UNIVERSITI TEKNOLOGI MARA

FLOOD FREQUENCY ANALYSIS USING EXTREME VALUE DISTRIBUTION IN PENINSULAR MALAYSIA BASED ON L-MOMENT METHOD

NUR AMIRA BINTI ABDUL AZIZ

BACHELOR OF SCIENCE (HONS.) STATISTICS
COLLEGE OF COMPUTING, INFORMATICS AND MATHEMATICS

FEBRUARY 2025

UNIVERSITI TEKNOLOGI MARA

FLOOD FREQUENCY ANALYSIS USING EXTREME VALUE DISTRIBUTION IN PENINSULAR MALAYSIA BASED ON L-MOMENT METHOD

NUR AMIRA BINTI ABDUL AZIZ (2023184923)

Final Year Project Submitted in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE (HONS.) STATISTICS

COLLEGE OF COMPUTING, INFORMATICS AND MATHEMATICS

FEBRUARY 2025

APPROVED BY:

(DR. BASRI BIN BADYALINA)
Supervisor
College of Computing, Informatics and Mathematics
Universiti Teknologi MARA Johor
Segamat Campus

ABSTRACT

This study explores the flood frequency analysis in Peninsular Malaysia. The L-Moment method are being use along with the extreme value distributions. It estimates the relevance of three probability distributions: Generalized Extreme Value (GEV), Generalized Pareto (GPA), and Generalized Logistic (GLO) for modelling flood magnitudes. Parameter estimation, performance evaluation, and quantile estimation are focused on this research to evaluate the flood risks. The main goal of this research is to identify the optimal probability distribution for flood frequency model. Given the ongoing monsoonal floods in Peninsular Malaysia, advanced analytical techniques are essential. The L-Moment method was applied to calculate the L-moment parameter estimate and estimate data precision from 1961 to 2023. Goodness-of-Fit tests, specifically the Anderson-Darling and Kolmogorov-Smirnov tests, along with performance metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE), and Coefficient of Determination (R^2) are being assess systematically. The GEV distribution exhibited superior performance in the majority of assessments, attaining the highest rank score, succeeded by the GLO and GPA distributions. The quantile estimates for different return periods further validated the reliability of the GEV model in forecasting flood risks. The results enhance the discipline of hydrology by offering dependable flood risk models and practical strategies for flood mitigation. Policymakers and urban planners can utilise these insights to enhance infrastructure design and disaster management strategies in flood-prone areas of Peninsular Malaysia.

ACKNOWLEDGEMENT

Alhamdulillah, all praise to Allah, the Almighty and the Most Gracious. His divine assistance has been helpful in the successful completion of this study report. Despite the numerous challenges encountered during the process, I am truly appreciative of His grace, which has permit me to get over each obstacle and keep going in enhancing my work.

My appreciation is extended to my supervisor, Dr. Basri Bin Badyalina. His dedicated attention to small detail in writing has leads to main factor of the completeness of this report. His patient and encouragement to guide me through every phase of research writing, regardless of day or night, were invaluable. His mentorship also left a huge impact on my academic journey.

Not to forget, I want to broaden the recognition to both my STA650 and MSP660 lecturers. Thankyou for giving me the opportunity to write this study. Their quick guidance and steady support throughout the course were essential in clarify my research. I am very much appreciate of their encouragement in fulfil this report writing.

Lastly, I wish to express my gratitude to my family for their tireless support in all my struggle to motivate me to continue writing this report. I am indebted of my parents for their emotional and financial support. I have always been aware of your belief in me and your desire for my well-being. To everyone who contributed to the success of this study, all your support, whether big or small, has played a crucial role in achieving this milestone. Thank you so much everyone.

TABLE OF CONTENTS

ABSTRACT **i**

ACKNOWLEDGMENT **ii**

Table of Contents **iv**

List of Tables **v**

List of Figures **vi**

Chapter 1 INTRODUCTION **1**
1.1 Background of study 1
1.2 Problem Statement 3
1.3 Research Questions 4
1.4 Research Objectives 4
1.5 Significance of Study 5
1.6 Scope and Limitation 5

Chapter 2 LITERATURE REVIEW **7**
2.1 Introduction 7
2.2 Flood Event in Peninsular Malaysia 7
2.3 Flood Frequency Analysis 8
2.4 Probability Distributions 9
2.4.1 Generalized Extreme Value 9
2.4.2 Generalized Pareto Distribution 10
2.4.3 Generalized Logistic Disribution 12
2.5 L-Moment 12
2.6 Performance Measurement 14
2.6.1 Accuracy Performance Measurement 14
2.6.2 L-Moment Ratio Diagram 15
2.6.3 Goodness-Of-Fit 15
2.7 Conclusion 16

Chapter 3 METHODOLOGY **17**
3.1 Introduction 17
3.2 Research Framework 18
3.3 Study Area 19
3.4 L-Moment 20
3.4.1 Generalized Extreme Value Distribution 22
3.4.2 Generalized Pareto Distribution 23
3.4.3 Generalized Logistic Distribution 24
3.4.4 Gringorten Plotting Position 25

3.4.5 Quantile Estimate using L-Moment 25
3.5 Performance Measurement 26
3.5.1 Accuracy Performance Measurement 27
3.5.2 L-Moment Ratio Diagram 28
3.5.3 Goodness Of Fit 29
3.5.3.1 Anderson-Darling Test 29
3.5.3.2 Kolmogorov-Smirnov Test 30
3.6 Measurement Ranking 31
3.7 Summary 31

Chapter 4 Results and Discussion 32
4.1 Introduction 32
4.2 Analysis of Flood Frequency 33
4.2.1 L-Moment 33
4.2.2 Estimation of Parameters using Extreme Value Distributions 33
4.2.3 Gringorten Plotting Position 35
4.2.4 Goodness-of-Fit 37
4.2.5 Accuracy Performance Measure 37
4.2.6 L-Moment Ratio Diagram 38
4.2.7 Measurement Ranking 40
4.2.8 Quantile Estimate 41
4.3 Summary of Results and Discussion 43

Chapter 5 Conclusion and Recommendations 44
5.1 Conclusion 44
5.2 Recommendations 45

Appendix A: Research Schedule 52

Appendix B: Turnitin Plagiarism Report 53

Appendix C: Turnitin Writing Report 54

LIST OF TABLES

3.1 Polynomial approximations of t_4 as a function of t_3 based on L-moment method 28
3.2 Summary of data analysis 31

4.1 L-Moment Components 33
4.2 Estimated Distribution Parameters Using L-Moment Methods 34
4.3 Euclidean Distance for Each Distributions 36
4.4 P-value of GOF Test for Each Distributions 37
4.5 Test Performance Measurement for Distributions 37
4.6 L-Moment Ratios of Candidate Distribution 38
4.7 Rank Score for Distributions 41
4.8 Quantile Estimates Based on Return Periods 42

5.1 Research Schedule 52

LIST OF FIGURES

3.1	Research Framework	18
3.2	Location of Flood Station in Study Area	20
3.3	L-Moment Ratio Diagram	29
4.1	Gringorten Plotting Position	35
4.2	L-Moment Ratio Diagram	39

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Malaysia was a Southeast Asian nation situated between Thailand and Singapore. It was geographically divided into two regions: Peninsular Malaysia and East Malaysia, which encompassed the states of Sabah and Sarawak. Malaysia was renowned for its tropical climate, characterised by elevated humidity, stable temperatures year-round, and plentiful rainfall. There were two primary monsoon seasons each year; Southwest Monsoon and the Northeast Monsoon. The Southwest Monsoon occurred from May to September, while the Northeast Monsoon usually occurred from November to March. The monsoons commonly resulted in increased precipitation the east coast of Peninsular Malaysia. The geographic features which include rivers that cross the country, the great variations in altitude of the land, the heavy and constant rains suffered during monsoon periods put Malaysia at high risk of floods. For example, flood events in Kelantan, Terengganu, and Johor Rivers frequently encountered bank overtopping at those periods, leading to huge flooding instances (Razali and Zolkeply, 2023; Zolkiply, 2023; Davies, 2023). More specifically, the Northeast Monsoon produced heavy and prolonged downpours which can result in sizeable floods regardless of the environment and affected both countryside and city.

These were some of the affects of flooding in Malaysia. For example affected lives, relocations of communities, destruction of infrastructure, and economic loses among others. Since these floods were annual or more frequent, the need to manage and assess the floods was important. Rentschler et al. (2022) found that the number of people affected by a once-in-a-century flood was about 1.81 billion people, or 23% of the world's population. To understand and control floods affecting areas like Peninsular Malaysia, it was important to appreciate the periodicity and intensity of floods as a vital resource, in disaster control, and in evaluation of infrastructure development. This understanding was important for risk profiling and to increase the simplification of decisions that needed to be made. Hence, an analysis for flood was needed in this study to improve flood risk planning and management.

Majority of hydrologists employed flood frequency analysis (FFA) to analyze the extreme runoff or river flow and their association with a certain return period based on probability distribution (Hamed & Rao, 2019). This study targeted the use of L-moment and other probability distributions including Generalized Pareto Distribution, Generalized Extreme Value Distribution, and Generalized Logistic Distribution method to estimate the probability of different flood magnitudes in flood frequency analysis of Peninsular Malaysia.

Previous studies have focused on different aspects of the Segamat River, Peninsular Malaysia, Pelabuhan Klang, and Kelantan River Basin, and so on (Badyalina et al., 2022; Mohd Baki et al., 2014; Ahmad et al., 2023; Hamzah et al., 2020). Mohd Baki et al. (2014) have previously investigated FFA However, the data collected showed that there was variation from one region to another, which was between 20 and 50 years.This research approach sought to estimate the peak flow reading of the river in Peninsular Malaysia for a period of 63 years using the L-Moment method. Hence, this research gave a brief analysis of the objectives, significances, and the problem of this flood analysis.

1.2 PROBLEM STATEMENT

Flood incident was a familiar and serious threat in Peninsular Malaysia. It posed threats to people, structures, and the environment. The increasing frequency and intensity of floods in the past year, underlined the requirement to develop better system of forecasting and the floods. The increase in the rate of urbanisation and deforestation had aggravated floods through altering the flow of water and the volume of water that runs over the surface. Climate change compounded this issue adding the dimension of irregular and extreme weather conditions due to evoked changes in precipitation (Abid et al., 2021).

Quantitative assessment and analysis of floods were important for effective water management, city development, and approaches to reduce the impacts of disasters. One of the ways of achieving this was by using FFA with L-moment method. FFA was a statistical approach that employs water discharge history in an attempt to estimate future instances of floods. L-Moment, a measure used in this study, provide a more accurate prediction of the probability distributions of the extreme events than the conventional moments. The strong statistical models are not sensitive to outliers and are more efficient in analysing the data with skewed distribution, which was particularly important for flood risk assessment.

The application of L-moment in Peninsular Malaysia to apply FFA enhanced the accuracy of the flood estimate, flood management and mitigation strategies. As a result , this study cover the assessment of the flood occurrence rate at different scales, the analysis of the streamflow data to determine the characteristics of floods, and the practical recommendations for flood risk mitigation in the area. The goal was to improve the analysis of flood and statistical characteristics of extreme value distributions and their parameters by applying L-moment. Using this sophisticated approach, policymakers and planners can improve the knowledge of the potential consequences of flooding as well as increase community, protecting infrastructure and ecosystems.

