INTI
International University
LAUREATE INTERNATIONAL UNIVERSITIES

INTI INTERNATIONAL UNIVERSITY

Faculty of Science, Technology, Engineering and Mathematics

COMPARISON OF PARAMETER ESTIMATION METHODS USED IN PROBABILITY FITTING TO HYDROLOGICAL EVENTS

Michelle Tai Pei Wen
BEng (Hons) in Civil Engineering

Project Supervisor: Dr. Deepak T. J.

Final Year BEng Project
2015

ACKNOWLEDGEMENT

I thank INTI International University, for providing me a good and friendly environment to concentrate in my studies in the past five years. I appreciate what is being taught to me along my university life. I would like to thank all the lecturers from FOSTEM for the time they spent for teaching me.

Next, I thank Department of Irrigation and Drainage Malaysia for providing me the data needed in this thesis efficiently.

I would like to express my gratitude to my supervisor, Dr. Deepak T.J. and programme officer of FOSTEM, Ms. Gan Chai Lian for their support and patience all this while. I am especially thankful to Prof. Dr. Ni Lar Win for her guidance and encouragement.

Finally, I wish to say thanks to all my family and friends for their unconditional support for me in finishing up this thesis.

ABSTRACT

In hydrology engineering, parameter estimation is of importance in the analysis of probability distribution. It is especially useful in predicting extreme hydrological events like flood so that an accurate and precise outcome can be obtained. In this study, distributions chosen to fit the annual maximum discharge of Sungai Kelantan are the log-normal distribution, the log-Pearson type III distribution and the generalized extreme value distribution. Based on the best-fit distribution chosen according to Chi-squared test, a number of statistical methods by MOM, MLM and MLEs are used to estimate parameters of the distribution. The required expectations and quantile estimates can then be obtained from the distribution parameters with a return period of 2, 10, 25, 50, 100 and 200 years.

Keywords: Parameter Estimation, Method of Moments, Method of *L*-Moments, Maximum Likelihood Estimation, Frequency Analysis, Log-Pearson Type III Distribution, Chi-Squared Test, Goodness of Fit

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

CHAPTER 2 LITERATURE REVIEW

CHAPTER 3 METHODOLOGY

CHAPTER 4 RESULTS AND DISCUSSIONS

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASCE	American Society of Civil Engineers
CDF	Cumulative Distribution Function
DID	Drainage and Irrigation Department
EV1	Extreme Value Type I
EV2	Extreme Value Type II
EV3	Extreme Value Type III
GEV	Generalized Extreme Value
GOF	Goodness of Fit
GPD	Generalized Pareto Distribution
GU	Gumbel Distribution
JKR	Jabatan Kerja Raya
LN2	Log-Normal Distribution
LN3	Three Parameter Log-Normal Distribution
LPT3	Log-Pearson Type III Distribution
MLEs	Maximum Likelihood Estimations
MLM	Method of *L*-moments
MOLS	Method of Least Squares
MOM	Method of Moments
NO	Normal Distribution
PDF	Probability Density Function
PT3	Pearson Type III Distribution
PWMs	Probability-Weighted Moments
UNESCO	United Nations Economics, Social and Cultural Organization
USDA	United States Department of Agriculture
WMO	World Meteorological Organization

LIST OF NOMENCLATURES

β_r	Probability Weighted Moments
C	Runoff Coefficient
C_e	Equivalent Runoff Coefficient
C_k	Coefficient of Kurtosis
C_s	Coefficient of Skewness
C_v	Coefficient of Variation
E_i	Expected Frequency
f	Frequency
$f(x)$	Prabability Density Function
$F(x)$	Cumulative Distribution Function
g	Coefficient of Skew
i	Rainfall Intensity
K	Frequency Factor
k	Number of Classes
λ_r	L-Moments
m	Order Number
N	Number of Events
O_i	Observed Frequency
P	Exceedance Probability
Q_p	Peak Discharge
S_n	Reduced Standard Deviation
s	Standard Deviation
s_z	Standard Deviation of Neperian Logarithm
s^2	Variance
T	Return Period
t_c	Time of Concentration
v	Degree of Freedom
$\bar{x}$	Arithmetic Mean

χ^2	Pearson Cumulative Test Statistics
$\bar{y}_n$	Reduced Mean
z	Normal Distribution z Score
$\bar{z}$	Arithmetic Mean of Neperian Logarithm

CHAPTER 1 INTRODUCTION

1.1 General

Flood happens when the level at which river overflows its bank and inundates the adjoining area. Flood occurs after a long duration of heavy downpour which causes the flow in the river exceeding its carrying capacity.

