PARAMETER ESTIMATION FOR GENERALIZED EXTREME VALUE DISTRIBUTION OF EXTREME RAINFALL IN JOHOR

NURHAZIMAH BINTI NAZMI

UNIVERSITI TEKNOLOGI MALAYSIA
PARAMETER ESTIMATION FOR GENERALIZED EXTREME VALUE DISTRIBUTION OF EXTREME RAINFALL IN JOHOR

NURHAZIMAH BINTI NAZMI

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

Faculty of Science
UniversitiTeknologi Malaysia

JANUARY 2014
To my beloved father, mother, brother and sisters
thank you for all your love and support.
(Hj. Nazmi Bin Hj. Othman and Hjh. Nooraini Bt Abdullah)
ACKNOWLEDGEMENT

Alhamdulillah, my gratefulness is to ALLAH S.W.T for His permission that I had successfully completed this academic project in the time given.

I was in contact with many people in preparing this thesis. They have contributed towards my understanding and thoughts. Firstly, I wish to express my sincere appreciation to my supervisor, Associate Professor Dr. Fadhilah Yusof for her encouragement, guidance, critics and advices. She gives me a lot of ideas to tackle the problem that I need to solve.

I wish to express my gratitude and my appreciation to my fellow friends who always support me and give motivation to complete this project especially Lee Mee Yung, Azwaniza Poharan and Heinzah Melati Jamri. I also like to thank Ibrahim Ali, a researcher, who helped me solve the programming in this project.

Amillion thank you to my parents who never stop praying for my success and lend me their shoulders. Lastly, thank you to my sisters, Noorazni Nazmi, Noorizan Nazmi and brother, Nasiruddin Nazmi. Without their continued support and interest, this thesis would not have been the same as presented here. Thank you to all and May God bless us.
ABSTRACT

Analysis of extreme rainfall events is essential since the results beneficial for civil engineers and planners to estimate the ability of building structures to survive under the utmost extreme conditions. The annual maximum series (AMS) data of daily rainfall in Johor were fitted to Generalized Extreme Value (GEV) distribution. A study was conducted to determine the best method to estimate parameters of GEV distribution by using method of moments (MOM), maximum likelihood estimators (MLE) and Bayesian Markov Chain Monte Carlo (MCMC) simulations. Previous study show that parameter estimation using MOM was a better method than MLE since MLE method had problem in handling small sample. However, some researchers identified that the performance of parameter estimation of MLE method can be improved when adopting Bayesian MCMC. The performances of parameter estimations by using MOM, MLE and Bayesian MCMC were compared by conducting Relative Root Mean Square Error (RRMSE) and Relative Absolute Square Error (RASE). The results showed that Bayesian MCMC method was better than MOM and MLE method in estimating GEV parameters. Bayesian also had the maximum value return level for 10, 25, 50 and 100 years for most of stations. Therefore, it can be concluded that Bayesian MCMC is the best method to estimate the distribution parameters of extreme daily rainfall amount in Johor.
ABSTRAK

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SUPERVISOR AUTHENTICATION</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TITLE</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Background Study</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Statement of Problem</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Objectives of Study</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Scope of study</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>Significance of Study</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>10</td>
</tr>
</tbody>
</table>
2.2 Data 10
2.3 Overview on Rainfall Modeling 11
2.4 Parameter Estimation Methods
  2.4.1 Method of Moments 14
  2.4.2 Maximum Likelihood Estimator 15
  2.4.3 Bayesian Approach 16
2.5 Goodness-of-fit test 18
2.6 Summary 19