1.3 RESEARCH QUESTIONS

The following research questions have been formulated to guide the investigation and address key aspects of flood frequency analysis in Peninsular Malaysia:

1. How can L-moment be effectively used to estimate the parameters of extreme value distributions for flood frequency analysis?

2. What is the most accurate extreme value distribution for developing a flood frequency model that can predict the occurrence and magnitude of floods in Peninsular Malaysia's river systems?

3. How can extreme value distributions be applied to analyse historical streamflow data to accurately estimate the return periods of different flood magnitudes in Peninsular Malaysia's river systems?

1.4 RESEARCH OBJECTIVES

The aim of this study was to create a reliable flood frequency model that utilises extreme value distributions to accurately estimate the occurrence and size of floods. This involves several key objectives:

1. To utilise L-moment for the estimation of parameters of extreme value distributions used in modelling flood data.

2. To develop and evaluate a precise flood frequency model using extreme value distributions for predicting the occurrence and magnitude of floods in Peninsular Malaysia's river systems.

3. To estimate the return periods of various flood magnitudes by analyzing historical streamflow data and applying extreme value distributions.

1.5 SIGNIFICANCE OF STUDY

The results of the method had been highly advantageous for residents, as it would provided flood analysis that would alert them to the upcoming occurrence of floods. This study provides engineers and town planners with valuable data for flood analysis, which enables them to develop and enhanced infrastructure such as roads, bridges, and dams as a means of mitigating the impact of floods. FFA assists the government in identifying flood-prone areas and high-risk zones to prevent development in those areas. Additionally, it had the potential to enhance emergency preparedness by utilising flood risk assessment to optimise disaster response efficiency.

1.6 SCOPE AND LIMITATION

This study encounters several limitations that could affect the outcomes. Gathering data independently for researchers was challenging due to the time-consuming nature of the data acquisition process, which relies on offline sources from the Department of Irrigation and Drainage Malaysia (DIDM). Furthermore, the study was limited to used a particular set of online databases that were easily accessible, while other potentially valuable databases necessitate a purchase and were inaccessible. Then the data is gathered from the period 1961-2023 and it could not include all events that took place in the given period and it also does not capture the history of the countries.

The objective of this study was to develop an accurate flood frequency model that could estimate the probability and magnitude of floods in the river systems of Peninsular Malaysia. This was accomplished by utilising extreme value distributions, namely the Generalized Pareto Distribution, Generalized Extreme Value Distribution, and

Generalized Logistic Distribution. The study sought to offered an optimised evaluation of three distributions' performance in flood frequency modelling by contrasting their efficiency and accuracy for Peninsular Malaysia.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This literature discussed the existing literature that were examined during the implementation of the FFA using the L-Moment method. At the outset, Section 2.2 introduced the concept of a flood, while Section 2.3 primarily focused around the FFA. Section 2.4 was divided into three subsections that specifically addressed the three main probability distributions considered in this study. Section 2.5 provided a comprehensive explanation of the primary methodology employed in this study, namely L-Moment. Section 2.6 was subdivided into three parts, covered accuracy measurements, the L-Moment Ratio Diagram, and Goodness-Of-Fit tests. The final section offered a comprehensive summary of the findings and conclusions drawn from this chapter.

2.2 FLOOD EVENT IN PENINSULAR MALAYSIA

Malaysia was regularly impacted by the annual flooding event triggered by the seasonal monsoon, which resulted in substantial damages. The escalation of flood risk, exposure, and damage potential led to to a corresponding increase in poverty and vulnerability levels (Jan et al., 2023). These floods may be devastating in major agricultural states and substantial destruction in crops and live stocks.Yield of crops was affected by soil dehydration, and the delivery of the harvested products was also affected by flood affected roads (Romali and Yusop, 2021; Safiah Yusmah et al., 2020).

The flood hazards on an annual events had rushed the residents to prepare to evacuate before the flood happened. This action allowed them to get back to normal life faster.

However, a study done by Sufian et al. (2022) revealed that there was scarce information on the extent of flood preparedness among the residents. The researchers recommended that the authorities and government to step up on public awareness and education programmes as the region was flooded on an annual basis.

Hence, it became imperative to make an assessment on flooding to enable a proper forecast and hence prepared the residents for emerging the flood disasters. Therefore, having a clear view regarding the varied and complicated nature of the flooding problem in Malaysia was crucial for designing efficient plans and measures to reduce the risks and negative consequences of this phenomena for the affected population.

2.3 FLOOD FREQUENCY ANALYSIS

Flood Frequency Analysis (FFA) underwent considerable development in recent years. In the early 1990s, several procedures were developed to improve the FFA technique by allowing censored and individual peak data to be incorporated into the record along with multiple tests to identify outliers. In the 2000s, there were some developments in using historical data for FFA. For example, research at that time showed that historical data were helpful for the estimation of flood quantiles. Nowadays, FFA remains the core method for assessment and management of flood risks. Many research works were devoted in developing more accurate estimates with a high degree of confidence (Ali and Rahman, 2022; Dalrymple, 1960; Stedinger, 1993). Literature on FFA was becoming huge and increasing every day. So, many authors have synthesized and reviewed FFA in their papers (Hamed & Rao, 2019). The fundamental concepts of frequency analysis of streamflow data were first introduced by Todd (1957), and later the accuracy of flood estimates was discussed by Linsley (1986) and Moughamian et al. (1987).

FFA played a very important role, especially in flood-prone areas, because it provided ways to forecast future floods with the help of streamflow records (Hamzah et al., 2020). In Malaysia, FFA had also been widely applied in critical sectors, including river basin management, climate change adaptation, and flood system management. These applications helped in mitigation planning, flood pattern prediction, and drainage system improvement (Bakri, 2022; Ahmad et al., 2023; Che Ilias et al., 2021). Predicting the maximum flood magnitude of a given return period thus usually became one of the crucial requirements for the planning, design, and management of hydraulic and other infrastructure at any particular region (Krishna & Veerendra, 2015). Thus, FFA played an important role in analysing events of high runoff or river flow and their relationship with specific recurrence intervals using probability distributions (Hamed & Rao, 2019).

2.4 PROBABILITY DISTRIBUTIONS

In order to properly depict the nature of streamflow measurements at a given place, it was first necessary to identify the distribution that best described the data set. For Malaysia, extreme value distributions such as the Generalized Pareto Distribution (GPA), Generalized Extreme Value Distribution (GEV) and Generalized Logistics Distribution (GLO) were in wide use to describe probabilities of various magnitudes of flood.

2.4.1 Generalized Extreme Value

The Generalized Extreme Value (GEV) served as an essential instrument in FFA, being pivotal for modelling the distribution of extreme values, including annual maximum flood levels. This distribution was characterised by three parameters: the location parameter (ξ), which determined the central tendency; the scale parameter (α), ruling the spread or dispersion, and the shape parameter (k), influencing the tail behaviour of the distribution, hence determining the type of extreme value distribution. The GEV was an extension of what was known as the Gumbel, Fréchet, and Weibull distributions, referred as Type I, Type II, and Type III distributions, respectively; this unification was based on the shape parameter (Hossain et al., 2021). The Gumbel distribution was employed to model maxima with a light tail where the value of k was zero. On the other hand, the Fréchet distribution has an appropriate role in heavy-tailed data representing more extreme events whenever the shape parameter, was suitable for heavy-tailed data that representing more extreme events whenever the shape parameter k exceeds zero. To sum up, the Weibull distribution suits for bounded upper-tailed data where there existed some finite upper limit whenever the shape parameter k was less than zero (Legrand, 2022).

Previous work by Ahmad et al. (2023) suggested applying the GEV method in the FFA of the annual maximum stream at Bukit Kenau station in Kuantan from 1977 to 2013. The analysis showed that the GEV model was a better fit compared to the GPA. A work by Yusoff et al. (2022) in Langat River, Selangor used the GEV distribution in calculating the FFA of the annual maximum streamflow. The study conducted by Hamzah et al. (2020) also included the application of GEV distribution in Sungai Langat. Furthermore, the conducted by Bakri (2022) on Regional Flood Frequency Analysis (RFFA) in Peninsular Malaysia also applied the GEV, GPA, and GLO distributions. Moreover, the authors also determined that the GEV distribution was the most suitable distribution for the RFFA. Another study by Hamzah et al. (2020) inferred the use of GEV and GLO distributions; however, results showed that the best fit model for the particular location was GEV because of its robust evaluation.

2.4.2 Generalized Pareto Distribution

The GPA distribution was employed widely in hydrological applications to represent the distribution of peak exceedance above a specified threshold in flood data. Accurate estimation of the probability and magnitude of severe flood event was essential for assessing flood risks, managed them effectively, and designed hydraulic structures. Probability-weighted moments were one of the methods for estimating GPA distribution;

used to estimate the parameters of the GPA distribution, L-moment were generally preferred due to their robustness and reliability in hydrological applications. This distribution was characterised by three parameters: location (ξ), which sets the threshold for the distribution; the scale parameter (α), which defined the spread or dispersion of the data; and the shape parameter (k), which affected the tail behaviour and severity of flood events. The shape parameter, k was an essential one in definin the tail behaviour of the distribution. When k was positive, the distribution of GPA has a heavy upper tail; hence, the probability of an extreme event woul be greater. On the other hand, for negative k value, the variable has finite upper tail. For k value equal to zero, the distribution became the exponential distribution characterised by a light tail. In general,the GPA distribution was an excellent tool for understanding and controlling flood risks and provided critical information on the frequency and severity of extreme flood events (Campos-Aranda, 2016).

Several studies have been conducted to use the GPA, such as Hassim et al. (2022) used GPA in Kelantan River Basin and indicated that GPA has the most desirable distribution compared with other variables on a study about FFA. The researchers concluded that the GPA distribution was more appropriate than the GEV distribution based on the performance evaluation carried out in the study. A different research project conducted by Hamzah (2020) applied the GPA distribution to study the annual maximum of high tides at the Port Klang station with a particular emphasis on the study of FFA. A study done by Badyalina et al. (2022) at Labis, Johor used the GPA distribution approach in estimating the flood return period. The researchers recommended the GEV and GLO distributions be used in future studies aimed to ascertaining the most fitting distribution to the data at the location. This was based on the fact that characteristics of the peak flow reading dataset change annually.

2.4.3 Generalized Logistic Disribution

The Generalized Logistic (GLO) distribution was a powerful and flexible tool for flood frequency analysis, providing valuable insights into the probability and magnitude of extreme flood events. Its ability to model different tail behaviours made it especially useful for designing flood protection measured, and conducting comprehensive flood risk assessments. Parameter of GLO can be characterised into three parameters which were location (ξ), scale (α) and shape (k) parameters. The shape parameter, k was important in defining the tail behaviour of the distribution. Positive k indicated a heavy upper tail and a higher probability of extreme value; negative k implied a bounded upper tail which indicated a finite upper limit for data. k equal to zero indicated a symmetric tail. Due to its effectively represented the tail behaviour, hence GLO was essential for comprehending the extreme flood events (Hamed & Rao, 2019).