According to Damodaran (2014) pointed out that a funding of 9300 million ringgit is required to build infrastructure to deal with flash flood problems in the urban areas in Malaysia based on the research done by Drainage and Irrigation Department (DID), 2013. DID has also determined 102 areas subjected to flash flood issues.

The massive flood events have occurred in the years 1886, 1926, 1967, 1971, 1973, 1979, 1983, 1993 and 1995 in Malaysia when Malaysia undergoes a significant of development. Flood affected approximately 15% of the population or about 2.7 million of residents which covers an area of 29,000 km^2. (Hiew, 1996).

In the beginning of year 2014, continuous heavy rainfall has caused floods in Kelantan, Sabah and Sarawak. Flood not only affects the daily activities of the people around the region, it also causes economic hardship due to the dying of livestock and food crops. In order to cope with this situation, the government has been establishing flood mitigation works throughout the past few decades. One of the most significant work done for flood management is SMART (Stormwater Management And Road Tunnel) Tunnel project to cater for major storm events in Kuala Lumpur.

In order to cope with flood issues so as to reduce the amount of funding spent on flood, a good forecasting of the frequency of occurrence of the floods is important. A proper parameter estimates can lead to an efficient prediction of flood based on the probability distribution fitting to provide best solution to flood problems.

1.2 Study Background

Malaysia, despite being one of the best country for their economy records in Asia, often strikes with hydrological events like flood and drought throughout the year. Malaysia consists total land area of 329,847 sq. km divided into two regions: Peninsular Malaysia and East Malaysia (Sabah and Sarawak). The topography of Peninsular Malaysia is separated into the eastern side and the western side of the Peninsula due to the existence of a central spine which is the Main Range/Banjaran Titiwangsa which extends from Thailand to Negeri Sembilan. The ground elevations in the states of Kelantan, Pahang and Terengganu can be up to 2000 meters above the mean sea level. The mountainous region relatively becomes flatter to the coastal plains. Similar to Peninsular Malaysia, East Malaysia also consists of similar topography in which the mountainous and higher grounds along a northeast-southeast are found in the interior of Sabah and Sarawak.

Due to the terrain of Malaysia, rainfall will flow from a ridge towards the lower elevations in the catchment area. Due to the steeper gradient in the upper stretches, the rivers are short compared to the rivers in lower stretches which are flat and meandering. For this reason, the duration of flow of water in rivers of lower stretches will be longer and a higher intensity towards the coastal plains. The concentration of population tends to be in lower and flatter areas, this makes them prone to the flood damage.

Due to its geographic location and equatorial climate, Malaysia has two monsoon seasons each year: southwest monsoon (April to October) and northeast monsoon (October to February). The average annual rainfall in Peninsular Malaysia is 2420mm, while for Sabah and Sarawak is about 2630mm and 3830mm respectively.

For this study, the comparison of parameter estimation methods will mainly focus in Kelantan River in Kelantan since Kelantan River Basin acts as a main river in state of Kelantan experiences flood frequently (DID, 2005). As for Kelantan, it is one of the states located in the north-eastern of the Peninsula, consists of 15,099 km^2 in total area. Based on Kanta and Mikaail (1990), Kelantan can be further divided into the northern part and the southern part according to its topography. The southern side of Kelantan is the mountainous area which forms the catchment of Kelantan River. This includes approximately 11,900 square kilometers or 85% of Kelantan. The four main river systems, namely Pergau, Galas, Lebir and Nenggiri rivers direct water out at Kuala Krai. For the northern part of Kelantan carries the flat and lowland plains.

Annual rain flow in Kelantan is 2100 mm in the centre-south and 3200 mm in the west. For coastal flood plains, average annual rainfall is between 2500 mm to 2800 mm. About 60% of the annual rainfall falls in the period of September to December due to the northeast monsoon. This follows by an abrupt stop in January or February. The Kelantan River often encounters major peak discharge during the monsoon season causing annual floods to occur in the northern plains around October to January since the water is merging to one river. The Mail Archive, 2006 stated that the flood occurred in 1967 has affected about 537,000 (84%) of the population in Kelantan. 125,000 people were evacuated and 38 people died from this incident. It has caused the area to be declared as Emergency Area.

