STATISTICAL MODELING OF EXTREME RAINFALL IN PENINSULAR MALAYSIA

LIEW WOON SHEAN

A project report submitted in partial fulfilment of the requirements for the award of Master of Mathematics

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

AUGUST 2020

DECLARATION OF ORIGINALITY

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : ____________________

Name : Liew Woon Shean

ID No. : 1903172

Date : 22 August 2020

APPROVAL FOR SUBMISSION

I certify that this project report entitled “**STATISTICAL MODELING OF EXTREME RAINFALL IN PENINSULAR MALAYSIA**” was prepared by **LIEW WOON SHEAN** has met the required standard for submission in partial fulfilment of the requirements for the award of Master of Mathematics at Universiti Tunku Abdul Rahman.

Approved by,

Signature : ______________________

Supervisor : DR. TAN WEI LUN

Date : 22 August 2020

Approved by,

Signature : ______________________

Co-Supervisor Dr. Ling Lloyd

Date : Aug. 22, 2020

ACKNOWLEDGEMENTS

I want to thank my family and friends for their support and help in finishing the project. Moreover, I want to thank my supervisor Dr. Tan Wei Lun and my co-supervisor Dr. Ling Lloyd for guiding me in completing the project. Lastly, I want to thank UTAR for giving me the opportunity in learning and doing this project.

LIEW WOON SHEAN

STATISTICAL MODELING OF EXTREME RAINFALL IN PENINSULAR MALAYSIA

LIEW WOON SHEAN

ABSTRACT

Flash floods are known as one of the common natural disasters that costs over billions of Ringgit Malaysia throughout the history, and monsoon season that is known as the rainy phase of a seasonally changing pattern is the main period of occurrence of flash floods. Academically, an extreme rainfall model is effective in modeling, so as to predict and prevent the occurrence of flash floods. A reliable extreme rainfall model would help in reduce the cost and rate of mortality in the occurrence of flash floods. This study is to compare four probability distributions, which include Exponential distribution, Generalized Extreme Value (GEV) distribution, Gamma distribution, and Weibull distribution, with the data for rainfall from 10 stations in Peninsular Malaysia for the period of northeast monsoon from November to February. The time span of the data is from 1975 to 2008. The comparison is based on the performance of descriptive and predictive analytics of models. Rainfall data is cleansed by applying peak-over-threshold approach to obtain data that are more suitable to be use in modelling extreme rainfall. Determination of the most effective model is relying on both numerical result through Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared Test, and graphical result through using the quantile-quantile plot. Result shows that GEV is the most preferred extreme rainfall model to the rainfall cases in Peninsular Malaysia.

Table of Contents

TITLE **i**

DECLARATION OF ORIGINALITY **ii**

ACKNOWLEDGEMENTS **vii**

ABSTRACT **viii**

CHAPTER 1 Introduction **1**

1-1 Introduction . . . 1
1-2 Problem Statement . . . 2
1-3 Objectives of Research . . . 2
1-4 Research Question . . . 3
1-5 Project Scope . . . 3

CHAPTER 2 Literature Review **4**

2-1 Introduction . . . 4
2-2 Overview on Extreme Rainfall Modeling . . . 4

CHAPTER 3 Research Methodology **7**

3-1 Introduction . . . 7
3-2 Data Cleansing . . . 7
3-3 Fitting of Probability Distribution . . . 7
3-3-1 Exponential Distribution . . . 8
3-3-2 Generalized Extreme Value (GEV) Distribution . . . 8
3-3-3 Gamma Distribution . . . 8
3-3-4 Weibull Distribution . . . 8
3-4 Parameter Estimator . . . 9

3-4-1 Exponential Distribution 10
3-4-2 Generalized Extreme Value (GEV) Distribution 10
3-4-3 Gamma Distribution 11
3-4-4 Weibull Distribution 12
3-5 Goodness-of-fit Test 12
3-5-1 Kolmogorov-Smirnov Test (K-S Test) 12
3-5-2 Anderson–Darling Test (A–D Test) 13
3-5-3 Chi-Squared Test 13
3-6 Quantile-quantile Plots 13

CHAPTER 4 Results and Discussion **15**
4-1 Introduction 15
4-2 Data Cleansing 15
4-3 Parameter Estimator 20
4-4 Goodness-of-fit Test 23
4-5 Quantile-Quantile Plots 29

CHAPTER 5 Conclusion and Recommendation **41**
5-1 Introduction 41
5-2 Conclusion 41
5-3 Recommendation 42

List of Figures

4.1 Boxplot of Monthly Rainfall Intensity for selected areas in Peninsular Malaysia . . . 16

4.2 Boxplot of Rainfall Intensity for November to February of selected areas in Peninsular Malaysia . . . 18

4.3 Boxplot of Rainfall Intensity for November to February of selected areas in Peninsular Malaysia . . . 19

4.4 Q-Q Plot for extreme rainfall data in Arau with threshold value of 90^{th} percentile. . . . 30

4.5 Q-Q Plot for extreme rainfall data in Pekan Merlimau with threshold value of 90^{th} percentile. . . . 30

4.6 Q-Q Plot for extreme rainfall data in JPS Wilayah Persekutuan with threshold value of 90^{th} percentile. . . . 31

4.7 Q-Q Plot for extreme rainfall data in Ladang Benut Rengam with threshold value of 90^{th} percentile. . . . 31

4.8 Q-Q Plot for extreme rainfall data in Mersing with threshold value of 90^{th} percentile. . . . 32

4.9 Q-Q Plot for extreme rainfall data in Ladang Boh with threshold value of 90^{th} percentile. . . . 32

4.10 Q-Q Plot for extreme rainfall data in Ipoh with threshold value of 90^{th} percentile. . . . 33

4.11 Q-Q Plot for extreme rainfall data in Gua Musang with threshold value of 90^{th} percentile. . . . 33

4.12 Q-Q Plot for extreme rainfall data in Bukit Berapit with threshold value of 90^{th} percentile. . . . 34

4.13 Q-Q Plot for extreme rainfall data in Kota Bharu with threshold value of 90^{th} percentile. . . . 34

4.14 Q-Q Plot for extreme rainfall data in Arau with threshold value of 95^{th} percentile. 35
4.15 Q-Q Plot for extreme rainfall data in Pekan Merlimau with threshold value of 95^{th} percentile. 35
4.16 Q-Q Plot for extreme rainfall data in JPS Wilayah Persekutuan with threshold value of 95^{th} percentile. 36
4.17 Q-Q Plot for extreme rainfall data in Ladang Benut Rengam with threshold value of 95^{th} percentile. 36
4.18 Q-Q Plot for extreme rainfall data in Mersing with threshold value of 95^{th} percentile. 37
4.19 Q-Q Plot for extreme rainfall data in Ladang Boh with threshold value of 95^{th} percentile. 37
4.20 Q-Q Plot for extreme rainfall data in Ipoh with threshold value of 95^{th} percentile. 38
4.21 Q-Q Plot for extreme rainfall data in Gua Musang with threshold value of 95^{th} percentile. 38
4.22 Q-Q Plot for extreme rainfall data in Bukit Berapit with threshold value of 95^{th} percentile. 39
4.23 Q-Q Plot for extreme rainfall data in Kota Bharu with threshold value of 95^{th} percentile. 39

List of Tables

4.1 Estimated parameter for threshold value of 90^{th} percentile. 21
4.2 Estimated parameter for threshold value of 95^{th} percentile. 22
4.3 Test statistics for K-S test with threshold value of 90^{th} percentile. . . . 24
4.4 Test statistics for A-D test with threshold value of 90^{th} percentile. . . 25
4.5 Test statistics for Chi-Squared test with threshold value of 90^{th} percentile. 26
4.6 Test statistics for K-S test with threshold value of 95^{th} percentile. . . . 27
4.7 Test statistics for A-D test with threshold value of 95^{th} percentile. . . 28
4.8 Test statistics for Chi-Squared test with threshold value of 95^{th} percentile. 29

CHAPTER 1

INTRODUCTION

1-1 Introduction

Each year, there is always some natural disasters happened around the world, and amongst the natural disaster, flash flood is the most common natural disaster occurs in Malaysia was flood. The reason for Malaysia being more vulnerable to flood are due to the high amount of rainfall generated each year in the country, in with come out to an average amount of around 2000mm to 4000mm rainfall. Throughout the year, most of the rainfalls in Malaysia happened during the monsoon season.

There are two monsoon seasons faced by Malaysia, which are the southwest monsoon from May to August, and the northeast monsoon from November to February. Under normal circumstances, northeast monsoon period that originates from China and the north Pacific will generate higher amount of rainfall as compare to southwest monsoon period that originates from the deserts of Australia.

In Malaysia, floods are the most frequently happen natural disaster, it is also considered as the costliest natural disaster risk. Whenever a flood happens, it cost a high amount of money to repair and reconstruct the affected areas, and the life of residence would also be threatened. Therefore, a suitable model that can be use to estimate the total rainfall amount during the next extreme rainfall event would be beneficial for Malaysia. The model would allow the government to determine the likelihood of a flood happening in the study area, and hence execute some appropriate preventions before the occurrence of the flood.

With the appropriate preventions done before the possible floods hit the country, government would be able to reduce the damage done by the floods, and hence reduce the cost of repair and reconstruct after the occurrence of the catastrophe and the number of injuries and death on the affected areas.

1-2 Problem Statement

Due to the heavy rainfall that occurred every year on the monsoon season, massive floods happened frequently all around Peninsular Malaysia. And one of the reasons for the occurrence of floods are with the increase in global warming problem, and the Southern Pacific that located beside Malaysia, it had made Malaysia more vulnerable in getting floods as compared to the previous years.

According to the online news from New Straits Times (Povera, 2019), as Antarctica melts, sea levels would be rising and it would bring along problems like thunderstorm happening and heavy rain above the sea. This cause major problem to Malaysia as it is a country that lie next to the Southern Pacific, which has been recorded for having high level of sea rise even among all other places around the world. As monsoon season occurs, thunderstorm would be brought along to Peninsular Malaysia, and possibly unleashing floods in few of the low-lying areas of Malaysia.

The costs for dealing with the problems made by floods are high, and with the current world situation, it is expected that in the near future, the extreme rainfall event will occur more and more frequently. If we were not able to make any precaution to deal with these problems, the costs in dealing with the floods will keep rising and it would affect the economy of the country.

To deal with the problems, this research was conducted in order to come out with a model with suitable distributions for estimating the future extreme rainfall data. By having a better estimation model, we would be able to come out with more suitable and reliable precaution in dealing with the extreme rainfall event that are expected to be occurring.

1-3 Objectives of Research

Throughout this research, we will focus on three objectives which are

- To analyze the rainfall characteristics at the study areas.
- To estimate the parameters of model distributions.
- To determine the best-fit distribution for the extreme rainfall data.

1-4 Research Question

To achieve the objective of this research, few problem statements are to be solve which consist of

- What are the rainfall characteristics of the study areas?
- What are the parameters of model distributions?
- What is the best-fit distribution for the extreme rainfall data?

1-5 Project Scope

This research will be focus on 10 rainfall stations in Peninsular Malaysia, from year 1975 to year 2008. All the historical rainfall data were obtained from Department of Irrigation and Drainage Malaysia.

CHAPTER 2

LITERATURE REVIEW

2-1 Introduction

This chapter will review the discussion on extreme rainfall modelling.

2-2 Overview on Extreme Rainfall Modeling

A lot of study regarding extreme cases event had been carried out over the years, which included the extreme rainfall event that lead to possible flood on certain area. In year 2017, Syafrina and Norzaida (2017) tested and compared the performance of Gamma and Weibull in a weather generator model. Advanced Weather Generator (AWE-GEN) model is employed to model the rainfall at hourly scale. The area of interest in the study involved five stations on eastern region of Peninsular Malaysia. Throughout the research, the author concluded that the Gamma distribution provided a better result for the hourly rainfall data. Different result by Kumar et al. (2017) are obtained on the research carried out in all district of Uttarakhand, India. The best fit distribution is applied and comparison are performed based on the goodness-of-fit test, which include Kolmogorov-Smirnov, Anderson-Darling, and Chi Squared, Weibull distribution outperformed other distribution while the Chi Squared (2P) and Log-Pearson are the next best distribution to be used. The usage of Mann–Kendall (MK), modified Mann–Kendall (MMK) and Theil–Sen's slope estimator was presented by Prabhakar et al. (2019), which was applied for trend analysis for rainfall data for a long period of times. The changing point for long-term rainfall time series is investigated by the usage of Standard normal homogeneity test (SNHT) and Mann–Whitney–Pettitt (MWP) test in 30 different districts on the state of Odisha, India. The result showed that there is a decreasing trend of rainfall beyond year 1945. Furthermore, all three of the research above showed that most of the rainfall occur on the monsoon period.