3 METHODOLOGY
3.1 Introduction 20
3.2 Data Sources 21
3.3 Descriptive Statistics
  3.3.1 Mean 26
  3.3.2 Standard Deviation 26
  3.3.3 Coefficient of Variation 27
  3.3.4 Kurtosis 27
  3.3.5 Skewness 28
3.4 Probability Distribution 29
3.5 Parameter Estimation
  3.5.1 Method of Moments 30
  3.5.2 Maximum Likelihood Estimator 34
  3.5.3 Bayesian approach 36
3.6 MCMC Simulation 38
  3.6.1 Gibbs Sampler 40
  3.6.2 Metropolis-Hastings Sampling 41
3.7 Goodness-of-fit test 43
3.8 Research Framework 45
3.9 Conclusion 46
4 RESULTS

4.1 Introduction 47
4.2 Descriptive Statistics 47
4.3 Analysis of the parameter estimation of GEV Distribution
4.3.1 Method of Moments 50
4.3.2 Maximum Likelihood Estimators 53
4.3.3 Bayesian MCMC 55
4.4 Analysis of the best method for describing the extreme rainfall events in Johor 61
4.5 Return level 64

5 CONCLUSION AND RECOMMENDATION

5.1 Introduction 71
5.2 Conclusion 72
5.3 Recommendations 73

REFERENCES 74
<table>
<thead>
<tr>
<th>TABLE</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The coordinates of 12 rain gauge station in Johor</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>Annual Maximum Series of 12 rain gauge station in Johor</td>
<td>22</td>
</tr>
<tr>
<td>4.1</td>
<td>Descriptive statistics for daily rainfall amount in Johor</td>
<td>47</td>
</tr>
<tr>
<td>4.2</td>
<td>Parameters estimation by using MOM</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Parameters estimation by using MLE</td>
<td>52</td>
</tr>
<tr>
<td>4.4</td>
<td>Parameters estimation by using MCMC Simulation</td>
<td>57</td>
</tr>
<tr>
<td>4.5</td>
<td>Relative Root Mean Squared Error (RRMSE) for each station of MOM, MLE and Bayesian techniques</td>
<td>60</td>
</tr>
<tr>
<td>4.6</td>
<td>Relative Absolute Square Error (RASE) for each station of MOM, MLE and Bayesian techniques</td>
<td>61</td>
</tr>
<tr>
<td>4.7</td>
<td>Estimates for 10 year return levels (mm) at each station</td>
<td>64</td>
</tr>
<tr>
<td>4.8</td>
<td>Estimates for 25 year return levels (mm) at each station</td>
<td>65</td>
</tr>
<tr>
<td>4.9</td>
<td>Estimates for 50 year return levels (mm) at each station</td>
<td>66</td>
</tr>
<tr>
<td>4.10</td>
<td>Estimates for 100 year return levels (mm) at each station</td>
<td>67</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Locations of 12 rain gauge stations in Johor</td>
<td>21</td>
</tr>
<tr>
<td>3.2</td>
<td>Time series plot of annual maximum daily rainfall in Johor</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>Probability density function of rainfall data in station S03</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>The Research Framework of this study</td>
<td>44</td>
</tr>
<tr>
<td>4.1</td>
<td>The values of (a) location, (b) shape and (c) scale parameters for GEV distribution for each station using MOM</td>
<td>50</td>
</tr>
<tr>
<td>4.2</td>
<td>The values of (a) location, (b) shape and (c) scale Parameters for GEV distribution for each station using MLE</td>
<td>53</td>
</tr>
<tr>
<td>4.3</td>
<td>Graph plot (a) location, (b) scale and (c) shape parameters of GEV parameters for 20000 iterations</td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>The values of (a) location, (b) shape and (c) scale parameters for GEV distribution for each station using Bayesian</td>
<td>58</td>
</tr>
</tbody>
</table>
4.5 The return level for 10, 25, 50 and 100 years of each station using MOM, MLE and Bayesian
CHAPTER 1

INTRODUCTION

1.1 Introduction

Malaysia’s climate is hot and humid throughout the year since it is situated near the equator. According to the Malaysian Meteorological Department, Malaysia will experience a rainy season almost every year neither drizzle nor heavy rainfall. In particular, Malaysia deals with two distinct monsoons that are northeast monsoon which blows a rainy season from October to March and Southwest monsoon which blows a dry season from June to September. Meanwhile, the two shorter periods of inter-monsoon seasons is marked by heavy rainfall as it yields uniform periodic changes in the wind flow patterns over the country.