1.3 Problem Statement

The use parameter estimation for analytical probability fitting based on empirical data to predict the frequency of flood occurrence is fairly important. The probability distribution which gives a close fit will lead to good estimation of flood events. The importance of parameter estimation lies in the fact that many statistical formulae are based on them, such as confidence interval and hypothesis testing, etc. Hence, it is rather critical to study the most efficient and the best method in parameter estimation for hydrological events.

1.4 Aim and Objectives

It is known that extreme hydrological events are often unpredictable. Apart from that, the effect of flood events can be reduced by preventive measures by estimating the value of return levels that might occur for the next 2, 10, 25, 50, 100 or maybe 200 years. Accurate statistical analysis of extreme rainfall data are often a result of good parameter estimation method used. A good estimate should be unbiased, have small variance, efficient and consistent.

The following are the objectives of this study:

(a) To determine the best-fit probability distribution for annual maximum flood events.

(b) To estimate the parameters using different methods for the chosen probability distribution in predicting hydrological events.

1.5 Scope of Study

Flood frequency analysis is done to estimate the peak discharge with different return period of floods. Several probability distributions which are the log-normal, log-Pearson Type III and generalized extreme value distribution are selected to get the best distribution to fit extreme hydrological events.

This study focuses on different parameter estimation methods used in frequency analysis for annual maximum flood events. Estimation methods considered in this study are method of moments (MOM), maximum likelihood estimations (MLEs), method of *L*-moment (MLM).

1.6 Report Organization

In Chapter 1 of this report, a general idea of the study background of this project and some information related to floods are described. The objectives and the scope are included as well.

In Chapter 2 of this report, it includes a more specific understanding of the study area of this project, Kelantan. It gives some basic knowledge that are required before conducting the research. It explains the different types of frequency analysis and parameter estimation methods by the others in a general way.

In Chapter 3 of this report, steps are provided in sequence so that the reader will get a better understanding knowing the methods used in achieving the aims and objectives of this project.

In Chapter 4 of this report, results of this research will be shown and discussed in this section. It includes the graph and tables used to present the outcomes of the study.

In Chapter 5 of this report, it shows conclusion of the study and recommendations for further study.

CHAPTER 2 LITERATURE REVIEW

2.1 The Study Area

Kelantan River Basin in Kelantan state, Malaysia is chosen as the study area for this project.

