A Comparative Analysis of Estimation Methods for Pearson Type 3 Distribution Using Rainfall Extremes

By

Sadia Fatima

Reg No: 00000402048

School of Interdisciplinary Engineering and Sciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

A Comparative Analysis of Estimation Methods for Pearson Type 3 Distribution Using Rainfall Extremes

By

Sadia Fatima

Reg No: 00000402048

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

in partial fulfillment of the requirements for the degree of

Master of Science in
Computational Science and Engineering

Supervisor: Dr. Zamir Hussain

School of Interdisciplinary Engineering and Sciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(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 **Sadia Fatima** Registration No. **00000402048** 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: 24-OCT-2024

Signature of HoD with stamp: DR. MUHAMMAD TARIQ SAEED ASSOCIATE PROFESSOR School of Interdisciplinary Engineering & Sciences (SINES) NUST, SECTOR H-12, ISLAMABAD

Date: 28/10/2024

Countersign by

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

Date: 29/10/24

AUTHOR’S DECLARATION

I **Sadia Fatima** hereby state that my MS thesis titled “**A Comparative Analysis of Estimation Methods for Pearson Type 3 Distribution Using Rainfall Extremes**” is my own work and has not been submitted previously by me for taking any degree from 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: **Sadia Fatima**

Date: 30th October 2024

DEDICATION

I dedicate this work to my family, who have been the foundation of my success with their everlasting love and support. Especially my mother, whose prayers, kindness and unwavering support have been a constant inspiration. And my husband, whose help and constant support have been invaluable.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Zamir Hussain, for his invaluable guidance, support, and encouragement throughout the course of this research. His expertise and insights were instrumental in shaping the direction and quality of this work. I am also grateful to my GEC members, Dr. Ammar Mushtaq and Dr. Salma Sherbaz for their constructive feedback. A special thank you to Dr. Syed Irtiza Ali Shah, Principal of SINES, for developing an environment that encourages academic growth and excellence.

Lastly, I extend my appreciation to the management and faculty of my department, SINES, for their continuous support and for providing the resources and facilities necessary to complete this research.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

1	AMRS	Annual Maximum Rainfall Series
2	MPS	Maximum Product of Spacing
3	MLE	Maximum Likelihood Estimation
4	LM	L Moments
5	RMSE	Root Mean Square Error
6	PE3	Pearson Type 3 Distribution
7	ZDist	Z Statistics

LIST OF EQUATIONS

ABSTRACT

Selecting an appropriate model and method of estimation is critical for precise parameter estimation, accurate representation, and reliable prediction in hydrological and extreme rainfall analysis. This study investigates the application of the Pearson Type III (PE3) distribution for extreme rainfall analysis on Annual Maximum Rainfall Series (AMRS) in Zone B of Pakistan which includes sites of upper Punjab, KPK and Kashmir according to PMD. Three parameter estimation methods Maximum Product of Spacings (MPS), L-moments (LM), and Maximum Likelihood Estimation (MLE) are evaluated for their efficiency in fitting the PE3 distribution to the extreme rainfall data. The superior performance of MPS is attributed to its ability to minimum value of RMSE and Bias almost in all stations of zone B for moderate sample size and where skewness and kurtosis is moderate to high and provide better estimates for the tail behavior of the distribution. The findings underscore the potential of the MPS method as a reliable estimation method for Pearson Type 3 distribution.

Keywords: *Annual Maximum Rainfall series, L-moments, Maximum Likelihood Estimation, Maximum Product of Spacing, Pearson Type III distribution, Root mean square value, Bias.*

Chapter 1: Introduction

Climate change is the most critical global environmental challenge that humans are facing these days. Events of extreme weather like storms, floods, and droughts are becoming more often and severe due to climate change, which is a result of human activity like burning fossil fuels and deforestation. As a result, there are serious hazards to ecosystems, public health, and infrastructure from the unpredictability and impact of heavy rainfall events[1]. In Pakistan, climate change is intensifying extreme events by increasing riverine flooding, coastal erosion, glacial melt, and monsoon intensity. These factors, along with the nation's increasing urbanization and lack of infrastructure, make Pakistan more vulnerable to severe weather impacts on its economy, which is heavily dependent on agriculture[2].

The accuracy and reliability of the results are greatly impacted by the model and estimation method choice, making it crucial for at-site analysis in hydrological research. Various statistical models, including the Pearson Type III (PE3) distribution, are designed to reflect the unique properties of hydrological data, like skewness and kurtosis. Making effective forecasts and evaluations of extreme hydrological occurrences requires that the distribution appropriately represents the underlying data, which can only be achieved by carefully choosing a model. Accurate modeling of such data is crucial for understanding the behavior of extreme events and their potential impacts. Choosing the wrong model can lead to significant misestimations, affecting flood risk assessments and management strategies. Parameter estimation methods such as Maximum Product of Spacings (MPS), L-moments (LM), and Maximum Likelihood Estimation (MLE) have varying degrees of efficacy based on sample sizes and data properties. Each method has particular advantages and disadvantages. MLE is well-known for its asymptotic qualities, this method produces effective estimates from large samples[3]. L-moments might be more suitable for smaller datasets and provides resilience to outliers and is especially helpful for data with long tails[4]. MPS performs better in scenarios involving moderate to high skewness and kurtosis, yielding reliable tail estimates essential for analyzing extreme values. It works well for intermediate sample sizes because it successfully reduces bias and root mean square error (RMSE)[5]. Accurate predictions of intense precipitation occurrences facilitate the development of efficient flood prevention and management strategies diminishing possible financial losses and augmenting public security. Extreme event frequency and severity might be overestimated or underestimated due to poor model or estimation method selection, which can result in insufficient planning and reaction.

To model the behavior of extreme events the parameters of Pearson Type III distribution using three estimation methods, L-moments, Maximum product of spacing and Maximum likelihood estimation, are compared, to get the best estimation method out of all the three methods for PE3 for real dataset.

1.1 Study Area and Variable:

The data for this study was obtained from the Pakistan Meteorological Department (PMD). It comprises the Annual Maximum Rainfall Series (AMRS) for Zone B of Pakistan, as delineated by the PMD based on geographical coordinates of longitude and latitude. Zone B includes 14 meteorological stations: Dera Ismail Khan (Dikhan), Lahore, Sialkot, Islamabad, Peshawar, Cherat, Faisalabad, Jhelum, Kohat, Kotli, Mianwali, Sargodha, Murree, and Parachinar. Zone B was selected for this study due to its highest zonal mean and median values for average rainfall over a 30-year period, which are 66.99 mm and 57.05 mm, respectively. Additionally, Zone B, along with Zone A, exhibits higher standard errors and wider 95% confidence intervals compared to other zones, indicating greater variability and uncertainty in rainfall measurements in these regions[6].

1.2 PE3 as a preferred distribution:

The Pearson Type III (PE3) distribution is preferred for extreme value analysis due to its flexibility, ability to handle skewness, and historical precedence in hydrological studies. PE3 is known for its flexibility in modeling a wide range of data shapes, including those with skewed and heavy-tailed distributions typical of extreme values. This flexibility allows it to accurately capture the variability and extremity of annual maximum values, which is crucial for understanding rare and extreme events in hydrology. It effectively handles the skewness often present in extreme rainfall and flood events, providing more accurate and realistic modeling[7]. It is present as one of the five candidate distributions in the renowned regional frequency analysis methodology proposed by Hosking and Wallis (1997) further highlights its importance and suitability for modeling extreme values. In addition, this study contains annual maxima's which are more often used for PE3 distribution.

1.3 MPS as an estimation method:

The Maximum Product of Spacings (MPS) method is considered reliable model for estimating parameters due to its ability to focus on the spacings between ordered data points, making it sensitive to the distribution's tail behavior. This sensitivity is particularly important in extreme value analysis, where accurate modeling of tail behavior is crucial. Compared to some other estimation methods, MPS tends to produce parameter estimates with lower bias, mean square error and variance, especially in moderate sample sizes [8]. Also, for PE3 distribution, MPS and LM yield closely identical parameter estimates [9]. When compared on simulated dataset MPS performs better than MLE [10].

1.4 Gap identification:

In previous studies, the data utilized for this type of research consisted of simulated datasets. Additionally, the comparison of these estimation methods was conducted using data from only four stations, employing the Pearson Type III (PE3) distribution as the basis for the analysis [11]. So here the dataset is real data and dataset of 14 stations of zone B of Pakistan is used.

1.5 Objectives:

Following are the objectives of the study:

1. To model the PE3 distribution using the Annual Maximum Rainfall Series (AMRS) of 14 sites of Pakistan.
2. To compare the performance of MPS, MLE and LM using bias and RMSE considering available sample size.
3. To analyze efficiency each estimation method handles the skewness and inherent variability of extreme rainfall data.

1.6 Relevance to the national needs:

1. Provides clear guidelines for choosing PE3 model with exact method for selection based on sample size, skewness, and kurtosis to ensure accuracy and reliability.
2. Accurate models of extreme rainfall are essential for creating national plans for responding to natural disasters.
3. More reliable model and method will provide accurate prediction of extreme rainfall events and in turn will lead to improved flood management.

Chapter 2: Literature Review

The literature consists of the comparison of three estimation methods LM, MLE and MPS in at site frequency analysis as well as consists of at site frequency analysis carried out in different areas.

The study assesses the effects of three parameter estimation methods LM, MLE and MPS (L-moments, maximum likelihood, and maximum product of spacing) for the Pearson Type-3 distribution in modeling extreme values. For this purpose, study estimates flood quantiles considering annual maxima of river discharges (AMRD) of four sites of Khyber Pakhtunkhwa (KPK), Pakistan. The outcomes show that the LM method has low bias for small samples and data with small to moderate skewness and kurtosis, while MPS is a reasonable alternative for data with large skewness and kurtosis and small to moderate sample sizes. MLE is useful for very large sample sizes with low values of shape characteristics[11].

The paper analyzes different assessment techniques and distinguishes the best-fit distribution for at-site flood frequency analysis in Pakistan using various goodness-of-fit tests. The study focuses on the annual maximum stream flow (AMSF) data for at-site flood frequency analysis in Pakistan. The study analyzes AMSF data from a total of 18 sites in Pakistan including Tarbela, Kalabagh, Chashma, Taunsa, Guddu, Sukkur, Kotri, Mangla, Rasul, Marala and Khanki. The study finds that the Generalized Pareto distribution is the most suitable distribution for most sites, and the L-moments estimation method is the most suitable for finding the best-fit distribution. The estimated flows based on the fitted distribution are in close agreement with observed flows[12].

The paper focuses on the selection of the most appropriate probability distribution and parameter estimation method for at-site flood frequency analysis in the Torne River in Sweden. The study compares the performance of different distributions, such as generalized extreme value, three-parameter log-normal, generalized logistic, Pearson type-III, and Gumbel, using goodness-of-fit tests and accuracy measures. The results show that the L-moments estimation method generally provides the best-fitted distributions at most sites[13].

The study conducted flood frequency analysis using L moments and identified the best fit distribution for both at-site and regional flood frequency analysis in Kerala, India. The best-fit distribution for same sites for regional frequency analysis is totally different from the at-site frequency analysis whereas regional frequency analysis gave growth curves that were useful for the estimation of flood magnitude and frequency at ungauged sites[14].

The study analyzed rainfall patterns in different climate zones of Pakistan over the past three decade utilizing information from 30 meteorological observatories. The general outcome showed a diminishing pattern in precipitation all over the country, with a significant decrease in stations located in North, North West, West, and Coastal areas. However, plain areas and South West of the country did not show a significant trend[6].

The paper analyzes the annual maximum rainfall in 10 regions of Pakistan using different probability distribution functions and goodness of fit tests. The log-logistics distribution is found to be the best-fitting probability distribution for most areas in Pakistan. The study also compares the goodness of fit of different probability distribution functions using chi-square test, Kolmogorov-Smirnov test, peak weight root means square error (PWRMSE), and root mean square error (RMSE). The value of RMSE is almost always smaller than PWRMSE because it considers all values equally, while PWRMSE gives more weight to extreme values[15].