Previous study by Che Ilias et al. (2021) in Peninsular Malaysia, regarding RFFA implemented GLO in the study and stated that GLO distribution was the most robust probability distribution in the Region III of the study area which cover the south area of the Peninsular Malaysia. Mohd Baki et al. (2014) also supported that the GLO provided the most accurate approximation of annual maximum flow compared to GEV distribution for Region 5 in the study. Previous study by Badyalina et al. (2021), also implied the use the GLO distribution in Segamat River. Bakri (2022) implemented the use of GLO in Peninsular Malaysia, and claimed that most data of the stations were found to follow the GLO distribution instead of GEV and GPA distributions.

2.5 L-MOMENT

L-moment (LMO) were commonly employed for parameter estimation of probability distributions that characterised the occurrence of flood events. This aided in the development of flood control infrastructure and the effective management of flood hazards. Greenwood et al. (1979) once introduced Probability Weighted Moments (PWM) as an alternative to conventional moments. PWM allowed for summarising probability distributions, estimating parameters, and conducting hypothesis testing. Nevertheless, these quantities pose challenges when it came to interpret them in relation to distributional characteristics. LMO then introduced by J. R. Hosking (1990), as a set of linear combinations of order statistics that offered a robust approach to summarise the probability distributions of hydrological data. LMO were more robust to outliers and yield more accurate parameter estimates for distributions, particularly when dealing with skewed or heavy-tailed data. LMO also nearly unbiased for all combination of sample size and population. LMO coverage were more quickly then as the sample size increased which would provide more reliable parameter estimate with relatively small datasets (Vivekanandan, 2015; J. R. Hosking, 1990).

Several studies conducted by various researchers have focused on the adoption of probability distributions for FFA. Badyalina et al. (2021) applied LMO method in FFA for Segamat River in Johor, Malaysia. The study conducted by Mohd Baki et al. (2014) also utilised the GEV and GLO distributions in regional flow frequency analysis, employed the LMO in Peninsular Malaysia. The result indicated that the GLO distribution was more appropriate for estimating design runoff. Hamzah et al. (2020) proposed the existence of a LMO in an application study performed on the FFA at Pelabuhan Klang, Malaysia. The GEV and GPA models were applied in order to estimate the ebst fit for the annual maximum data of high tide in that specific location. The study carried out by Bakri (2022) applied the LMO approach in estimating the parameters of a model adopted for the study of extreme rainfall regional frequency analysis in Peninsular Malaysia. It was evidenced in the findings that the GEV distribution was the most fitted for analysing the data obtained from monthly records at 28 rain gauge stations.

2.6 PERFORMANCE MEASUREMENT

The performance of measurement had always been a prerequisite for model evaluation. This section was further divided into three subsections with regard to the measurement of performance: accuracy performance measurement, L-Moment Ratio Diagram (LMRD), and Goodness-Of-Fit (GOF) test.

2.6.1 Accuracy Performance Measurement

The accuracy measurement were necessary in the model studying model evaluation. These measurements were the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE), and the Coefficient of Determination (R^2). A lower value of these measurements indicated a better fit between the estimation from the candidate distribution and the actual data, while a higher value of R^2 indicated a better fit of the model. .

Badyalina et al. (2021) used MAPE, MAE, RMSE, RMSPE, and R^2 as evaluation criteria to assess the FFA at Segamat, Johor. The MAE, MAPE, RMSE, and RMSPE had been used in many studies to measure the difference between the actual observed peak flow with the theoretical estimated value while the R^2 measure how well the theoretical estimated from data fit to the actual observed data. The study conducted by Hassim et al. (2022) used various performance measures such as MAPE, MAE, and RMSE, to evaluate the Kelantan River Basin. The rank score method was used in this study to find the best distribution, where the minimum values of RMSE, MAE, and MAPE being assigned a maximum score of 5. On the other hand, the distribution that had the maximum values for MAPE, RMSE, and MAE received got a score of 1. which represents the minimum possible rank. The results reveal that the GPA distributions was the best methodology in characterising the annual peak flow data, while GLO was not suitable because of its low rank score. In another work by Baidya et al. (2020), the authors used the RMSE as the criterion to determine the most adequate distribution (GEV, GPA, or GLO) for estimating flood quantiles in the Mahandi River Basin, India.

2.6.2 L-Moment Ratio Diagram

The L-Moment Ratio Diagram (LMRD) was one of the methods used to determine the most suitable regional frequency distribution by examining the ratio of L-moments. A study by Prahadchai et al. (2024) used the LMRD method to explore the RFFA. The aim of the study was to determine the most suitable frequency distribution among three distributions: GEV, GPA and GLO. The relation of L-skewness and L-kurtosis showed that the L-moment ratios were more akin to the GEV distribution. As the result, the author concluded that the GEV distribution was the best distribution for the region. The LMRD method was also employed in the research conducted by Baidya et al. (2020) to ascertain the most appropriate distribution in the study. The study concluded that the GEV distribution outperformed other candidate distributions.

2.6.3 Goodness-Of-Fit

Acharya and Joshi (2020) appraised the Anderson Darling (AD) and Kolmogorov-Smirnov (K-S) tests in conjunction with Flood Frequency Analysis (FFA) within the Modi Khola area of Nepal. The aim of the study was to determine the suitable probability distribution function for the dataset. Researcher examined three distributions which include the EV, GA and GLO distributions. It was proven that GEV distribution was the suitable fit to the data in comparison to the other distribution types. In the same vein, Kousar et al. (2020) employed K-S and AD approaches as well to examine levels of FFA. The GEV was determined to demonstrate an optimal level of accuracy for the considered sites in their study utilizing the Goodness-of-Fit (GOF) test. Furthermore, Weerabangsa et al. (2023) recommended the approach including the GOF test to be used for FFA analysis. Authors also validated that GEV was the most suitable model for the data analysed.

A GOF test was also performed in Malaysia . On the other hand, Yusoff et al. (2022) classified the GOF concept as an analysis of FFA. The Kolmogorov-Smirnov (K-S) and Anderson Darling (AD) were employed by researchers to compare the data, in the case, against the annual maxima of streamflow data. In addition, Ahmad et al. (2023) suggested use of the K-S test for the frequency analysis of the annual maximum streamflow. Two extreme value distributions were analysed in this study: the GEV and GPA distributions. The output employed a ranking system, with the GEV method was the ranked highest, followed later by the GPA method.

2.7 CONCLUSION

In a nutshell, LMO method was widespread in Malaysia and it had proven by above literature that all these three distribution (GEV, GPA and GLO), can be described as those distributions which were suitable to be applied in Malaysia. The GEV, GPA, and GLO distributions together with L-moment, were useful and appropriate techniques for fitting flood frequency analysis. This technique helped in enhancing better flood control and risk mitigation in Peninsular Malaysia.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

Utilising the LMO method, flood frequency analysis was determined and ranking was a fundamental step on this process. This research applied ranking using the score method to find the best distribution for the target site which ensured better evaluation. Each distribution was ranked based on the level of correspondence between a distribution and the reality (Badyalina et al., 2021). The LMO method was applied to in charge of ranking and parameter estimation of probability models and data samples. These methods that originate from the linear combinations of given order data allow accurate estimates of mean, variance and coefficient of kurtosis.

As the result, this specific method was incorporated within this section of the methodology because of its use in assessing flood frequency. The rest of the paragraph goes on to introduces GEV, GPA and GLO alongside with each distribution parameters which were estimated using the scores derived from the LMO method enabling for stronger quantile analysis to take place. So, in order to determine which distribution was more reliable, particularly in improving the reliability and accuracy of flood frequency estimates, three subsections of performance measures, which were accuracy measure of distributions, Goodness-Of-Fit (GOF), and L-Moment Ratio Diagram (LMRD), were being covered in this study.

3.2 RESEARCH FRAMEWORK

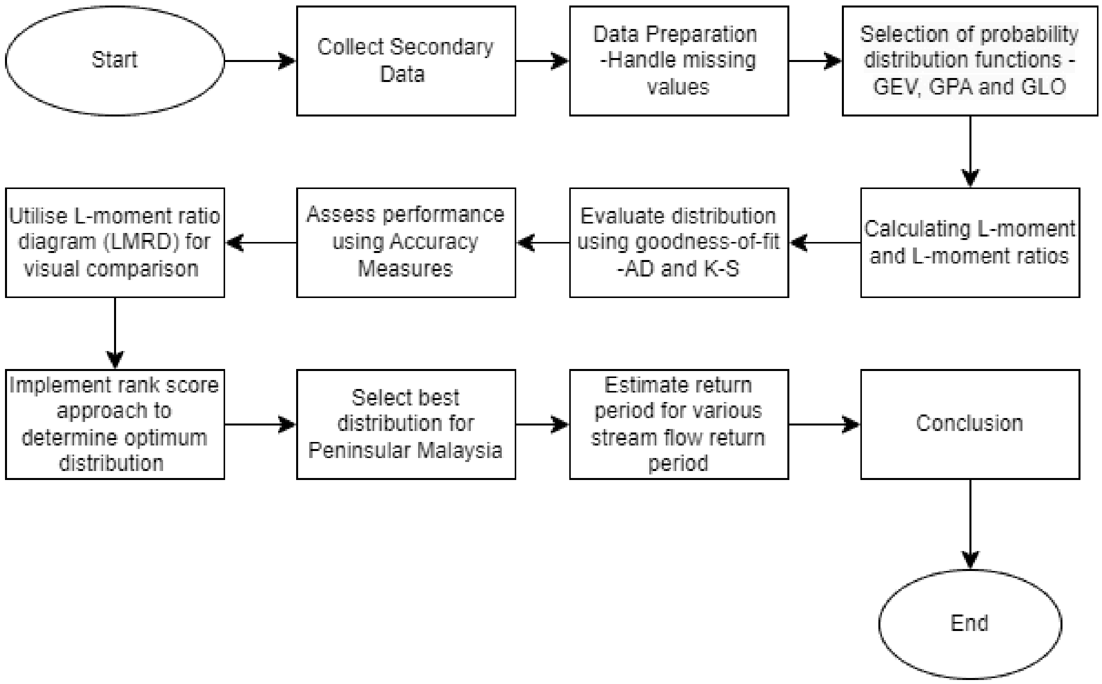


Figure 3.1: Research Framework

The framework depicted in Figure3.1 illustrated the methodology employed in the study of FFA in Peninsular Malaysia, utilising data collected from the period covering 1961 to 2023. The process commenced with the collection of data from offline sources. Subsequently, the data underwent a cleaning process during data preparation, where any missing values were addressed. Researchers then chose to include the extreme value distribution in the study, specifically the GEV, GPA, and GLO distributions. The next step in the process required the researcher to compute the LMO and LMO ratio values, and subsequently assessed the parameter estimate obtained from the distribution using the GOF test. The conducted study also included the use of Goodness-Of-Fit tests namely Anderson Darling (AD) and Kolmogorov-Smirnov(K-S).

The process then begin to evaluate diagnostics and prediction performance of the model using accuracy metrics such as Mean Absolute Error (MAE), The Mean Absolute

Percentage Error (MAPE), The Root Mean Square Error (RMSE), The Root Mean Square Percentage Error (RMSPE), and The Coefficient of Determination (R^2). Researcher might used LMO ratio to compare visually of LMRD of each distribution. The best distribution was awarded with a high score 3, while the distribution that scored less was awarded a low score 1. As the result of the total rank scores, researchers were able to identify the preferred distribution used in this study located in Peninsular Malaysia. Finally, researchers formed conclusions with respect to the findings achieved.

