(X) = g_d(\bar{x}, \overline{x^2}, \dots, \overline{x^d}).$$

3.3.2 Maximum Likelihood Estimation

3.3.2.1 The Likelihood function

Consider an iid sample $x_1, x_2, \dots, x_n$ having a density function f(x_i ; θ), where θ is a vector of parameters of order (k ×1). For example, if X~N(μ, σ^2) , then f(x_i ; θ) = $(2\pi\sigma^2)^{\frac{-1}{2}} \exp\left(-\frac{1}{2\sigma^2}\right)(x_1 - \mu)^2)$ $and\ \theta = (\mu, \sigma^2)$. By independence, the joined function is defined as the product of marginal densities i.e.

$$f(x_1, x_2, \dots, x_n ; \theta) = f(x_1 ; \theta), f(x_2 ; \theta), \dots, f(x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

The joint function is n dimensional function of the data $x_1, x_2, \dots, x_n$ with parametric vector θ. The density of the joint distributions must obey the following:

$$f(x_1, x_2, \dots, x_n ; \theta) \geq 0$$

$$\int \dots \int f(x_1, x_2, \dots, x_n ; \theta) d_{x_1}, \dots, d_{x_n} = 1$$

The joint density as a function of parameter θ is defined as the likelihood function.

$$L(\theta / x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

One thing that is of most importance and must be noted that the likelihood function is basically a function of distribution parameters rather than data. Although, it can never be negative, but not a proper pdf and thus

$$\int \dots \int L(\theta / x_1, x_2, \dots, x_n) \, d_{\theta_1}, \dots, d_{\theta_n} \neq 1$$

For the simplicity of notation let we suppose X= ($x_1, x_2, \dots, x_n$), where $x_1, x_2, \dots, x_n$ are sample observations. Now the joint density in simplified form can be articulated as $f(X;\theta)$ or L(θ / X).

3.3.3 The method of L-moment

L-moments defined for first time by Hosking in 1990, which are defined as expectations of linear arrangement of orderly statistic. These moments can be calculated for all random variables having finite mean. Consider a random sample $X_1, \ldots, X_r$ of size r with distribution function F(X) and corresponding quantile function X(F). Again let we consider $X_{1:r} \leq X_{2:r} \leq \cdots \leq X_{r:r}$ as a random sample of ordered statistic. The explained r^th population moment of random sample by Hosking is given by:

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

Here λ_1 is defined as measure of location, λ_2 is termed as measure of scale, τ is known as measure of Linear- coefficient of variation (L-CV), τ_3 is Linear- Skewness (L-Skewness) and τ_4 is the Linear-Kurtosis (L-Kurtosis).

3.3.3.1 Estimation of L-moments

Usually L-moments are estimated on the basis of random sample obtained from any well-known probability distribution. λ_r means the r^th population moment, which is the mixture of expectations of ordered statistics based on sample of magnitude r. consider a random sample of size n such that $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ is orderly arrangement of the sample data, then The equation of L-moments indicates the linear combination of expected value of ordered statistic $X_{r-k:r}$. The first few L-moments are:

$$\lambda_1 = E(X_{1;1})$$

$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$$

$$\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3})$$

$$\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} + X_{1:4})$$

The ratio of L-moments is given by:

$$\tau = {}^{\lambda_2}/_{\lambda_1}$$

$$\tau_3 = {}^{\lambda_3}/_{\lambda_2}$$

$$\tau_4 = {}^{\lambda_4}/_{\lambda_2}$$

According to Asquith (2007) the rth L-moment for sample data is defined as:

$$l_{r=}\frac{1}{r}\sum_{i=1}^{n}\left[\sum_{j=0}^{r-1}\frac{(-1)^j\binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-1}{j}}{\binom{n}{r}}\right]X_{i:n} \qquad r = 1, 2 \ldots$$

By Wang (1996) the beginning four L-moments for sample data are mentioned as:

$$l_1 = \sum_{i=1}^{n}\left[\frac{x_{i:n}}{\binom{n}{1}}\right]$$

$$l_2 = \frac{1}{2}\sum_{i=1}^{n}\left[\frac{\binom{i-1}{1} - \binom{n-i}{1}}{\binom{n}{2}}\right]x_{i:n}$$

$$l_3 = \frac{1}{3}\sum_{i=1}^{n}\left[\frac{\binom{i-1}{2} - 2\binom{i-1}{1}\binom{n-i}{1} + \binom{n-i}{2}}{\binom{n}{3}}\right]x_{i:n}$$

$$l_4 = \frac{1}{4}\sum_{i=1}^{n}\left[\frac{\binom{i-1}{3} - 3\binom{i-1}{2}\binom{n-i}{1} + 3\binom{i-1}{1}\binom{n-i}{2} - \binom{n-i}{3}}{\binom{n}{4}}\right]x_{i:n}$$

The L-ratios for sample data are discussed by

$$t = \frac{\ell_2}{\ell_1}$$

$$t_3 = \frac{\ell_3}{\ell_2}$$

$$t_4 = \frac{\ell_4}{\ell_2}$$

Where l_1, l_2, t, t_3, and t_4 respectively denote the location, scale, L-coefficient of variation (LCV), L-skewess and L-kurtosis. In statistics these measures are considered very useful. These are measures involved in the classification of parent probability distribution.

3.4 The main objectives of the research

- To find the distribution that best fit the data of AMP flows in the class of five distributions by the application of ADC, AIC and BIC.
- To compare the three estimation methods i.e. MLE, MLM and MME on the basis of suitability for the AMP flows while performing FFA.
- To recognize the site that will receive the maximum flows in forthcoming.

3.5 The goodness-of-fit tests

The selection of a particular probability distribution involves number of factors. Of them some are very important such as the method which is used for comparison of various distribution performance, the estimation method which we use to estimate the distribution parameter(s), and the accessibility of (flood) data. The underling study is based on three well known goodness-of-fit criteria to assess the appropriateness of probability distributions. These tests are (i) Akaike Information Criterion (ii) Bayesian Information Criterion and (iii) Anderson Darling test goodness-of-fit. Also the L-moment ratio diagram is involved for visual chromatic evaluation of tests.

3.5.1 Akaike Information Criterion

The Akaike information criterion (AIC) measures the relative excellence of Statistical model for a data set. The AIC estimates the performance of each model in the set of under study models, relative to other models. Thus we can say that AIC provide a basis for the selection of an appropriate model.

AIC is based over information philosophy; it offers a comparative of information lost when a model is used for signifying the procedure that produces data. AIC is not a test that signifies a null and alternative hypothesis for the excellence of model, so the criterion provides nothing about the quality of the model. If all underling models poorly fit the data, the criterion will give no warning.

The Akaike Information Criterion is defined by the relation:

$$AIC = 2k - 2\ln(\hat{L})$$

Where

$\hat{L}$ = the value that maximized the likelihood function of given model. By definition $\hat{L} = p(x/\hat{\theta}, M)$, here M is the involved model and x is observed data and

k = the number of estimated parameters.

In the class of under study models the appropriate model is one that has minimum AIC value. The AIC value contains the penalty which is growing function of number of estimated parameters. The penalty disappoints over fitting, because growing number of parameters always rise the goodness-of-fit.

The AIC is practically applied to a set of models and then the AIC values of the corresponding models are obtained. The presentation of true model always contains the information loss that is caused by using a model. The best model to be selected among candidate models is the one that minimalize estimated information loss.