The paper investigates the suitability of fifteen different probability distributions and three parameter estimation methods including Method of moments (MOM) Maximum likelihood estimates (MLE) L-moments for at-site flood frequency analysis in Australia, using a large annual maximum flood data set. Four goodness-of-fit tests, including the Akaike information criterion, Bayesian information criterion, Anderson-Darling test, and Kolmogorov-Smirnov test, are used to identify the best-fit probability distributions. The Pearson 3, generalized extreme value, and generalized Pareto distributions are identified as the top three best-fit distributions[16].

The paper discusses the use of extreme value analysis (EVA) to estimate the likelihood of extreme values in partial coverage inspection (PCI) data, and presents a method to determine the return level and its 95% confidence intervals. The study demonstrates that extrapolations to areas larger than 1000 times the inspected area can result in unacceptable uncertainties in the return level, highlighting the importance of considering the uncertainties associated with EVA extrapolation[17].

In this paper the author has estimated the regional rainfall quantiles of 23 sites in the monsoon region of Pakistan using L-moment based index flood regional frequency analysis. This study considered three regions in the monsoon region of Pakistan and evaluated three probability distributions, namely GEV, GNO, and GLO, to determine the best choice for each region. The study found that GNO is the best choice for robust regional quantile estimation at larger return periods of 50, 100, 500, and 1000 for all three regions and that GEV is the best choice for return periods of 1, 2, 5, 10, and 20 for all three regions. The study involves testing assumptions of independence, stationarity, and identical distribution, and using different statistical measures to obtain accurate regional estimates[18]

Chapter 3: Methodology

In this section we will discuss all the methods that were conducted to carry out this research to achieve the objectives that are mentioned. The main objective of the study is to compare three methods of parameter estimation including L-moments (LM), Maximum likelihood (MLE) and Maximum Product of Spacing (MPS) Methods using Pearson type 3 distribution. Therefore, zone B of climatic zones of Pakistan comprising of 14 sites is used for the study.

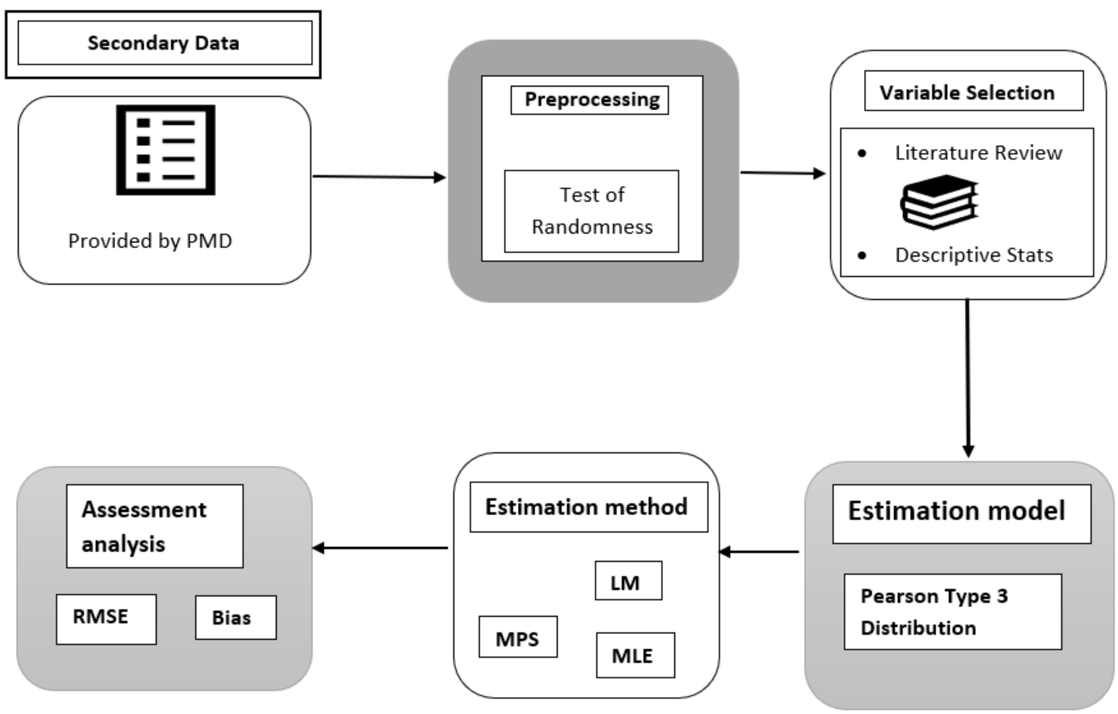

Figure 3. 1 Flow chart of methodology

3.1 Descriptive analysis:

There are generally two types of statistical analysis descriptive analysis and inferential analysis. Descriptive analysis tells about the fundamental characteristics of the data. It summarizes the involves summarizing and understanding the distribution of data. This can be done through measures of central tendency (location), measures of dispersion (shape), and measures of symmetry or skewness (scale).

Location parameters provide information about the central tendency of the data, giving an idea of where the "center" of the data lies. Common measures of location include the mean, median, and mode. Location of the data provides a sense of its typical or average value.

Shape parameters describe the dispersion or spread of the data around the center. Measures of dispersion include the range, variance, standard deviation, and interquartile range. Shape determine how spread out the data points are and whether they cluster tightly around the center or spread out more widely.

Scale parameters are related to the symmetry or skewness of the data distribution. They provide information about the relative sizes of the data values. Skewness measures whether the distribution is symmetric or skewed to one side. Kurtosis measures the "tailedness" of the distribution, indicating whether the data are concentrated around the mean or spread out more widely in the tails[19].

3.1.1 Skewness:

Skewness is the measure of asymmetry in the data. The data can be normal, moderately skewed or highly skewed. Skewness is normal which refers to ‘no skewness’ when its value is between -0.5 to 0.5. Data is moderately skewed when the values are value between -1 and -0.5 or between 0.5 and 1. And high skewness means less than -1 or greater than 1[19].

3.1.2 Kurtosis:

The value of kurtosis shows the "tailedness" of a distribution, which indicates how much of the data is in the tails of the distribution. It can be mesokurtic, leptokurtic and platykurtic. Mesokurtic means kurtosis is equal to 3, when adjusted by subtracting 3 (excess kurtosis), the value is zero, which means the distribution has neither too many outliers nor too few compared to a normal distribution. Leptokurtic means kurtosis is grater then 3 (excess kurtosis > 0). Platykurtic which is kurtosis less than 3 (excess kurtosis < 0)[19].

3.2 Pearson type 3 distribution:

Pearson type 3 distribution is the generalized gamma distribution, which is very useful for hydrologic frequency analysis and extreme value analysis. So, it is a skewed distribution and is the member of family of Pearson distributions which were introduced by Karl Pearson. If there is a random variable Y that is having Pearson type 3 distribution then its probability density function is

$$f(Y) = \frac{1}{a\Gamma(b)}\left(\frac{y-\varepsilon}{a}\right)^{b-1} e^{-\left(\frac{y-\varepsilon}{a}\right)} \qquad \textbf{(3. 1)}$$

Where ε, a and b are location, scale and shape parameters respectively. If a<0 and y≤ε then the shape parameter of the distribution is negatively skewed. If a>0 and y≥ε then the shape parameter of the distribution is positively skewed.

Given below is the key standardization of random variable Y having PE3 distribution with (a, b and ε)

$$Z = \frac{Y-\varepsilon}{a}, K = \frac{Y-\varepsilon-ab}{ab^{1/2}} \qquad \textbf{(3. 2)}$$

Where K is the frequency factor having variance is 1, mean is 0 and skewness is $2b^{1/2}$ and Z has Gamma distribution with one parameter which is equal to PE3(1, b, 0).

3.3 Three methods of estimation:

3.3.1 L-moments:

L-moments is one of the conventional moments' method utilized for the estimation of the parameters of a probability distribution. L-minutes are brilliant in tracking down the distributional properties of the severely skewed data. In depth description of L moments approach is present in [20], but here L-moments for Pearson type 3 distribution are discussed.

Here standardized parametrization is used where y>0. If y< 0 then corresponding results can be achieved through changing the signs of $\boldsymbol{\lambda}_1$, $\boldsymbol{\tau}_3$ and ε on the places where they occur in the equation.

L- moments and L-skewness according to [4]are given as follows

$$\lambda_1 = \varepsilon + ab \quad \textbf{(3. 3)}$$

$$\lambda_2 = \pi^{-1/2} a\Gamma(b+\frac{1}{2}) / \Gamma(b) \quad \textbf{(3. 4)}$$

$$\tau_3 = 6I_{1/3}(b, 2b) - 3 \quad \textbf{(3. 5)}$$

The incomplete beta function ratio is denoted here by $I_x(p,q)$.

3.3.2 Maximum likelihood method of estimation:

Maximum Likelihood Estimation (MLE) is a strong statistical method used to estimate the parameters of a probability distribution[21]. Maximum likelihood for Pearson type3 distribution is used here. So, the Maximum likelihood density function for PE3 distribution is as follows

$$lnL(a,b,\varepsilon) = -nln(a) - nln\Gamma(b) + (b-1\sum_{i=1}^{n}\ln(y_i - \varepsilon) - n(b-1)\ln(a) - \frac{1}{a}\times\sum_{i=1}^{n}(y_i - \varepsilon) \quad \textbf{(3. 6)}$$

The above equation is differentiated partially with respect to a, b and ε to get the estimates. This equation can be iteratively solved according to [22]. The value of skewness coefficient should be less than 2. For greater than 2 value of skewness will require the conditional use of Maximum likelihood parameter estimation for PE3 distribution.

3.3.3 Maximum Product of Spacing Method:

The Maximum Product of Spacings (MPS) method is a used for estimating parameters of certain probability distributions, specifically for location and scale parameters. The MPS method can be particularly useful for the Pearson Type III distribution because it simplifies the likelihood function and can give effective estimates of the parameters including location, scale and shape the product of spacing method can be advantageous in situations where maximum likelihood calculations are complicated.

The optimum log estimator of MPS is given by

$$K_{opt}(\theta) = \log \frac{1}{n+1} \qquad \textbf{(3. 7)}$$

The above equation shows that MPS has more effective results as compared to MLE as log-likelihood can reach to the infinity. The estimates of PE3 distribution parameters can be obtained through maximum product of spacing method.

$$K_{opt}(a,b,\varepsilon) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log \left[y_{i-1} \overline{\int} \left(\frac{1}{a\Gamma(b)} \left(\frac{y-\varepsilon}{a} \right)^{b-1} e^{-\left(\frac{y-\varepsilon}{a}\right)} \right) dy \right] \qquad \textbf{(3. 8)}$$

The maximization of log estimator of MPS given in the above equation is used to get estimates of PE3 distribution parameters[23].

3.4 At site frequency analysis

At site frequency analysis gives insight into the probability of future flood events. It also helps in determining the return period, which is the average time interval between floods of a certain magnitude or greater. Furthermore, at-site frequency analysis is essential for establishing flood insurance rates and developing emergency response plans.

At- site frequency analysis involves the selection of appropriate probability distribution and an appropriate method for estimation of parameters of the distribution. And Goodness of fit measures are used for the selection of distribution[13]. Typically, at site frequency analysis steps involve analyzing historical flood data, fitting statistical distributions to the data, estimating flood magnitudes and return periods, and assessing potential risks and vulnerabilities associated with flood events at a specific location.

3.5 Empirical analysis:

Empirical analysis is a crucial tool for understanding and predicting real-world phenomena. Using empirical analysis allows researchers to gather and analyze data from the real world, providing evidence-based insights into various aspects of the phenomena under study [24]. By conducting empirical analysis, researchers can uncover patterns, relationships, and trends that may not be apparent through theoretical or speculative approaches alone

Empirical analysis includes the comparison of three estimation methods LM, MLE and MPS performance for fitting the Pearson type3 distribution. For this purpose, real-life data is used of 14 sites of Pakistan which is chosen on the basis of trends and tendencies of scale and shape characteristics. The parameter estimates are calculated with two accuracy measure RMSE and bias for each site. It is further discussed in the next chapter.

Chapter 4: Results

In this chapter, the methodology which was proposed in the previous chapter is implemented here in a step by step process. A conclusion is also drawn at the end of the chapter.