Extreme rainfall events are ranked according to Weibull's method in the study done by Sabarish et al. (2017), while the Chi Square and Kolmogorov-Smirnov test are used to investigate the relevancy of distribution. The study area involved are the Tiruchirappalli City, South India, and the results shows that log-Pearson type III distribution are suitable to be adopted for estimation of rainfall amounts at various probability levels due to the least value of Chi square value for the return period up to 100 years. A daily rainfall disaggregation model was adopted by Paola et al. (2014) to evaluate an IDF curves of rainfall. The IDF curves were obtained by using the probability distribution of Gumbel and shorter duration of rainfall data that are less than 24 hours have been obtained by using two different models of disaggregation in the historical data. Three cities in African are advised in the research and the result shows that the effect of climate change would affect in the increase in frequency of extreme events. Future rainfall intensity is assumed to be subjected to decreases or increases depending on which area are used in the research, but the frequency of rainfall will be kept on increasing.

Ten commonly used probabilities distribution for extreme rainfall were considered in the study made by Nguyen and Nguyen (2016), and a further investigation is made by Nguyen et al. (2017) with the same distribution were tested on the same area. The study area involved are 21 weather stations at Ontario, Canada. The results of four goodness-of-fit tests are relied on the findings, which including root mean square error, relative root mean square error, maximum absolute error, and correlation coefficient. Throughout the studies, Generalized Extreme Value (GEV), Generalized Normal (GNO) and Pearson Type III (PE3) are the best overall distribution that provide the best goodness-of-fit and sturdy quantile extrapolations. Since GEV is having more solid theoretical basis and the inherent scale-invariance property of its non-central moments over different time scales, GEV is more preferred as compared to two other best overall distribution. Kar et al. (2017) used a regional approach based on L-moments to give estimation of hourly rainfall frequency estimation and goodness-of-fit measure. Five rainfall stations in Jeju Island, Korea are set as the study area and the motive of the study are to provide helpful information for water system design, water resources management and hydro-meteorological emergencies. Amongst all the distribution tested, this study shows that Gumbel and GEV distribution would be considered as a more reliable and successful models for the studied area due to its lower root mean square error value. This study showed that the model is suitable and can be implicated in other areas with similar characteristics, limited rainfall data and steep land slope.

With the usage of Program R, Smith (2015) set the time series data with the monthly observations to do modification for seasonality. A weighted least-squares regression was implied to measure the rainfall trends of Barker in Southeast Texas. Data obtained from four stations located around the Houston metropolitan area are used and the result fail to demonstrate with a logical confidence that less frequent, more extreme annual rainfall events would be occurring now than occurred in the past, therefore in the near future, Houston are presumed to be dry without greater increases in their annual rainfall amounts. Mehr et al. (2017) developed and applied a novel classification-forecasting model, namely Binary GP (BGP), for teleconnection studies between sea surface temperature (SST) variations and maximum monthly rainfall (MMR) events. Data are obtained from two rain gauge stations on northwest of Iran, and the sea surface temperature of the surrounding seas are also used. The model are train and test using SST series and the proposed model are claims to be able to capture nonlinear feature of potential teleconnection signals satisfactorily. The few limitation found throughout the studies are the model only suitable for maximum monthly rainfall classification/forecasting and there will be binary classification issues using genetic programming.

In contrast, there is no consensus as to which of the tested distribution is more suitable for the extreme rainfall series. Generally, different areas with unique characteristics for the available rainfall data would affect the choice of appropriate distribution to be used.

CHAPTER 3

RESEARCH METHODOLOGY

3-1 Introduction

This chapter discusses the methodology of the research. We begin with data cleansing. The fitting of the probability distribution is also described in detail. Parameter(s) of the distribution is then estimated. After that, goodness-of-fit test are used to determine which distribution fit better into the extreme rainfall data. Lastly, quantile-quantile plots will be use to visualize the suitability of the distributions and the extreme rainfall data.

3-2 Data Cleansing

Peak-over-threshold (POT) approach is one of the many methods that are used in extreme value analysis and this method is apply by looking at the extreme values from the given data that surpass a certain threshold value. For current study, the rainfall data are cleansed by using peak-over-threshold (POT) approach to obtain a list of extreme rainfall data for fitting into some selected probability distributions.

First, by applying POT approach, all the zero rainfall are withdrawn from the data. Then, rainfall amount that exceed the pre-set threshold will be included into the model. The thresholds used in this study are determined by the 90% quartile and 95% quantile of the data.

3-3 Fitting of Probability Distribution

The extreme rainfall data are determined after applying the peak-over-threshold approach. The extreme rainfall data are fitted to few selected probability distributions. The probability distributions used in this research are Exponential, Generalized Extreme Value (GEV), Gamma, and Weibull.

3-3-1 Exponential Distribution

The probability density function for exponential distribution is as follow:

$$f(x) = \lambda e^{-\lambda x} \tag{3.1}$$

where the variable $x \geq 0$ and the parameter λ represent the rate.

3-3-2 Generalized Extreme Value (GEV) Distribution

The cumulative distribution function for GEV distribution is as follow:

$$F(x; \mu, \sigma, \xi) = e^{-[1+\xi(\frac{x-\mu}{\sigma})]^{\frac{-1}{\xi}}}, \xi \neq 0 \tag{3.2}$$

where the three parameters μ, σ and ξ represents the location, scale and shape of the distribution function respectively.

3-3-3 Gamma Distribution

The probability density function for gamma distribution is as follow:

$$f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \tag{3.3}$$

where the variable x is positive real quantities, and the parameter k and θ represents the shape and scale of the distribution respectively.

3-3-4 Weibull Distribution

The probability density function for Weibull distribution is as follow:

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \tag{3.4}$$

where the variable x and the parameters λ and k that represents scale and shape respectively are all positive real numbers.

3-4 Parameter Estimator

Maximum Likelihood Estimator (MLE) is selected to be used in estimating the parameter(s) of the probability distributions used in this project.

When there is a joint density function with a set of variates with a sample size of n, if the function is viewed as a function for the parameters alone, and not for the provided set of variates, it would be called as a likelihood function. Therefore, if the parameters that are to be estimate is θ, then the likelihood function, $L(\theta)$ could be defined as:

$$L(\theta) = f_\theta(x) \tag{3.5}$$

and when the variates are independent and identically distributed, the likelihood function, $L(\theta)$ could be defined as:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) \tag{3.6}$$

The MLE, $\hat{\theta}$ would be the value of the parameter θ that maximizes the likelihood function, $L(\theta)$.

To find the MLE, the first derivative of the function is set to be zero to solve for the value of $\hat{\theta}$. Next, the second derivative of the function is calculated, if the value for the second derivative is negative, it would confirm that $\hat{\theta}$ maximizes the likelihood function.

MLE was chosen among all the available estimators due to the few properties of the sequence of MLE as the sample size increase to infinity under quite general conditions:

(i) Consistency:
MLE would converge to the true parameter value.

(ii) Functional Invariance:
The MLE of $g(\theta)$, in which represent any transformation of function of θ is equal to that function evaluate at the MLE of θ. Hence, the MLE for a given function $\alpha = g(\theta)$ is $\hat{\alpha} = g(\hat{\theta})$.

(iii) Efficiency:

Since MLE attains the Cramér–Rao lower bound when the sample size increase to infinity, there is no compatible estimator that has lower asymptotic means squared error than MLE, which also conclude that MLE has asymptotic normality.

3-4-1 Exponential Distribution

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a vector of n observations from an exponential distribution with parameter rate = λ.

The MLE of λ is given by:

$$\hat{\lambda} = \frac{1}{\bar{x}} \tag{3.7}$$

where $\bar{x}$ denotes the sample mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.8}$$

in which the MLE represent the reciprocal of the sample mean.

3-4-2 Generalized Extreme Value (GEV) Distribution

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample of n observations from a generalized extreme value distribution with parameters location = μ, scale = σ, and shape = ξ.

The log-likelihood function is given by:

$$L(\mu, \sigma, \xi) = -\mu log(\sigma) - (1 - \xi)\sum_{i=1}^{\mu} y_i - \sum_{i=1}^{\mu} e^{y_i} \tag{3.9}$$

where

$$y_i = -\frac{1}{\xi} log\left[\frac{1 - \xi(x_i - \mu)}{\sigma}\right] \tag{3.10}$$

The MLE's of μ, σ, and ξ are those values that maximize the likelihood function, subject to the following constraints:

$$\sigma > 0 \tag{3.11}$$

$$\xi \leqslant 1 \tag{3.12}$$

$$x_i < \mu + \frac{\sigma}{\xi} \text{ if } \xi > 0 \tag{3.13}$$

$$x_i > \mu + \frac{\sigma}{\xi} \text{ if } \xi < 0 \tag{3.14}$$

A constraint $\xi \leqslant 1$ is imposed due to the likelihood can be made infinite and cause the MLE to not exist when $\xi > 1$.

3-4-3 Gamma Distribution

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample of n observations from a gamma distribution with parameters shape = k and scale = θ. The relationship between these parameters and the mean (μ) and coefficient of variation (τ) of this distribution is given by:

$$k = \tau^{-2} \text{ and } \theta = \frac{\mu}{k} \tag{3.15}$$

and

$$\mu = k\theta \text{ and } \tau = k^{-\frac{1}{2}} \tag{3.16}$$

The MLE's of k and θ are the solutions of the simultaneous equations:

$$\hat{k} = \frac{1}{n}\sum_{i=1}^{n} log(x_i) - log(\bar{x}) = \psi(\hat{k}) - log(\hat{k}) \tag{3.17}$$

$$\hat{\theta} = \frac{\bar{x}}{\hat{k}} \tag{3.18}$$

where ψ denotes the digamma function, and $\bar{x}$ denotes the sample mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.19}$$

3-4-4 Weibull Distribution

Let $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample of n observations from a Weibull distribution with parameters scale = λ and shape = k.

The MLE's of λ and k are the solutions of the simultaneous equations:

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{k}}\right]^{\frac{1}{\hat{k}}} \tag{3.20}$$

$$\hat{k} = \frac{n}{\left\{\left(\frac{1}{\hat{\lambda}}\right)^{\hat{k}} \sum_{i=1}^{n}\left[x_i^{\hat{k}} log(x_i)\right]\right\} - \sum_{i=1}^{n} log(x_i)} \tag{3.21}$$

3-5 Goodness-of-fit Test

The goodness-of-fit test are used to obtain the best probability distribution for the extreme rainfall in this study. The suitability of the selected probability distributions and the extreme rainfall data.

Amongst the available goodness-of-fit test, Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared Test are used for the goodness of fit tests, with the significant level of 5%. The techniques for determine the best-fit distribution with the minimum error produced are:

3-5-1 Kolmogorov-Smirnov Test (K-S Test)

The K-S Test compares the empirical distribution function ($F_n(x)$) with a specified cumulative distribution function ($F(x)$). The equation for computing the Kolmogorov-Smirnov statistic (D) is:

$$D_n = max|F_n(x) - F(x)| \tag{3.22}$$

where the equation is used to compute the distance between the two functions, $F_n(x)$ and $F(x)$. The larger the value of the test statistics, the higher the inconsistency between the observed data.

3-5-2 Anderson–Darling Test (A–D Test)

The A-D Test is the modified version of the K-S Test that put higher weight on the tails of the tested distributions. This would provide a better result when the tail of the tested distributions is on a higher significancy. The equation for the test statistics (AD) is:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)(\ln(F(x_i)) + \ln(1 - F(x_{(n+1-i)}))) \tag{3.23}$$

where $x_{(1)}$ to $x_{(n)}$ is the ordered sample of size n from smallest to largest, and $F(x)$ is the cumulative distribution function for the specified distribution. The null hypothesis would be rejected when the AD is larger than the critical value of AD_α with the given significant level of α.

3-5-3 Chi-Squared Test

The Chi-squared test is used to check the suitability of a specific distribution by observing the frequency in a sample. By using O as the "observed count" and E as the "expected count", the equation to calculate Chi-squared is:

$$\chi^2 = \frac{\sum(O-E)^2}{E} \tag{3.24}$$

The null hypothesis for the test states that there is insignificant evidence to conclude that there is dissimilarity in the observed frequencies and the expected frequencies, while the alternative hypothesis states that the frequencies are dissimilar.

3-6 Quantile-quantile Plots

The quantile-quantile plot (Q-Q plot) is a graphical tool that helps in assess the plausibility of a set of data to a theoretical distribution. Q-Q plot is a scatterplot that use two sets of quantiles against one another and if both sets are come from the same distribution, the points form by the scatterplot would be forming a line that is roughly straight.

Q-Q plots use the data from the sample and sort it in an ascending order. Then a scatterplot is plotted by using the sample data versus the quantiles obtained from the theoretical distribution. The quantity of quantiles would be selected to complement the size of the sample data.