3.5.2 Bayesian Information Criterion

Bayesian Information Criterion (BIC) also known as Schwarz criterion (SBIC) is the criterion used to differentiate an appropriate model in the class of finite models. On the basis of this criterion, the model is considered to be the best whose BIC value is minimum as compare to other models involved in the underlying research. It has a closed relation with the AIC as both contain the likelihood function in calculation, the only difference is that it contain ln(n) instead of "2". As ln(n) = 2.08 for n = 8, and the research is based on almost more than 8 data points, that is why the AIC value is often less than BIC.

While fitting models, the likelihood can be increased by the addition of parameters, but this will cause over fitting. AIC and BIC both resolve the problem by defining a penalty term concerned with number of model parameters. This penalty term is lesser for AIC than BIC. The BIC is defined by the relation

$BIC = \ln(n)k - 2\ln(\hat{L})$, Where

$\hat{L}$ = the value that maximized the likelihood function of given model. By definition $\hat{L} = p(x/\hat{\theta}, M)$, here M is the involved model and x is observed data.

n = strength of data points

k = number of estimated parameters

BIC is very close in relation with other criteria that penalize the likelihood, such as AIC and RIC. The Bayesian Information Criterion is suffering with two limitations that are, the number of data points in the study must be greater than the number of parameters to be estimated. The second limitation is, that in case of high dimension and complex collection of candidate models, BIC is hopeless the variable selection.

3.5.3 Anderson Darling test goodness-of-fit

Anderson Darling test is the modified form of Kolmogorov- Smirnov (K-S) test that assign more weight to tails than that of K-S test. We use the Anderson Darling test whether, the data sample belong to the population having a definite distribution. In K-S test is independent of specific distribution, on the other hand the Anderson Darling test is used for specific distributions to calculate values for the establishment of critical values. The basic and notable advantage of the AD test is that it is more sensitive.

The disadvantage of the AD test is that there exists no unique table for critical values to be used for all distributions. We have to calculate the critical values on distribution basis i.e. the table that containing critical values varies from distribution to distribution. The critical value tables are available for exponential, uniform, normal, lognormal, weibull, logistic, generalized pareto and extreme value type 1 distributions.

The null and alternative hypotheses in the Anderson Darling test are classified as

H_0: The data belong to a specific distribution, against the alternative hypothesis,

H_1: The data don't belong to a specific distribution.

The test statistic for Anderson Darling test is given by:

$$A^2 = - n - s$$

Where

$$S = \sum_{i=1}^{n} \frac{(2i-1)}{n} [\ln F(Y_i) + \ln (1\text{-}F(Y_{n+1-i})]$$

F being a distribution function, Y_i are order data, ln is the natural log and n is data size.

3.6 The probability distributions of the study

3.6.1 The Normal Distribution (NOR)

Normal distribution is one of the most important distribution in the class of continuous distributions. It is also named as the Gaussian distribution. The probability density function of the distribution is given by:

$$f(x/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\frac{x-\mu}{2\sigma})^2}, \qquad -\infty \leq x \leq \infty$$

where μ *and* σ are two parameters of the normal distribution.

μ = mean of distribution and σ = standard deviation of distribution.

The notation $N(\mu, \sigma^2)$ discusses the normal distribution with mean μ and variance σ^2.

3.6.2 The Log Normal Distribution (LNO)

The log normal distribution is basically a continuous distribution of such a random variable whose log values are normally distributed. The distribution is some time referred as the Galton distribution. If X is a random variable, distributed log normally, then ln(X) will be normally distributed. similarly, as X= exp(ln(X)), thus exp(ln(X)) must be distributed log normally. The pdf of the distribution is mentioned by the function:

$$f(x/\mu,\sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\frac{\ln x-\mu}{2\sigma})^2}, \qquad x > 0$$

where μ is mean of distribution and σ is stander deviation of distribution. It has two parameters μ and σ.

3.6.3 The Exponential Distribution (EXP)

The exponential distribution is in the class of continuous distributions whose density function is given by

$$f(x/\lambda) = \lambda\, e^{-\lambda x}, \; x > 0, \; \lambda > 0 \qquad \text{or}$$

$$f(x/\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \; x > 0, \beta > 0$$

where λ is the only parameter for first function and β is the only parameter for second function.

3.6.4 The Generalized Extreme value Distribution (GEV)

The generalized-extreme-value distribution is the continuous family of probability distributions. It has been developed by the combination of Weibull, Gumbel and Frechet families. It is also known as Fisher- Tippett distribution. The pdf of the distribution is given by

$$f(x;\mu,\sigma,\xi) = \exp(-(1+\ \xi(\frac{x-\mu}{\sigma}))^{-\frac{1}{\xi}})$$

where μ, σ, ξ are the three parameters defined as

μ = location parameter , σ = scale parameter and ξ = shape parameter

3.6.5 The Gumbel Distribution (GUM)

The Gumbel distribution is also named as the extreme value distribution type-1. It is discussed in two forms. The one that is based on smallest extremes, is referred as minimum case and the other based on largest extremes is referred as maximum case. The probability density of gumbel (minimum) is given by

$$f(x;\mu,\beta) = \frac{1}{\beta} e^{\frac{x-\mu}{\beta}} e^{-e^{\frac{x-\mu}{\beta}}}$$

where μ = location parameter and β = scale parameter. Similarly for the largest extremes the density function is given by

$$f(x;\mu,\beta) = \frac{1}{\beta} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}}$$

CHAPTER 4

Results and Discussions

4.1 Assumptions

In the early stage of the analysis, we applied the assumptions of FFA to examine the independence, stationary and homogeneity of the Annual Maximum Peak (AMP) flows data by using Lag-one correlation, Mann-Whitney U test and Ljung Box test respectively. The results obtained by micro soft exile for the seven sites of Indus River are given in table 4.1.1. On the basis of results the hypothesis of independence, homogeneity and stationary are accepted as shown in the table 4.1.1. After the satisfaction of the assumption the analysis is performed by Statistical Programming language say 'R'.

Table 4.1.1 Basic assumptions of FFA applied on the AMP flows

Criterion	Lag-one correlation		Mann-Whitney U test		Ljung box test	
Site	P-Value	Decision	P-Value	Decision	P-Value	Decision
Tarbela	0.423	accept H_0	0.904	accept H_0	0.759	accept H_0
Kalabagh	0.670	accept H_0	0.106	accept H_0	0.958	accept H_0
Chashma	0.620	accept H_0	0.356	accept H_0	0.536	accept H_0
Taunsa	0.870	accept H_0	0.857	accept H_0	0.890	accept H_0
Kotri	0.135	accept H_0	0.656	accept H_0	0.869	accept H_0
Guddu	0.157	accept H_0	0.171	accept H_0	0.298	accept H_0
Sukkar	0.170	accept H_0	0.904	accept H_0	0.303	accept H_0

4.2 Sample L-moments and L-moments Ratios:

L-moments as well as L-moments ratios for 7 sites of Indus River are analyzed with the use of renowned programming language, say 'R'. the tabulation of results is given in table 4.2.1. The table shows the largest mean of 532466.59 and the smallest mean of 386962.96, that are for Guddo Barrage and tarbela respectively. The values of largest and smallest variance are of magnitude 173169.96 and 42836.48 respectively, that are for the sites Sukker and Tarbela respectively. The skewness and Kurtosis ranges from 0.0563-0.4047 and 0.0119-0.3237 respectively. The most consistent site is Tarbela.