3.3 STUDY AREA

Figure 3.2 illustrated the geographical placement of the stations across Peninsular Malaysia. This research concentrated on a station located in Jelebu, Malaysia, with a specific emphasis targeting river stations due to their heightened susceptibility to recurrent and severe flooding triggered by monsoon seasons, which have notably affected both rural and urban populations (Bernama, 2021). The data concerning extreme streamflow was obtained from the Department of Irrigation and Drainage Malaysia, a government entity responsible for the management of water resources in Malaysia. The study area included stations with longitudinal streamflow records extending over 63 years (1961–2023), rendering it suitable for a thorough flood frequency analysis.

The existing of this historical peak flow facilitated precise parameter estimation and reinforced the suitability of the extreme value distributions in modelling flood magnitudes. Additionally, the region's varied hydrological traits, characterised by diverse topographies and extensive river systems, established a solid foundation for assessing the effectiveness of statistical models. The conclusions derived from this area were anticipated to posses considerable practical ramifications for flood risk management, infrastructure development, and disaster preparedness,thereby making it a pertinent selection.

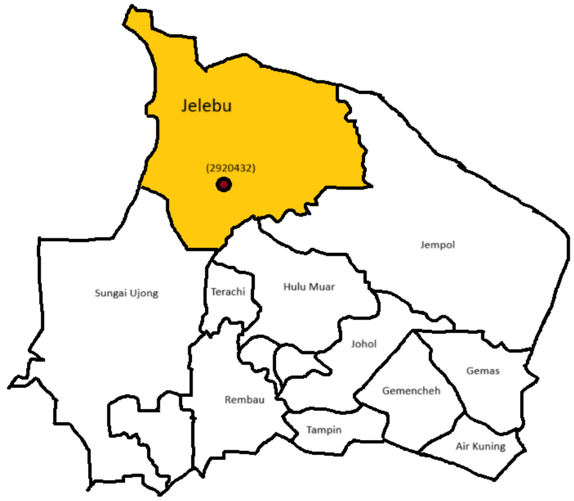


Figure 3.2: Location of Flood Station in Study Area

3.4 L-MOMENT

L-moment (LMO), which was derived from the Probability Weight Moment (PWM), had been used broadly in hydrological investigations due to its robustness and ability in the determination of the scale and shape of probability distributions (Hamed & Rao, 2019). LMO had some advantages over the traditional moments: it was less sensitive to outliers and thus more effective when dealing with small sample sizes. Their importance in hydrology was especially strong in the estimation of flood recurrence intervals, the severity of droughts, and trends in precipitation. LMO improved the accuracy of fitting probability distributions to hydrological data by providing reliable estimates of statistical characteristics. This, improvement, in turn, enabled better risk assessment and more effective management of water resource. J. R. Hosking (1990) provided the expression of PWM to estimate the r-th LMO, which was denoted by b_r, where r was 1,2,3,..., n

$$b_r = \frac{1}{n}\binom{n-1}{r}^{-1}\sum_{i=r+1}^{n}\binom{i-1}{r}x_{i:n} \tag{3.1}$$

where $x_{i:n}$ was the ordered reading of streamflow, b_r was the PWM, n was the sample size. Thus, the first four moments of an unbiased sample estimator were given as followed:

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} x_{i:n} \tag{3.2}$$

$$b_1 = \frac{1}{n}\sum_{i=2}^{n} \frac{(i-1)}{(n-1)} x_{i:n} \tag{3.3}$$

$$b_2 = \frac{1}{n}\sum_{i=3}^{n} \frac{(i-1)(i-2)}{(n-1)(n-2)} x_{i:n} \tag{3.4}$$

$$b_3 = \frac{1}{n}\sum_{i=4}^{n} \frac{(i-1)(i-2)(i-3)}{(n-1)(n-2)(n-3)} x_{i:n} \tag{3.5}$$

The first four sample estimates for LMO were the mean of distribution (l_1), measure of scale (l_2), measure of skewness (l_3) and measure of kurtosis (l_4) respectively, which referred as:

$$l_1 = b_0 \tag{3.6}$$

$$l_2 = 2b_1 - b_0 \tag{3.7}$$

$$l_3 = 6b_2 - 6b_1 + b_0 \tag{3.8}$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{3.9}$$

LMO also could be used to derive the LMO ratios, which were anologous to convectional moment ratios, (J. R. Hosking, 1990). Hence, the sample of the LMO ratio could be referred as:

$$t_2 = \frac{l_2}{l_1} \tag{3.10}$$

$$t_3 = \frac{l_3}{l_2} \tag{3.11}$$

$$t_4 = \frac{l_4}{l_2} \tag{3.12}$$

where t_2 was the measure of scale and dispersion, t_3 was the measure of skewness and t_4 was the measure of kurtosis.

3.4.1 Generalized Extreme Value Distribution

The GEV distribution—widely used for the analysis of extreme events— has been widely applied in the frequency analysis of both flood and drought phenomena, (Zeng et al., 2015). The Eq. (3.13) presented the probability density function of GEV distribution (Hamed & Rao, 2019).

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{-1/k-1}\exp\left\{-\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{-1/k}\right\} \tag{3.13}$$

where ξ was the location parameter, $\alpha > 0$ was the scale parameter, k was the shape parameter. The cumulative density function of GEV, was given by Eq. (3.14) and was denoted by $F(x)$,

$$F(x) = \exp\left\{-\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{-1/k}\right\} \tag{3.14}$$

where ξ represented the location parameter, $\alpha > 0$ denoted the scale parameter, and k signified the shape parameter. Consequently, the parameters ξ, α, and k associated with GEV distribution can be estimated according to the methodology outlined in, (J. R. M. Hosking et al., 1985).

$$\hat{k} = 7.8590 + 2.9554C^2 \tag{3.15}$$

$$\hat{\alpha} = \frac{l_2\hat{k}}{\Gamma(1+\hat{k})(1-2^{-k})} \tag{3.16}$$

$$\hat{\xi} = l_1 + \frac{\hat{\alpha}}{\hat{k}}\left[\hat{\Gamma}(1+\hat{k})-1\right] \tag{3.17}$$

which $C = \frac{2}{3+t_3} - \frac{log2}{log3}$, where t_3 was derived from Eq. (3.11), while l_2 and l_1 were derived from Eq. (3.7) and Eq. (3.6) correspondingly.

3.4.2 Generalized Pareto Distribution

The Generalized Pareto distribution (GPA) had been quite used in hydrology for modelling of probability related to the occurrence of an extreme flood. This distribution was especially good in the fitting of high-accuracy tail data of the flood distributions. The probability density function of the three-parameters GPA distribution was given by: where ξ represented the location parameter, α the scale parameter, and k the shape parameter (Hamed & Rao, 2019).

$$f(x) = \frac{1}{\alpha}\left[1 - \frac{k}{\alpha}(x-\xi)\right]^{\frac{1}{k}-1} \tag{3.18}$$

Equation (3.19) below depicted the cumulative density function of GPA given by

$$F(x) = 1 - \left[1 - \frac{k}{\alpha}(x-\xi)\right]^{\frac{1}{k}} \tag{3.19}$$

where ξ represented the location parameter, α was the scale parameter, k was the shape parameter. Thus, the three parameters of ξ, α and k in GPA distribution can be estimated below by, (Hamed & Rao, 2019).

$$\hat{k} = \frac{1-3t_3}{1+t_3} \tag{3.20}$$

$$\hat{\alpha} = l_2(1+\hat{k})(2+\hat{k}) \tag{3.21}$$

$$\hat{\xi} = l_1 - l_2(2+\hat{k}) \tag{3.22}$$

where Eq. (3.6) and Eq. (3.7) provide the values of l_1 and l_2, respectively and t_3 value can be derived using Eq. (3.11).

3.4.3 Generalized Logistic Distribution

The Generalized Logistic (GLO) distribution was the statistical distribution used in the estimation of the highest annual flood peak values. It was mainly adopted to estimate the parameters of data sets that exhibit a distribution pattern similar to the logistic distribution. This method was quite useful in predicting the frequency and severity of extreme flooding events. Hence, the probability density function of the above distribution can be expressed as:

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{\left(\frac{1}{k}-1\right)}\left[1+\left\{\ 1-k\left(\frac{x-\xi}{\alpha}\right)\right\}^{\ 1/k}\right]^{-2} \tag{3.23}$$

where x was a random variable, ξ was the location parameter, α was the scale parameter, and k was shape parameter.

Next, the cumulative density function for GLO was given by:

$$F(x) = \left[1+\left\{\ 1-k\left(\frac{x-\xi}{\alpha}\right)\right\}^{\ 1/k}\right]^{-1} \tag{3.24}$$

where x was a random variable, ξ was the location parameter, α was the scale parameter, and k was shape parameter. Therefore, the three parameters of ξ, α and k in GLO distribution could be estimated as given below, (Hamed & Rao, 2019):

$$\hat{k} = -t_3 \tag{3.25}$$

$$\hat{\alpha} = \frac{l_2}{\Gamma(\hat{k})\left[\Gamma(1-\hat{k})-\Gamma(2-\hat{k})\right]} \tag{3.26}$$

$$\hat{\xi} = l_1 - \frac{\hat{\alpha}}{\hat{k}} + \hat{\alpha}\hat{\Gamma}(\hat{k})\Gamma(1-\hat{k}) \tag{3.27}$$

where t_3, and l_1 was obtained by Eq. (3.11) and Eq. (3.6) respectively.

3.4.4 Gringorten Plotting Position

Gringoten formula mainly used for plotting a set of ordered observations in a simple manner. According to Gringorten (1963), the best plotting position was achieved by using the Gringorten plotting position. To show the effectiveness of plotting position, even when the sample size was less than 20, this method was applied since it was good. Therefore, the formula for the Gringorten plotting was given by:

$$P_i = \frac{i - 0.44}{n + 0.12} \tag{3.28}$$

where P_i was the plotting position for the ith value, i was the rank of the value in the ordered dataset, and n was the sample size in the database.

Euclidean Distance

Euclidean Distance was adopted with the Gringorten Plotting Position in this paper to assess how ccuracy and dependable each distributions model was with the observed data. The formula of Euclidean distance between two point were as follow (Academy, 2023):

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{3.29}$$

where d was the Euclidean distance between the points, (x_1, y_1) was the coordinates for the first point and (x_2, y_2) for the second point.

3.4.5 Quantile Estimate using L-Moment

Quantile estimate were being computed after the estimation of parameter which related with different return periods. Return period of probability non-exceedance (F) refer as (Acharya & Joshi, 2020),

$$F = 1 - \frac{1}{T} \tag{3.30}$$

where T was the return period such as T= 2, 10, 50 and 100 years. Thus, the quantile estimate for GEV using L-Moment refer as shown below:

$$x_T = \xi + \frac{\alpha}{k}\left\{1 - [-\log(F)]^k\right\} \tag{3.31}$$

where ξ was the location parameter, k was the shape parameter, and F was the probability of non-exceedance.

