FLOOD FREQUENCY ANALYSIS USING PL-MOMENTS APPROACH

ZAHRAHTUL AMANI BINTI ZAKARIA

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

MAY 2013
To my beloved parents

My father, Zakaria bin Omar

My mother, Allahyarhamah Ramlah binti Udin

To my lovely brothers and sisters

Salahudin bin Zakaria
Muhammad Saifuldin bin Zakaria
Mohd Shaharuddin bin Zakaria
Zahrahtul Hazwani binti Zakaria
Norfadillah binti Hassan Basuri
Siti Ariffah binti Mustajap

Alhamdulillah, I thank God for blessing me with all your presence.
ACKNOWLEDGEMENT

I would like to express my sincere thanks and gratitude to Almighty Allah for his blessings. Now, as my journey in the postgraduate school comes to an end, I am thankful to so many people.

Dr. Ani bin Shabri, for helping me sail my boat gently down the stream of statistics in hydrology application, and patiently guiding me with his bird’s eye view of the field. He has been the wind behind my sail. Dr. Ruhaidah binti Shamsuddin, who helped me navigate my boat with her advices and motivation. I certainly would not have made it safe to the shore without their constant support and wise guidance.

I am also indebted to Universiti Sultan Zainal Abidin and Ministry of Higher Education, Malaysia for fueling my ship and supporting my visions. Special thanks to Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia and Universiti Teknologi Malaysia for being my primary sources of relevant information for my trip.

Two special peoples in my life have always been the wind behind my sail and never left me alone in the ocean of my exploration. My late mother is my guardian angel. She passed away during the end course of my sail, but her spirits always live in my soul. My father never ceased believing in me. This is the moment of validation of all their dreams and hopes.

My beloved brothers and sisters, who continuously supporting my sail so that it continues its course down the stream despite the occurring of floods. I am deeply grateful to them. All my fellow friends who helped me in accomplishing my journey, thank you so much. From the bottom of my heart I express my gratitude to everyone involved. May Allah repay all the kindness that you have given thus far.

After all “success is a journey, not a destination”.
ABSTRACT

Estimation of flood magnitude is a crucial component in planning, designing, and managing of water resources projects. Flood frequency analysis (FFA) provides a practical means of determining a robust probability distribution that fits streamflow data at a location of interest. The main focus in hydrology design is the estimation of high flow quantile. L-moments, popular among hydrologist in FFA is known to be oversensitive towards the lower part of the distribution and give insufficient weight to large sample values. As an alternative, the method of partial L-moments (PL-moments) is proposed to give weightage to the upper part of distribution and large values in censored sample. The aim of this study is to compare the performance of PL-moments and L-moments in FFA. The method of PL-moments was developed for generalized extreme value (GEV), generalized logistic (GLO), generalized pareto (GPA), extreme value type 1 (EV1) and logistic (LOG) distributions. Monte Carlo simulations from population distributions of known and unknown samples were conducted to assess the performance of PL-moments compared to L-moments. Simulation results showed that PL-moments give comparable and slightly better parameter estimates than those by L-moments particularly when estimating the high flow quantiles. In regional flood frequency analysis, new statistical tests based on PL-moments were developed to measure discordancy, regional homogeneity and identify a best regional distribution. The quantile estimates based on the regional distribution using PL-moments are more efficient than L-moments in estimating flood quantiles at higher return periods. The overall results strongly support that PL-moments method would improve the flood quantiles estimation particularly for higher quantiles and thus serves as a useful tool for application in flood frequency analysis.
ABSTRAK

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</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>xi</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xv</td>
</tr>
<tr>
<td></td>
<td>LIST OF SYMBOLS</td>
<td>xx</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Research Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Problems Statement</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Research Objectives</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Research Scope</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Research Contribution</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>Organization of Thesis</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
<td>12</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>Flood Frequency Analysis</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Parameter Estimator</td>
<td>14</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Simulation of the Estimator</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>Censored Data in Hydrology</td>
<td>21</td>
</tr>
</tbody>
</table>
2.4 Regional Flood Frequency Analysis
   2.4.1 Regional Homogeneity Measure
   2.4.2 Selection of Regional Distribution

### 3 CONCEPTS OF L-MOMENTS AND PL-MOMENTS

#### 3.1 Introduction

#### 3.2 Background of L-Moments
   3.2.1 L-Moments of Distributions
   3.2.2 Sample Estimates of L-Moments

#### 3.3 Background of PL-Moments
   3.3.1 PL-Moments of Distributions
   3.3.2 Sample Estimates of PL-Moments

#### 3.4 Ordered Statistics

#### 3.5 Complete and Censored Samples

#### 3.6 Flowchart of Research Methodology

### 4 L-MOMENTS AND PL-MOMENTS METHODS FOR PROBABILITY DISTRIBUTION FUNCTION

#### 4.1 Introduction

#### 4.2 Probability Distribution Function

#### 4.3 Generalized Extreme Value Distribution
   4.3.1 Parameter Estimation of L-Moments
   4.3.2 Parameter Estimation of PL-Moments

