Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation

By

Sofia Khan

Reg No: 00000402527

Supervisor

Dr. Zamir Hussain

Department of sciences

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

July, 2024

Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation

By

Sofia khan

Reg No: 00000402527

Supervisor

Dr. Zamir Hussain

A thesis submitted to the National University of Sciences and Technology, Islamabad,

In partial fulfillment of the requirements for the degree of

Master of Computational Science and Engineering

Supervisor: Dr. Zamir Hussain

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

June, 2024

Annex A to NUST Letter No.
0972/102/Exams/Thesis-Cert
dated 23 Dec 16.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr/Ms Sofia Khan

Registration No. 00000402527 of SINES has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: Associate Professor SINES - NUST, Sector H-12 Islamabad

Name of Supervisor: Dr ZAMIR HUSSAIN

Date: ____________________

Signature of HoD with stamp: Dr. Mian Ilyas Ahmad HoD Engineering Professor SINES - NUST, Sector H-12 Islamabad

Date: 22 July 2024

Countersign by

Signature (Dean/Principal): Dr SYED IRTIZA ALI SHAH Principal & Dean SINES • NUST, Sector H-12 Islamabad.

Date: 23 JUL 2024

FORM TH-4

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by Sofia Khan (00000402527) Titled: Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation be accepted in partial fulfillment of the requirements for the award of Master of Computational Science and Engineering degree with (Grade).

Examination Committee Members

1. **Name:** **Signature:** _

2. **Name:** **Signature:** _

3. **Name:** **Signature:** _

Supervisor's name: Dr. Zamir Hussain **Signature:**________________

Date: ____________________

________________________ ________________

Head of Department **Date**

COUNTERSIGNED

Date: ___________

____

Dean/Principal

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the SINES in partial fulfillment of the requirements for the degree of Master of Science in the Field of Computational Science and Engineering.
Department of Science, SINES, National University of Sciences and Technology, Islamabad.

Student Name: Sofia Khan Signature:

Examination Committee:

a) External Examiner 1: Name Signature: ______________
(Designation & Office Address)
……………………………….

b) External Examiner 2: Name Signature: ______________
(Designation & Office Address)
…………………………………

Supervisor Name: Dr. Zamir Hussain Signature: _________
Name of Dean/HOD: Signature:

AUTHOR'S DECLARATION

I, Sofia *Khan* hereby state that my MS thesis titled "Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation" is my own work and has not been submitted previously by me for taking any degree from the National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Sofia Khan ____________

Date: ______________________________

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in the thesis titled "Optimizing Quantiles Estimation of Annual Maximum Rainfall in Pakistan Considering Pearson Type-III Distribution with Different Methods of Estimation" is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged, and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and the National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above-titled thesis, even after the award of the MS degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and NUST, Islamabad have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: ______________________________

Name: ________________________________

DEDICATION

"My thesis is a heartfelt homage to my cherished family members, who have played an integral role in shaping my academic journey. I express gratitude to my father and brother unwavering support and words of motivation echo in my mind. Moreover, I support my mother, whose prayers, inspirational guidance, and rent less determination have left an indelible mark on me."

Acknowledgment

With the name of Allah, the most compassionate, the most merciful, I would like to express my deepest gratitude to my esteemed supervisor, "Dr. Zamir Hussain," for his unwavering support, insightful guidance, and profound knowledge, which were instrumental in the successful completion of this research. His constant motivation, enthusiasm, and patience were invaluable throughout the entire process. Without a doubt, his dedication to research continues to be a perpetual source of inspiration for me.

Furthermore, I acknowledge the pivotal role of the School of Interdisciplinary Engineering and Sciences (SINES) at NUST in providing the essential research infrastructure and supportive environment for this project. I am deeply grateful to the members of my Guidance and Examination Committee (GEC) "Dr. Tariq Saeed" and "Dr. Salma Sherbaz," for their expert guidance, constructive critiques, and invaluable support throughout the research process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

S.N0.	Abbreviations/symbols	Meaning
1	LM	L-Moments
2	MPS	Maximum Product Spacing
3	MLE	Maximum Likelihood
4	AMRS	Annual Maximum Rainfall Analysis
5	PE3	Pearson type III distribution

Abstract

Particularly considering the rapidly intensifying effects of global climate change, extreme rainfall events provide challenging obstacles to the effective management of water resources and the advancement of infrastructure. It is becoming increasingly important to precisely anticipate these events as Pakistan struggles with growing vulnerabilities to more intense extreme weather events, such as the devastating floods of 2010 and 2022. To assess extreme rainfall occurrences in Pakistan, this study examines the effectiveness of the Maximum Product Spacing (MPS) approach in conjunction with the Pearson Type III distribution. This research tries to improve the accuracy and dependability of extreme rainfall models by examining MPS with other estimating techniques. Pakistan is at the forefront of the effects of climate change, with increasing susceptibilities to more intense extreme weather events. The study examines differences in the annual maximum rainfall series in the Pakistan Meteorological Department demarcated zones A and C within this framework. The study concluded that when the data shows minor to moderate skewness and kurtosis and when the samples are small, the estimates produced by the LM approach show little Bias. When there is significant skewness and kurtosis in the data and a small to moderate sample size, the MPS approach is an acceptable substitute that yields accurate estimates. When data from characteristic values are low, and sample sizes are big, the MLE approach offers benefits. The superior performance of MPS is attributed to its ability to minimum the value of RMSE and Bias in all stations of zone A and C. It provides better estimates for the behavior of the tail of the distribution, which is significant for extreme value analysis. These findings provide useful guidance that the MPS method is reliable when fitted with the PE3 distribution, especially for extreme values.

Keywords: Maximum Product spacing, L-Moments, Maximum Likelihood, sectorial effects.

Chapter 1

1. Introduction

Climate change, a phenomenon defined by long-term changes in regional or global climate patterns, is one reason extreme weather events, such as temperature extremes and changes in precipitation patterns, have become more frequent. The frequency and severity of extreme weather events have grown, as has ocean acidification; ecosystems are changing; sea levels are rising because of melting polar ice; and threats to agriculture and food security are just a few of the significant repercussions of these changes. To reduce these catastrophic events, addressing climate change necessitates collective effort [1].

Hydrological research requires accurate estimation of the frequency of extreme rainfall events to manage water resources, identify risks, and construct infrastructure. To quantify extreme rainfall events at specific places using yearly maximum rainfall series data, at-site frequency analysis is an essential tool in this field. Furthermore, it offers a method for estimating the likelihood that extreme rainfall events may transpire at certain places based on yearly maximum rainfall series data, which includes the annual maximum rain intensity observed. Certain techniques can be used to estimate the parameters of probability distributions that describe intense rainfall. One well-known method for effectively capturing the characteristics of severe occurrences is Maximum Product Spacing (MPS), which is also quite simple to implement. For extreme value situations, MPS offers trustworthy parameter estimations [2].

Hydrological studies require precise estimation of extreme rainfall events for efficient water resource management, infrastructure design, and risk assessment. Using annual maximum rainfall series data, at-site frequency analysis is a crucial tool in this field because it allows one to quantify extreme rainfall events at specific locations and determine the likelihood of these events occurring based on the highest observed rainfall intensity per year. Extreme rainfall is described by probability distributions with parameters that can be estimated using various techniques. A popular method for recording the features of extreme events is Maximum Product Spacing (MPS). It is easy to use and highly successful [3].

To provide reliable parameter estimates for extreme value analysis, MPS maximizes the product of the spacing between ordered observations. MPS increases the product of space between ordered observations and has robust parameter estimations that are useful in extreme value analysis.

Extreme rainfall intensity modeling can be effectively accomplished by combining MPS with the Pearson Type III distribution. As skewed data are typical in extreme rainfall occurrences, the Pearson Type III distribution is adaptable and can describe them. Because of its improved fit to observed data, this combination can improve the accuracy of at-site frequency analysis. The reliability of frequency analysis results is impacted by the chosen estimation method, which impacts crucial decisions related to infrastructure design and water resource management. Consequently, thoroughly examining this methodology's practical usefulness and accuracy is necessary [4].

This study uses yearly maximum rainfall series data to assess if Maximum Product Spacing is a suitable estimation approach for at-site frequency analysis. This research aims to enhance the precision and dependability of modeling extreme rainfall events by evaluating the effectiveness of MPS in estimating the Pearson Type III distribution parameters and contrasting it with other widely utilized techniques [5].

1.1 Flood Patterns in Pakistan:

Climate change's projected effects on sensitive ecosystems have made it a major worldwide worry for the last 20 years. Water resources, agriculture, and disaster management are all directly impacted by changes in rainfall. Pakistan's vulnerability to natural catastrophes, such as cyclones, floods, droughts, heavy rains, and earthquakes, is highlighted in reports published by the Task Force on Climate Change (2010). One-third of the people of Pakistan now experiences numerous disasters regularly because of variations in rainfall, storms, floods, and droughts. These occurrences are now more common and powerful in recent decades [6].

The devastating Flood of 2010:

The greatest flood in over eight decades, for instance, struck Pakistan in July 2010 after intense monsoon rains across Khyber Pakhtunkhwa, Sindh, Punjab, and portions of Balochistan. An estimated 2,000 people are thought to have died in this calamity, which also damaged or destroyed over 700,000 dwellings. This event saw 274 mm (about 10.79 in) of rainfall in Peshawar in 24 hours, exceeding the previous record set in April 2009. On the other hand, from 1998 to 2001, Pakistan's central and southern regions experienced severe droughts [7].

The Devastating Flood of 2022:

Extreme flooding struck Pakistan in the summer of 2022, causing massive destruction. Large areas of land were submerged by intense rains, resulting in the deaths of over 1,730 people and the displacement of over 33 million others. The total damage caused exceeded USD 14.9 billion, and the economic losses were estimated to be around USD 15.2 billion [8]. The Maximum Product Spacing estimate method is used in this work to match the Pearson Type III distribution and evaluate fluctuations in the yearly maximum rainfall series.

Pakistan's river systems are vital to the nation's agricultural output and way of life, but they are becoming more vulnerable to extreme weather events like floods and droughts due to climate change. The disastrous floods of 2010 and 2022 are proof of their terrible effects. Analyzing and comprehending fluctuations in annual maximum rainfall series is vital given the urgent need for efficient catastrophe management and adaptation techniques. This knowledge can improve forecasting abilities and guide resilient water resource management strategies. The present study aims to bridge this knowledge gap by utilizing the Maximum Product Spacing estimation method to examine variations in the annual maximum rainfall series. This analysis will aid in the development of tailored adaptation and mitigation strategies as well as a better understanding of the impacts of climate change on Pakistan's water resources [9].

1.2 Geography of Pakistan:

Pakistan's wide range of landforms and unpredictable rainfall patterns are hallmarks of its geographical variety. Pakistan experiences a variety of rainfall levels throughout the year, ranging from the vast desert regions like the Thar Desert in the southeast to the arid plateaus of the Balochistan Plateau in the southwest, the fertile plains of the Punjab region irrigated by the Indus River, and the towering ranges of mountain of the Himalayas, Karakoram, and Hindu Kush in the north (Ghazi et al., 2011). Depending on the area and monsoon patterns, Pakistan's annual rainfall varies greatly (Khan et al., 2022). Every year, the nation receives 50 to 300 millimeters (about

11.81 in) of rainfall. Rainfall is essential to maintaining agricultural output and livelihoods, particularly for people residing along river basins [10].

Seasonal variations and regional topography are two elements that affect the geographical distribution of rainfall. Rainfall amounts vary around the nation, mostly due to the monsoon rains, which mostly fall in the summer. Pakistan's largest river, the Indus, is a perfect illustration of this unpredictability. Significant precipitation, frequently enhanced by glacier melt, falls on the upper reaches of the Indus basin in the northern mountainous regions, where orographic effects and melting are common. In contrast, there is less rainfall in the lower ranges, especially in the provinces of Sindh and Punjab [11].

Annual precipitation in Pakistan varies from about 100 millimeters (about 3.94 in) in the desert regions of Balochistan to more than 1,000 millimeters in locations like Azad Kashmir and the northern mountainous regions. Particularly during dry seasons when rainfall is rare, these patterns of rainfall impact water resource availability, essential for domestic consumption, industry, and agriculture.

1.3 Pakistan Meteorological Department:

The research employs secondary data obtained from the Pakistan Meteorological Department (PMD) to guarantee thorough coverage. Using the Maximum Product Spacing (MPS) approach and the Pearson Type-3 (PE3) distribution model, the analysis focuses on the Annual Maximum Rainfall Series (AMRS) using this data. Through a thorough analysis of precipitation patterns and their statistical properties, this method contributes to a better understanding of climatic variability and trends. The work intends to enhance the resilience and adaptive capacity of water resource management policies in Pakistan by integrating AMRS data and modern statistical methodologies, thereby contributing significant insights into precipitation patterns.[12].

Study Rationale:

Pakistan's location and topography provide a wide variety of weather conditions all year round. Pakistan experiences a variety of climates, some of which are: Pakistan's summers are hot and dry, with frequently over 40°C temperatures, particularly in the southern parts [13].

1.5 Preference of PE3 as a model:

Examining the frequency of rainfall becomes particularly important during the monsoon season, which typically lasts from June to September and contributes significantly to Pakistan's annual precipitation. Experts can improve their capacity to predict the likelihood and severity of floods by closely examining rainfall data during this period. A probability distribution function used in rainfall frequency analysis is the PE3 (Pearson Type III) Distribution. Current studies suggest that PE3 might be a good choice for predicting the frequency of extreme rainfall events in some parts of Pakistan, particularly for high return periods. In comparison, distributions like the Generalized Extreme Value (GEV) and Generalized Normal (GNO) may be better suited for shorter return periods [14].

1.6 Preference of MPS as an estimation method:

A less well-established general estimating technique for extreme rainfall analysis is maximum product spacing. To address the tails of the rainfall frequency analysis and distribution of extreme events, it is important to space products as much as possible. The power of MPS lies in its capacity to examine a distribution's extreme values or tails. This capability is especially apparent in large datasets when the tails are heavily packed with data points and are crucial for comprehending uncommon occurrences [15]. By maintaining a consistent spacing distribution throughout the data range, MPS improves its capacity to record extremely high rainfall frequency analysis values.

Even with this, the traditional approach, Maximum Likelihood Estimation (MLE), may have difficulties with the tails, giving priority to fitting the central mass of the data and jeopardizing the accuracy of extreme value estimations, particularly in big datasets where outliers are more prevalent. It is important to recognize that MPS has several limits, even if it exhibits resilience to outliers, reducing their impact in comparison to MLE. Although MLE is a less well-established method and has less experience than MPS, it is still effective for a variety of applications, even if MPS has higher computational requirements [16].

Pakistan's Rainfall-Rich Regions:

Fig 1.1: Map showing the climatic zones of Pakistan

Stations of Zone A	Astor, Gilgit, Skardu, Muzaffarabad, Chitral, Gupis, Drosh, Bunji, Balakot, GhariDupatta, Kakul

Stations of Zone C	Quetta, Kalat and Dalbaddin

Fig 1.2 Selected zones of Pakistan

According to OCHA following are the affected vulnerable areas of Punjab, KPK, and Balochistan.