Next, the quantile estimate for GPA using L-Moment was referred as:

$$x_T = \xi + \frac{\alpha}{k}\left[1 - (1-F)^k\right] \tag{3.32}$$

where ξ was the location parameter, α was the scale parameter, $k > 0$ was the shape parameter, T was the return period of formula of probability of non-exceedance $F = 1 - 1/T$.

Lastly, for quantile estimate of GLO using L-Moment is:

$$x_T = \xi + \frac{\alpha}{k}\left[1 - \{\ (1-F)/F\}\ ^k\right] \tag{3.33}$$

where x was random variable, ξ was the location parameter, α was the scale parameter, and k was shape parameter, and F was the probability of non-exceedance.

3.5 PERFORMANCE MEASUREMENT

Performance measurement were used to calculate the accuracy of the distributions for any improvement soon in the area of study. There were three sub-sections of this performance measurement in this study, which include the accuracy performance measurement, the L-Moment Ratio Diagram (LMRD), and the Goodness-Of-Fit (GOF) test.

3.5.1 Accuracy Performance Measurement

Five accuracy performance measurement used in this research were The Mean Absolute Error (MAE), The Mean Absolute Percentage Error (MAPE), The Root Mean Square Error (RMSE), The Root Mean Square Percentage Error (RMSPE), and The Coefficient of Determination (R^2) (Badyalina et al., 2022).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \mid F(y_i) - F(\hat{y}_i) \mid \tag{3.34}$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{F(y_i) - F(\hat{y}_i)}{F(y_i)} \right| \tag{3.35}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(F(y_i) - F(\hat{y}_i))^2}{n}} \tag{3.36}$$

$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{F(y_i) - F(\hat{y}_i)}{F(y_i)}\right)^2} \times 100\% \tag{3.37}$$

$$R^2 = \frac{\sum_{i=1}^{n} F(\hat{y}_i) - F(\bar{y}_i)^2}{\sum_{i=1}^{n} F(\hat{y}_i) - F(\bar{y}_i)^2 + \sum_{i=1}^{n} F(y_i) - F(\hat{y}_i)^2} \tag{3.38}$$

where n was the number of observations, $F(y_i)$ represented the actual values, and $F(\hat{y}_i)$ represented the predicted values for the i-th observation, and $F(\bar{y}_i)$ represented the mean of the actual values.

3.5.2 L-Moment Ratio Diagram

The L-Moment Ratio Diagram (LMRD), proposed by J. R. M. Hosking and Wallis (1997), was a useful tool for determining an appropriate distribution that accurately depicted the catchment's streamflow series. To create a ratio diagram, one must had a straightforward explicit expression for t_4 in terms of t_3 for the chosen probability distributions. The followed types of polynomial approximations had produced:

$$t_4 = A_0 + A_1 t_3 + A_2 (t_3)^2 + A_3 (t_3)^3 + A_4 (t_3)^4 + A_5 (t_3)^5 + A_6 (t_3)^6 + A_7 (t_3)^7 + A_8 (t_3)^8 \quad (3.39)$$

Table 3.1: Polynomial approximations of t_4 as a function of t_3 based on L-moment method

	GPA	GEV	GLO
A_0	0	0.10701	0.16667
A_1	0.20196	0.11090	0
A_2	0.95924	0.84838	0.83333
A_3	-0.20096	-0.06669	0
A_4	0.04061	0.000567	0
A_5	0	-0.04208	0
A_6	0	0.03763	0
A_7	0	0	0
A_8	0	0	0

Table 3.1 provided the coefficients of Ak for the GEV, GPA,and GLO distributions, respectively, based on L-Moment, where k = 0,1,2,...8. By substituted t_3 over the range $-0.5 \leq t_3 \leq 0.9$ into Eq. (3.39) yield values of t_4 for each distribution respectively. The range selected because mostly t_3 from fall between the range -0.5 and 0.9. The constructed LMRD for GEV, GPA, and GLO distributions respectively will be illustrated in Figure 3.3.

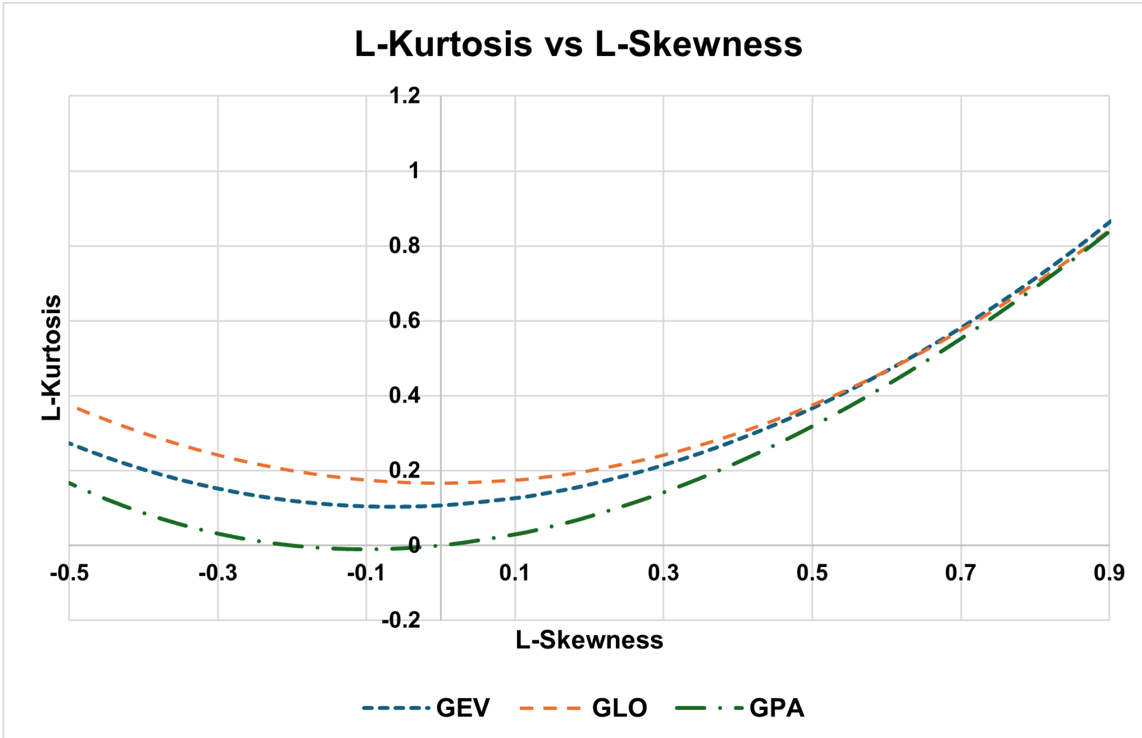


Figure 3.3: L-Moment Ratio Diagram

3.5.3 Goodness Of Fit

There were several goodness-of-fit tests commonly used in this field of study. However, this study only focused on the Anderson-Darling Test and Kolmogorov-Smirnov Test because they demonstrated greater rejection power compared to other tests to determined an appropriate model (Heo et al., 2013).

3.5.3.1 Anderson-Darling Test

The Anderson-Darling (AD) test was a statistical test specifically developed to assess the degree of fit between a distribution and a given dataset. It was particularly useful in flood frequency analysis due to its ability to identify deviations in the extreme ends of the distribution. This was especially important when analysed the extreme events such as flooding.

H_0: The data follow specified distribution.

H_1: The data does not follow the specified distribution.

P-value must be less than α to reject null hypothesis.

Hence, AD test statistic was as followed:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\ln(F(X_{(i)})) + \ln(1 - F(X_{(n-i+1)}))\right] \tag{3.40}$$

where n was the sample size, $F(X_{(i)})$ was the empirical cumulative distribution function of the ordered sample $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.

3.5.3.2 Kolmogorov-Smirnov Test

On the other hand, the Kolmogorov-Smirnov (K-S) test was another nonparametric test that was commonly employed in statistics for comparing distributions, (Hamed & Rao, 2019).

H_0: The data follow specified distribution.

H_1: The data does not follow the specified distribution.

P-value must be less than α to reject null hypothesis.

Hence, K-S test statistic, D_{N_j} was defined in Eq. (3.41) (Hamed & Rao, 2019)

$$D_{N_j} = max|F_N(x) - F_o(x)| \tag{3.41}$$

where N_j was the cumulative number of sample events at class limit j, and the values of $F_N(x)$ were estimated. 1/k, 2/k,..., etc. were the values of $F_o(x)$, where k was the number of class intervals.

3.6 MEASUREMENT RANKING

This study concluded with a ranking of the distributions to identify which was best suited for the target location. The performance measurements of every distribution determined in the previous step were compared using this method. Accuracy measurement (MAE, MAPE, RMSE, RMSPE, R^2), LMRD, and GOF (AD and K-S) were among the measurements. Three point were given to the distribution that fit the best, and one point was given to the distribution that fit the least. The best distribution then found by adding the scores of the three potential distributions (GEV, GPA, and GLO). Since it best satisfied the performance requirements for the target location, the distribution with the highest overall score was determined to be the best distribution for this study.

3.7 SUMMARY

The method of analysis of the research of flood frequency analysis based on L-Moment in Peninsular Malaysia can be summarised as followed:

Table 3.2: Summary of data analysis

Research Objectives	**Method of Analysis**
To utilise L-moment for the estimation of parameters of extreme value distributions used in modelling flood data	L-Moment
To develop and evaluate a precise flood frequency model using extreme value distributions for predicting the occurrence and magnitude of floods in Peninsular Malaysia's river systems.	Measurement Ranking
To estimate the return periods of various flood magnitudes by analyzing historical streamflow data and applying extreme value distributions	Quantile Estimate

CHAPTER 4

RESULTS AND DISCUSSION

4.1 INTRODUCTION

Flood frequency analysis is essential for comprehending and forecasting the incidence and intensity of flood events over time. It is extensively utilised in hydrological research to assess the likelihood of different flood magnitudes occurring within designated return periods, thus guiding flood risk management and planning. The main purpose of flood frequency analysis mainly focusses on the design of flood mitigation infrastructure, land-use planning, and disaster risk management.

This research involves the estimation of the frequency at which floods of different magnitudes occur, the examination of past streamflow data to identify the statistical properties of floods, and the provision of practical recommendations for the management of flood risks in the area. The aim of this study is to enhance the understanding of flood patterns and their statistical properties by employing extreme value distributions and estimating their parameters through L-moments.As a result, this approach is expecting to lead to improve flood preparedness and resilience in Peninsular Malaysia.

Hence, this chapter delineates the outcomes of the Flood Frequency Analysis study conducted in Peninsular Malaysia, using the L-Moment approach. The analyses include the derivation of flood frequency using extreme value distributions such as Generalized Pareto (GPA), Generalized Extreme Value (GEV), and Generalized Logistic (GLO). As such, estimation of parameters under these distributions is to be evaluated, together with their performances under a ranking scheme using various criteria.

4.2 ANALYSIS OF FLOOD FREQUENCY

This segment includes the examination performed at Sg. Triang in Kg. Chenor, Negeri Sembilan (2920432), which comprises L-Moment (LMO) analysis, parameter estimation for various distribution, Gringorten Plotting Position, comparisons of Goodness-of-Fit (GOF) tests, assessments of accuracy performance measures, L-Moment Ratio Diagrams (LMR), ranking procedures, and quantile estimation methodologies.