#### 4.4 Generalized Logistic Distribution
   4.4.1 Parameter Estimation of L-Moments
   4.4.2 Parameter Estimation of PL-Moments

#### 4.5 Generalized Pareto Distribution
   4.5.1 Parameter Estimation of L-Moments
   4.5.2 Parameter Estimation of PL-Moments

#### 4.6 Extreme Value Type I Distribution
   4.6.1 Parameter Estimation of L-Moments
   4.6.2 Parameter Estimation of PL-Moments

#### 4.7 Logistic Distribution
4.7.1 Parameter Estimation of L-Moments 69
4.7.2 Parameter Estimation of PL-Moments 71
4.8 Simplified Parameter Estimation of L-Moments and PL-Moments 72

5 SIMULATIONS STUDY 76
5.1 Introduction 76
5.2 Monte Carlo Simulation for Known Parent Distribution Function 78
5.3 Monte Carlo Simulation for Unknown Parent Distribution Function 80
5.4 Analysis of Monte Carlo Simulation Study 82
  5.4.1 Known Parent Distribution Function 82
  5.4.2 Conclusion 93
  5.4.3 Unknown Parent Distribution Function 94
  5.4.4 Conclusion 104

6 REGIONAL FLOOD FREQUENCY ANALYSIS 106
6.1 Introduction 106
6.2 Screening of the Data 107
6.3 Identification of Homogenous Region 109
  6.3.1 Kappa Distribution 113
    6.3.1.1 L-Moments Kappa Distribution 114
    6.3.1.2 PL-Moments Kappa Distribution 115
  6.3.2 Modification of Region 116
6.4 Selection of a Frequency Distribution 117
  6.4.1 Ratio Diagram 117
  6.4.2 Goodness-of-fit Measure, Z-test 119
6.5 Estimation of the Regional Distribution 122
6.6 Analysis Regional Flood Frequency Analysis 124
6.7 Discordancy Measure 127
6.8 Identification of Homogenous Region 130
  6.8.1 The L-Moments Method 130
6.8.2 The PL-Moments Method 134
6.9 Selection of a Frequency Distribution 136
  6.9.1 L-Moment Ratio Diagram 136
  6.9.2 PL-Moment Ratio Diagram 138
6.10 Selection Based on Z-test 139
  6.10.1 The L-Moments Method 139
  6.10.2 The PL-Moments Method 140
6.11 Estimation of the Regional Distribution 141
  6.11.1 Parameter Estimation of the Regional Distribution 141
  6.11.2 Simulation of Flood Quantiles 143
6.12 Conclusion 161

7 CONCLUSION 162
  7.1 Conclusion 162
  7.2 Recommendations 165

REFERENCES 167
Appendices A - G 175 - 209
### LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Parameter estimation using L-moments method</td>
<td>73</td>
</tr>
<tr>
<td>4.2</td>
<td>Parameter estimation using PL-moments method</td>
<td>74</td>
</tr>
<tr>
<td>5.1</td>
<td>Statistical characteristics of six Wakeby distributions</td>
<td>82</td>
</tr>
<tr>
<td>5.2</td>
<td>Performances of PL-moments compared to L-moments for simulation of known parent distribution based on RRMSE and MAE</td>
<td>92</td>
</tr>
<tr>
<td>5.3</td>
<td>Performances of PL-moments compared to L-moments for simulation of unknown parent distribution based on Efficiency and MAE</td>
<td>104</td>
</tr>
<tr>
<td>6.1</td>
<td>Critical values for the $D$-statistic</td>
<td>108</td>
</tr>
<tr>
<td>6.2</td>
<td>Polynomial approximations of $\tau_4$ as a function of $\tau_3$ based on L-moments method</td>
<td>118</td>
</tr>
<tr>
<td>6.3</td>
<td>Polynomial approximations of $\tau_4$ as a function of $\tau_3$ based on PL-moments method</td>
<td>118</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparisons of discordancy measure, homogeneity measure and selection of distribution using L-moments and PL-moments methods</td>
<td>121</td>
</tr>
</tbody>
</table>
6.5 The sites and statistics of annual maximum daily streamflow for study area in Peninsular Malaysia