2.1.1 River Systems in Kelantan

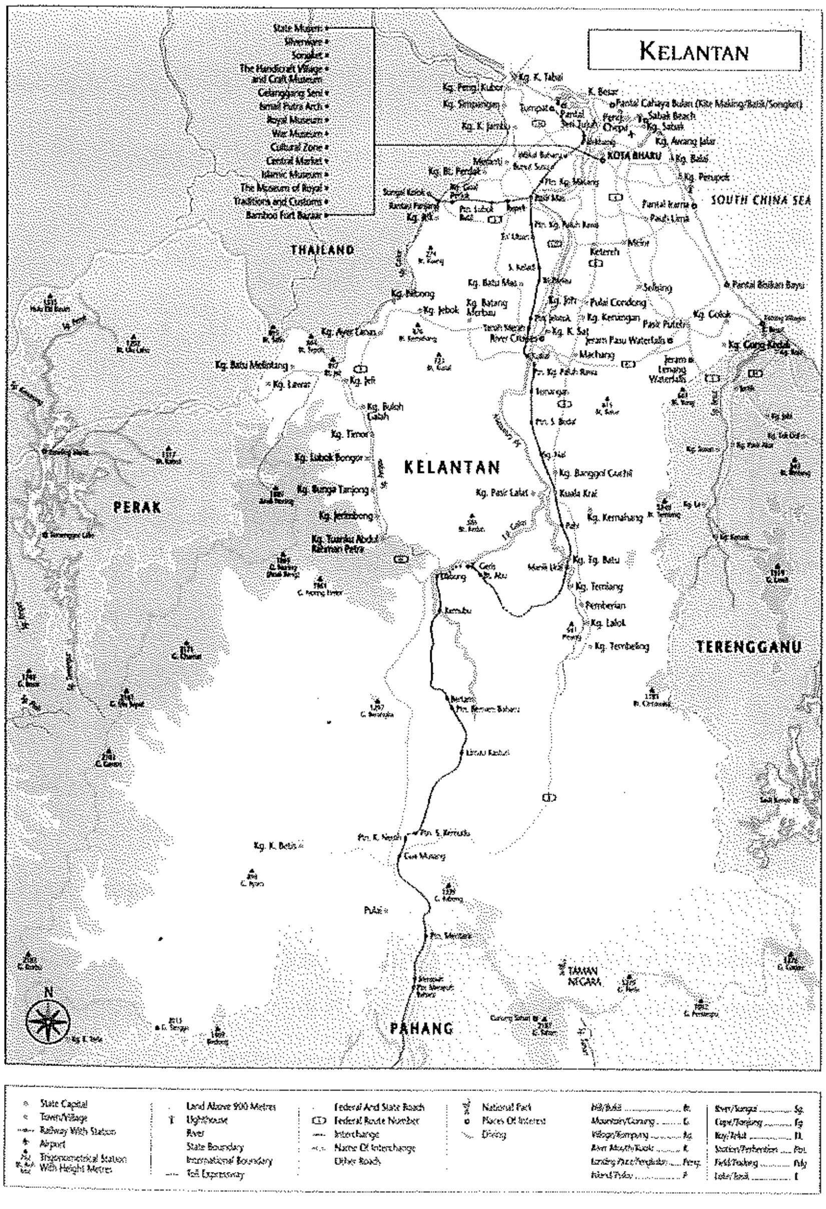

Figure 2.1: Map of River in Kelantan (Wonderful Malaysia, 2015)

The major rivers in Kelantan is made up of Kelantan River, Galas River and Lebir River. Kelantan River which acts as the main river of Kelantan is of length 248 km and covers a catchment area of 11,900 km^2. The highest elevation in the catchment area of Kelantan River is Mt. Korbu of height 2,183 m. For Galas River, a tributary of Kelantan covers the second large catchment area for Kelantan constitutes for about 7,770 square kilometers and has a length of 178 km. The highest peak around the region is Mt. Setong (1,422 m). While the third major river in Kelantan is the Lebir River of length 91 km and covers a catchment area of 2,430 km^2. The highest point in the catchment area is Cintawasa Hill of 1,185 m high.

Table 2.1: Characteristics of the River and the Main Tributaries. (UNESCO-IHP, 2002)

No.	Name of River	Length (km) Catchment Area (km^2)	Highest Peak (m)	Land Use (%) [1990]
1	Kelantan River (Main River)	248 11,900	Mt. Korbu (2,183 m)	F (73.2), U (0.1), L (10), OP (4), R (10), P(2.2), A (0.5).
2	Galas River (Tributary)	178 7,770	Mt. Setong (1,422 m)	F (85), OP (2), R (4), L (9).
3	Lebir River (Tributary)	91 2,430	Cintawasa Hill (1,185 m)	F (66), OP (12), R (13), L (9).

A: Other agricultural field (vegetable, grass) F: Forest L: Lake, River, Marsh O: Orchard

U: Urban R: Rubber OP: Oil Palm P: Paddy field

2.1.2 Kelantan River

Kelantan River basin lies between the latitudes N 4° 40' to N 6° 12' and longitudes E 101° 20' to E 102° 20'. Kelantan River has a maximum length of 150 km and a maximum breadth of 140 km. Located at the north eastern side of Peninsula, water flows northward passing main towns in Kelantan such as Kota Bharu, Tanah Merah, Pasir Mas and Kuala Krai before converging to South China Sea.

About 100 km from the estuary, it is divided into Galas River and Lebir River at Kuala Krai. The Nenggiri River which flows from the central mountain range of Peninsular Malaysia meets Pergau River at the junction and forming Galas River. For the case of Lebir River, water flows from the Tahan mountainous zone. The topography of the catchment area in Kelantan which consists of about 95% of the area is of steep mountainous zone with an elevation of 2,135 m in height. The remaining 5 percent is made up of flat undulating land.

An average runoff of Kelantan River taken at Guillemard Bridge is 557.5 m^3/s generated due to an elevation difference of more than 2,100 meters. (UNESCO-IHP, 2002)

2.1.3 Major Floods in Kelantan

Floods can be classified into two types in Malaysia, namely monsoon floods and flash floods. Monsoon floods for northeast monsoon and southeast monsoon normally affects large areas in Malaysia. This kind of floods can last for a few weeks causing damage to the infrastructures, proper lifestyle of the affected population and even endanger the lives of many. Monsoon floods usually happen in inland areas, coastal flood plains and urban areas (Sinha, 2007). While flash floods will occur without prior warning in city areas. Flash floods can be destructive since it occurs within a very short time and there is no time for preparation.