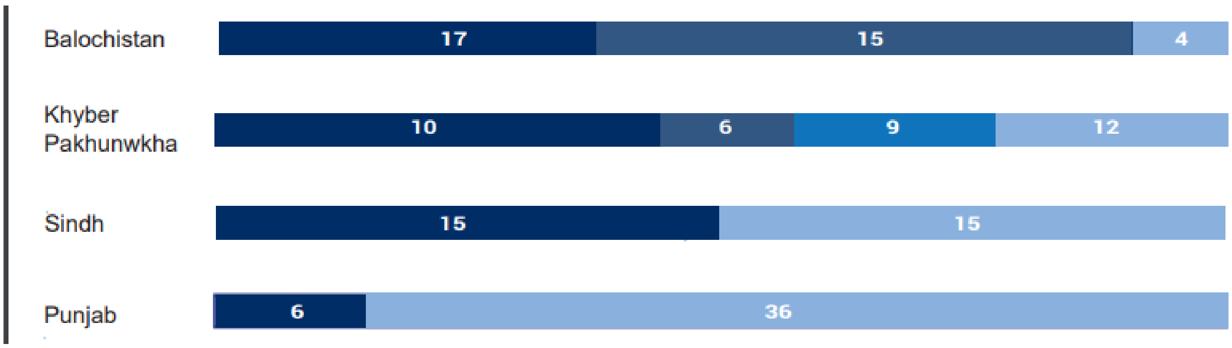

Fig 1.3: Vulnerable area of Pakistan by OCHA

The detailed analysis is given below in graphs in Fig 1.3, 1.4 and 1.5.

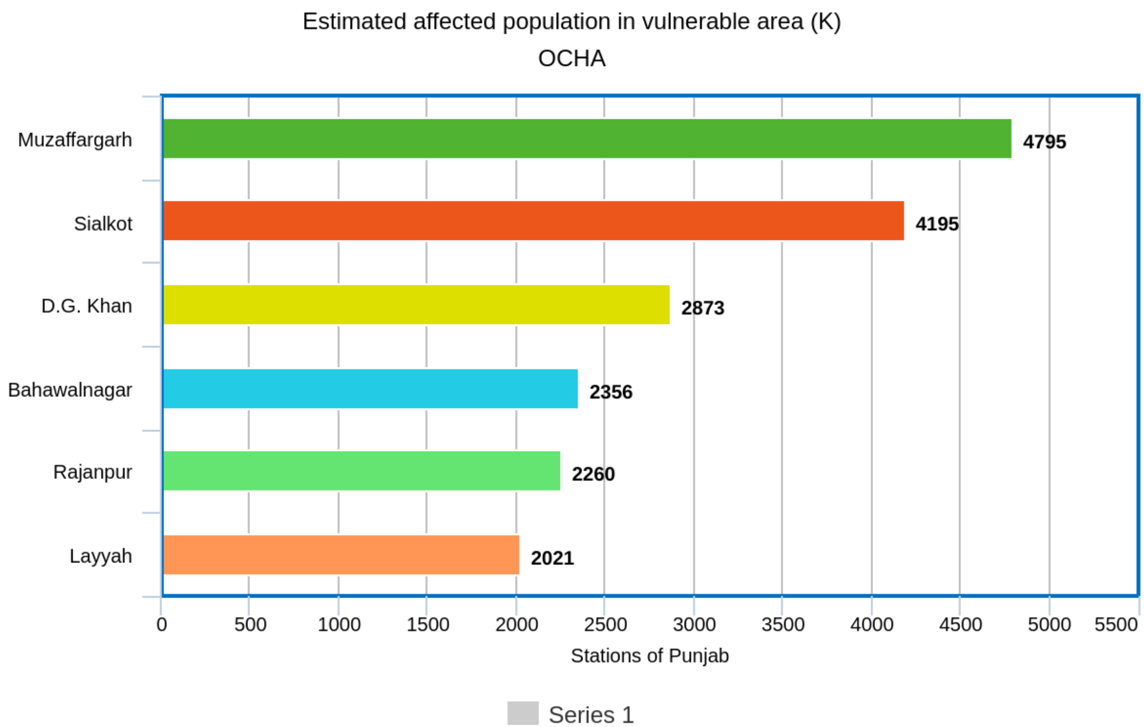

Fig 1.4: Estimated affected population in vulnerable stations of Punjab (K)

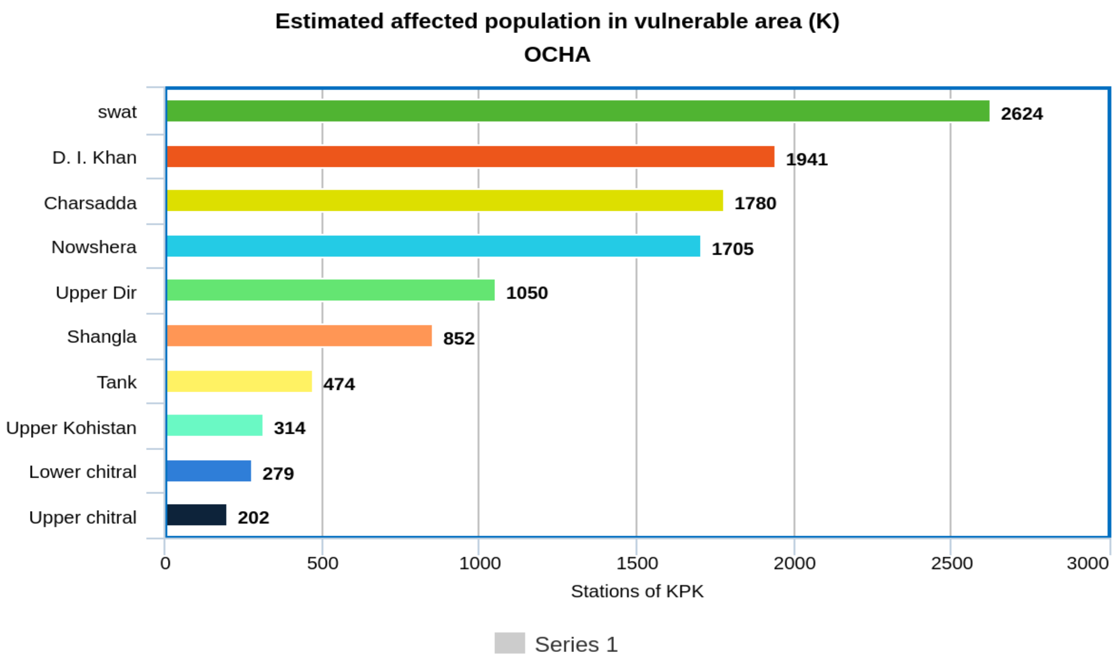

Fig 1.5: Estimated affected population in vulnerable stations of KPK (K)

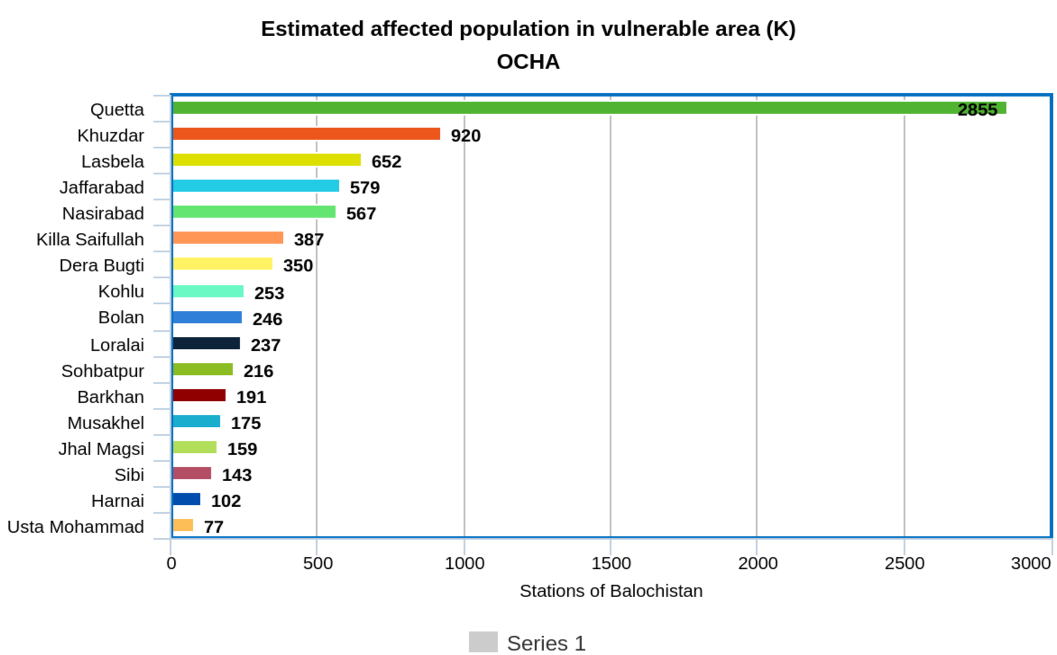

Fig 1.6: Estimated affected population in vulnerable stations of Balochistan (K)

1.6 Problem Statement:

The absence of reliable methodology may impede the ability to make accurate forecasts in the analysis of the annual maximum rainfall using the maximum product estimate method and the Pearson type III distribution. Current methods may need to account for important variables, resulting in inaccurate assessments of heavy precipitation occurrences. Closing this gap will help communities and infrastructure become more resilient to the effects of extreme weather. Therefore, creating a thorough framework that best incorporates these techniques is imperative to guarantee more accurate evaluations of the annual maximum rainfall.

1.7 Motivation:

This research aims to improve policy development at the National Disaster Management Authority (NDMA) and the Pakistan Meteorological Department (PMD). Precisely predicting the return period of extreme rainfall events in zones A and C is essential. This capacity for prediction allows for proactive preparedness for future events, protecting millions of priceless lives and human property. Authorities can put safety and security measures in place for communities against future disasters and reduce risks using accurate forecasts.

1.8 Objectives:

The study aims to achieve the following objectives:

- Evaluate the performance of Maximum Product Spacing in estimating the parameters of the Pearson Type III distribution for extreme rainfall events.

- Compare the results obtained using Maximum Product Spacing with other commonly used estimation methods.
- Assess the goodness-of-fit of the Pearson Type III distribution to the observed annual maximum rainfall series data.
- events. Provide insights into the practical implications of using Maximum Product Spacing on the accuracy of at-site frequency analysis for extreme rainfall

1.9 Relevance to national needs:

Particularly in the context of hydrological studies and flood management, the Annual Maximum Rainfall (AMRS) data utilizing Pearson Type III (PE3) distribution on Maximum Product Spacing (MPS) is highly relevant to national demands. Analyzing AMRS data with PE3 distribution helps accurately model extreme rainfall events.

Understanding the distribution and frequency of maximum rainfall is essential for designing effective flood defense systems, such as drainage systems, levees, and dams.

By improving the accuracy of early warning systems for extreme weather, this analysis helps to minimize the loss of life and property by facilitating prompt evacuations and preparations.

Data also facilitates better disaster recovery planning, aiding the speedy reconstruction and rehabilitation of communities following severe weather occurrences.

1.10 Content organization of thesis:

The remaining content of the thesis is organized below in the following order:

- In Chapter 2, we delve into the literature review, examining previous studies addressing similar problems. We discuss the several factors' data analysts have used in the past and elucidate the goals of the applied theories.
- Chapter 3 elaborates on the proposed methodology, providing a comprehensive discussion alongside the illustration of the available dataset and the study area.
- Chapter 4 elaborates on the performed analysis of annual maximum rainfall data.
- Chapter 5 presents the study's results, describes work that can be carried out in the Future, and forecasts and summarizes our research's results.

Chapter 2

2 Literature review

2.1 Previous Research:

There is a variety of literature on the analysis of the annual maximum rainfall using estimate techniques like maximum likelihood and L-moments. Still, there needs to be more information regarding the maximum product spacing utilizing at-site analysis. The investigations by Khan et al. [17] serve as benchmark studies for the start of this study project. The objective is to use relevant models to generate appropriate estimates that the relevant officials can use to influence planning and policymaking concerning agricultural management and water resources. Below is a summary of a few of the published studies:

In many applications, extreme values, especially annual maxima, are modeled using extreme value analysis. Another use of extreme value analysis, namely the Pearson Type-3 (PE3) distribution for modeling extreme values, is Khan et al. The impact of several estimate techniques (maximum product of spacing, maximum likelihood estimation, and L-moments) on fitting the PE3 distribution to extreme values. In addition to offering suggestions for fitting the PE3 distribution to extreme values based on sample size and form properties of the data, extreme value analysis aids in comprehending the features of extreme events [18].

Kim et al. estimated extreme rainfall events under nonstationary settings, considering how urbanization and climate change affect hydrological data. Based on nonstationary generalized extreme value (NS-GEV) distributions, it suggests a nonstationary population index flood (NS-

PIF) method that considers the nonstationary statistical characteristics at each location. To estimate extreme rainfall occurrences in South Korea, the NS-PIF method is assessed and contrasted with current approaches. Government officials may find the study's findings helpful in putting preventative measures and future advances into practice [19].

Hussain et al. performed a regional frequency analysis of Sindh, Pakistan's Annual Maximum Monthly Rainfall Totals (AMMRT), and suggested a linear regression model to calculate rainfall quantiles for ungauged locations. The analysis discovers that there is no serial correlation and no discordant sites in the data set at the specified sites, which is random. The quantiles and parameters of three distributions (GNO, PE3, and GPA) are computed once they satisfy the goodness-of-fit requirement. It is discovered that the constructed regression model which considers the site elevation is adequate for explaining the variation in rainfall. To get better results, the study recommends adding more data and site features and stresses the significance of employing formal tests for confirming regression estimates [20].

Qin et al. also aimed to research guiding national or international coastal engineers, managers, and planners in attaining consistent outcomes from extreme value evaluations. It contrasts the POT approach with the BM method for measuring extreme still water levels, emphasizing the need for indirect methods and problems with underestimating. The essential procedures, including parameter estimation and detrending approaches, for calculating extreme water level probability using the BM and POT methods are highlighted in the article. The theoretical distributions of GEV and GPD, which are frequently employed in extreme value investigations, are covered, along with a description of the functions and formulas needed. The POT approach is preferred for more accurate estimations of extreme still water levels, and a 1-year moving average is suggested for detrending [21].

Nawaz et al. primary emphasis is on applying L-moment analysis to regional rainfall frequency analysis in seven different northern Punjab, Pakistan locales. After the data series from these seven sites were examined, numerous important conclusions were made. First, there was no evidence of serial correlation, and the data were shown to be independent, uniformly distributed, and random. The area also appeared uniform, with not one of the seven locations showing discrepancies. There were several goodness-of-fit tests used to evaluate the suitability of various probability distributions for quantile estimation. The L-moment Ratio diagrams, the Z DIST statistic, and the Mean Absolute Deviation Index were all examined in these experiments. According to the findings, the best distributions to use when estimating quantiles were the Pearson Type III (PE3), Generalized Normal (GNO), and Generalized Extreme Value (GEV) [22].

Shahzadi et al. aimed to calculate the regional quantiles of rainfall at 23 distinct Pakistani sites. The L-moment-based index flood regional frequency analysis is used to carry out this estimation. The molds of independence, stationarity, and identical distribution are tested using a variety of techniques to guarantee the analysis's validity. Three separate regions are identified within the research area, mostly due to the features of highly elevated locations that experience heavier rainfall. GEV (Generalized Extreme Value), GNO (Generalized Normal), and GLO are the probability distributions that are most appropriate for calculating regional quantiles within these regions (Generalized Logistic). More precisely, it is found that GNO is the most reliable option for estimating regional quantiles for longer return times, while GEV works better at shorter return periods. Measures such as relative root mean square error, relative absolute Bias, and relative Bias are used to assess the accuracy of these regional estimations (RMSE). The study uses the regional L-moments algorithm described by Hosking and Wallis in 1997 to estimate the regional frequency distributions [23].