4.1 Study Area:

Pakistan is located in the western region of South Asia, roughly between 23- and 37-degrees north latitude and 62- and 75-degrees east longitude. Pakistan experiences a high degree of climate variability. The north and northwest high mountain ranges experience bitterly harsh winters, but April through September is a nice summertime. The plains of the Indus Valley experience scorching summers and arid winters. The climate of the southern coast strip is moderate. Pakistan experiences the following four distinct seasons.

1. December to February is winter
2. June to September is summer
3. March to May is spring;
4. October to November is post-monsoon

While august and September are hot and humid due to the monsoon season.

Although Pakistan experiences a wide range of rainfall distributions, most of which are related to monsoon winds and western disturbances, the country does not see year-round rainfall. For example, the provinces of Baluchistan and Khyber Pakhtunkhwa (the northern mountains) have the most rainfall from December to March, whereas Punjab and Sindh have the most rainfall from 50% to 75% throughout the monsoon season[25].There are two main seasons for the precipitation that the country receives: summer, or monsoon, and winter. From July through September, Pakistan receives the monsoon rains, which comes in from the east and northeast. The north and northeastern regions of the nation receive a significant amount of rainfall during this time. The main sources of winter precipitation (December to March) are western disturbances that enter from Afghanistan and Iran. To detect the rainfall trends in the country a dataset of 30 years is utilized containing 30 stations chosen from extreme north to south and east to west covers the entire nation. These stations were selected on the basis of latitudinal position, elevation above sea level, duration of record, accuracy, and reliability of data to enable the creation of a synoptic picture of the whole nation. Additionally, the chosen stations have been split up into five distinct climatic zones A, B, C, D and E. In this study, we have selected zone B for analysis. Zone B that is mainly the region of upper Punjab and KPK including 14 meteorological stations: Dera Ismail Khan (Dikhan), Lahore, Sialkot, Islamabad, Peshawar, Cherat, Faisalabad, Jhelum, Kohat, Kotli, Mianwali, Sargodha, Murree, and Parachinar. Zone B was selected for this study because it had the highest zonal mean and median rainfall values over a 30-year period (66.99 mm and 57.05 mm, respectively). Furthermore, compared to other zones, Zone B and Zone A show larger 95% confidence intervals and higher standard errors, indicating increased variability and uncertainty in rainfall observations in these areas.[6] Also Dera Ismail Khan and Mianwali in Zone B are affected in the floods of 2022 in Pakistan.

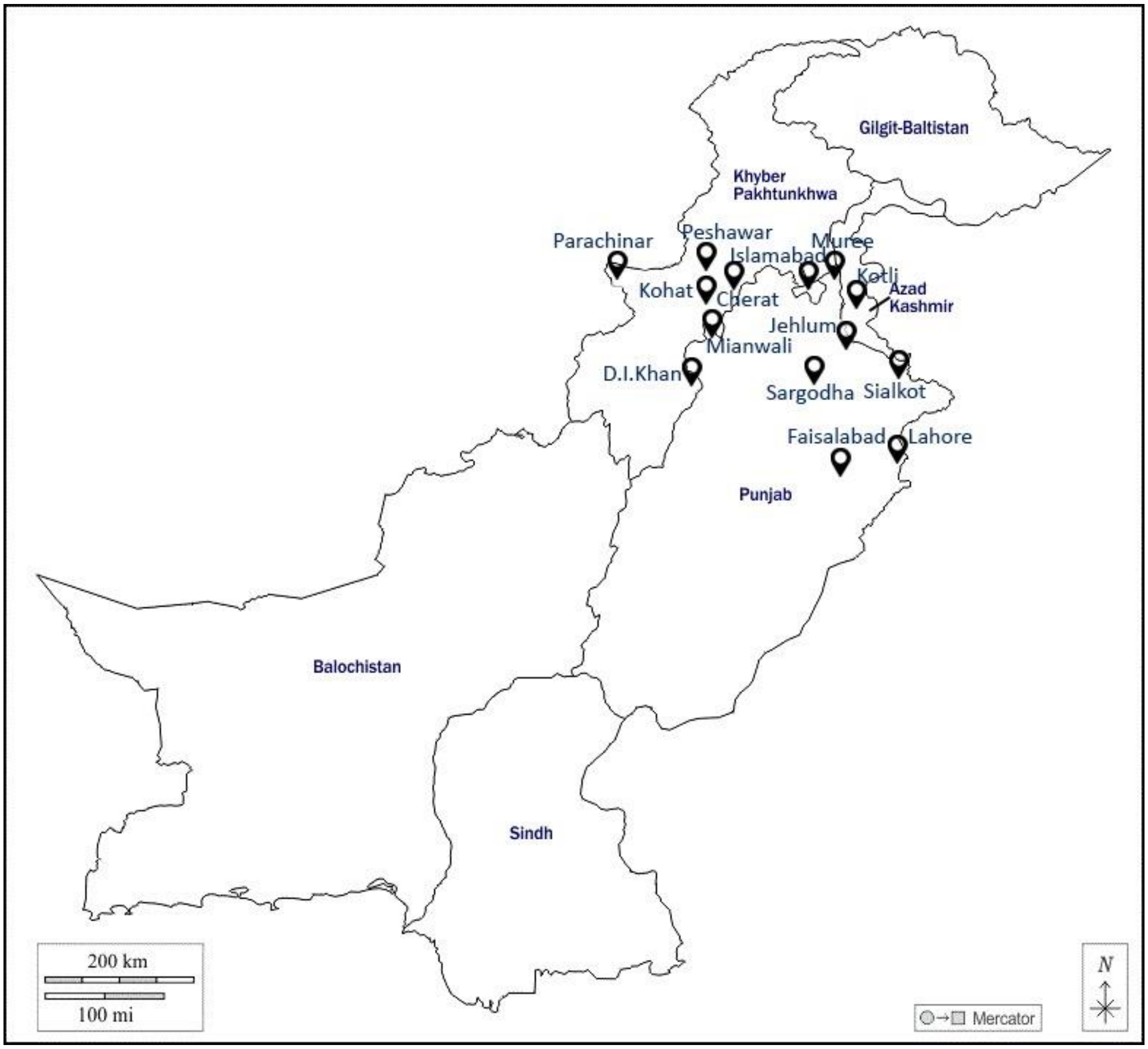

Figure 4. 1: Map of Pakistan indicating sites of zone B

4.2 Dataset and Variable:

The dataset is obtained from Pakistan Meteorological Department. It contains data of 36 years (1980-2015) and total sites included for the study are 14 distinct meteorological stations of zone B. The variable used in the study is AMRS (Annual Maximum Rainfall Series). The annual maximum rainfall series will capture the highest rainfall recorded in the year at a particular location, focusing on extreme values critical for hydrological studies and flood risk assessment. As the series is recorded on a yearly basis, most often assumes a right-skewed distribution, thereby signifying that very large rainfalls events, though quite infrequent but possible. The values are always non-negative and represent measurable amounts of rainfall. One of the main assumptions in this series is that annual maxima are all independent of one another. The methods used in parameter estimation for fitting distributions, which would be in this case Pearson Type 3, involve the maximum product of spacing, normally known as MPS, Maximum Likelihood Estimation (MLE) and L-moments (LM). Underlying the analysis is Extreme Value Theory, normally due to its availability as a robust framework in the modeling of extreme events. In general, long-term records are more homogeneous and their statistical inferences are more reliable. During the annual maxima, at the same time, the series might record some seasonal patterns or any kind of changes and variability that might indicate broader climatic changes[26].

Table 4. 1. Sites and their geographical coordinates

Sr #	Sites	Longitude	Latitude
1	**D.I. Khan**	70.9115°	31.8314°
2	**Lahore**	74.3293°	31.5820°
3	**Sialkot**	74.5361°	32.4972°
4	**Islamabad**	73.0845°	33.7380°
5	**Peshawar**	71.5601°	34.0259°
6	**Cherat**	71.8904°	33.8225°
7	**Faisalabad**	73.0791°	31.4187°

8	**Jhelum**	73.7276°	32.9405°
9	**Kotli**	73.9022°	33.5183°
10	**Kohat**	71.1667°	33.3333°
11	**Mianwali**	71.5285°	32.5776°
12	**Sargodha**	72.6742°	32.0859°
13	**Murree**	73.3903°	33.9083°
14	**Parachinar**	70.1098°	33.8836°

4.3 Empirical analysis:

For the purpose of analysis here R studio is used due to its extensive package support and comprehensive statistical analysis environment. The implementation of the approaches is facilitated by specialized packages such as lmomco, which makes it simple to undertake thorough and rigorous comparisons. It performs exceptionally well while managing and manipulating data, guaranteeing precise outcomes and effective preparation of rainfall data.

Here in this study, the RMSE and Bias of LM, MLE and MPS methods are compared using Pearson type 3 distribution to determine the robustness of these methods. RMSE (Root Mean Square Error) is a commonly used statistical tool to measure the difference between observed and predicted values. RMSE is computed through the square root of the average squared difference between the values that were predicted and those that were observed.

Lower RMSE values shows that the model fits the data better since there are less differences between the observed and predicted values. Whereas higher RMSE values shows a poorer fit reflecting higher differences between expected and actual observations. However, the difference between the expected and actual values of a predictive model's predictions is known as bias, which is a measure of the systematic error in the model. It illustrates a model's ability to frequently overestimate or underestimate the real values. The difference between the mean of the observed values and the mean of the expected results is known as bias in statistics.

The ability to provide thorough evaluations of the performance and accuracy of these statistical methods when estimating parameters for distributions like the Pearson Type 3 in extreme rainfall analysis makes RMSE (Root Mean Square Error) and bias significant when comparing L-moments (LM), Maximum Likelihood Estimation (MLE), and Method of Moments (MPS). If we talk about RMSE, RMSE is useful in evaluating the accuracy of parameter estimates produced by LM, MLE, and MPS in order to identify which approach provides the closest values in a consistent manner because by calculating the average magnitude of the errors between predicted and observed values, root mean square error (RMSE) measures a model's overall accuracy. Also, Large deviations are penalized more than small deviations by RMSE since it requires squaring the errors. Because of this feature, it's especially helpful for identifying techniques that could occasionally result in estimations that are noticeably incorrect. Now any consistent tendencies of each method that diverge from the true parameter values can be found by comparing the bias of LM, MLE, and MPS. The method having least bias is the one that implies no systematic error[27].

For calculation real data of rainfall is given as input and by using lmomco library in R studio, parameters are estimated for LM, MPS and MLE. Then by comparing the predicted and actual parameters, RMSE and bias are calculated.