6.6 L-moment ratios and $D$-statistic values for L-moments

6.7 PL-moment ratios and $D$-statistic values for PL-moments

6.8 Maximum values of $D$-statistic and $H$-test for each region based on L-moments

6.9 Maximum values of $D$-statistic and $H$-test for each region after region modifications of step (iii) based on L-moments

6.10 $H$-test values for each region after region modifications of step (i) and (ii) based on L-moments

6.11 Maximum values of $D$-statistic and $H$-test for each region based on PL-moments

6.12 Maximum values of $D$-statistic and $H$-test for each region after region modifications of step (iii) based on PL-moments

6.13 $H$-test values for each region after region modifications of step (i) and (ii) based on PL-moments

6.14 Regional average L-moment ratios of homogeneous regions

6.15 Regional average PL-moment ratios of homogeneous regions

6.16 Goodness-of-fit test, Z-test based on L-moments method

6.17 Goodness-of-fit test, Z-test based on PL-moments method

6.18 Regional parameters and quantile estimates of the GEV, GLO, GPA, EV1 and LOG distributions for L-moments and PL-moments
6.19  RBIAS values for different quantiles for L-moments and PL-moments for R1 (East Coast)  144

6.20  RBIAS values for different quantiles for L-moments and PL-moments for R2 (Southern)  145

6.21  RBIAS values for different quantiles for L-moments and PL-moments for R3 (Northern)  146

6.22  RBIAS values for different quantiles for L-moments and PL-moments for R4 (West Coast I)  147

6.23  RBIAS values for different quantiles for L-moments and PL-moments for R5 (West Coast II)  148

6.24  RRMSE values for different quantiles for L-moments and PL-moments for R1 (East Coast)  150

6.25  RRMSE values for different quantiles for L-moments and PL-moments for R2 (Southern)  151

6.26  RRMSE values for different quantiles for L-moments and PL-moments for R3 (Northern)  152

6.27  RRMSE values for different quantiles for L-moments and PL-moments for R4 (West Coast I)  153

6.28  RRMSE values for different quantiles for L-moments and PL-moments for R5 (West Coast II)  154
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Flowchart of research methodology</td>
<td>45</td>
</tr>
<tr>
<td>5.1</td>
<td>RBIAS for GEV quantile estimator $x(F = 0.980)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>83</td>
</tr>
<tr>
<td>5.2</td>
<td>RBIAS for GEV quantile estimator $x(F = 0.995)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>83</td>
</tr>
<tr>
<td>5.3</td>
<td>RBIAS for GLO quantile estimator $x(F = 0.980)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>84</td>
</tr>
<tr>
<td>5.4</td>
<td>RBIAS for GLO quantile estimator $x(F = 0.995)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>84</td>
</tr>
<tr>
<td>5.5</td>
<td>RBIAS for GPA quantile estimator $x(F = 0.980)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>84</td>
</tr>
<tr>
<td>5.6</td>
<td>RBIAS for GPA quantile estimator $x(F = 0.995)$ plotted against censoring level $(F_0)$ for GEV shape parameter, $k = -0.2$ and $k = +0.2$</td>
<td>85</td>
</tr>
</tbody>
</table>
5.7 RBIAS for EV1 quantile estimator \(x(F = 0.980)\) and 
\(x(F = 0.995)\) plotted against censoring level \((F_a)\) for 
different sample size \((n)\)

5.8 RBIAS for LOG quantile estimator \(x(F = 0.980)\) and 
\(x(F = 0.995)\) plotted against censoring level \((F_a)\) for 
different sample size \((n)\)

5.9 RRMSE for GEV quantile estimator \(x(F = 0.980)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.10 RRMSE for GEV quantile estimator \(x(F = 0.995)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.11 RRMSE for GLO quantile estimator \(x(F = 0.980)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.12 RRMSE for GLO quantile estimator \(x(F = 0.995)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.13 RRMSE for GPA quantile estimator \(x(F = 0.980)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.14 RRMSE for GPA quantile estimator \(x(F = 0.995)\) plotted 
against GEV shape parameter \((k)\) for sample size of \(n = 15\) 
and \(n = 50\)

5.15 RRMSE for EV1 quantile estimator \(x(F = 0.980)\) and 
\(x(F = 0.995)\) plotted against sample size \((n)\) for different 
censoring level \((F_a)\)

5.16 RRMSE for LOG quantile estimator \(x(F = 0.980)\) and 
\(x(F = 0.995)\) plotted against sample size \((n)\) for different 
censoring level \((F_a)\)
5.17 RBIAS of quantile estimator $x(F = 0.980)$ using L-moments and PL-moments, fitting GEV distribution to generated Wakeby samples