Q-Q plots is endorsed to envision the suitability of specific distribution to the available rainfall data. Thus, Q-Q plot is useful to check the fitness of the extreme rainfall data with the certain probability distribution.

CHAPTER 4

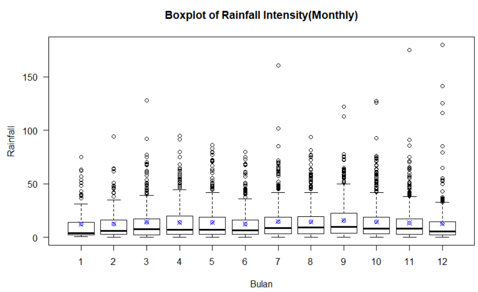
RESULTS AND DISCUSSION

4-1 Introduction

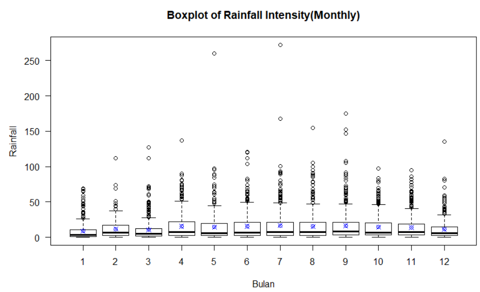
This chapter presents the results from the extreme rainfall modeling of this study. First starts by data cleansing. Then the parameter(s) for the distributions that would be use in fitting the data are estimated. Goodness-of-fit test are used to determine the best-fit distribution for the data. Lastly, Q-Q plots is adopted to visualised the suitability of the data with the selected distribution.

4-2 Data Cleansing

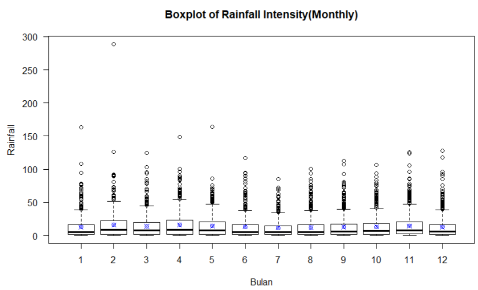
All the zero rainfall are withdrawn from the data. After withdrawing all of the zero rainfall from the data, boxplots are plotted to analyze the trend of monthly rainfall for 10 areas from year 1975 to year 2008. Figure 4.1 shows the boxplot for the 10 rainfall stations in Peninsular Malaysia:



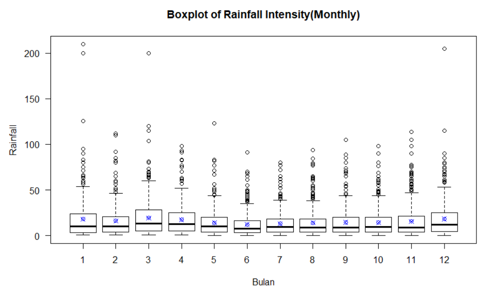
(a) Arau



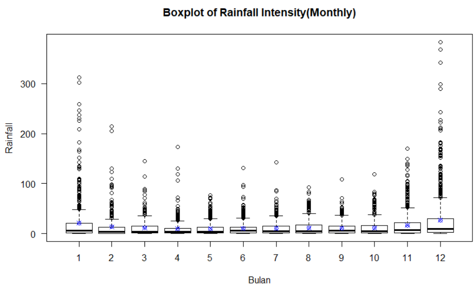
(b) Pekan Merlimau



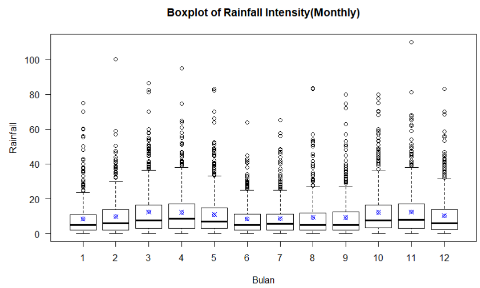
(c) JPS Wilayah Persekutuan



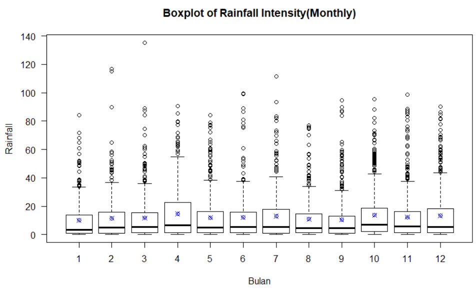
(d) Ladang Benut Rengam



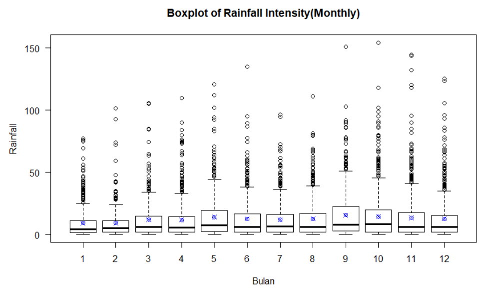
(e) Mersing



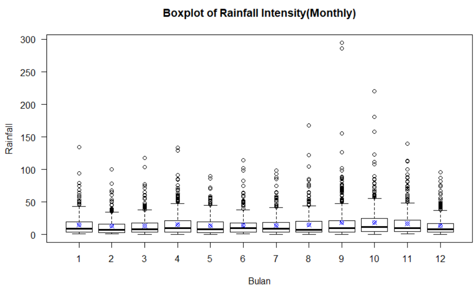
(f) Ladang Boh



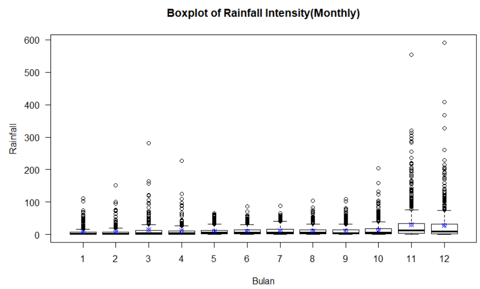
(g) Ipoh



(h) Gua Musang



(i) Bukit Berapit



(j) Kota Bharu

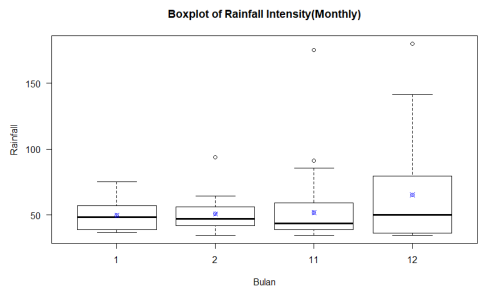
Figure 4.1: Boxplot of Monthly Rainfall Intensity for selected areas in Peninsular Malaysia

From Figure 4.1, we can see that all of the boxplots show a positive skewed trend and there exist lots of data above the maximum of the boxplot. To deal with the situation, peak-over-threshold (POT) approach are used to focus on the data that are critical for current research.

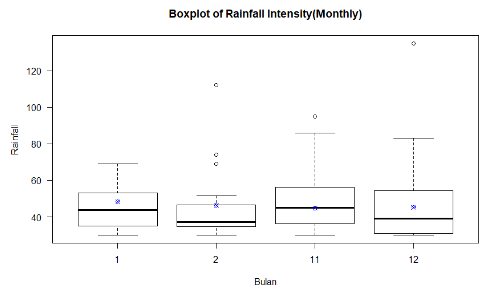
Throughout the research, only northeast monsoon is taken into consideration in getting the best-fit distribution for modelling for the extreme rainfall data. Therefore, the rainfall data from November to February are considered into the study.

The POT value that will be used are 90^{th} percentile and 95^{th} percentile of the data excluding all the zero rainfall. Two thresholds are used to further analyze whether there is any different in result for taking consideration of more data and less data for the selected areas.

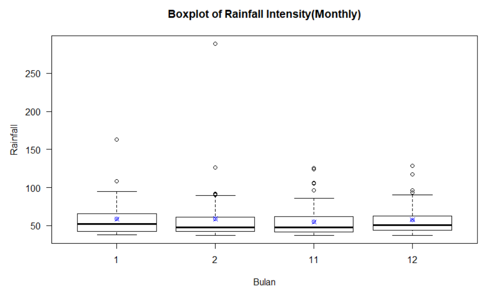
For 90^{th} percentile as the POT value, the boxplot for the rainfall data for November to February are shown in Figure 4.2:



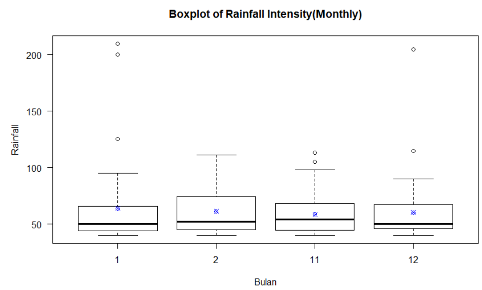
(a) Arau



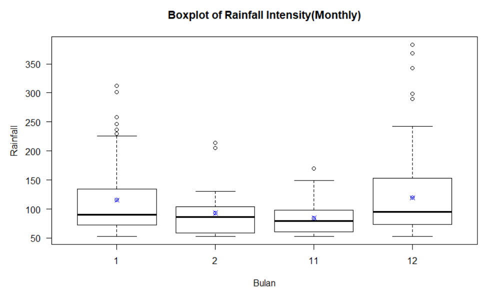
(b) Pekan Merlimau



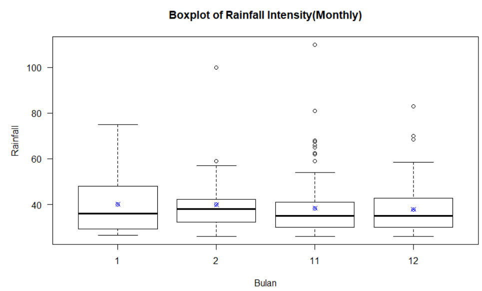
(c) JPS Wilayah Persekutuan



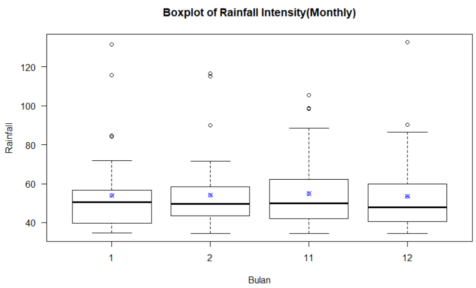
(d) Ladang Benut Rengam



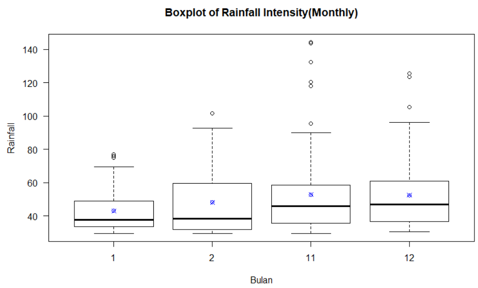
(e) Mersing



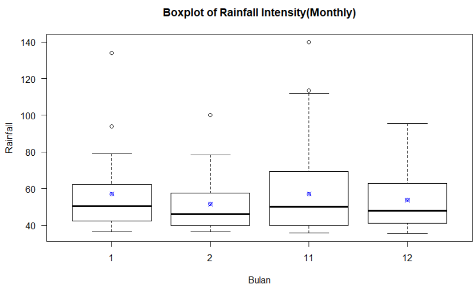
(f) Ladang Boh



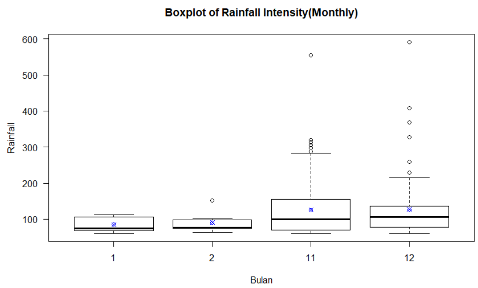
(g) Ipoh



(h) Gua Musang



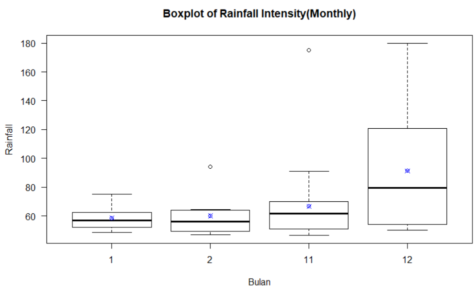
(i) Bukit Berapit



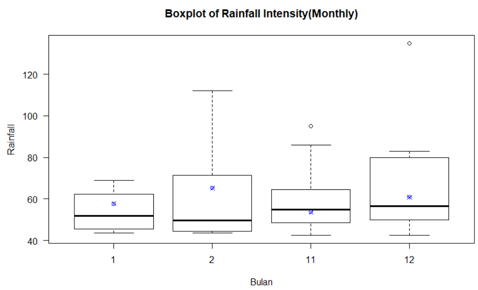
(j) Kota Bharu

Figure 4.2: Boxplot of Rainfall Intensity for November to February of selected areas in Peninsular Malaysia