Table 4. 2 Estimates of parameters of PE3 distribution along with their RMSE and Bias

S#	Estimation Methods		LM			MLE			MPS		
	Sites	**Parameter**	**Location**	**Scale**	**Shape**	**Location**	**Scale**	**Shape**	**Location**	**Scale**	**Shape**
1	Sialkot	Estimate	43.05	23.70	1.77	43.20	24.73	1.89	44.14	25.30	1.79
		RMSE	3.90	4.88	0.51	4.27	5.95	0.31	4.38	5.66	0.32
		Bias	0.07	0.22	0.03	0.30	2.26	0.21	1.19	2.41	0.08
2	D.I. Khan	Estimate	37.68	23.27	1.41	35.65	26.64	2.10	38.81	26.45	1.73
		RMSE	3.87	4.42	0.49	4.67	6.73	0.32	4.56	5.79	0.29
		Bias	0.08	0.13	0.09	0.26	2.14	0.19	1.39	2.71	0.03
3	Cherat	Estimate	76.17	31.96	0.06	76.17	30.86	0.27	76.49	33.43	0.33
		RMSE	5.13	3.99	0.46	4.81	3.86	0.57	5.63	4.65	0.49
		Bias	0.45	0.07	0.002	0.37	0.13	0.02	0.52	2.50	0.52
4	Peshawar	Estimate	70.64	26.41	0.34	70.65	25.71	0.26	70.74	27.98	0.19
		RMSE	4.30	3.34	0.45	4.12	3.29	0.56	4.75	3.67	0.49

		Bias	0.02	0.15	0.02	0.11	0.24	0.03	0.40	2.08	0.01
5	Islamabad	Estimate	60.19	37.00	1.90	60.37	40.52	2.12	61.87	39.35	1.91
		RMSE	6.34	8.17	0.49	6.67	10.15	0.34	6.79	8.83	0.31
		Bias	0.16	0.36	0.013	0.014	3.26	0.21	1.66	3.59	0.08
6	Lahore	Estimate	29.16	16.37	1.59	30.18	19.40	2.04	29.86	17.64	1.80
		RMSE	2.69	3.18	0.49	3.34	4.50	0.29	3.11	3.98	0.32
		Bias	0.10	0.05	0.04	0.32	1.47	0.17	0.78	1.71	0.10
7	Faisalabad	Estimate	35.95	33.02	2.48	37.25	36.31	2.59	37.78	33.41	2.31
		RMSE	5.48	8.78	0.55	6.10	10.92	0.47	6.09	8.59	0.28
		Bias	0.13	0.43	0.03	0.70	3.01	0.16	1.90	3.87	0.11
8	Jhelum	Estimate	48.24	26.16	1.77	48.91	31.34	2.26	49.69	30.40	2.09
		RMSE	4.32	5.48	0.50	5.49	8.39	0.35	5.35	7.63	0.30
		Bias	0.05	0.03	0.11	0.45	2.67	0.18	1.58	3.43	0.03
9	Kotli	Estimate	59.91	25.55	1.79	59.87	24.11	1.48	60.90	25.86	1.43

		RMSE	4.42	5.52	0.50	4.16	5.49	0.38	4.51	5.11	0.34
		Bias	0.02	0.19	0.02	0.02	1.22	0.17	1.34	2.47	0.07
10	Kohat	Estimate	77.98	29.33	0.54	77.98	29.41	0.80	78.78	32.18	0.79
		RMSE	4.77	3.93	0.47	5.03	4.63	0.48	5.23	5.07	0.45
		Bias	0.35	0.10	0.035	0.036	0.09	0.06	0.68	2.53	0.02
11	Mianwali	Estimate	52.83	32.63	1.60	52.85	33.17	1.59	54.40	35.34	1.53
		RMSE	5.57	6.24	0.48	5.48	7.54	0.38	6.06	7.13	0.34
		Bias	0.11	0.27	0.07	0.18	2.36	0.21	1.36	3.07	0.03
12	Sargodha	Estimate	38.82	34.10	2.72	42.94	34.88	2.15	40.47	31.97	2.0
		RMSE	5.52	9.83	0.62	6.001	8.78	0.31	5.59	7.74	0.30
		Bias	0.09	0.43	0.004	0.31	2.86	0.19	1.69	3.40	0.10
13	Murree	Estimate	75.72	37.03	1.40	75.48	44.32	2.05	77.47	41.76	1.73
		RMSE	6.26	6.73	0.47	7.57	10.56	0.33	7.17	9.16	0.32
		Bias	0.008	0.29	0.07	0.21	4.10	0.25	1.85	3.80	0.03

14	Parachinar	Estimate	37.78	11.41	1.14	37.78	11.37	1.32	38.16	12.08	1.23
		RMSE	1.89	1.86	0.48	1.97	2.33	0.43	2.06	2.23	0.39
		Bias	0.02	0.11	0.05	0.05	0.46	0.16	0.38	1.01	0.02

The estimated parameters along with RMSE and bias are given in the Table 4.2. The two metrics of accuracy, bias and RMSE, were computed using real data. For each site PE3 distribution is fitted using LM, MLE and MPS methods.

4.4 Descriptive statistics:

4.4.1 Skewness:

Skewness is a statistical metrics that quantifies how asymmetrically a data distribution is around its mean. It shows whether there is a tail on one side of the mean due to a greater concentration of data points on the other side. when the distribution's right side has a longer or fatter tail than its left. This indicates that a small number of larger values are located on the right side of the data, while the rest of the data points are clustered on the left. This means the distribution is right- skewed or the skewness is positive and vice versa for the left-skewed distribution or negative skewness.

Extreme weather events often cause skewed distributions in meteorological data, especially rainfall. Meteorologists and hydrologists can more effectively characterize the data and choose the right statistical techniques to use when they have a better understanding of the skewness. Skewness is significant when evaluating risks associated with floods. An increased likelihood of extreme rainfall events, which might result in flooding, is indicated by positive skewness in rainfall data. An accurate assessment of skewness aids in the prediction of flood frequencies and magnitudes, which in turn assists in the design of infrastructure and flood control systems. The estimation of parameters for different hydrological models is impacted by skewness. For example, the capacity of the Pearson Type 3 distribution to describe skewed data makes it a popular choice in hydrology[3].

The data can be normal, moderately skewed or highly skewed. Skewness is normal which refers to ‘no skewness’ when its value is between -0.5 to 0.5. Data is moderately skewed when the values are value between -1 and -0.5 or between 0.5 and 1. And high skewness means less than -1 or greater than 1[19].

4.4.2 Kurtosis:

Kurtosis is a statistical metric that emphasizes the extreme or "tailedness" of the data by characterizing the shape of a distribution's tails in respect to its overall shape.

In meteorology and rainfall analysis, kurtosis is essential because it gives insight into data distribution, with a particular emphasis on tails and the possibility of extreme results. Kurtosis is a useful tool for determining anomalies in meteorological data. This is crucial for rainfall analysis since prolonged rainfall have the potential to trigger flooding and other natural disasters. Meteorologists are better able to identify unusual trends and forecast extreme weather situations when they comprehend the kurtosis of rainfall data[3].

Kurtosis can be mesokurtic, leptokurtic and platykurtic. Mesokurtic is similar to normal distribution with a kurtosis value of 3. When adjusted by subtracting 3 (excess kurtosis), the value is zero. The moderate tails and modest peak of a mesokurtic distribution suggest that the likelihood of extreme values is comparable to that of a normal distribution. A mesokurtic distribution's bell curve indicates that the data are uniformly distributed around the mean. Compared to a normal distribution, a leptokurtic distribution has a larger peak and fatter tails. A kurtosis value more than 3 (excess kurtosis > 0) characterizes them. This suggests that extreme values are more common in the data. Leptokurtic distributions are particularly significant in domains like finance and meteorology where comprehending the frequency of extreme events is vital since they imply a higher possibility of outliers and extreme deviations from the mean. In comparison to normal distribution, platykurtic distributions feature thinner tails and a lower peak. Their kurtosis value is smaller than 3 (excess kurtosis < 0), which defines them. This suggests that there is less of an extreme value tendency in the data[19].

Table 4. 3 Descriptive statistics of AMRS of 14 sites of zone B. Min and Max are minimum and maximum values in the data series. Skewness and Kurtosis are the moments of measure of skewness and kurtosis

Sr #	Site Name	Min	Max	Mean	Standard Deviation	Skewness	Kurtosis
1	**Dikhan**	10.22	112.26	37.67	22.65	1.24	1.91
2	**Lahore**	11.20	65.93	29.16	15.50	1.00	0.11
3	**Sialkot**	17.17	114.42	43.05	23.23	1.58	2.84
4	**Islamabad**	22.17	157.27	60.19	35.51	1.42	1.57
5	**Peshawar**	17.22	123.42	70.65	26.03	0.19	-0.32
6	**Cherat**	25.62	141.09	76.17	31.18	0.11	-0.64
7	**Faisalabad**	9.24	140.74	35.95	30.65	1.62	2.55
8	**Jhelum**	21.13	127.40	48.24	25.41	1.47	2.19

9	**Kohat**	30.65	164.46	77.98	29.40	0.70	1.04
10	**Kotli**	28.72	129.28	59.91	24.55	1.28	1.07
11	**Mianwali**	12.72	177.27	52.83	33.12	1.80	4.72
12	**Sargodha**	10.44	133.81	38.82	31.75	1.82	2.74
13	**Murree**	32.15	184.27	75.72	35.78	1.12	1.19
14	**Parachinar**	21.89	66.51	37.78	10.99	0.81	0.14
	Average	19.33	131.29	53.15	26.84	1.16	1.65

Table 4.3 shows the average values min, max, mean, standard deviation, skewness and kurtosis in the last row of the table and average of minimum and maximum rainfall in zone B is 19.33 mm and 131.29 mm respectively. Whereas average skewness of the zone is 1.16 and average kurtosis is 1.65 which shows that skewness and kurtosis is generally high.

Table 4.3 describes the descriptive statistics of 14 sites of zone B and Table 4.2 tells about the parameter estimation. So, these two tables combine generate following discussion about the under discussion 14 sites. Here kurtosis value below 0(platykurtic) is considered low and above 0 high(leptokurtic) and skewness above 1 is considered high and below 0.5 is low.

- Skewness and kurtosis value for Sialkot site is high. For location parameter the values of RMSE and bias are significantly less for LM method. Whereas for scale parameter the value of RMSE is less for MLE and the value of bias is less for LM method. While for shape parameter MLE gives lowest value for RMSE and LM gives lowest for bias.

- The D.I. Khan site has high skewness and kurtosis value. The RMSE and bias values for the location parameter are substantially lower using the LM approach. While for shape parameter MPS gives lowest value for RMSE and bias. In contrast, the RMSE and bias values for the MLE and LM methods are lower for the scale parameter.
- The value of skewness and kurtosis is low for Cherat. For location parameter MLE method gives smaller values of RMSE and bias. Whereas for shape parameter MLE gives smaller value for RMSE and LM gives smaller value for bias. And LM gives lowest values for shape parameter for RMSE and bias.
- The Peshawar site has low values for skewness and kurtosis. The location and scale parameter for the MLE approach yields a smaller RMSE value, while the bias value for the LM method is the lowest. The LM method yields lowest value for RMSE and MPS method gives lowest for bias.
- Skewness and kurtosis value for Islamabad is high. For location parameter LM method gives smallest value of RMSE and MLE method gives smallest value for bias. For scale parameter LM method gives lowest value for RMSE and bias. For shape parameter MPS method gives smallest value for RMSE and LM gives smallest value for bias.
- Skewness and kurtosis for Lahore is high. The LM technique yields the lowest values for RMSE and bias for location and shape parameters. The shape parameter

yields the lowest values for RMSE and bias when using the MLE and LM methods, respectively.

- Skewness and kurtosis are high for Faisalabad. For location parameter LM method gives lowest value for RMSE and bias. For scale and shape parameter MPS method gives lowest value RMSE and LM method gives lowest value for bias.
- Skewness and kurtosis are high for Jhelum. For location and scale parameter LM method gives lowest value for RMSE and bias. While for shape parameter MPS gives lowest value for RMSE and bias.
- It is visible that in Kotli there is high skewness and kurtosis. For location parameter MLE gives lowest value RMSE and LM gives lowest values for bias. For scale and shape parameter MPS method gives smallest value for RMSE and LM method gives lowest value for bias.
- While for Kohat there is Moderate skewness and high kurtosis. For location and scale parameter LM method gives lowest value for RMSE and MLE method gives lowest value for bias. Whereas for shape parameter MPS gives lowest value for RMSE and bias.
- For Mianwali the skewness and kurtosis both are high. For location parameter MLE method gives lowest value of RMSE and LM method gives lowest value for bias. For scale parameter LM method gives smallest value for RMSE and bias. For shape parameter MPS method gives lowest value for RMSE and bias.
- In Sargodha the_skewness and kurtosis are high. For location parameter LM give lowest value of RMSE and bias. For scale and shape parameter MPS gives smallest value for RMSE and LM gives smallest value for bias.

- In Murree the skewness and kurtosis are high. For location and scale parameter LM gives smallest value of RMSE and bias. For shape parameter MPS gives smallest value of RMSE and bias.
- In Parachinar the skewness is moderate and kurtosis is high. For location and scale parameter LM gives smallest value of RMSE and bias. For shape parameter MPS gives smallest value for RMSE and bias.