5.18 RBIAS of quantile estimator $x(F = 0.995)$ using L-moments and PL-moments, fitting GEV distribution to generated Wakeby samples

5.19 RBIAS of quantile estimator $x(F = 0.980)$ using L-moments and PL-moments, fitting GLO distribution to generated Wakeby samples

5.20 RBIAS of quantile estimator $x(F = 0.995)$ using L-moments and PL-moments, fitting GLO distribution to generated Wakeby samples

5.21 RBIAS of quantile estimator $x(F = 0.980)$ using L-moments and PL-moments, fitting GPA distribution to generated Wakeby samples

5.22 RBIAS of quantile estimator $x(F = 0.995)$ using L-moments and PL-moments, fitting GPA distribution to generated Wakeby samples

5.23 RBIAS of quantile estimator $x(F = 0.980)$ using L-moments and PL-moments, fitting EV1 distribution to generated Wakeby samples

5.24 RBIAS of quantile estimator $x(F = 0.995)$ using L-moments and PL-moments, fitting EV1 distribution to generated Wakeby samples

5.25 RBIAS of quantile estimator $x(F = 0.980)$ using L-moments and PL-moments, fitting LOG distribution to generated Wakeby samples

5.26 RBIAS of quantile estimator $x(F = 0.995)$ using L-moments and PL-moments, fitting LOG distribution to generated Wakeby samples
5.27 Efficiency of quantile estimator \( x(F = 0.980) \) using L-moments and PL-moments, fitting GEV distribution to generated Wakeby samples

5.28 Efficiency of quantile estimator \( x(F = 0.995) \) using L-moments and PL-moments, fitting GEV distribution to generated Wakeby samples

5.29 Efficiency of quantile estimator \( x(F = 0.980) \) using L-moments and PL-moments, fitting GLO distribution to generated Wakeby samples

5.30 Efficiency of quantile estimator \( x(F = 0.995) \) using L-moments and PL-moments, fitting GLO distribution to generated Wakeby samples

5.31 Efficiency of quantile estimator \( x(F = 0.980) \) using L-moments and PL-moments, fitting GPA distribution to generated Wakeby samples

5.32 Efficiency of quantile estimator \( x(F = 0.995) \) using L-moments and PL-moments, fitting GPA distribution to generated Wakeby samples

5.33 Efficiency of quantile estimator \( x(F = 0.980) \) using L-moments and PL-moments, fitting EV1 distribution to generated Wakeby samples

5.34 Efficiency of quantile estimator \( x(F = 0.995) \) using L-moments and PL-moments, fitting EV1 distribution to generated Wakeby samples

5.35 Efficiency of quantile estimator \( x(F = 0.980) \) using L-moments and PL-moments, fitting LOG distribution to generated Wakeby samples

5.36 Efficiency of quantile estimator \( x(F = 0.995) \) using L-moments and PL-moments, fitting LOG distribution to generated Wakeby samples
6.1 Location of streamflow stations located throughout Peninsular Malaysia

6.2 L-moment ratio diagram

6.3 PL-moment ratio diagram

6.4 Results of the RRMSE for sample size, \( n = 30 \) computed for different quantiles for L-moments and PL-moments for R1 (East Coast)

6.5 Results of the RRMSE for sample size, \( n = 30 \) computed for different quantiles for L-moments and PL-moments for R2 (Southern)

6.6 Results of the RRMSE for sample size, \( n = 30 \) computed for different quantiles for L-moments and PL-moments for R3 (Northern)

6.7 Results of the RRMSE for sample size, \( n = 30 \) computed for different quantiles for L-moments and PL-moments for R4 (West Coast I)

6.8 Results of the RRMSE for sample size, \( n = 30 \) computed for different quantiles for L-moments and PL-moments for R5 (West Coast II)
LIST OF SYMBOLS

\( D_i \) - Discordancy measure
\( E[X] \) - Expectation of order statistic
\( F \) - Non-exceedance Probability
\( f_i \) - Plotting position
\( F_0 \) - Level of censoring
\( F(x) \) - Cumulative distribution function
\( f(x) \) - Probability distribution function
\( H_i \) - Heterogeneity measure
\( N \) - Number of sites in a region
\( n_i \) - Record length of a site
\( Q_T \) - Magnitude of flood at \( T \)-year
\( T \) - Return period, year
\( x \) - Random variable
\( x_0 \) - Censored data point
\( x(F) \) - Quantile function
\( Z_{dist} \) - Goodness-of-fit measure, \( Z \)-test
\( \mu \) - Population mean
\( \sigma \) - Population standard deviation
\( \nu \) - Population coefficient of variation
\( \delta \) - Population coefficient of skewness
\( \kappa \) - Population coefficient of kurtosis
\( \alpha \) - Scale parameter
\( \xi \) - Location parameter
$k, h$ - Shape parameter