Table 4.2.1 Sample L-moments and L-moments Ratios

Sites	l_1	l_2	l_3	l_4	t	t_3	t_4
Tarbela	386962.96	42836.48	10548.19	9095.72	0.1106	0.2462	0.2123
Kalabagh	464719.96	81158.68	14085.12	18309.99	0.1746	0.1735	0.2256
Chashma	475333.05	79990.80	9055.05	16374.09	0.1682	0.1132	0.2046
Taunsa	452791.55	76857.48	4325.55	13131.20	0.1697	0.0563	0.1709
Kotrri	395262.07	161959.43	65544.90	52419.16	0.4098	0.4047	0.3237
Guddo	532466.59	171201.83	35531.21	2033.71	0.3215	0.2075	0.0119
Sukker	498102.32	173169.96	37007.35	4941.75	0.3477	0.2137	0.0285

4.3 Comparison of estimation methods

The analysis of the three estimation methods on the basis of 5 distributions is discussed in table 4.3.1 as shown below. The results of the table 4.3.1 encourage the log normal (LNO) as a best fitted distribution with overall proportion 32/63 = 50.80%. Under the estimation method MLM, the best fitted distributions are NOR and GEV with individual proportion of 6/21 = 28.6% each. The next distribution best fitted under this method is GUM with proportion 5/21 = 23.8%.

Under the estimation method MLE, the best fitted distribution for the data is LNO with overall proportion of 13/21 = 61.9%. The next best fitted distribution under the method is GEV with proportion of 6/21 = 28.6%. The third estimation method also revealed LNO as a best fitted distribution with overall proportion of 15/21 = 71.4%. The next best distribution to the data is GEV with frequency of 3 that results 14.3%.

Table 4.3.1 Comparison of estimation methods

Method	MLM			MLE			MME		
Criterion / Site	AIC	BIC	ADC	AIC	BIC	ADC	AIC	BIC	ADC
Tarbela	GUM	GUM	GEV	GEV	GUM	GUM	GEV	GEV	GEV
Kalabagh	NOR	NOR	GUM	GUM	GUM	GEV	LNO	LNO	LNO
Chashma	GEV	GEV	LNO	GUM	LNO	LNO	LNO	LNO	LNO
Taunsa	GEV	GEV	LNO	LNO	LNO	NOR	LNO	LNO	GUM
Kotrri	GUM	GUM	GEV	LNO	LNO	LNO	LNO	LNO	LNO
Guddo	NOR	NOR	LNO	LNO	LNO	LNO	LNO	LNO	NOR
Sukker	NOR	NOR	LNO	LNO	LNO	LNO	LNO	LNO	NOR

4.4 Comparison of goodness-of-fit Criteria

All the three goodness-of-fit criteria as shown in the following table 4.4.1,fitted the LNO as best among the class of selected distributions. The LNO appeared best using the criteria AIC, BIC and ADC with corresponding frequencies 10, 11, 11 respectively that result 47.6%, 52.4% and 52.4%. Thus the overall proportion for LNO under the three selection criteria is 32/63 = 50.8%. The second best fitted distribution on overall basis is GEV and GUM that repeated 11 times each which results 52.4%.

Under the AIC the 2nd best distributions that fit the data are also GEV and GUM with corresponding percentages of 19.0% each. Under the BIC the 2nd best fitted distribution is GUM followed by GEV and NOR with frequencies 4, 3 and 3 respectively that results 19.0%, 14.3% and 14.3% with accordance. Under the ADC the 2nd distribution that fit the data best is the GEV followed by GUM and NOR with corresponding proportions 19.0%, 14.3% and 14.3%.

Table 4.4.1 Comparison of goodness-of-fit Criteria

Criterion	AIC			BIC			ADC		
Method / Site	MLM	MLE	MME	MLM	MLE	MME	MLM	MLE	MME
Tarbela	GUM	GEV	GEV	GUM	GUM	GEV	GEV	GUM	GEV
Kalabagh	NOR	GUM	LNO	NOR	GUM	LNO	GUM	GEV	LNO
Chashma	GEV	GUM	LNO	GEV	LNO	LNO	LNO	LNO	LNO
Taunsa	GEV	LNO	LNO	GEV	LNO	LNO	LNO	NOR	GUM
Kotrri	GUM	LNO	LNO	GUM	LNO	LNO	GEV	LNO	LNO
SGuddo	NOR	LNO	LNO	NOR	LNO	LNO	LNO	LNO	NOR
Sukker	NOR	LNO	LNO	NOR	LNO	LNO	LNO	LNO	NOR

4.5. Comparison of distribution for best fit

The table 4.5.1 discusses the frequency distribution and %age frequency distribution of 5 distributions under 7 sites for the AMP flows. The results in the table 4.5.1 showed that the GEV and GUM each are equally suitable for 17.5% of the sites, the NOR is appropriate for 14.3% of the sites and the LNO is the one that is suitable for more than half i.e. 50.8% of the sites is

appeared to be the best. One main point that is clear in the table 4.5.1 is that the EXP distribution has played no role in fitting the AMP flows data in performing FFA.

The results in the table 4.5.1 for individual sites signaled the LNO to be the best in four sites i.e. Chashma, Kotrri, Guddo and Sukker with proportion of 66.7% each. The importance of the distribution for the individual sites except EXP, ranges from 11.1% to 66.7%.

Table 4.5. 1 Comparison of distribution for best fit

Distribution	GEV		GUM		EXP		NOR		LNO		total
Site	Freq:	%	Freq:	%	Freq:	%	Freq:	%	Freq:	%	
Tarbela	5	55.6	4	44.4	0	0	0	0	0	0	9
Kalabagh	1	11.1	3	33.3	0	0	2	22.2	3	33.3	9
Chashma	2	22.2	1	11.1	0	0	0	0	6	66.7	9
Taunsa	2	22.2	1	11.1	0	0	1	11.1	5	55.6	9
Kotrri	1	11.1	2	22.2	0	0	0	0	6	66.7	9
Guddo	0	0	0	0	0	0	3	33.3	6	66.7	9
Sukker	0	0	0	0	0	0	3	33.3	6	66.7	9
total	11	17.5	11	17.5	0	0	9	14.3	32	50.8	63
%age	17.5		17.5		0		14.3		50.8		100.0

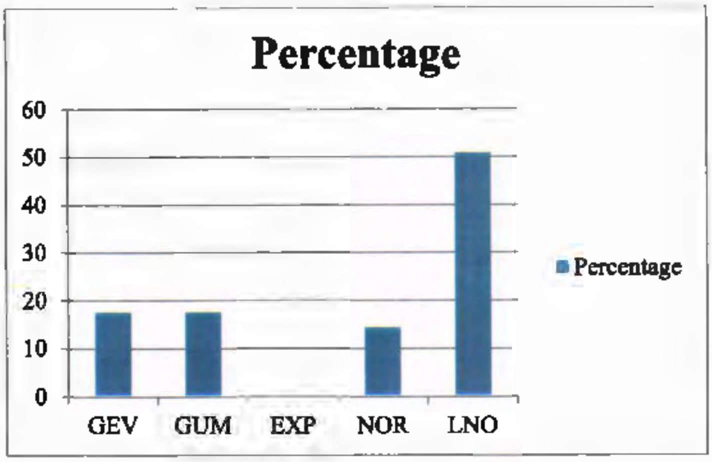

Figure 4.5.1 Comparison of distribution for best fit

4.6 Importance of distributions on the basis of estimation methods

The table 4.6.1 explains the importance of a particular distribution under an estimation method. As from the table 4.6.1 it is obvious LNO is the best under MME with corresponding score of 71.4% then for next time LNO is again the best under MLE with corresponding score of 61.9% and the next maximum value in the rest is 28.6% that is of GEV and NOR under MLM. On the basis of results of table 4.6.1, out of 5 distributions LNO is the best by using MME and MLE with corresponding average score of 66.7%, whereas, the MLM did not give any indication for the importance of LNO.