4.5 Time series plots:

Time series plots are graphical depictions of data points gathered or recorded at subsequent times, typically at consistent intervals. The time is plotted on the x-axis and the other variable on the y-axis, to illustrate how the variable changes over time. These plots are extremely useful for a variety of analytical reasons since they can highlight trends, seasonal patterns, cycles, and anomalies in the data. Plotting time series data facilitates the identification of long-term trends in hydrological data, like patterns of rising or falling precipitation, river discharge, or groundwater levels. Understanding the natural cycles of water systems depends on being able to visualize the seasonal variations and periodic fluctuations in hydrological data that they enable. Plots of time series can be used to show the frequency, length, and intensity of extreme rainfall occurrences. These graphs are shown in Fig 4.2 to 4.15.

In time series graph of D.I Khan the peak value of 112 mm in the year 1989 as shown in the Fig 4.3. Whereas Lahore and Sialkot show peak values of 66 mm and 114 mm respectively in the year 1988 in Fig 4.2 and Fig 4.4. Islamabad and Jhelum show maximum rainfall of 157 mm and 127 mm respectively in the year 1997 as shown in the Fig 4.5 and Fig 4.9. Parachinar, Peshawar, Kohat and Mianwali show maximum rainfall of 66 mm, 123 mm, 164 mm and 177 mm respectively in the year 1995 as shown in the Fig 4.15, Fig 4.6, Fig 4.10 and Fig 4.12. In the year 1994 Cherat, Faisalabad and Sargodha experienced maximum rainfall of 141 mm, 140 mm and 133 mm respectively as shown in the Fig 4.7, Fig 4.8 and Fig 4.13. Kotli shows maximum peak value of 129 mm in the year 2006 in Fig

4.11. Whereas in 2000 Murree experienced maximum rainfall of 184 mm as shown in the Fig 4.14.

Heavy precipitation is evident from the time series graphs in 1995 and Mid-July 1995 had heavy rainfall during the monsoon season. As a result, flooding began in the Indus River and other rivers and canals.