$\alpha^R$ - Regional scale parameter

$\xi^R$ - Regional location parameter

$k^R$ - Regional shape parameter

$\mu_r$ - $r^{th}$ moments

$\beta_r$ - $r^{th}$ probability weighted moments

$\beta_r'$ - $r^{th}$ partial probability weighted moments

$m_r$ - $r^{th}$ sample moments

$b_r$ - $r^{th}$ sample probability weighted moments

$b_r'$ - $r^{th}$ sample partial probability weighted moments

$\lambda_r$ - $r^{th}$ L-moments

$\lambda_r'$ - $r^{th}$ PL-moments

$l_r$ - $r^{th}$ sample L-moments

$l_r'$ - $r^{th}$ sample PL-moments

$C_v$ - Coefficient of variation (CV)

$C_s$ - Coefficient of skewness (CS)

$C_k$ - Coefficient of kurtosis (CK)

$\tau$ - L-coefficient of variation (L-CV)

$\tau_3$ - L-coefficient of skewness (L-CS)

$\tau_4$ - L-coefficient of kurtosis (L-CK)

$\hat{\tau}_r$ - $r^{th}$ sample L-moment ratios

$\nu$ - PL-coefficient of variation (PL-CV)

$\nu_3$ - PL-coefficient of skewness (PL-CS)

$\nu_4$ - PL-coefficient of kurtosis (PL-CK)

$\hat{\nu}_r$ - $r^{th}$ sample PL-moment ratios

$\eta^{(i)}$ - Coefficient of variation for site $i$

$\eta_3^{(i)}$ - Coefficient of skewness for site $i$

$\eta_4^{(i)}$ - Coefficient of kurtosis for site $i$

$\eta_r^R$ - $r^{th}$ regional average L-moment ratios
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>Euler’s constant</td>
</tr>
<tr>
<td>$E(.)$</td>
<td>Exponential Integral function</td>
</tr>
<tr>
<td>$\Gamma(.)$</td>
<td>Gamma function</td>
</tr>
<tr>
<td>$\gamma(.)$</td>
<td>Incomplete Gamma function</td>
</tr>
<tr>
<td>$B_{1-f_s}(.)$</td>
<td>Incomplete Beta function</td>
</tr>
<tr>
<td>GEV</td>
<td>Generalized extreme value distribution</td>
</tr>
<tr>
<td>GLO</td>
<td>Generalized logistic distribution</td>
</tr>
<tr>
<td>GPA</td>
<td>Generalized Pareto distribution</td>
</tr>
<tr>
<td>EV1</td>
<td>Extreme value type 1 distribution</td>
</tr>
<tr>
<td>LOG</td>
<td>Logistic distribution</td>
</tr>
<tr>
<td>PWMs</td>
<td>Probability Weighted Moments</td>
</tr>
<tr>
<td>PPWMs</td>
<td>Partial Probability Weighted Moments</td>
</tr>
<tr>
<td>RBIAS</td>
<td>Relative bias</td>
</tr>
<tr>
<td>RRMSE</td>
<td>Relative root mean square error</td>
</tr>
<tr>
<td>RFFA</td>
<td>Regional flood frequency analysis</td>
</tr>
<tr>
<td>WA</td>
<td>Wakeby distribution</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Mathcad programming for Monte Carlo known (GEV distribution) using PL-moments</td>
<td>175</td>
</tr>
<tr>
<td>B</td>
<td>Mathcad programming for Monte Carlo unknown (GEV distribution) using PL-moments</td>
<td>178</td>
</tr>
<tr>
<td>C</td>
<td>RRMSE values for Monte Carlo simulation of known parent distribution</td>
<td>181</td>
</tr>
<tr>
<td>D</td>
<td>MAE values for Monte Carlo simulation of known parent distribution</td>
<td>189</td>
</tr>
<tr>
<td>E</td>
<td>Efficiency values for Monte Carlo simulation of unknown parent distribution</td>
<td>197</td>
</tr>
<tr>
<td>F</td>
<td>MAE values for Monte Carlo simulation of unknown parent distribution</td>
<td>202</td>
</tr>
<tr>
<td>G</td>
<td>Academic contributions</td>
<td>207</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

The main interest of this research is the application of flood frequency analysis based on L-moments and Partial L-moments approaches. Therefore, this chapter basically introduces the background of flood frequency analysis and highlights the problems arising in the analysis. This chapter also includes the objectives, scope, contribution of the research and organization of the thesis.