Rainfall is the amount of precipitation that has fallen within a specific period of time. The rainfall processes had always been determined which include the estimation of the rainfall distribution and the identification of wet or dry events on a particular day. The process and structures require to be built to resist the extreme behaviour. For instance, a reservoir should be capable of storing the amount of rainfall expected to fall in the region of interest. If the reservoir is not large enough, there is a risk that the water may overtop the dam. The rainfall to be stored in the reservoir may have come from a large region and may also build up over a number of days. Even though hot or raining season is determined by the flow of the monsoon wind, the changes in world’s climate that happen in the past several years have caused the season to be unpredictable all year round.
As Malaysia is located in a tropical climate zone, extreme rainfall is estimated to take place regularly every year resulting from local tropical wet season. Extreme rainfall and extremely dry long spell of dry days (drought) were the most disruptive of atmospheric phenomena. Extreme rainfall event is often related with the changes of climate, which may be followed by a series of natural disasters like flash flood and landslides. The rapid changes in the climate have consistently increased the number of extreme floods in Malaysia, especially in the highly populated areas. Unpredictable extreme rainfall events phenomenon that increase in frequency lately has brought damages costing millions of Malaysian ringgit.

Extreme rainfalls data need to be modelled by suitable statistical distributions which give the best inferences of the patterns of extreme rainfall. There are many distributions that were able to be used in extreme rainfall analysis such as Generalized Extreme Value (GEV) distribution (Kim et. al, 2013), Generalized Pareto (GP) distribution (Coles et. al, 2003), and Lognormal distribution (Adetan and Afullo, 2012). The studies that have been conducted in Malaysia commonly used GP and Mixed exponential for representing the daily rainfall amount. Rainfall distributions occurred in Malaysia is categorized as extreme events since Malaysia is located in tropical climate zone. Therefore, for this study GEV distribution will be used to fit the extreme daily rainfall in Johor.

Heavy rainfall and the associated floods occur frequently in Malaysia and have caused huge economic losses as well as social problems. Typhoons sometimes occur in Malaysia from July to mid-November and contribute to heavy damage, flooding and erosion. The high loss in economic especially in agriculture sectors, has significant impacts to the country development and that inspires the researchers to investigate and model the rainfall process to further understand the changes and the characteristics of rainfall in the past. This is important to be highlighted as it will not only be used in planning the water resources, but it can also be used to improve the sensitivity of the rainfall systems and managements in a country.
The occurrence of the two most extreme El Niño’s in year 1982/1983 and year 1997/1998 were reported to be the factor of extreme climate changes in Peninsular Malaysia (Zin et. al, 2010). Based on United Nations Framework on Climate Change (UNFCCC), water resources, agriculture and food security, ecosystems and biodiversity, human health and coastal zones will be affected by the changes of climate in Asia. Zin et. al (2010) added that the impact of extreme rainfall events may be decreased by preventive measures based on the results from statistical analysis of extreme rainfall data. Therefore, the inferences on the extremes of environmental processes are essential for design-specification in civil engineering. Modeling the extremes of rainfall data can also be used to assess the risk of urban flooding and this information is often used for insurance purposes.