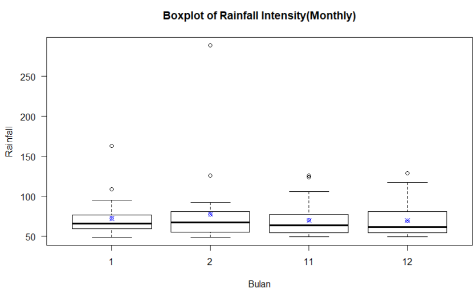
For 95^{th} percentile as the POT value, the boxplot for the rainfall data for November to February are shown in Figure 4.3:



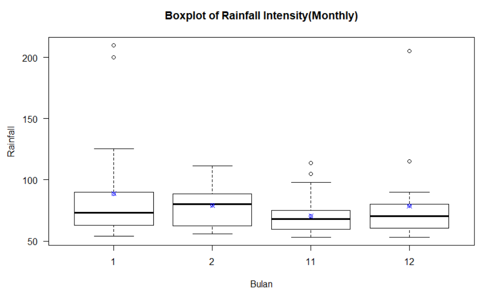
(a) Arau



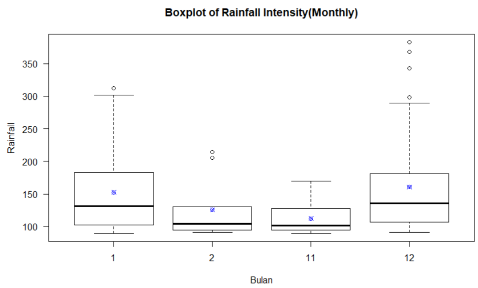
(b) Pekan Merlimau



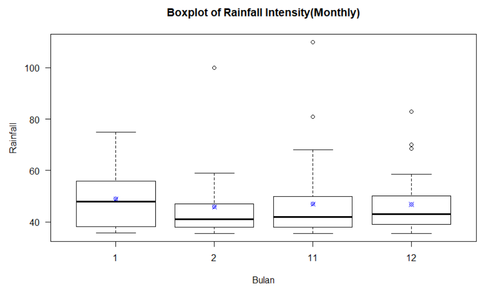
(c) JPS Wilayah Persekutuan



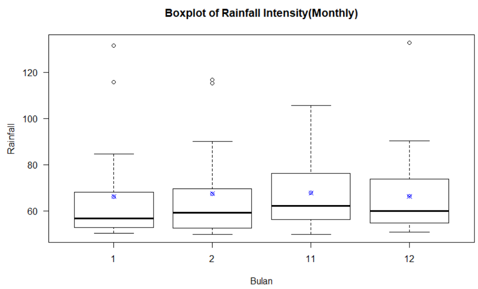
(d) Ladang Benut Rengam



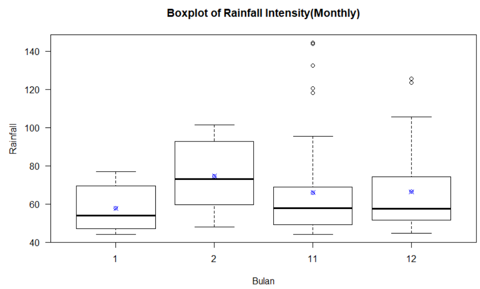
(e) Mersing



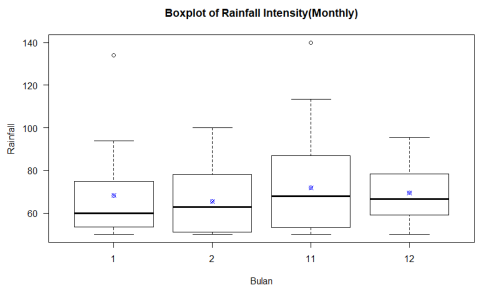
(f) Ladang Boh



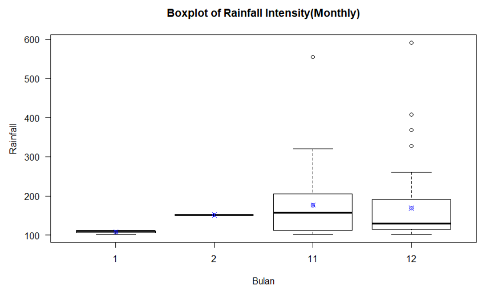
(g) Ipoh



(h) Gua Musang



(i) Bukit Berapit



(j) Kota Bharu

Figure 4.3: Boxplot of Rainfall Intensity for November to February of selected areas in Peninsular Malaysia

The Figure 4.2 and Figure 4.3 show both thresholds are suitable to be used in modelling for extreme rainfall cases. By comparing the trend of extreme rainfall data as shown in the boxplot, it could be seen that the trend shows similar pattern for all 10 stations for both thresholds which is most of the extreme rainfall data having positive skewed trend. Since both the thresholds would probably lead to the same results due to the similarity in trend from the boxplots, the 95^{th} percentile are preferred as it is more time efficient to generate results from lesser data.

4-3 Parameter Estimator

The parameter(s) for each of the distribution shown in the previous section are estimated for both data with threshold value of 90^{th} percentile and 95^{th} percentile in the selected areas by using maximum likelihood estimator (MLE). After applying the MLE, the estimated parameter(s) for each of the distribution are as shown in the table below:

Threshold Value = 90^{th} Percentile	Estimated Parameter(s)	
Area Name	Exponential Distribution	Generalized Extreme Value Distribution
Arau	$\lambda = 0.018439$	$\mu = 40.930375, \sigma = 7.532072, \xi = -0.744269$
Pekan Merlimau	$\lambda = 0.021381$	$\mu = 37.383810, \sigma = 8.180785, \xi = -0.469732$
JPS Wilayah Persekutuan	$\lambda = 0.017493$	$\mu = 44.763463, \sigma = 8.514707, \xi = -0.607598$
Ladang Benut Rengam	$\lambda = 0.016446$	$\mu = 47.661947, \sigma = 8.684274, \xi = -0.625206$
Mersing	$\lambda = 0.009248$	$\mu = 75.960770, \sigma = 25.187830, \xi = -0.525310$
Ladang Boh	$\lambda = 0.025871$	$\mu = 32.172431, \sigma = 6.156170, \xi = -0.384057$
Ipoh	$\lambda = 0.018443$	$\mu = 44.762656, \sigma = 9.880943, \xi = -0.324186$
Gua Musang	$\lambda = 0.019780$	$\mu = 38.448843, \sigma = 9.302372, \xi = -0.553023$
Bukit Berapit	$\lambda = 0.017975$	$\mu = 43.942123, \sigma = 8.937590, \xi = -0.573548$
Kota Bharu	$\lambda = 0.008088$	$\mu = 81.994504, \sigma = 26.346240, \xi = -0.706841$

Threshold Value = 90^{th} Percentile	Estimated Parameter(s)	
Area Name	Gamma Distribution	Weibull Distribution
Arau	$k = 6.698420, \theta = 8.096231$	$k = 2.137523, \lambda = 61.376439$
Pekan Merlimau	$k = 9.573792, \theta = 4.885238$	$k = 2.746517, \lambda = 52.400516$
JPS Wilayah Persekutuan	$k = 8.118586, \theta = 7.041046$	$k = 2.207335, \lambda = 64.272211$
Ladang Benut Rengam	$k = 8.122050, \theta = 7.486347$	$k = 2.274670, \lambda = 68.503700$
Mersing	$k = 4.511983, \theta = 23.965971$	$k = 1.945663, \lambda = 122.803088$
Ladang Boh	$k = 12.155040, \theta = 3.179993$	$k = 2.950212, \lambda = 43.044884$
Ipoh	$k = 11.016434, \theta = 4.921835$	$k = 2.972338, \lambda = 60.523314$
Gua Musang	$k = 7.030617, \theta = 7.190917$	$k = 2.368762, \lambda = 57.123778$
Bukit Berapit	$k = 9.365283, \theta = 5.940380$	$k = 2.769707, \lambda = 62.415913$
Kota Bharu	$k = 3.922303, \theta = 31.523720$	$k = 1.760565, \lambda = 140.128654$

Table 4.1: Estimated parameter for threshold value of 90^{th} percentile.

Threshold Value = 95^{th} Percentile	Estimated Parameter(s)	
Area Name	Exponential Distribution	Generalized Extreme Value Distribution
Arau	$\lambda = 0.014254$	$\mu = 54.291925, \sigma = 9.036329, \xi = -0.738278$
Pekan Merlimau	$\lambda = 0.017004$	$\mu = 50.112093, \sigma = 8.048719, \xi = -0.413445$
JPS Wilayah Persekutuan	$\lambda = 0.013866$	$\mu = 58.789960, \sigma = 10.279920, \xi = -0.496930$
Ladang Benut Rengam	$\lambda = 0.012805$	$\mu = 64.075697, \sigma = 10.835179, \xi = -0.471028$
Mersing	$\lambda = 0.006810$	$\mu = 107.601563, \sigma = 22.834922, \xi = -0.833175$
Ladang Boh	$\lambda = 0.021215$	$\mu = 40.020977, \sigma = 4.943260, \xi = -0.628783$
Ipoh	$\lambda = 0.014897$	$\mu = 57.293432, \sigma = 7.766114, \xi = -0.536825$
Gua Musang	$\lambda = 0.015335$	$\mu = 52.904782, \sigma = 9.540268, \xi = -0.539583$
Bukit Berapit	$\lambda = 0.014271$	$\mu = 58.120913, \sigma = 10.195953, \xi = -0.517098$
Kota Bharu	$\lambda = 0.005824$	$\mu = 123.074747, \sigma = 27.001167, \xi = -0.836290$

Threshold Value = 95^{th} Percentile	Estimated Parameter(s)	
Area Name	Gamma Distribution	Weibull Distribution
Arau	$k = 7.793649, \theta = 9.001875$	$k = 2.367737, \lambda = 79.236446$
Pekan Merlimau	$k = 16.005415, \theta = 3.674432$	$k = 3.357202, \lambda = 64.953610$
JPS Wilayah Persekutuan	$k = 9.899709, \theta = 7.285099$	$k = 2.354248, \lambda = 80.810852$
Ladang Benut Rengam	$k = 9.962612, \theta = 7.838902$	$k = 2.497066, \lambda = 87.689650$
Mersing	$k = 6.885273, \theta = 21.327076$	$k = 2.417365, \lambda = 166.005290$
Ladang Boh	$k = 17.059947, \theta = 2.762971$	$k = 3.396004, \lambda = 52.002488$
Ipoh	$k = 17.812432, \theta = 3.768701$	$k = 3.661288, \lambda = 73.918180$
Gua Musang	$k = 10.80851, \theta = 6.03319$	$k = 2.894411, \lambda = 72.942719$
Bukit Berapit	$k = 14.213285, \theta = 4.929905$	$k = 3.457429, \lambda = 77.581044$
Kota Bharu	$k = 5.835814, \theta = 29.424126$	$k = 2.102253, \lambda = 194.569794$

Table 4.2: Estimated parameter for threshold value of 95^{th} percentile.

Table 4.1 and Table 4.2 show all of the estimated parameter(s) that are to be used in the future calculation for the selected probability distributions for both thresholds.

4-4 Goodness-of-fit Test

After all of the estimated parameter(s) for the selected distribution are obtained from using MLE, goodness-of-fit test are used to determine the best fit distribution for the selected extreme rainfall data. The goodness-of-fit test that are chosen to be used are Kolmogorov-Smirnov test (K-S test), Anderson-Darling Test (A-D test), and Chi-Squared test.

For K-S test, the test statistics D_n is found. D_n represent the absolute maximum distance between the true cumulative distribution function and the tested distribution. Therefore, the larger the value of D_n, the higher the inconsistency between the observed data and vice versa.

For A-D test, the test statistics AD is found. AD represent the area between the true cumulative distribution function and the tested distribution. Therefore, the larger the value of AD, the higher the inconsistency between the observed data and vice versa.

For Chi-Squared test, the test statistics χ^2 is found. χ^2 is based on the dissimilarity between the real data obtained from the station and the expectation obtained when there was absolutely no relationship connecting the variables. Therefore, a low value for χ^2 means there is a high correlation between the observed data and vice versa.

In contrast, the lower the value for all of the selected test for a selected distribution, the better the selected distribution are to be used in fitting the extreme rainfall data into a modal.