1.1 Research Background

A flood is an unusually high stage of water levels that generally happened in river that overflows and submerges land and inundates the adjoining area. 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. Flooding can caused damages of private and public properties, loss of life and economic problems. In terms of the number of population affected, frequency, area extent, duration and social economic damage, flooding is the most natural hazard in Malaysia. The flood events occur in various states in Peninsular Malaysia.

Several efforts have been done to overcome the floods including flood protection projects such as construction of barrages, dams, water reservoirs and widening or deepening the rivers. The barrages which act as an embankment have
been constructed along a body of water to prevent water flooding onto the land from the sea or river. Dams and water reservoirs which are generally built to store water can also be used to prevent floods. However, the constructions of barrages and dams require high financial cost from the government while the process of widening and deepening the rivers have high potential in destroying the ecosystems of the river itself. Thus, these flood protection projects without proper planning and designing will only create drawbacks. In order to reduce the drawbacks, information related to these aspects need to be carefully considered. First, how long and high should the barrages be built? Second, how big should the water reservoirs and dams be constructed? Third, how wide and deep should the rivers be excavated?

Consequently, a clear knowledge related to magnitude and frequencies of the flood occurrences are fundamentally needed to deal with all of those questions. A similar circumstance is applied in designing of hydraulic structures and water-related projects such as spillways, culvert, highways, etc. Information regarding accurate estimation of flood magnitudes and their frequency of occurrence are of great importance in the planning, designing and management of such structures at the location or station of interest. The structures need to be designed by considering the maximum flows that exceed certain level in a given return period. On the other hand, the flows below the critical value are less important since they do not negatively affect the structure.

Estimating flood magnitudes and their frequencies need knowledge related to distributions of flood flow series. 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). In this case, flood frequency analysis is a most suitable method in order to determine a robust probability distribution that fit to streamflow data at a certain location of interest. The most widely used methods in predicting the flood magnitudes are at-site and regional flood frequency analyses. In general, flood frequency analysis is defined as an estimation of how often a specified event, in this case, flood will occur (Hosking and Wallis, 1997).
There are two important components in frequency analysis which are parameter estimation method (or estimator) and probability distribution being used for describing flood occurrences. Before a probability distribution can be fitted to the data, parameters of the particular distribution need to be estimated from the data samples. Many estimators have been introduced in estimating distributional parameters, but the most commonly used estimators in hydrology are method of moments (MOM), the maximum likelihood estimation (MLE) and the probability weighted moments (PWMs) or generally known as L-moments. The MOM is a conventional and relatively easy parameter estimation method. Although long established in statistics, this method is not always satisfactory. The MOM estimates are usually inferior in quality and generally not as efficient as the MLE and L-moments especially in the case where the distributions have a large number of parameters. The MLE often regarded as the most efficient method. However, the MLE might be hard to compute and involves numerical algorithm especially if the number of parameters is larger than three. This will in turn make it hard and might also be impossible to obtain MLE of small samples.

During the past three decades, major developments in flood frequency analysis revolved around the idea of probability weighted moments (PWMs) introduced by Greenwood et al. (1979) and the theory of L-moments proposed by Hosking (1986, 1990) as the parameter estimation method. L-moment estimators are an exact analogue to conventional MOM estimators, but are weighted linear sums of the expected order statistics. The L-moment estimates are comparable to the MLE estimates and in certain cases superior than the MLE particularly in small sample sizes. Recent studies on statistical analysis of annual maximum flood series have shown that L-moments provide simple and reasonably efficient estimators of characteristics of hydrologic data and of a distribution’s parameters (Hosking, 1990; Hosking and Wallis, 1993; Stedinger et al., 1992).

The second component in frequency analysis is the probability distributions function being used. In determining a correct distribution function, the concern is to find the one that would be capable of describing the recorded sample and more importantly, extrapolating correctly to large return periods. Many distribution function forms have been proposed for describing flood occurrences. However,
previous studies have been reported in the literature that there is no ‘universal’
distribution function that is best representing floods trend at all streamflow locations
of interest. The extreme value Type I (EV1), II, and III distributions have recently
 gained considerable acceptance for describing annual maximum flows on the
theoretical consideration that the distribution of the maximum of a sample tends to
converge to one of the three extreme value distributions as the sample size increases
(Wang, 1996). Other distributions also have been proposed to be used in flood
frequency analysis including three-parameter distributions; generalized extreme
value type I (GEV), generalized logistic (GLO), generalized Pareto (GPA),
lognormal (LN3), Pearson Type III (P3) and log-Pearson Type III (LP3), and two-
parameter distributions; extreme value Type I (EV1), logistic (LOG), normal and
Pareto distributions.

