REGIONAL FREQUENCY ANALYSIS OF MAXIMUM DAILY RAINFALLS USING TL-MOMENT APPROACH

NORATIQAH BINTI MOHD ARIFF

A dissertation submitted in partial fulfillment of the requirements for the award of the degree of Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

OCTOBER 2009

To those whose moral supports and love had helped me countless of time to overcome each and every obstacle

Father, Mohd Ariff bin Omar

Mother, Hajar binti Hawari

My lovely sisters,
Norlina binti Mohd Ariff
Norasyiqin binti Mohd Ariff

I thank God for blessing me with all your presence.

ACKNOWLEDGEMENT

First of all, I am grateful to the Almighty Allah S.W.T. because with His blessings, I am able to finish my Master's dissertation in the allocated time given. I thank God again for giving me good health in order for me to successfully complete this thesis.

I wish to express my sincere gratitude to my supervisor, Dr. Ani Shabri for his guidance in helping me throughout this thesis and not forgotten to Encik Abu Salim from "Jabatan Pengairan dan Saliran, Malaysia" who made it easy for me to collect the data for my thesis.

I am very much indebted to my beloved family members who had helped me in each and every step of the way. All your patience, understanding, love and kindness have blessed my life in more ways than one.

Last but not least, I am grateful to all my friends and all those who had helped me to accomplish this project. May Allah repay all the kindness that you have given me thus far.

ABSTRACT

Analyzing rainfalls data are important in order to obtain the probability distribution of flood. The main aim of the study is to perform regional frequency analysis of maximum daily rainfalls measured over stations in Selangor and Kuala Lumpur by using the TL-moment method with $t = 0$, $t = 1$ and $t = 2$. Initially, the maximum of each daily rainfall for each year were obtained. Then, parameters of every distributions considered including the normal (N), logistic (LOG), generalized logistic (GLO), extreme value type I (EV), generalized extreme value (GEV) and generalized Pareto (GPA) distribution were estimated using TL-moment approach. TL-moments with $t = 0$ are known as L-moments while TL-moments with $t = 1$ and $t = 2$ imply TL-moments that are symmetrically trimmed by one and two conceptual sample values respectively. The most suitable distribution were determined according to the mean absolute deviation index (MADI), mean square deviation index (MSDI) and correlation, *r*. L-moment and TL-moment ratio diagrams provided visual proofs of the results. The L-moment method showed that the generalized logistic (GLO) distribution is the best distribution whilst TL-moment method with $t = 1$ and $t = 2$ concluded that the extreme value type I (EV) and generalized extreme value (GEV) distributions are the most suitable distributions to fit the data of maximum daily rainfalls for stations in Selangor and Kuala Lumpur.

.

ABSTRAK

Penganalisaan taburan hujan adalah penting untuk mendapatkan taburan kebarangkalian banjir. Objektif utama kajian ini adalah untuk menjalankan analisis frekuensi rantau terhadap data hujan harian maksimum yang diukur pada stesen-stesen hujan di Selangor dan Kuala Lumpur menggunakan pendekatan TL-momen dengan t =0, t = 1 dan t = 2. Pada mulanya, jumlah maksimum hujan bagi setiap tahun dikenalpasti. Kemudian, parameter untuk setiap taburan yang diambil kira termasuk taburan normal (N), logistik (LOG), logistic teritlak (GLO), nilai ekstrim Jenis I (EV), nilai ekstrim teritlak (GEV) dan Pareto teritlak (GPA) dikira melalui kaedah TL-momen. TL-momen dengan t = 0 merupakan kaedah L-momen manakala TL-momen dengan $t = 1$ dan $t = 2$ menunjukkan TL-momen yang ditrim secara simetri oleh satu dan dua nilai sampel masing-masing. Taburan yang paling sesuai untuk mewakili data stesen-stesen ini dikenalpasti melalui sisihan indeks min mutlak (MADI), sisihan indeks min kuasa dua (MSDI) dan korelasi, r. Rajah nisbah L-momen dan TL-momen digunakan sebagai bukti dapatan kajian. Hasil dari kajian ini didapati bahawa apabila menggunakan kaedah L-momen, taburan logistik teritlak (GLO) adalah taburan terbaik manakala penggunaan kaedah TL-momen dengan $t = 1$ dan $t = 2$ menunjukkan taburan nilai ekstrim Jenis I (EV) dan nilai ekstrim teritlak (GEV) adalah kaedah paling sesuai bagi mewakili data taburan hujan harian maksimum di stesen-stesen Selangor dan Kuala Lumpur.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