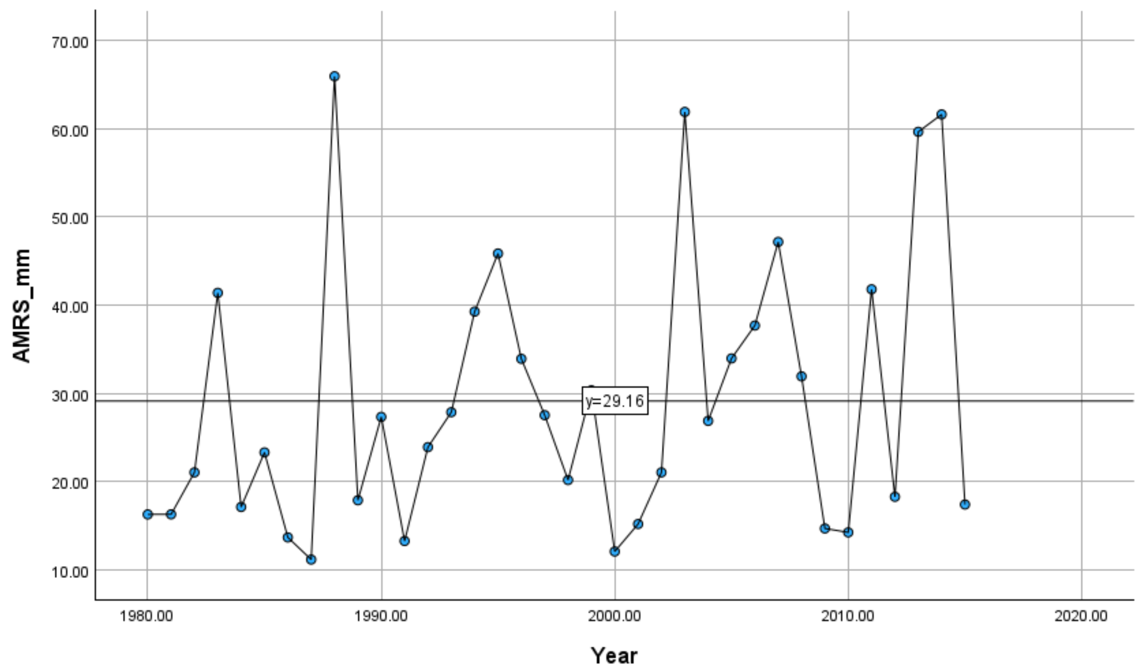

Figure 4. 2: Time Series Plot for the site Lahore

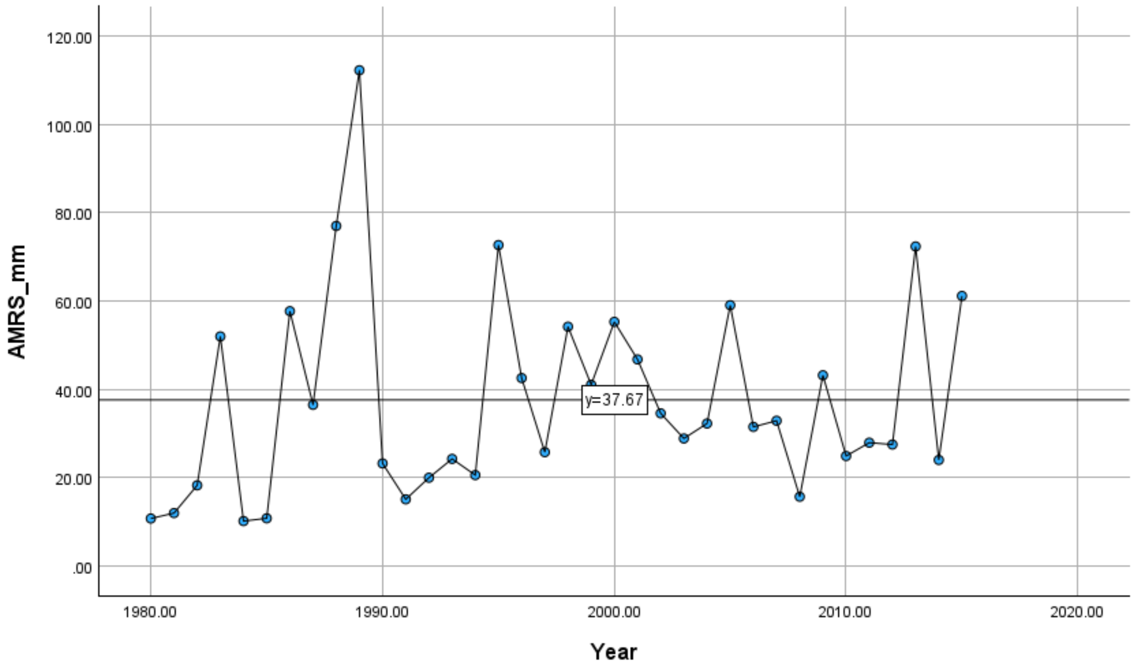

Figure 4. 3: Time Series Plot for site D.I.Khan

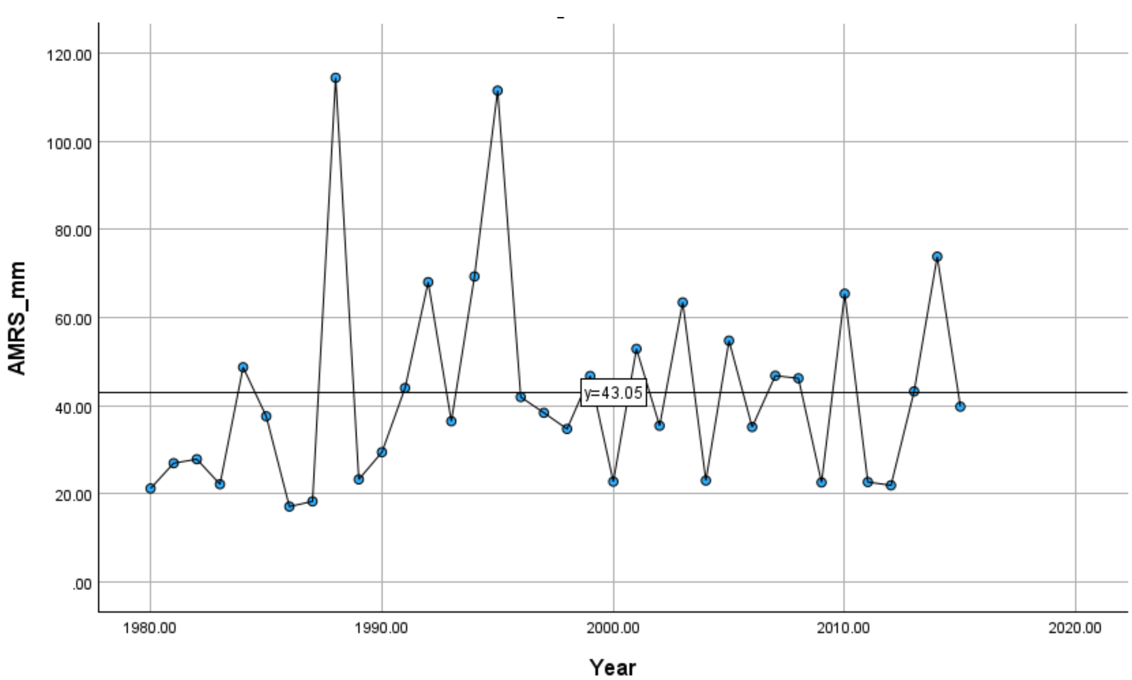

Figure 4. 4: Time Series Plot for the site Sialkot

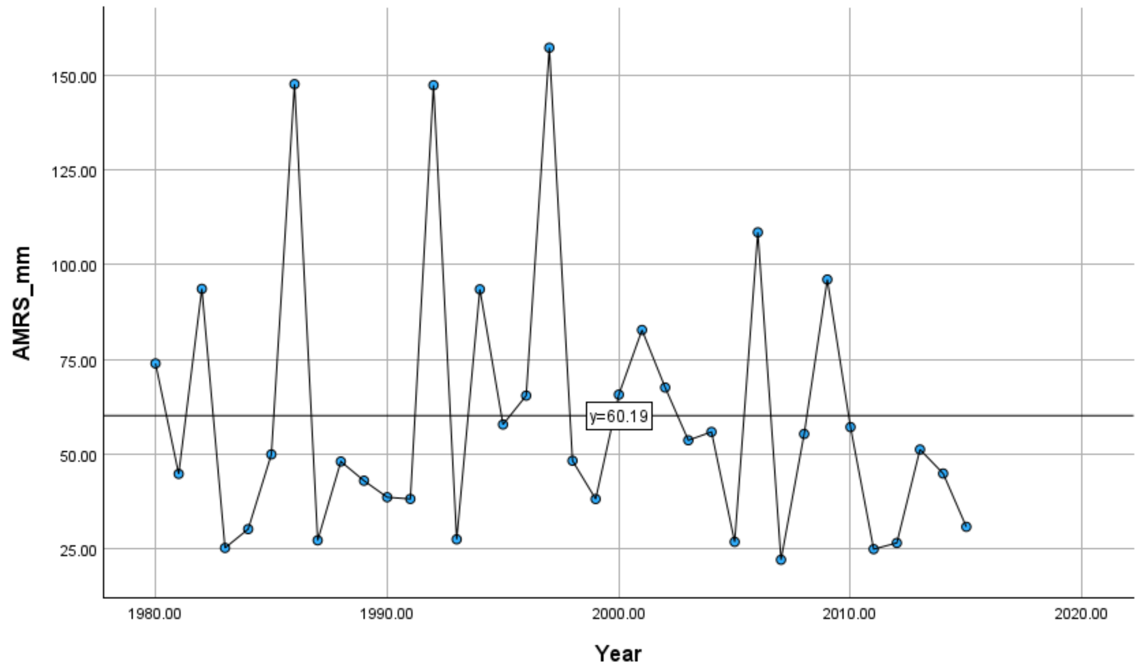

Figure 4. 5: Time Series Plot for the site Islamabad

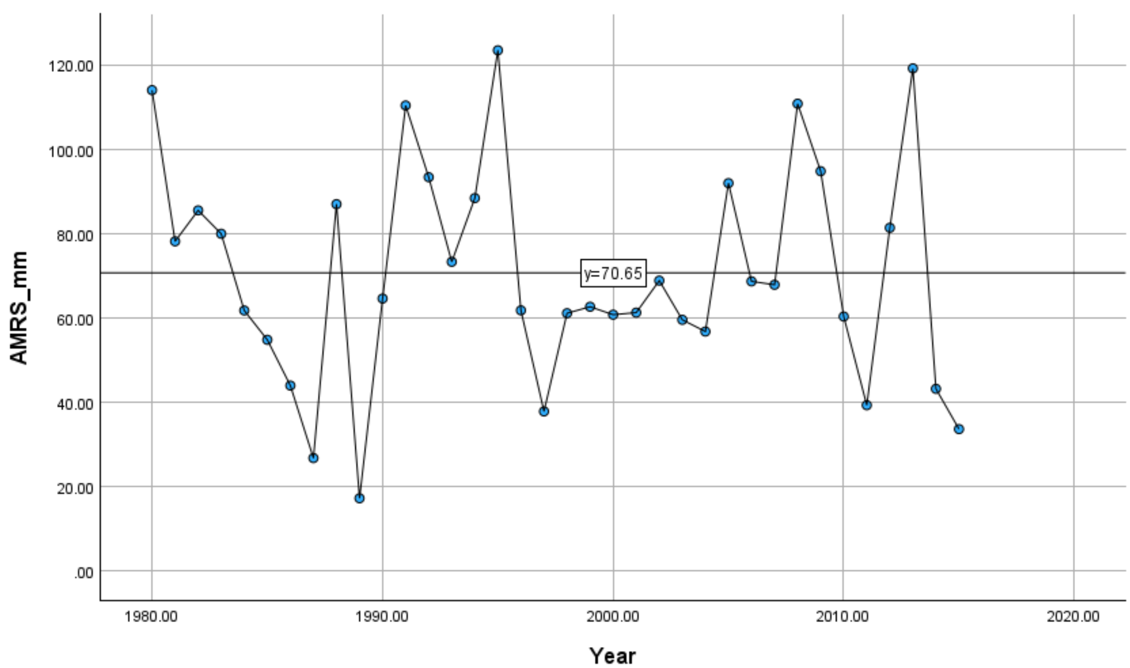

Figure 4. 6: Time Series Plot for the site Peshawar

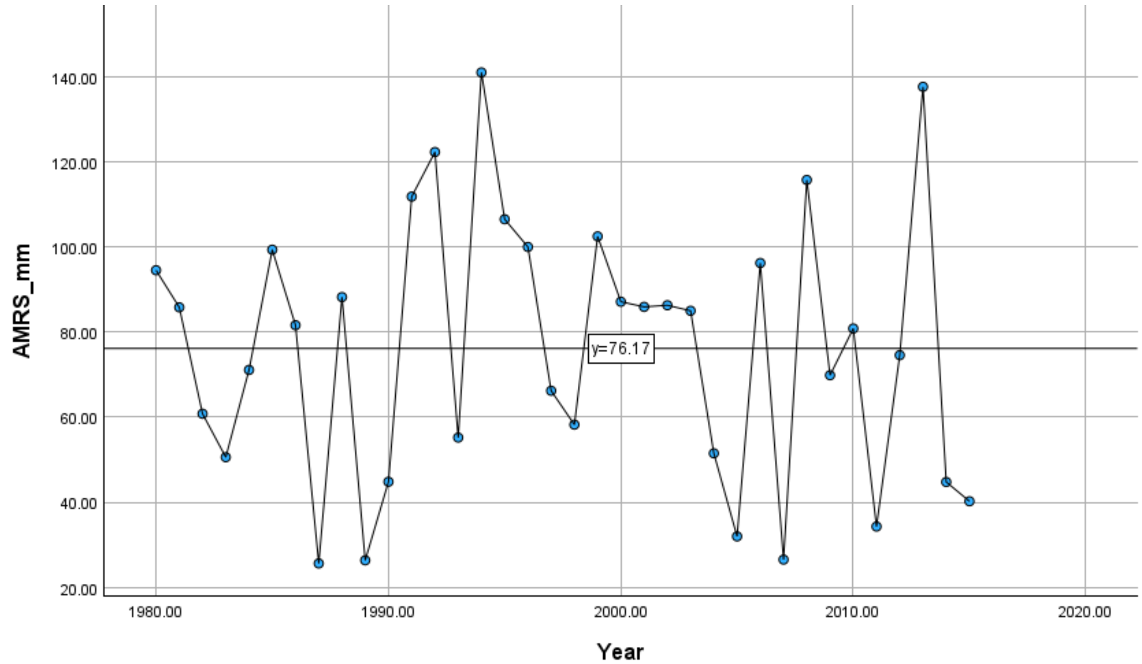

Figure 4. 7: Time Series Plot for the site Cherat

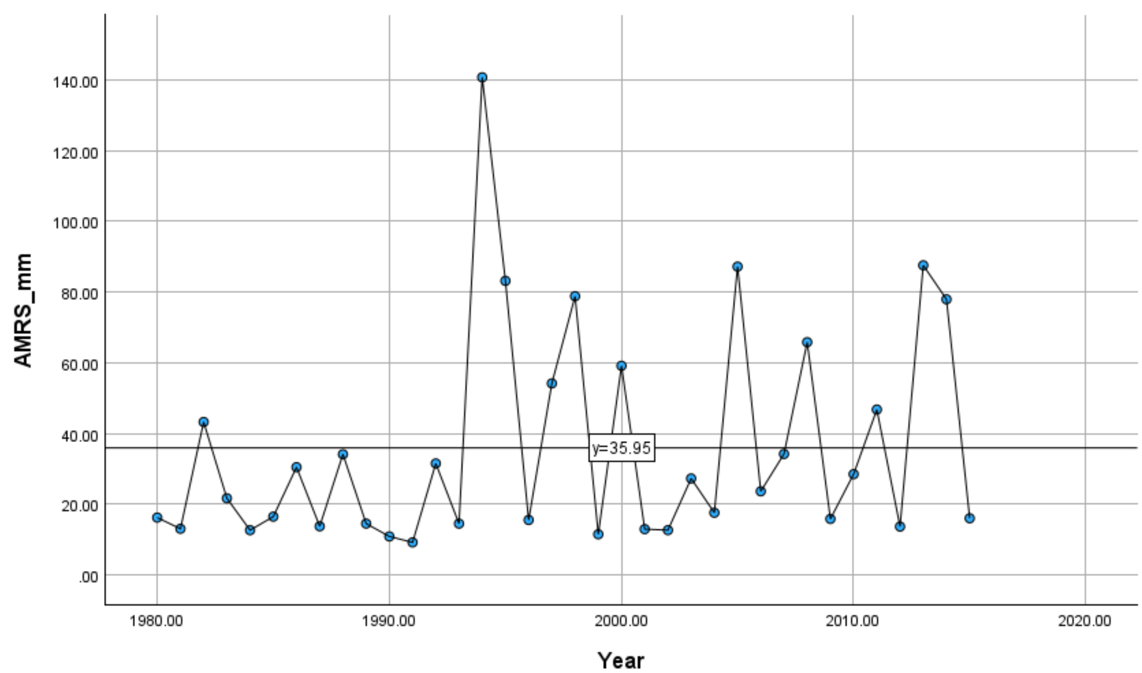

Figure 4. 8: Time Series Plot for the site Faisalabad

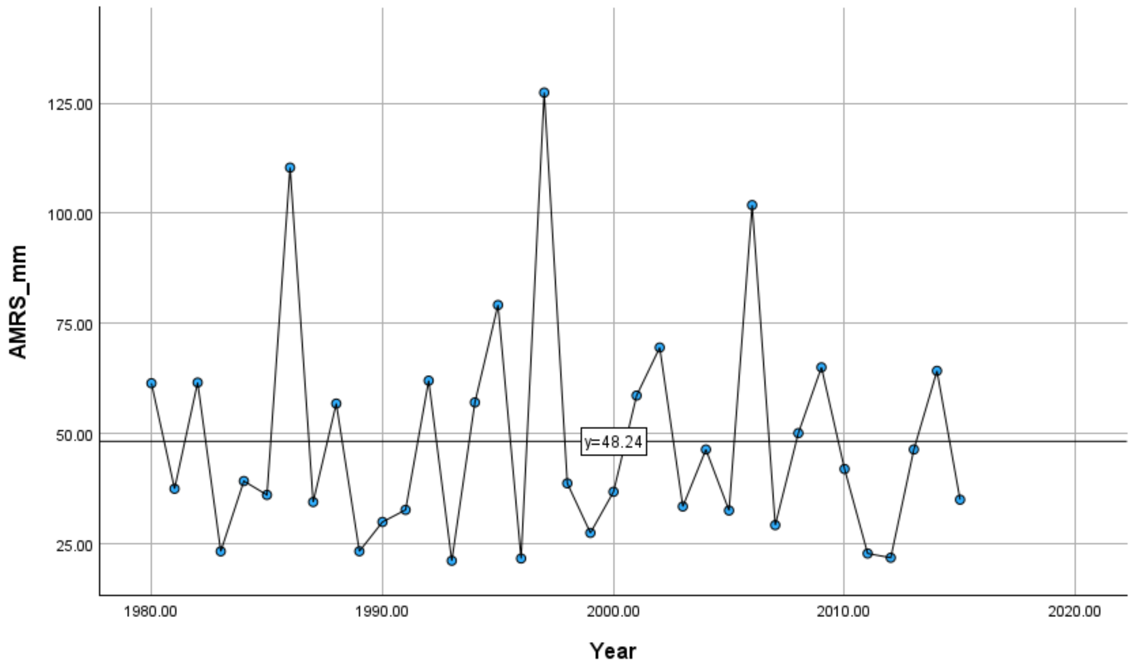

Figure 4. 9: Time Series Plot for the site Jhelum

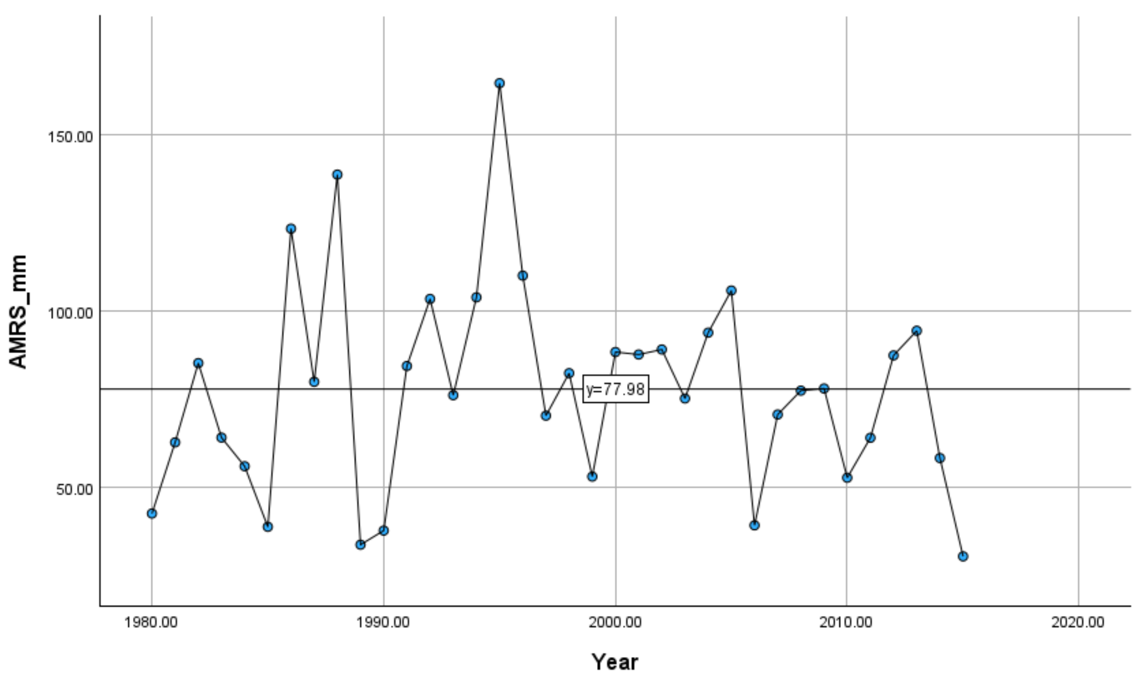

Figure 4. 10: Time Series Plot for the site Kohat

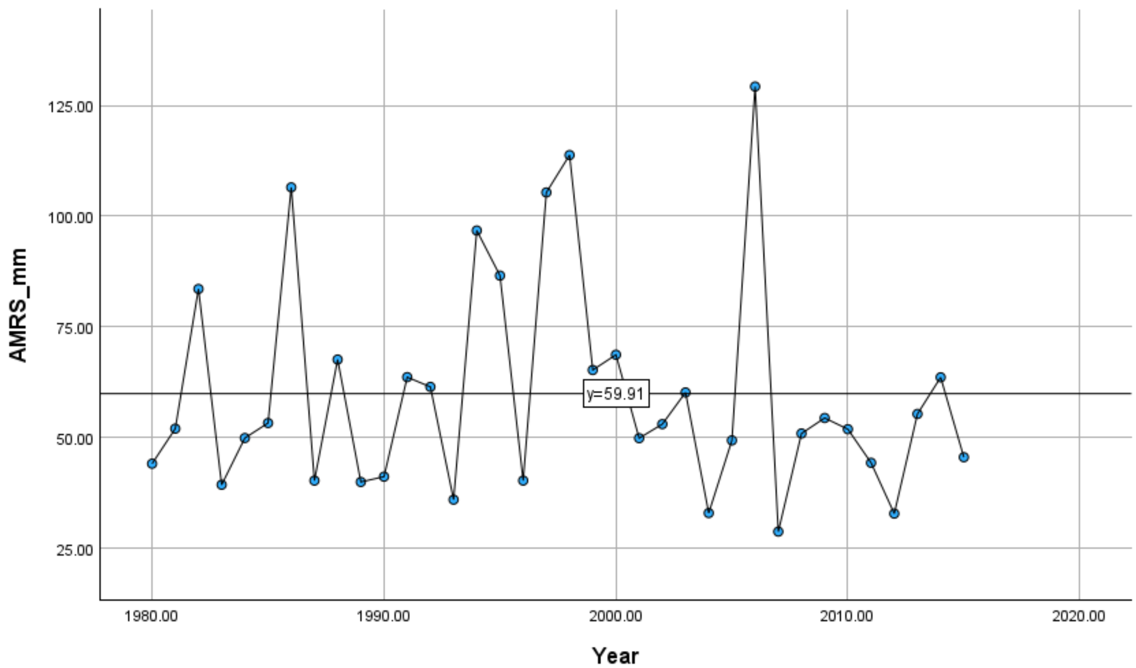

Figure 4. 11: Time Series Plot for the site Kotli

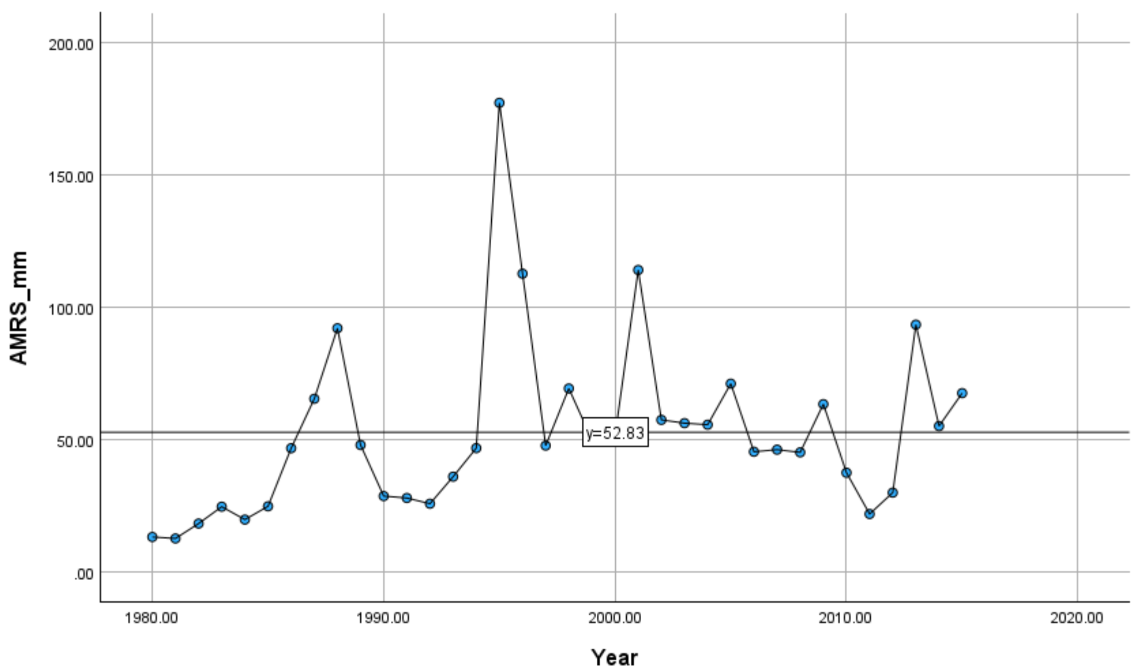