2.2 Classical Methods:

Forestieri et al. also noticed two important aspects: the necessity of accurate precipitation forecasts, especially for hydraulic engineering decision-making, and the increasing influence of extreme rainfall events because of climate change. In this study, we utilize regional frequency analysis (RFA) to analyze precipitation data that we have gathered from multiple rainfall stations located throughout Sicily, Italy. There is variation in the length of rainfall in these statistics. The RFA methodology classifies stations with similar precipitation patterns by using principal component analysis (PCA) and k-means clustering algorithms. We evaluate three probability distributions (LN3, GEV, TCEV) for their applicability in modeling extreme rainfall, and we use L-moments to estimate regional parameters. We utilize criteria such as relative Bias and relative root-mean-square error to assess the accuracy of growth curves (RMSE). Our results show that the LN3 distribution performs better than the others over longer return times. Considering this, this study offers an updated reference for gauging extreme precipitation levels in Sicily, which is essential for creating depth–duration–frequency (DDF) curves. [24].

Rajeevan et al., using a dataset that covers 104 years of gridded daily rainfall data and the NOAA Extended Reconstructed SST (ERSST) Version 2 dataset, investigate the changes and patterns of extreme rainfall events in India. A strong relationship has been found by the researchers between the equatorial Indian Ocean's sea surface temperatures (SST) and the frequency of very heavy rainfall (VHR) episodes in central India. The study indicates an increase in the frequency of VHR episodes and the corresponding flood danger in central India in the context of continuing global warming. From 1901 to 2004, there was an increase in the number of VHR events, according to the data. This increase was especially noticeable after the mid-1970s. Geographically, the west coast, northeastern India, and central India are the areas where the monsoon season's intense rainfall events are concentrated. [25].

Chapter 3

3. Methodology

In this chapter, we have discussed the methodology used to perform this research. Figure 3.1 gives the overall outline. The details of each step of the methodology are given in the following sections.

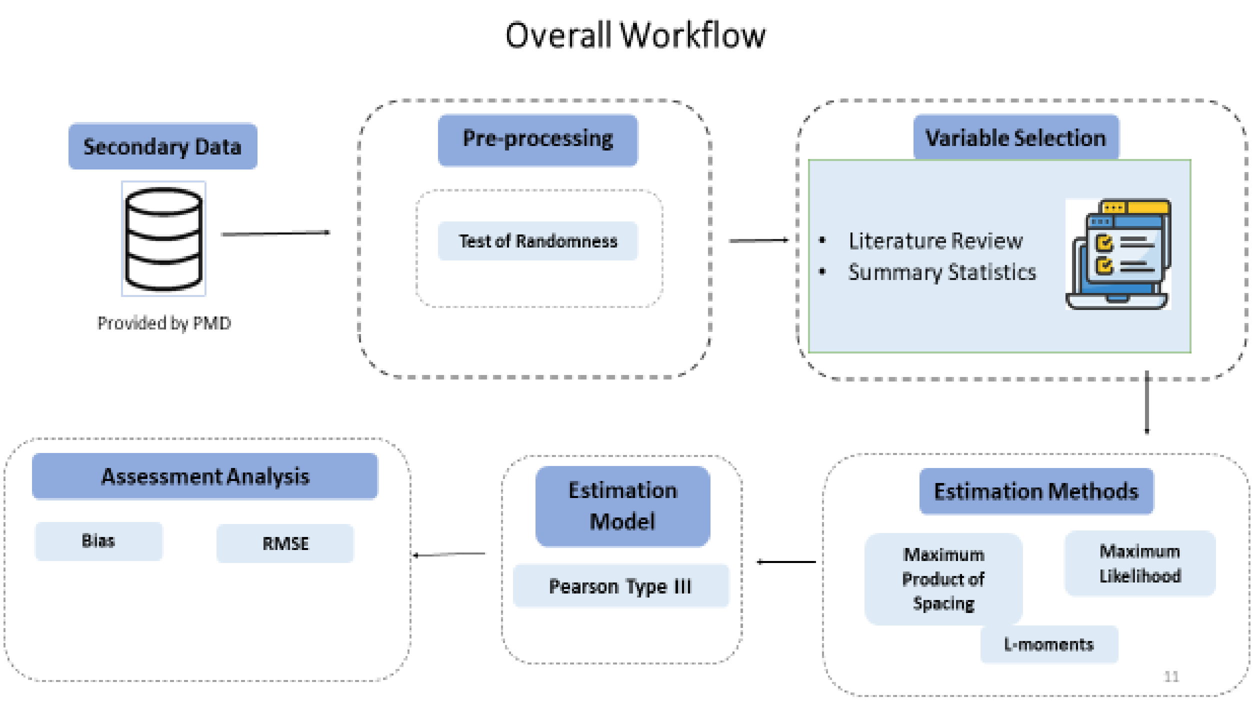

Fig 3.1: Flowchart of methodology steps

3.1 At-site Analysis:

To achieve the research goals, at-site frequency analysis is used to look at rainfall data directly at a particular spot. To comprehend the distinct patterns, intensities, and behaviors of rainfall occurrences at that location, this method entails studying annual or daily precipitation data. It also entails researching a range of environmental factors, including soil properties, water quality, and air quality, all of which are essential to comprehending the specific effects of heavy rainfall in each area. At-site analysis, which focuses on the individual characteristics of a given area, offers crucial insights for site-specific decision-making and management. This allows for precise and efficient planning and mitigation techniques that are customized to the site's particular requirements.

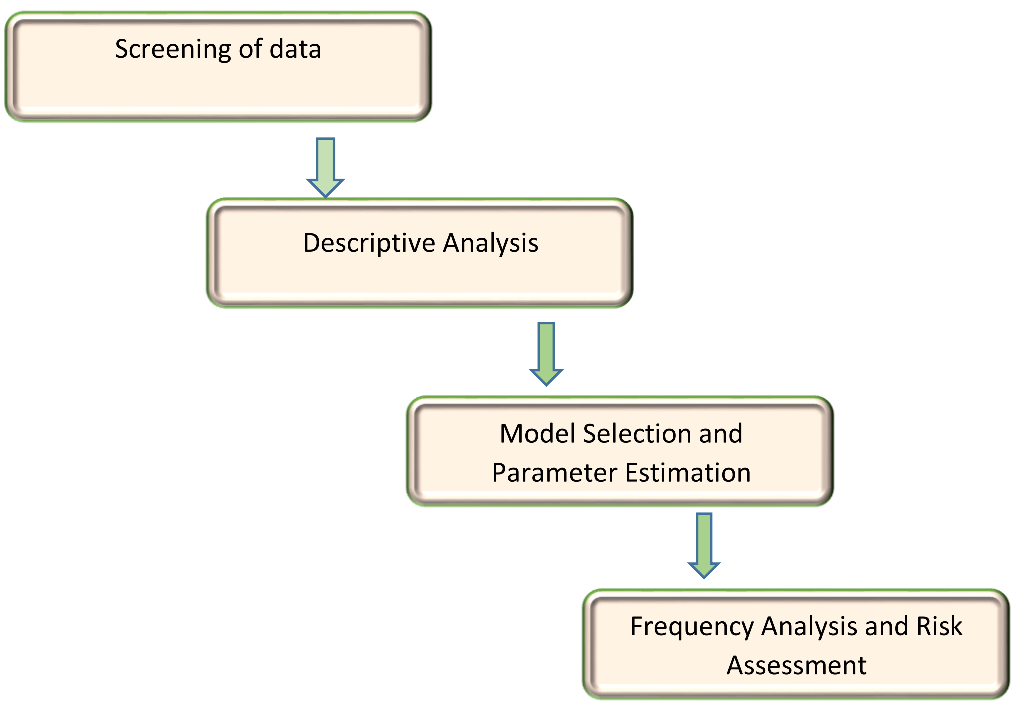

Fig 3.2: Methodology steps for at-site frequency analysis

3.2 Initial screening of data:

Initial screening is done to check whether the given data is suitable for at-site frequency analysis of rainfall. According to Hosking and Wallis (1997) and [26], useful information can be obtained by comparing MPS on AMRS with available evidence of estimation methods.

3.3 Statistical Analysis:

Three different types of data (annual rainfall, yearly daily maximum rainfall, and annual monthly maximum rainfall) are subjected to descriptive statistical analysis. Organization and summarization of large-scale data are the focus of descriptive statistics. The raw data is presented via tables, graphs, and numbers (Ott and Longnecker, 2010). Using rainfall data, we conducted descriptive statistical analyses to investigate its skewness, peakedness, central tendency (mean, median, and mode), variability (standard deviation), and symmetry (kurtosis) [27]. The following lists the numerous statistical moments that were employed in this investigation:

First moment (mean):

$$\bar{y} = \Sigma\, yi\, /n \qquad 3.1$$

Second moment (variance):

$$S\,2 = \Sigma\,(yi - \bar{y})\,2/ \qquad 3.2$$

$n - 1$

Third moment (skewness):

$$g = n\,\Sigma\,(yi - \bar{y})\,3/\,(n - 1)(n - 2)s\,3 \quad 3.3$$

Fourth moment (kurtosis):

$$\text{Ɣ}2 = [(\mu 4)(\mu 2)2] - 3 \quad 3.4$$

3.4 Descriptive analysis:

To analyze the data, extreme value analysis will be used for our results since data contains the annual maximum value of precipitation. Therefore, we will adopt the block maxima approach to check the positive and negative trend of maximum values. The generalized extreme value distribution model described below will be used to verify the results above. In addition, statistical tests will be performed to ascertain which model best fits our data. The maximum likelihood (MLE), maximum product spacing (MPS), and L-moment (LM) approaches are some of these assessments. In this evaluation, the data's location, scale, and form are considered when estimating Bias and the root mean square error (RMSE).

3.5 Model of Estimation:

Pearson type III distribution (PE3)

The Pearson type III distribution (PE3) is one of the seven varieties of probability distributions that Pearson proposed. To be mostly applied in survival analysis in biostatistics. As a parent model of the regular Gamma distribution, this distribution is sometimes called the three-parameter generalized Gamma distribution. It is one of the distributions that is most likely to describe skewed or asymmetric data. The distribution of Pearson type III has two exceptional situations. When the shape parameter of the first example becomes zero, it reduces to a normal distribution with all its characteristics. The second instance of Pearson type III pertains to a reduction in the exponential distribution when the shape parameter is set to two. It is a widely used distribution for forecasting and modeling hydrology. Not only are extreme events, such as strong floods and excess rainfall, typically not symmetrical around their means, but the logarithm makes things much easier to understand when the underlying variable is used.

Generalized Extreme Value Distribution:

To simulate a random variable's extreme values (high or low), statisticians employ the Generalized Extreme Value (GEV) distribution, a probability distribution [28]. It is typically utilized in fields where severe events are of interest, like hydrology, meteorology, finance, and environmental research. The probability density function (pdf) of the GEV distribution is given by:

$$f(x) = \sigma 1[1 + \xi(\sigma x - \mu)] - 1/\xi - 1\exp\{-[1 + \xi(\sigma x - \mu)] - 1/\xi\} \quad 3.5$$

Where:

- x is the random variable.
- μ is the location parameter, which shifts the distribution along the x-axis.
- σ is the scale parameter, which controls the spread or variability of the distribution.
- ξ is the shape parameter. It determines the shape of the distribution:
- When ξ >0, the distribution is right-skewed (Type II extreme value distribution).

- When ξ <0, the distribution is left-skewed (Type III extreme value distribution).

When ξ =0, when modeling the maximum (or minimum) of a collection of samples from a random variable, the distribution becomes the Gumbel distribution. Gumbel, Frechet, and Weibull are the three extreme value distribution types included in the GEV distribution. The type of selection is based on the shape parameter's value. ξ

Because it describes the tail behavior of distributions, the GEV distribution is helpful for modeling extreme events. It is frequently used in risk assessment, insurance, and environmental research to analyze uncommon but severe catastrophes, including floods, droughts, extremely elevated temperatures, or financial market crises.

3.6 Method of estimation:

L-Moment:

Statistical measurements called "L-moments" are employed in hydrology, meteorology, and related domains to analyze data distributions. They have some benefits over more conventional moments like the mean, variance, skewness, and kurtosis, especially when analyzing data with heavy tails or skewed distributions. A distribution's probability-weighted moments (PWMs) are the source of the L-moments.

Three L-moments exist:

1. L1 (mean): Equivalent to the mean of the distribution.
2. L2 (variance): Equivalent to half the variance of the distribution.

3. L3 (skewness): Measures the asymmetry of the distribution.
4. L4 (kurtosis): Measures the tail heaviness of the distribution.

The Generalized Pareto Distribution (GPD) and the Generalized Extreme Value (GEV) distribution, which are frequently used in the modeling of extreme events like floods and droughts, are two examples of distributions for which L-moments are frequently used to estimate parameters.

The advantage of L-moments is superior to regular moments because they are typically more resilient and unaffected by outliers. This property makes them especially helpful for studying data with skewed distributions or heavy tails. Furthermore, L-moments are useful for practical applications since they may be promptly estimated from data without necessitating an understanding of the underlying distribution.

Maximum Likelihood:

The statistical technique known as Maximum Likelihood Estimation (MLE) is flexible and can be used with small, big, or moderate-sized data sets. MLE is determined not by the data quantity but by its distributional assumptions, regardless of its size. MLE can estimate parameters for tiny data sets, however, because of possible limits in precision and accuracy, care should be used. For moderate-sized data sets, MLE usually yields more trustworthy estimates by balancing computational viability and accuracy. On the other hand, MLE scales well and converges to true parameter values with a large sample size. Therefore, it is still effective for large data sets, even with computing difficulties [29]. MLE is frequently utilized for parameter estimation in a variety of domains, providing flexibility and dependability for a range of data sizes.

Mathematically, the formula for MLE can be expressed as:

$$\theta\text{\textasciicircum}MLE = \text{argmax}\theta(L(\theta \mid x)) \qquad 3.6$$

Where,

- θ^MLE represents the maximum likelihood estimate of the parameter vector θ.
- L (θ|x) denotes the likelihood function, the joint probability density function (pdf) of the observed data x given the parameter vector θ.

Maximum product spacing:

Irrespective of the underlying distributional shape—normal, skewed, or otherwise the Maximum Product Spacing (MPS) approach is especially well-suited for examining data sets that display severe events. However, it works particularly well for distributions with constrained support, which might be difficult for more conventional techniques like Maximum Likelihood Estimation (MLE). Even with small sample sizes, MPS is robust in terms of sample size and can produce accurate parameter estimates and confidence ranges. This means that it can be used in a variety of contexts, such as the analysis of annual maximum rainfall in hydrology or severe temperature occurrences in climatology, as well as small-scale research with sparse data sets and larger-scale analyses with abundant data sets.

This can be calculated as:

$$\theta\text{\textasciicircum}MPS = \text{argmax}\theta(\sum i = 1n + 1\log(F(X(i);\theta) - F(X(i-1);\theta))) \qquad 3.7$$

Where,

- θ^ MPS is the estimated parameter value obtained using MPS.

- $F(X_{(i)};\theta)$ represents the cumulative probability of the i-th ordered data point based on the parameter θ.
- $X_{(i)}$ denotes the i-th ordered data point.
- n is the sample size.

The essence of MPS is summed up in this formula, which maximizes the product of spacing between order statistics to estimate parameters with accuracy. This makes MPS a useful tool in a variety of statistical investigations, particularly when working with extreme occurrences or small data sets.