K	-	shape parameter
r	-	correlation
s_{xx}^2	-	sample variance for observed flows
s_{zz}^2	-	sample variance for predicted flows
s_{xz}^2	-	sample covariance
x_i	-	observed flows
z_i	-	predicted flows
α	-	scale parameter
μ	-	mean of the x series
σ	-	standard deviation of the x series
ξ	-	location parameter
τ_3	-	L-moment skewness
τ_4	-	L-moment kurtosis
τ_3^1	-	TL-moment skewness with $t = 1$
τ_4^1	-	TL-moment kurtosis with $t = 1$
τ_3^2	-	TL-moment skewness with $t = 2$
τ_4^2	-	TL-moment kurtosis with $t = 2$
Φ	-	standard normal distribution function
$\Phi^{-1}(F)$	-	the inverse of standard normal distribution function

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Flood in Malaysia

Human society faces great problems due to extreme environmental events. For example, floods, rainstorms, droughts and high winds that cause tornadoes and such destroy almost anything that is in their vicinity at the moment of occurrences. Flood, also known as deluge, is a natural disaster that could diminish properties, infrastructures, animals, plants and even human lives.

In terms of the number of population affected, frequency, area extent, duration and social economic damage, flooding is the most natural hazard in Malaysia (Ministry of Natural Resources and Environment, Malaysia, June 2007). According to the Ministry of Natural Resources and Environment, Malaysia in June 2007, Malaysia has experienced major floods since 1920 especially in the years 1926, 1963, 1965, 1967, 1969, 1971, 1973, 1979, 1983, 1988, 1993, 1998, 2005, December 2006 and January 2007. These flood events occurred in various states including Selangor and the capital city of Malaysia, Kuala Lumpur.

The basic cause of river flooding is the incidence of heavy rainfall (monsoon or convective) and the resultant large concentration of runoff, which exceeds river capacity (Ministry of Natural Resources and Environment, Malaysia, June 2007). Flood had resulted in a loss of millions in Malaysia. For example, the 1971 flood that hit Kuala Lumpur and many other states had caused more than RM200 million losses and 61 deaths. Furthermore, the massive floods due to a few abnormally heavy rains in 2006 and 2007 cost RM 1.5 billion and hence deemed as the most expensive flood events ever to occur in Malaysian history. This includes the cost of damage in infrastructures, bridges, roads, agriculture and private commercial and residential properties. During this flood event, 18 people unfortunately died and around 110,000 people were evacuated from their homes and were sheltered in relief centers.

1.2 Introduction to Flood Frequency Analysis

Analyzing rainfalls and stream flows data are important in order to obtain the probability distribution of flood and other phenomenon related to them. By knowing the probability distribution, prediction of flood events and their characteristics can be determined. With this, prevention acts and measures can be taken and flash flood warning models can be built easily.

The study of water related characteristics and modeling throughout the Earth such as the movement, distribution, resources, hydrologic cycle and quality of water is called hydrology. By knowing and analyzing statistical properties of hydrologic records and data like rainfall or river flow, hydrologists are able to estimate future hydrologic phenomena. A very active area of investigation in Statistical Hydrology is the frequency of floods (Rao et al., 2000).