Below shows the table for the test statistics of all the selected goodness-of-fit tests for exponential distribution (Exp), Generalized Extreme Value distribution (GEV), Gamma distribution (Gamma) and Weibull distribution (Wei):

Threshold Value = 90^{th} Percentile	Test Statistics (D_n)			
Area Name	Exp	GEV	Gamma	Wei
Arau	0.470677	0.071557	0.169765	0.253154
Pekan Merlimau	0.473463	0.076519	0.121446	0.194385
JPS Wilayah Persekutuan	0.476526	0.060610	0.150429	0.255879
Ladang Benut Rengam	0.482034	0.081088	0.169482	0.254808
Mersing	0.386321	0.054495	0.144165	0.175962
Ladang Boh	0.489647	0.069824	0.124019	0.202260
Ipoh	0.469768	0.050248	0.096832	0.170147
Gua Musang	0.442060	0.067608	0.119040	0.188617
Bukit Berapit	0.471708	0.084471	0.160547	0.189031
Kota Bharu	0.384462	0.085352	0.135799	0.201181

Table 4.3: Test statistics for K-S test with threshold value of 90^{th} percentile.

Threshold Value = 90^{th} Percentile	Test Statistics (AD)			
Area Name	Exp	GEV	Gamma	Wei
Arau	5.220377	0.566878	5.232899	7.294910
Pekan Merlimau	3.439769	1.263812	3.450424	5.527289
JPS Wilayah Persekutuan	8.541317	1.126472	8.563472	12.857350
Ladang Benut Rengam	7.027613	1.108373	7.043768	10.391750
Mersing	6.619743	1.340040	6.650962	9.771360
Ladang Boh	6.157516	1.067121	6.177817	11.124070
Ipoh	4.270140	0.970005	4.287341	8.237348
Gua Musang	6.006791	1.258900	6.027548	9.319557
Bukit Berapit	4.430056	1.172779	4.441655	6.467238
Kota Bharu	6.500320	2.312418	6.527284	8.940657

Table 4.4: Test statistics for A-D test with threshold value of 90^{th} percentile.

Threshold Value = 90^{th} Percentile	Test Statistics (χ^2)			
Area Name	Exp	GEV	Gamma	Wei
Arau	235.500000	9.354167	74.625000	94.937500
Pekan Merlimau	309.836700	18.408160	41.877550	68.408160
JPS Wilayah Persekutuan	425.833300	21.666670	88.500000	170.500000
Ladang Benut Rengam	419.589700	29.435900	90.153850	173.846200
Mersing	286.741400	28.534480	118.379300	147.862100
Ladang Boh	533.428600	11.277310	111.260500	142.571400
Ipoh	412.973100	26.820630	56.363230	120.452900
Gua Musang	306.741500	18.653660	104.897600	136.243900
Bukit Berapit	300.444400	29.333330	91.555560	109.111100
Kota Bharu	212.333300	38.333330	112.000000	131.000000

Table 4.5: Test statistics for Chi-Squared test with threshold value of 90^{th} percentile.

Threshold Value = 95^{th} Percentile	Test Statistics (D_n)			
Area Name	Exp	GEV	Gamma	Wei
Arau	0.484591	0.096843	0.208430	0.246553
Pekan Merlimau	0.514539	0.078918	0.125636	0.213957
JPS Wilayah Persekutuan	0.489562	0.047015	0.146238	0.259630
Ladang Benut Rengam	0.492700	0.074288	0.184557	0.247550
Mersing	0.455263	0.098376	0.141037	0.199718
Ladang Boh	0.529113	0.076247	0.150643	0.239288
Ipoh	0.523767	0.076246	0.164369	0.209832
Gua Musang	0.490714	0.071615	0.147658	0.206675
Bukit Berapit	0.510108	0.108498	0.132700	0.196649
Kota Bharu	0.447249	0.095979	0.154311	0.226012

Table 4.6: Test statistics for K-S test with threshold value of 95^{th} percentile.

Threshold Value = 95^{th} Percentile	Test Statistics (AD)			
Area Name	Exp	GEV	Gamma	Wei
Arau	3.082355	0.304698	3.088934	4.184783
Pekan Merlimau	1.962797	0.554088	1.967830	3.502498
JPS Wilayah Persekutuan	4.446193	0.339081	4.457478	7.227890
Ladang Benut Rengam	4.229512	0.259279	4.239559	6.499587
Mersing	4.267576	1.710341	4.277907	5.768144
Ladang Boh	5.251070	0.777562	5.259986	7.794504
Ipoh	4.216441	0.674641	4.224041	6.463023
Gua Musang	3.984001	0.605622	3.994395	6.220754
Bukit Berapit	1.586243	1.309774	1.589369	2.499491
Kota Bharu	3.983300	1.270503	3.993577	5.431606

Table 4.7: Test statistics for A-D test with threshold value of 95^{th} percentile.

Threshold Value = 95^{th} Percentile	Test Statistics (χ^2)			
Area Name	Exp	GEV	Gamma	Wei
Arau	84.702130	5.553191	25.553190	40.872340
Pekan Merlimau	195.513500	12.270270	21.351350	35.621620
JPS Wilayah Persekutuan	272.333300	7.629630	60.777780	97.592590
Ladang Benut Rengam	182.864900	4.810811	36.270270	69.351350
Mersing	198.299100	24.076920	92.521370	81.991450
Ladang Boh	353.588200	19.705880	80.176470	94.058820
Ipoh	307.108100	13.990990	54.603600	104.297300
Gua Musang	218.156900	14.490200	42.529410	65.725400
Bukit Berapit	127.552200	14.925370	23.791040	45.791040
Kota Bharu	172.274500	20.607840	88.156860	84.333330

Table 4.8: Test statistics for Chi-Squared test with threshold value of 95^{th} percentile.

According to Table 4.3 to Table 4.8, its show GEV distribution is the best overall result from all 3 of the goodness-of-fit tests for both extreme rainfall data with 90^{th} percentile and 95^{th} percentile. Therefore, it led to a conclusion that GEV distribution is the best fit distribution for the given extreme rainfall data in the 10 selected areas.

4-5 Quantile-Quantile Plots

The quantile-quantile (Q-Q) plots are adopted into the extreme rainfall data to further visualize the suitability of the selected distribution. The following figures show the Q-Q plots of the selected distribution for 10 rainfall stations in Malaysia:

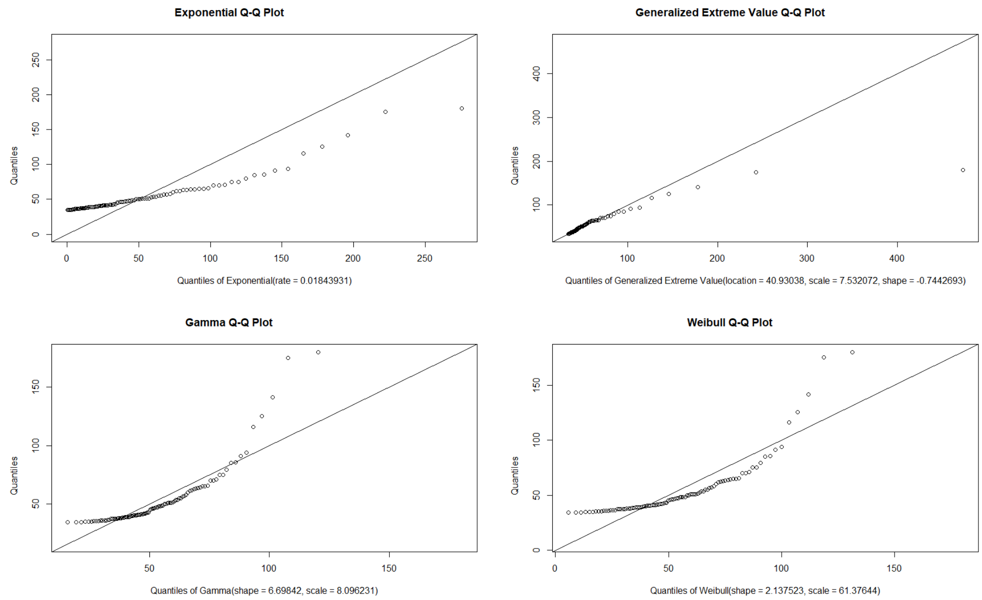


Figure 4.4: Q-Q Plot for extreme rainfall data in Arau with threshold value of 90^{th} percentile.

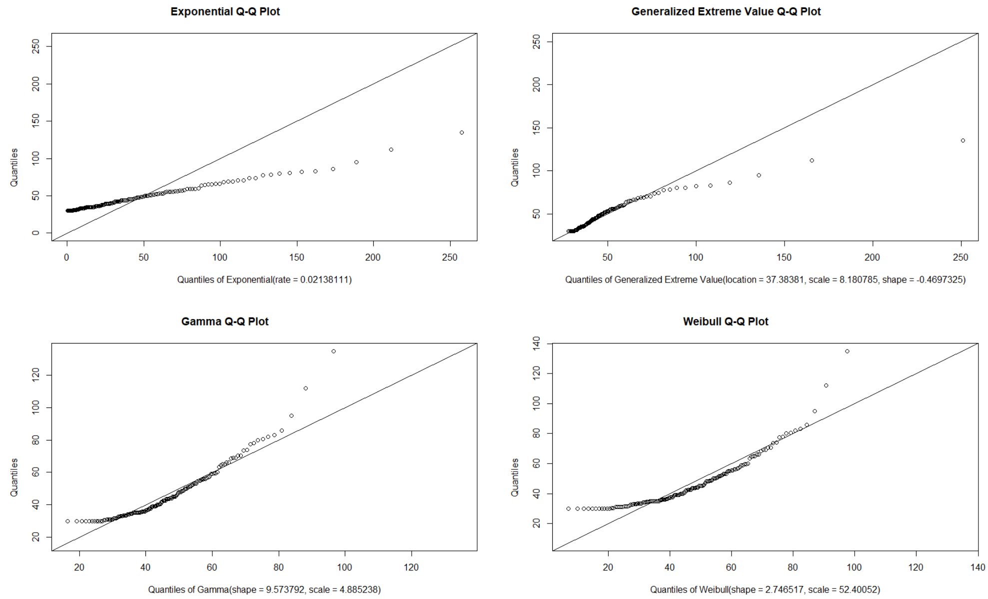


Figure 4.5: Q-Q Plot for extreme rainfall data in Pekan Merlimau with threshold value of 90^{th} percentile.

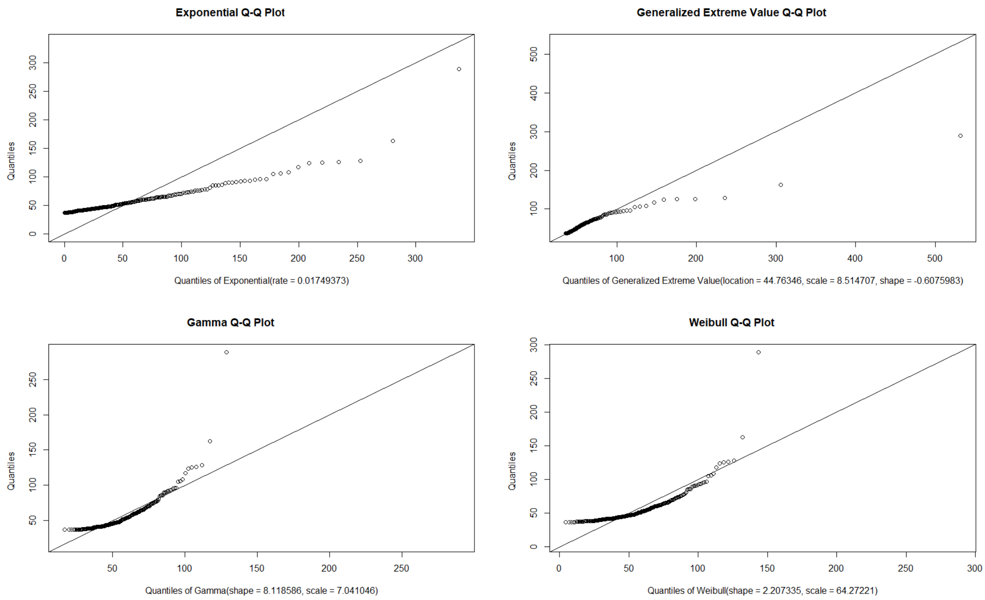


Figure 4.6: Q-Q Plot for extreme rainfall data in JPS Wilayah Persekutuan with threshold value of 90^{th} percentile.