Chapter 4

4 Results and Discussions

We implemented the proposed methodology of At-site frequency analysis based on Maximum product spacing, Maximum Likelihood, and L-moments using the step-by-step processes that are shown in Figure. In this section, the results are presented in the same sequence, along with their statistical reasoning and discussion.

4.1 Study Area and Available Data

In terms of latitude, Pakistan sits between 23 degrees 35 minutes and 37 degrees 05 minutes north and between 60 degrees 50 minutes and 77 degrees 50 minutes east. Its boundaries meet the Hindukush Mountains in the north, and it stretches from the Arabian Sea in the south to the Pamirs in the northwest [30]. Pakistan's terrain demonstrates dramatic temperature variations over time, which characterize the country's climate during the many seasons. Pakistan saw a rise in precipitation from 223.41 mm (about 8.8 in) in 2021 to 442.88 mm in 2022 (https://tradingeconomics.com/pakistan/precipitation). Pakistan saw 282.50 mm (about 11.12 in) of precipitation on average between 1901 and 2022; the country's record low was 181.50 mm (about 7.15 in) in 2018, and its highest point was 442.88 mm in 2022.

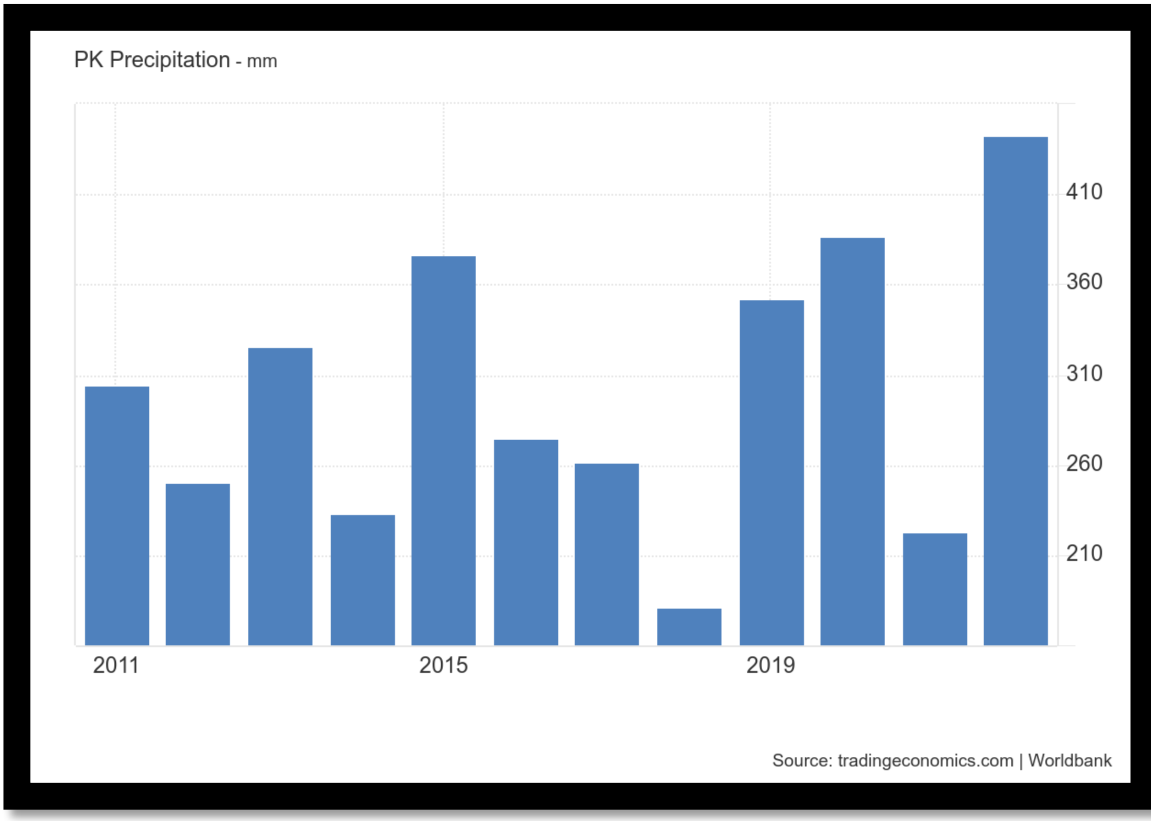

Fig 4.1 Rainfall record from 2011-2022

Pakistan has four different seasons, which are as follows:

1) Winter (December to March)
2) Summer (April to June)
3) Monsson season (July to September)
4) Post Monsoon season (October to November)

There are restrictions in detecting rainfall trends because the distribution of rainfall lacks a distinct altitudinal trend. To address this, a thorough dataset from 14 stations spread over the nation for 36 years (1980–2015) from east to west and from the far north to the far south was used. The five microclimatic zones that PMD separated these stations into are designated as A, B, C, D, and E in Figure 4.2, along with their corresponding latitudinal extend (https://www.researchgate.net/figure/Map-of-Pakistan-Showing-different-climate-zones-of-Pakistan-along-with-their-latitude-and_fig1_276060366). To provide an extensive picture of the nation's climate, stations were chosen according to their latitudinal position, height above sea level, duration of recording period, data completeness, and record dependability [31].

4.2 Zone A

Zone A comprises stations with cold climates and high mountains in northern Pakistan. These stations are Chitral, Gilgit, Muzaffarabad, Said-u-Sharif, Skardu, Astor, Dir, Chilas Parachinar, and Kakul. They are mostly hill stations located between 34 N and 38 N in the Himalaya, Hindukush, and Koh-e-Sufaid mountain ranges, as described in the table below [32].

4.3 Zone C

The climate is cold in winter and hot in summer. Most of them are mountainous stations with high elevations from the mean sea level and cover an area between 27 N to 32 N and 64 E to 70 E, as described in the table below. Stations included in this zone are Quetta, Zhob, Kalat and Khuzdar [33].

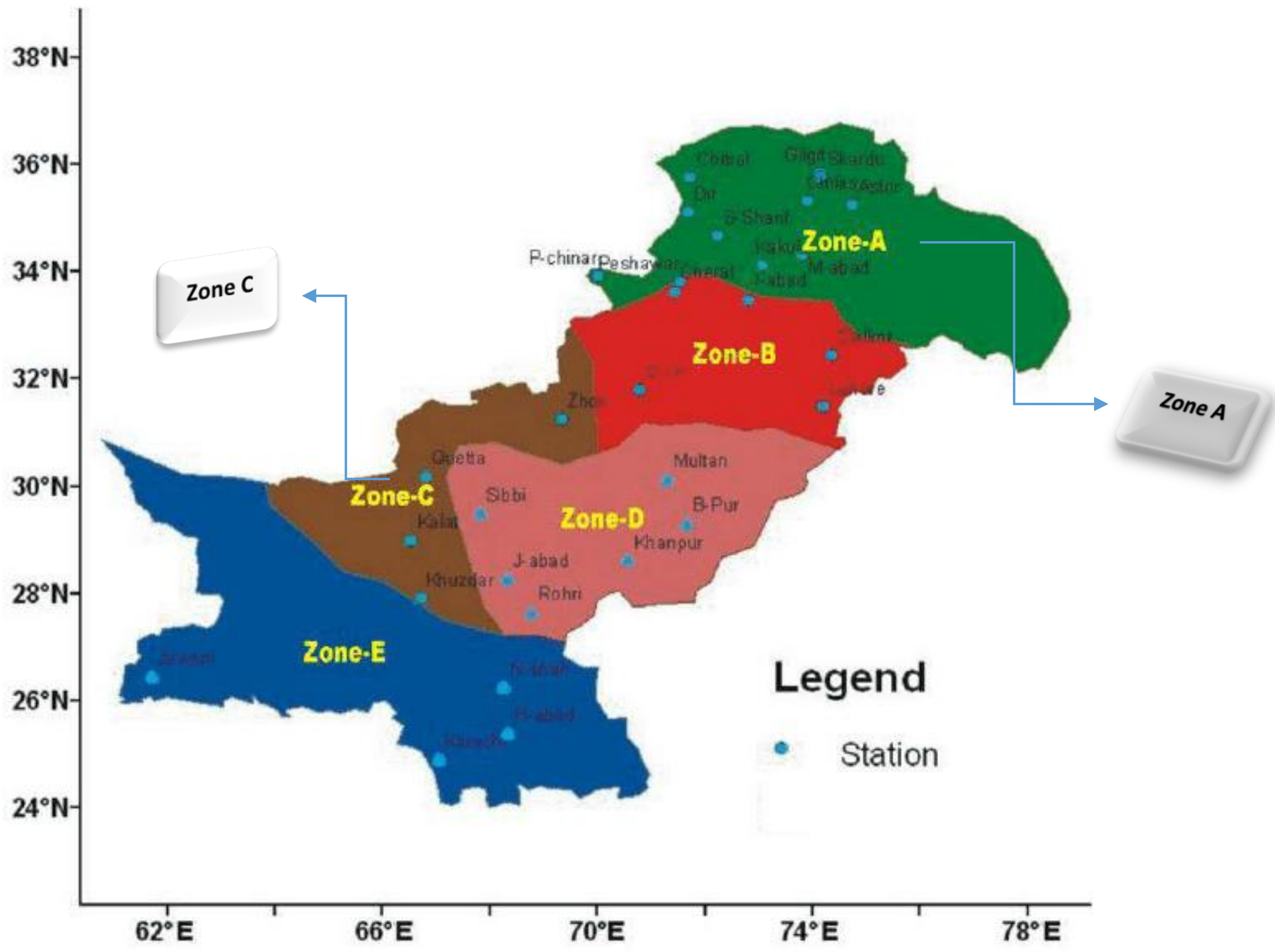

Fig 4.2: Map of Pakistan Showing different climate zones of Pakistan along with their latitude and longitude.

Table 4.1 Geographical stations of Zone A

Sr.No	Sites	Longitude (E)	Latitude (N)
1	GILGIT	74.31°	35.92°
2	SKARDU	75.63°	35.30°
3	ASTORE	74.84°	35.36°
4	MUZAFFARABAD	73.47°	34.36°
5	CHITRAL	71.77°	35.77°
6	GUPIS	74.51°	35.97°
7	DROSH	71.80°	35.57°
8	BUNJI	74.38°	35.38°
9	BALAKOT	73.35°	34.54°
10	GHARIDUPATTA	73.61°	34.22°
11	KAKUL	73.16°	34.21°

ZONE C

Table 4.2 Geographical stations of Zone C

Sr.NO	Sites	Longitude	Latitude
1	QUETTA	66.99°	30.18°
2	KALAT	66.59°	29.03°
3	DALBANDIN	64.42°	28.88°

4.4 Empirical Analysis:

The performance of three estimation methods, L-Moments, Maximum Product Spacing, and Maximum Likelihood has been tested using Pearson type III distribution. Annual maximum rainfall series data (AMRS) of zones A and C is used for fitting distribution on Pearson type III. PMD (Pakistan Meteorological Department) has divided the rainfall stations of Pakistan into five zones [34]. Zone A and C are selected based on trends and tendencies of scale and shape characteristics based on their skewness and kurtosis values. None of the established studies so far, to the best of the authors' knowledge, performed the at-site frequency analysis using AMRS using Pearson type III distribution of zones A and C.

To observe the general trends and tendencies of AMRS at each site, a few descriptive measures are calculated and presented in Table 4.3. The information reveals that for the sites of Zone A, the sample size is fixed; that is 36 years (1980-2015) of data is used for descriptive analysis. The shape of the data series for stations of zone A is positively skewed, with the value of skewness ranging from 0.36 to 1.37 and leptokurtic behavior as kurtosis values are showing more spread ranging from 0.002 to 3.24. These descriptive statistics indicate that the trends and tendencies of AMRS at zone A stations differ significantly, particularly in distribution shape. Consequently, this data is well-suited for evaluating the effectiveness of various estimation methods for fitting the PE3 distribution.

Table 4.3: Descriptive statistics of AMRS of Zone A.

S. No.	Site Name	Min	Max	Mean	Standard Deviation	Skewness	Kurtosis
1	GILGIT	13.08	37.15	21.37	5.19	0.854	0.943
2	SKARDU	13.20	36.28	21.05	5.82	0.798	0.002
3	ASTORE	13.82	55.23	26.21	8.39	1.338	2.882
4	MUZAFFARABAD	30.55	153.98	74.463	31.43	1.091	0.718
5	CHITRAL	18.69	51.67	29.67	7.60	0.829	0.529
6	GUPIS	16.55	44.84	27.65	6.32	0.365	0.181
7	DROSH	28.48	79.38	47.17	12.40	0.752	0.213
8	BUNJI	16.24	78.97	33.46	12.73	1.378	3.242
9	BALAKOT	30.55	153.98	74.46	31.43	1.091	0.718

10	GHARIDUPATTA	32.15	184.27	75.72	35.78	1.121	1.188
11	KAKUL	32.15	184.27	75.72	35.78	1.121	1.188
	Average	23.23	96.36	46.08	17.53	0.976	1.07

Note: Here, n is the number of observations, Min and Max are the minimum and maximum values in the data series, and Skewness and Kurtosis are moment's measures of skewness of kurtosis.

Table 4.4: Descriptive statistics of AMRS of Zone C.

S. No.	SITES	Min	Max	Mean	Standard Deviation	Skewness	Kurtosis
1	QUETTA	4.74	28.52	15.08	6.15	0.801	-0.099
2	KALAT	5.37	61.30	15.99	11.85	2.286	6.239
3	DALBANDIN	4.57	32.04	12.70	7.87	1.224	0.356
	Average	4.89	40.62	14.59	8.62	1.43	2.16

Notes: Here is the number of observations, Min and Max are the minimum and maximum values in the data series, and Skewness and Kurtosis are moment's measures of kurtosis's skewness.

The shape of the data series for stations of zone C is positively skewed, with a skewness value ranging from 0.801 to 2.28. Leptokurtic behavior is also evident, as kurtosis values show more spread, ranging from -0.099 to 6.239.

4.5 Time series plots:

The time series graphs show trends and variations over time. They are drawn for the descriptive analysis of all 14 zone A and C stations under the study from Fig 4.3 to 4.16. The time of 36 years is plotted on the x-axis, and annual maximum rainfall values are plotted on the y-axis. The mean line is drawn to see the trends fluctuating across the mean.

In the time series graphs of zone A for Astor, the peak value of 56mm (about 2.2 in) occurred in 1991. The average value is 26.5mm (about 1.04 in), as shown in Figure 4.3. Chitral showed a peak value of 53.5mm (about 2.11 in) in 1998 with an average value of 29.55mm (about 1.16 in), as shown in Fig 4.4. Gilgit showed a peak value of 37mm (about 1.46 in) occurred in 1993 with an average value of 22mm (about 0.87 in) as shown in Fig 4.5. Muzaffarabad showed a peak value of 155mm (about 6.1 in) in 2002 with an average value of 74mm (about 2.91 in), as shown in Fig 4.6. Skardu showed a peak value of 37mm (about 1.46 in) in 1992 with an average value of 22mm (about 0.87 in), as shown in Fig 4.7. Kakul showed a peak value of 180mm (about 7.09 in) in 2001 with an average value of 75mm (about 2.95 in), as shown in Fig 4.8. Gupis showed a peak value of 45mm (about 1.77 in) in 1994 with an average value of 27.5mm (about 1.08 in), as shown in Fig 4.9. Drosh showed a peak value of 79.5mm (about 3.13 in) in 2008 with an average value of 47mm (about 1.85 in), as shown in Fig 4.10. Bunji showed a peak value of 79.5mm (about 3.13 in) in 2001 with an average value of 32mm (about 1.26 in), as shown in Fig 4.11. Balakot showed a peak value of 54mm (about 2.13 in) in 2002 with an average value of 74mm (about 2.91 in), as shown in Fig 4.12. GhariDupatta showed a peak value of 180mm (about 7.09 in) in 2002 with an average value of 75mm (about 2.95 in), as shown in Fig 4.13.