4.2.1 L-Moment

The following table delineates the initial four elements involved in the computation of L-Moments. These components—mean (l_1), scale (l_2), skewness (l_3), and kurtosis (l_4)—provide insights into the statistical characteristics of the flood data at this station.

Table 4.1: L-Moment Components

l_1 **(Mean)**	l_2 **(Scale)**	l_3 **(Skewness)**	l_4 **(Kurtosis)**
4.3715365	1.6051485	0.2264257	0.2397717

Table 4.1 displays the initial four components of LMO, outlining the statistical attributes of the flood data. The l_1 (4.3715) as in Eq. (3.6), signifying the central tendency of the data and reflecting the average magnitude of the flood events. The l_2 (1.6051) obtained from Eq. (3.7), indicates the variability of the data around the mean, demonstrating moderate variation in flood magnitudes. The l_3 (0.2264) from Eq. (3.8) signifies a minor positive asymmetry in the data distribution, implying that high-magnitude floods occur infrequently but are not common. Finally, the l_4 (0.2398) obtained from Eq. (3.9), signifies light tails in the distribution, suggesting that extreme flood events are rare.

4.2.2 Estimation of Parameters using Extreme Value Distributions

This section presents the estimated distribution parameters calculated using the L-Moment method. The parameters include the scale parameter ($\hat{\alpha}$), location parameter ($\hat{\xi}$), and shape parameter ($\hat{k}$) for three different probability distributions: GEV, GPA and GLO

Table 4.2: Estimated Distribution Parameters Using L-Moment Methods

Distribution	$\hat{\alpha}$	$\hat{\xi}$	$\hat{k}$
GEV	2.41087887	3.08395543	0.04542615
GPA	6.0547059	0.34982765	0.5055058
GLO	1.5531227	4.0027214	-0.1410622

Table 4.2 presents the estimated distribution parameters for GEV, GPA, and GLO distributions, derived using the LMO method. The GEV $\hat{\alpha}$ value of 2.4101, as indicated by Eq. (3.16), signifies the scale and variability of flood magnitude. The value 3.0840 of $\hat{\xi}$, based on Eq. (3.17), signifies the central tendency of flood magnitude. The positive value 0.0454 of $\hat{k}$ indicates a slightly heavy tailed behaviour, suggesting potential occurrence of significant floods in the study area.

The estimated parameters for GPA distribution using L-moment method has the highest $\hat{\alpha}$ (6.0547) among other distributions, obtained from Eq. (3.21). This value indicates a wider spread in flood magnitudes, reflecting greater variability flood events. This broad range may necessitates adaptable flood prevention measures to handle varying flood sizes (Hamed & Rao, 2019). The value of $\hat{\xi}$ (0.3498), as per Eq. (3.22), is the lowest among the distributions, indicating the low flood magnitude. A positive $\hat{k}$ (0.5055), derived from Eq. (3.20), GPA exhibits a heavy tail, implying a higher likelihood of rare and extreme flood occurrences.

The GLO distribution exhibits the lowest value for $\hat{\alpha}$ in comparison of other distributions as in Eq. (3.26). This value indicating that the limited variability in flood magnitude. The GLO distribution has the highest value $\hat{\xi}$ (4.0027) as indicated in Eq. (3.27). This signify that flood magnitude modelled by GLO is the highest and GLO may be utilised to forecast more significant and impactful flood occurrences. By using Eq. (3.25), GLO distribution shows a negative $\hat{k}$ value that specify a bounded tail distribution. This value suggests that the model have more frequent events with fewer extremes, which may require consistent but moderate flood management strategies (Hamed & Rao, 2019).

4.2.3 Gringorten Plotting Position

This section intends to illustrate the plotting position of each distribution at the study area. As a result, it will display the plotting position of the study area along with a brief explanation related to each distribution.

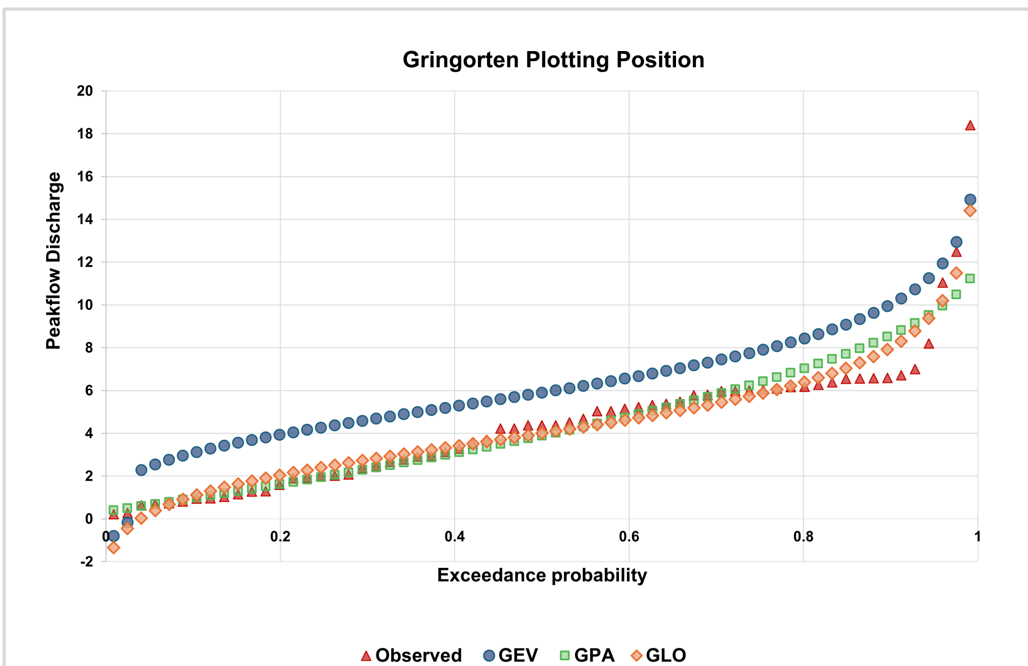


Figure 4.1: Gringorten Plotting Position

Figure 4.1 illustrates the Gringorten Plotting Position graph of the comparison between observed peak flow measurements and predicted values derived from the GEV, GPA, and GLO distributions. The x-axis denotes the Gringorten plotting position probability, computed using Gringorten formula in Eq. (3.28), whereas the y-axis represent the peakflow discharge m^3/s. The graph shows that both GPA and GLO distributions closely related with the observed values, specify a strong fit to the data (Samat & Othman, 2023).

In order to evaluate the accuracy and reliability of the plotting positions relative to the fitted extreme value distributions, the Euclidean distance metric is employed as in Eq. (4.3). This approach quantified the deviation between observed plotting positions and the theoretical quantiles obtained from the distributions. The Euclidean distance provide a quantitative measure of the goodness-of-fit, allowing a direct comparison between observed data and predictions of GEV, GPA and GLO distributions.

Table 4.3: Euclidean Distance for Each Distributions

Distribution	**Euclidean Distance**
GEV	7.050357
GPA	9.196464
GLO	6.224768

Table 4.3 displays the distance between the observed and predicted values for the GEV, GPA, and GLO distributions utilising Gringorten plotting positions. Results shows that the GLO distributions is the most suitable model for describing the flood frequency analysis at the analysed station. With the smallest Euclidean distance of 6.2248, the GLO distribution aligns most closely with the observed data, indicating a better predictive model. The GEV distribution with Euclidean distance of 7.0504 indicate a better model than GPA with 9.1965 but did not outperform the GLO distribution. GPA distribution with the highest Euclidean distance indicate the least fit distribution with the observed data among the three distributions.

4.2.4 Goodness-of-Fit

This section highlights the outcomes of the Goodness-of-Fit (GOF) tests for the three probability distributions — GEV, GPA and GLO— using the Anderson-Darling (AD) test and Kolomogorov-Smirnov (K-S) test. The result are as follow:

Table 4.4: P-value of GOF Test for Each Distributions

Distribution	GEV	GPA	GLO
AD	0.3979	0.000009524	0.3684
K-S	0.4244	0.1567	0.5923

Table 4.4 presents the p-value from the GOF tests for GEV, GPA and GLO distributions. In terms of GOF, as in Eq. (3.40), the AD test p-value for GEV (0.3979) indicate the best fit, while the GPA has the smallest AD p-value (0.000009524), suggesting a poor fit to the data. Similarly, as per Eq. (3.41) the K-S test p-value further confirm that GLO (0.5923) fit the data well, while the GPA distribution (0.1567) performs the least in this test than GEV and GLO.

4.2.5 Accuracy Performance Measure

This section entails the addition of several performance metrics including MAPE, MAE, RMSE, RMSPE, and R^2. These measures evaluate the suitability and precision of the distributions in modelling the flood data at this station.

Table 4.5: Test Performance Measurement for Distributions

Distribution	GEV	GPA	GLO
MAPE	0.2275573	0.1270529	0.3044755
MAE	0.536894	0.5826525	0.5396244
RMSE	0.8882615	1.158646	0.784247
RMSPE	65.05872	19.93939	99.92112
R^2	0.9180464	0.858907	0.9366164

Regarding performance metrics in Table 4.5, as per Eq. (3.35) the GPA distribution shows the lowest MAPE (0.1271), indicate that it has the smallest average percentage error, thus providing the most accurate predictions on average than other distributions. For MAE in Eq. (3.34), the GEV (0.5369) performing better than other distribution, indicating more accurate absolute estimates. The RMSE in Eq. (3.36) values highlight that the GLO distribution (0.7842) has the lowest value, indicating it performs best in minimising large prediction errors.

For metric RMSPE using Eq. (3.37), the GPA distribution (19.9394) has the lowest value, which suggests it performs better than GEV (65.05872) and GLO (99.9211). Finally, the R^2 as per Eq. (3.38) values indicate that the GLO distribution (0.9366) explains the highest proportion of variance in the data, followed by GEV (0.9180) and the GPA distribution (0.8589).

4.2.6 L-Moment Ratio Diagram

This section intends to elucidate the L-Moment Ratio (LMR) for three distributions in Peninsular Malaysia. This section will present the output for all three distributions, accompanied by a brief discussion for each distribution.

Table 4.6: L-Moment Ratios of Candidate Distribution

L-CV (t_2)	**L-Skewness (t_3)**	**L-Kurtosis (t_4)**
0.3672	0.1411	0.1494

Table 4.6 presents LMR for a river station in Peninsular Malaysia, capturing essential attributes of the hydrological data via three LMO statistics: t_2 (L-CV), t_3 (L-skewness), and t_4 (L-kurtosis) which describe the standard deviation, asymmetry and tailed of the distributions, respectively.

The t_2 value of 0.3672 as per Eq. (3.10) signifies moderate variability in the data, demonstrating a balance between consistency and dispersion in the observed values. The t_3 value, 0.1411, from Eq. (3.11), indicates moderate right skewness, reflects moderate right skewness, therefore, suggest that the distribution has a tendency to have a longer tail on the right side, though not markedly pronounced. From Eq.(3.12), t_4 value of 0.1494 indicates that the tails in the distribution are moderately heavy, suggesting that extreme values can be present but not to a marked degree (Malá et al., 2022; Peel et al., 2001).