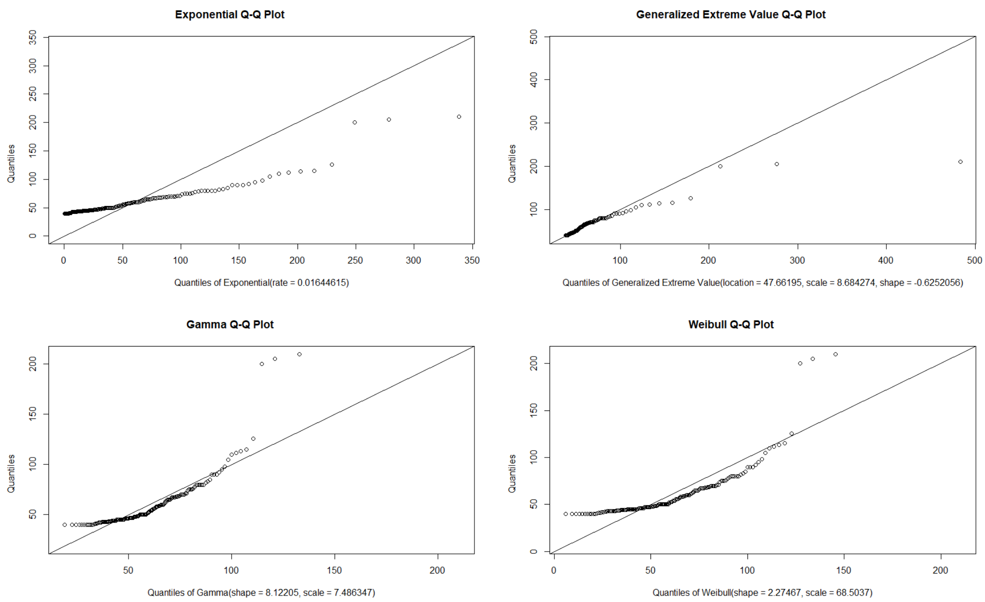


Figure 4.7: Q-Q Plot for extreme rainfall data in Ladang Benut Rengam with threshold value of 90^{th} percentile.

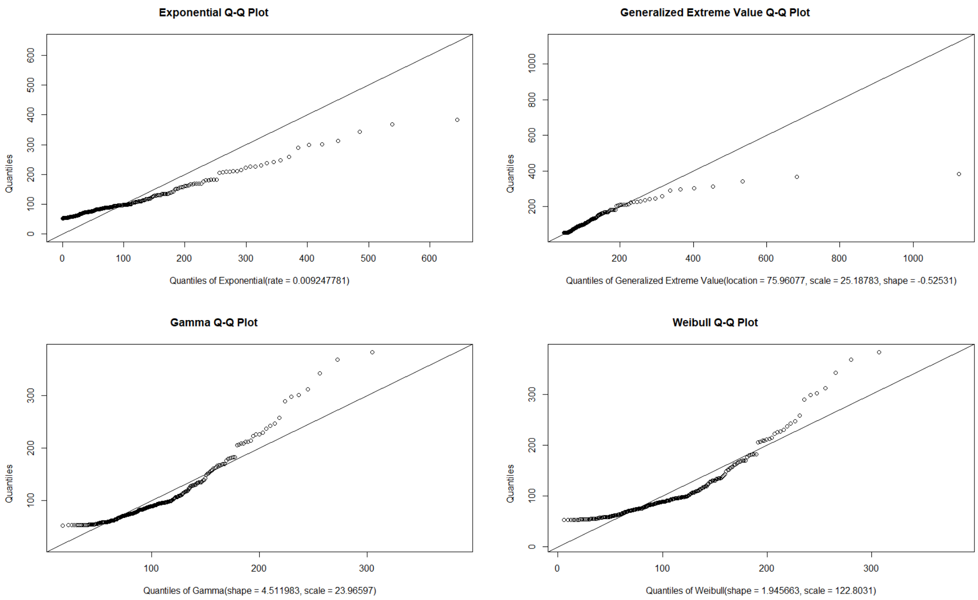


Figure 4.8: Q-Q Plot for extreme rainfall data in Mersing with threshold value of 90^{th} percentile.

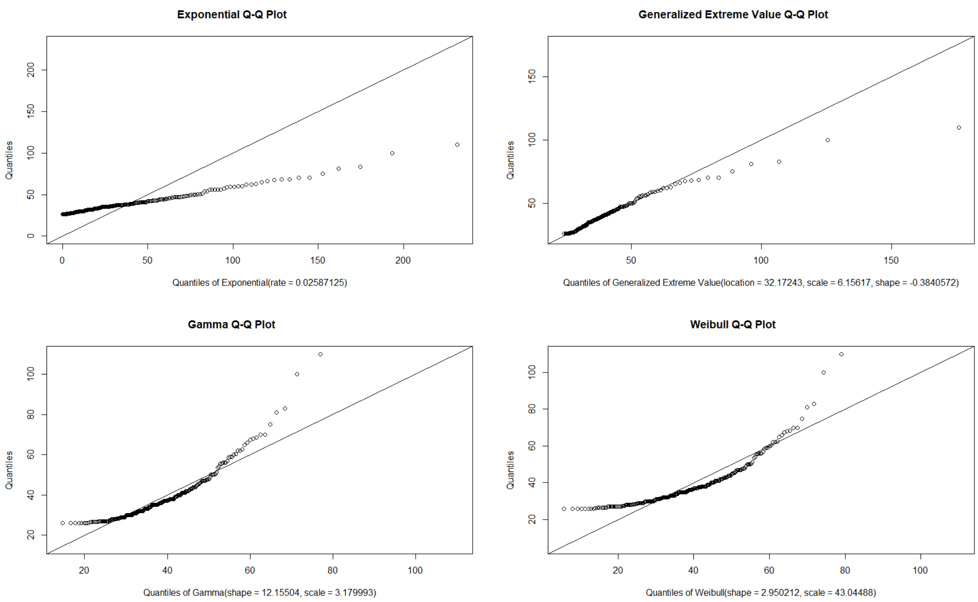


Figure 4.9: Q-Q Plot for extreme rainfall data in Ladang Boh with threshold value of 90^{th} percentile.

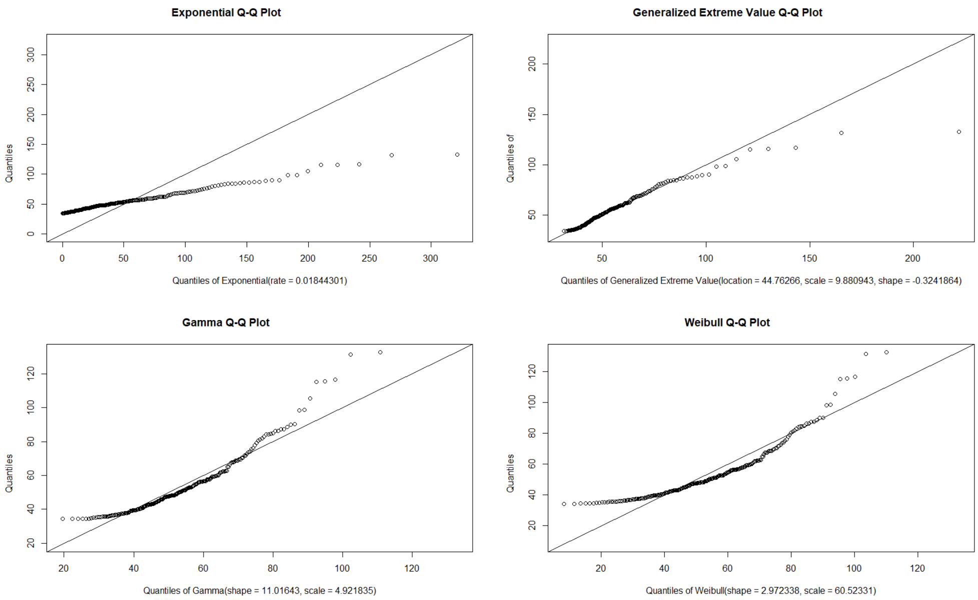


Figure 4.10: Q-Q Plot for extreme rainfall data in Ipoh with threshold value of 90^{th} percentile.

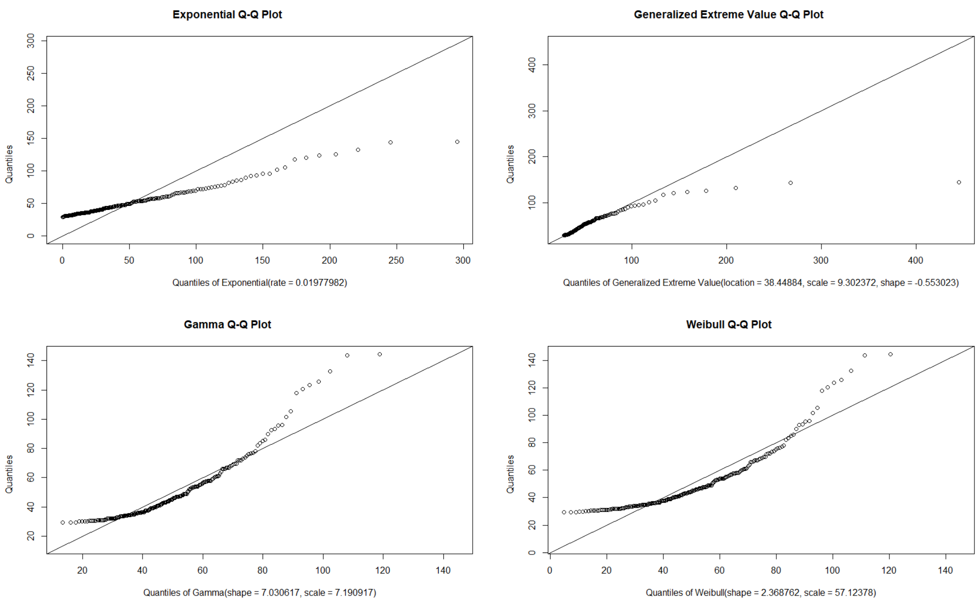


Figure 4.11: Q-Q Plot for extreme rainfall data in Gua Musang with threshold value of 90^{th} percentile.

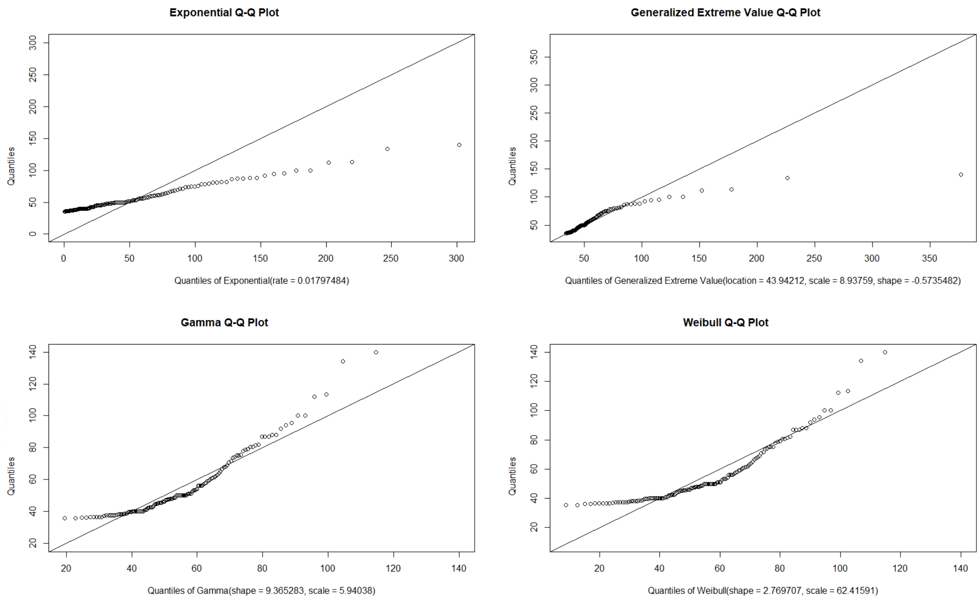


Figure 4.12: Q-Q Plot for extreme rainfall data in Bukit Berapit with threshold value of 90^{th} percentile.

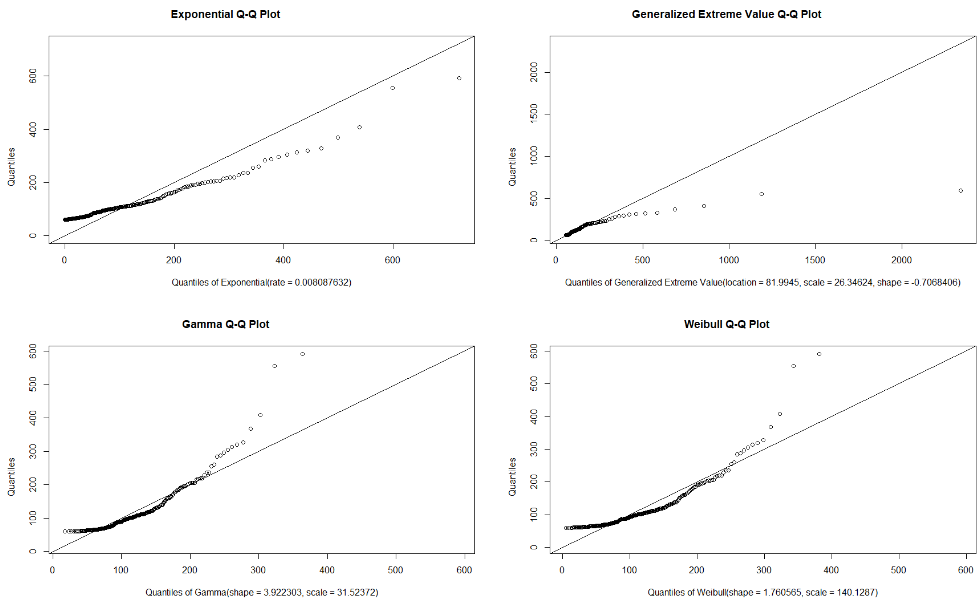


Figure 4.13: Q-Q Plot for extreme rainfall data in Kota Bharu with threshold value of 90^{th} percentile.