For zone C, the time series graph for Quetta showed a peak value of 28mm (about 1.1 in) in 2008 with an average value of 15mm (about 0.59 in), as shown in Fig 4.14. Kalat showed a peak value of 61.5mm (about 2.42 in) in 2012 with an average value of 17.5mm (about 0.69 in), as shown in Fig 4.15. Dalbaddin showed a peak value of 33mm (about 1.3 in) in 1986 with an average value of 12.5mm (about 0.49 in), as shown in Fig 4.16.

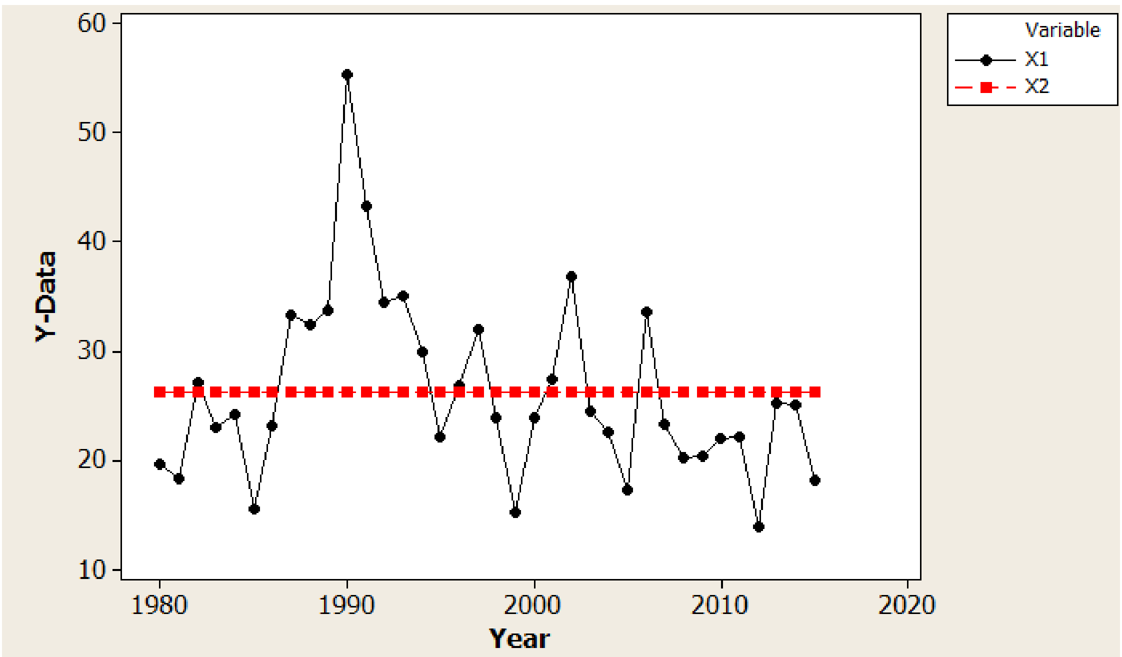

Fig 4.3: Time series plot of the data for the site of Astor

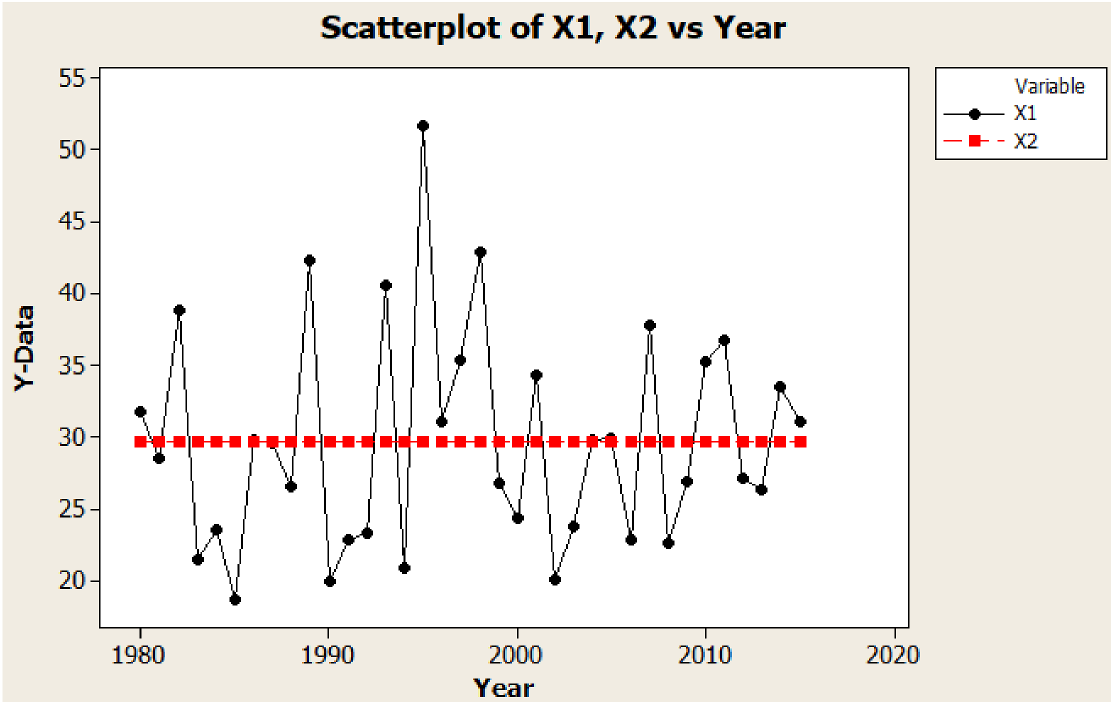

Fig 4.4: Time series plot of the data for the site of Chitral

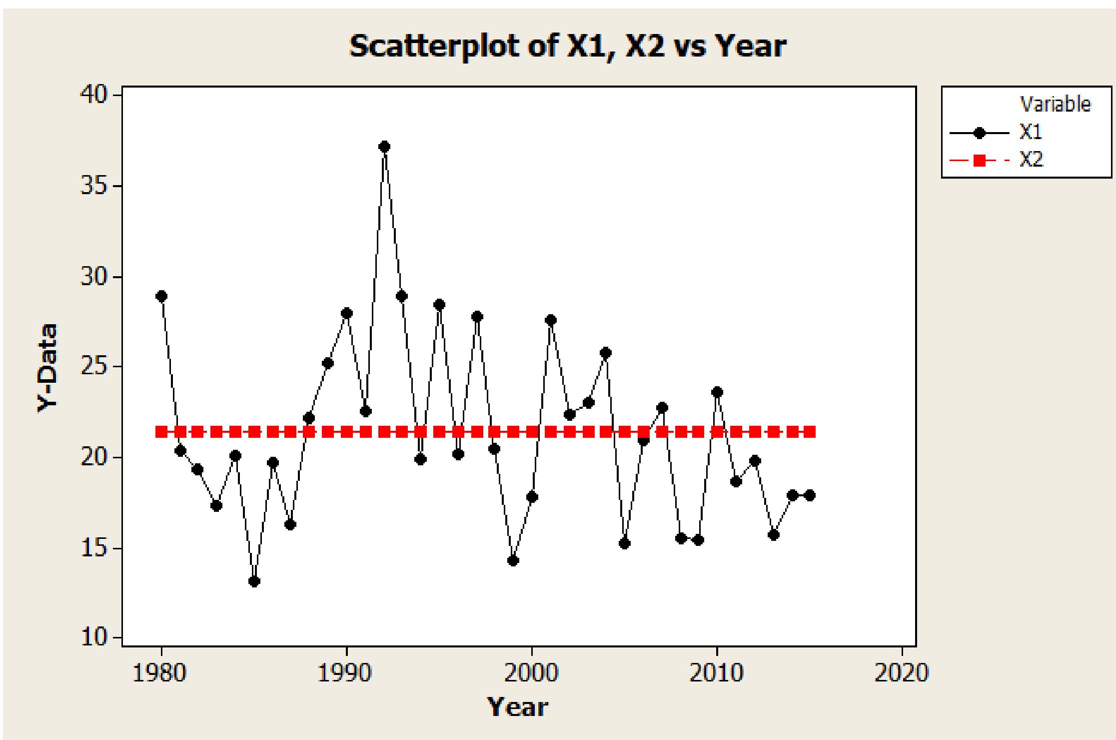

Fig 4.5: Time series plot of the data for the site of Gilgit

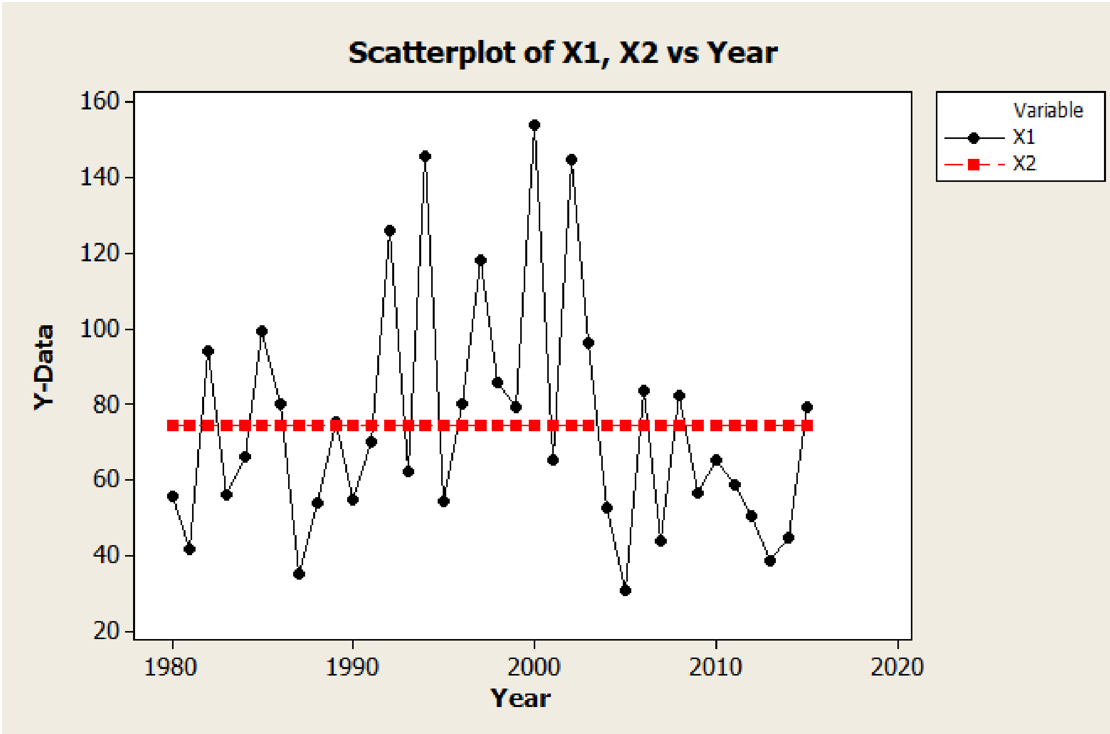