These L-Moment Ratios (LMR) add to a deeper understanding of the behaviour of the hydrological data and aid in informing flood prediction models. The consideration of these three L-Moment statistics brings about a much more holistic view of how the data behaves, especially in relation to predict uncommon hydrological extremes like flooding events. The constructed L-Moment Ratio Diagram (LMRD) for the three distributions is illustrated in Figure 4.6.

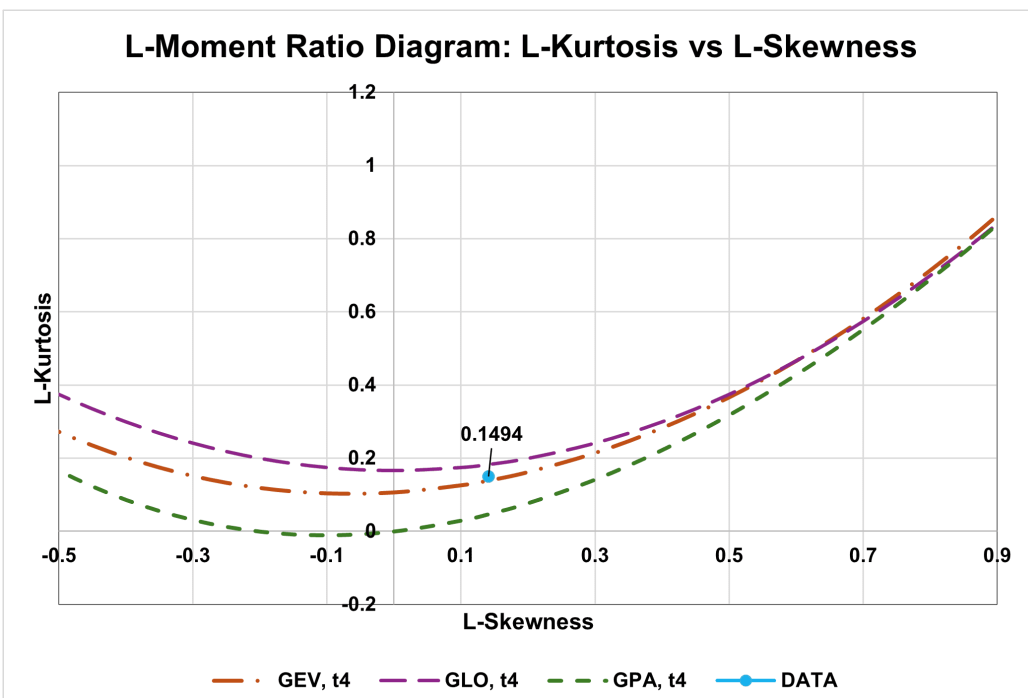


Figure 4.2: L-Moment Ratio Diagram

Figure 4.2 shows the LMRD depicting the relation of L-Skewness (along x-axis) and L-Kurtosis (along y-axis) for the observed data along with three theoretical extreme value distributions: GEV, GPA and GLO distributions. Theoretical curves for each distribution are shown in the diagram, and the observed data point (0.1494) is plotted at a position determined by its respective L-Skewness and L-Kurtosis values.

The observed data point (0.1494) is closest to the curve of GEV, which implies that GEV model is the most suitable for the data. Observed data shows more scatter from the GLO and GPA curves. This indicates that both distributions are less accurate in modelling the observed data. The major outcome from the LMRD reveals that the GEV distribution fits best for this dataset, hence effective in flood frequency analysis and promising to represent correctly the statistical properties of observed data.

4.2.7 Measurement Ranking

A score three, which is the topmost point, is given to the best fits distribution of each goodness-of-fits and accuracy performance measure, while the least fit gets a point value of one. The optimum distribution is obtained by summing the scores of the three candidate distributions namely, Generalized Extreme Value – GEV, Generalized Pareto – GPA, and Generalized Logistic – GLO. The distribution scoring the highest overall is considered the most suited for this study since it satsifies the performance measures best at the location in question.

Table 4.7: Rank Score for Distributions

Distribution	**GEV**	**GPA**	**GLO**
AD	3	1	2
K-S	2	1	3
MAPE	2	3	1
MAE	3	1	2
RMSE	2	1	3
RMSPE	2	3	1
R^2	2	1	3
Total Score	**16**	**11**	**15**

Table 4.7 puts the Generalized Extreme Value (GEV) distribution at the forefront as the best suitable model for flood frequency model for this study. This can be supported by its outstanding performance in the GOF test specifically AD and K-S Test, combined with its good results in other accuracy measures. Besides, GEV obtained the highest cumulative rank score in Table 4.7, further proving its appropriateness for this station. By contrast, GLO and GPA distributions were the least favourable ones with regard to their total ranks scores being the lowest among all. The present study then concludes that the GEV distribution is the best and most appropriate model for the purpose of flood frequency analysis in the river system of Peninsular Malaysia.

4.2.8 Quantile Estimate

Flood quantile estimates are very important in assessing flood risk and formulating mitigation plans. In this study, quantile estimates of the study station are estimated using three probability distributions: GEV, GPA, and GLO. Table 4.8 are derived for the study station using three probability distributions: GEV, GPA and GLO. These distributions are employed to estimate the expected flood discharge for designated return periods (10, 50, and 100 years) at specified probabilities (p). The results are presented in cubic meters per second (m^3/s), indicating the potential level of flood events at various intervals.

The flood discharge of GEV distributions, as indicated in Table 4.8 exhibits a rising pattern with longer return period. For a return period 10 years (p = 0.90), the estimated

Table 4.8: Quantile Estimates Based on Return Periods

Return Period (Years)	**Probability (p)**	**Estimated Flood Discharge (m^3/s)**		
		GEV	**GPA**	**GLO**
10	0.90	8.241226	8.228045	8.382849
50	0.98	11.704503	12.111101	10.652548
100	0.99	13.092262	14.074898	11.153635

flood discharge is 8.2412 m^3/s as per in Eq. (3.31) conjunction with Eq. (2.7). The value escalates to 50 years (p = 0.98) at 11.7045 m^3/s and reaches 13.0923 at return period 100 years with probability of 0.99. This consistent progression highlights the GEV distribution's ability to model increasing flood magnitudes for less frequent, extreme events.

In comparison to alternatives distributions, the GEV model predicts slightly lower flood discharges than the GLO distribution for shorter return periods (8.241 m^3/s vs 8.3828 m^3/s for a 10-years), yet higher than the GPA distribution (8.2280 m^3/s) as calculated using Eq. (3.32). For longer return periods, such as 100 years, the GEV distribution provides more conservative estimates (13.0923 m^3/s) compared to GPA (14.0749 m^3/s). For a 50-year return period, GEV gives a higher magnitude compared to GLO.It means that the GEV distribution might be especially appropriate for applications where conservative flood risk estimates are needed, not leading to overestimation of extreme flood magnitudes.

In general, the GEV distribution is a robust choice for flood modelling, as it provides realistic estimates over a wide range of return periods. This makes it specially valuable for long-term flood risk management and the design of infrastructure to withstand extreme hydrological events because of its ability to provide consistent and conservative predictions.

4.3 SUMMARY OF RESULTS AND DISCUSSION

The L-Moment analysis provided a lot of interesting statistical information about the flood data: the mean value (l_1) of 4.3175, indicates the average size of floods; the scale parameter (l_2) of 1.6051, suggests moderate variability. Moreover, slight positive assymmetry is indicate by skewness(l_3) 0.2264, and light tails by kurtosis (l_4) 0.2398, suggesting the extreme events are relatively rare. The parameter estimation for the three distributions (GEV, GPA and GLO), exhibited significant variations.The GPA had the highest scale parameter ($\hat{\alpha} = 6.0547$), indicating larger variability, while the GLO had the highest location parameter ($\hat{\xi} = 4.0027$), reflecting more extreme flood magnitudes.. The shape parameter ($\hat{k}$) was positive for GPA, which suggesting a heavy tail, while for GLO shows negative value, which implies a bounded tail.

The Gringorten plotting position analysis revealed that GPA and GLO distributions closely matched observed flood values. However, after evaluation by using Euclidean distance it is proven that GLO distribution is better predictive model. The LMRD provided additional validation, with L-CV (t_2) at 0.3672 indicating a moderate right variability, L-Skewness (t_3) at 0.1411 reflecting moderate right-skewness and, L-kurtosis (t_4) at 0.1494 suggesting moderately heavy tails. These results confirmed the suitability of GEV, GPA, and GLO distributions for flood frequency modelling. Lastly, ranking the distributions based on GOF and accuracy measures, GEV emerged as the optimal distribution with the highest total score (16), followed by GLO (15) and GPA (11).

Finally, the quantile estimate highlighted the possible discharge levels of flooding for various return periods. Overall, the Generalized Extreme Value (GEV) distribution proved to have the best reliability as a model for flood frequency analysis at the investigated station due to its good performance in most evaluation criteria.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

This research provides a holistic framework for understanding and managing flood risks in Peninsular Malaysia, which is significantly prone to floods caused by monsoons. This study applies the state-of-the-art statistical techniques, such as L-moments and extreme value distributions (Generalized Extreme Value, Generalized Pareto, and Generalized Logistic), in order to develop a robust model for flood magnitude and frequency predictions. The results demonstrate the efficiency of these approaches in analyzing long-term records, estimating return periods, and identifying appropriate probability distributions for different flood situations. This result contributes to the academic understanding of flood frequency analysis and also has practical implications in disaster mitigation applications.

The importance of detailed data analysis to inform policy decisions that will improve infrastructure resilience and community preparedness is shown throughout the research. It calls for its findings to be integrated into regional planning and flood management strategies—especially in view of challenges presented by rapid urbanization and climate change. The results shows that the Generalized Exreme Value (GEV) distribution is the most reliable model for predicting flood magnitudes across different return periods due to its consistent in outperform other distributions based on goodness-of-fit tests and accuracy measures. In addition, the quantile estimate from GEV also gives a valuable insight into potential flood magnitudes, showing an increase in food discharge for 10, 50, and 100 years at 8.24 m^3/s, 11.70 m^3/s, and 13.09 m^3/s, respectively. These findings provide a solid foundation to more effective preparedness and mitigation efforts in the study area.

5.2 RECOMMENDATIONS

From the results, a number of recommendations can be put forward in order to improve flood management and preparedness. Firstly, policymakers and engineers should emphasize the use of GEV distribution in designing flood control structures such as dams and bridges. This is because the research identifies GEV distribution as the most optimal distribution for modelling extreme flood events. Next, government and policymakers should employ quantile estimates to mitigate flood impacts in high-risk areas. This approach would help the development of targeted strategies and efficient allocating resources. Collaboration of government with any relevant academic institution also essential to enhance the acccuracy of real-time streamflow data. This approach will improve the precision of flood frequency models and lead to more effective flood management in future studies.

For future research, researcher are encouraged to apply the L-Moment method in other regions with different hydrological conditions or limited data availability. This would allow for broader validation in using L-Moment method. Expanding the analysis to other regions in Malaysia will help in validating the adaptability and robustness of GEV model, especially for areas not covered in study. Lastly, future studies should also consider to cover more river systems and regions to examine the impacts of climate changes on return periods and flood magnitudes for future scenario planning. This action will provide valuable information for future scenario planning and help to understand how climate change may influence flood risks and preparedness.