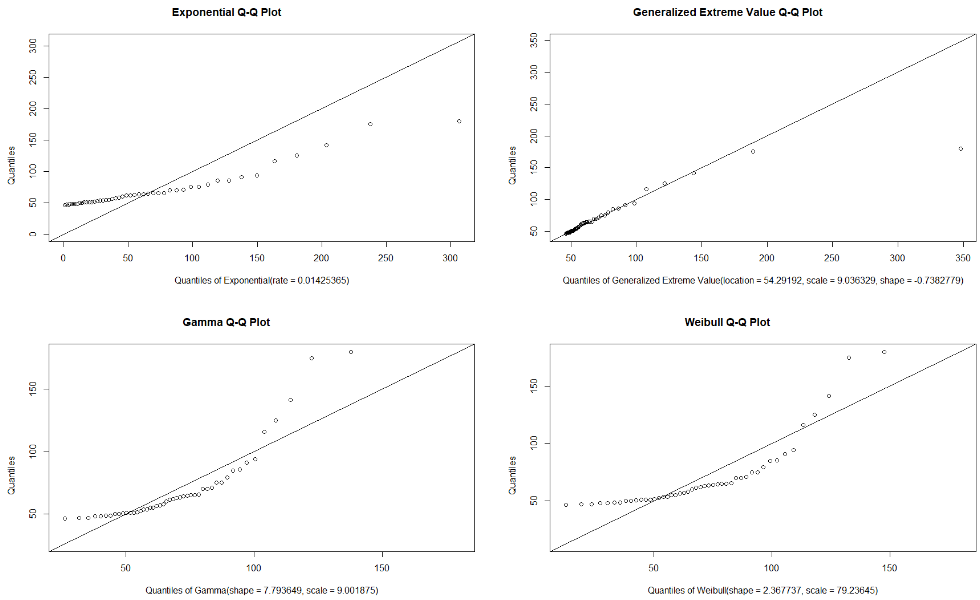


Figure 4.14: Q-Q Plot for extreme rainfall data in Arau with threshold value of 95^{th} percentile.

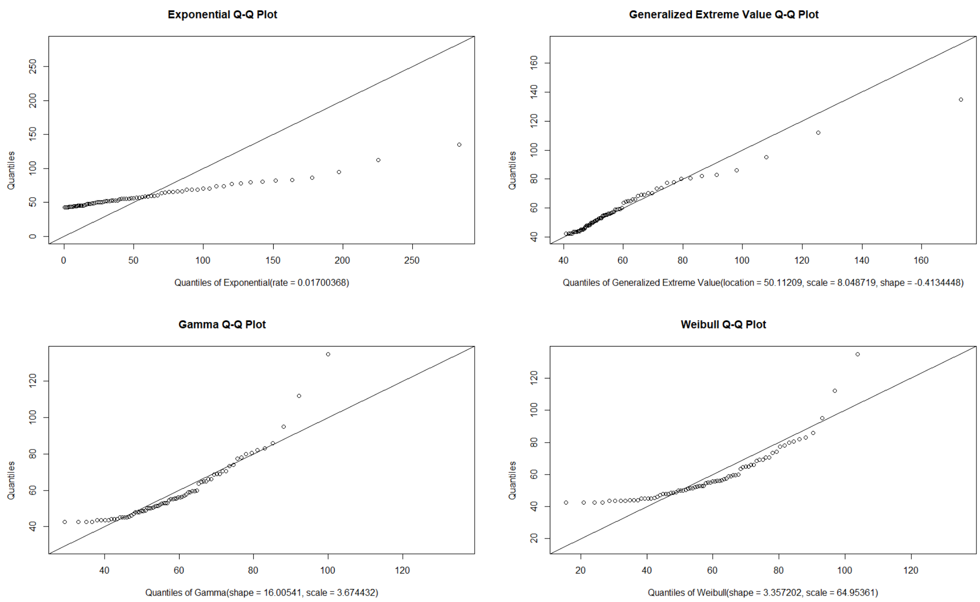


Figure 4.15: Q-Q Plot for extreme rainfall data in Pekan Merlimau with threshold value of 95^{th} percentile.

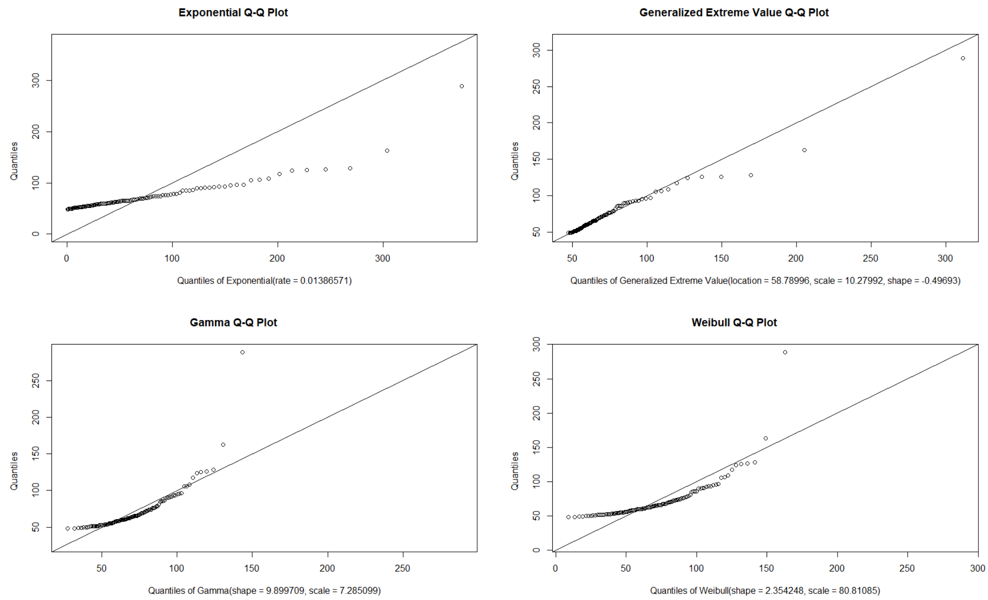


Figure 4.16: Q-Q Plot for extreme rainfall data in JPS Wilayah Persekutuan with threshold value of 95^{th} percentile.

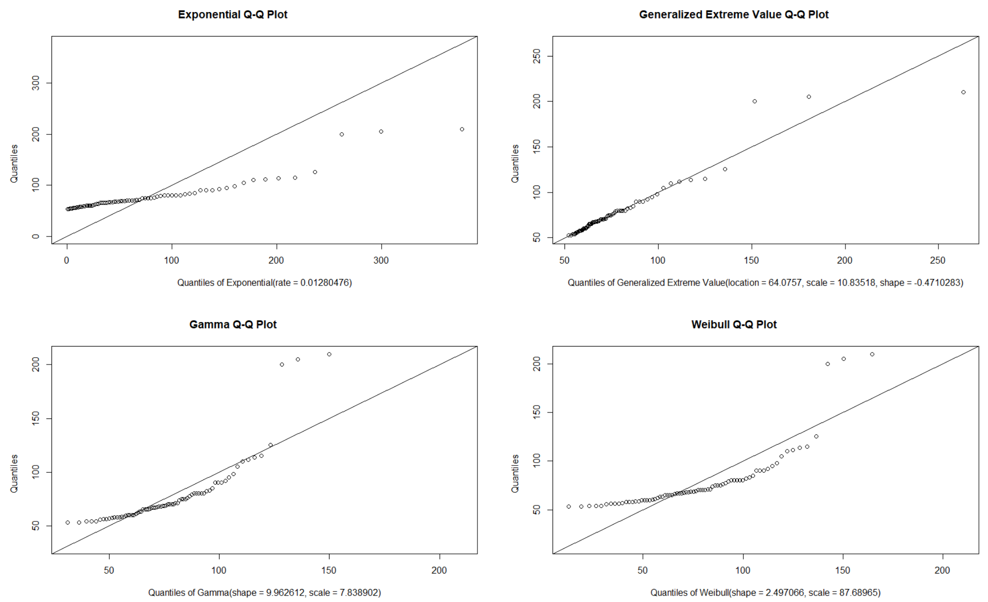


Figure 4.17: Q-Q Plot for extreme rainfall data in Ladang Benut Rengam with threshold value of 95^{th} percentile.

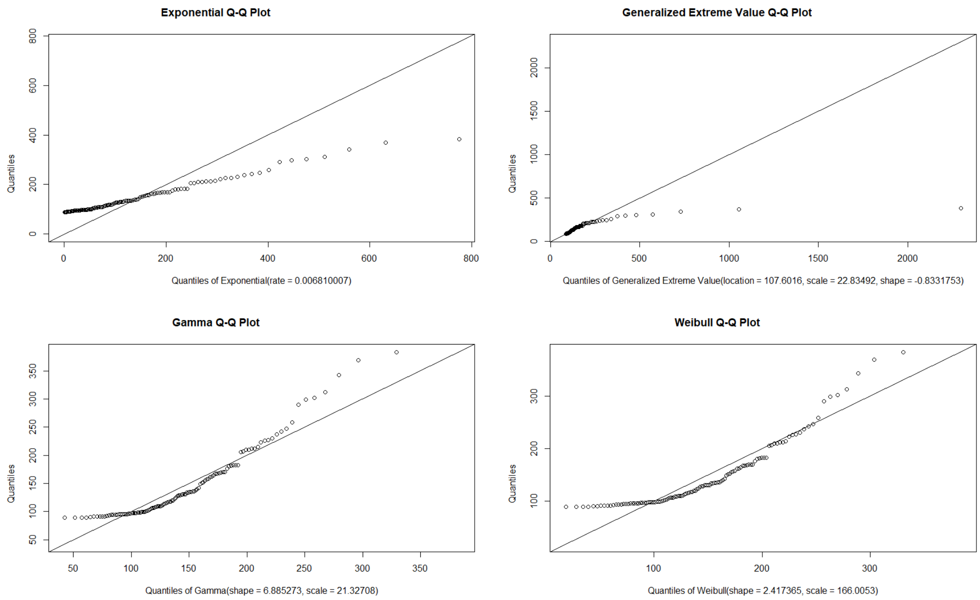


Figure 4.18: Q-Q Plot for extreme rainfall data in Mersing with threshold value of 95^{th} percentile.

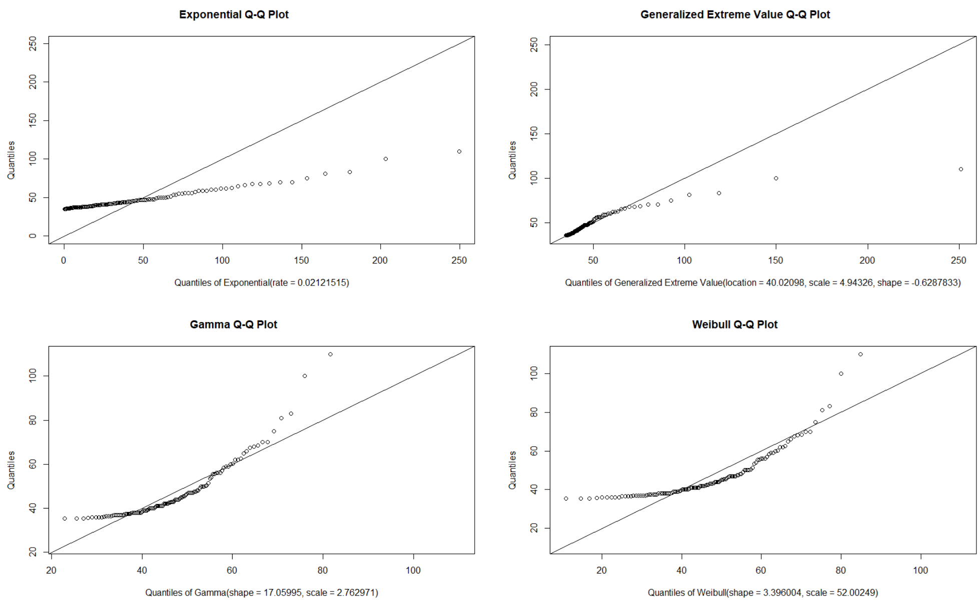


Figure 4.19: Q-Q Plot for extreme rainfall data in Ladang Boh with threshold value of 95^{th} percentile.