Figure 4. 12: Time Series Plot for the site Mianwali

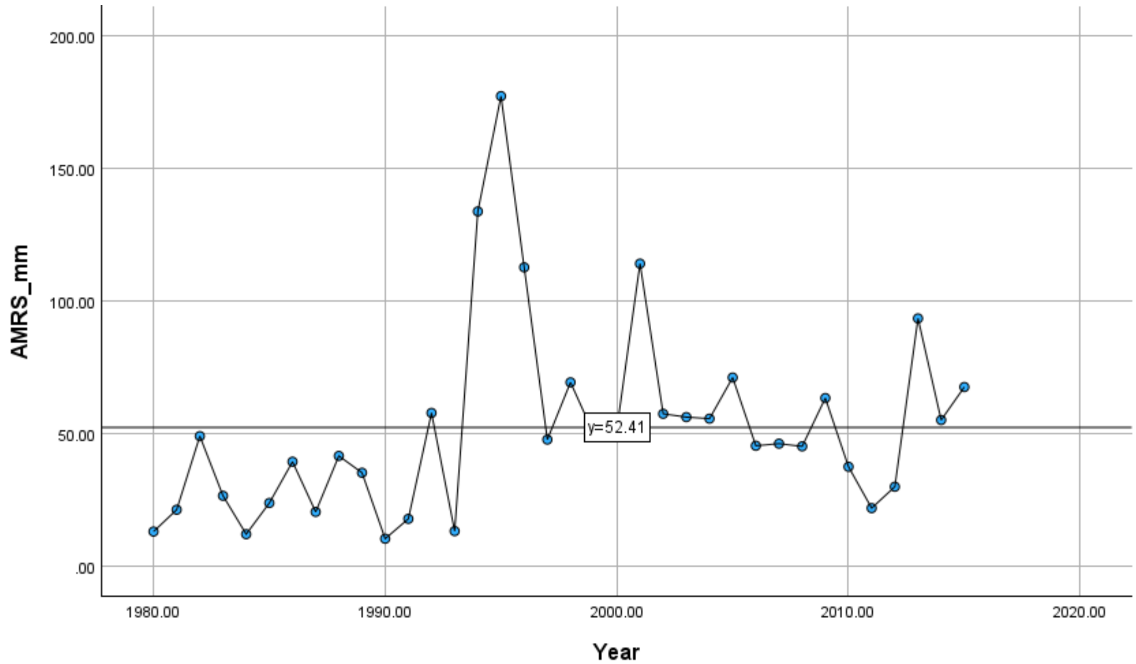

Figure 4. 13: Time Series Plot for the site Sargodha

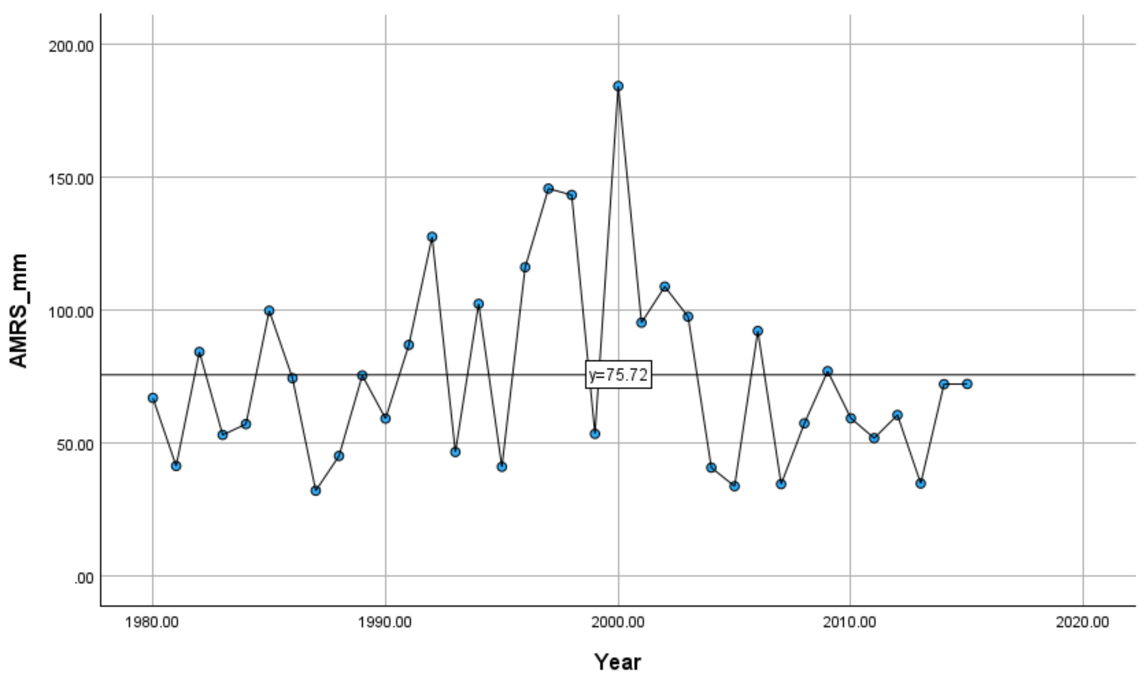

Figure 4. 14: Time Series Plot for the site Murree

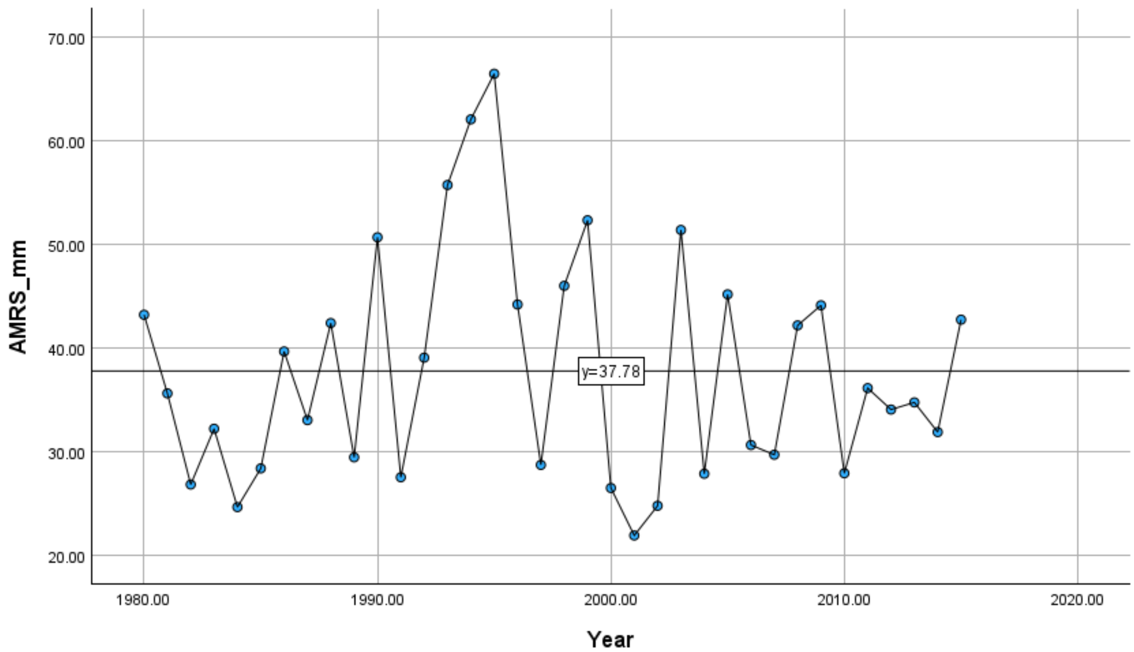

Figure 4. 15: Time Series Plot for the site Parachinar

Chapter 5: Conclusion and Future Recommendations:

This study examined the fit of the Pearson Type III (PE3) distribution to extreme rainfall data from the Annual Maximum Rainfall Series (AMRS) in Zone B of Pakistan using three parameter estimation techniques: Maximum Product of Spacings (MPS), L-moments (LM), and Maximum Likelihood Estimation (MLE). In terms of minimizing RMSE across the majority of stations in the region, the results show that the MPS method performs better than LM and MLE, especially for moderate sample sizes and data with moderate to high skewness and kurtosis. Whereas in case of Bias both LM and MPS perform equally good.