Rainfall data for Southern Peninsular Malaysia, Johor had been analyzed for the modelling of daily rainfall in this study. Johor experiences a tropical rainforest climate with monsoon rain from November until February which blows from the South China Sea. The mean annual rainfall is 1778 mm with average temperatures ranging between 25.5°C and 27.8°C. Furthermore, continuous heavy downpour had occurred in Johor on December 2006, which led to the 2006/2007 Malaysia floods. Seriously flooded with water level as high as 10 feet (3.0m) above the ground level had occurred in Muar, Kota Tinggi and Segamat.
1.2 Background of the Study

Floods are the natural hazards than can occur in many region of the world including Malaysia. The river flooding is the incidence of heavy rainfalls (monsoon season or convective) and the resultant of large concentration of runoff, which exceeds river capacity (Ministry of Natural Resources and Environment Malaysia, 2007). Over flown floods or inundation that come from river or other body of water can affect or threaten damage based on US Geological Survey.

In Malaysia, a continuous heavy downpour on December 2006 was the most tragic event that had occurred in Johor where it led to the 2006-2007 Malaysia floods. The incidence of heavy rainfall assisted with waste disposal into the river had caused the heavy downpour. The impact of these disasters is a decrease in quality of life and economic growth in the country has disrupted, many properties are damaged and occasionally it causes a loss of human lives. Without knowledge on the frequency of occurrence of rainfall and magnitude of flooding, this can cause an extreme flood event.

Recently, the increase of global temperatures has led to regional increases in the amount and intensity of rainfall. Furthermore, many activities such as deforestation caused by logging, clearance of land for agriculture, housing and industrial purposes had corrupted the ecosystem. Hence, the analysis of rainfall behaviour, particularly in terms of rainfall occurrence amount beneficial of managing the consumption of water. Moreover, this information encourages the experts to give a reliable prediction about extreme weather events.

In this context, this study provides an insight to the potential change in the rainfall extreme and extreme dry spell for the past 34 years. Statistical modelling of extreme rainfall is important since the results can help civil engineers and planners to estimate the ability of building structures to survive under the utmost extreme conditions. In addition, it is beneficial for engineer’s design of drainage, bridge,
retaining wall and dam systems based on the expected rain amount expected during certain time period.

Ailliot et. al (2008) had mentioned that for many analysed, the data that available is often limited on extremes. Therefore, the fitting an extreme value distribution can lead to difficulties, particularly for the estimation of extreme quantiles. In addition, for some cases there is a need to better understand the asymptotic and small sample properties of the parameters of the extreme value distribution chosen.

In order to estimate the parameter for GEV distribution, research conducted in Malaysia has focused on Method of Moments (MOM), Maximum Likelihood Estimator (MLE) and Bayesian techniques. MOM has fairly simple and yield consistent estimators which under very weak assumption but are often biased. Meanwhile, MLE has higher probability of being close to the quantities and more often unbiased. Bayesian analysis provides a method in which observations are used to update estimates unknown parameters.

Parameter estimation via moment-based techniques may produce better estimation than maximum likelihood estimator since MLE contributes to large deviation error because MLE method incapable to obtain the parameter estimation method for small sample (Adetan and Afullo, 2012). However, Bayesian approach has no advantage over the MLE from the results of the parameter estimation because MLE is all-around utility and adaptability to model change (Coles, 2001).

MLE is dependent on regularity assumptions required by asymptotic theory. Asymptotic properties associated with MLE estimator are violated when estimating parameters for GEV distribution due to the assumption of a restricted parameter space. Afterward, an alternative way to estimate the parameters of GEV distribution is by using Bayesian approach. Bayesian analysis offers a way of dealing with information conceptually different from all other statistical methods. Eli et. al (2012)
reported that even though extreme data are limited in nature, Bayesian inferences have the ability to incorporate other source of information via prior distribution.

1.3 Statement of Problem

The parameter estimation by using Method of Moments (MOM), the estimators performed well when the sample sizes are modest. However, MOM is the oldest method of deriving point estimators and not be the best estimators since MOM always produces some asymptotically unbiased estimators.

Maximum likelihood estimators (MLE) is emerged as a flexible and powerful modeling tool in such applications, but its performance with small samples has been shown to be poor relative to an alternative fitting procedure based on probability weighted moments. Even though MLE method had many advantages, its poor performance when dealing with small samples had been improved by conducting Bayesian MCMC.