4.7. Importance of distributions on the basis of goodness-of-fit criteria

Table 4.7.1 clearly discussed the importance of LNO over other distribution of the class under all three estimation methods. All the three methods AIC, BIC and ADC respectively assigned a score of 47.6%, 52.4% and 52.4%. The 2nd highest score 19.0% is that of GUM under AIC and BIC and GEV under AIC and ADC.

Table 4.7.1 Importance of distributions on the basis of goodness-of-fit criteria

Distribution	GEV		GUM		EXP		NOR		LNO		total
Site	Freq:	%	Freq:	%	Freq:	%	Freq:	%	Freq:	%	
AIC	4	19.0	4	19.0	0	0	3	14.3	10	47.6	21
BIC	3	14.3	4	19.0	0	0	3	14.3	11	52.4	21
ADC	4	19.0	3	14.3	0	0	3	14.3	11	52.4	21
total	11	17.5	11	17.5	0	0	9	14.3	32	50.8	63
% age	17.5		17.5		0		14.3		50.8		100.0

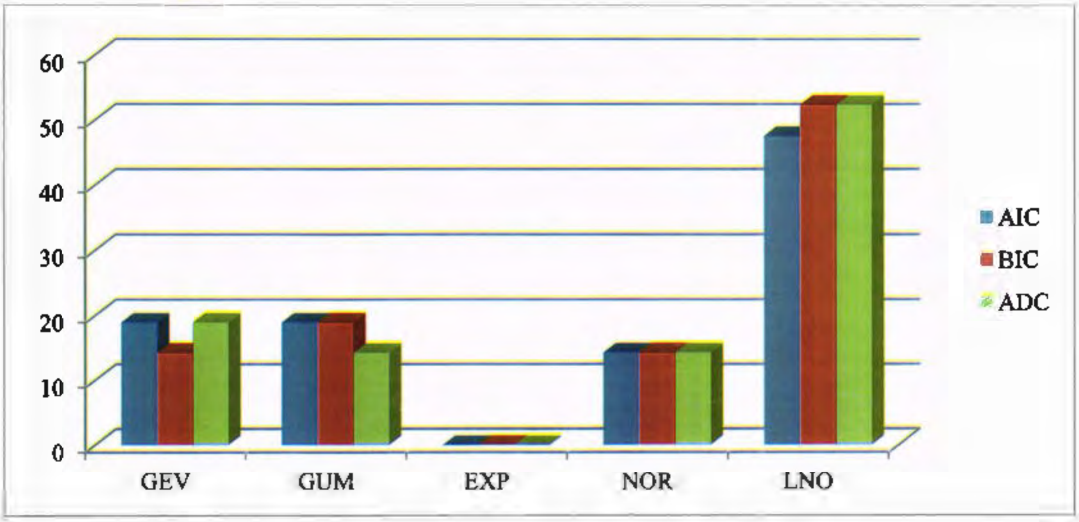

Figure 4.7.1 Importance of distributions on the basis of goodness-of-fit criteria

4.8 Summary of best fit dist: on the basis of ADC for the sites

As in the above tables we mentioned that on the basis of overall study the LNO is the best one. This does not means that for all sites of the underlying study LNO must be applied for prediction. It means that for most of the sites of Indus river LNO is suitable. Now we need a further study to signify most suitable distributions for various sites.

The following table 4.8.1 is based on the AD test statistic. The AD test statistic of distribution for each site is ranked from 1 to 3 such that minimum value is assigned with a rank of 1 and so on. The table 4.8.1 clarify that under ADC the estimation method MLM is the best one as on the basis of the method 6 sites (out of 7) are assigned with rank 1. In other words we can say that for 85.8% of the sites the most suitable distributions are obtained by MLM.

The next point that is clear from the table is that under this goodness-of-fit criterion the most appropriate distributions are LNO and GEV respectively. For the four sites namely, Chashma,Taunsa, Guddo and Sukker the LNO is most suitable, whereas for the remaining three sites namely, Tarbela, Kalabagh and Kitrri GEV is the best. Thus for 57% of the sites LNO is preferred and 43% of the sites GEV is suitable.

Table 4.8.1 Summary of best fit distribution on the basis of ADC for the sites

Site	Rank 1		Rank 2		Rank 3	
	Distribution	Parameter Est:Method	Distribution	Parameter Est:Method	Distribution	Parameter Est:Method
Tarbela	GEV	MLM	GEV	MLE	LNO	MLM
Kalabagh	GEV	MLE	LNO	MLE	GUM	MLE
Chashma	LNO	MLM	GEV	MLM	LNO	MME
Taunsa	LNO	MLM	GEV	MLM	NOR	MME
Kotrri	GEV	MLM	LNO	MLE	LNO	MME
Guddo	LNO	MLM	GUM	MLM	GEV	MLM
Sukker	LNO	MLM	LNO	MLE	GEV	MLM

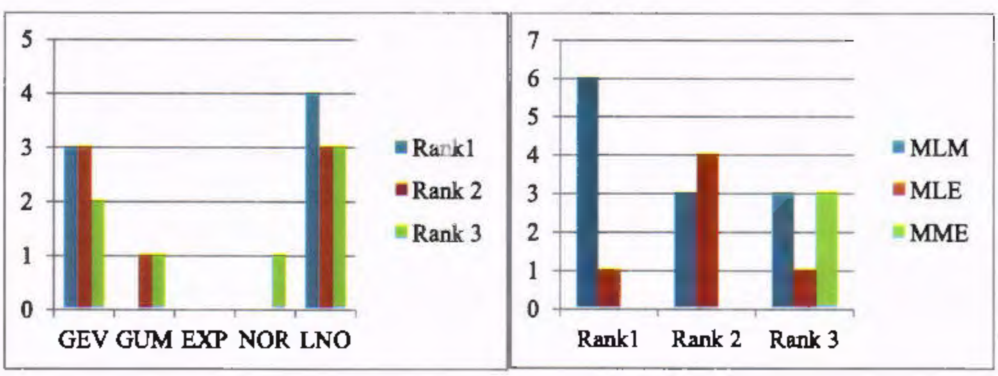

Figure 4.8.1 Summary of best fit distribution on the basis of ADC for the site

4.9 Summary of best fit dist: on the basis of AIC and BIC for the sites

The table 4.9.1 is based on AIC and BIC, the two widely used goodness-of-fit criteria for the selection of an appropriate distribution for a study. From the table it is obvious that LNO and GUM are best for 3 sites each and for one site NOR is assigned with rank 1. For the 2nd position under Rank 2 the next suitable distribution is LNO, as it is next to the best for 71% sites. so here also the indication of best probability is toward LNO.

The next point being noted from the table is that the best estimation methods are MLE and MLM respectively. For 43% of the sites the best distributions are obtained by MLM and MLE both respectively. Only in one case for the first time MME appeared to categories a best fit distribution for a site.