Flood frequency analysis may suffer from sampling variability when applied
to data for a single site, especially for estimating return periods that exceed the length
of the observed record at a site (Hosking and Wallis, 1993; Cunnane, 1988). The
observed flood data at a particular station are generally insufficient to obtain reliable
estimates of the flood quantiles, especially in developing and undeveloped countries.
This is due to lack of technology and other problems which affect the process of data
collection. In a relatively young country like Malaysia, majority of stations having
data record dated back from 1960 with the average record length of 35 years. This is
inadequate to allow for reliable estimation of flood magnitudes especially for larger
return period than the available length of data record.

One way of providing more reliable estimation is to use several records from
a region with identical behavior of flood, rather than only single site information
(Hussain and Pasha, 2009). This is known as regional flood frequency analysis
(RFFA). In RFFA, estimates at a single site can be enhanced by pooling the data
from other sites which confirmed to have similar frequency distribution. The
information from other sites however, only can be appropriately transferred within a
homogenous region. Studies have shown that, even though a region may be
moderately heterogeneous, regional frequency analysis will still yield much more
accurate quantile estimates than at-site frequency analysis (Lettenmaier et al., 1987).
Recent advance in RFFA involves the use of L-moment estimators as reported by Hosking and Wallis (1997). In RFFA, the objectives are to identify a robust regional distribution for each identified homogenous region and to estimate the quantiles at the station of interest for a given return period. Therefore, the following procedures of RFFA based on L-moments approach have been employed to attain the goals. The procedures include the detection of outliers, identification and verifying of homogenous regions, identification and testing of regional frequency distribution, and estimation of flood quantiles at stations of interest. This methodology has been applied successfully in modeling floods in a number of case studies from Malaysia (Lim and Lye, 2003; Zin et al., 2009; Shabri et al., 2011), New Zealand (Pearson, 1991, 1995; Madsen et al., 1997), South Africa (Kachroo et al., 2000; Kjeldsen et al., 2002), China (Jingyi and Hall, 2004), UK (Fowler and Kilsby, 2003), Pakistan (Hussain and Pasha, 2009; Hussain, 2011), Turkey (Saf, 2009a, 2009b), Iran (Abolverdi and Khalili, 2010) and Italy (Norbiato et al., 2007; Noto and Loggia, 2009).

1.2 Problems Statement

The purpose of analyzing hydrological extreme events such as annual maximum series of floods is, in most cases, to predict magnitude of flood of relatively large return period such as 100 years and above (Wang 1990). Hence, it is actually advantageous to intentionally censor (or eliminate) low-value observations because using only the larger value flood ensures that the extrapolation to large return periods flood is carried out by exploring the trend of these larger flows only. Cunanne (1987) suggested that in such cases a censored sample should be used and the analysis will be based on only those floods whose magnitudes have exceeded a certain threshold.

Since L-moments were first introduced by Hosking (1990) as a parameter estimation method, it has been widely applied in many fields of hydrology. Although L-moments result in quite efficient estimate in parameter estimation, this may not be so for predicting large return period events. The question arose whether L-moments
are oversensitive to the lower part of distributions and give insufficient weight to large data values that actually contain useful information on the upper distribution tail (Wang, 1997; Bobee & Rasmussen, 1995).

Wang (1990) has introduced the L-moments method based on the concept of partial probability weighted moments (PPWMs), which are called partial L-moments (PL-moments) for fitting distribution functions to censored samples. PL-moments are variants of L-moments and also analogous to the PPWMs. PL-moments are introduced for characterizing the upper part of distributions and larger events in data. Using PL-moments reduce undesirable influences that small sample events may have on the estimation of large return period events.

However, there is no further research investigating on flood frequency analysis of censored sample thus far. Hence, this research will provide further investigation and more comprehensive evaluation of censored sample based on PL-moments approach in flood frequency analysis especially on evaluating the performance of PL-moments compared to L-moments.
1.3 Research Objectives

In this research, a comprehensive evaluation of PL-moments estimator in flood frequency analysis will be investigated particularly on evaluating the performance of PL-moments compared to L-moments. The PL-moments at various levels of censoring, $F_0$ will be considered in this research. Five probability distributions with three-parameter; GEV, GLO and GPA distributions and two-parameter; EV1 and LOG distributions will be used in flood frequency analysis of this research.

The main objectives of this research are:

i. To derive the parameters estimation models of PL-moments approach for GLO, GPA, EV1 and LOG distributions and to enhance the parameters estimation for GEV distribution.

ii. To evaluate the sampling properties of PL-moments compared to L-moments in characterizing larger events in sample using Monte Carlo simulation data generated from known and unknown parent distribution function.

iii. To develop the PL-moments approach in regional flood frequency analysis based on L-moments approach in modeling the annual maximum streamflow over stations in Peninsular Malaysia.

iv. To assess performances of PL-moments compared to L-moments in all stages of regional flood frequency analysis.