Fig 4.6: Time series plot of the data for the site of Muzaffarabad

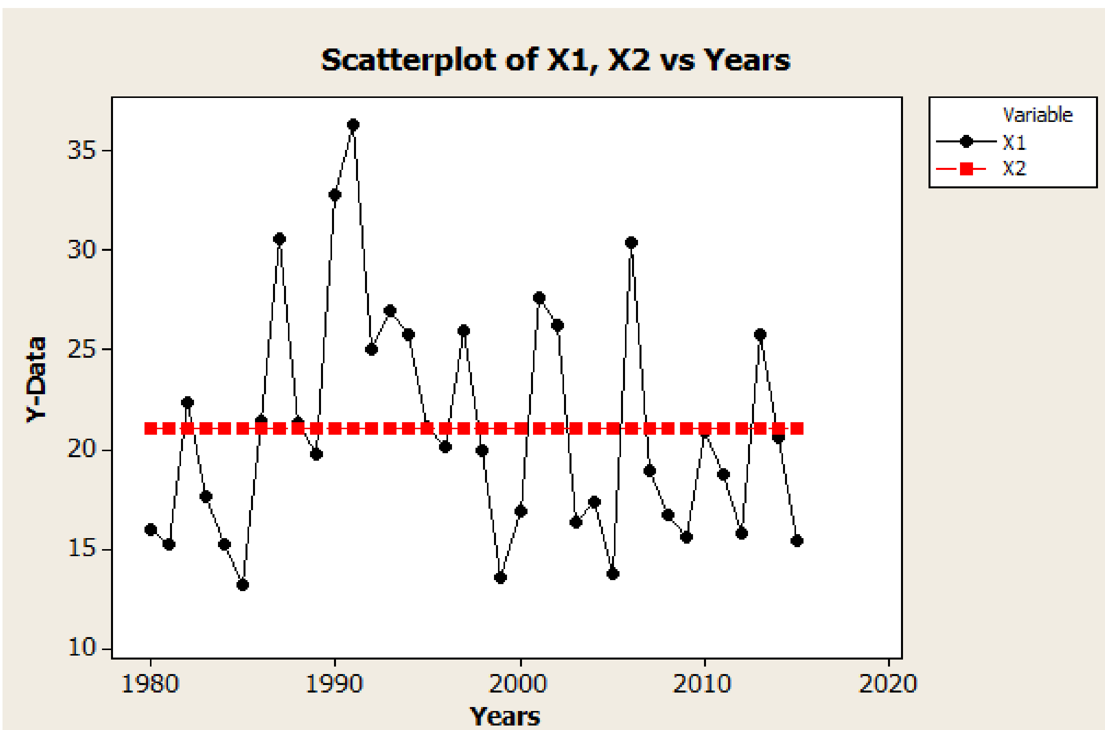

Fig 4.7: Time series plot of the data for the site of Skardu

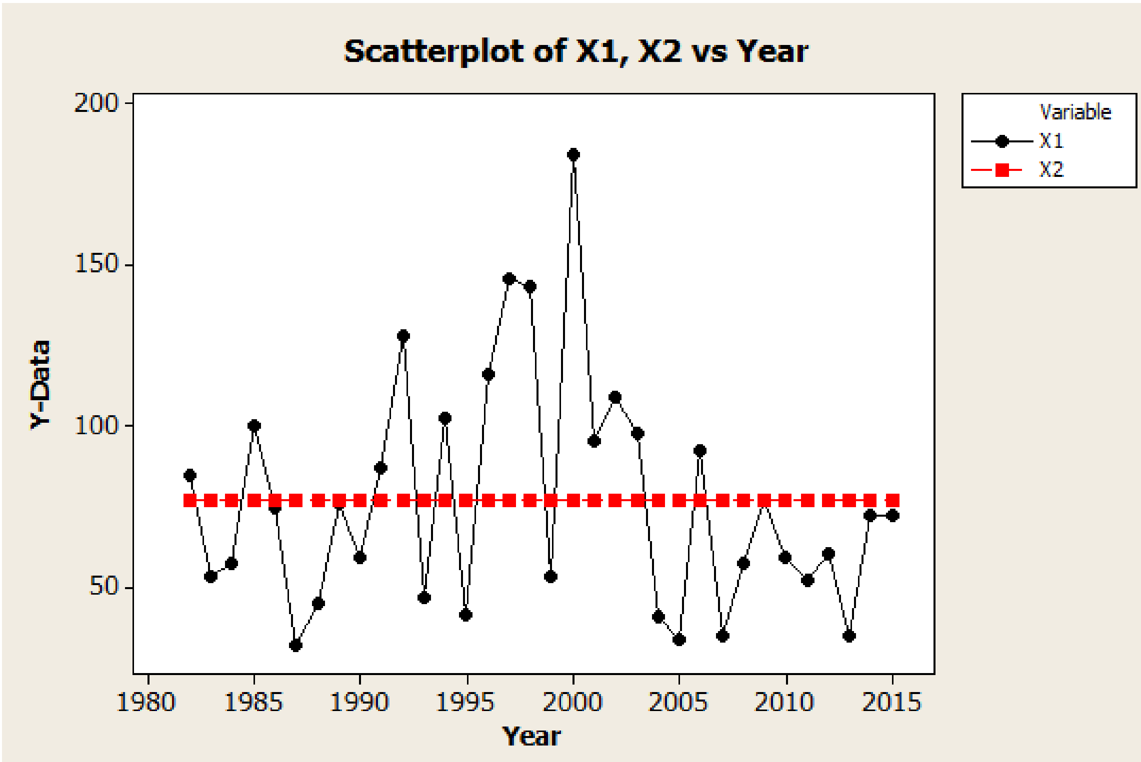

Fig 4.8: Time series plot of the data for the site of Kakul

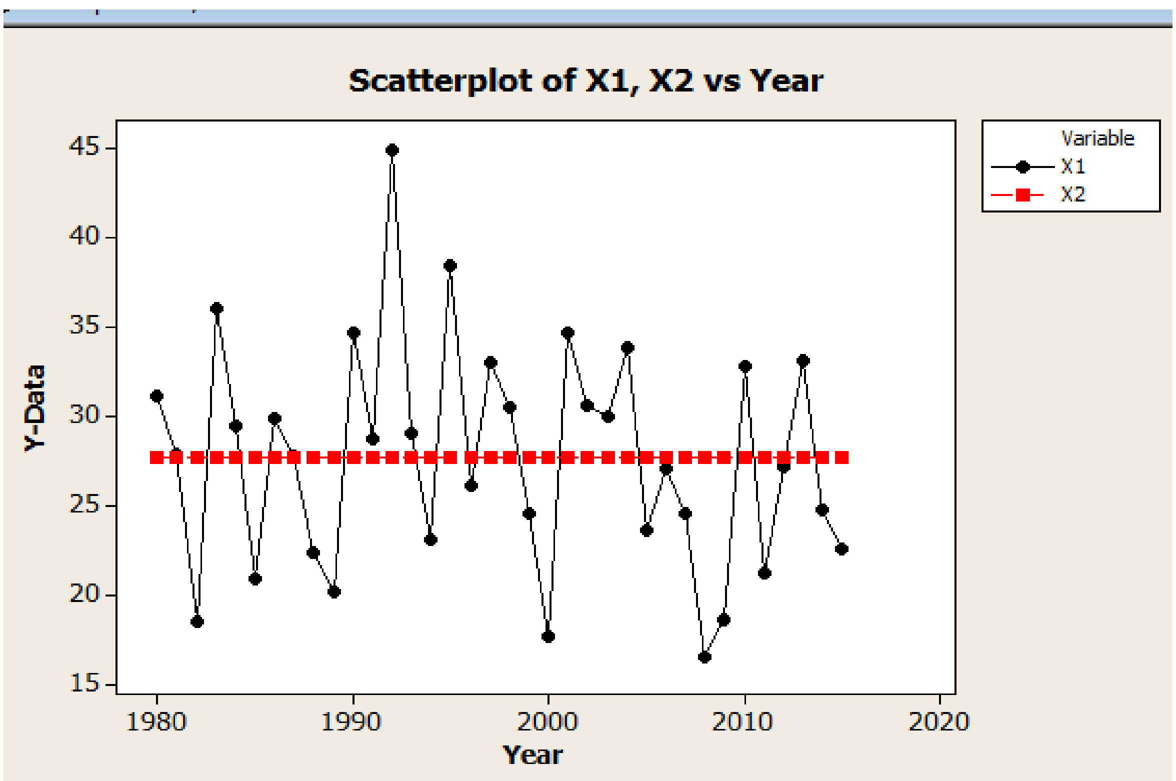

Fig 4.9: Time series plot of the data for the site of Gupis

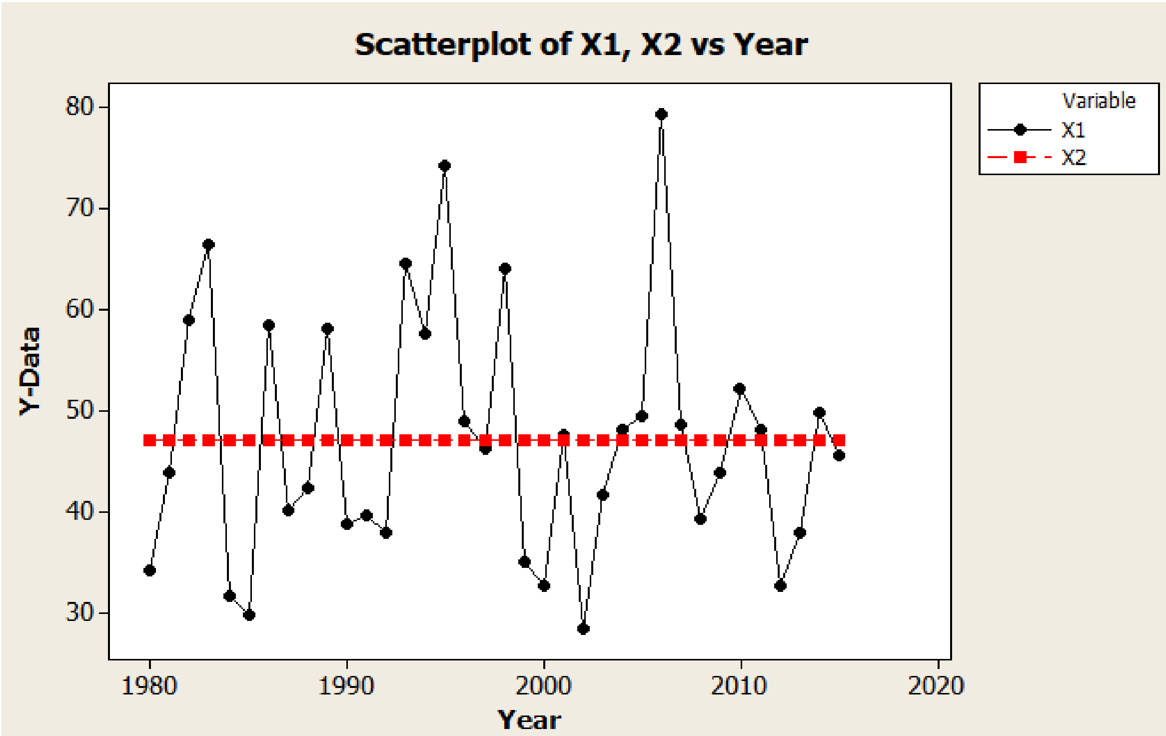

Fig 4.10: Time series plot of the data for the site of Drosh

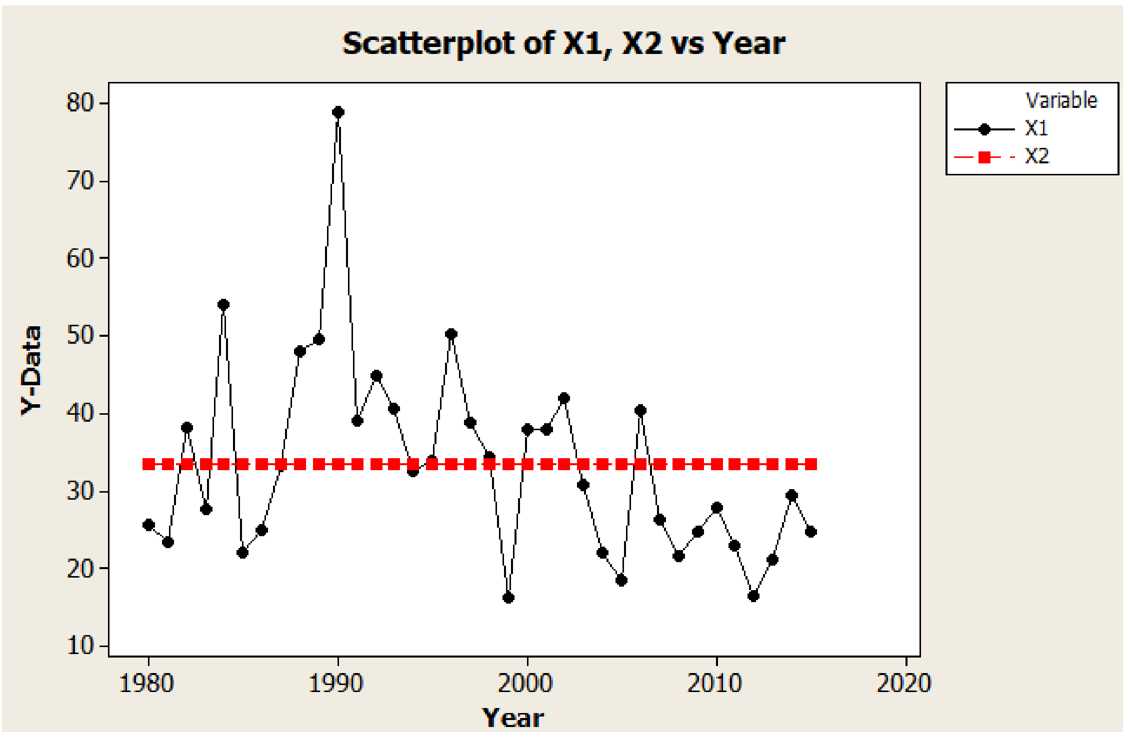

Fig 4.11: Time series plot of the data for the site of Bunji

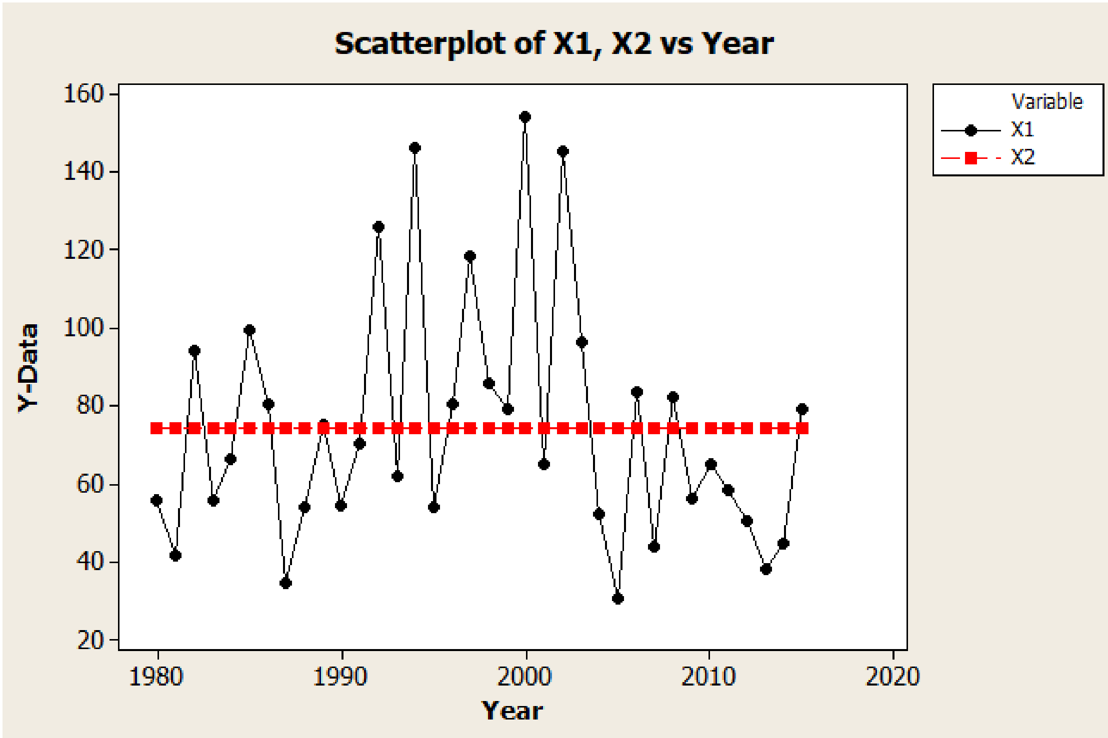

Fig 4.12: Time series plot of the data for the site of Balakot

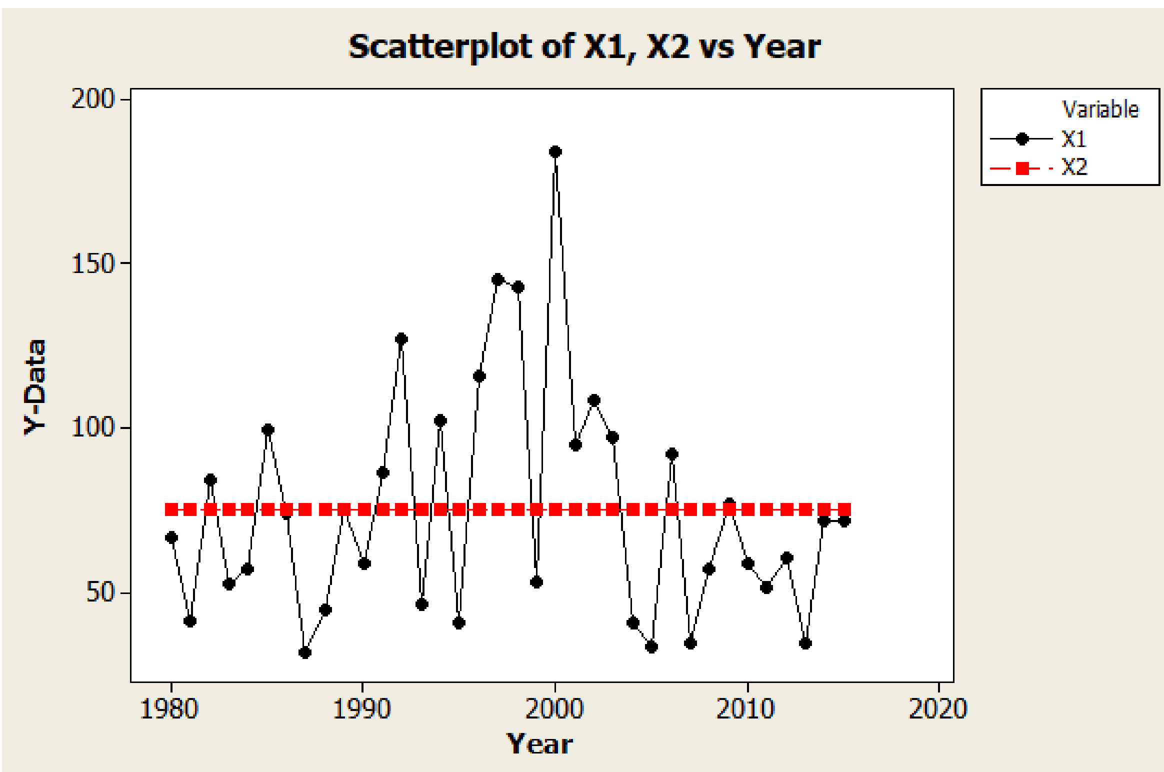

Fig 4.13: Time series plot of the data for the site of GhariDupatta

For Zone C:

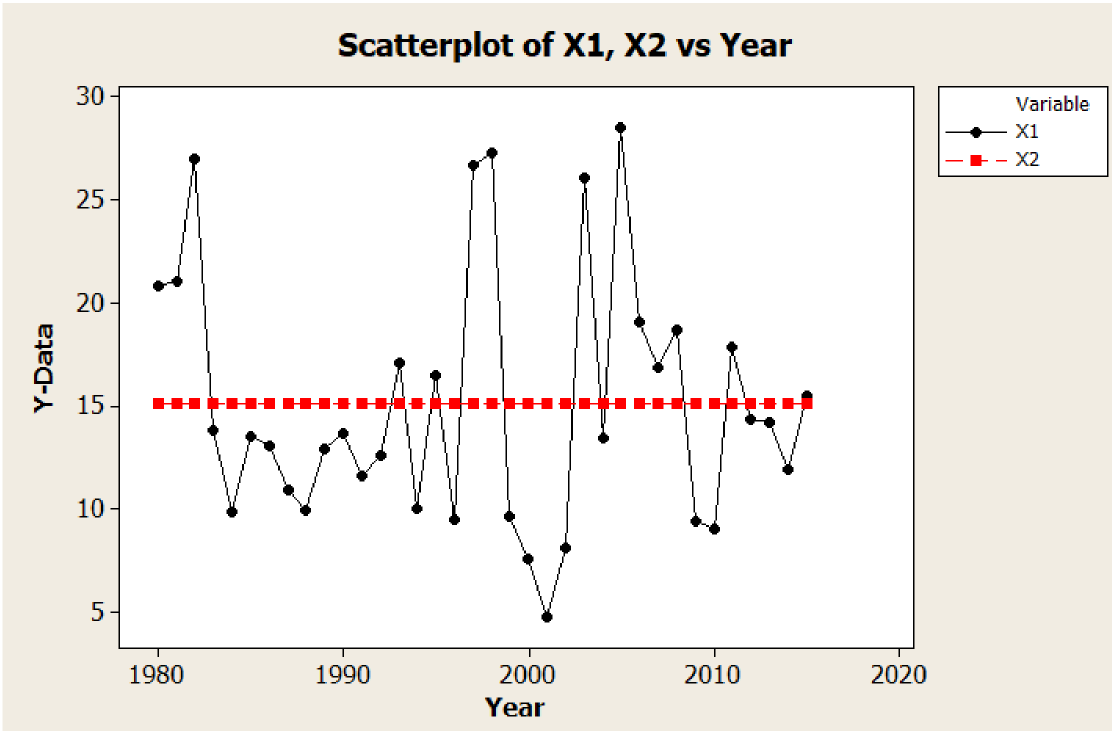

Fig 4.14: Time series plot of the data for the site of Quetta

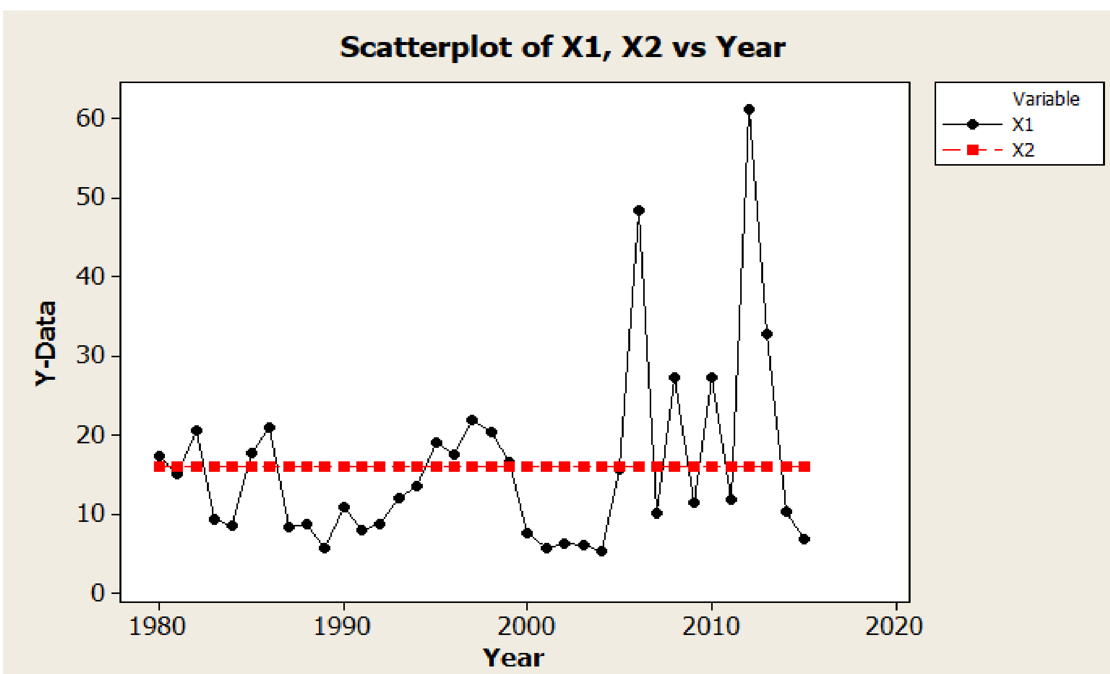

Fig 4.15: Time series plot of the data for the site of Kalat

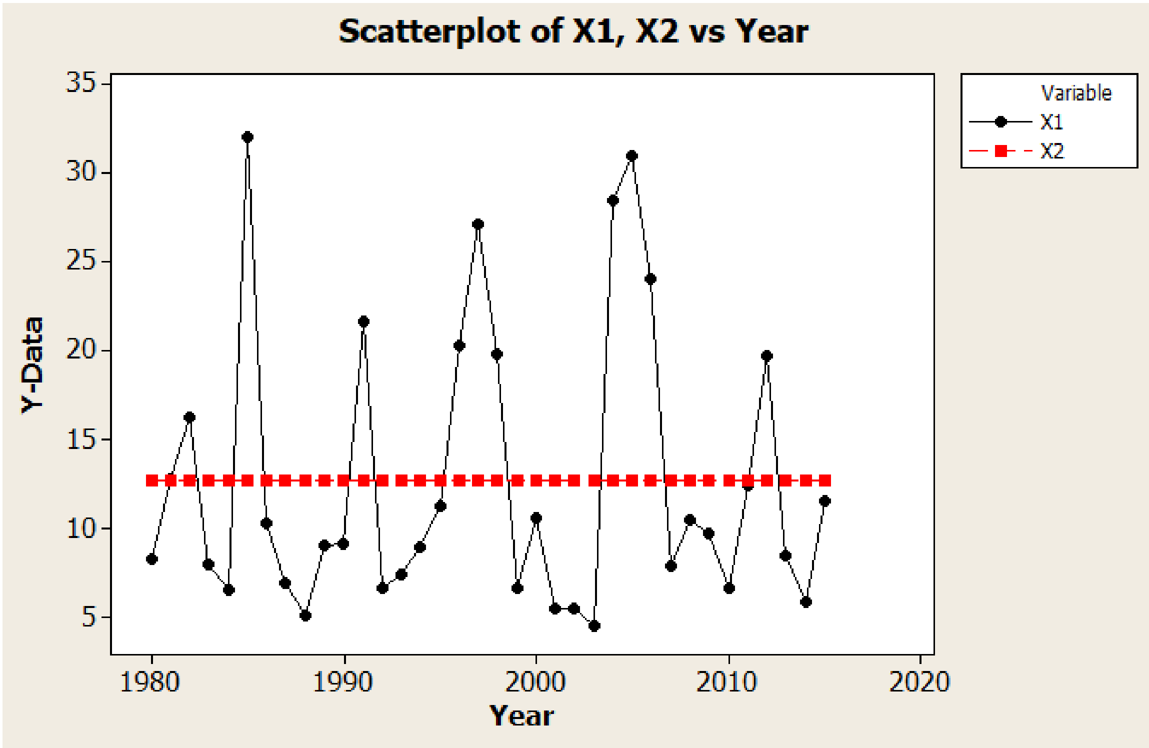

Fig 4.16: Time series plot of the data for the site of Dalbaddin

Table 4.5 Values of Bias and RMSE of the parameters estimated through LM, MLE and MPS for Zone A

	ESTIMATION METHODS		LM			MLE			MPS		
	SITES	PARAMETER	LOCATION (Mu)	SCALE (Sigma)	SHAPE (Gamma)	LOCATION	SCALE	SHAPE	LOCATION	SCALE	SHAPE
1	ASTORE	ESTIMATE	26.21	8.36	1.24	26.21	8.22	1.13	26.52	8.97	1.10
		RMSE	1.34	1.42	0.48	1.34	1.46	0.45	1.56	1.62	0.39
		BIAS	0.0033	0.0428	0.0287	0.0270	0.1072	0.1166	0.2783	0.6954	0.0350
2	CHITRAL	ESTIMATE	29.68	7.82	1.02	29.68	7.99	1.33	29.95	8.46	1.22
		RMSE	1.27	1.17	0.44	1.37	1.66	0.42	1.43	1.58	0.38
		BIAS	0.0381	0.0260	0.0283	0.0038	0.2946	0.1373	0.3097	0.6819	0.0339
3	GILGIT	ESTIMATE	21.37	5.26	0.91	21.37	5.21	1.02	21.52	5.63	0.94
		RMSE	0.87	0.81	0.48	0.85	0.91	0.46	0.98	0.96	0.44

		BIAS	0.0169	0.0494	0.0472	0.0032	0.0508	0.0940	0.1420	0.4639	0.0441
4	MUZAFFARABAD	ESTIMATE	74.46	32.40	1.44	74.46	31.48	1.31	75.51	33.54	1.24
		RMSE	5.36	5.85	0.48	5.38	6.68	0.43	6.08	6.45	0.39
		BIAS	0.2134	0.0842	0.0333	0.0547	1.5757	0.1830	1.2949	2.6385	0.0287
5	SKARDU	ESTIMATE	21.05	6.05	1.15	21.05	6.36	1.55	21.27	6.59	1.42
		RMSE	1.02	0.97	0.46	1.09	1.39	0.36	1.17	1.35	0.36
		BIAS	0.0186	0.0119	0.0511	0.0230	0.3752	0.1771	0.3244	0.5954	0.0496
6	KAKUL	ESTIMATE	75.72	37.03	1.40	75.48	44.32	2.05	77.47	41.76	1.73
		RMSE	6.15	6.52	0.48	7.62	11.08	0.34	7.25	9.05	0.31

		BIAS	0.3108	0.4190	0.0387	0.4030	4.3029	0.2288	1.9796	3.8813	0.0810
7	GUPIS	ESTIAMTE	27.65	6.41	0.24	27.65	6.28	0.50	27.77	6.89	0.52
		RMSE	1.07	0.81	0.46	1.04	0.86	0.51	1.16	1.02	0.49
		BIAS	0.0387	0.0210	0.0071	0.0094	0.0280	0.0641	0.1004	0.4993	0.0085
8	DROSH	ESTIMATE	47.18	12.73	0.95	47.18	12.80	1.17	47.55	13.61	1.07
		RMSE	2.14	1.99	0.47	2.17	2.45	0.48	2.42	2.45	0.40
		BIAS	0.0732	0.1622	0.0331	0.0093	0.3377	0.1096	0.3413	1.0409	0.0213
9	BUNJI	ESTIMATE	33.46	12.72	1.21	33.46	12.85	1.35	33.98	13.84	1.31
		RMSE	2.08	2.11	0.47	2.30	2.75	0.40	2.36	2.67	0.39
		BIAS	5.2135	0.1198	0.0172	0.0572	0.6258	0.1786	0.6424	1.2380	0.0416
10	BALAKOT	ESTIMATE	74.46	32.40	1.44	74.46	31.48	1.31	75.51	33.54	1.24

		RMSE	5.40	6.004	0.49	5.30	6.46	0.43	5.98	6.24	0.39
		BIAS	0.0756	0.2411	0.0330	0.0782	1.1029	0.1427	1.2976	2.9564	0.04515
11	GHARIDUPATTA	ESTIMATE	75.72	37.03	1.40	75.48	44.32	2.05	77.47	41.76	1.73
		RMSE	6.10	6.46	0.48	7.58	11.21	0.33	7.29	9.01	0.31
		BIAS	0.3651	0.4054	0.0456	0.1050	4.5821	0.2577	1.8617	3.7728	0.0855

The major findings of the study of zone A are:

1. For the site of Astor, with a fixed sample size, high skewness value, and low kurtosis value, the estimates of location parameters show that the MLE method shows the smaller value for RMSE, and the LM method gives the lowest value for Bias. For the scale parameter, LM gives the smallest value for Bias and RMSE. For the shape parameter, MPS gives the lowest value for RMSE, and LM gives the lowest value for Bias.

2. For the site of Skardu, with a fixed sample size, moderate skewness, and low kurtosis in the estimates of the location parameter, the LM method shows the lowest value for both

Bias and RMSE. For the scale parameter, LM shows the lowest value for both RMSE and Bias. While for the shape parameter, MPS shows the lowest value for both RMSE and Bias.

3. For the Gilgit site, which has a fixed sample size with moderate skewness and low kurtosis value, the estimates of the location parameter MLE show the lowest value for both Bias and RMSE. For the scale parameter, LM shows the lowest value for both Bias and RMSE. While for the shape parameter, MPS shows the lowest value for both RMSE and Bias.

4. For the site of Muzaffarabad, which has a fixed sample size with high skewness and low kurtosis value, the estimates of the location parameter show that the LM method shows the smaller value for RMSE, and the MLE method gives the lowest value for Bias. For the scale parameter, LM shows the lowest value for both Bias and RMSE. While for the shape parameter, MPS shows the lowest value for both RMSE and Bias.

5. For the Gupis site, which has a fixed sample size with no skewness and low kurtosis value in the estimates of location parameters, the MLE method shows a smaller value for both RMSE and Bias. For the scale parameter, LM shows the lowest value for both Bias and RMSE. For the shape parameter, LM shows the lowest value for both RMSE and Bias.

6. For the site of Chitral, which has a fixed sample size with moderate skewness and low kurtosis value in the estimates of the location parameter, the LM method shows the lowest value for RMSE, and MLE shows the lowest value for Bias. For the scale parameter, LM

shows the lowest value for both RMSE and Bias. For the shape parameter, MPS shows the lowest value for RMSE, and LM shows the lowest value for Bias.

7. For the site of Drosh, which has a fixed sample size with moderate skewness and low kurtosis in the estimates of location parameters, the LM method shows the lowest value for RMSE, and MLE shows the lowest value for Bias. For the scale parameter, the LM method shows the lowest value of both RMSE and Bias. For the shape parameter, MPS shows the lowest value for both RMSE and Bias.

8. For the site of Bunji, which has a fixed sample size with high skewness and high kurtosis value in the estimates of location parameters, LM shows the lowest value for RMSE, and MLE shows the lowest value for Bias. For the scale parameter, LM shows the lowest value for both RMSE and Bias. For the shape parameter, MPS shows the lowest value for RMSE, and LM shows the lowest value for Bias.

9. For the site of Balakot, which has a fixed sample size with high skewness and low kurtosis value in the estimates of location parameters, MLE shows the lowest value for RMSE, and LM shows the lowest value for Bias. For the scale parameter, LM shows the lowest value for both RMSE and Bias. For the shape parameter, MPS shows the lowest value for RMSE, and LM shows the lowest value for Bias.