If we concentrate on the combine results of table 4.8.1 and 4.9.1, we observe that for 64% of the sites the best distributions are obtained by MLM, whereas, for 29% of the sites the appropriate distributions are decided by MLE. Thus on the basis of individual goodness-of-fit criterion , as well as overall basis the parameter estimation method MLM is the best estimation method for selecting the best probability distribution for sites of Indus river in Pakistan.

Table 4.9.1 Summary of best fit dist: on the basis of AIC and BIC for the sites

	Rank 1		Rank 2		Rank 3	
Site	Distribution	Parameter Est:Method	Distribution	Parameter Est:Method	Distribution	Parameter Est:Method
Tarbela	GUM	MLM	GEV	MLE	GUM	MLE
Kalabagh	GUM	MLE	LNO	MLE	LNO	MME
Chashma	LNO	MME	GUM	MLE	NOR	MLM
Taunsa	NOR	MLM	LNO	MLE	LNO	MME
Kotrri	GUM	MLM	LNO	MLE	LNO	MME
Guddo	LNO	MLE	LNO	MME	GUM	MLE
Sukker	LNO	MLE	LNO	MME	GUM	MLE

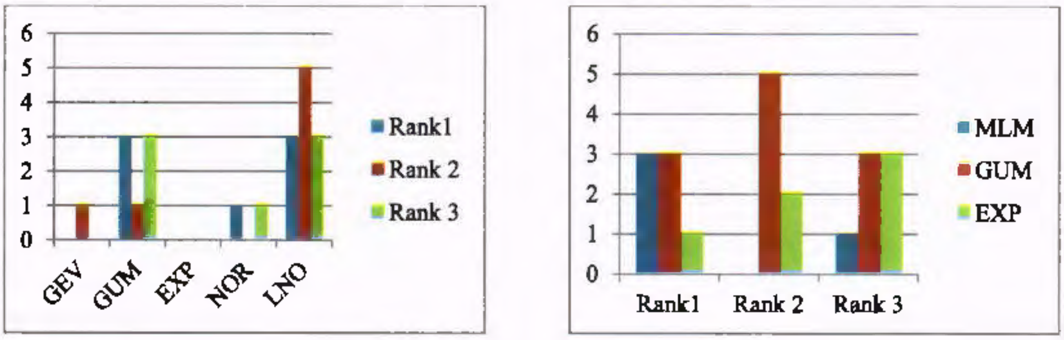

Figure 4.9.1 Summary of best fit dist: on the basis of AIC and BIC for the sites

4.10 Summary of best fit dist: on the basis of Ranking for the sites

Table 4.10 discusses the cross tabulation of distribution and ranks. The first entry of the table i.e. on the intersection of Rank 1 and GEV "8" means that GEV is repeated for 8 times for Rank 1 and so on. From the table it is clear that LNO is repeated 21 times for Rank 1, so without any doubt we ca: ooldly say that LNO is the best.

4.10.1 Summary of best fit dist: on the basis of Ranking for the sites of Indus River

Distribution	Rank 1	Rank 2	Rank3	Rank 4	Rank 5	total
GEV	8	18	9	6	1	42
GUM	7	8	9	4	14	42
EXP	0	4	7	16	15	42
NOR	6	3	12	15	6	42
LNO	21	9	5	1	6	42
total	42	42	42	42	42	

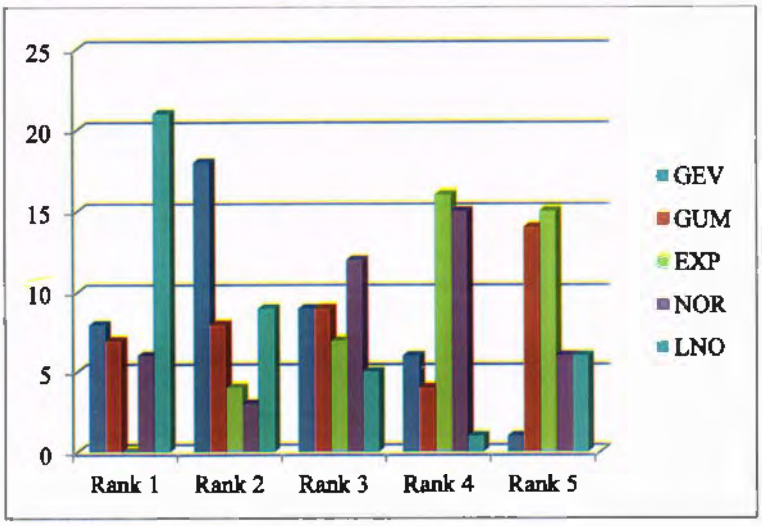

Figure 4.10.1 best fit dis on the basis of Ranking for the sites

4.11 comparisons of distributions for best fit using ranks for GOF test statistic

The table 4.11.1 discusses the ranks assigned to distributions on the basis of goodness-of-fit test statistic criterion. The values in the body of table against each estimation method for individual criterion are ranked from 1 through 5 in such a way that rank "1" is allotted to minimum value in the row and so on. It is open from the table that LNO has the minimum rank total of "88" in the distributions. It is a clear indication that LNO is the most suitable for most of the sites of the study.

Table 4.11.1 comparison of distributions for best fit using ranks for GOF test

Site	Method	Criterion	GEV	GUM	EXP	NOR	LNO
Tarbela	MLM	AIC & BIC	4	1	3	2	5
	MLM	ADC	1	3	5	4	2
	MLE	AIC & BIC	1	2	5	4	3
	MLE	ADC	2	1	4	5	3
	MME	AIC & BIC	1	5	4	3	2
	MME	ADC	1	5	4	3	2
Kalabagh	MLM	AIC & BIC	4	2	3	1	5
	MLM	ADC	2	1	5	4	3

	MLE	AIC & BIC	3	1	5	4	2
	MLE	ADC	1	3	4	5	2
	MME	AIC & BIC	2	5	4	3	1
	MME	ADC	2	5	4	3	1
Chashma	MLM	AIC & BIC	1	3	4	2	5
	MLM	ADC	2	3	5	4	1
	MLE	AIC & BIC	2	1	5	4	3
	MLE	ADC	2	4	3	5	1
	MME	AIC & BIC	2	5	4	3	1
	MME	ADC	2	5	4	3	1
Taunsa	MLM	AIC & BIC	1	2	4	3	5
	MLM	ADC	2	4	5	3	1
	MLE	AIC & BIC	3	2	5	4	1
	MLE	ADC	2	4	5	1	3
	MME	AIC & BIC	3	5	4	2	1
	MME	ADC	3	5	4	1	2
Kotrri	MLM	AIC & BIC	4	1	2	3	5
	MLM	ADC	1	3	4	5	2
	MLE	AIC & BIC	2	4	3	5	1
	MLE	ADC	3	2	4	5	1
	MME	AIC & BIC	3	5	2	4	1
	MME	ADC	2	5	3	4	1
Guddo	MLM	AIC & BIC	5	1	2	3	4
	MLM	ADC	3	2	5	4	1
	MLE	AIC & BIC	3	2	5	4	1
	MLE	ADC	2	3	5	4	1
	MME	AIC & BIC	2	5	4	3	1
	MME	ADC	4	5	3	1	2
Sukker	MLM	AIC & BIC	4	3	2	1	5
	MLM	ADC	2	3	5	4	1
	MLE	AIC & BIC	3	2	5	4	1
	MLE	ADC	2	3	5	4	1
	MME	AIC & BIC	2	5	4	3	1
	MME	ADC	4	5	3	1	2
		Rank total	100	136	168	138	88