There are several major flood events that happened in the past few decades in the state of Kelantan. The disastrous flood, Bah Air Merah, meaning reddish water flood, happened in the year 1926 which destroyed the homes of many. Reddish flood occurred due to the heavy rainfall inland and at the upstream of the river where the excess water washed the hills and slopes in a high velocity. The resulting flood water that covered the towns and villages is red with high silt content (Rashid, 2012).

Another major flood event happened in Jan 1967. During the flood, approximately 84% of the Kelantan population were affected and 125,000 people have been evacuated to a safer place. It had caused a damage of RM 30,000,000 roughly (Sani, 1974). Major floods also occurred in the year of 1983, 1988, 1993 and 2004 causing more than ten thousands of people to move to a safer place. Most of these events is due to heavy rainfall in a long duration.

2.2 Factors Affecting Peak Discharge

Peak discharge for a river is governed by complex relationships of the meteorological and watershed factors. Meteorological factors control the rising limb of the flood hydrograph and the recession limb is dependent on watershed factors only.

2.2.1 Meteorological Factors

With the increase in intensity of rainfall, rainfall magnitude, duration of storm over a catchment area, the peak discharge will generally increase. Thus, it changes the shape of a flood hydrograph. The peak discharge and the volume of the surface runoff are both proportional to the rainfall intensity for a given duration (Subramanya, 2009). The shorter duration of a storm rainfall with the same magnitude, the higher the peak discharge. Peak discharge will decrease with the same amount of rainfall happens in a longer duration (USDA, 1989).

The movement or direction of the storm will has an effect on the flood hydrograph. When the storm moves from an upstream to a downstream in a river, there will be a higher concentration of flow at the river mouth. Inversely, if the storm moves towards the upstream of a river, this will results in lower peak discharge and a longer time base (Subramanya, 2009).

2.2.2 Watershed Characteristics

Based on what is stated in USDA (1989), a smaller watershed or a more obvious watershed characteristics will have a significance influence in peak discharges. Typically, for a larger watershed with a similar characteristics will give a larger peak discharge. Overland flow is more prominent over the channel flow for a rainfall in a small catchment. The peak discharge in a small catchment will depends on the land use and the rainfall intensity. While for larger catchment area, channel flow is much more dominant than the overland flow (Subramanya, 2009).

On the other side, the time taken for the water from the furthest point of a catchment basin to reach the river outlet depends on the shape of the basin. Basin shape will affect the shape of a hydrograph and also the occurrence of peak discharge (Subramanya, 2009).

The velocity of flow is governed by the slope of the main arteries in the channel. Stream channel will affect the recession limb of the hydrograph which shows the depletion of storage (Subramanya, 2009). Catchment area of steeper slopes produce larger volume of peak discharge (USDA, 1989).

In Subramanya (2009), cover of the catchment area of a stream which directly influence the runoff coefficient C also affects the amount of runoff. The coefficient C which relies on rainfall intensity, slope of the surface and the nature of the surface shows the total effect of the catchment losses. In short, having a higher vegetative cover or an impervious surface results in a higher surface runoff. A decreased density of vegetation also imply a lower rate of interception and infiltration to the ground. This in turn increase the surface runoff rate and the peak discharge rate. The usage of land can affect the discharge to the river. Land usage that promotes the infiltration and the surface water storage capability decrease the peak flow by increase the duration of the flow to reach the catchment outlet (USDA, 1989).

The availability of the surface storage like lakes, ponds and swamps can greatly affect the peak discharge. Surface depression can capture the rainfall later infiltrate over a longer duration. According to USDA (1989), the flow of water takes a longer time to reach the basin outlet which causes the overall discharge to be smaller.

The soil type and the geological condition of the catchment area affect the peak discharge and the lag time. When a stream is surrounded by impervious soil layer, rain water travels down to the stream via overland flow. The lag time between the peak rainfall and the peak discharge is reduced significantly. The rate of infiltration will increase with the unconsolidated soils surrounding the river allowing much of the water to pass through.