REFERENCES

Abid, S. K., Sulaiman, N., Wei, C. S., & Nazir, U. (2021). Flood vulnerability and resilience: Exploring the factors that influence flooding in sarawak. *IOP Conference Series: Earth and Environmental Science*, *802*(1), 012059. https://doi.org/https://doi.org/10.1088/1755-1315/802/1/012059

Academy, E. (2023, August). How does euclidean distance help measure the similarity between data points in machine learning? - eitca academy. https://eitca.org/artificial-intelligence/eitc-ai-mlp-machine-learning-with-python/programming-machine-learning/euclidean-distance/examination-review-euclidean-distance/how-does-euclidean-distance-help-measure-the-similarity-between-data-points-in-machine-learning/

Acharya, B., & Joshi, B. (2020). Flood frequency analysis for an ungauged himalayan river basin using different methods: A case study of modi khola, parbat, nepal. *Meteorology Hydrology and Water Management. Research and Operational Applications*, *8*(2), 46–51.

Ahmad, A. M., Romali, N. S., & Sulong, S. (2023). Flood frequency analysis of annual maximum stream flows for kuantan river basin. *AIP Conference Proceedings*, *2688*(1).

Ali, S., & Rahman, A. (2022). Development of a kriging-based regional flood frequency analysis technique for south-east australia. *Natural Hazards*, *114*(3), 2739–2765.

Badyalina, B., Mokhtar, N. A., Mat, J. N. A., Hassim, N. H., & Yusop, H. (2021). Flood frequency analysis using l-moment for segamat river. *Matematika*, 47–62.

Badyalina, B., Mokhtar, N. A., Ya'acob, F. F., Ramli, M. F., Majid, M., & Jan, N. A. M. (2022). Flood frequency analysis using l-moments at labis in bekok river station. *Applied Mathematical Sciences*, *16*(11), 529–536.

Baidya, S., Singh, A., & Panda, S. N. (2020). Flood frequency analysis. *Natural Hazards*, *100*(3), 1137–1158.

Bakri, S. N. F. A. (2022). At-site and regional frequency analysis of extreme rainfall modelling in peninsular malaysia. *Applied Mathematics and Computational Intelligence (AMCI)*, *11*(2), 534–549.

Bernama. (2021). N. sembilan's driest district, jelebu, struck by worst-ever floods. https://thesun.my/malaysia-news/n-sembilan-s-driest-district-jelebu-struck-by-worst-ever-floods-XE8701123

Campos-Aranda, D. F. (2016). Fitting of the gev, glo and gpa distributions with trimmed l moments (1, 1) in 21 annual flood records of the hydrological region no. 10 (sinaloa), mexico. *Agrociencia*, *50*(1), 15–31.

Che Ilias, I. S., Wan Zin, W. Z., & Jemain, A. A. (2021). Regional frequency analysis of extreme precipitation in peninsular malaysia.

Dalrymple, T. (1960). *Flood-frequency analyses*. US Government Printing Office.

Davies, R. (2023, January). Malaysia – floods in johor and pahang after 430mm of rain in 24 hours. https://floodlist.com/asia/malaysia-floods-johor-pahang-sabah-january-2023

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water resources research*, *15*(5), 1049–1054.

Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, *68*(3), 813–814.

Hamed, K., & Rao, A. R. (2019). *Flood frequency analysis*. CRC press.

Hamzah, F. M. (2020). Flood frequency analysis of annual maximum of high tides. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1.2), 6–11. https://doi.org/https://doi.org/10.30534/ijatcse/2020/0291.22020

Hamzah, F. M., Tajudin, H., Abdullah, S. M. S., Toriman, E., & Juahir, H. (2020). Flood frequency analysis of tidal data at pelabuhan klang, malaysia. *International Journal of Engineering Trends and Technology (IJETT)*, (3), 92–96. https://doi.org/10.14445/22315381/CATI3P215

Hassim, N. H., Badyalina, B., Mokhtar, N. A., Abd Jalal, M. Z. H., Zamani, N. D. B., Kerk, L. C., Zainoddin, A. I., Fadzil, A. S. A., & Shaari, N. F. (2022). In situ flood frequency analysis used for water resource management in kelantan river basin. *AL JOUR*, *11*(5), 9.

Heo, J.-H., Shin, H., Nam, W., Om, J., & Jeong, C. (2013). Approximation of modified anderson–darling test statistics for extreme value distributions with unknown shape parameter. *Journal of Hydrology*, *499*, 41–49.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis*.

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251–261.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *52*(1), 105–124.

Hossain, I., Imteaz, M., & Khastagir, A. (2021). Study of various techniques for estimating the generalised extreme value distribution parameters. *IOP Conference Series: Materials Science and Engineering*, *1067*(1), 012065.

Jan, N. A. M., Marsani, M. F., Thiruchelvam, L., Abidin, N. B. Z., Shabri, A., & Sani, S. A. A. (2023). Mitigating infectious disease risks through non-stationary flood frequency analysis: A case study in malaysia based on natural disaster reduction strategy. *Geospatial Health*, *18*(2).

Kousar, S., Khan, A. R., Ul Hassan, M., Noreen, Z., & Bhatti, S. H. (2020). Some best-fit probability distributions for at-site flood frequency analysis of the ume river. *Journal of Flood Risk Management*, *13*(3), e12640.

Krishna, G. S., & Veerendra, G. (2015). Flood frequency analysis of prakasam barrage reservoir krishna district, andhra pradesh using weibull, gringorten and l-moments formula. *Int. J. Civil. Struct. Env. Infr. Eng. Res. Dev*, *5*.

Legrand, J. (2022, September). *Simulation and assessment of multivariate extreme models for environmental data* (Publication No. 2022UPASJ015) [Theses]. Université Paris-Saclay. https://theses.hal.science/tel-03827820

Linsley, R. K. (1986). Flood estimates: How good are they? *Water Resources Research*, *22*(9S), 159S–164S.

Malá, I., Sládek, V., & Habarta, F. (2022). Comparison of estimates using l and tl moments and other robust characteristics of distributional shape and tail heaviness. *REVSTAT-Statistical Journal*, *20*(5), 529–546.

Mohd Baki, A., Mat Yusof, D. A., Atan, I., & Mohd Halim, N. F. (2014). Regional flow frequency analysis on peninsular malaysia using lmoments. *Jurnal Intelek*, *9*(1), 63–68.

Moughamian, M. S., McLaughlin, D. B., & Bras, R. L. (1987). Estimation of flood frequency: An evaluation of two derived distribution procedures. *Water resources research*, *23*(7), 1309–1319.

Peel, M. C., Wang, Q., Vogel, R. M., & McMAHON, T. A. (2001). The utility of l-moment ratio diagrams for selecting a regional probability distribution. *Hydrological sciences journal*, *46*(1), 147–155.

Prahadchai, T., Busababodhin, P., & Park, J.-S. (2024). Regional flood frequency analysis of extreme rainfall in thailand, based on l-moments. *Communications for Statistical Applications and Methods*, *31*(1), 37–53.

Razali, H., & Zolkeply, Z. I. (2023). Floods worsen in kelantan, number of evacuees surpasses 17,000 people. https://www.nst.com.my/news/nation/2023/12/994434/floods-worsen-kelantan-number-evacuees-surpasses-17000-people

Rentschler, J., Salhab, M., & Arga Jafino, B. (2022). Flood risk already affects 1.81 billion people. climate change and unplanned urbanization could worsen exposure. https://blogs.worldbank.org/en/climatechange/flood-risk-already-affects-181-billion-people-climate-change-and-unplanned

Romali, N. S., & Yusop, Z. (2021). Flood damage and risk assessment for urban area in malaysia. *Hydrology Research*, *52*(1), 142–159.

Safiah Yusmah, M., Bracken, L. J., Sahdan, Z., Norhaslina, H., Melasutra, M., Ghaffarianhoseini, A., Sumiliana, S., & Shereen Farisha, A. (2020). Understanding urban flood vulnerability and resilience: A case study of kuantan, pahang, malaysia. *Natural Hazards*, *101*, 551–571.

Samat, S., & Othman, N. (2023). Plotting position for low flow frequency analysis at jempol river streamflow station. *IOP Conference Series: Earth and Environmental Science*, *1140*(1), 012021.

Stedinger, J. R. (1993). Frequency analysis of extreme events. *Handbook of hydrology*.

Sufian, A., Chi, C. J., Azman, H., Aziz, N. A. A., Fen, F. S., & Zamri, A. A. M. (2022). Assessing residents' flood preparedness through adaption of protective behaviour in melaka, malaysia. *Environment and Ecology Research*, *10*(3), 334–335.

Todd, D. K. (1957). Frequency analysis of streamflow data. *Journal of the Hydraulics Division*, *83*(1), 1166–1.

Vivekanandan, N. (2015). Flood frequency analysis using method of moments and l-moments of probability distributions. *Cogent engineering*, *2*(1), 1018704.

Weerabangsa, M., Himanujahn, S., Iresh, A., & Athapattu, B. (2023). Flood frequency analysis for the upper kalu river basin at ratnapura, sri lanka. *2023 Moratuwa Engineering Research Conference (MERCon)*, 403–408.

Yusoff, S. H. M., Hamzah, F. M., & Jaafar, O. (2022). Multiparameter probability distributions of at-site l-moment-based frequency analysis in malaysia. *Int J Mech Eng*, *7*, 724–737.

Zeng, X., Wang, D., & Wu, J. (2015). Evaluating the three methods of goodness of fit test for frequency analysis. *Journal of Risk Analysis and Crisis Response*, *5*(3).

Zolkiply, Z. I. (2023). Flood update: Over 3,000 still affected by floods in terengganu — new straits times. https://www.nst.com.my/news/nation/2023/12/995197/flood-update-over-3000-still-affected-floods-terengganu#google_vignette

APPENDIX A: RESEARCH SCHEDULE

Table 5.1: Research Schedule

No.	Activity	2024										2025	
		March	April	May	June	Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb
1.	Discussing overall topics and title												
2.	Introduction												
3.	Literature Review												
4.	Design Research Methodology												
5.	Submission proposal to supervisor												
6.	Submission proposal to panel												
7.	Presentation proposal												
8.	Result of study												
9.	Interpretation of result												
10.	Conclusion												
11.	Presentation report												
12.	Submission report												

APPENDIX B: TURNITIN PLAGIARISM REPORT

turnitin Page 2 of 69 - Integrity Overview Submission ID trn:oid:::13381:79067561

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

107 Not Cited or Quoted 12%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
6% Publications
9% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

turnitin Page 2 of 69 - Integrity Overview Submission ID trn:oid:::13381:79067561

APPENDIX C: TURNITIN AI WRITING REPORT

28% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

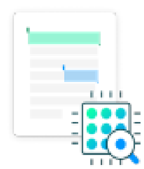
1 AI-generated only 28%
Likely AI-generated text from a large-language model.

2 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.



False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

turnitin Page 2 of 63 - AI Writing Overview Submission ID trn:oid:::13381:79067561