10. For the site of GhariDupatta, which has a fixed sample size with high skewness and low kurtosis value in the estimates of location parameters, LM shows the lowest value for

RMSE, and MLE shows the lowest value for Bias. For scale parameters, LM shows the lowest value for both RMSE and Bias. For the shape parameter, MPS shows the lowest value for RMSE, and LM shows the lowest value for Bias.

11. For the site of Kakul, which has a fixed sample size with high skewness and low kurtosis value in the estimates of location parameters, LM shows the lowest values for RMSE and Bias. For scale parameters, LM shows the lowest values for both RMSE and Bias. For the shape parameter, MPS shows the lowest value for RMSE, and LM shows the lowest value for Bias.

Table 4.6 Choice of estimation method for zone A based on distributional shape characteristic for PE3 Distribution considering Bias

Sr.No	Site Name	Skewness	Kurtosis	Distributional shape	Shape
1	**GILGIT**	0.85	0.94	Slightly heavier tail and moderate positive skewness	MPS
2	**SKARDU**	0.80	0.002	Slightly heavier tail and moderate positive skewness	MPS
3	**ASTORE**	1.33	2.88	Significantly heavier tail and Significant positive skewness	LM
4	**MUZAFFARABAD**	1.10	0.72	Significantly heavier tail and moderate positive skewness	MPS
5	**CHITRAL**	0.83	0.53	Significantly heavier tail and moderate positive skewness	LM
6	**GUPIS**	0.36	0.18	Slightly heavier tail and Slight positive skewness	LM
7	**DROSH**	0.75	0.21	Slightly heavier tail and Significant positive skewness	MPS

8	**BUNJI**	1.38	3.24	Significantly heavier tail and Significant positive skewness	LM
9	**BALAKOT**	1.09	0.71	Slightly heavier tail and Significant positive skewness	LM
10	**GHARIDUPATTA**	1.12	1.19	Slightly heavier tail and Significant positive skewness	LM
11	**KAKUL**	1.12	1.19	Slightly heavier tail and Significant positive skewness	LM

Table 4.7 Choice of estimation method for zone A based on distributional shape characteristic for PE3 Distribution considering RMSE

Sr.No	**Site Name**	**Skewness**	**Kurtosis**	**Distributional shape**	**Shape**
1	**GILGIT**	0.85	0.94	Slightly heavier tail and moderate positive skewness	MPS
2	**SKARDU**	0.80	0.002	Slightly heavier tail and moderate positive skewness	MPS
3	**ASTORE**	1.33	2.88	Significantly heavier tail and Significant positive skewness	MPS
4	**MUZAFFARABAD**	1.10	0.72	Significantly heavier tail and moderate positive skewness	MPS
5	**CHITRAL**	0.83	0.53	Significantly heavier tail and moderate positive skewness	MPS

6	**GUPIS**	0.36	0.18	Slightly heavier tail and Slight positive skewness	LM
7	**DROSH**	0.75	0.21	Slightly heavier tail and Significant positive skewness	MPS
8	**BUNJI**	1.38	3.24	Significantly heavier tail and Significant positive skewness	MPS
9	**BALAKOT**	1.09	0.71	Slightly heavier tail and Significant positive skewness	MPS
10	**GHARIDUPATTA**	1.12	1.19	Slightly heavier tail and Significant positive skewness	MPS
11	**KAKUL**	1.12	1.19	Slightly heavier tail and Significant positive skewness	MPS

The major findings of the study of zone C are:

1. The Quetta site has a fixed sample size with moderate skewness and low kurtosis value in the estimates of the location parameter. MLE shows the lowest value for RMSE, and LM shows the lowest value for Bias. For scale parameters, MPS shows the lowest value for RMSE, and MLE shows the lowest value for Bias. MPS shows the lowest value for the shape parameter for RMSE and Bias.

2. The Dalbaddin site, with a fixed sample size, high skewness, and low kurtosis value in the estimates of location parameter LM, shows the lowest values for RMSE and Bias. For scale parameters, MLE shows the lowest value for RMSE, and LM shows the lowest for Bias. For shape parameters, MLE shows the lowest value for RMSE, and LM shows the lowest for Bias.

Table 4.8 Choice of estimation method for zone C based on distributional shape characteristic for PE3 Distribution considering Bias

SR. NO	SITES	Skewness	Kurtosis	Distributional shape	Shape
1	QUETTA	0.80	-0.10	Slightly heavier tail and moderate positive skewness	MPS
2	DALBADDIN	1.22	0.35	Slightly heavier tail and significant positive skewness	LM

Table 4.9 Choice of estimation method for zone C based on distributional shape characteristic for PE3 Distribution considering RMSE

SR. NO	SITES	Skewness	Kurtosis	Distributional shape	Shape
1	QUETTA	0.80	-0.10	Slightly heavier tail and moderate positive skewness	MPS
2	DALBADDIN	1.22	0.35	Slightly heavier tail and significant positive skewness	MLE

The estimates of the PE3 distribution parameters, along with their RMSE and Bias for MLE, LM, and MPS, are generated using R language in R Studio. Additionally, SPSS and Minitab were used for statistical analysis. The results are presented in Tables 4.3 and 4.4.

Table 4.10 Values of Bias and RMSE of the parameters estimated through LM, MLE and MPS for Zone C

ESTIMATION METHODS		LM			MLE			MPS		
SITES	PARAMETER	LOCATION	SCALE	SHAPE	LOCATION	SCALE	SHAPE	LOCATION	SCALE	SHAPE
QUETTA	ESTIMATE	15.08	6.36	1.20	15.08	6.07	0.95	15.21	6.51	0.87
	RMSE	1.06	1.06	0.47	1.01	1.04	0.46	1.08	1.03	0.43
	BIAS	0.0127	0.0622	0.0262	0.0591	0.0347	0.0591	0.1537	0.5043	0.0129
DALBANDIN	ESTIMATE	12.71	8.45	2.13	12.86	8.62	2.08	13.06	8.49	1.93
	RMSE	1.37	1.92	0.51	1.41	1.89	0.23	1.53	2.03	0.30
	BIAS	0.0629	0.0172	0.0135	0.0803	0.3697	0.1090	0.4819	0.8869	0.0868

Chapter 5

5 Summary and conclusion:

This study investigated the Annual Maximum Rainfall Series (AMRS) of 14 stations in Pakistan's Zones A and C.

- Evaluate the efficiency of Pearson Type III distribution parameter estimation for extreme rainfall occurrences using Maximum Product Spacing.
- Compare its results to those of other popular estimating techniques.
- Assess the degree to which the recorded annual maximum rainfall series fits the Pearson Type III distribution.
- Explain how the accuracy of the at-site frequency analysis for extreme rainfall occurrences is affected when Maximum Product Spacing is used.

This study uses at-site frequency analysis and the Maximum Product Spacing (MPS) estimation approach to determine the best-fit distributions. Three estimating methods are assessed: Three methods are used to estimate probability: 1) Maximum Likelihood Estimation (MLE), 2) L-Moments (LM), and 3) Maximum Product Spacing (MPS) for fixed sample sizes and PE3 distribution shape features. The literature on MPS application is scarce despite the well-established nature of the MLE and LM approaches [35].

When empirical studies of real data are used, the results show clear trends: MPS works better with bigger samples and severe skewness and kurtosis, while LM performs well with smaller sample numbers and mild skewness and kurtosis. Especially, MPS shows consistency and effectiveness in determining the PE3 distribution's shape parameter. On the other hand, MLE yields fewer desirable outcomes for modest to small sample sizes. For low-bias estimates with smaller samples and intermediate data features, the study suggests LM, whereas MPS offers a good substitute, particularly for highly skewed and kurtotic data. For large sample sizes with few form features, MLE is still applicable. These results provide useful guidance for fitting the PE3 distribution to extreme values. Further refining of the model can be achieved by experimenting with different probability distributions, sample sizes, and parameter adjustments.

6 Limitations:

An additional phase that must be covered in this research is estimating annual maximum rainfall-extreme events for sites not gauged. There has been a problem with statistical hydrology. This study is restricted to the analysis of real data from ungauged sites.

This study's estimated methodology is restricted to 36 years of actual observed data from a small number of stations of Pakistan. With good prediction, it is simple to examine all Pakistani stations when true AMRS data is available.

7 References:

1. Hussain, M., Butt, A. R., Uzma, F., Ahmed, R., Irshad, S., Rehman, A., & Yousaf, B. (2020). A comprehensive review of climate change impacts, adaptation, and mitigation on environmental and natural calamities in Pakistan. *Environmental monitoring and assessment*, *192*, 1-20.
2. Akhtar, M., Ahmad, N., & Booij, M. J. (2008). The impact of climate change on the water resources of Hindukush–Karakorum–Himalaya region under different glacier coverage scenarios. *Journal of hydrology*, *355*(1-4), 148-163.
3. Ali, M. A. At-site and Regional Frequency Analysis of Annual Maximum Rainfall for Gilgit-Baltistan Region of Pakistan.
4. Blain, G. C. (2011). Standardized precipitation index based on Pearson type III distribution. *Revista Brasileira de Meteorologia*, *26*, 167-180.
5. Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. *New York Science Journal*, *3*(9), 40-49.
6. Hussain, Z., Shahzad, M. N., & Abbas, K. (2017). Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan. *KSCE Journal of Civil Engineering*, *21*, 1812-1819.
7. Khan, M. S. U. R., Hussain, Z., & Ahmad, I. (2021). Effects of L-moments, maximum likelihood and maximum product of spacing estimation methods in using Pearson type-3 distribution for modeling extreme values. *Water Resources Management*, *35*, 1415-1431.
8. Arns, A., Wahl, T., Haigh, I. D., Jensen, J., & Pattiaratchi, C. (2013). Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practice. *Coastal Engineering*, *81*, 51-66.
9. Khan, M. K. U., Iqbal, M. F., Mahmood, I., Shahzad, M. I., Zafar, Q., & Khalid, B. (2023). Evaluation of precipitation products over different climatic zones of Pakistan. *Theoretical and Applied Climatology*, *151*(3), 1301-1321.

10. Farooq, M., Shafique, M., & Khattak, M. S. (2018). Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. *Arabian Journal of Geosciences*, *11*, 1-10.
11. Adnan, S., Ullah, K., Khan, A. H., & Gao, S. (2017). Meteorological impacts on evapotranspiration in different climatic zones of Pakistan. *Journal of Arid Land*, *9*, 938-952.
12. Nawaz, U., Hussain, Z., Nihal, T., & Usman, S. (2020). Regional Rainfall Frequency Analysis on Seven Sites of Punjab, Pakistan using L-moments. *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 2(2), 25-35.*
13. Yonus, M., Jan, B., Khan, H., Nawaz, F., & Ali, M. (2023). Study the seasonal trend analysis and probability distribution functions of rainfall for the atmospheric region of Pakistan. *MethodsX*, *10*, 102058.
14. Alahmadi, F. (2017). Regional rainfall frequency analysis by L-moments approach for Madina Region, Saudi Arabia. *International journal of engineering research and development*, *13*(7), 39-48.
15. Aurangzeb, A., Abbas, K., Iqbal, A., Altaf, M., Nadeem, M. T., & Shahzad, F. (2022). Statistical Analysis of Rainfall Trends in Balochistan and Sindh. *Technical Journal of University of Engineering & Technology Taxila*, *27*(1).
16. Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic change*, *100*(1), 71-76.
17. Sultana, H., Ali, N., Iqbal, M. M., & Khan, A. M. (2009). Vulnerability and adaptability of wheat production in different climatic zones of Pakistan under climate change scenarios. *Climatic change*, *94*, 123-142.
18. Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2004). Statistics of extremes. Wiley series in probability and statistics. In *Chichester, Theory and Applications, with Contributions from Daniel De Waal and Chris Ferro*. John Wiley & Sons, Ltd..
19. Drissia, T. K., Jothiprakash, V., & Anitha, A. B. (2019). Flood frequency analysis using L moments: a comparison between at-site and regional approach. *Water Resources Management*, *33*, 1013-1037.
20. Hussain, Z. (2017). Estimation of flood quantiles at gauged and ungauged sites of the four major rivers of Punjab, Pakistan. *Natural hazards*, *86*, 107-123.

21. Mat Jan, N. A., & Shabri, A. (2017). They are estimating distribution parameters of annual maximum streamflows in Johor, Malaysia using TL-moments approach. *Theoretical and applied climatology*, *127*, 213-227.
22. Khan, M. S. R., Hussain, Z., & Ahmad, I. (2019). A comparison of Quadratic Regression and Artificial Neural Networks for the estimation of quantiles at ungauged sites in regional frequency analysis. *Applied Ecology & Environmental Research*, *17*(3).
23. Lei, G. J., Yin, J. X., Wang, W. C., & Wang, H. (2018). The analysis and improvement of the fuzzy weighted optimum curve-fitting method of Pearson–type III distribution. *Water resources management*, *32*, 4511-4526.
24. Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological applications*, *6*(2), 119-132.
25. Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 93-112.
26. Singh, U., Singh, S. K., & Singh, R. K. (2014). A comparative study of traditional estimation methods and maximum product spacings method in generalized inverted exponential distribution. *Journal of Statistics Applications & Probability*, *3*(2), 153.
27. Wong, T. S. T., & Li, W. K. (2006). A note on the estimation of extreme value distributions using maximum product of spacings. *Lecture Notes-Monograph Series*, 272-283.
28. Bali, T. G. (2003). The generalized extreme value distribution. *Economics letters*, *79*(3), 423-427.
29. Le Cam, L. (1990). Maximum likelihood: an introduction. *International Statistical Review/Revue Internationale de Statistique*, 153-171.
30. Dimri, A. P., Niyogi, D., Barros, A. P., Ridley, J., Mohanty, U. C., Yasunari, T., & Sikka, D. R. (2015). Western disturbances: a review. *Reviews of Geophysics*, *53*(2), 225-246.
31. Bhatti, A. S., Wang, G., Ullah, W., Ullah, S., Fiifi Tawia Hagan, D., Kwesi Nooni, I., ... & Ullah, I. (2020). Trend in extreme precipitation indices based on the long term in situ precipitation records over Pakistan. *Water*, *12*(3), 797.
32. Burauskaite-Harju, A., Grimvall, A., & Brömssen, C. V. (2012). A test for network-wide trends in rainfall extremes. *International journal of climatology*, *32*(1), 86-94.

33. Nanditha, J. S., Kushwaha, A. P., Singh, R., Malik, I., Solanki, H., Chuphal, D. S., ... & Mishra, V. (2023). The Pakistan flood of August 2022: causes and implications. *Earth's Future*, *11*(3), e2022EF003230.
34. Pan, J. X., Fang, K. T., Pan, J. X., & Fang, K. T. (2002). Maximum likelihood estimation. *Growth curve models and statistical diagnostics*, 77-158.
35. Abdulsalam, H. A., Bashiru, S. O., Isa, A. M., & Ojirobe, Y. A. Maximum Product of Spacing Parameter Estimation of Gompertz Rayleigh Distribution and Application to Rainfall Datasets.
36. Lane, J., Kasparis, T., Jones, L., Merceret, F., Glito, P., McFarquar, G., & Fisher, B. (1998, June). Steps toward improved radar estimates of convective rainfall using spatial averages obtained from rain gauge clusters. In *Proceedings of the First International Conference on Geospatial Information in Agriculture and Forestry. Lake Buena Vista, Florida*.
37. Momani, P. E. N. M., & Naill, P. E. (2009). Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. *American Journal of Environmental Sciences*, *5*(5), 599.