2.3 Peak Discharge

The highest rate of surface runoff or discharge from a catchment area for a given rainfall is the peak discharge. Determination of peak discharge is rather important in order to design for drainage systems, sewerage systems, dam design, road and bridges design. It is useful in flood prevention works also (Mishra et al., 2003). Due to the increase in the runoff coefficient, the peak discharge is higher in the urban area. According to Johnston et al. (1980), for storms on Canadian Sandy Creek watershed, the peak discharge increase three folds compared to its historical data. This further confirms the need to come out with an efficient and accurate parameter estimation methods for floods.

2.3.1 Determination of Peak Discharge

These are the alternatives in determination of flood peak: Rational method, unit hydrograph technique and flood frequency based methods.

In Subramanya (2009), in order to determine which methods to be used, it depends on how important the project is, the data availability and also the objective of the project. For a catchment area of less than 50 square kilometers, rational method should be used. While for a moderate size of watershed which is less than area of 5000 square kilometers, the usage of unit hydrograph technique is suitable.

2.3.2 Rational Method

A few limitations when using rational method have been described by McPherson (1984). Rational method is highly dependent to the runoff coefficient which is taken as constant throughout the storm event and the time of concentration of rainfall t_c. This method is used with the assumption that rainfall occurs for a constant duration and is uniformly distributed in the catchment area. The peak discharge from the rational method will assume to have the same return period with the rainfall intensity used.

The rational equation is

$$Q_p = \frac{1}{3.6} C\left(i_{tc,P}\right) A \qquad (2.1)$$

where Q_p is the peak discharge in m^3/s, C is a dimensionless coefficient of runoff, $i_{tc,P}$ is the intensity of rainfall for a t_c duration and an exceedence probability of P.

Time of concentration t_c is the time taken for a drop of water from the most hydraulically furthest point of the catchment area to flow to the outlet (Subramanya, 2009). Maximum flow distance, rainfall intensity, infiltration rate, surface slope and roughness are the major factors affecting the time of concentration for overland flow (ASCE, 1992). The most common method used relating time of concentration of travel length is the following equation which was developed by Kirpich (1940):

$$t_c = 0.01947\, L^{0.77} S^{-0.385} \qquad (2.2)$$

where L is the maximum hydraulic length in meter, S is the mean slope along the hydraulic length expressed as a fraction and t_c is the time of concentration in minutes.

Runoff coefficient is a difficult factor to determine since it depends on factors such as interception, surface detention, infiltration and antecedent conditions. It is not a constant coefficient but changes with the frequency of the runoff event. For composite areas having different runoff coefficient, the equivalent runoff coefficient C_e can be calculated on an area weighted basis from

$$C_e = \frac{\sum_1^N C_i A_i}{A} \tag{2.3}$$

where C_i is the coefficient with respect to the area A_i. When large parts of an area has the same runoff coefficient, it can be considered as a single type of layout.

In Design Hydrology and Sedimentology for Small Catchments by Haan et al. (1994) stated that a surface runoff rate will increase to a state where the entire catchment area is contributing to the runoff when a uniformly distributed rainfall occurs in a catchment area. When a rainfall duration is less than the time of concentration t_c, the whole of the catchment area is not contributing to the runoff.

2.3.3 Unit Hydrograph Technique

According to Ghosh (2014), a hydrograph is a graphical representation of a stream's discharge versus the time. A hydrograph generally consists of a rising limb, crest segment, recession limb. Unit hydrograph technique acts as a tool for peak discharge analysis. This method was first suggested by Sherman (1932). Unit hydrograph is a hydrograph that gives direct runoff volume of one unit depth (1 cm) for a rainfall of specified duration distributed uniformly over the catchment area.

The main advantage of using unit hydrograph is that each unit hydrograph remains constant shape in a catchment area unless the characteristics of basin are changed. The actual flood hydrograph is obtained by multiplying each of the discharge ordinate in unit hydrograph by the rainfall excess from the observed rainfall record (Weaver, 2003).

2.3.4 Flood Frequency Based Methods

Floods are complicated natural events that depend on a number of component parameters. This results in estimation of flood peak discharge rather complex. Statistical method of frequency analysis is another method used in annual maximum discharge estimation. By using of probability distributions, frequency analysis relates the flood magnitude to their frequency of occurrence (Chow et al., 1988). There are a few assumptions made when analyzing flood frequency analysis using data obtained for a period of time of a river system. It is taken that the floods have not been affected by manmade changes or natural in the hydrological regime. The data also assumed to be stochastic, time and space independent (Hamed and Rao, 1999).