As stated earlier, flood is the most costly natural hazard in Malaysia. It is also one of the oldest natural hazards in the world. Hence, its characteristics and the magnitude-recurrence interval relationship are important for hydrologist to design or plan hydrological projects. In order to be able to plan and design these projects such as hydraulic or water resources projects, continuous hydrological data, for example, rainfalls data or river flow data is necessary. With the help of the data, flow pattern or trend can be determined to make sure the design and planning can be done accordingly. Hydraulic structures such as weirs, barrages, dams, spillways and bridges can be modeled and damages can be minimized with a reliable and good estimation of magnitude and frequency of occurrence of such extreme events. Many aspects of water resources engineering and hydraulic studies need to estimate region or for a group of sites (Rao et al., 2000). However, to select a reliable design quantile, which has affect on design, operation, management and maintenance of hydraulic structure depends on statistical methods used in parameter estimation belonging to probability distribution (Hosking and Wallis, 1993).

Estimating flood and designing water related structure, erosion and agricultural considerably need knowledge related to distributions of extreme rainfall depths. Probability for future events can be predicted by fitting past observations to selected probability distributions. The primary objective is to relate the magnitude of these extreme events to their frequency occurrence through the use of probability distributions (Chow et al., 1988).

However, extreme events are usually too short and too rare for a reliable estimation to be obtained. This also includes the difficulties of identifying the appropriate statistical distribution to describe the data and estimating the parameters of the selected distribution. Hence, regional frequency analysis which was developed by Hosking and Wallis (1991) is used since it can resolve this problem by trading space for time.

With this method, this problem will be resolved. According to Cunnane (1989), regional analysis is based on the concept of regional homogeneity which assumes that annual maximum flow populations at several sites in a region are similar in statistical characteristics and are not dependent on catchments size.

1.3 Introduction to L-Moment and TL-Moment

Extreme events such as flood are rare and often occur in a short amount of time. Hence, it is difficult to analyze the characteristics of its statistical probability distributions. By replacing space for time, frequency analysis is used to obtain the probability distributions for extreme events. Outliers are common to be found in data related to flood which is an extreme natural hazard.

Recently, the most popularized method in frequency analysis is the L-moment approach introduced by Hosking in 1990 (Rao et al., 2000). The main role of the L-moments is for estimating parameters for probability distributions. L-moments' estimates are superior to standard moment-based estimates generally and especially for small samples. They are also relatively insensitive to outliers compared to conventional moments. Their small sample bias tends to be very small. L-moments are also preferable when maximum likelihood estimates are unavailable, difficult to compute or have undesirable properties.

Probability distributions are used to analyze data in many disciplines and are often complicated by certain characteristics such as large range, variation or skewness. Hence, outliers or highly influential values are common (Asquith, 2007). Outliers can have undue influence on standard estimation methods (Elamir and Seheult, 2003). According to Elamir and Seheult, if there is a concern about extreme observations having undue influence, a robust method of estimation which is developed to reduce the said influence of outliers on the final estimates should be preferable. TL-moments are derived by Elamir and Seheult in 2003 from L-moments and might have additional robust properties compared to L-moments. In other words, TL-moments are claimed to be more robust than the L-moment. Hence, for extreme data, TL-moments are also considered for estimating the parameters of the selected probability distributions.

Thus, this study focused on identifying a suitable probability distribution, including normal (N), logistic (LOG), generalized logistic (GLO), extreme value type I (EV), generalized extreme value type I (GEV) and generalized Pareto (GPA) by using TL-moments technique for maximum daily rainfalls selected for each year among daily rainfalls measured over the regions in Selangor and Kuala Lumpur, Malaysia. The TL-moments for all the said distributions were derived in order to be able to fit the rainfall data to the probability distributions.