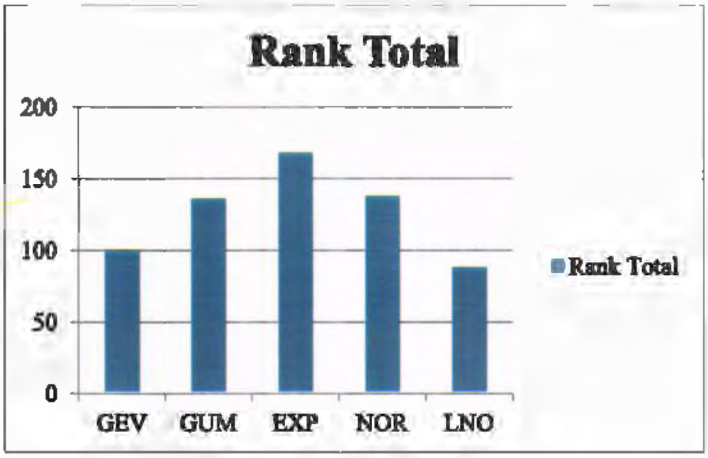

Figure 4.11.1 comparison of distributions for best fit using ranks

CHAPTER 5

RESULTS AND DECISIONS:

The present study is based on 7 sites of Indus River in Pakistan. The data being processed for analysis has obtained from WAPDA and Federal Floods Commission (FFA) of Pakistan. the unit of measurement used for AMP flows data of 7 sites is cusec. The AMP flows recorded length for sites ranges from 22 to 56 years.

The study is based on three estimation methods (MME, MLE and MLM), three goodness-of-fit criteria (ADC, BIC and AIC) processed over five probability distributions (GEV, GUM, EXP, NOR and LNO) for the purpose to find out the distribution that fit the data best.

The results revealed that among the five distributions LNO is the best one. The table 4.5.1 clearly described that on overall basis for 50.8% of the sites followed by GEV and GUM which are best for 17.5% of the sites each. Under the estimation methods as shown in table 4.6.1 again the LNO is best at 61.9% and 71.4% of the sites respectively under MME and MLE. The NOR and GEV are the next distributions that are best for 28.6% of the sites under MLM.

All the goodness-of-fit criteria favor LNO as a best distribution at 47.6%, 52.4% and 52.4% sites respectively under AIC, BIC and ADC. The overall ranking on the basis of AIC,ADC and BIC statistics the minimum rank total of 88 indicates that LNO is the best distribution followed by GEV with rank total of 100. The individual ranking in table 4.10 LNO is assigned 21 times for rank 1 which is 50% of the assigned to all distribution.

The AD test statistic of distribution for each site are ranked from 1 to 3 such that minimum value is assigned with a rank of 1 and so on. The table 4.8.1 clarify that under ADC the estimation method MLM is the best one as on the of the method 6 sites (out of 7) are assigned with rank 1. In other words we can say that for 85.8% of the sites the most suitable distributions are obtained by MLM.

From the table it is obvious that LNO and GUM are best for 3 sites each and for one site NOR is assigned with rank 1. For the 2nd position under Rank 2 the next suitable distribution is LNO, as it is next to the best for 71% sites. so here also the indication of best probability is toward LNO.

The next point being noted from the table is that the best estimation methods are MLE and MLM respectively. For 43% of the sites the best distributions are obtained by MLM and MLE both respectively. Only in one case for the first time MME appeared to categories a best fit distribution for a site.

If we concentrate on the combine results of table 4.8.1 and 4.9.1, we observe that for 64% of the sites the best distributions are obtained by MLM, whereas, for 29% of the sites the appropriate distributions are decided by MLE. Thus on the basis of individual goodness-of-fit criterion , as well as overall basis the parameter estimation method MLM is the best estimation method for selecting the best probability distribution for sites of Indus river in Pakistan.

REFERENCES

Abida, H., & Ellouze, M. (2008). Probability distribution of flood flows in Tunisia. *Hydrology and Earth System Sciences Discussions*, *12*(3), 703-714.

Ahmad, U. N., Shabri, A., & Zakaria, Z. A. (2011). Flood frequency analysis of annual maximum stream flows using L-moments and TL-moments approach. *Applied Mathematical Sciences*, *5*(5), 243-253.

Ahmad, I., Fawad, M., & Mahmood, I. (2015). At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. *Polish Journal of Environmental Studies*, *24*(06), 2345-2353.

Bhattara, K. P. (2005). Flood frequency analysis of Irish river flow data using variants of L-Moment. In *National Hydrology Seminar*.

Bobee, B., Cavadias, G., Ashkar, F., Bernier, J., & Rasmussen, P. (1993). Towards a systematic approach to comparing distributions used in flood frequency analysis. *Journal of Hydrology*, *142*(1-4), 121-136.

Bobée, B., & Rasmussen, P. F. (1995). Recent advances in flood frequency analysis. *Reviews of Geophysics*, *33*(S2), 1111-1116.

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of hydrology*, *18*(3-4), 257-271.

Davies, L., Mister, R., Spence, D. P., Calverley, P. M., Earis, J. E., & Pearson, M. G. (1997). Cardiovascular consequences of fibreoptic bronchoscopy. *European Respiratory Journal*, *10*(3), 695-698.

Deka, S., Borah, M., & Kakaty, S. C. (2009). Distributions of annual maximum rainfall series of north-east India. *European Water*, *27*(28), 3-14.

Dupuis, D. J. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, *1*(3), 251-261.

Galoie, M., Zenz, G., & Motamedi, A. Rainfall analysis for the Schoeckelbach Basin (Australia) and determining its best-fit probability distribution model. *DOI, 10*, 43-52.

Gocic, M., & Trajkovic, S. (2014). Drought characterisation based on water surplus variability index. *Water resources management*, *28*(10), 3179-3191.

Gubareva, T. S., & Gartsman, B. I. (2010). Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water Resources*, *37*(4), 437-445.

Haddad, K., & Rahman, A. (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment*, *25*(3), 415-428.

Haddad, K., Rahman, A., & Green, J. (2011). Design rainfall estimation in Australia: a case study using L moments and generalized least squares regression. *Stochastic Environmental Research and Risk Assessment*, *25*(6), 815-825.

Hosking, J. R. M. (1992). Moments or L moments? An example comparing two measures of distributional shape. *The American Statistician*, *46*(3), 186-189.

Howson, C. P., Kinney, M. V., McDougall, L., Lawn, J. E., Blencowe, H., Cousens, S., ... & Kinney, M. (2017). Born too soon. *Collections*, *10*, 1.

Kabiri, R. (2014). *Assessment of climate change impact on runoff and peak flow: a case study on Klang watershed in West Malaysia* (Doctoral dissertation, University of Nottingham).

Karim, M. A., & Chowdhury, J. U. (1995). A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal*, *40*(1), 55-66.

Klemeš, V., Srikanthan, R., & McMahon, T. A. (1981). Long-memory flow models in reservoir analysis: What is their practical value?. *Water Resources Research*, *17*(3), 737-751.

Kurunç, A., Yürekli, K., & Çevik, O. (2005). Performance of two stochastic approaches for forecasting water quality and streamflow data from Yeşilırmak River, Turkey. *Environmental Modelling & Software, 20*(9), 1195-1200.