Information obtained from the flow record is used to estimate the relationship between Q and T. Frequency analysis using frequency factors through probability distribution helps in knowing the relationship between the extreme flood magnitude and the number of occurrence. A frequency factor, K shows the frequency of occurrence of an event which relies on the distribution of flood events (Singh et al., 2012). Chow (1951) has shown that most frequency distribution can be expressed by the hydrologic frequency analysis general equation:

$$x_T = \bar{x} + K\sigma \tag{2.4}$$

where x_T is the maximum value of event of a hydrologic series with a recurrence interval T, $\bar{x}$ is the mean, K is the frequency factor which depends on assumed frequency distribution and the return period T and σ is the standard deviation. Normal distribution, Log normal distribution, Log-Pearson Type III distribution and Gumbel's extreme-value distribution are some of the common frequency distribution functions for predictions of flood events. The statistical parameters used are:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x \tag{2.5}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x-\bar{x})^2}{N-1}} \tag{2.6}$$

where $\bar{x}$ is the mean of the annual maximum discharges (m^3/s), x is annual maximum discharges (m^3/s), σ is the standard deviation (m^3/s) and N is the number of flood events.

Generalized extreme value (GEV) distribution is the most suitable distribution for annual maximum rainfall in Peninsular Malaysia (Zalina et al., 2002). While Zin et al. (2009) stated that GEV and generalized Pareto distribution (GPD) fits well in the annual series in Peninsular Malaysia. In the work of Jain and Singh (1987), some flood frequency distributions such as Gumbel's method, 3-parameter log-normal, Pearson type III, log-Pearson type III and Boughton distributions are compared using 55 flood data from various parts of U.S.A. Log-Pearson type III was the most accurate distribution based on empirical criteria.

2.4 Rating Curve

Most often in river or stream, the discharges are estimated from measured water stages of the stream. A rating curve therefore is needed to relate the water stages measured to discharges. Hornberger et al. (1998) stated that rating curves are usually approximated using a power function due to the nonlinear characteristics. The usage of rating curve to estimate higher discharges may become unreliable during the occurrence of overbank flow which involves multiple major flood paths. The analysis of peak discharge using rating curve is an approximation rather than an exact solution. Peak annual flood level is converted to peak annual discharges after the rating curve is done and continues by doing a frequency of analysis of the discharges (Csiro, 2000).

2.4.1 Plotting Position

In flood frequency analysis method, the sample data can be plotted on probability paper to analyse the distribution curve, to check for outliers and to detect errors. In order to have probability plots, the plotting position or probability of non-exceedence are required. Hence, by using the probability-weighted moments (PMWs) method, parameters estimation can be done using plotting positions. Table 2.2 below shows some commonly used plotting position formulas where m is the order number of floods arranged in decreasing magnitude and N is the total number of flood events.

Table 2.2: Commonly Used Plotting Position Formulas (Cunnane, 1989)

Method	Plotting Position, P
Weibull	$\frac{m}{N+1}$
Gringorton	$\frac{m-0.44}{N+0.12}$
Hazen	$\frac{m-0.5}{N}$
Hosking	$\frac{m-0.35}{N}$
Blom	$\frac{m-0.375}{N+0.25}$
Cunnane	$\frac{m-0.4}{N+0.2}$
California	$\frac{m-1}{N}$
Chegodayev	$\frac{m-0.3}{N+0.4}$
Adamowski	$\frac{m-0.24}{N+0.5}$

2.4.2 Return Period

Return period or recurrence interval is the average time interval of extreme rainfall of magnitude will be equaled to or larger than a specified magnitude at least once (Patra, 2001). Flood peaks will not occur in a same period of time and magnitude, therefore they have a different time intervals. The return period is calculated as

$$T = \frac{1}{P} \tag{2.7}$$

2.5 Normal Family

2.5.1 Normal Distribution (NO)

Normal distribution is one of the most popular continuous probability distribution. It is also recognised as the Gaussian distribution or error function. Hazen (1914) introduced the normal probability paper for hydrologic data analysis made it one of the earliest applications of normal distribution in hydrology. Slack et al. (1975) recommended the use of normal distribution where there are loss in the flood distribution information and losses due to economic related with the flood retardation design. Markovic (1965) fitted the normal distribution to annual rainfall and runoff data.