1.4 Research Scope

This research covers the following aspects:

i. This research covers the derivation of the parameters estimation models for GLO, GPA, EV1 and LOG distributions and to enhance the parameters estimation for GEV distribution based on PL-moments approach. The
parameter estimation models for GEV, GLO, GPA, EV1 and LOG distributions based on L-moments are revisited.

ii. Two types of data are utilized in this study. The first data are synthetic “flood-like” data obtained from Monte Carlo simulation data. The Monte Carlo simulation generates synthetic flows from various background distributions of known and unknown parent distribution function. The GEV, GLO, GPA, EV1 and LOG distributions are assumed as known parent distribution function while six Wakeby distributions are assumed as unknown parent distribution function.

iii. The second data are used in regional flood frequency analysis. Data of annual maximum streamflow over stations located throughout Peninsular Malaysia which ranges from 1960 to 2009 has been used. Records of daily streamflow from 56 stations with record lengths of 15 to 50 years were acquired from the Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia.

iv. PL-moments with various levels of censoring, \( F_0 \) are investigated in Monte Carlo simulation study ranging from \( F_0 = 0.1, 0.2, 0.3, 0.4 \) and 0.5.

1.5 Research Contribution

This research offers several contributions. The main contributions are:

i. This research contributes to the development of the several three-parameter estimation models of GEV, GLO, GPA distributions and two-parameter estimation models of EV1 and LOG distributions based on PL-moments method to be used in application of flood frequency analysis. The PL-moments method is developed as similar to L-moments method in estimating
parameters of various probability distributions for extreme events in hydrology.

ii. Since the choice of censoring values, \( F_0 \) are still under discussion, by investigating PL-moments with censoring levels, \( F_0 \) ranging from 0.1 to 0.5, the readers will have some ideas in choosing the suitable censoring value to improve the estimation of extreme events particularly in high return period events in frequency analysis studies.

iii. This research also contributes to the development of regional flood frequency analysis (RFFA) based on PL-moments approach in each stages of RFFA. These include the process of screening out the data, verifying the homogenous region using statistical measure, selecting suitable regional probability distribution and estimating regional parameters and flood quantiles according to L-moments approach.

iv. The results of this study give benefits to hydrological studies. The direct beneficiaries of the study are the 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. By knowing the information regarding flood magnitudes and corresponding frequencies of occurrence, engineering projects such as dams, spillways, highways, etc can be planned, designed and managed effectively. 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.
1.6 Organization of Thesis

The rest of the thesis is organized as follows:

Chapter 2 reviews main subjects used in the study that includes flood frequency analysis, parameter estimator, simulation of the estimator, censored data in hydrology, regional flood frequency analysis, regional homogeneity measure and selection of regional probability distribution.

Chapter 3 describes in detail the related theories and methodologies for the development of flood frequency analysis. The main ideas behind the building of flood frequency model are also discussed. The background of L-moments and Partial L-moments are defined by explaining their population and sample theories.

Chapter 4 discusses on flood frequency analysis and quantile estimation using probability distribution function. Several distributions, namely GEV, GLO, GPA, EV1 and LOG distributions are considered to be used as possible candidates in this study. The details of each distribution will be presented including their probability distribution function (pdf), cumulative distribution function (cdf) and quantile function. The parameter estimation using the methods of L-moments is revisited and parameter estimation using the methods of PL-moments is derived for each distribution.

Chapter 5 presents the results of Monte Carlo simulation study to investigate the sampling properties of the proposed parameter estimation methods of L-moments and PL-moments. The analyses of the simulations are for the cases of known parent distribution function and unknown parent distribution function.

Chapter 6 develops the procedures of regional flood frequency analysis (RFFA) for PL-moments based on the L-moments approach. The procedures include four stages of RFFA such as screening of the data, identification of homogeneous regions, identification and testing of regional frequency distributions, and estimation of flood quantiles at recurrence intervals of interest.
Chapter 7 presents the analysis of the regional flood frequency analysis (RFFA) based on L-moments and PL-moments for study data of daily streamflow from 56 stations located throughout Peninsular Malaysia. Finally, capabilities of the L-moments and PL-moments estimators in estimation of design flood quantiles are evaluated at specific recurrence intervals.

Chapter 8 summarizes the procedures and analysis in the research, draws some conclusions of the research and provides suggestions and recommendation for future research.
REFERENCES