Lu, L. H., & Stedinger, J. R. (1992). Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. *Journal of Hydrology, 138*(1-2), 223-245.

Madsen, T. K., & Servais, P. (1997). The internationalization of born globals: an evolutionary process?. *International business review, 6*(6), 561-583.

Markiewicz, D., Lawford, H., Doyle, A. B., & Haggart, N. (2006). Developmental differences in adolescents' and young adults' use of mothers, fathers, best friends, and romantic partners to fulfill attachment needs. *Journal of Youth and Adolescence, 35*(1), 121-134.

Nair, R. P., Duffin, K. C., Helms, C., Ding, J., Stuart, P. E., Goldgar, D., ... & Ruether, A. (2009). Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. *Nature genetics, 41*(2), 199-204.

Önöz, B., & Bayazit, M. (1999). GEV-PWM model for distribution of minimum flows. *Journal of Hydrologic Engineering, 4*(3), 289-292.

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural hazards, 69*(3), 1803-1813.

Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology, 313*(1), 97-116.

Salem, A. Z., Canann, S. A., & Saigal, S. (2001). Mid-node admissible spaces for quadratic triangular 2D finite elements with one edge curved. *International Journal for Numerical Methods in Engineering, 50*(1), 181-197.

Stedinger, J. R. (1980). Fitting log normal distributions to hydrologic data. *Water Resources Research, 16*(3), 481-490.

Strupczewski, W. G., Singh, V. P., & Feluch, W. (2001). Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *Journal of Hydrology*, *248*(1), 123-142.

Talma, A. S., & Vogel, J. C. (1993). A simplified approach to calibrating 14 C dates. *Radiocarbon*, *35*(02), 317-322.

Tasker, G. D. (1987). A comparison of methods for estimating low flow characteristics of streams. *JAWRA Journal of the American Water Resources Association*, *23*(6), 1077-1083.

Zalina, M. D., Desa, M. N. M., Nguyen, V. T. A., & Kassim, A. H. M. (2002). Selecting a probability distribution for extreme rainfall series in Malaysia. *Water science and technology*, *45*(2), 63-68.

Zin, W. Z. W., Jemain, A. A., & Ibrahim, K. (2009). The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment. *Theoretical and applied climatology*, *96*(3-4), 337-344.

Appendix

Table 4.12.1 Anderson Darling GOF Test using method of L-moments

Dist:	GEV		GUM		EXP		NOR		LNO	
Sites	AD	P-value	AD	P-value	AD	P-value	AD	P-value	AD	P-value
Tarbela	0.2421	0.3854	0.3295	0.4380	7.5210	0.9400	1.7128	1.0000	0.2916	0.5550
Kalabagh	0.4398	0.8747	0.4349	0.6540	8.9710	0.9570	1.0540	0.9917	0.4733	0.9012
Chashma	0.3351	0.7417	0.4465	0.6930	6.9220	0.9490	0.5616	0.8504	0.3198	0.6724
Taunsa	0.5202	0.9480	0.8677	0.9600	11.590	0.9780	0.5567	0.8456	0.4612	0.9280
Kotrri	0.2855	0.4870	1.2080	0.9920	1.8340	0.7680	3.5363	1.0000	0.6283	0.9218
Guddo	0.5208	0.9389	0.5159	0.8070	1.0480	0.7120	0.7702	0.9564	0.4903	0.9262
Sukker	0.4093	0.8410	0.4222	0.6940	0.9192	0.6760	0.7412	0.9481	0.3808	0.8258

Table 4.12.2 Akaike and Bayesian Information Criteria for method of L-moment

Site	Dist:	GEV	GUM	EXP	NOR	LNO
Tarbela	aic	2156.18	1236.62	1716.50	1370.92	4273.51
	bic	2162.14	1240.60	1718.49	1374.90	4277.49
Kalabag	aic	17325.4	1228.33	1369.15	1227.20	2850.17
	bic	17330.9	1231.99	1370.98	1230.86	2853.82
Chashm	aic	64.9260	1183.87	1288.04	1146.18	2692.87
	bic	70.2096	1187.39	1289.80	1149.70	2696.40
Taunsa	aic	715.570	1440.80	1669.51	1487.02	3487.67
	bic	721.646	1444.85	1671.54	1491.067	3491.72
Kotrri	aic	1299.48	1099.62	1225.95	1234.82	2353.68
	bic	1304.83	1103.19	1227.73	1238.39	2357.24
Guddo	aic	1309.67	0666.38	0631.15	0621.85	1231.05
	bic	1312.94	0668.56	0632.24	0624.03	1233.24

Sukker	aic	1204.67	0635.79	0626.51	0622.36	1217.08
	bic	1207.94	0637.97	0627.60	0624.54	1219.27

Table 4.12.3 Anderson Darling GOF Test using method of maximum likelihood

Dist:	GEV		GUM		EXP		NOR		LNO	
Site	AD	P-value	AD	P-value	AD	P-value	AD	P-value	AD	P-value
Tarbela	0.2540	0.4079	0.2508	0.2429	0.6604	0.4386	1.7128	0.9998	0.5812	0.8653
Kalabagh	0.4297	0.8213	0.4334	0.6943	0.5034	0.6544	1.0540	0.9917	0.4333	0.6967
Chashma	0.3947	0.7661	0.4374	0.7011	0.4198	0.6935	0.5615	0.8504	0.3926	0.6220
Taunsa	0.5797	0.9349	0.7565	0.9487	0.9560	0.9609	0.5566	0.8456	0.7208	0.9385
Kotrri	0.9279	0.9978	0.7977	0.9600	1.8455	0.9923	3.5363	1.000	0.3162	0.4380
Guddo	0.6204	0.9646	0.6753	0.9224	5.0276	0.8071	0.7702	0.9563	0.5596	0.8512
Sukker	0.5484	0.9377	0.5647	0.8541	2.9336	0.6945	0.7412	0.9481	0.3990	0.6347

Table 4.12.4 AIC and BIC for method of maximum likelihood

Site \ Dist:		GEV	GUM	EXP	NOR	LNO
Tarbela	aic	1360.95	1361.06	1499.54	1385.56	1367.10
	bic	1366.92	1365.04	1501.53	1389.54	1371.08
Kalabag	aic	1224.56	1222.56	1294.53	1231.15	1222.63
	bic	1230.04	1226.22	1296.35	1234.80	1226.29
Chashma	aic	1146.05	1144.44	1212.17	1149.79	1147.58
	bic	1151.34	1147.96	1213.93	1153.31	1144.05
Taunsa	aic	1488.42	1488.27	1572.60	1489.68	1487.22
	bic	1494.50	1492.32	1574.62	1493.73	1491.27
Kotrri	aic	1213.00	1225.61	1224.08	1258.38	1211.37
	bic	1218.36	1229.18	1225.87	1261.95	1214.94
	aic	0618.60	0616.94	0626.15	0620.35	0614.90

	bic	0621.88	0619.12	0627.24	0622.53	0617.08
Sukker	aic	0618.82	0617.22	0623.22	0620.91	0614.95
	bic	0622.10	0619.40	0624.31	0623.10	0617.13