Table 5. 1 Preference table of RMSE for shape parameter

S.no	Site name	Skewness	Kurtosis	Shape
1	**Dikhan**	1.24	1.91	MPS
2	**Lahore**	1.00	0.11	MLE
3	**Sialkot**	1.58	2.84	MLE
4	**Islamabad**	1.42	1.57	MPS
5	**Peshawar**	0.19	-0.32	LM
6	**Cherat**	0.11	-0.64	LM
7	**Faisalabad**	1.62	2.55	MPS
8	**Jhelum**	1.47	2.19	MPS
9	**Kohat**	0.70	1.04	MPS

10	**Kotli**	1.28	1.07	MPS
11	**Mianwali**	1.80	4.72	MPS
12	**Sargodha**	1.82	2.74	MPS
13	**Murree**	1.12	1.19	MPS
14	**Parachinar**	0.81	0.14	MPS

Table 5. 2 Preference table of bias for shape parameter

S.no	**Site name**	**Skewness**	**Kurtosis**	**Shape**
1	**Dikhan**	1.24	1.91	MPS
2	**Lahore**	1.00	0.11	LM
3	**Sialkot**	1.58	2.84	LM
4	**Islamabad**	1.42	1.57	LM
5	**Peshawar**	0.19	-0.32	MPS
6	**Cherat**	0.11	-0.64	LM
7	**Faisalabad**	1.62	2.55	LM
8	**Jhelum**	1.47	2.19	MPS
9	**Kohat**	0.70	1.04	MPS

10	**Kotli**	1.28	1.07	LM
11	**Mianwali**	1.80	4.72	MPS
12	**Sargodha**	1.82	2.74	LM
13	**Murree**	1.12	1.19	MPS
14	**Parachinar**	0.81	0.14	MPS

Table 5.1 and 5.2 show that for the majority of stations in Zone B, the MPS approach continuously showed better results in reducing the Root Mean Square Error (RMSE) while for Bias it can be seen that LM and MPS both perform well for different sites. For moderate sample sizes and data with moderate to high skewness and kurtosis, this performance was very noteworthy. For the purpose of properly forecasting extreme rainfall events which are essential for assessing flood risk MPS's capacity to offer improved tail estimates of the PE3 distribution is essential.

The stronger ability of MPS to precisely predict the distribution's tail behavior a critical component for extreme value analysis is responsible for its higher performance. Predicting the frequency and intensity of extreme rainfall events is crucial for managing flood risks, and accurate tail estimation plays a key role in this process. The results of the study imply that the MPS method can be a trustworthy instrument for planning and assessing hydrological risk in Zone B of Pakistan, potentially leading to advancements in flood control and mitigation techniques.

However, the study admits a few limitations such as relying on the quantity and quality of historical rainfall data, focusing only on Zone B geographically, and assuming data stationarity. To overcome these constraints, future studies should investigate different distributions and estimating techniques, incorporate non-stationary models to account for the effects of climate change, and expand the investigation to different areas and data sets.

Overall, this study highlights the potential of MPS and LM as a useful strategy for extreme rainfall analysis while fitting Pearson type 3 model. Its use may lead to more precise and trustworthy flood forecasts, which in turn may facilitate the creation of successful flood control and prevention plans in the area.

5.1 Limitation:

Here are some limitations for the study;

- The only 14 sites of Pakistan are included in the study. The results might not apply to other areas with distinct rainfall patterns or climatic conditions due to this geographical focus. The applicability of results can be greatly impacted by variations in geography, land use, and local weather systems.
- For moderate sample sizes and when skewness and kurtosis are moderate to high, the performance of the estimation methods especially MPS is observed to be superior. This implies that different sample sizes and distributional features may have varied effects on the effectiveness of approach. The results might not apply to very big or very small datasets, or datasets with different statistical characteristics.
- The study makes use of annually maximum rainfall data, which might not account for more temporary extreme occurrences that take place during a year. Though they

might not be available or consistent across all stations in Zone B, high resolution data (such as monthly or daily maxima) could offer more in-depth insights into the patterns and effects of significant rainfall events.

The limitations of the study emphasize on the importance for continued research and breakthroughs in methodology to effectively tackle the complex and ever-changing nature of extreme rainfall phenomena.

REFERENCES

[1] A. G. Yilmaz, I. Hossain, and B. J. C. Perera, "Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne," *Hydrol. Earth Syst. Sci.*, vol. 18, no. 10, pp. 4065–4076, Oct. 2014, doi: 10.5194/hess-18-4065-2014.

[2] G. Rasul, M. Zahid, A. Syed, and A. Bukhari, *Climate Change in Pakistan Focused on Sindh Province*. 2012. doi: 10.13140/2.1.2170.6560.

[3] D. Wilks, *Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition (International Geophysics)*. 2005.

[4] J. R. M. Hosking and J. R. Wallis, *Regional frequency analysis: an approach based on L-moments*, 1. publ. Cambridge ; New York: Cambridge University Press, 1997.

[5] "Handbook of Hydrology." Accessed: Jul. 28, 2024. [Online]. Available: https://www.mhprofessional.com/handbook-of-hydrology-9780070397323-usa

[6] S. Rehman and M. A. Shah, "Rainfall Trends in Different Climate Zones of Pakistan," vol. 9, no. 17.

[7] S.-K. Chong and S. M. Moore, "Flood frequency analysis for small watersheds in southern illinois1," *JAWRA J. Am. Water Resour. Assoc.*, vol. 19, no. 2, pp. 277–282, Apr. 1983, doi: 10.1111/j.1752-1688.1983.tb05326.x.

[8] T. H. Soukissian and C. Tsalis, "The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction," *Nat. Hazards*, vol. 78, no. 3, pp. 1777–1809, Sep. 2015, doi: 10.1007/s11069-015-1800-0.

[9] W. Asquith, J. Kiang, and T. Cohn, *Application of At-Site Peak-Streamflow Frequency Analyses for Very Low Annual Exceedance Probabilities*. 2017. doi: 10.3133/sir20175038.

[10] E. M. Almetwally, H. M. Almongy, M. K. Rastogi, and M. Ibrahim, "Maximum Product Spacing Estimation of Weibull Distribution Under Adaptive Type-II Progressive Censoring Schemes," *Ann. Data Sci.*, vol. 7, no. 2, pp. 257–279, Jun. 2020, doi: 10.1007/s40745-020-00261-5.

[11] M. S. ul R. Khan, Z. Hussain, and I. Ahmad, "Effects of L-Moments, Maximum Likelihood and Maximum Product of Spacing Estimation Methods in Using Pearson Type-3 Distribution for Modeling Extreme Values," *Water Resour. Manag.*, vol. 35, no. 5, pp. 1415–1431, Mar. 2021, doi: 10.1007/s11269-021-02767-w.

[12] "(PDF) At-Site Flood Frequency Analysis of Annual Maximum Stream Flows in Pakistan Using Robust Estimation Methods." Accessed: Mar. 16, 2024. [Online]. Available: https://www.researchgate.net/publication/283432897_At-Site_Flood_Frequency_Analysis_of_Annual_Maximum_Stream_Flows_in_Pakistan_Using_Robust_Estimation_Methods

[13] M. Ul Hassan, O. Hayat, and Z. Noreen, "Selecting the best probability distribution for at-site flood frequency analysis; a study of Torne River," *SN Appl. Sci.*, vol. 1, no. 12, p. 1629, Nov. 2019, doi: 10.1007/s42452-019-1584-z.

[14] T. K. Drissia, V. Jothiprakash, and A. B. Anitha, "Flood Frequency Analysis Using L Moments: a Comparison between At-Site and Regional Approach," *Water Resour. Manag.*, vol. 33, no. 3, pp. 1013–1037, Feb. 2019, doi: 10.1007/s11269-018-2162-7.

[15] T. khan, Y. Wang, and M. Anwar, *Analysis of Annual Maximum Rainfall for Frequency Distribution to Determine the Best-fitted Probability Distribution for Different Sites in Pakistan.* 2021. doi: 10.21203/rs.3.rs-780243/v1.

[16] A. S. Rahman, A. Rahman, M. A. Zaman, K. Haddad, A. Ahsan, and M. Imteaz, "A study on selection of probability distributions for at-site flood frequency analysis in Australia," *Nat. Hazards*, vol. 69, no. 3, pp. 1803–1813, Dec. 2013, doi: 10.1007/s11069-013-0775-y.

[17] D. Benstock and F. Cegla, "Extreme value analysis (EVA) of inspection data and its uncertainties," *NDT E Int.*, vol. 87, pp. 68–77, Apr. 2017, doi: 10.1016/j.ndteint.2017.01.008.

[18] A. Shahzadi, A. S. Akhter, and B. Saf, "Regional Frequency Analysis of Annual Maximum Rainfall in Monsoon Region of Pakistan using L-moments," *Pak. J. Stat. Oper. Res.*, vol. 9, no. 1, pp. 111–136, Feb. 2013, doi: 10.18187/pjsor.v9i1.461.

[19] S. M. Ross, *Introduction to probability and statistics for engineers and scientists*, Fifth edition. Amsterdam ; Boston: Elsevier, AP, 2014.

[20] J. R. M. Hosking and J. R. Wallis, "Some statistics useful in regional frequency analysis," *Water Resour. Res.*, vol. 29, no. 2, pp. 271–281, 1993, doi: 10.1029/92WR01980.

[21] "Maximum Likelihood," in *The Concise Encyclopedia of Statistics*, New York, NY: Springer, 2008, pp. 334–336. doi: 10.1007/978-0-387-32833-1_248.

[22] "Eureka! It fits a Pearson type: 3 distribution - Matalas - 1973 - Water Resources Research - Wiley Online Library." Accessed: Mar. 22, 2024. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR009i002p00281

[23] S. Thongkairat, W. Yamaka, and S. Sriboonchitta, "Maximum product spacings method for the estimation of parameters of linear regression," *J. Phys. Conf. Ser.*, vol. 1053, no. 1, p. 012110, Jul. 2018, doi: 10.1088/1742-6596/1053/1/012110.

[24] L. Liu, B. F. Jones, B. Uzzi, and D. Wang, “Data, measurement and empirical methods in the science of science,” *Nat. Hum. Behav.*, vol. 7, no. 7, pp. 1046–1058, Jul. 2023, doi: 10.1038/s41562-023-01562-4.

[25] M. Latif, A. Hannachi, and F. Syed, “Analysis of rainfall trends over Indo-Pakistan summer monsoon and related dynamics based on CMIP5 climate model simulations: RAINFALL TRENDS OVER INDO-PAKISTAN SUMMER MONSOON AND DYNAMICS,” *Int. J. Climatol.*, vol. 38, Dec. 2017, doi: 10.1002/joc.5391.

[26] R. W. Katz, M. B. Parlange, and P. Naveau, “Statistics of extremes in hydrology,” *Adv. Water Resour.*, vol. 25, no. 8, pp. 1287–1304, Aug. 2002, doi: 10.1016/S0309-1708(02)00056-8.

[27] T. O. Hodson, “Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not,” *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.