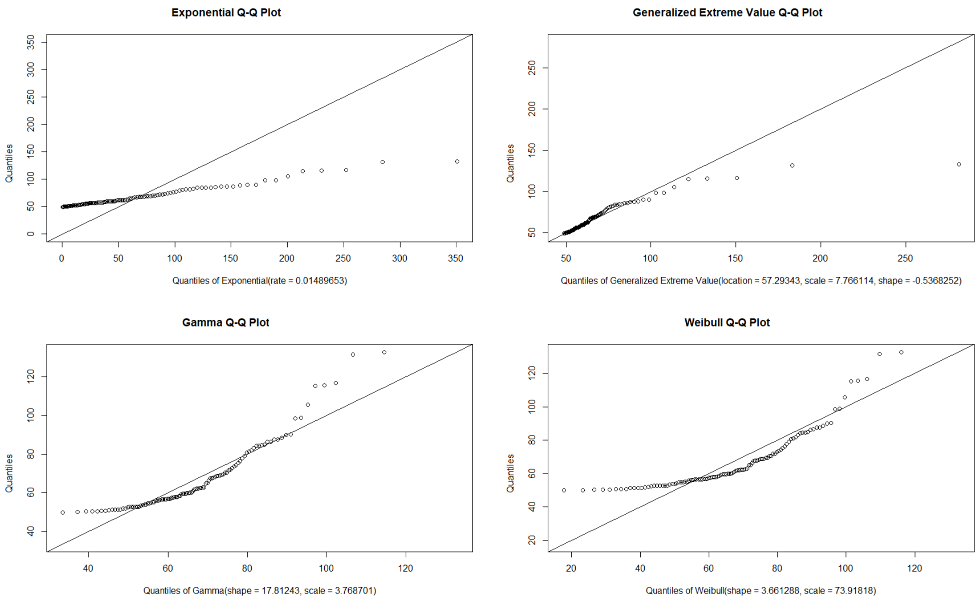


Figure 4.20: Q-Q Plot for extreme rainfall data in Ipoh with threshold value of 95^{th} percentile.

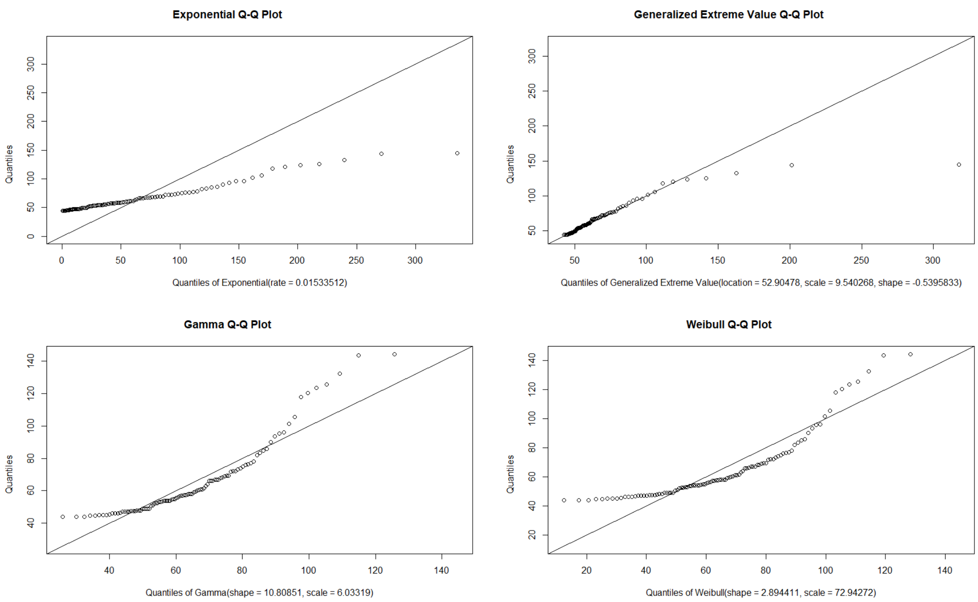


Figure 4.21: Q-Q Plot for extreme rainfall data in Gua Musang with threshold value of 95^{th} percentile.

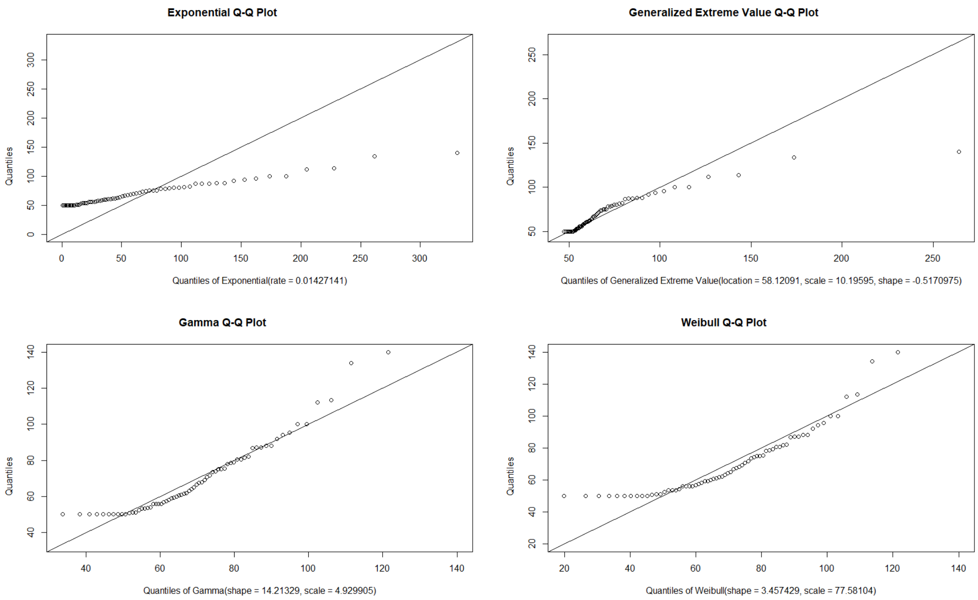


Figure 4.22: Q-Q Plot for extreme rainfall data in Bukit Berapit with threshold value of 95^{th} percentile.

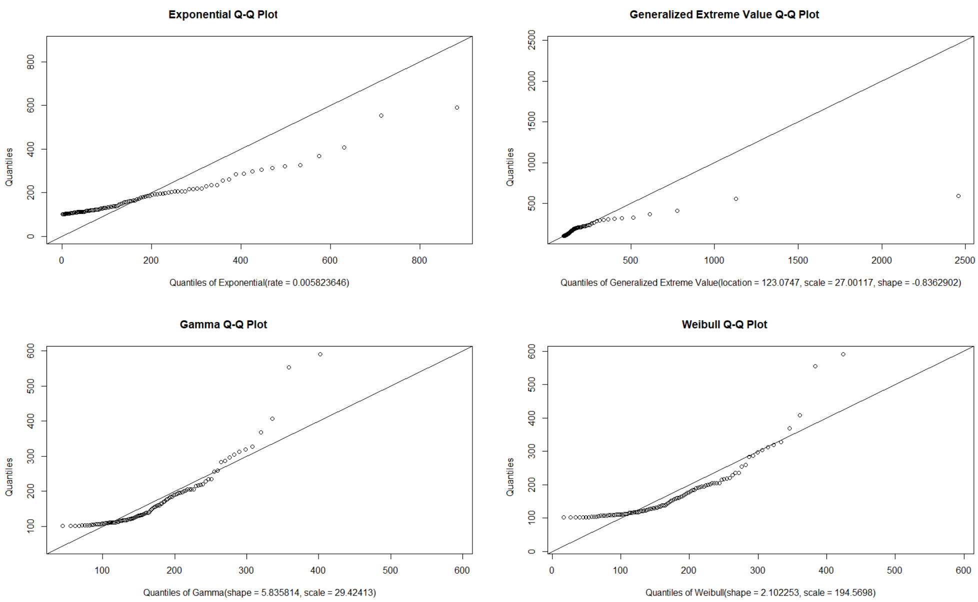


Figure 4.23: Q-Q Plot for extreme rainfall data in Kota Bharu with threshold value of 95^{th} percentile.

From Figure 4.4 to Figure 4.23, its show the Q-Q plot shows that the extreme rainfall data fit into generalized extreme value distribution the best, with the majority of the data fall around the straight line, while the gamma distribution and the Weibull distribution are the second best feasible choice with similar Q-Q plots, and the exponential distribution would be the least favourable distribution to be chosen for fitting the extreme rainfall model.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5-1 Introduction

This chapter concludes the results and findings in this study and the recommendation for the further study.

5-2 Conclusion

To conclude and make a final judgement from the large amount of rainfall data, both the numerical result obtained from using the goodness-of-fit test and the graphical result from the Q-Q plots are to be combined and compared.

For the goodness-of-fit test that rank different probability distribution based on which of the distributions are having the minimum test statistics value. GEV distribution shows the best result by having the lowest test statistics value in all three of the goodness-of-fit tests. Which means that by the numerical data obtained, GEV distribution would be the best probability distribution to be used in modelling the extreme rainfall data, and it would result in least amounts of error as compared with the other selected probability distributions.

For the Q-Q plots, the suitability of the selected probability distribution is visualized through the plotted graph. By comparing all the Q-Q plots, GEV distribution shows the best in fitting the extreme rainfall data with the least amount of data in further location of the fitted line shown in the plots.

In conclusion, both the numerical result from goodness-of-fit test and the graphical result from Q-Q plots shows that generalized extreme value distribution would be the best fit distribution for modelling the extreme rainfall data.

With the knowledge of generalized extreme value are the best to be use for modelling the extreme rainfall data amongst other distributions used in this project, it could be used to forecasting the future extreme rainfall event to helps Malaysian in making precautions and preventions before the happening of extreme rainfall that would possibly lead to flood. With some good precautions and preventions done before the hit of extreme rainfall, the risk of flood happening in the selected areas would be highly reduced, and would reduce the rate of mortalities and cost of repairing and rebuilding needed when the flood happens.

5-3 Recommendation

Some recommendations are suggested as a result of this study. The recommendations for the further research include:

1. Adoption of southwest monsoon for the study.

2. Include the climate information into the model.

References

Kar, K., Yang, S., Lee, J. and Khadim, F., 2017. 'Regional frequency analysis for consecutive hour rainfall using l-moments approach in jeju island, korea', *Geoenvironmental Disasters* **4**, 1–13.

Kumar, V., Khan, S. and Jahangeer, 2017. 'Statistical distribution of rainfall in uttarakhand, india', *Applied Water Science* **7**, 4765–4776.

Mehr, A., Nourani, V., Hrnjica, B. and Molajou, A., 2017. 'A binary genetic programming model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events', *Journal of Hydrology* **555**, 397–406.

Nguyen, T., Outayek, S. E., Lim, S. and Nguyen, V., 2017. 'A systematic approach to selecting the best probability models for annual maximum rainfalls – a case study using data in ontario (canada)', *Journal of Hydrology* **553**, 49–58.

Nguyen, V. and Nguyen, T., 2016. 'Statistical modeling of extreme rainfall processes (smexrain): A decision support tool for extreme rainfall frequency analyses', *Procedia Engineering* **154**, 624–630.

Paola, F., Giugni, M., Topa, M. and Bucchignani, E., 2014. 'Intensity-duration-frequency (idf) rainfall curves, for data series and climate projection in african cities', *SpringerPlus* **3**(1), 1–18.

Povera, A., 2019. 'New straits times: Storms, floods to worsen in malaysia, warns climate change expert', Available at: `https://www.nst.com.my/news/nation/2019/10/532908/storms-floods-worsen-malaysia-warns-climate-change-expert`. [Accessed 18 November 2019].

Prabhakar, A. K., Singh, K. K., Lohani, A. K. and Chandniha, S. K., 2019. 'Assessment of regional-level long-term gridded rainfall variability over the odisha state of india', *Applied Water Science* **9**, 1–15.

Sabarish, R., Narasimhan, R., Chandhru, A., Suribabu, C., Sudharsan, J. and Nithiyanantham, S., 2017. 'Probability analysis for consecutive-day maximum rainfall for tiruchirapalli city (south india, asia)', *Applied Water Science* **7**, 1033–1042.

Smith, R., 2015. 'Changing rainfall and humidity within southeast texas', *SpringerPlus* **4**(1), 1–11.

Syafrina, A. and Norzaida, A., 2017. 'Extreme rainfall analysis in the eastern region of peninsular malaysia', *RGN Publications* **8**(3), 241–251.