In the case of TL-moments which are symmetrically trimmed by one conceptual sample value, i.e $t_1 = t_2 = t = 1$, for normal (N) and logistic (LOG) distributions, the TL-moments and their parameter estimates were computed and checked with those obtained by Elamir and Seheult in 2003. Meanwhile, the TL-moments and their parameter estimates for generalized logistic (GLO), extreme value type I (EV), generalized extreme value type I (GEV) and generalized Pareto (GPA) distributions were derived since none had been done before. However, for TL-moments which are symmetrically trimmed by two conceptual sample values, i.e $t_1 = t_2 = t = 2$, all six distributions' TL-moments and their respective parameter estimates were all derived in this study. The results from both cases ($t = 1$ and $t = 2$) were then compared with those obtained using the method of L-moments similar to the previous study by Shabri and Ariff (2009).

1.4 Objectives of the Study

The objectives of this study are:

i. To derive the TL-moments for the selected distributions to be considered. Not all distributions' TL-moments have been derived thus far. New derivation will be done for generalized logistic (GLO), extreme value type I (EV), generalized extreme value type I (GEV) and generalized Pareto (GPA) distributions.
ii. To obtain the respective parameter estimates for each distribution.
iii. To find the most suitable distribution to fit the maximum daily rainfalls data by using the goodness-of-fit tests.
iv. To compare the results with the ones obtained from using the method of L-moments in the previous study.

1.5 Scope of the Study

The scope of this study consisted of the TL-moment approach on maximum daily rainfalls through regional frequency analysis. The data of maximum daily rainfalls were selected each year and measured over stations in Selangor and Kuala Lumpur. The aims of this study were to derive new TL-moments population and to determine the best probability distribution among the selected distributions whose parameters were estimated using the method of TL-moment. Furthermore, this study included the comparison between the results achieved and those that have been obtained through the L-moment approach.

1.6 Significance of the Study

The results of this study give benefits to statistical and hydrological studies. The direct beneficiaries of the study are the statisticians, applied mathematicians, engineers and hydrologists working in the research areas of applications from the result of specifying the probability distribution of extreme events which in this case is flood. Thus, this also helps our country from unnecessary cost and economic losses as well as preventing possible danger due to overflow of water in the country. This study widened the scope of TL-moments to distributions that have not been considered before which were the generalized logistic (GLO), extreme value type I (EV), generalized extreme value type I (GEV) and generalized Pareto (GPA) distributions and used all these TL-moments in estimating the probability distribution of rainfalls data. The comparison of the TL-moment method and the L-moment method is also useful in helping statisticians and mathematicians to determine the most suitable method for different situations.

1.7 Chapters' Overview

Chapter 1 highlights introductions to flood occurences in Malaysia, flood frequency analysis and the TL-moment method. It also covers the objectives, scope and significance of the study.

Chapter 2 explains on frequency analysis, parameter estimations and selection of distributions. It also includes a brief history and introductions on the statistical distributions considered in this study which consists of normal, logistic, generalized logistic, extreme value, generalized extreme value and the generalized Pareto distribution.

Chapter 3 covers the probability density functions, distribution functions, quantile functions, L-moments, L-moment ratios and parameter estimates using the L-moment method for each statistical distributions that are being considered which are normal (N), logistic (LOG), generalized logistic (GLO), extreme value type I (EV), generalized extreme value type I (GEV) and generalized Pareto (GPA) distribution. This chapter also includes the derivation of the TL-moments and TL-moment ratios in the case of $t = 1$ and $t = 2$ for all the six distributions. In addition, it also covers the goodness of fit test.

Chapter 4 includes a brief overview of Selangor and Kuala Lumpur, the data collection process, the name and information on all the 55 stations considered and the descriptive statistics of the data. It also presents the L-moments, L-moment ratios, TL-moments and TL-moment ratios for both $t = 1$ and $t = 2$ cases.

Chapter 5 discusses the analysis of the data using the TL-moments and L-moment methods and their ratios that had been given in Chapter 3. It also covers the results of the data using the mean absolute deviation index (MADI), mean square deviation index (MSDI) and correlation, *r*.

Chapter 6 presents the conclusions made from the analyzed data. Recommendations for future research are also given in this section.