One of the advantages of using normal distribution is that it is extensively tabulated and the standardized normal variate is the same as the frequency factor. The rainfall data are arranged according to their plotting positions to perform cumulative frequency analysis. The probability density function is

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.8}$$

where μ is the mean or median or mode in this distribution (location parameter), σ is the standard deviation and σ^2 is the variance (scale parameter). The range of value x is between $-\infty$ and $+\infty$.

When $\mu = 0$ and $\sigma = 1$, normal distribution will yield standard normal distribution. The graphical plot of probability density function of normal distribution is symmetrical with the mean value at its center. Hence it is not suitable for data that has a high value of skewness. Such distribution may described by log-normal distribution or the Pareto distribution. The cumulative distribution function is

$$f(x, \mu, \sigma) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right] dx \tag{2.9}$$

Variable X is normalized as $u = (x - \mu)/\sigma$, the equation becomes

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) \tag{2.10}$$

Table 2.3: Summary of Normal Distribution (WMO, 2009)

Normal Distribution (μ, σ^2)	
Parameters	μ – location (mean), σ^2 – squared scale (variance).
Support	$-\infty < x < \infty$
Probability Density Function	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Cumulative Distribution Function	$\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right] dx$

Quoted from Bhim Singh et al. (2012), frequency factor K is expressed by the following equation (Chow et al., 1988)

$$K = \frac{x - \mu}{\sigma} \tag{2.11}$$

This is the same as standard normal variate z. The value of z corresponding to an exceedence of P can be calculated by knowing the intermediate variable w,

$$w = \left[ln\left(\frac{1}{P^2}\right)\right]^{\frac{1}{2}}, 0 < P \leq 0.5 \tag{2.12}$$

$$z = w - \left[\frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}\right] \tag{2.13}$$

where P is the probability of exceedence. When $P > 0.5$, $(1 - P)$ is substituted for P. The calculated z value will give a negative sign (Bhakar et al., 2006). The error obtained is less than 0.00045 when using the above equation to estimate the frequency factor (Chin, 2006).

The estimated floods with various return periods is calculated by

$$x_T = \bar{x} + K\sigma \tag{2.14}$$

2.5.2 Log-Normal Distribution (LN2)

Basak et al. (2009) stated that log-normal distribution has a normally distributed logarithm and it is effective to model data which is skewed positively and long-tailed. For instance, $Y = \log(X)$ has a normal distribution if the random variable X has a log-normal distribution. $X = \exp(Y)$ has a log-normal distribution if Y is normally distributed. It has a probability density function of

$$Fx\ (x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}} \tag{2.15}$$

and a cumulative distribution function of

$$Fx\ (x; \mu, \sigma) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right] \tag{2.16}$$

Table 2.4: Summary of Log-Normal Distribution (WMO, 2009)

Log-Normal Distribution (μ, σ)	
Parameters	σ – shape, μ – log scale.
Support	$0 < x < \infty$
Probability Density Function	$\frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}$
Cumulative Distribution Function	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right]$

From Reddy (1997), if a random variable x is made up with the sum of many small effects, the x is expected to be normally distributed in the central limit theorem. Likely, the $\ln x$ can say to be normally distributed if x is equal to the products of many small effects (Ewemoje and Ewemooje, 2011). The sample data is transformed into logarithmic form. Subramanya (2009) states that if X is the variate of random hydrologic series, the series of Y variates is

$$y = \log x \tag{2.17}$$

$$y_T = \bar{y} + K_y \sigma_y \tag{2.18}$$

$$K_y = \frac{y_T - \bar{y}}{\sigma_y} \tag{2.19}$$

The frequency factor K_y can be determined from Appendix B with coefficient of skew of variate equal to zero.

$$x_T = antilog\ (y_T) \tag{2.20}$$

2.5.3 Three Parameter Log-Normal Distribution (LN3)

When a log-normal distribution has three parameters, it becomes three parameter log-normal distribution (LN3). Hoshi and Burges (1978) proposed approximating the normal populations with three parameter log-normal distributions to facilitate multivariate hydrologic disaggregation or generation schemes in cases where mixed normal and log-normal populations existed. The equation of $\mathrm{Y} = \log(X - \gamma)$ is a normal distribution with two parameters of mean μ and standard deviation σ, it becomes a log-normal distribution for X with the parameters γ, μ, σ. The probability density function is