In many areas of applications, Bayesian approached is quite famous. A challenge when adopting this approach is the computational difficulties which can be solved by the application of Markov Chain Monte Carlo as mentioned by Coles (2001). Markov chain Monte Carlo (MCMC) concepts and techniques and shown how to apply them to the estimation of a Bayesian hierarchical model of interdependent extreme operational risks.

Bayesian analysis provides a more convenient and direct way of managing and expressing uncertainties instead of give a radically different interpretation of the data as declared by Coles et. al (2003) on the study of fully probabilistic approach to extreme rainfall modeling. Besides that, to reduce the amount of uncertainties in the model, Bayesian approach has the ability to embrace other source of information and to derive the prediction of future return levels of extreme rainfall.
The trends of rainfall in Malaysia are different where it depends on the location, area and surrounding factors. Flash floods, droughts and landslides had happened frequently at certain regions in Malaysia. Smith (2005) had explained that the inference on the extremes of environmental processes is essential for design-specification in civil engineering.

Therefore, the prediction of future rainfall distribution for different climate changes scenarios is needed to provide information for high quality climate-related studies. The probability distribution for modelling of regional data is one of the concerns of hydrologists and engineers. Information related to distributions of rainfall amounts is beneficial for designing water related structures.

1.4 Objectives of Study

1. To estimate and compare the parameters of Generalized Extreme Value (GEV) distribution using Bayesian Markov Chain Monte Carle (MCMC), Method of Moments (MOM) and Maximum Likelihood Estimation (MLE).
2. To determine the best-fit parameters values of GEV distribution using Relative Mean Square Error (RRMSE) and Relative Absolute Square Error (RASE).
3. To predict the values of return levels for the next 10, 25, 50 and 100 years based on 30-year history.
1.5 Scope of Study

Extreme rainfall data need to be modelled by a suitable statistical distribution. For this study, GEV distribution is the most suitable as it gives the best inferences of the behaviour of extreme rainfall events (Zalina et al., 2002 and Ibrahim (2004)). The daily rainfall data of 12 rain gauge stations in Johor which have records for 34 years from year 1975 to year 2008 is analyzed. The data of annual maximum flow is selected each year in Johor.

Method of Moments, Maximum Likelihood Estimation and Bayesian approaches are applied in order to estimate the parameters of GEV distribution. MCMC simulation is used to estimate parameters for the purpose of facilitating Bayesian calculation as approved by Eli et al. (2012). Stimulation of MCMC is supported by R statistical software as suggested by Gamerman and Lopes (2006). To find the best model, Relative Root Mean Square Error (RRMSE) and Relative Absolute Square Error (RASE) are carried out. Lastly, by using the estimated parameters, the predicted values of return levels for the next 10, 25, 50 and 100 years would be calculated based on 30-year history annual maximum rainfall.
1.6 Significance of Study

It is known that the analysis of rainfall occurrence is beneficial for managing the consumption of water, to estimate the ability of building structures and also could help to predict the extreme weather events. Information about rainfall can be used to assess danger of heavy downpour flood and landslides.

Furthermore, civil engineer can predict the material that should be used to build bridges, design of drainage, retaining wall and dam systems referred to rainfall distributions received during certain time duration. Moreover, it may help our country from unnecessary costs and economic losses as well as avoiding danger due to overflow of water in the country.

The predicted values of return levels had been calculated by using parameter estimation of MOM, MLE and Bayesian approach for the next 10, 25, 50 and 100 years which were calculated based on 30-year history annual maximum rainfall. Thus, the results can facilitate civil engineers and planners to estimate the ability of building structures which can to survive under utmost extreme events Eli et al (2012).
REFERENCES

Adetan, O., and Afullo, T. J. (2012). Comparison of Two Methods to Evaluate the Lognormal Raindrop Size Distribution Model in Durban.