Table 4.12.5 Anderson Darling GOF Test using method of moments

Dist:	GEV		GUM		EXP		NOR		LNO	
Site	AD	P-value	AD	P-value	AD	P-value	AD	P-value	AD	P-value
Tarbela	0.6604	0.8650	38.313	0.9994	16.718	0.2428	1.7527	0.4073	1.0017	0.4308
Kalabagh	0.5034	0.6969	32.043	0.9915	10.648	0.6947	1.0540	0.8231	0.4470	0.6514
Chashma	0.4198	0.6225	28.359	0.8508	10.179	0.7018	0.5616	0.7616	0.3914	0.6933
Taunsa	0.9560	0.9387	36.642	0.8453	12.958	0.9486	0.5567	0.9394	0.7285	0.9612
Kotrri	1.8455	0.4381	34.650	0.9566	2.3690	0.9623	3.5363	0.9987	0.4963	0.9922
Guddo	5.0276	0.8517	18.278	0.9562	2.1429	0.9252	0.7702	0.9664	0.8985	0.8067
Sukker	2.9336	0.6347	18.015	0.9481	1.6837	0.8553	0.7412	0.9371	0.8160	0.6943

Table 4.12.6 Akaike and Bayesian Information Criteria for method of moments

Dist: / Site		GEV	GUM	EXP	NOR	LNO
Tarbela	aic	1365.50	1570.93	1499.54	1385.56	1369.00
	bic	1371.47	1574.90	1501.53	1389.54	1372.98
Kalaba	aic	1225.67	1419.76	1294.53	1231.15	1222.66
	bic	1231.63	1423.42	1296.35	1234.80	1226.32
Chash	aic	1147.09	1334.54	1212.17	1149.79	1144.06
	bic	1153.06	1338.06	1213.93	1153.31	1147.58
Taunsa	aic	1490.07	1750.97	1572.60	1489.68	1487.50
	bic	1496.04	1755.02	1574.62	1493.73	1491.55

Kotrri	aic	1235.56	1385.33	1224.08	1258.38	1212.03
	bic	1241.53	1388.90	1225.87	1261.95	1215.60
Guddo	aic	0619.06	0711.25	0626.15	0620.35	0615.44
	bic	0625.02	0713.44	0627.24	0622.53	0617.62
Sukker	aic	0619.86	0710.99	0623.22	0620.91	0615.87
	bic	0625.83	0713.17	0624.31	0623.10	0618.05

Table 4.12.7 Goodness-of-fit test statistics under distributions for sites

Site	Method	Citerion	GEV	GUM	EXP	NOR	LNO
Tarbela	MLM	AIC	2156.18	1236.62	1716.50	1370.92	4273.51
		BIC	2162.14	1240.60	1718.49	1374.90	4277.49
		ADC	0.2421	0.3295	7.5210	1.7128	0.2916
	MLE	AIC	1360.95	1361.06	1499.54	1385.56	1367.10
		BIC	1366.92	1365.04	1501.53	1389.54	1371.08
		ADC	0.2540	0.2508	0.6604	1.7128	0.5812
	MME	AIC	1365.50	1570.93	1499.54	1385.56	1369.00
		BIC	1371.47	1574.90	1501.53	1389.54	1372.98
		ADC	0.6604	38.313	16.718	1.7527	1.0017
Kalabagh	MLM	AIC	17325.4	1228.33	1369.15	1227.20	2850.17
		BIC	17330.9	1231.99	1370.98	1230.86	2853.82
		ADC	0.4398	0.4349	8.9710	1.0540	0.4733
	MLE	AIC	1224.56	1222.56	1294.53	1231.15	1222.63
		BIC	1230.04	1226.22	1296.35	1234.80	1226.29
		ADC	0.4297	0.4334	0.5034	1.0540	0.4333
	MME	AIC	1225.67	1419.76	1294.53	1231.15	1222.66
		BIC	1231.63	1423.42	1296.35	1234.80	1226.32
		ADC	0.5034	32.043	10.648	1.0540	0.4470
	MLM	AIC	64.9260	1183.87	1288.04	1146.18	2692.87
		BIC	70.2096	1187.39	1289.80	1149.70	2696.40

		ADC	0.3351	0.4465	6.9220	0.5616	0.3198
	MLE	AIC BIC	1146.05 1151.34	1144.44 1147.96	1212.17 1213.93	1149.79 1153.31	1147.58 1144.05
		ADC	0.3947	0.4374	0.4198	0.5615	0.3926
	MME	AIC BIC	1147.09 1153.06	1334.54 1338.06	1212.17 1213.93	1149.79 1153.31	1144.06 1147.58
		ADC	0.4198	28.359	10.179	0.5616	0.3914
Taunsa	MLM	AIC BIC	715.570 721.646	1440.80 1444.85	1669.51 1671.54	1487.02 1491.06	3487.67 3491.72
		ADC	0.5202	0.8677	11.590	0.5567	0.4612
	MLE	AIC BIC	1488.42 1494.50	1488.27 1492.32	1572.60 1574.62	1489.68 1493.73	1487.22 1491.27
		ADC	0.5797	0.7565	0.9560	0.5566	0.7208
	MME	AIC BIC	1490.07 1496.04	1750.97 1755.02	1572.60 1574.62	1489.68 1493.73	1487.50 1491.55
		ADC	0.9560	36.642	12.958	0.5567	0.7285
Kotrri	MLM	AIC BIC	1299.48 1304.83	1099.62 1103.19	1225.95 1227.73	1234.82 1238.39	2353.68 2357.24
		ADC	0.2855	1.2080	1.8340	3.5363	0.6283
	MLE	AIC BIC	1213.00 1218.36	1225.61 1229.18	1224.08 1225.87	1258.38 1261.95	1211.37 1214.94
		ADC	0.9279	0.7977	1.8455	3.5363	0.3162
	MME	AIC BIC	1235.56 1241.53	1385.33 1388.90	1224.08 1225.87	1258.38 1261.95	1212.03 1215.60
		ADC	1.8455	34.650	2.3690	3.5363	0.4963
Guddo	MLM	AIC BIC	1309.67 1312.94	0666.38 0668.56	0631.15 0632.24	0621.85 0624.03	1231.05 1233.24
		ADC	0.5208	0.5159	1.0480	0.7702	0.4903
	MLE	AIC BIC	0618.60 0621.88	0616.94 0619.12	0626.15 0627.24	0620.35 0622.53	0614.90 0617.08

		ADC	0.6204	0.6753	5.0276	0.7702	0.5596
	MME	AIC	0619.06	0711.25	0626.15	0620.35	0615.44
		BIC	0625.02	0713.44	0627.24	0622.53	0617.62
		ADC	5.0276	18.278	2.1429	0.7702	0.8985
Sukker	MLM	AIC	1204.67	0635.79	0626.51	0622.36	1217.08
		BIC	1207.94	0637.97	0627.60	0624.54	1219.27
		ADC	0.4093	0.4222	0.9192	0.7412	0.3808
	MLE	AIC	0618.82	0617.22	0623.22	0620.91	0614.95
		BIC	0622.10	0619.40	0624.31	0623.10	0617.13
		ADC	0.5484	0.5647	2.9336	0.7412	0.3990
	MME	AIC	0619.86	0710.99	0623.22	0620.91	0615.87
		BIC	0625.83	0713.17	0624.31	0623.10	0618.05
		ADC	2.9336	18.015	1.6837	0.7412	0.8160

Table 4.12.8 number of years/ site processed in the study

Site	Tarbela	Kalabagh	Chashma	Taunsa	Kotrri	Guddo	Sukkers
No.of years	54	46	43	56	44	22	22